Sur quelques résultats d'inférence pour les processus fractionnaires et les processus ponctuels spatiaux de Gibbs

Jean-François Coeurjolly

MÉMOIRE D'HABILITATION

présenté par
Jean-François COEURJOLLY

en vue de l'obtention du diplôme
d'Habilitation à Diriger des Recherches de Université de Grenoble

Spécialité : **Mathématiques Appliquées**

Sur quelques résultats d'inférence pour les processus fractionnaires et les processus ponctuels de Gibbs

Soutenue le 23 Novembre devant le jury composé de

Rapporteurs :	Anne ESTRADE	-	Université Paris Descartes
	Jesper MØLLER	-	Aalborg University
	Eric MOULINES	-	Ecole Nationale Supérieure des Télécommunications
Examinateur :	Patrick FLANDRIN	-	Ecole Normale Supérieure de Lyon
Président :	Anestis ANTONIADIS	-	Université Joseph Fourier, Grenoble

Remerciements

Il est certaines pages qu'on rechigne à écrire ou pour lesquelles on parvient à s'auto-persuader qu'il vaut mieux ne pas les écrire. Ce n'est pas le cas de celle-ci et c'est avec plaisir que je me plie à cet exercice.

En tout premier lieu, je voudrais remercier les trois rapporteurs Anne Estrade, Eric Moulines et Jesper M\o{}ller (mange tak for deres bidrag og frugtbare diskussioner om fri model!). Accepter de rapporter une habilitation me semble une tâche inconcevable tant parfois les thèmes peuvent être variés. Je suis très honoré qu'ils aient accepté dès mon premier contact de rapporter ce travail. Et encore plus flatté qu'ils y aient trouvé un intérêt personnel. De ce fait, j'ai aussi été ravi qu'ils aient pu faire le déplacement pour participer au jury de soutenance. Je suis également très reconnaissant à Patrick Flandrin d'avoir pris part dans ce jury.

J'exprime toute ma gratitude à Anestis Antoniadis pour avoir présidé ce jury. A la fin de mon stage de DEA, je lui avais annoncé que je partais à l'ONERA effectuer une thèse, que je ne comptais pas faire une carrière universitaire et que ce n'était donc pas la peine que je sois classé pour obtenir une bourse du ministère. Je n'avais alors ressenti de sa part aucun signe de désaccord par rapport à ce projet (ni de soulagement d'ailleurs!). Alors que tout fut quasiment signé, la porte de l'ONERA s'est refermée. La gueule enfarinée, j'ai repris contact avec Anestis qui a fait des pieds et des mains (certains diront de la voix) pour me réintégrer dans le classement pour finalement faire une thèse sous sa direction. Cet épisode est très révélateur du personnage qui n'a jamais cessé ses encouragements et sa confiance depuis toutes ses années. Finalement, je me dis qu'une explication possible du fait que nous n'ayons jamais publié ensemble est qu'au moins sa participation au jury ne pouvait être remise en cause. A nouveau, merci à lui.

Une recherche du mot "Drouilhet" dans le document montre 54 occurences (y compris celle-ci)!!! Nous collaborons depuis mon recrutement à l'UPMF et mon arrivée dans son bureau sur de nombreux sujets appliqués et théoriques : enseignement, processus ponctuels, contrat avec notre sponsor psychologue de ces dernières années, co-direction de la thèse du grand Jeff (ni lui ni moi n'avions alors l'Habilitation à Diriger des Relous) ... Je lui dois beaucoup autant d'un point de vue professionnel que personnel. Un profond merci amical à lui.

Merci à tous mes co-auteurs de la Gibbs Team aux plus récents co-auteurs du Gipsa et de l'Est qui rendent ce ""travail"" plus agréable et enrichissant ! Je remercie également Jean-Marc Chassery ainsi que Pierre-Olivier, Sophie, Nicolas,... pour leur accueil au sein du Gipsa dans le cadre de ma délégation et leurs encouragements dans mes démarches. Ce changement d'amtosphère, leur bonne humeur et les conditions de travail ont été très propices pour les nombreuses discussions, collaborations et perspectives de travail et pour avoir pu préparer sereinement cette habilitation.

Merci enfin à ma famille et mes amis pour leur soutien. Merci à eux d'avoir fait le déplacement (merci aussi à certains amis de n'avoir pas fait le déplacement, ça aurait été un sacré bazar et ils m'auraient fini le fût en vingt minutes!). Je n'oublie pas Anne-Claude pour sa présence à mes côtés ; ces dernières semaines où mon niveau d'attention à certaines questions semblait être particulièrement en adéquation avec un processus à faible régularité Höldérienne ont été plus faciles grâce à elle et à mes deux estimateurs presque normaux Martin et Prune. Je remets toujours mon activité en question quand je me rappelle que Martin, à la question qu'est-ce qu'il fait ton papa, répond : "Il va au boulot avec ses copains et il lit ses mails!".

Table des matières

Introduction (en français)

Ce mémoire présente une synthèse de mes activités de recherche depuis mon doctorat. Ces travaux sont organisés en trois parties distinctes. Les deux premières parties ont pour point commun l'*inférence statistique de quelques processus stochastiques*. Les processus centraux en question sont respectivement **le mouvement Brownien fractionnaire** (et quelques unes de ses extensions) et **les processus ponctuels spatiaux de Gibbs**. Comme, nous le verrons par la suite, bien que ces processus soient de nature très différente, ils s'inscrivent dans la modélisation de données dépendantes qu'elles soient temporelles ou spatiales. Nos travaux ont pour objectifs communs d'établir des propriétés asymptotiques de méthodes d'estimation ou de méthodes de validation, classiques ou originales. Par ailleurs, une autre similitude est la mise en perspective de ces processus avec des applications faisant intervenir des *systèmes complexes*. La troisième partie, quant à elle, regroupe des thèmes satellites regroupés sous la dénomination *contributions à la statistique appliquée*.

Pour avoir une lecture plus facile et accentuer les points communs entre les deux premières parties, nous avons adopté le même plan de présentation. Les Chapitres 1 proposent une introduction à chacune de ces parties en présentant des références générales (d'où le fait que cette introduction n'en contienne aucune), les modèles auxquels nous nous sommes intéressés. Les Chapitres 2 développent nos contributions en terme d'inférence (simulation, estimation, validation) tandis que les Chapitres 3 présentent des perspectives de recherche et travaux en cours ainsi que les deux projets de recherche en cours qui s'inscrivent dans la thématique des systèmes complexes.

Dans ce mémoire, nous présentons nos contributions en présentant les résultats principaux et en indiquant les outils sur lesquels ils s'appuient. Nous renvoyons systématiquement aux articles correspondants pour les résultats exhaustifs ainsi que pour les différentes preuves. Enfin précisons qu'un résumé de mon activité scientifique (publications, encadrement, projets de recherche, animation) est diponible à la fin de ce mémoire, en page 551).

Partie I : Autour du mouvement Brownien fractionnaire et quelques unes de ses extensions

La Partie I est donc dédiée à l'étude de quelques processus fractionnaires. Ces travaux, initiés au cours de mon doctorat, ont ensuite été développés et étoffés. Ils ont débuté avec l'étude du mouvement Brownien fractionnaire, processus à accroissements stationnaires et autosimilaires, caractérisé notamment par l'exposant $H \in (0,1)$, dit exposant de Hurst, paramétrisant sa fonction de covariance. Cet exposant caractérise non seulement la dimension fractale du processus gaussien mais également sa régularité, le caractère longue mémoire de ses accroissements et l'autosimilarité du processus. Ce processus permet notamment de modéliser des signaux temporels très irréguliers possédant des propriétés de fractalité et trouve des applications dans de nombreux domaines comme la turbulence, les neurosciences ...

Pour ce processus, nous avons réalisé une large étude bibliographique des problèmes de simulation de ses trajectoires et de l'estimation du paramètre de Hurst. Nous avons également étendu (d'un point de vue méthodologique et théorique) la méthode des variations discrètes (méthode de moments, d'ordre deux, basés sur des versions filtrées-dilatées d'une trajectoire d'un mouvement

brownien fractionnaire) et montré son optimalité (en termes de vitesse de convergence) vis-à-vis de l'estimateur du maximum de vraisemblance pour lequel nous avons retrouvé la vitesse de convergence en démontrant un résultat d'algèbre original sur l'inverse de matrices localisées. Faisant suite à ce travail, nous avons défini et étudié une nouvelle méthodologie d'estimation de l'exposant de Hurst robuste à des données aberrantes. Cette procédure consiste (principalement) à remplacer l'estimateur empirique des moments par un quantile empirique ou une moyenne tronquée. La difficulté de ce travail a résidé essentiellement dans l'obtention de représentation de type Bahadur de quantiles empiriques pour des processus Gaussien subordonnés dont la fonction de corrélation decroît hyperboliquement. Récemment, nous avons également exploité la méthode des variations discrètes pour définir une méthodologie robuste à un bruit gaussien additif. Toutes ces procédures ont été implémentées sous la forme de packages `R`. Enfin, dans un travail actuellement soumis, nous avons tenté une approche non asymptotique pour proposer un intervalle de confiance du paramètre H. Ces nouveaux intervalles utilisent des inégalités de concentration pour des intégrales stochastiques de processus gaussiens récemment obtenues et que nous avons améliorées (ou plutôt optimisées) et exploitées.

Nous nous sommes également intéressés à deux extensions du mouvement brownien fractionnaire : le mouvement Brownien multifractionnaire et le mouvement Brownien fractionnaire multivarié. Le premier est une extension bien connue maintenant consistant essentiellement à autoriser la régularité des trajectoires à pouvoir évoluer avec le temps. L'exposant de Hurst devient maintenant une fonction du temps. Ce processus n'est plus que localement asymptotiquement autosimilaire. Nous avons alors exploité localement la méthodologie standard et obtenu des résultats partiels [1]. La seconde extension définit un processus autosimilaire multivarié. L'introduction de ce processus est très récente et extrêmement prometteuse dans le cadre d'applications en neurosciences. Nous apportons plusieurs contributions : conditions d'existence, calcul de la densité spectrale, propriétés de type longue mémoire des accroissements, transformée en ondelettes, simulation...

Les travaux de cette partie ont été réalisés en collaboration avec Pierre-Olivier Amblard, Sophie Achard, Jean-Christophe Breton, Jacques Istas, Frédéric Lavancier et Anne Philippe.

Partie II : Inférence pour les processus ponctuels spatiaux de Gibbs

A la suite de mon doctorat, j'ai été recruté à l'Université Pierre Mendès-France et intégré le LABSAD où j'y ai rencrontré entre autres Etienne Bertin, Jean-Michel Billiot et Rémy Drouilhet. C'est à leur contact que je me suis intéressé, après quelques années d'exercice, au sujet qui les motivait (d'un point de vue théorique) depuis quelques années déjà, à savoir les processus ponctuels. Un processus ponctuel est une collection aléatoire de points localement finie dans l'espace. Il est dit marqué si à chaque point de l'espace est associée une marque aléatoire. Le domaine des processus ponctuels constitue un domaine à part entière à l'interface de la physique statistique, des probabilités et de la statistique. Le processus ponctuel de référence est le processus de Poisson qui permet de modéliser une configuration de points aléatoire dans l'espace sans aucune interaction entre ces points. Une manière d'introduire de la dépendance est de considérer la classe des modèles de Gibbs. De manière assez simple, dans un domaine borné de $\mathbb{R}^d$, un processus de Gibbs est défini via sa mesure de probabilité dont la densité par rapport à la mesure de Poisson est proportionnelle à $e^{-V(\varphi)}$, où $V(\varphi)$ correspond à la fonction énergie (appelée également dans la littérature Hamiltonien ou fonction d'interaction) d'une configuration de points φ. Mes trois collègues avaient travaillé sur les problèmes d'existence de modèles Gibbsiens dans $\mathbb{R}^d$ originaux par rapport à la littérature classique (interactions de paires mesurés sur des graphes structurés tels que le graphe de Delaunay ou le graphe des k-plus proches voisins, interaction mesurée à partir de caractéristiques géométriques comme des cellules de Voronoï). La première collaboration est née du souhait de

1. nous avons obtenu des résultats asymptotiques dont une partie contient un certain nombre d'erreurs. Voir la Section 2.3 pour plus de détails

pouvoir inférer sur les modèles qu'ils proposaient. Le premier constat a été que les travaux existants sur l'obtention de propriétés asymptotiques de méthodes d'estimation paramétrique (de la fonction énergie V) ou de méthodes de validation de modèles étaient peu nombreux et ne permettaient pas d'inclure ces modèles. Nos différentes contributions vont dans ce sens. Sur le problème d'estimation, nous avons au cours de deux travaux fourni des résultats concernant la méthode du maximum de la pseudo-vraisemblance (alternative à la vraisemblance qui pose de réelles difficultés pour les modèles Gibbsiens dû au calcul prohibitif de la constante de normalisation définissant ces modèles). Ces résultats incluent nos modèles d'intérêt, de nombreux modèles classiques possédant la propriété d'être localement stable et de portée finie ainsi que des modèles appartenant à la classe des potentiels superstables (voir Chapitre 1 pour plus de détails sur ces termes). Au cours d'une autre collaboration, nous nous sommes également intéressés à la méthode de Takacs-Fiksel (dont la pseudo-vraisemblance est un cas particulier) basée sur l'équation d'équilibre de Campbell. Cette méthode pour laquelle nous avons fourni des propriétés asymptotiques est très riche et intéressante ; elle permet dans certaines situations de définir un estimateur rapidement calculable. Nous lui trouvons également un intérêt pour le modèle Quermass, modèle Gibbsien où les points ne sont pas observés ; la fonction énergie de ce processus dépend uniquement des caractéristiques géométriques - périmètre, volume, caractéristique d'Euler-Poincaré - d'un certain ensemble aléatoire. L'équation d'équilibre de Campbell est à l'origine de l'introduction de la notion de résidus pour les processus ponctuels. Nous avons obtenu des résultats asymptotiques pour le processus des résidus tout à fait intéressants qui nous permettent de pouvoir dériver des tests d'adéquation pour des modèles Gibbsiens. Soulignons que nos travaux ont nécessité l'obtention d'un TCL multivarié dans un contexte non-stationnaire et pour des tableaux triangulaires pour des fonctionelles de champs markoviens possédant une certaine propriété de centrage conditionnel que nous présentons. Une partie des résultats obtenus a trouvé une application en informatique graphique pour capturer l'arrangement spatial d'objets géométriques dessinés par un utilisateur.

Les travaux théoriques de cette Partie II ont été réalisés en collaboration avec Jean-Michel Billiot, David Dereudre, Rémy Drouilhet et Frédéric Lavancier.

Lien avec les Systèmes Complexes

Depuis Septembre 2009, je suis en délégation CNRS au sein du GIPSA-lab, où j'y ai côtoyé des chercheurs structurés autour de cette notion de système complexe. Et c'est une volonté de ma part que d'essayer d'inscrire désormais mon activité dans ce cadre. De nos jours, plusieurs travaux s'accordent à définir un *système complexe* comme un système composé d'un grand nombre d'entités en interaction, présentant des phénomènes d'auto-organisation et d'où émerge des propriétés macroscopiques. L'étude d'un système complexe requiert nécessairement plusieurs compétences issus de domaines pluridisciplinaires et mon objectif est de m'inscrire dans un axe modélisation et inférence dans le cadre de projets de recherche en cours.

Ainsi mes travaux de la Partie I peuvent être utilisés pour mesurer la connectivité et la fractalité de signaux d'Imagerie par Résonance Magnétique fonctionelle (IRMf) issus de l'organisation des neurones. Les processus ponctuels de Gibbs, quant à eux, permettent de modéliser des systèmes de particules (ou plus généralement des objets géométriques) en interaction. La configuration de particules observée est celle qui minimise une fonction énergie engendrant ainsi des phénomènes d'auto-organisation. Par essence, ces modèles permettent donc de modéliser des systèmes complexes. Les récents modèles étudiés (modèle Quermass, mosaïque de Voronoï) pourraient être utilisés pour modéliser l'organisation des cellules, de fluides complexes ou l'interface entre deux matériaux. Par ailleurs, nous souhaitons également utiliser ce type d'outils pour modéliser une propriété (reconnue par certains comme émergente) de l'activité solaire à savoir le phénomène des taches solaires. Ce projet pourrait au final être un pont entre les deux premières parties de ce mémoire car les données consistent en un processus ponctuel spatio-temporel et de récents

travaux tendent à montrer un certain caractère longue mémoire de la série chronologique associée au nombre de taches observées. Ces aspects seront développés dans les Chapitres 3 de chacune des Parties I et II.

Partie III : Contributions à la statistique appliquée

Cette partie regroupe des thèmes plutôt indépendants les uns des autres ayant pour dénominateur commun la statistique appliquée. Dans mes premières années d'exercice de la fonction de maître de conférences dans une UFR d'Economie, j'ai travaillé en étroite collaboration avec Rémy Drouilhet pour tenter de développer une nouvelle approche (approche expérimentale des probabilités) et de nouveaux outils (outils logiciels, système de notations) pour faire passer des notions tels que les tests d'hypothèses simples dans un cadre non nécessairement gaussien, à un public peu enclin aux langage et techniques mathématiques. Quelques unes de ces contributions sont présentées dans la Section1.1. J'ai également participé à deux co-encadrements de thèse (de JF Robineau et M Nguile Makao). Le premier a réalisé un travail sur la sélection de variables et la théorie de l'information et le second est issu du domaine de la biostatistique et de l'épidémiologie et a pour objectif la modélisation de la pneumonie nosocomiale dans le service de réanimation. Les productions résultant de ce travail sont détaillées dans les Sections 1.2 et 1.3. Enfin, ma position dans une université de sciences sociales m'a amené à collaborer avec des économistes et psychologues. Cette activité est détaillée dans la Section 1.4.

Introduction (in english)

This document is a synthesis of my research activity since my PhD. These works are organized in three different parts. The first and second parts are dedicated to the *statistical inference of stochastic processes*. The main processes we study are the **fractional Brownian motion** and **spatial Gibbs point processes**. The nature of these processes is very different but both are used to model data with strong (temporal or spatial) dependence. My contributions have for common points the study of asymptotic properties of estimation and validation methods. Another common point is the perspective to use these processes in applications dealing with *complex systems*. Finally, the third part of this document gathers some works under the name *contributions to applied statistics*.

In order to facilitate the reading process and to underline the common points between Parts I and II, we adopt the same plan. Chapters 1 provide an introduction with general references and the different models we study. Chapters 2 develop our contributions in terms of inference (simulation, estimation, validation) whereas Chapters 3 present research perspectives, working papers and research projects in the field of complex systems.

In this document, we present our contributions by providing the main results and the different required tools. We refer the reader to the associated papers for the details on exhaustive results and the proofs. Finally, let us underline that a summary of my research activity (publications, supervision of PhD students, research projects) is available at the end of this document, page 551.

Partie I : Around the fractional Brownian motion and some of its extensions

Part I is devoted to the study of fractional processes. These works initiated during my PhD have then been improved and extended. They began with the study of the fractional Brownian motion. This stochastic process is a centered self-similar Gaussian process with stationary increments, notably characterized by the Hurst exponent $H \in (0,1)$ parameterizing its covariance function. This exponent is related to the fractal dimension, the regularity of the process, the long-memory property of its increments and the self-similarity. In particular, this process is commonly used to model irregular time series exhibiting fractal properties and finds applications in several domains such as turbulence, neurosciences,...

Concerning this process, we have proposed a large bibliographical study of simulation and estimation (of the parameter H) problems. We have also improved (from a methodological and theoretical point of view) the discrete variations method (which is a variogram method based on filtered-dilated version of a discretized sample path of a fractional Brownian motion) and proved its optimality (in terms of rate of convergence) with respect to the maximum likelihood estimator for which we have retrieved its rate of convergence by establishing an algebra result concerning the inverse of localized matrices. Following this work, we have defined and studied a new estimation procedure which is robust to outliers. This procedure mainly consists in replacing the empirical variance with a sample quantile or a trimmed-mean in the previous procedure. The key-ingredient of this work is the obtention of a Bahadur type representation of sample quantiles for subordinated

Gaussian stationary processes with correlation function decaying as $|i|^{-\alpha}$. Recently, we have also exploited the discrete variations method to propose a procedure which is robust to an additive noise. The whole of these procedures are implemented in `R` packages. Finally, in a submitted work, we attempt to use a non asymptotic approach to propose a confidence interval for the parameter H. These new intervals make use of concentration inequalities for stochastic integrals of Gaussian processes which have recently been established and that we have improved (actually optimized) and exploited.

We are also interested in two extensions of the fractional Brownian motion : the multifractional Brownian motion and the multivariate fractional Brownian motion. The first one is now a well-known extension which mainly consists in allowing the regularity to vary with time. The Hurst exponent now becomes a function of time. This process is only locally asymptotically self-similar. We therefore exploit locally the standard method and obtain partial results [2]. The second extension deals with a multivariate self-similar process. The introduction of this process is very recent and extremely promising for possible applications in neurosciences. We propose several contributions : existence conditions, computations of the spectral density, long-memory type properties, wavelet transform,...

The results in this part consitute joint works with Pierre-Olivier Amblard, Sophie Achard, Jean-Christophe Breton, Jacques Istas, Frédéric Lavancier and Anne Philippe.

Partie II : Inference for spatial Gibbs point processes

In 2001, I obtained a position of lecturer at the University Pierre Mendès-France. I have joined the LABSAD where I met Etienne Bertin, Jean-Michel Billiot and Rémy Drouilhet. Through contact with them, I started (after several years) to take an active interest in the topic that motivated them from a theoretical point of view, that is the domain of spatial Gibbs point processes. A point process is a locally finite set of points in space. The reference process is the Poisson process which allows to model a random configuration of points without any interaction between points. A way to introduce dependence is to consider the class of Gibbs points processes. In a bounded domain of $\mathbb{R}^d$, a Gibbs point process is defined through its probability measure whose density with respect to the Poisson measure is proportional to $e^{-V(\varphi)}$, where $V(\varphi)$ corresponds to the energy function (also called hamiltonian or interaction function in the literature) of a configuration of points φ. My three colleagues worked a lot on conditions of existence of original Gibbsian models (*e.g.* pairwise interaction point processes with interaction measured on the Delaunay or the k-nearest neighbour graph, interactions based on Voronoï cells). The first collaboration takes its origin in the objective to make inference on such models. As we have observed that existing results could not take into account such interactions, our contributions aim at filling this gap. Concerning the problem of estimating a parametric energy function, we have provided general results for the maximum pseudo-likelihood procedure (alternative to the MLE that avoids the computation of the normalizing constant). These results include our models of interest, a large class of classical models having the property of local stability and finite range and also superstable potentials (see Chapter 1 for more details on these concepts). Then, we focus on the Takacs-Fiksel method (which is a general method including the pseudo-likelihood method) based on the Campbell equilibrium equation. This method, for which we have also proposed asymptotic results, is very rich and interesting ; in certain situations, it allows us to propose a very quick estimator. We also find an application of this procedure for the Quermass model which is a Gibbsian model for which the points are not observed (the energy function of this process is based on geometric characteristics - perimeter, volume, Euler-Poincaré characteristic - of a certain random set.
The Campbell equilibrium equation is also the basis for the (very recent) concept of residuals for

2. some asymptotic results have been obtained but some of them contain mistakes. See Section 2.3 for more details

spatial point processes. We obtain promising asymptotic results for the residuals process which allow us to derive goodness-of-fit tests for spatial Gibbs point processes. Let us underline that some of these works require the obtention of a multivariate CLT for triangular arrays of functional of Markov random fields that are conditionally centered that we present at the end of Chapter 2.

The papers of this part have been written in collaboration with Jean-Michel Billiot, David Dereudre, Rémy Drouilhet and Frédéric Lavancier.

Relations with Complex Systems

Since September 2009, I am in secondment at the GIPSA laboratory. In particular, I have met several searchers working around the notion of complex systems. My future objective is to include my research activity in this framework. A *complex system* is a system composed of interconnected elements that as a whole exhibit one or more properties and from which may emerge macroscopic properties. The study of a complex system naturally requires various skills and my objective is to work in the field of modelling and inference via current research perspectives.

My contributions from Part I can find interesting issues in measuring the connectivity and the fractality of functional Magnetic Resonance Imaging (fMRI) time series coming from the organization of neurons.
By definition, spatial Gibbs point processes are suitable for complex systems. Indeed, these processes can model a system of particles (or more generally geometric objects) in interaction. The Voronoï tessellation and the Quermass model could be used to model the organization of cells, complex fluids or materials interface. Furthermore, we aim at exploiting spatial point processes for modelling the phenomenon of sunspots which is known to be representative of the solar activity. This project could be a bridge between the two first parts of this document. The data consist in a spatio-temporal point process and recent works have shown a long-memory type property of the time series related to the number of sunspots. These aspects are developped in Chapters 3 of Parts I and II.

Partie III : Contributions to applied statistics

During my first years as a teacher of statistics in the department of Economy, I have developped, with Rémy Drouilhet, an experimental approach and new tools (software tools, system of notation) in order to propose to a non mathematician public and in a simple way, notions such as a confidence interval, a hypothesis test in a non necessarily Gaussian framework. Some of these contributions are presented in Section1.1.

I have co-supervised two students (J.-F. Robineau et M. Nguile Makao). The first one has accomplished a work on the problem of variables selection using information theory and the second one aims at modelling the nosocomial pneumony in internal care units using multi-state models. These themas are described in Sections 1.2 and 1.3. Finally, my position in an university of social sciences, has lead me to collaborate with economists and psychologists. This activity is detailed in Section 1.4.

Première partie

Inférence pour le mouvement Brownien fractionnaire et quelques unes de ses extensions

Chapitre 1

Principaux modèles considérés

Sommaire

En termes d'introduction à cette partie, ce chapitre se propose de détailler succinctement les principaux modèles auxquels nous nous sommes intéressés durant la thèse et à la suite de celle-ci : mouvement Brownien fractionnaire (FBM), mouvement Brownien multifractionnaire (MBM) et mouvement Brownien fractionnaire multivarié (MFBM). Les deux premiers modèles sont maintenant bien connus dans la littérature et les Sections 1.1 et 1.2 rappellent simplement et brièvement leurs principales caractéristiques. Le MFBM, extension au cas multivarié, est beaucoup plus récent et la Section 1.3 résume quelques propriétés obtenues dans [CAA10] et [ACLP10].

1.1 Mouvement Brownien fractionnaire (FBM)

Les définitions développées ici peuvent être retouvées dans le travail de Mandelbrot et Van Ness [84] et dans les livres [106] et [45]. Le FBM est défini comme l'unique processus Gaussien (noté X) centré, nul à l'origine, à accroissements stationnaires et autosimilaires. L'autosimilarité est entendue ici comme l'égalité en lois finies-dimensionnelles (f.d.) suivante : pour $\lambda > 0$ et $H \in (0,1)$ (appelé paramètre de Hurst), $X(\lambda t) \overset{f.d.}{=} \lambda^H X(t)$. Ce processus gaussien paramétré par $(H, \sigma^2) \in (0,1) \times \mathbb{R}^+$ est entièrement déterminé par sa fonction de covariance :

$$EX(s)X(t) = \frac{\sigma^2}{2}\left(|s|^{2H} + |t|^{2H} - |t-s|^{2H}\right).$$

Rappelons que lorsque $H = 1/2$, le processus s'identifie à un mouvement Brownien standard. La fonction de covariance des accroissements d'un FBM (appelé bruit gaussien fractionnaire) de taille 1 s'écrit alors

$$\begin{aligned}\gamma(h) &= E\Delta X(t)\Delta X(t+h)\\ &= \frac{\sigma^2}{2}\left(|h-1|^{2H} - 2|h|^{2H} + |h+1|^{2H}\right)\\ &\sim \sigma^2 H(2H-1)|h|^{2H-2}, \text{ lorsque } |h| \to +\infty.\end{aligned}$$

La dernière équation, valable pour $H \neq 1/2$, montre que les accroissements discrétisés constituent un processus à courte mémoire lorsque $H < 1/2$ (i.e. de covariance sommable) et à longue mémoire lorsque $H > 1/2$. Notons enfin, que le paramètre H gouverne également la régularité des trajectoires puisque la dimension de Hausdorff du FBM est presque sûrement égale à $2 - H$.

De part ces différentes caractéristiques, ce processus a été utilisé dans de nombreux domaines tels que l'hydrologie [92], la biologie [29], l'économie [62], en analyse du traffic réseau [118].

Terminons cette section, en précisant que le FBM admet les représentations stochastiques intégrales suivantes (e.g. [106]) :

$$X(t) = \sigma\sqrt{K(H)}\int_{\mathbb{R}} k_H(s,t)B(ds) \tag{1.1}$$

$$\widetilde{X}(t) = \sqrt{\widetilde{K}(H)}\int_{\mathbb{R}} \frac{e^{itx}-1}{|x|^{H+1/2}}\widetilde{B}(dx), \tag{1.2}$$

où B et $\widetilde{B}$ sont deux mouvements Browniens standard, où pour $H \in (0,1)$, $K(H) = \frac{\Gamma(2H+1)\sin(\pi H)}{\Gamma(H+1/2)^2}$, $\widetilde{K}(H) = \frac{H\Gamma(2H)\sin(\pi H)}{\pi}$ et où $k_H(s,t) = (t-s)_+^{H-1/2} - (-s)_+^{H-1/2}$ (pour $u \in \mathbb{R}, (u)_+ = \max(0,u)$). Ces représentations appelées classiquement réprésentation moyenne mobile et spectrale sont égales en distribution.

Un certain nombre de problèmes et de résultats obtenus (en particulier l'estimation d'exposants de Hurst) sont valables pour une classe plus large que le FBM. Il s'agit de la classe des processus Gaussiens localement autosimilaires en 0 : processus gaussiens à accroissements stationnaires et dont la fonction variance satisfait

$$EX(t)^2 = \sigma^2|t|^{2H}(1+r(t)),$$

avec $r(t) \to 0$ lorsque $|t| \to 0$.

1.2 Mouvement Brownien multifractionnaire (MBM)

De nombreux travaux ont montré la nécessité d'étendre le FBM de manière à relâcher la contrainte de stationnarité des accroissements, à autoriser la régularité des trajectoires à évoluer avec le temps. Deux approches naturelles ont été originellement proposées pour étendre le FBM en ce sens consistant à remplacer les paramètres H et σ par deux fonctions du temps ([99], [15]). Ces versions (et d'autres) sont largement discutées dans le papier très riche de Stoev et Taqqu [110]. Nous présentons ci-après la version moyenne mobile : soient H et σ deux fonctions höldériennes d'ordre $\eta > 0$ sur $(0,1)$ (et vérifiant $0 < \inf H(t) \leq \sup H(t) < \min(1,\eta)$) et $\mathbb{R}^+$ respectivement

$$X(t) = \sigma(t)\sqrt{K(H(t))}\int_{\mathbb{R}} k_{H(t)}(s,t)B(ds).$$

La fonction de covariance du MBM est donée par

$$EX(s)X(t) = \frac{\sigma(s)\sigma(t)}{2}\frac{\sqrt{K(H(s))K(H(t))}}{K(H_{s,t})}\left(|s|^{H_{s,t}} + |t|^{H_{s,t}} - |t-s|^{H_{s,t}}\right),$$

où $H_{s,t} = H(s) + H(t)$. Les propriétés de stationnarité des accroissements et d'autosimilarité ne sont vérifiées qu'asymptotiquement localement : en effet dans [15], il est prouvé que

$$\lim_{\varepsilon\to 0^+}\left(\frac{X(t+\varepsilon u) - X(t)}{\varepsilon^{2H(t)}}\right)_{u\in\mathbb{R}^+} \overset{d}{=} \sigma(t)\left(B_{H(t)}(u)\right)_{u\in\mathbb{R}^+}. \tag{1.3}$$

1.3 Mouvement Brownien fractionnaire multivarié (MFBM)

Le recueil de données temporelles multivariées intervient dans de nombreuses applications (neurosciences, économie, sociologie, physique,...) et de telles séries exhibent souvent des propriétés spécifiques telles que la fractalité, la longue dépendance, l'autosimilarité. En ce sens, il paraît pertinent de s'intéresser à une alternative au MBF dans un cadre multivarié. Un mouvement Brownien

fractionnaire multivarié (MFBM) est un processus nul à l'origine, gaussien, à accroissements stationnaires et autosimilaires. L'autosimilarité est ici entendue au sens vectoriel

$$(X_1(\lambda t),\dots,X_p(\lambda t))_{t\in\mathbb{R}} \stackrel{f.d.}{=} \left(\lambda^{H_1}X_1(t),\dots,\lambda^{H_p}X_p(t)\right)_{t\in\mathbb{R}}.$$

Cette définition peut être vue comme un cas particulier des opérateurs autosimilaires [41]. Résumons ici quelques propriétés déjà établies dans [41, 78] et d'autres obtenues dans [CAA10, ACLP10] et [CAA10a]. Les différentes contributions (d'un point de vue chronologique) sont précisées à la fin de ce paragraphe.

Proposition 1 *Les covariances croisées d'un MFBM possèdent nécessairement la représentation suivante pour tout $(i,j)\in\{1,\dots,p\}^2$, $i\neq j$,*

1. *Si $H_i+H_j\neq 1$, il existe $(\rho_{i,j},\eta_{i,j})\in[-1,1]\times\mathbb{R}$ avec $\rho_{i,j}=\rho_{j,i}=\mathrm{corr}(X_i(1),X_j(1))$ et $\eta_{i,j}=-\eta_{j,i}$ tels que*

$$\begin{aligned}EX_i(s)X_j(t) \;=\; &\frac{\sigma_i\sigma_j}{2}\left\{(\rho_{i,j}+\eta_{i,j}\mathrm{sign}(s))|s|^{H_i+H_j}+(\rho_{i,j}-\eta_{i,j}\mathrm{sign}(t))|t|^{H_i+H_j}\right.\\ &\left.-(\rho_{i,j}-\eta_{i,j}\mathrm{sign}(t-s))|t-s|^{H_i+H_j}\right\}. \quad (1.4)\end{aligned}$$

2. *Si $H_i+H_j=1$, il existe $(\tilde\rho_{i,j},\tilde\eta_{i,j})\in[-1,1]\times\mathbb{R}$ avec $\tilde\rho_{i,j}=\tilde\rho_{j,i}=\mathrm{corr}(X_i(1),X_j(1))$ et $\tilde\eta_{i,j}=-\tilde\eta_{j,i}$ tels que*

$$EX_i(s)X_j(t) \;=\; \frac{\sigma_i\sigma_j}{2}\left\{\tilde\rho_{i,j}(|s|+|t|-|s-t|)+\tilde\eta_{i,j}(t\log|t|-s\log|s|-(t-s)\log|t-s|)\right\}. \quad (1.5)$$

Les paramètres $\rho_{i,j}$ et $\widetilde{\rho}_{i,j}$ sont obtenus en prenant $s=t=1$ et peuvent donc être interprétés comme des corrélations instantanées. Les paramètres $\eta_{i,j}$ et $\widetilde{\eta}_{i,j}$ sont plus difficilement interprétables (mais peuvent être liés par exemple à $\rho_{i,j}$ et $\widetilde{\rho}_{i,j}$ et $EX_i(1)X_j(-1)$). La fonction de covariance croisée des accroissements (de taille 1) d'un MFBM, notée $\gamma_{i,j}(h):=E\Delta X_i(t)\Delta X_j(t+h)$, se déduit facilement de la précédente proposition :

$$\gamma_{i,j}(h)=\frac{\sigma_i\sigma_j}{2}\Big(w_{i,j}(h-1)-2w_{i,j}(h)+w_{i,j}(h+1)\Big). \quad (1.6)$$

où la fonction $w_{i,j}(\cdot)$ est donnée par

$$w_{i,j}(h)=\begin{cases}(\rho_{i,j}-\eta_{i,j}\mathrm{sign}(h))|h|^{H_i+H_j} & \text{si } H_i+H_j\neq 1,\\ \tilde\rho_{i,j}|h|+\tilde\eta_{i,j}h\log|h| & \text{si } H_i+H_j=1.\end{cases} \quad (1.7)$$

Notons que les conditions sur les différents coefficients de la Proposition 1 sont des conditions nécessaires pour définir une fonction de covariance. Pour établir une condition nécessaire et suffisante, nous nous sommes basés dans [ACLP10, CAA10] sur les représentations stochastiques intégrales, en particulier la représentation spectrale d'un MFBM obtenue par Didier et Pipiras [41], et sur le calcul de la densité spectrale des accroissements d'un MFBM de fonction de covariance donnée par la Proposition 1.

Théorème 2 (Didier and Pipiras, [41]) *Soit $(X(t))_{t\in\mathbb{R}}$ un MFBM de paramètre $(H_1,\cdots,H_p)\in(0,1)^p$, alors il existe une matrice (p,p) complexe A telle que chaque composante admette la représentation stochastique intégrale suivante*

$$X_i(t)=\sum_{j=1}^{p}\int\frac{e^{\mathbf{i}tx}-1}{\mathbf{i}x}(A_{ik}x_+^{-H_i+1/2}+\bar A_{ik}x_-^{-H_i+1/2})\tilde B_j(dx), \quad (1.8)$$

où pour tout $j=1,\dots,p$, $\tilde B_j$ est une mesure Gaussienne complexe telle que $\tilde B_j=\tilde B_{j,1}+\mathbf{i}\tilde B_{j,2}$ avec $\tilde B_{j,1}(x)=\tilde B_{j,1}(-x)$, $\tilde B_{j,2}(x)=-\tilde B_{j,2}(x)$, $\tilde B_{j,1}$ et $\tilde B_{j,2}$ sont indépendantes et $E(\tilde B_{j,i}(dx)\tilde B_{j,i}(dx)')=dx$, $i=1,2$. Réciproquement tout processus satisfaisant (1.8) est un MFBM.

Soulignons que [41] proposent également une représentation de type moyenne mobile. Celle-ci n'est pas aussi intéressante que la version spectrale car elle souffre d'un problème de construction lorsque $H_i = 1/2$.

Proposition 3 *Soit $S_{i,j}(\cdot)$ la densité spectrale croisée des accroissements (de taille 1) des composantes i et j d'un MFBM (voir [ACLP10] pour les conventions prises pour la transformée de Fourier)*

$$S_{i,j}(\omega) = \frac{\sigma_i \sigma_j}{\pi} \Gamma(H_i + H_j + 1) \frac{1 - \cos(\omega)}{|\omega|^{H_i + H_j + 1}} \times \tau_{i,j}(\mathrm{sign}(\omega)), \tag{1.9}$$

où

$$\tau_{i,j}(\mathrm{sign}(\omega)) = \begin{cases} \rho_{i,j} \sin\left(\frac{\pi}{2}(H_i + H_j)\right) - \mathbf{i}\eta_{i,j}\mathrm{sign}(\omega) \cos\left(\frac{\pi}{2}(H_i + H_j)\right) & \text{si } H_i + H_j \neq 1, \\ \tilde{\rho}_{i,j} - \mathbf{i}\frac{\pi}{2}\tilde{\eta}_{i,j}\mathrm{sign}(\omega) & \text{si } H_i + H_j = 1. \end{cases} \tag{1.10}$$

En calculant la densité spectrale des accroissements à partir de la représentation spectrale, i.e. (1.8), nous avons alors obtenu la condition nécessaire et suffisante suivante.

Proposition 4 *La matrice $\Sigma(s,t) = (EX_i(s)X_j(t))$ est une matrice de covariance si et seulement si la matrice Hermitienne $Q = (\Gamma(H_i + H_j + 1)\tau_{i,j}(1))$ avec $\tau_{i,j}$ defini par (1.10), est positive.*

Lorsque $p = 2$, une caractérisation graphique s'en déduit. L'ensemble des valeurs possibles de (ρ, η) correspond à un domaine décrit par une ellipse dont l'équation dépend évidemment des paramètres H_1, H_2 (cf [ACLP10]). Cette contrainte énoncée par la Proposition 4 est relativement contraignante, car elle précise entre autres qu'il n'est pas possible (en général) de corréler comme on le souhaite deux FBMs. Lorsque tous les paramètres H_i sont égaux, nous montrons qu'il n'y a pas de contrainte sur les paramètres mais à titre d'exemple lorsque $p = 2$, $H_1 = 0.1$ et $H_2 = 0.8$, le coefficient $|\rho|$ ne peut excéder 0.514.

Le MFBM possède plusieurs propriétés remarquables. Par exemple, nous avons montré qu'il apparaît naturellement (tout comme le FBM) comme la limite de sommes partielles de processus linéaires. Notons aussi que la covariance croisée des accroissements (ainsi que la densité spectrale croisée) possède des caractéristiques de type longue mémoire, liées aux coefficients H_i et H_j comme le souligne le résultat suivant.

Proposition 5 *Lorsque $|h| \to +\infty$, nous avons*

$$\gamma_{i,j}(h) \sim \sigma_i \sigma_j \delta^2 |h|^{H_i + H_j - 2} \kappa_{i,j}(\mathrm{sign}(h)), \tag{1.11}$$

avec

$$\kappa_{i,j}(\mathrm{sign}(h)) = \begin{cases} (\rho_{i,j} - \eta_{i,j}\mathrm{sign}(h))(H_i + H_j)(H_i + H_j - 1) & \text{si } H_i + H_j \neq 1, \\ \tilde{\eta}_{i,j}\mathrm{sign}(h) & \text{si } H_i + H_j = 1. \end{cases} \tag{1.12}$$

Bien évidemment lorsque $\rho_{i,j} = \eta_{i,j} = 0$ (ou $\tilde{\rho}_{i,j} = \tilde{\eta}_{i,j} = 0$), les composantes X_i et X_j sont indépendantes. Dans les autres situations, le résultat (1.11) amène les remarques suivantes : si $H_i + H_j \neq 1$, la covariance croisée est absolument sommable ssi $H_i + H_j < 1$ (en particulier elle peut l'être avec une composante à courte mémoire et la seconde à longue mémoire). Si $H_i + H_j = 1$, $H_i \neq 1/2$, la covariance croisée n'est jamais sommable.

Terminons ce paragraphe en tentant de définir nos différentes contributions : dans [41], Didier et Pipiras étudient le mouvement Brownien fractionnaire-opérateur au travers de représentations moyenne mobile et spectrale. Dans [78], les auteurs partent de la définition d'un processus multivarié autosimilaire (non nécessairement gaussien) et montrent que la fonction de covariance est

décrite, à une reparamétrisation près, par la Proposition 1. Par ailleurs partant de la représentation moyenne mobile obtenue par [41], Lavancier et al. explicitent les différents paramètres du modèle. Dans [CAA10],[CAA10a], nous nous sommes focalisés sur deux cas particuliers de MFBM issus eux aussi de la représentation moyenne mobile : le cas causal (extension naturelle de (1.1) au cas multivarié) et le cas "well-balanced" (obtenue en prenant $A = \overline{A}$ dans le Théorème 2). Nous étudions les propriétés des accroissements : calcul des fonctions de covariance croisées, de la matrice de densité spectrale, comportements limites. Par ailleurs, nous effectuons une analyse en ondelettes de ce processus (voir la section 2.4 pour plus de détails). Nous avons réalisé un travail un peu plus général dans [ACLP10]. Partant de la reparamétrisation de la fonction de covariance décrite par la Proposition 1, nous avons notamment obtenu un critère nécessaire et suffisant d'existence du MFBM. Nous montrons également que le MFBM apparaît comme la limite de sommes partielles, détaillons un algorithme de simulation exacte de ce processus (obtenu par [26]) et proposons quelques trajectoires dans des cas plus généraux que ceux traités dans [CAA10].

Chapitre 2

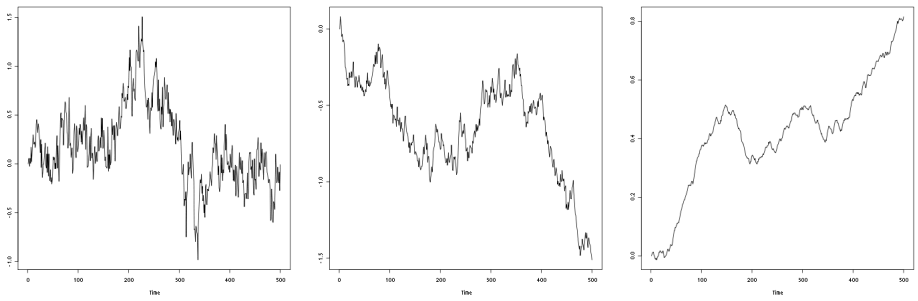
Problèmes d'inférence considérés

Sommaire

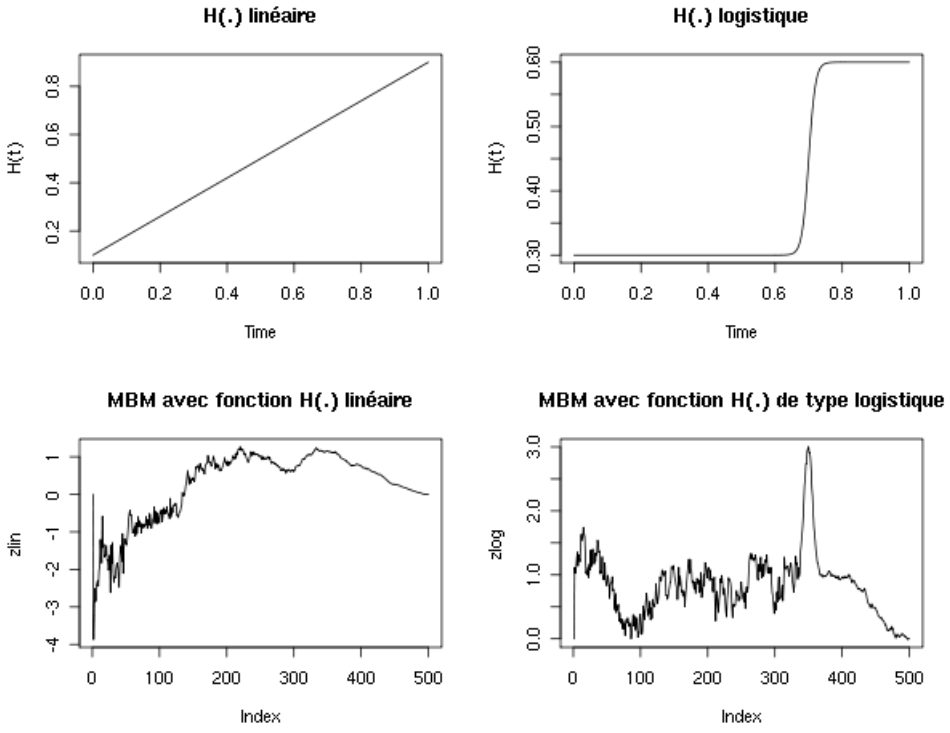
Dans ce chapitre, nous résumons les différentes contributions en termes d'inférence relatives aux modèles présentés dans le Chapire 1.

2.1 Simulation des modèles considérés au Chapitre 1

Le tout premier travail, associé à la publication [Coe00], a consisté en une étutde bibliographique et comparative des méthodes de simulation du FBM et d'estimation de ses paramètres (H et σ^2). L'ensemble des méthodes (simulation/estimation) décrites constituent un ensemble de fonctions R/S-plus disponibles en ligne sur le site du journal. Depuis, ce problème a également été entrepris par [45] et [42]. Dans [Coe00], nous montrons que pour simuler de manière exacte et extrêmement rapide les accroissements d'un FBM (sur une grille régulière), la méthode la plus performante est celle dite de la matrice circulante (adaptée pour simuler des trajectoires discrétisées de processus gaussiens stationnaires), initiée par Davies et Harte (1987) et reprise et améliorée par Wood et Chan [119]. Pour simuler un bruit gaussien fractionnaire aux instants $0, 1, \ldots, n-1$, on peut naturellement penser à calculer la racine carrée de la matrice de Toeplitz G et à définir $Y = G^{1/2}Z$ où $Z \rightsquigarrow \mathcal{N}(0, I_n)$. Cette approche est coûteuse en temps de calculs et peut être source de problèmes numériques pour certaines valeurs de H. Pour contourner cette difficulté, l'idée consiste à plonger la matrice de Toeplitz $G = (\gamma(j-i))$ dans une matrice circulante, C, de taille une puissance de deux dont la première ligne s'écrit $(\gamma(0), \gamma(1), \ldots, \gamma(m/2-1), \gamma(m/2), \gamma(m/2-1), \ldots, \gamma(1))$. Puisque, G est une sous-matrice de C, il suffit alors de prendre les n premières composantes du vecteur $Y \rightsquigarrow \mathcal{N}(0, C)$. La caractéristique circulante de C s'exprime par le fait qu'elle peut être décomposée sous la forme $C = Q\Lambda Q^*$, où Q est la matrice unitaire définie par $Q = (m^{-1/2}e^{-2i\pi jk/m})$ et Λ diagonale. Le calcul de $C^{1/2}$ (lorsqu'il est possible) est bien plus simple et rapide que celui de $G^{1/2}$ car il passe par une double utilisation de la Transformée de Fourier rapide. Les résultats de [31] et [43] montrent que théoriquement les valeurs propres de C sont strictement positives dès la première puissance de 2 supérieure à $2(n-1)$. Le coût de cet algorithme est en $\mathcal{O}(n \log n)$. Voici quelques trajectoires pour différents paramètres de Hurst ($H = 0.3$ à gauche, 0.5 au milieu et 0.8 à droite).



Comme précisé dans le Chapitre précédent, les accroissements d'un MBM ne sont plus stationnaires. En conséquence, la méthode de Wood et Chan n'est plus adaptée pour ce processus. La seule méthode (exacte en théorie) consiste à obtenir une décomposition de Cholesky de la matrice de covariance (voir [Coe05]). Clairement, cette méthode ne peut être implémentée de manière efficace au-delà de $n = 10000$. La figure ci-dessous illustre quelques trajectoires de MBM pour une fonction $H(\cdot)$ linéarie (à gauche) et une fonction logistique (à droite).

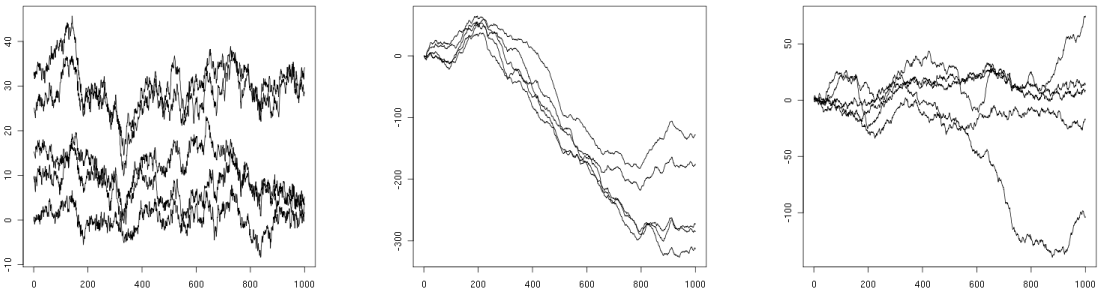


En ce qui concerne le problème de simulation d'un MFBM, notons ici $\Delta X^{(n)}$ le vecteur agrégé $\Delta X^{(n)} = (\Delta X(1)', \ldots, \Delta X(n)')'$ (où rappelons-le $X(t) = (X_1(t), \ldots, X_p(t))'$) et par $\mathbb{G}$ sa matrice de covariance. $\mathbb{G}$ est une matrice de Toeplitz par blocs de taille $np \times np$ donnée par $\mathbb{G} = (G(|i-j|))_{i,j=1,\ldots,n}$ où pour $h = 0, \ldots, n-1$, $G(h)$ est la matrice de taille $p \times p$ donnée par $G(h) := (\gamma_{i,j}(h))_{j,k=1,\ldots,p}$. Le problème de simulation des accroissements d'un MFBM peut alors être vu comme la simulation d'un vecteur suivant une loi $\mathcal{N}_{np}(0, \mathbb{G})$. De manière similaire à la simulation d'un FBM, l'idée de Chan et Wood [26] est de plonger la matrice $\mathbb{G}$ dans une matrice circulante par blocs $C = circ\{C(j), j = 0, \ldots, m-1\}$, où m est à nouveau une puissance de 2 supérieure à

$2(n-1)$ et où $C(j)$ est la matrice $p \times p$ donnée par

$$C(j) = \begin{cases} G(j) & \text{si } 0 \leq j < m/2 \\ \frac{1}{2}\big(G(j) + G(j)'\big) & \text{si } j = m/2 \\ G(j-m) & \text{si } m/2 < j \leq m-1. \end{cases} \tag{2.1}$$

Une telle définition assure que C est symmétrique et circulante par blocs et que $\mathbb{G} = \{C(j), j = 0, \ldots, n-1\}$ est une sous-matrice de C. Il suffit donc à nouveau de savoir simuler une loi $\mathcal{N}_{mp}(0, C)$, i.e. de calculer $C^{1/2}$. Et ceci est à nouveau plus abordable que le calcul de $\mathbb{G}^{1/2}$ car nous avons le résultat suivant : il existe m matrices Hermitiennes $B(j)$ de taille $p \times p$ telles $C = (J \otimes I_p)\ \mathrm{diag}(B(j), j = 0, \ldots, m-1)\ (J^* \otimes I_p)$, où $\otimes$ désigne le produit de Kronecker. Ce résultat permet (lorsque $C(j)$ est positive) à nouveau une utilisation très efficace de la Transformée de Fourier rapide permettant de réduire le coût de $\mathcal{O}(pn^3)$ à $\mathcal{O}(p^2 m \log m) + \mathcal{O}(mp^3)$. Ci-dessous quelques exemples de trajectoires de MFBM causaux ($p = 5$, H_i uniformément répartis dans l'intervalle $[0.3, 0.4]$ (à gauche), $[0.8, 0.9]$ (au milieu) and $[0.4, 0.8]$ (à droite), $\rho_{ij} = 0.6$ (à gauche et au milieu) et 0.3 à droite).



2.2 Inférence autour du FBM

2.2.1 Estimation rapide et efficace - Technique de filtrage

Depuis le travail initial de Mandelbrot et Van Ness, de nombreuses méthodes d'estimation du paramètre de Hurst ont été proposées, voir par exemple [Coe00], [45] ou [54] pour une revue des méthodes existantes. Durant mon travail de thèse, nous nous sommes intéressés à une classe de méthodes spécifique basée sur des techniques de filtrage. Cette méthode initiée simultanément par [77] et [69] dans le contexte des processus gaussiens localement autosimilaires en zéro a été reprise et améliorée dans [Coe01]. Décrivons assez brièvement les idées principales de cette méthode qui ont été exploitées dans d'autres contextes par nous-même (voir les paragraphes suivants) et d'autres auteurs ([28], [102], [68], [25]).

2.2.1.1 Quelques notations

Soit $\mathbf{X} = (X(0), \ldots, X((n-1)/n))$ un échantillon d'un processus stochastique à accroissements stationnaires (de variance finie) discrétisé aux instants $i = 0, \ldots, (n-1)/n$. Soit a un filtre de longueur $\ell + 1$ et d'ordre $p \geq 1$, i.e. un vecteur ayant $\ell + 1$ composantes réelles satisfaisant

$$\sum_{q=0}^{\ell} q^j a_q = 0 \text{ pour } j = 0, \ldots, p-1 \text{ et } \sum_{q=0}^{\ell} q^p a_q \neq 0. \tag{2.2}$$

Quelques exemples de filtres les plus utilisés : accroissements d'ordre 1 - $a = i1 = (-1, 1)$, accroissements d'ordre 2 - $a = i2 = (1, -2, 1)$, daublets 4 - $a = d4 = (-0.09150635, -0.15849365, 0.59150635, -0.34150635)$.

Notons $\mathbf{X}^a$ le vecteur $\mathbf{X}$ filtré par a et donné pour $i = \ell, \ldots, n-1$ par $X^a(i/n) := \sum_{q=0}^{\ell} a_q X(i - q/n)$. Ainsi, $\mathbf{X}^{i2}$ représente les accroissements d'ordre 2 du vecteur $\mathbf{X}$. Nous noterons également $\widetilde{\mathbf{X}}^a$ le vecteur $\mathbf{X}^a$ normalisé, défini par $\widetilde{X}^a(i) = \frac{X^a(i/n)}{E(X^a(i/n)^2)^{1/2}}$. Enfin, $\overline{\mathbf{X}}$ désignera la moyenne empirique de $\mathbf{X}$ et pour une fonction $g(\cdot)$, $\mathbf{g}(\mathbf{X}) := (g(X(0)), \ldots, g(X(n-1/n)))$.

2.2.2 Applications au FBM

Dans ce paragraphe, $\mathbf{X}$ est une trajectoire discrétisée d'un FBM de paramètres $(H, \sigma) \in (0,1) \times \mathbb{R}^+$. Les fonctions de covariance et corrélation de $\mathbf{X}^a$ sont données pour $i, j \in \mathbb{Z}$ par (voir [Coe01])

$$EX^a\left(\frac{j}{n}\right) X^a\left(\frac{i+j}{n}\right) = \frac{\sigma^2}{n^{2H}} \times \pi_H^a(i) \text{ avec } \pi_H^a(i) = -\frac{1}{2} \sum_{q,r=0}^{\ell} a_q a_r |q - r + i|^{2H}$$

et $\rho_H^a(i) := E\widetilde{X}^a(j)\widetilde{X}^a(i+j) = \frac{\pi_H^a(i)}{\pi_H^a(0)}$. L'intérêt de filtrer un FBM est révélé par le résultat suivant (voir e.g. [Coe01]) :

$$\rho_H^a(i) \sim k_H |i|^{2H-2p}, \text{ lorsque } |i| \to +\infty,$$

exprimant que plus l'ordre du filtre est élevé plus la fonction de corrélation décroît rapidement. Définissons la collection suivante $(a^m)_{m \geq 1}$ de filtres dilatés du filtre a : a^m est le filtre de longueur $m\ell + 1$ et d'ordre p défini pour $i = 0, \ldots, m\ell$ par

$$a_i^m = \begin{cases} a_{i/m} & \text{si } i/m \text{ est un entier} \\ 0 & \text{sinon.} \end{cases} \tag{2.3}$$

A titre d'exemple, si $a := a^1 = (1, -2, 1)$, alors $a^2 := (1, 0, -2, 0, 1)$. On peut alors montrer que

$$\pi_H^{a^m}(0) = -\frac{1}{2} \sum_{q,r=0}^{m\ell} a_q^m a_r^m |q - r|^{2H} = -\frac{1}{2} \sum_{q,r=0}^{\ell} a_q a_r |mq - mr|^{2H} = m^{2H} \pi_H^a(0).$$

Ainsi, $E\left(\overline{(\mathbf{X}^{a^m})^2}\right) = \sigma^2 \pi_H^{a^m}(0) = m^{2H}\gamma$, où $\gamma := \frac{\sigma^2}{n^{2H}} \pi_H^a(0)$ (qui est indépendant de m), ce qui nous amène à

$$\log \overline{(\mathbf{X}^{a^m})^2} = 2H \log m + \log \gamma + \underbrace{\log\left(\frac{\overline{(\mathbf{X}^{a^m})^2}}{E\left(X^{a^m}(1/n)^2\right)}\right)}_{:=\varepsilon_m^{ST}}. \tag{2.4}$$

On obtient alors une estimation explicite du paramètre de Hurst en estimant H par moindres carrés ordinaires pour $m = M_1, \ldots, M_2$:

$$\widehat{H}^{ST} = \frac{\mathbf{A}'}{2\|\mathbf{A}\|^2} \left(\log\left(\overline{(\mathbf{X}^{a^m})^2}\right)\right)_{m=M_1,\ldots,M_2}, \tag{2.5}$$

où $A_m = \log(m) - \frac{1}{M_2 - M_1 + 1} \sum_{m=M_1}^{M_2} \log(m)$ et $1 \leq M_1 \leq M_2$. Notons que cette procédure est extrêmement intéressante car la définition de l'estimateur est indépendante du paramètre σ^2 et du pas de discrétisation. Le principal résultat de [Coe01] est le suivant :

Proposition 6 *Lorsque $n \to +\infty$, $\widehat{H}^{ST}$ converge presque sûrement vers H et lorsque $p > H + 1/4$, il existe $\sigma_{ST}^2(H)$ tel que $\sqrt{n}(\widehat{H}^{ST} - H) \xrightarrow{d} \mathcal{N}(0, \sigma_{ST}^2(H))$.*

Pour démontrer ces résultats, nous nous sommes essentiellements basés sur les travaux de Doob [44] et sur un théorème central limite pour des processus gaussiens stationnaires subordonnés obtenu par Breuer et Major [24].

Pour terminer cette section énonçons les quelques remarques suivantes associées à ce résultat : la condition $p > H + 1/4$ est toujours vérifiée lorsque $p \geq 2$ et ne l'est que pour $H \in (0, 3/4)$ lorsque $p = 1$. Ensuite, lorsque le paramètre σ est supposé connu, un seul filtre suffit pour estimer H et nous avons obtenu un résultat équivalent à la Proposition 6 (cf Proposition 2 de [Coe01]) avec une vitesse de convergence en $\sqrt{n}\log n$. Estimer H en utilisant le k-ème moment absolu empirique ($k > 0$) plutôt que la variance empirique, n'améliore pas l'estimation puisque nous avons montré (cf Corollaire 2 de [Coe01]) que la variance asymptotique était minimale lorsque $k = 2$. Enfin, en estimant la matrice de covariance asymptotique du vecteur (ε_m^{ST}), on peut définir un estimateur par moindres carrés généralisés de H ayant des propriétés similaires à la Proposition 6. Notons pour terminer que les estimateurs développés sont très rapides même pour des tailles d'échantillon élevés et sont plutôt compétitifs par rapport à l'estimateur du maximum de vraisemblance.

2.2.3 Bornes de Cramèr-Rao pour les paramètres d'un FBM

Pour étudier l'efficacité d'un estimateur on compare usuellement son écart quadratique moyen avec la borne de Cramèr-Rao (CRB). Soit $\mathbf{Y} = (Y_1, \ldots, Y_n)$ un vecteur aléatoire dont la densité de probabilité dépend d'un paramètre inconnu θ. Sous des conditions de régularité (*e.g.* Dacunha-Castelle et *al* [33], p.178), on obtient pour tout estimateur non biaisé $\widehat{\theta}$ l'inégalité suivante :

$$E\left(\widehat{\theta} - \theta\right)^2 \;\geq\; I_n(\theta)^{-1} \equiv CRB_n(\theta)\;, \tag{2.6}$$

où $I_n(\theta)$ est la matrice d'information de Fisher définie par :

$$I_n(\theta) \;= \left\{ E\left(\frac{\partial}{\partial\theta_i} \log L(\mathbf{Y},\theta)\, \frac{\partial}{\partial\theta_j} \log L(\mathbf{Y},\theta) \right) \right\}_{1\leq i,j\leq k} \tag{2.7}$$

où L est la vraisemblance du modèle. Une inégalité semblable pour les estimateurs biaisés est également disponible (e.g.[33], p.178). Les résultats que nous avons obtenus dans [CI01] (en collaboration avec J. Istas) sur le comportement des bornes de Cramèr-Rao pour le(s) paramètre(s) d'un FBM sont partiellement connus. Dans [34], Dahlhaus explicite analytiquement les bornes de Cramèr-Rao des estimateurs de paramètres pour certains processus autosimilaires, à longue mémoire. En particulier, il étudie les accroissements d'un FBM. Le point-clé de sa preuve réside dans le fait qu'il contrôle $\mathrm{Tr}\left(\frac{d}{dH}G.G^{-1}\right)$, où $G := G(H,\sigma^2) = (\gamma(j-i/n))$ désigne la matrice de covariance des accroissements d'un FBM. Notre contribution est d'une part d'avoir regardé le cas σ connu et d'autre part de fournir un résultat moins fort mais avec une preuve originale basée sur le contrôle du comportement de coefficients d'inverses de matrices dont les coefficients décroissent hyperboliquement en séloignant de la diagonale.

Dans la suite, le modèle statistique sera constitué d'un échantillon des accroissements d'un FBM discrétisé aux instants i/n pour $i = 0, \ldots, n-1$. Nous utilisons ici la notation suivante pour x_n and y_n, deux suites réelles : $x_n \asymp y_n$, s'il existe $c_1, c_2 > 0$ telles que $c_1|y_n| \leq |x_n| \leq c_2|y_n|$). Pour deux matrices A_n et B_n, $A_n \asymp B_n$ si pour tout i, j, on a $(A_n)_{i,j} \asymp (B_n)_{i,j}$.

Théorème 7 *(i) Si σ est connu :*

$$CRB_n(H) \;\sim\; \frac{1}{2}\,\frac{1}{n\log^2(n)}\;.$$

(ii) Si σ est inconnu :

$$CRB_n\left(\, H,\sigma^2 \,\right) \;\asymp\; \begin{pmatrix} 1/n & \log(n)/n \\ \log(n)/n & \log^2(n)/n \end{pmatrix}\;.$$

Sans préciser la preuve soulignons que l'outil principal est un résultat sur les suites d'inverse de matrice qui généralise celui obtenu par Jaffard [71] pour des opérateurs. Commençons par la définition suivante : pour $\alpha > 0,\ c > 0$ (indépendantes de n), soit $(Q_{\alpha,n})_{n\geq 1}$ la suite de sous-ensembles de matrices $n \times n$ telles que :

$$A \in Q_{\alpha,n} \ \Leftrightarrow\ |(A)_{k,k'}| \ \leq\ c\,(\,1+|k'-k|\,)^{-\alpha}\ ,\ \forall k,k'=1,\ldots,n.$$

Remarquons, comme dans [71], que $Q_{\alpha,n}$ est une algèbre pour $\alpha > 1$. Nous avons alors établi le résultat suivant.

Théorème 8 *Soit $\alpha \in]1,3/2[$ et soit $(A_n)_{n\geq 1}$ une suite de matrices $n \times n$ symétriques définies positives telles que $A_n \in Q_{\alpha,n}$. Soit A_∞ l'opérateur défini, si la limite existe, pour $(i,j) \in \mathbb{Z}^2$ par*

$$(A_\infty)_{i,j}\ =\ \lim_{n\to+\infty}(A_n)_{i,j}\ .$$

Supposons maintenant les conditions suivantes

$(\mathbf{A_1})$ $(A_\infty)_{i,i+j} = a(j)$ *avec* $a(j) \sim c|j|^{-\alpha}$ *lorsque* $|j| \to +\infty$
$(\mathbf{A_2})$ $\widehat{a}(\lambda) > 0$ *pour* $\lambda \neq 0$
$(\mathbf{A_3})$ $\sum_{i,j=1}^{n}\{(A_\infty)_{i,j}\ -\ (A_n)_{i,j}\}^2\ = o(1)\ ,$ *lorsque* $n \to +\infty$

alors $A_n^{-1} \in Q_{\alpha,n}$.

2.2.4 Estimation robuste

2.2.4.1 Robustesse aux données aberrantes

Stoev et al. [111] ont mis en avant le fait qu'une des méthodes les plus connues pour estimer l'exposant de Hurst (d'un processus autosimilaire et/ou à longue mémoire) à savoir la méthode d'ondelettes est particulièrement sensible à l'ajout de données aberrantes. Cette remarque est également valable pour la méthode des variations discrètes décrite dans la Section 2.2.1. Dans [111], les auteurs proposent alors de remplacer la variance empirique des coefficients d'ondelettes à chaque échelle par la médiane des coefficients d'ondelettes au carré. Cette procédure, pour laquelle les auteurs précisent n'avoir aucune justification théorique est clairement plus robuste. Notre contribution dans ce contexte a été d'étendre la procédure proposée dans [111] en étudiant des combinaisons convexes de quantiles empiriques et des moyennes tronquées et en fournissant des résultats de convergence.

Partons des notations établies dans la Section 2.2.1. Soit $(\mathbf{p},\mathbf{c}) = (p_k,c_k)_{k=1,\ldots,K} \in ((0,1)\times\mathbb{R}^+)^K$ pour un entier K fini. Définissons la statistique suivante : $\widehat{\xi}(\mathbf{p},\mathbf{c},\mathbf{X}) = \sum_{k=1}^{K} c_k\,\widehat{\xi}(p_k,\mathbf{X})$, où $\widehat{\xi}(p,\mathbf{X})$ désigne le quantile empirique d'ordre $p \in (0,1)$ basé sur le vecteur $\mathbf{X}$. De la même façon, pour $\beta = (\beta_1,\beta_2)$ avec $0 < \beta_1,\beta_2 < 1/2$, on notera $\overline{\mathbf{X}}^{(\beta)}$ la moyenne tronquée d'ordre β donnée par

$$\overline{\mathbf{X}}^{(\beta)} = \frac{1}{n-[n\beta_2]-[n\beta_1]}\sum_{i=[n\beta_1]+1}^{n-[n\beta_2]}(\mathbf{X}^a)_{(i),n}\,,$$

où $[\cdot]$ désigne la partie entière et où $(\mathbf{X})_{(1),n} \leq (\mathbf{X})_{(2),n} \leq \ldots \leq (\mathbf{X})_{(n),n}$ sont les statistiques d'ordre. Dans la suite $Y \rightsquigarrow \mathcal{N}(0,1)$. L'idée des procédures d'estimation robuste part des calculs suivants :

$$\widehat{\xi}(\mathbf{p},\mathbf{c},(\mathbf{X}^a)^2) = E\left(X^a(1/n)^2\right)\times\widehat{\xi}(\mathbf{p},\mathbf{c},(\widetilde{\mathbf{X}}^a)^2) \quad \text{et} \quad \overline{(\mathbf{X}^a)^2}^{(\beta)} = E\left(X^a(1/n)^2\right)\times\overline{(\widetilde{\mathbf{X}}^a)^2}^{(\beta)}.$$

On peut s'attendre à ce que $\widehat{\xi}(\mathbf{p},\mathbf{c},(\widetilde{\mathbf{X}}^a)^2)$ et $\overline{(\widetilde{\mathbf{X}}^a)^2}^{(\beta)}$ convergent respectivement lorsque $n\to+\infty$ vers $\xi_{Y^2}(\mathbf{p},\mathbf{c})=\sum_{k=1}^K c_k\xi_{Y^2}(p_k)$ et $\overline{Y^2}^{(\beta)}=(1-\beta_2-\beta_1)^{-1}\int_{\beta_1}^{1-\beta_2}\xi_{Y^2}(p)dp$. Par conséquent, en utilisant à nouveau la collection de filtres dilatés, on obtient (lorsque n est grand)

$$\widehat{\xi}(\mathbf{p},\mathbf{c},(\widetilde{\mathbf{X}}^{a^m})^2)\simeq m^{2H}\gamma\xi_{Y^2}(\mathbf{p},\mathbf{c})\quad\text{et}\quad\overline{(\widetilde{\mathbf{X}}^{a^m})^2}^{(\beta)}\simeq m^{2H}\gamma\overline{Y^2}^{(\beta)},$$

ce qui suggère d'estimer H par une régression log-linéaire en prenant un certain nombre de filtres dilatés et ainsi définir

$$\widehat{H}^Q = \frac{\mathbf{A}'}{2\|\mathbf{A}\|^2}\left(\log\left(\widehat{\xi}\left(\mathbf{p},\mathbf{c},\overline{(\mathbf{B}_H^{a^m})^2}\right)\right)\right)_{m=M_1,\ldots,M_2} \tag{2.8}$$

$$\widehat{H}^{TM} = \frac{\mathbf{A}'}{2\|\mathbf{A}\|^2}\left(\log\left(\overline{(\mathbf{B}_H^{a^m})^2}^{(\beta)}\right)\right)_{m=M_1,\ldots,M_2}. \tag{2.9}$$

On remarquera la très forte analogie entre (2.5), (2.8) et (2.9). De manière analogue à la Proposition 6, nous avons obtenu dans [Coe08b] le résultat suivant.

Proposition 9 *Pour $\bullet=Q,TM$, lorsque $n\to+\infty$, $\widehat{H}^\bullet$ converge presque sûrement vers H et lorsque $p>H+1/4$, il existe $\sigma^2_\bullet(H)$ tel que $\sqrt{n}(\widehat{H}^\bullet-H)\xrightarrow{d}\mathcal{N}(0,\sigma^2_\bullet(H))$.*

Précisons que les constantes asymptotiques sont explicites et ne dépendent que de H et des autres paramètres des estimateurs. Les Théorèmes 4 et 5 établis dans [Coe08b] sont en réalité plus complets que la proposition précédente. D'une part, nous explicitons une borne presque sûrement de la vitessse de convergence des estimateurs. D'autre part, les résultats ont été démontrés (moyennant des hypothèses de régularité supplémentaires) pour une classe plus large que celle du FBM, à savoir la classe des processus gaussiens localement autosimilaires en zéro.

Les preuves sont établies en utilisant deux ingrédients complémentaires : des représentations de type Bahadur pour des quantiles empiriques de processus gaussiens stationnaires subordonnées (qui permettent de ramener le comportement d'un quantile à celui de la fonction de répartition empirique) et un TCL multivarié obtenu par Arcones [5]. Décrivons un peu plus en détail le premier ingrédient.

Soit g une fonction telle que $Eg(Y)^2<+\infty$ et $p\in(0,1)$. Pour établir une représentation de type Bahadur (voir *e.g.* [107]), il nous faut assurer que $F'_{g(Y)}(\xi(p))>0$ (où $F_{g(Y)}$ est la fonction répartition de $g(Y)$ et $\xi(p):=\xi_{g(Y)}(p)$ pour simplifier) et $F''_{g(Y)}(\cdot)$ existe et est bornée dans un voisinage de $\xi(p)$. Nous supposons donc que g satisfait (see *e.g.* [32], p.33) :

$\mathbf{A}(\xi(\mathbf{p}))$: il existe U_i, $i=1,\ldots,L$, des ensembles ouverts disjoints tels que U_i contienne une unique solution à l'équation $g(t)=\xi_{g(Y)}(p)$, tel que $F'_{g(Y)}(\xi(p))>0$ et tel que g soit un $\mathcal{C}^2-$difféomorphisme sur $\cup_{i=1}^L U_i$.

Définissons $\overline{\tau}_p=\inf_{\gamma\in\cup_{i=1}^L g(U_i)}\tau(\gamma)$ où pour un réel u, $\tau(u)$ est le rang d'Hermite de la fonction $h_u(t)=\mathbf{1}_{\{g(t)\le u\}}(t)-F_{g(Y)}(u)$.

Nous avons alors obtenu le résultat suivant (Théorème 2 de [Coe08b])

Théorème 10 *Soit $\{Y(i)\}_{i=1}^{+\infty}$ un processus Gaussien stationnaire centré de variance 1 et de fonction de corrélation $\rho(\cdot)$ telle que lorsque $i\to+\infty$*

$$|\rho(i)|\sim L(i)\,i^{-\alpha}, \tag{2.10}$$

pour un certain $\alpha>0$ et pour une fonction L à variations lentes. Alors sous l'hypothèse $\mathbf{A}(\xi(\mathbf{p}))$, nous avons presque sûrement lorsque $n\to+\infty$

$$\widehat{\xi}(p,\mathbf{g}(\mathbf{Y}))-\xi_{g(Y)}(p)=\frac{p-\widehat{F}(\xi_{g(Y)}(p);\mathbf{g}(\mathbf{Y}))}{f_{g(Y)}(\xi_{g(Y)}(p))}+\mathcal{O}_{p.s.}(r_n(\alpha,\overline{\tau}_p)), \tag{2.11}$$

où la suite $\left(r_n(\alpha,\overline{\tau}_p)\right)_{n\geq 1}$ *est définie par*

$$r_n(\alpha,\overline{\tau}_p) = \begin{cases} n^{-3/4}\log(n)^{3/4} & si \quad \alpha\overline{\tau}_p > 1, \\ n^{-3/4}\log(n)^{3/4}L_{\overline{\tau}_p}(n)^{3/4} & si \quad \alpha\overline{\tau}_p = 1, \\ n^{-1/2-\alpha\overline{\tau}_p/4}\log(n)^{\overline{\tau}_p/4+1/2}L(n)^{\overline{\tau}_p/4} & si \quad 2/3 < \alpha\overline{\tau}_p < 1, \\ n^{-\alpha\overline{\tau}_p}\log(n)^{\overline{\tau}_p}L(n)^{\overline{\tau}_p} & si \quad 0 < \alpha\overline{\tau}_p \leq 2/3, \end{cases} \tag{2.12}$$

où pour $\tau \geq 1$, $L_\tau(n) = \sum_{|i|\leq n} |\rho(i)|^\tau$.

Notons que si $L(\cdot)$ est une fonction croissante $L_\tau(n) = \mathcal{O}(\log(n)L(n)^\tau)$.

Le comportement de la suite $r_n(\cdot,\cdot)$ est relié à la nature courte ou longue-mémoire du processus $\{h_u(Y(i))\}_{i=1}^{+\infty}$ pour u dans un voisinage de $\xi(p)$. Lorsque $\alpha\overline{\tau}_p > 1$, le résultat est semblable à celui obtenu par Bahadur, see *e.g.* [107], dans le cas i.i.d. Pour des processus linéaires à courte mémoire, en utilisant un résultat du type loi du logarithme itéré, Wu [121] a obtenu la borne $n^{-3/4}\log\log(n)^{3/4}$ sous l'hypothèse que $F'(\cdot)$ et $F''(\cdot)$ existent et sont uniformément bornées.

Le comportement de la convergence presque sûre de $\widehat{H}^Q$ est essentiellement basé sur l'application du Théorème 10 en prenant $g(t) = t^2$. Pour la convergence de $\widehat{H}^{TM}$, une représentation uniforme de Bahadur est nécessaire (voir Théorème 3 de [Coe08b]).

Dans le cas i.i.d., Ghosh [58] a obtenu un résultat plus simple mais suffisant pour des applications statistiques. Il a établi via une preuve originale, en supposant que F' existe et est bornée dans un voisinage de $\xi(p)$, que le terme de reste dans la représentation satisfait $r_n = o_P(n^{-1/2})$ (autrement dit $n^{1/2}r_n \to 0$ en probabilité). Dans le même esprit, nous avons obtenu dans [Coe08a] un résultat analogue pour des fonctionelles non-linéaires de processus gaussiens stationnaires.

Théorème 11 *Sous les conditions du Théorème 10 avec* $L(\cdot) = 1$ *et en supposant* $\mathbf{A}(\xi(\mathbf{p})$ *avec* g *un* $\mathcal{C}^1-$*difféomorphisme sur* $\cup_{i=1}^L U_i$*, alors une relation de type (2.11) prévaut en remplaçant* $\mathcal{O}_{p.s.}(r_n(\alpha,\overline{\tau}_p))$ *par* $o_P(r_n(\alpha,\overline{\tau}_p))$ *avec*

$$r_n(\alpha,\overline{\tau}_p) = \begin{cases} n^{-1/2} & si \quad \alpha\overline{\tau}_p > 1, \\ n^{-1/2}\log(n)^{1/2} & if \quad \alpha\overline{\tau}_p = 1, \\ n^{-\alpha\overline{\tau}_p/2} & if \quad \alpha\overline{\tau}_p > 1. \end{cases} \tag{2.13}$$

2.2.4.2 Robustesse à un bruit additif

Dans le même esprit que la recherche d'un estimateur qui soit robuste à des données aberrantes, rapide à implémenter, indépendant du coefficient σ^2 et simple à étudier nous avons regardé (en collaboration avec Sophie Achard) comment définir un tel estimateur en présence d'un bruit additif. Ceci a fait l'objet d'un travail publié dans [AC10] (où nous reprenons également les méthodes développées dans la section précédente) et du package `R dvfbm`. Nous présentons ici simplement l'idée générale. Les problèmes évoqués (en particulier le modèle $B0$ ci-dessous) ont été regardés par d'autres auteurs [108] et [13] dans un contexte d'ondelettes.

Dans cette section, $\mathbf{B_H}$ représente un FBM discrétisé aux instants $i = 0,\ldots,n-1$. Deux modèles de bruit additif ont été considérés conduisant à deux modifications de la méthode des variations discrètes originale.

Modèle $B0$: les accroissements sont perturbés par un bruit Gaussien additif. Autrement dit

$$X(i) = B_H(i) \ + \ \sigma_b \, B^{(0)}(i),$$

où $H \neq 1/2$, $\sigma_b > 0$ et où $B^{(0)}(i)$ pour $i = 0,\ldots,n-1$ est une discrétisation d'un Brownien standard. Cette perturvation affecte la variance de la série filtrée de la manière suivante (puisque $B^{(0)}$ est un FBM de paramètre $1/2$)

$$EX^a(i)^2 = \sigma^2 \, \pi_H^a(0) \ + \ \sigma_b^2 \, \pi_{1/2}^a(0),$$

ce qui conduit à

$$E\left(\overline{(\mathbf{X}^{a^m})^2}\right) = m^{2H}\gamma + m\sigma_b^2\pi^a_{1/2}(0).$$

Définissons alors $Y^{a^m}(i) = \frac{X^{a^m}(i)}{\sqrt{m}}$, l'estimation de H est alors basée sur le calcul suivant valable lorsque $H \neq 1/2$

$$\begin{aligned} E\left(\overline{(\mathbf{Y}^{a^{2m}})^2} - \overline{(\mathbf{Y}^{a^m})^2}\right) &= \left((2m)^{2H-1} - m^{2H-1}\right)\gamma + \frac{2m}{2m}\sigma^2\pi^a_{1/2}(0) - \frac{m}{m}\sigma^2\pi^a_{1/2}(0) \\ &= m^{2H-1}\left(2^{2H-1}-1\right)\gamma. \end{aligned}$$

On retrouve une certaine log-linéarité qui nous permet de proposer comme estimateur, l'estimateur

$$\widehat{H}^{B0-ST} = \frac{1}{2} + \frac{\mathbf{A}'}{2\|\mathbf{A}\|^2}\left(\log\left(\left|\overline{(\mathbf{Y}^{a^{2m}})^2} - \overline{(\mathbf{Y}^{a^m})^2}\right|\right)\right)_{m=M_1,\ldots,M_2}. \tag{2.14}$$

Modèle $B1$: le FBM est contaminé par un bruit Gaussien additif

$$X(i) = B_H(i) \; + \; \sigma_b \, B^{(1)}(i),$$

où $\sigma_b > 0$ et où $B^{(1)}(i)$ pour $i = 0, \ldots, n-1$ sont des variables aléatoires Gaussiennes standard. De la même façon que précédemment commençons par remarquer que

$$\begin{aligned} EX^a(i)^2 &= \sigma^2\,\pi^a_H(0) \; + \; \sigma_b^2 \sum_{q,r=0}^{\ell} a_q a_r E\left(B^{(1)}(j-q)B^{(1)}(i+j-r)\right), \\ &= \gamma + \sigma_b^2 \sum_{q,r=0}^{\ell} a_q a_r \delta_{q,r} = \gamma + \sigma_b^2|a|^2, \end{aligned}$$

où $|a|^2 = \sum_{q=0}^{\ell} a_q^2$. Puisque $|a^m|^2 = |a|^2$, il vient

$$E\left(\overline{(\mathbf{X}^{a^m})^2}\right) = m^{2H}\gamma + \sigma^2|a|^2.$$

ce qui suggère d'estimer H de la façon suivante :

$$\widehat{H}^{B1-ST} = \frac{\mathbf{A}'}{2\|\mathbf{A}\|^2}\left(\log\left(\left|\overline{(\mathbf{X}^{a^{2m}})^2} - \overline{(\mathbf{X}^{a^m})^2}\right|\right)\right)_{m=M_1,\ldots,M_2}. \tag{2.15}$$

Les estimateurs (2.14) et (2.15) sont en tout point ressemblants à ceux développés dans les sections précédentes. On peut également combiner toutes ces idées en remplaçant dans (2.14) et (2.15) les variances empiriques par des combinaisons convexes de quantiles ou des moyennes tronquées. Nous renvoyons le lecteur à [AC10] pour plus de détails. Toutes ces procédures y sont discutées et une large étude en simulation montrant l'intérêt pratique de chacune de ces méthodes y est proposée. Précisons enfin que nous avons pour le moment obtenu un résultat partiel à savoir la consistance des estimateurs développés (en fonction du modèle), voir pour cela la Proposition 1 [AC10].

2.2.5 Estimation par intervalles de confiance à taille d'échantillon finie

Si l'on souhaite obtenir un intervalle de confiance pour l'exposant de Hurst d'un FBM, une stratégie assez simple consiste à construire un intervalle de confiance asymptotique en utilisant le résultat de normalité asymptotique établi dans la Proposition 6 par exemple. Une autre stratégie consiste à dériver un intervalle en utilisant des inégalités de type concentration. Ceci a été

envisagé par [23] (dans le cas où le paramètre σ régissant un FBM est connu) en utilisant des inégalités de concentration assez générales établies par Nourdin et Viens [98]. En collaboration avec Jean-Christophe Breton, nous avons amélioré et complété ces différents travaux dans un article actuellement soumis [BC10] à plusieurs niveaux : nous avons tout d'abord cherché à améliorer les inégalités de concentration en optimisant les constantes associées, ensuite amélioré singulièrement l'intervalle proposé par [23] qui n'était pas calculable pour des tailles d'échantillon relativement faible et pour toutes les valeurs de H. Nous avons proposé un intervalle de confiance du paramètre H en supposant le coefficient σ inconnu (en utilisant des techniques de filtrage) et enfin comparé les intervalles de confiances obtenues par les inégalités de concentration avec une approche asymptotique.

Commençons par présenter l'inégalité de concentration obtenue. Cette inégalité est relativement générale (voir par exemple Prposition 3 de [BC10]) et est énoncée ici dans le cadre des H_2-variations d'un processus gaussien stationnaire (où $H_2(t) = t^2 - 1$ est le second polynôme d'Hermite).

Proposition 12 *Soit $\{X_i\}_{1\leq i\leq n}$ un processus gaussien stationnaire de variance 1 et de fonction de corrélation ρ. Définissions V_n la statistique $V_n = \frac{1}{n}\sum_{i=1}^n H_2(X_i)$. Enfin soit $\kappa_n = 2\|\rho\|_{\ell^1_n} = 2\sum_{|i|\leq n}|\rho(i)|$, alors pour tout $t > 0$, on a :*

$$P\left(\sqrt{n}V_n \geq t\right) \leq \varphi_{r,n}(t;\kappa_n) := e^{-\frac{t\sqrt{n}}{\kappa_n}}\left(1+\frac{t}{\sqrt{n}}\right)^{\frac{n}{\kappa_n}} \tag{2.16}$$

$$P\left(\sqrt{n}V_n \leq -t\right) \leq \varphi_{l,n}(t;\kappa_n) := e^{\frac{t\sqrt{n}}{\kappa_n}}\left(1-\frac{t}{\sqrt{n}}\right)^{\frac{n}{\kappa_n}}\mathbf{1}_{[0,\sqrt{n}]}(t). \tag{2.17}$$

Par la suite, ce genre d'inégalités a été appliquée au vecteur $\mathbf{X}^a$ (FBM filtré par a). Pour obtenir des bornes indépendantes du paramètre H, un des premiers travail a été de calculer des constantes du type : $\kappa_a := \sum_{H\in I}\|\rho_H^a\|_{\ell^1(\mathbb{Z})}$ (où $I = (0, 1/2]$ pour des filtres d'ordre $p = 1$ et $(0, 1)$ lorsque $p \geq 2$). Nous sommes parvenus (voir Section 3 de [BC10]) à calculer analytiquement ce genre de constantes pour des filtres associés aux accroissements et leurs versions dilatées. A titre d'exemple $\kappa_{i1} = 2, \kappa_{i2} = 8/3, \kappa_{i3} = 16/3$. Munis de ces majorations, nous avons utilisé des idées issues des techniques de filtrage pour définir des intervalles de confiance à taille d'échantillon finie. Nous présentons ici la version avec σ inconnu.

Pour $\alpha \in (0, 1)$, notons $q^a_{\bullet,n}(\alpha) := (\varphi_{\bullet,n})^{-1}(\alpha;\kappa^a)$ pour $\bullet = l, r$ et définissons

$$x^a_{l,n-\ell}(\alpha) := 1 - \frac{q^a_{l,n-\ell}(\alpha)}{\sqrt{n-\ell}} \quad \text{et} \quad x^a_{r,n-\ell}(\alpha) := 1 + \frac{q^a_{r,n-\ell}(\alpha)}{\sqrt{n-\ell}}.$$

Soit $M \geq 2$ et considérons le vecteur $\mathbf{d} = (d_1, \ldots, d_M)'$ de réels non nuls tels que $\sum_{i=1}^M d_i = 0$ et $\mathbf{d}'\mathbf{L_M} > 0$, où $\mathbf{L_M} = (\log(m))_{m=1,\ldots,M}$. Notons enfin I^- et I^+ les sous-ensembles de $\{1, \ldots, M\}$ définis par $I^- = \{i \in \{1, \ldots, M\} : d_i < 0\}$ et $I^+ = \{i \in \{1, \ldots, M\} : d_i > 0\}$. Enfin, soit $\mathbf{L_{S_n}} := \left(\log\left(\overline{(X^{a^m})^2}\right)\right)_{m=1,\ldots,M}$.

Proposition 13 *Soient $\mathbf{L}_{\mathbf{X_n^{inf}}}$ et $\mathbf{L}_{\mathbf{X_n^{sup}}}$ les vecteurs dont les composantes sont données par*

$$\left(\mathbf{L}_{\mathbf{X_n^{inf}}}\right)_m = \begin{cases} \log\left(x^{a^m}_{l,n-m\ell}(\alpha/2M)\right) & si\ m \in I^- \\ \log\left(x^{a^m}_{r,n-m\ell}(\alpha/2M)\right) & si\ m \in I^+ \end{cases}, \left(\mathbf{L}_{\mathbf{X_n^{sup}}}\right)_m = \begin{cases} \log\left(x^{a^m}_{r,n-m\ell}(\alpha/2M)\right) & si\ m \in I^- \\ \log\left(x^{a^m}_{l,n-m\ell}(\alpha/2M)\right) & si\ m \in I^+. \end{cases}$$

1. *Si $n \geq M\ell + 1$, on a*

$$P\left(H \in \left[\widetilde{H}_n^{\inf}(\alpha), \widetilde{H}_n^{\sup}(\alpha)\right]\right) \geq 1 - \alpha \tag{2.18}$$

où

$$\begin{aligned}\widetilde{H}_n^{\inf}(\alpha) &= \max\left(0, \frac{1}{2\mathbf{d}'\mathbf{L_M}}\left(\mathbf{d}'\mathbf{L}_{\mathbf{S_n}} - \mathbf{d}'\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\inf}}\right)\right)\\ \widetilde{H}_n^{\sup}(\alpha) &= \min\left(1, \frac{1}{2\mathbf{d}'\mathbf{L_M}}\left(\mathbf{d}'\mathbf{L}_{\mathbf{S_n}} - \mathbf{d}'\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\sup}}\right)\right).\end{aligned}$$

2. *Lorsque $n \to +\infty$, l'intervalle défini par (2.18) satisfait presque sûrement*

$$\left[\widetilde{H}_n^{\inf}(\alpha), \widetilde{H}_n^{\sup}(\alpha)\right] \to \{H\}$$

et sa longueur, notée μ_n vérifie

$$\mu_n := \frac{\mathbf{d}'\left(\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\inf}} - \mathbf{L}_{\mathbf{X}_\mathbf{n}^{\sup}}\right)}{2\mathbf{d}'\mathbf{L_M}} \sim \frac{1}{\sqrt{n}}\,\frac{\mathbf{d}'\mathbf{q_M}(\alpha/2M)}{\mathbf{d}'\mathbf{L_M}}$$

où $\mathbf{q_M}(\alpha/2M)$ est le vecteur de longueur M donné par

$$\left(\mathbf{q_M}(\alpha/2M)\right)_m := \begin{cases} -q^{a^m}(\alpha/2M) & si\ m \in I^- \\ q^{a^m}(\alpha/2M) & si\ m \in I^+ \end{cases}$$

avec $q^a(\alpha) := \sqrt{2\kappa^a \log(1/\alpha)}$.

La définition de cet intervalle possède des similitudes importantes avec la définition de l'estimateur (2.5). L'intérêt de ce résultat est qu'il semble fournir un résultat intéressant à taille d'échantillon finie et qu'asymptotiquement la vitesse de convergence de la longueur de l'intervalle correspond à celle d'un intervalle défini via un résultat de normalité asymptotique (à savoir $1/\sqrt{n}$).

Nous renvoyons le lecteur à la Section 5 de [BC10] où cet intervalle est comparé via une étude en simulations à celui obtenu via la Proposition 6. Bien qu'à notre sens les avancées théoriques de ce travail sont non négligeables et sont la base de plusieurs perspectives, les résultats de simulation montrent clairement que l'on n'améliore pas les choses lorsque le paramètre σ est supposé connu et que l'intervalle est trop conservateur (i.e. le taux de couverture est trop supérieure au seuil $1-\alpha$) lorsque le paramètre σ est inconnu.

2.3 Estimation de la fonction de régularité d'un MBM

Rappelons qu'un MBM est une extension du FBM dans le sens où la régularité H peut maintenant dépendre du temps. Le résultat (1.3) exprime qu'à tout instant t, le processus tangent au MBM (noté encore dans cette partie $X(\cdot)$) est un FBM de paramètre de Hurst $H(t)$. Dans l'optique d'estimer la fonction de régularité $H(\cdot)$, ce résultat suggère d'adapter "localement" une méthode d'estimation disponible pour le FBM. Ceci a été entrepris par [15] puis étendue (d'un point de vue méthodologique) dans [Coe05]. Pour $\alpha \in (0,1)$, définissons

$$\mathcal{V}_{n,\alpha}(t) = \left\{k \in \mathbb{N}, |k/n - t| \le n^{-\alpha}\right\} \quad \text{et} \quad v_{n,\alpha}(t) = |\mathcal{V}_{n,\alpha}(t)|$$

où pour un ensemble dénombrable nous utilisons ici la notation $|\cdot|$ pour désigner son cardinal. Dans le même esprit que les notations des sections précédentes, notons $\overline{\left(X^a_{n,\alpha}(t)\right)^2} = \frac{1}{v_{n,\alpha}(t)}\sum_{k\in\mathcal{V}_{n,\alpha}(t)} X^a(k/n)^2$. De manière analogue à la procédure standard, on peut définir un estimateur ponctuel de la fonction de régularité d'un MBM de la façon suivante en utilisant la technique des filtres dilatés :

$$\widehat{H}(t) := \frac{A^T}{2\|A\|^2}\left(\log \overline{\left(X^{a^m}_{n,\alpha}(t)\right)^2}\right)_{m=M_1,\ldots,M_2}$$

Dans notre papier [Coe05], nous avions considéré des fonctions de Hurst höldériennes d'ordre $\eta > 0$ sur l'intervalle $[0,1]$ et proposé une procédure adaptative pour choisir le voisinage optimal. Bien que la méthodologie développée ne souffre d'aucun problème, ce travail contient un certain nombre d'erreurs. Premièrement les convergences annoncées (de l'estimateur de la fonction $H(\cdot)$) ont lieu en lois finies-dimensionelles, comme cela nous a été notifié par Arnaud Begyn [14]. Par ailleurs, comme cela nous a été souligné par Jean-Marc Bardet et Donatas Surgailis, nos résultats souffrent d'une erreur fondamentale dans le calcul des covariances $EV^a(k/n)V^a(k'/n)$ pour $k, k' \in \mathcal{V}_{n,\alpha}(t)$. Notre développement asymptotique n'est correct que lorsque $\eta \geq 1$. Le développement correct et les résultats asymptotiques qui s'en déduisent sont établis dans la prépublication [12]).

2.4 Analyse en ondelettes du MFBM

Dans la Section 1.3, nous avions présenté une alternative au FBM dans un cadre multivarié que nous avons récemment étudiée d'un point de vue théorique dans [CAA10] et [ACLP10]. Nous voudrions ici compléter un peu plus ces différentes propriétés en précisant quelques résultats obtenus en analysant les covariances et densités spectrales de la transformée en ondelettes continue d'un MFBM (résultats issus de [CAA10]), ce qui généralise dans un certain sens les travaux de Flandrin [51, 52], Tewfik et Kim [114], Wornell [120] Kato et Masry [75] obtenus pour le FBM. Ces propriétés sont particulièrement intéressantes et seront certainement la base de futurs travaux pour identifier le modèle par exemple.

Soit ψ une ondelette complexe, soient $a > 0$ et $b \in \mathbb{R}$ et définissons $\psi_{ab}(.) = a^{-1/2}\psi((.-b)/a)$. Nous définissons alors

$$d^i_{a,b} \quad := \quad \left\langle X_i \middle| \psi_{ab} \right\rangle_{L^2} = \int_{\mathbb{R}} X_i(t)\overline{\psi_{ab}}(t)dt \tag{2.19}$$

le coefficient en ondelettes de la i-ème composante d'un MFBM. $\overline{\psi}$ désigne le complexe conjugué de ψ. Considérons les hypothèses suivantes :

[C1] Condition d'admissibilité : $\psi(t) \in L^2$ et $|\widehat{\psi}(\omega)|^2/|\omega| \in L^1$, où $\widehat{\psi}$ est la transformée de Fourier de ψ.

[C2(K)] $t^m\psi(t) \in L^1$ pour $m = 0, 1, \ldots, K$.

[C3] ψ possède M moments nuls, i.e. $\int_{\mathbb{R}} t^k\psi(t) = 0$ pour $k = 0, \ldots, M-1$.

Rappelons que [C1] assure que $\widehat{\psi}(0) = 0$ et que $\int_{\mathbb{R}} \psi(t)dt = 0$. Sous l'hypothèse [C2(2)], on peut écrire

$$E[d^i_{a_1,b_1}\overline{d^j_{a_2,b_2}}] = -\frac{\sigma_i\sigma_j}{2}\sqrt{a_1a_2}\int_{\mathbb{R}^2} w_{i,j}(a_2t_2 - a_1t_1 + b_2 - b_1)\overline{\psi}(t_1)\psi(t_2)dt_1dt_2, \tag{2.20}$$

où la fonction $w_{i,j}$ est définie par (1.7). Pour a_1, a_2 fixés, la quantité $E[d^i_{a_2,b_2}\overline{d^j_{a_2,b_2}}]$ peut s'interpréter comme une covariance croisée entre deux signaux et on observe qu'elle ne dépend alors que de la différence $b_2 - b_1$.

Proposition 14 *(i) Sous les hypotèses [C1] et [C2(2)], soient $b_1 = b_2 = b$ et $a_1 = a_2 = a > 0$, alors*

$$E(d^i_{a,b}\overline{d^j_{a,b}}) = \frac{\sigma_i\sigma_j}{2}\, z_{i,j}\, a^{H_i+H_j+1}, \qquad \text{où } z_{i,j} := -\int_{\mathbb{R}^2} w_{i,j}(t_2 - t_1)\overline{\psi}(t_1)\psi(t_2)dt_1dt_2. \tag{2.21}$$

(ii) Sous les hypothèses [C1], [C2(2M+1)] et [C3], alors lorsque $|h| \to +\infty$, on a

$$E\left(d^i_{a_1,b}\overline{d^j_{a_2,b+h}}\right) \sim -\frac{\sigma_i\sigma_j}{2}\kappa(\psi, M)|h|^{H_i+H_j-2M}\widetilde{\tau}_{i,j}(\mathrm{sign}(h))$$

où $\kappa(\psi, M) := \binom{2M}{M}(a_1a_2)^M\left|\int t^M\psi(t)dt\right|^2$ et où $\widetilde{\tau}_{i,j}(\mathrm{sign}(h))$ est une constante ne dépendant que des paramètres du modèle et du signe de h.

La propriété (i) exprime qu'à une même échelle, la covariance instantanée entre les transformées des deux composantes X_i et X_j exhibe une propriété de type autosimilarité. La seconde propriété exprime, quant à elle, le pouvoir de décorrélation des ondelettes. A échelles fixées, la covariance croisée décroît hyperboliquement et d'autant plus vite que le nombre de moments nuls est grand. Ce dernier résultat repose abondamment sur le développement en séries entières des fonctions $(1+x)^\alpha$ et $\log(1+x)$ pour $|x|<1$ et du théorème de convergence dominée. Ces deux propriétés sont finalement des extensions de celles obtenues pour un FBM (dans le cas d'ondelettes à support compact), cf [51, 52].

Sur la base du travail de [75], nous avons également fourni une preuve simple de l'existence au sens L^1 de la densité spectrale croisée d'ondelettes à différentes échelles entre les composantes i et j d'un MFBM.

Proposition 15 *Sous les hypothèses [C1], [C2(M)] et [C3] (avec $M \geq 2$), alors la transformée de Fourier de $E(d^i_{a_1,b_1}\overline{d^j_{a_2,b_2}})$ (en fonction de $h = b_2 - b_1$) existe et est donnée par*

$$\widetilde{S}_{i,j}(\omega) = \frac{\sigma_i\sigma_j}{4\pi}\frac{\Gamma(H_i+H_j+1)}{|\omega|^{H_i+H_j+1}}\sqrt{a_1a_2}\,\overline{\widehat{\psi}(-a_1\omega)}\widehat{\psi}(-a_2\omega)\times\tau_{i,j}(\mathrm{sign}(\omega)) \tag{2.22}$$

où $\tau_{i,j}(\mathrm{sign}(\omega))$ est définie par (1.10).

En regardant le développement limité de $\widetilde{S}_{i,j}$ en zéro, on retrouve un résultat analogue à celui de la Proposition 14 (ii). Cette dernière proposition utilise le théorème de Bochner et une représentation de type Bahr et Essen des fonctions $v_+^\alpha = \max(0,v)^\alpha$ et $v_-^\alpha = max(0,-v)^\alpha$. En 1965, Von Bahr et Essen [11] ont obtenu la représentation suivante pour $\alpha\in(0,2)$:

$$|v|^\alpha = \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi}\int_{\mathbb{R}}\frac{1-\cos(\omega v)}{|\omega|^{\alpha+1}}d\omega.$$

Sur la même idée, nous avons obtenu les représentations suivantes.

Lemme 16 *Pour tout $\alpha\in(0,2)$, $\alpha\neq 1$*

$$v_+^\alpha = \frac{\Gamma(\alpha+1)}{2\pi}\int_{\mathbb{R}}\frac{\sin\left(\pi\frac{\alpha}{2}\right)(1-\cos(\omega v))+\cos\left(\pi\frac{\alpha}{2}\right)\mathrm{sign}(\omega)\,(\sin(\omega v)-g(\omega v))}{|\omega|^{\alpha+1}}d\omega$$

et

$$v_-^\alpha = \frac{\Gamma(\alpha+1)}{2\pi}\int_{\mathbb{R}}\frac{\sin\left(\pi\frac{\alpha}{2}\right)(1-\cos(\omega v))-\cos\left(\pi\frac{\alpha}{2}\right)\mathrm{sign}(\omega)\,(\sin(\omega v)-g(\omega v))}{|\omega|^{\alpha+1}}d\omega$$

où la fonction g est égale à zéro lorsque $\alpha\in(0,1)$ et à la fonction identité lorsque $\alpha\in(1,2)$.

Chapitre 3

Perspectives de recherche et travaux en cours

Sommaire

3.1 En lien direct avec le chapitre précédent

Autour de la simulation d'un MFBM

Comme nous l'avons souligné dans la section 2.1, le problème de simulation des accroissements d'un FBM est parfaitement résolu par la méthode de la matrice circulante puisque les travaux joints de [43] et [31] montrent que la matrice circulante de taille m première puissance de 2 supérieure à $2(n-1)$ est symétrique définie positive pour tout $H \in (0,1)$. La simulation d'un processus gaussien stationnaire multivarié (et donc en particulier les accroisssements d'un MFBM) requiert plus d'attention. L'algorithme proposé par [26] est exact en théorie, à la condition que la matrice circulante par blocs définie par (2.1) soit symétrique positive. La possibilité d'obtenir des résultats similaires au cas du FBM est un problème particulièrement intéressant en termes de coûts de calculs.

Dans le cas où la matrice (définie par (2.1)) n'est pas positive, [26] proposent de tronquer les valeurs propres négatives à zéro. Cela induit naturellement une erreur dans la simulation. Nous pensons alors qu'une utilisation des inégalités de concentration étudiées dans [BC10] pourrait s'avérer intéressante pour mesure une erreur ℓ^2 par exemple.

Autour de l'estimation robuste

Dans le récent travail [AC10], nous avons exploité la technique de filtrage à de nombreuses situations où un FBM pouvait être contaminé. Nous sommes parvenus à définir un estimateur consistant, rapide et facile à implémenter. La variance de ces estimateurs (en particulier ceux définis en présence d'un bruit additif) dépend intrinsèquement des paramètres de ce bruit. Afin d'estimer cette variance, il serait intéressant soit d'estimer ces paramètres soit d'estimer la variance des estimateurs par une technique de type bootstrap par blocs. Une analyse des performances théoriques pourrait alors s'appuyer sur les résultats d'Arcones [5].

Intervalles de confiance à taille d'échantillon finie

Il s'agit ici de pistes de réflexion menées avec Jean-Christophe Breton. La Proposition 13 est une première tentative pour définir un intervalle de confiance du paramètre H (indépendamment de σ) pour des tailles d'échantillon raisonnables. En analysant attentivement la preuve de cette

proposition (voir preuve de la Proposition 7 [BC10]), on s'aperçoit que nous utilisons une minoration certainement trop forte (Equation (26)). Pour s'en affranchir, il nous faudrait pouvoir obenir une inégalité multivariée de concentration du type

$$P(\sqrt{n}\,\overline{(\mathbf{X}^{a^m})^2} \geq s, \sqrt{n}\,\overline{(\mathbf{X}^{a^{m'}})^2} \geq t).$$

Ce type de problèmes nécessite sans aucun doute une extension du travail de Nourdin et Viens [98].

Une autre approche pour définir un intervalle de confiance meilleur que celui basé sur un théorème central limite pourrait résider dans l'obtention d'un développement d'Edgeworth (à un ordre supérieur à 2) pour un estimateur de l'exposant de Hurst. Assez vraisemblablement, ce problème devrait passer par l'obtention d'un développement d'Edgeworth de statistiques de type H_2-variations d'un processus gaussien stationnaire.

Classification en termes de variance de signaux de turbulence

Certains signaux issus de la turbulence exhibent plus ou moins un caractère monofractal connu, en général 1/3. Si l'on modélise ce type de signaux par un mouvement Brownien fractionnaire, le problème d'estimation qui en résulte est l'estimation du coefficient σ^2 à exposant de Hurst fixé. Les techniques asymptotiques et celles utilisant les inégalités de concentration peuvent être adaptées pour ce problème. En collaboration avec Brani Vidakovic (Georgia Technology) et Gabriel Katul (Duke University), nous avons pour projet de classifier des signaux de turbulence en termes de variance. Un article est en préparation sur ce thème.

Identification d'un MBM bruité avec application pour des signaux EEG

Dans le cadre d'une collaboration avec Sky Lee (Georgia Technology) et Brani Vidakovic, nous avons modélisé des signaux d'Electro-encéphalographie (EEG) par un MBM contaminé par un bruit aléatoire. Nous avons adapté la méthodologie des variations discrètes locales. Nous espérons pouvoir établir des propriétés asymptotiques à partir des travaux de [14] et [12]. Un article est en préparation sur ce thème.

3.2 Outils d'analyse fractale multivariée pour l'analyse d'IRMfs

Ce paragraphe présente succinctement le contexte et les axes de recherche d'ordre méthodologique du projet ANR jeunes chercheurs que nous venons d'obtenir (début en septembre 2010). Ce projet nommé "InfoNetComaBrain" s'intitule : méthodes statistiques pour l'étude des réseaux de connectivité fonctionnelle cérébrale, fusion avec la connectivité anatomique. Vers un nouvel outil diagnostique et pronostique pour l'évaluation des désordres de la conscience. Il est à l'interface de plusieurs domaines de compétence : traitement statistique du signal, analyse et visualisation de réseaux complexes, neurosciences. Ce projet fédère le GIPSA-lab (S. Achard, porteur du projet), le LJK (JF Coeurjolly), le Grenoble Institut des Neurosciences (C. Delon-Martin) et le CHU de Strasbourg (S. Kremer et F. Schneider).

Contexte

Le cerveau demeure aujourd'hui encore un sujet de fascination et d'interrogations pour la science et la compréhension de son fonctionnement constitue un enjeu formidable. Le cerveau humain est clairement un *système complexe*, constitué de plus de 10^{11} neurones, organisés en structures interconnectées, qui forment un réseau gigantesque, dont le fonctionnement en boucle fermée est

essentiellement dynamique et non-linéaire, réseau qui réagit aux stimuli extérieurs, possède une mémoire et une structure propre à chaque individu. De plus, c'est un organe vivant, capable d'évoluer au cours de processus d'apprentissage, du vieillissement ou de conditions pathologiques. Par conséquent, il est difficile, voire illusoire, de comprendre son fonctionnement global par l'analyse de petites structures isolées, et sur un seul individu. La recherche sur le cerveau s'est développée par phases successives. Les premières recherches par dissection et coloration sélective ont permis de décrire anatomiquement le cerveau. Le fonctionnement détaillé des neurones ou de petits réseaux a ensuite pu être mis en avant grâce aux travaux de neurophysiologie via l'utilisation de micro-électrodes. Plus récemment, l'Imagerie par Résonance Magnétique fonctionnelle (IRMf) a rendu possible l'observation globale du cerveau lors de son fonctionnement, avec une excellente résolution spatiale permettant par exemple l'identification des régions du cerveau impliquées dans une tâche motrice, sensorielle ou cognitive. Ainsi, le résultat d'une expérience d'IRMf fournit une masse d'information considérable constituée de signaux temporels échantillonnés à peu près toutes les secondes et acquis dans environ 50000 voxels (i.e. 32 coupes de 64×64 voxels).

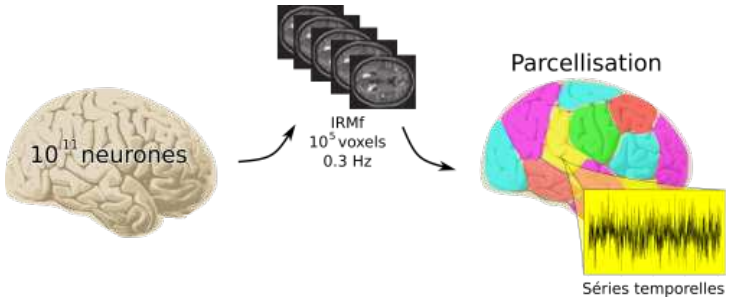


FIGURE 3.1 – Les données résultant d'une expérience d'IRMf correspondent à des images volumiques du cerveau acquises au cours du temps et peuvent aussi être vues comme l'acquisition de signaux temporels acquis dans chaque voxel.

Les contextes d'étude d'expérience d'IRMf sont très diversifiés et les challenges et méthodes statistiques induits par ces expériences sont tout aussi nombreux et variés [79] : détection-estimation de régions activées par un stimulus associé à une tâche particulière par des modélisations temporelles de type GLM (Modèles Linéaires Généralisées) couplées à des procédures de test multiples (de type False Discovery Rate [16]), par des méthodes de type ICA (Analyse en Composantes Indépendantes), par des modélisations spatio-temporelles de type GLM à effets fixes et aléatoires et couplées à des méthodes d'ondelettes ou des méthodes bayésiennes,.... Les comparaisons sujet à sujet constituent également un autre volet impliquant de nouvelles difficultés (renormalisation des cerveaux, analyse de type ANOVA,...). Enfin que cela soit au repos ou sous l'action d'un stimulus, l'analyse de la régularité des signaux et l'observation de la connexion de zones du cerveau par des analyses de type corrélation sont des problèmes majeurs. Notons que le traitement de tous ces problèmes est d'autant plus difficile que la masse de données est importante et leur visualisation relativement complexe. L'acquisition même de ces données nécessite une compréhension des phénomènes physiques mis en jeu et un pré-traitement important (par exemple pour effectuer une éventuelle correction de mouvement du sujet analysé).

Dans le cadre du projet ANR InfoNetComaBrain (qui va débuter en Septembre 2010), nous étudions des patients dans le coma. Certains patients étant dans le coma suite à un accident grave, il s'avère primordial pour les médecins de disposer d'outils de compréhension complémentaires à une technique d'imagerie anatomique ou métabolique (technique de Tomographie par Emission de Positons - TEP) pour qualifier et quantifier l'activité du cerveau et établir les connexions. Parmi

les questions en suspens, on peut lister les suivantes : quelles sont les zones qui sont complètement inactives ? par rapport à un patient sain, les connexions s'établissent-elles de la même façon ou l'accident a-t-il induit une nouvelle organisation des connexions ?

Dans ce contexte d'analyse de sujets au repos, notre objectif est de fournir aux médecins de nouveaux outils de diagnostic à caractère prédictif de ces patients en utilisant des outils d'analyse fractale pour caractériser globalement puis localement les signaux temporels de chaque voxel, les résumer sous forme de paramètres interprétables qui permettraient de comprendre en quel sens une région est plus active qu'une autre, et des outils d'analyse de connectivité statique puis dynamique de régions du cerveau.

▶ Analyse fractale voxel par voxel

Depuis le travail de Zarahn et ses coauteurs [122], il semble admis que le décours temporel d'un voxel d'une expérience d'IRMf d'un sujet au repos exhibe des propriétés de bruit en $1/f$, c'est-à-dire présente des fortes irrégularités, des caractéristiques de type fractal et des caractéristiques longue mémoire. Dans de nombreux travaux, *e.g.* [105] [109], [47], [91], il est alors paru normal de considérer le modèle le plus standard ayant ses caractéristiques : le bruit gaussien fractionnaire. En particulier dans [91], après pré-traitements des données, les auteurs modélisent chaque série temporelle par un bruit gaussien fractionnaire, établissent des cartes d'estimation de H et se servent de cette caractéristique de la série pour discriminer des patients atteints de la maladie d'Alzheimer de patients sains. En utilisant des estimations de l'exposant de Hurst robustes (définis dans la Section 2.2.4), nous avons également utilisé cette approche pour décrire l'activité du cerveau de 11 patients dans le coma [AC10] ainsi que d'une vingtaine de sujets sains (voir Figure 3.2). Cependant cette approche d'analyse de régularité fractale voxel par voxel souffre d'un problème de modélisation spatiale car les décours temporels de deux voxels voisins par exemple sont fortement dépendants l'un de l'autre. Et cette corrélation spatiale est susceptible d'engendrer une mauvaise estimation des intervalles de confiance.

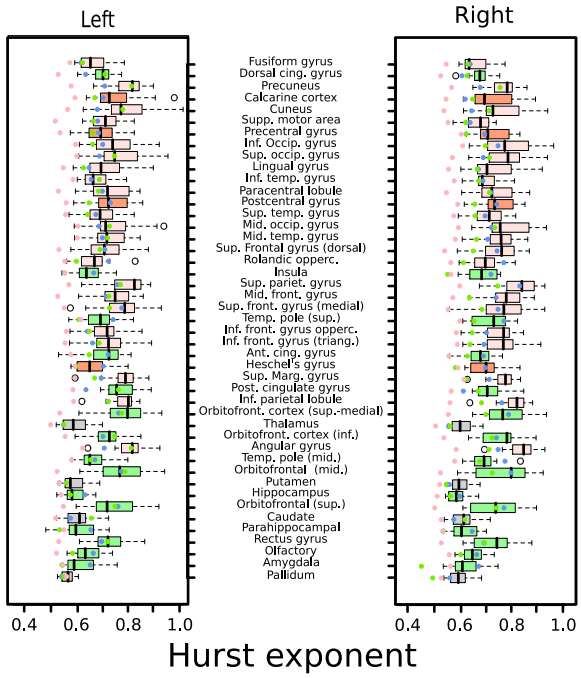


FIGURE 3.2 – Exposants de Hurst calculés dans 90 régions (45 de chaque hémisphère) pour 18 patients sains (boxplots) et 3 patients dans le coma (points).

► Connectivité du cerveau

L'évaluation de la connectivité fonctionnelle dans le cerveau est également un problème majeur, traité ces dernières années en parallèle du problème de détection. Là encore, plusieurs approches ont été proposées (voir [79] pour une revue) : analyse des corrélations simples des signaux, analyse en composantes principales ou indépendantes,... Parmi ces méthodes, la méthode développée dans [2] semble particulièrement efficace et rigoureuse. L'analyse est réalisée en plusieurs étapes : (1) identification des régions d'intérêt du cerveau (2) modélisation de deux décours temporels de deux régions disons A et B par un processus à longue mémoire bivarié [117] (3) Décomposition en ondelettes des deux signaux (de manière à supprimer d'éventuels petits artefacts non stationnaires et à détruire la corrélation des observations) (4) Test d'hypothèses de positive corrélation à une échelle fixée des deux tranformées en ondelettes. L'intérêt d'avoir modélisé les deux décours temporels initiaux trouve tout son sens ici puisque que grâce à cette modélisation, le risque d'erreur de première espèce du test d'hypothèses est contrôlé de manière rigoureuse. (5) Répétition des précédentes étapes pour toutes les paires de régions d'intérêt et construction d'une matrice d'adjacence des corrélations significativement positives. (6) Représentation sous la forme d'un graphe de connexions relativement à cette matrice d'adjacence et détermination de descripteurs topologiques de ce graphe en utilisant entre autres les travaux de Watts et Strogatz [116]. Nous avons également appliqué cette méthodologie à notre application pour des patients dans le coma. En collaboaration avec Ed Bullmore, Sophie Achard et Chantal Delon-Martin, un article est en préparation autour des Figures 3.2 et 3.3.

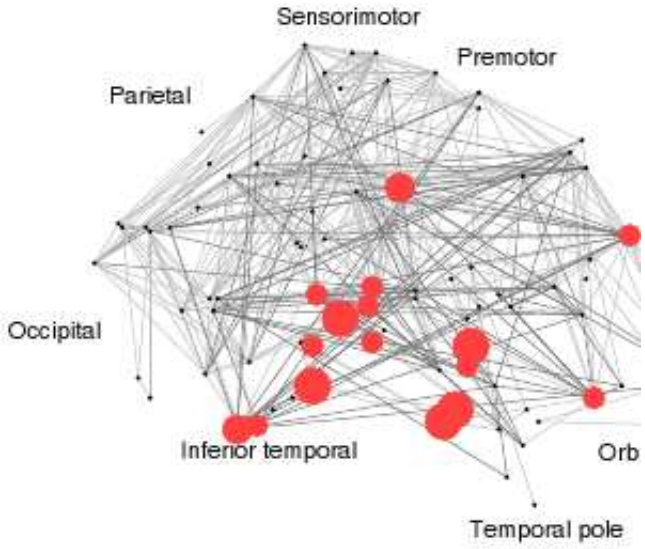


FIGURE 3.3 – Exemple de graphe de connectivité.

Dans [2], les auteurs proposent également de définir le sous-graphe du graphe précédent en ne conservant que les arêtes dont les corrélations partielles (i.e. la corrélation entre deux signaux conditionnés à l'ensemble des autres) sont significativement positives. La méthodologie souffre ici d'un problème de modélisation profond. En effet, pour comparer (de manière rigoureuse) les corrélations et corrélations partielles, il est absolument nécessaire que l'ensemble des signaux des voxels soient modélisés simultanément en particulier pour que l'étape (4) précédente conserve tout son sens rigoureux.

► Approche considérée

La suite du projet repose sur l'utilisation de modèles type MFBM qui devraient permettre de contourner les lacunes évoquées précédemment.

Identification

L'idée sera ici d'étendre les méthodes d'identification disponibles pour le cas uni-dimensionnel. Des méthodes de type décomposition en ondelettes multivariée ou filtrage discret multivarié devraient pouvoir être mises en place en exploitant la propriété auto-similaire du processus et la propriété de décorrélation attendue. Développons un peu nos idées. Rappelons qu'un MFBM est défini au travers des paramètres σ_i, H_i, $\rho_{i,j}$ et $\eta_{i,j}$ (on supposera ici pour simplifier que $H_i+H_j \neq 1$). Puisque chaque composante d'un MFBM constitue un FBM, on peut penser estimer H_i et σ_i en utilisant les techniques déjà développées. Basés alors ces estimations, on peut alors estimer $\rho_{i,j}$ et $\eta_{i,j}$. Une autre stratégie peut consister à utiliser les résultats établis dans la Section 2.4. Ceux-ci suggèrent que la covariance croisée entre deux composantes possède des propriétés autosimilaires (d'ordre $(H_i+H_j)/2$) et peut être estimée efficacement en utilisant des transformées en ondelettes ou variations discrètes. Poursuivons un peu plus notre idée et pour cela définissons la statistique

$$C_{i,j}^a = \overline{\mathbf{X_i}^a \mathbf{X_j}^a}(0), \qquad \text{où } \overline{\mathbf{X_i}^a \mathbf{X_j}^a}(h) := \frac{1}{n-h} \sum_{k=0}^{n-h-1} X_i(k) X_j(k+h).$$

En utilisant la technique des filtres dilatés, on peut montrer qu'il existe $\gamma_{i,j} \in \mathbb{R}$ indépendant de m, tel que $EC_{i,j}^{a^m} = m^{H_i+H_j}\gamma_{i,j}$. Ainsi, par exemple dans le cas $p=2$ (pour simplifier), on peut construire le modèle de régression log-linéaire suivant :

$$\begin{pmatrix} \log \overline{(\mathbf{X_1}^{a^m})^2} \\ \log \overline{(\mathbf{X_2}^{a^m})^2} \\ \log |C_{i,j}^{a^m}| \end{pmatrix} = \begin{pmatrix} 2H_1 \\ 2H_2 \\ H_1+H_2 \end{pmatrix} \log m + \begin{pmatrix} \log \gamma_{1,1} \\ \log \gamma_{1,1} \\ \log |\gamma_{1,2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,1}^m \\ \varepsilon_{2,2}^m \\ \varepsilon_{1,2}^m \end{pmatrix}.$$

Etudier numériquement, en simulation et asymptotiquement l'estimation des exposants de Hurst par moindres carrés ordinaires issus du modèle précédent constitue un travail en cours et un papier en préparation.

Construction d'un graphe de connectivité de régions du cerveau

(a) En utilisant les résultats de la partie précédente, une série de tests d'hypothèses pourrait être mise en place pour savoir si deux régions sont significativement positivement corrélées et/ou positivement partiellement corrélées. Un graphe dont les arêtes correspondent à l'acceptation de ce test (après correction due à la multicplicité des tests) pourrait résumer l'ensemble des tests. La stratégie proposée dans [2] est d'utiliser les travaux de Watts et Strogatz [116] pour résumer le graphe de connexions par des descripteurs topologiques tels que la longueur moyenne (arithmétique ou harmonique) du plus court chemin entre deux noeuds du graphe. Dans la modélisation proposée, ce descripteur est une variable aléatoire et en obtenir des propriétés asymptotiques permettrait sans doute de réaliser de manière efficace une comparaison entre différents sujets. Ce dernier point semble particulièrement difficile et l'adapatation dans ce contexte de méthodes Bootstrap pourrait être un point d'entrée plus abordable.

(b) La définition d'un modèle fractal multivarié permet d'obtenir des résultats asymptotiques pour des mesures de dépendance qui peuvent être bien plus complexes que les mesures de corrélation linéaire. Tenter de généraliser le précédent point à des mesures plus axées sur la théorie de l'information telle que l'information mutuelle, les mesures entropiques que nous avons étudiées

dans [CDR07] ou l'entropie de Rényi étudiée (sur le plan statistique) par [3] est une perspective à envisager.

Perspectives à plus long terme

Modèle multivarié avec a priori : dans le cadre du projet ANR, un des objectifs sera d'utiliser une technique de tractographie pour obtenir un a priori anatomique sur les connexions du cerveau. Ces informations pourraient être utiles pour affiner le modèle initial du MFBM et définir un modèle paramétrique spatial sur la matrice des coefficients $\rho_{i,j}$ et $\eta_{i,j}$. Le choix d'un modèle, son identification et les procédures de test en résultant sont naturellement les problèmes qui vont en découler.

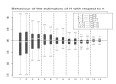
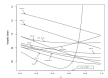
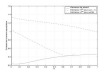
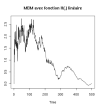
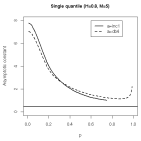
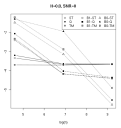
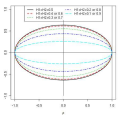
Connectivité dynamique : le modèle considéré suppose que la régularité de chaque composante du mouvement Brownien fractionnaire vectoriel est constante au cours du temps tout comme les corrélations entre les différentes composantes (exprimées par les coefficients $\rho_{i,j}$). Pour aller vers des graphes dynamiques permettant de visualiser les connexions évoluant au cours du temps, il sera indispensable d'assouplir le modèle initial en autorisant les corrélations et régularités à dépendre du temps et définir et étudier un processus analogue au mouvement Brownien multifractionnaire vectoriel.

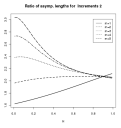
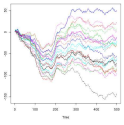
Vers la non-gaussianité : en dimension 1, le processus de Rosenblatt qui reste un processus autosimilaire et à accroissements stationnaires constitue une alternative intéressante au modèle gaussien. Les travaux récents sur ce processus d'un point de vue simulation et identification [1, 115, 27] laissent entrevoir une possibilité de pouvoir définir et étudier un tel processus en multivarié.

Expérience avec stimulation : les données d'IRMf à la base de ce projet de recherche sont des données au repos. Un travail plus conséquent est à fournir pour intégrer dans la modélisation envisagée pour la stimulation (en relation avec une tâche cognitive). Des rapprochements dans le cadre de l'ANR avec P. Ciuciu (LNAO, Neurospin, CEA-Saclay), qui a récemment proposé une modélisation très fine dans un cadre bayésien [83], sont envisagés.

Chapitre 4

Liste des articles en relation avec la Partie I

Articles parus/ à paraître	
	**[Coe00]** J.-F. Coeurjolly. Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *J. Stat. Softw.*, 5(7):1-53, 2000.
	**[Coe01]** J.-F. Coeurjolly. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Infer. Stoch. Process.*, 4(2):199-227, 2001.
	**[CI01]** J.-F. Coeurjolly and J. Istas. Cramèr-Rao bounds for fractional Brownian motions. *Stat. Probab. Lett.*, 53(4):435-447, 2001.
	**[Coe05]** J.-F. Coeurjolly. Identification of multifractional Brownian motion, *Bernoulli*, 11(6):987-1008, 2005. (Erratum 12(2):381, 2006).
	**[Coe08a]** J.-F. Coeurjolly. Bahadur representation of sample quantiles for functional of Gaussian dependent sequences under a minimal assumption. *Stat. Probab. Lett.*, 78(15):2485-2489, 2008.
	**[Coe08b]** J.-F. Coeurjolly. Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. *Annals of Statistics*, 36(3):1404-1434, 2008.
	**[AC10]** S. Achard and J.-F. Coeurjolly. Discrete variations for the fractional Brownian motion in presence of outliers and/or an additive noise. *Statistics Surveys*, 4:117-147, 2010.
	**[ACLP10]** P.O. Amblard and J.-F. Coeurjolly and F. Lavancier and A. Philippe. Basic properties of the Multivariate Fractional Brownian Motion. *to appear in Bulletin de la Société Mathématique de France, Séminaires et Congrés* (available at `http://arxiv.org/abs/1007.0828`), 2010.

Articles soumis	
	**[BC10]** J.-C. Breton and J.-F. Coeurjolly. Confidence intervals for the Hurst parameter of a fractional Brownian motion based on concentration inequalities. *submitted for publication* (available at `http://arxiv.org/abs/0910.3088`, 2010.
	**[CAA10]** J.-F. Coeurjolly and P.O. Amblard and S. Achard. Normalized causal and well-balanced multivariate fractional Brownian motion. *submitted for publication* (available at `http://arxiv.org/abs/1007.2109`), 2010.

Remarques :

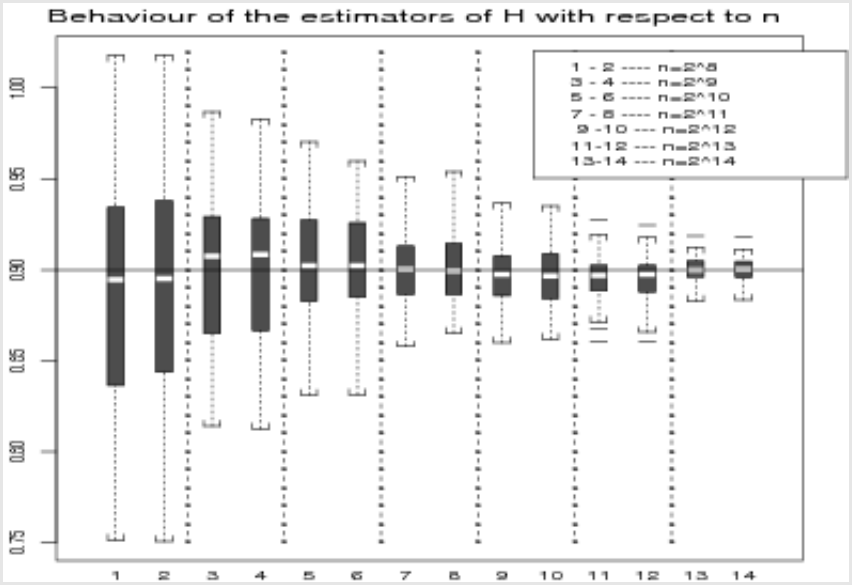
• L'article [Coe05] souffre de plusieurs erreurs (voir paragraphe 2.3) qui ont été corrigées dans [14] (concernant la convergence des variations quadratiques localisées) et dans [12].

• Une version longue du papier [Coe08] (contenant un certain nombre de preuves et de simulations additionnelles) est déposée dans hal : `http://hal.archives-ouvertes.fr/hal-00005371/fr/`.

Chapitre 5

Annex Part I : published and submitted papers

This chapter contains a copy of published or submitted papers related to Part I. These papers are chronologically ordered.



[Coe00]
J.-F. Coeurjolly. Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *J. Stat. Softw.*, 5(7):1-53, 2000.

SIMULATION AND IDENTIFICATION OF THE FRACTIONAL BROWNIAN MOTION: A BIBLIOGRAPHICAL AND COMPARATIVE STUDY

COEURJOLLY Jean-François [1]

Date: June 7, 2000

Abstract : We present a non exhaustive bibliographical and comparative study of the problem of simulation and identification of the fractional Brownian motion. The discussed implementation is realized within the software S-plus 3.4. A few simulations illustrate this work. Furthermore, we propose a test based on the asymptotic behavior of a self-similarity parameter's estimator to explore the quality of different generators. This procedure, easily computable, allows us to extract an efficient method of simulation. In the Appendix are described the S-plus scripts related to simulation and identification methods of the fBm.

Keywords: Fractional Brownian motion, simulation, identification of a parametric model, quality of a generator, S-plus.

[1] LMC-IMAG, BP 53 - 38041 Grenoble cedex 9 - FRANCE, Jean-Francois.Coeurjolly@imag.fr. To whom correspondence should be addressed.

1 Introduction

Many naturally occurring phenomena can be effectively modeled using self-similar processes. For such processes observations that are far apart (in time or space) are correlated too strongly indicating the presence of a long-range dependence. As a result self-similar processes have been used to successfully model data exhibiting long-range dependence and arising in a variety of different scientific fields, including hydrology, see *e.g.* [6], geophysics [12], biology [8], telecommunication networks [28] and economics [24]. The empirical presence of long-memory in such series is found in a local version of the power spectrum which behaves, as $|\lambda|^{1-2H}$, as $\lambda \to 0$, where $H \in]1/2, 1[$ is the long-memory parameter. Among the simplest models that display long-range dependence, one can consider the fractional Brownian motion (fBm), introduced by Kolmogorov in a theoretical context [19], and by Mandelbrot and his co-workers [21] for its statistical applications. The fractional Brownian motion (in short fBm), denoted by $\{B_{H,C}(t), t \geq 0\}$, with parameters $(H, C) \in (]0, 1[\times\mathbb{R}_+^*)$, is the process defined as the fractional integration of a Gaussian pure white noise, or equivalently by the stochastic integral:

$$B_H(t) = C\ V_H^{1/2} \int_{\mathbb{R}} f_t(s) dB(s) \qquad (1)$$
$$\text{with} \quad f_t(s) = \frac{1}{\Gamma(H+1/2)} \left\{ |t-s|^{H-\frac{1}{2}} \mathbb{1}_{]-\infty,t]}(s) - |s|^{H-1/2} \mathbb{1}_{]-\infty,0]}(s) \right\},$$

with $B_H(0) = 0$ and $V_H = \Gamma(2H+1)\sin(\pi H)$. Due to the non stationarity of the fBm and the presence of long-memory, simulation and identification of a fBm is a delicate task. A vast literature has been published on these subjects. A good survey can be found in Beran [6] where historical and statistical aspects are considered. We refer to Adler and *al.* [4] or Taqqu and *al.* [27] for an empirical study on estimation methods.

Through a bibliographical study, we intend to draw up a non exhaustive list of methods for simulating a fBm, and for estimating the self-similarity parameter H. Firstly in Section 2, we recall fundamental properties of fBm: covariance and autocovariance functions, spectral density, Hausdorff dimension. We then describe five simulation methods in Section 3: the method of Mandelbrot and *al.* [21], that of Sellan and *al.* [2], the Choleski method, the Levinson one [24] and finaly the method of Wood and Chan [29]. Section 4 discusses several methods for estimating the self-similarity parameter: spectral methods, maximum likelihood, time-scale methods and temporal methods. Section 5 presents a few simulation results with Boxplots, thus illustrating numerically the notions of Sections 3 and 4. Section 6 explores the quality of pseudo-random generators of fBm. A similar study has been undertaken by Jennane and *al.* [17]: three testing procedures defined independently of the model's identification were considered there. Our approach is slightly different: we provide a theoretical test based on the asymptotic behavior of a parametric estimator of H which allows us to point out a good simulation method of the fBm. Finally we display in the Appendix the S-plus scripts implementing the methods considered in Sections 3 and 4.

2 Properties of fractional Brownian motion

As an alternative way to (1), the fBm can be defined as the unique mean-zero Gaussian process, null at the origin, with stationary and self-similar increments, such that

$$\mathbb{E}\,(\,B_{H,C}(t) - B_{H,C}(s)\,)^2 \;=\; C^2\,|t-s|^{2H}\;, \quad \forall\, s,t \in \mathbb{R}^+\;. \tag{2}$$

Hereafter, we shall call standard fractional Brownian motion, the fBm with scale parameter $C \equiv 1$. Let us briefly review some fundamental results about the fBm. From the self-similarity property, we deduce the covariance and the autocovariance functions, given by

$$\begin{aligned}
\Gamma(t,s) &= \mathrm{Cov}\big(\,B_H(t),\,B_H(s)\,\big) \;=\; \frac{C^2}{2}\,\big(\,|t|^{2H} \;+\; |s|^{2H} \;-|t-s|^{2H}\,\big) &(3)\\
\gamma(t-s) &= \mathrm{Cov}\big(\,B_H(t+1)-B_H(t),\,B_H(s+1)-B_H(s)\,\big)\\
&= \frac{C^2}{2}\big(\,|t-s-1|^{2H} \;-\; 2\,|t-s|^{2H} \;+\; |t-s+1|^{2H}\,\big). &(4)
\end{aligned}$$

In the particular case $H = 1/2$, the fBm is identical to the Brownian motion; consequently $\gamma(k) = 0$, for $|k| \geq 1$. When $H \neq 1/2$, an asymptotic expansion, as $|k| \to +\infty$ exhibits an hyperbolic decrease of $\gamma(\cdot)$:

$$\gamma(k) \sim\; C^2\; H(2H-1)\;|k|^{2H-2}\;, \qquad \text{as } |k| \to +\infty\;. \tag{5}$$

This asymptotic behavior clearly shows that a path deviating from its mean, will have tendency to deviate more when $H > 1/2$, or to return closer to the mean when $H < 1/2$.

The increments process of the fBm is called the fractional Gaussian noise (in short fGn). The fGn constitutes a stationary time-series, and admits a spectral density, defined as the Fourier transform of $\gamma(.)$, explicitly given by

$$f(\lambda) \;=\; 2\,c_\lambda\,(1-\cos\lambda)\,\sum_{j\in\mathbb{Z}} |\,2\pi j + \lambda)\,|^{-1-2H}, \qquad \forall\,\lambda \in [0,2\pi]\;, \tag{6}$$

with $c_\lambda = \frac{C^2}{2\pi}\sin(\pi H)\Gamma(2H+1)$. A Taylor expansion of f near 0 shows that the spectral signature of the fGn is $|\lambda|^{1-2H}$, indicating a pole at zero for $H > 1/2$, a characteristic fact of long-memory processes.

Concerning the paths' regularity, the fBm similarly to the Brownian motion, has continuous and almost surely non differentiable sample paths. The fractal approach refines the difference. Indeed, the Hausdorff dimension of a fBm with parameter $H \in]0,1[$ is almost surely equal to $2-H$, which implies that for $H < 1/2$, the paths are more irregular than those of the Brownian motion and conversely for $H > 1/2$. Figure Fig.1 illustrates this remark.

Finally, let us mention a property of continuity in H, in the sense of Kolmogorov, for the fBm proved by Peltier and Lévy-Véhel [22]:

$$\forall\;\; [a,b] \subset]0,1[\text{ and } K>0, \quad \lim_{h\to 0}\;\sup_{\substack{a\leq H,H'\leq b\\ |H-H'|<h}}\;\sup_{t\in[0,K]}\; |B_{H,C}(t) - B_{H',C}(t)| \;=\; 0\;. \tag{7}$$

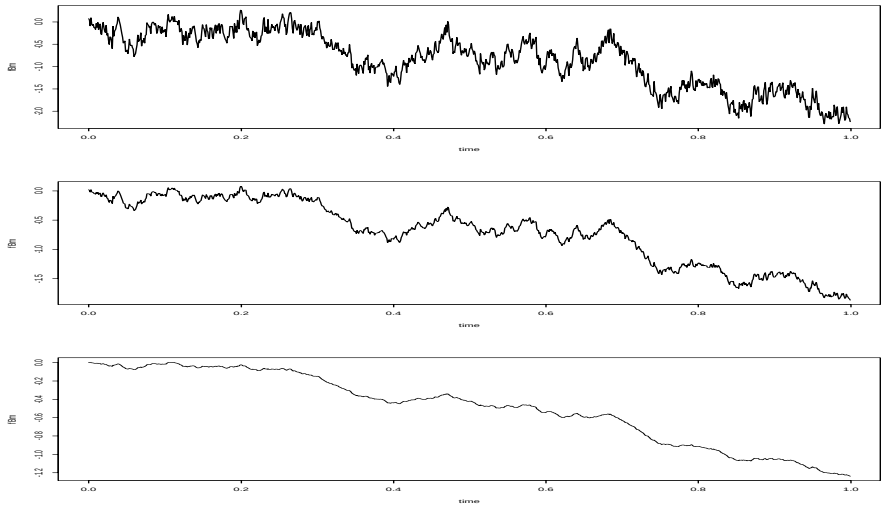


Figure 1: Samples of a fractional Brownian motion on [0,1], simulated by Wood-Chan's method for values $H = 0.3, 0.5, 0.8$ from top to bottom.

3 Simulating the fractional Brownian motion

3.1 Statement of the problem

In this Section, our aim is to describe few methods for simulating a fBm. We adopt the following framework: simulation of a sample of a standard fractional Brownian motion ($C = 1$), of length N at times $i/N,\ i = 0, \ldots, N-1$.

Two approaches are distinguished: the first one consists in using only the properties of the fBm. It gives rise to three methods: the first one is based on a stochastic representation of the fBm [21], the second one consists in extracting the square root of the covariance matrix of the fBm, and the last one consists in the fractional integration of a Gaussian white noise, decomposed on a multiresolution analysis and relies upon a method of Sellan and *al* [2].

The second approach is in fact an indirect way. The idea is to generate a fGn, $\widetilde{X}$, at times i/N, for $i = 0, \ldots, N-1$ and then to define a sample of a fBm via the cumulated sums of $\widetilde{X}$, that is to say to define $\widetilde{B}_H(i/N) = \sum_{k=0}^{i} \widetilde{X}(i/N)$ for $i = 1, \ldots, N-1$ and $\widetilde{B}_H(0) = 0$. The interest to use the increments process rests in its stationarity. The covariance matrix of the fGn at times $i/N,\ i = 0, \ldots, N-1$ is a Toeplitz matrix. The two methods presented (method of Levinson [24], method of Wood-Chan [29]) consist in computing in an exact way the square root of a Toeplitz matrix.

3.2 Stochastic representation of fBm

By considering the fBm's representation of Mandelbrot and Van Ness [21], the first natural idea to simulate a fBm consists in discretizing the stochastic integral (1). We have to approximate the integral

(1) by a Riemann sum truncated to a bound $-a_N \to -\infty$. For $t = 1, \ldots, N-1$, one gets

$$\widetilde{B}_H\left(\frac{t}{N}\right) =$$

$$\frac{V_H^{1/2}}{\Gamma(H+1/2)}\frac{1}{N^H}\left\{\sum_{k=-a_N}^{0}\left[(t-k)^{H-1/2}-(-k)^{H-1/2}\right]B_1(k)+\sum_{k=0}^{t}(t-k)^{H-1/2}B_2(k)\right\},$$

where B_1 (resp. B_2) is a vector of a_N+1 (resp. N) zero-mean standard Gaussian variables independent, and independent of B_2. The choice of a_N results from a compromise between the desired precision and the number of temporal points. In practice, and for the illustration that follows, we have chosen $a_N = N^{1.5}$. This approach is purely historic, and owing to several approximations is not a good way to generate a fBm.

3.3 Method of Sellan, Meyer and Abry

This method has been established by Sellan, Meyer and Abry [2]. Since the fBm is derived by fractional integration of a Gaussian pure white noise, the idea is to start from the decomposition of a pure white noise onto a multiresolution analysis (in short MRA), see *e.g.* Daubechies [9] for generalities on wavelets and MRA:

$$w(t) = \sum_{k\in\mathbb{Z}}\lambda(k)\,\phi_0(t-k) + \sum_{j\geq 0}\sum_{k\in\mathbb{Z}}\gamma_j(k)\,\psi_{j,k}(t)\,, \tag{8}$$

where $\lambda(k)$ and $\gamma_j(k)$ are standard independent Gaussian variables. ϕ_0 is the scaling function and $\{\psi_{j,k}\}_{j\geq 0,k\in\mathbb{Z}}$ the wavelets associated to the MRA. Applying the operator of fractional integration, denoted by D^{-s} (with $s = H+1/2$), to (8) leads to

$$B_H(t) = \sum_{k\in\mathbb{Z}}\lambda(k)\,(D^{-s}\phi_0)(t-k) + \sum_{j\geq 0}\sum_{k\in\mathbb{Z}}\gamma_j(k)\,(D^{-s}\psi_{j,k})(t)\,. \tag{9}$$

The following result, due to Sellan [26], describes explicitly how to integrate fractionally a MRA, and the necessity to introduce biorthogonal wavelets.

Theorem 1 *Let $V_0(\phi_0)$ be an orthogonal MRA of $L^2(\mathbb{R})$ with regularity $r \in \mathbb{N}^*$, ϕ_0 and ψ_0 representing respectively the scale function and the mother wavelet deduced from this analysis. Assume that $s \in]\frac{1}{2}, \frac{3}{2}[$, then*

$$V_0^{(s)} = \{\, f \in L^2(\mathbb{R}), D^s f \in V_0 \,\}\,, \quad \text{and} \quad V_0^{(-s)} = \overline{\{\, f \in L^2(\mathbb{R}), \overline{D^{-s}} f \in V_0 \,\}}\,,$$

define two biorthogonal MRAs, admitting for scale functions

$$\phi_0^{(s)} = U_s(\phi_0) \quad \text{for} \quad V_0^{(s)}(\phi_0)\,, \quad \text{and} \quad \phi_0^{(-s)} = \overline{U}_{-s}(\phi_0) \quad \text{for} \quad V_0^{(-s)}(\phi_0)\,,$$

where $g = U_s(f)$ has for Fourier transform $(i2\pi\nu)^{-s}(1-\exp(i2\pi\nu))^2\,\widehat{f}(\nu)$, and for mother wavelets:

$$\psi_0^{(s)} = 4^s\, D^{-s}(\psi_0)\,, \quad \text{and} \quad \psi_0^{(-s)} = 4^{-s}\,\overline{D}^s(\psi_0)\,.$$

$\overline{D^{-s}}$ denotes the conjugate operator of D^{-s}, and $\overline{E}$, for a set E, is the adherence of E. Under the conditions of Theorem 1, Sellan proves that there exists a Gaussian white noise with variance σ^2, allowing to construct an $ARIMA(0, s, 0)$, denoted by b_H, and for $j \in \mathbb{Z}^+$, a Gaussian discrete white noise, with variance $2^j\sigma^2$, denoted by $(\gamma_j(k))_{k\in\mathbb{Z}}$ such that the restriction of B_H to the interval $]0, T]$, $T > 0$ admits the following decomposition

$$\forall\, t \in]0, T], \qquad B_H(t) - b_0 \;=\; \sum_{k\in\mathbb{Z}} b_H(k)\; \phi_0^{(s)}(t-k) + \sum_{j\geq 0}\sum_{k\in\mathbb{Z}} 4^{-s}\; 2^{-js}\; \gamma_j(k)\; \psi_{j,k}^{(s)}(t)\;. \qquad (10)$$

The computing implementation is then realized in three steps:

1. Estimation of the filters related to $\phi_0^{(s)}$ and $\psi_0^{(s)}$: let $u(k)$ be the filter of the initial MRA related to ϕ_0, and $v(k)$ the associated quadrature mirror filter. The above cited authors show, by denoting $u^{(s)}$ and $v^{(s)}$ (resp. $u^{(-s)}$ and $v^{(-s)}$), the filters associated to $\phi_0^{(s)}$ and $\psi_0^{(s)}$ (resp. $\phi_0^{(-s)}$ and $\psi_0^{(-s)}$), the following relations:

$$\begin{array}{lcllcl} u^{(s)} & = & f^{(s)} * u, & F^{(s)}(z) & = & 2^{-s}(1+z^{-1})^s\;. \\ v^{(s)} & = & g^{(s)} * v, & G^{(s)}(z) & = & 2^{s}(1-z^{-1})^{-s}\;. \\ u^{(-s)} & = & f^{(-s)} * u, & F^{(-s)}(z) & = & 2^{s}(1+z)^{-s}\;. \\ v^{(-s)} & = & g^{(-s)} * v, & G^{(-s)}(z) & = & 2^{-s}(1-z)^{s}\;. \end{array}$$

   where $*$ denotes the convolution product and where $F^{(\epsilon s)}$ (resp. $G^{(\epsilon s)}$) denotes the z-transform of $f^{(\epsilon s)}$ (resp. $g^{(\epsilon s)}$), $\epsilon = \pm 1$. However, a numerical problem appears: for $s = H + 1/2$, the functions $f^{(s)}$ and $g^{(s)}$ have, in general, an infinite support and the series (10) diverges. To avoid this problem, the following approximations are proposed:

$$\begin{array}{lcllcl} u^{(s)} & = & u * f^{(1)} * tf^{(d)}\;, & u^{(-s)} & = & -\delta_{-1} * (\widetilde{v}^{(s)})^{\vee}\;, \\ v^{(s)} & = & v * g^{(1)} * tg^{(d)}\;, & v^{(-s)} & = & \delta_1 * (\widetilde{u}^{(s)})^{\vee}\;, \end{array}$$

   where $d = s - 1$ and where $tf^{(d)}$ and $tg^{(d)}$ are versions of $f^{(d)}$ and $g^{(d)}$ truncated up to an order chosen a priori.

2. Simulation of b_H: the process $ARIMA(0, s, 0)$ results from the convolution of a Gaussian white noise and a filter with impulsive response defined by:

$$\alpha_k^{(s)} \;=\; \sum_{p=0}^{k} (-1)^p \binom{-s}{p} \quad \text{with} \binom{-s}{p} = \frac{\Gamma(-s+1)}{\Gamma(p+1)\Gamma(-s-k+1)}$$

   Since the same numerical difficulty noted previously appears, the following approximation is proposed: $b_H = \gamma_j * \alpha^{(1)} * t\alpha^{(d)}$, where $t\alpha^{(d)}$ is a version of $\alpha^{(d)}$ truncated up to an order chosen a priori.

3. One then truncates the series (10) at some resolution J to get,

$$B_H^J(t) \;=\; \sum_{k\in\mathbb{Z}} b_H(k)\; \phi_{-J,k}^{(s)}(t) \;+\; \sum_{j=0}^{J}\sum_{k\in\mathbb{Z}} \gamma_j(k)\; 4^{-s}\; 2^{-js}\; \psi_{j,k}^{(s)}(t)\;. \qquad (11)$$

One can now use the pyramidal algorithm of Mallat adapted to biorthogonal wavelets, see Daubechies [9]: decomposition with the help of filters $u^{(-s)}$ and $v^{(-s)}$, and synthesis with the help of $u^{(s)}$ and $v^{(s)}$. The diagram in Fig.2 illustrates these operations. To remain coherent, getting a sample of length N through a resolution J needs to generate a fBm on a duration $T = 2^{-J}N$; the self-similarity property is used to get a sample at times i/N, for $i = 0, \ldots, N-1$. In practice, we have, as Abry and Sellan recommend, chosen a resolution of 6 or 7, and used the filters of a Daubechies wavelet of order 20, for its regularity properties.

Caption

$\triangleright$ $[\uparrow 2]$: dilatation operator defined by $[\uparrow 2]x_k = x_{2k}$.

$\triangleright$ b_H : $ARIMA(0, s, 0)$ simulated.

$\triangleright$ $g_0, \ldots, g_{J-1}$: simulated standard Gaussian variables.

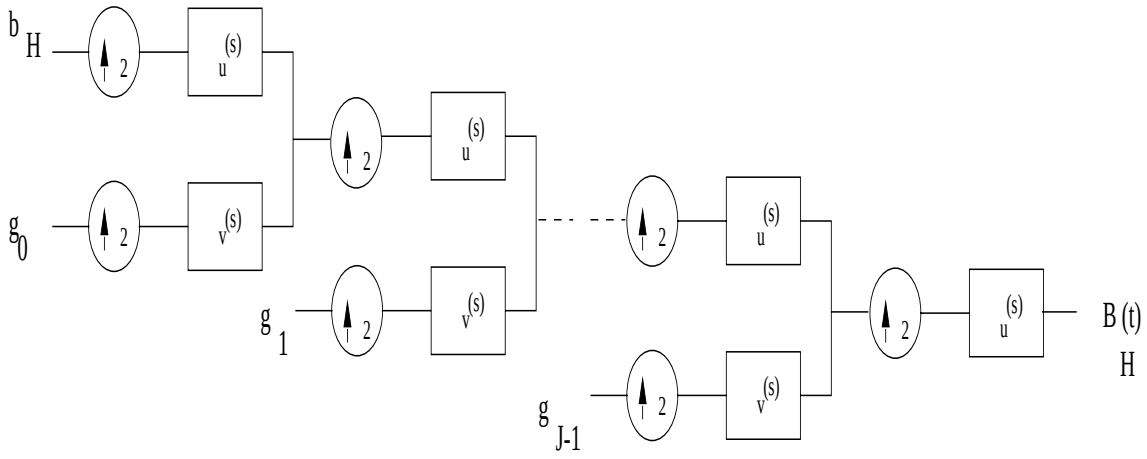


Figure 2: Diagram for the simulation of a fBm by wavelet synthesis.

3.4 Method of Choleski

Let Γ be the covariance matrix of a standard fBm, discretized at times i/N, for $i = 0, \ldots, N-1$. From (3),

$$(\Gamma)_{i,j} \;=\; \Gamma\left(\frac{i}{N}, \frac{j}{N}\right), \quad \text{for } i, j = 0, \ldots, N-1.$$

Define Γ' as the matrix Γ deprived of its first row and its first column. Since Γ' is a symmetric definite positive matrix, it admits a Choleski decomposition $\Gamma' = LL^t$, where L is a lower triangular matrix.

Thus simulating a sample of a fBm at times i/N for $i = 1, \ldots, N-1$ is equivalent to generate a vector Z of $(N-1)$ standard independent Gaussian variables and apply the product LZ. Indeed, LZ is a centered Gaussian vector and $\mathbb{E}\big((LZ)(LZ)^t\big) = \Gamma'$. Define $\widetilde{B} = \big(0, (LZ)^t\big)^t$, $\widetilde{B}$ is a sample of a fBm at times i/N for $i = 0, \ldots, N-1$. This method is the only one exact in theory, but due to a computational complexity of order $\mathcal{O}(N^3)$ and to the fact that Γ' is extremely ill conditioned, it is of interest to derive methods that are less computational demanding.

3.5 Method of Levinson

Let G be the autocovariance matrix of a standard fBm, discretized at times i/N, for $i = 0, \ldots, N-1$. From (4),

$$(G)_{i,j} \;=\; \gamma(j-i), \quad \text{for } i, j = 0, \ldots, N-1.$$

To avoid the computation of the Choleski decomposition of G (which would lead to a method identical to the one presented in Section 3.4), it suffices to remark that G is a Toeplitz matrix. The Toeplitz nature means that the first row of G suffices to reconstruct G, which leads to advanced algorithms to extract the square root of G. The following one can be found in [24].
At step 1 define:

$$\longrightarrow\; k_1 \;=\; -\gamma(\tfrac{1}{N})\,,\; \sigma_1^2 \;=\; 1\,,$$

$$\longrightarrow\; L^1 \;=\; \big(\,1,\; \gamma\left(\tfrac{1}{N}\right),\; \ldots,\; \gamma\left(\tfrac{N-1}{N}\right)\big)^t \;\text{ and }\; \widehat{L}^1 \;=\; \big(\,0,\; \gamma\left(\tfrac{1}{N}\right),\; \ldots,\; \gamma\left(\tfrac{N-1}{N}\right)\big)^t\,.$$

Then define vectors $L^j \;=\; \sigma_j^2 \left(\,0,\; \ldots,\; 0,\; 1,\; \ell_2^j,\; \ldots,\; \ell_{N-j+1}^j\,\right)^t$ and
$\widehat{L}^j \;=\; \sigma_j^2 \left(\,0,\; \ldots,\; 0,\; 0,\; \hat{\ell}_2^j,\; \ldots,\; \hat{\ell}_{N-j+1}^j\,\right)^t$.
At step $j+1$, one has:

$$\longrightarrow\; k_{j+1} \;=\; \tfrac{-\hat{\ell}_1^j}{\sigma_j^2} \quad \sigma_{j+1}^2 \;=\; \sigma_j^2(1-k_{j+1}^2)\,,$$

$$\longrightarrow \begin{pmatrix} \widehat{L}^{j+1} \\ L^{j+1} \end{pmatrix} = \begin{pmatrix} I_N & k_{j+1}\, Z \\ k_{j+1}\, I_N & Z \end{pmatrix} \begin{pmatrix} \widehat{L}^j \\ L^j \end{pmatrix}, \quad \text{where } Z = \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}.$$

Let us denote by D the matrix $D = diag(\sigma_1, \ldots, \sigma_N)$ and let $L = (L^1, \ldots, L^N)\; D^{-1}$. One can check that $LL^t = G$. Denoting by Z a vector of N standard independent Gaussian variables, the vector (LZ) defines a sample of a fGn at times i/N, for $i = 0, \ldots, N-1$. The cumulated sums of this sample define a sample of a fBm, $\widetilde{B}_H$ at the desired times (setting moreover $\widetilde{B}_H(0) = 0$). This method generates exactly a fGn with a computation cost $\mathcal{O}(N^2 \log(N))$ but still remains particularly slow within S-plus, as soon as $N > 1000$.

3.6 Method of Wood-Chan

Initially proposed by Davis and Harte [10], this method, available for any stationary Gaussian process, has been recently improved by Wood and Chan [29]. In order to extract the square root of the autocovariance matrix G, the idea is to embed G in a circulant matrix C, of size $m = 2^g$, $g \in \mathbb{N}^*$, and then to generate a vector $Y = (Y_0, \ldots, Y_{m-1})^t \rightsquigarrow \mathcal{N}(0, C)$, and thanks to an appropriate construction of C, to generate $(Y_0, \ldots, Y_{N-1})^t \rightsquigarrow \mathcal{N}(0, G)$. Let C be the matrix defined by:

$$C = \begin{pmatrix} c_0 & c_1 & \ldots & c_{m-1} \\ c_{m-1} & c_0 & \ldots & c_{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & \ldots & c_0 \end{pmatrix} \quad \text{where} \quad c_j = \begin{cases} \gamma(\frac{j}{N}) & \text{if} \quad 0 \leq j < \frac{m}{2} \\ \gamma(\frac{m-j}{N}) & \text{if} \quad \frac{m}{2} < j < m-1 \, . \end{cases}$$

By construction, C is symmetric and circulant. One chooses m the first power of two, for questions of algorithmic rapidity, $m \geq 2(N-1)$ such that C is definite positive. The authors suggest an approximation when this condition can not be fulfilled. For a fBm, this condition is satisfied for the value $m = 2 * 2^\nu$, where 2^ν is the first power of two superior to N. Then, in order to diagonalize C, one uses a result of Brockwell and Davis: C can be decomposed as $C = Q \Lambda Q^*$, where Λ is the diagonal matrix of eigenvectors of C, and Q is the unitary matrix defined by

$$(Q)_{j,k} = m^{-1/2} \exp\left(-2i\pi \frac{jk}{m}\right) , \qquad \text{for} \quad j,k = 0, \ldots, m-1 \, .$$

Because Q is unitary, if $Y = Q\Lambda^{1/2} Q^* Z$ with $Z \rightsquigarrow \mathcal{N}(0, I_m)$, one has $Y \rightsquigarrow \mathcal{N}(0, C)$. Thus the simulation procedure of a fBm's sample at times i/N, for $i = 0, \ldots, N-1$ reduces itself to the three following steps:

1. Estimation of the eigenvalues of C: a matrix calculus shows that

$$\lambda_k = \sum_{j=0}^{m-1} c_j \exp\left(-2i\pi \frac{jk}{m}\right) = \sum_{j=0}^{m-1} c_j \exp\left(2i\pi \frac{jk}{m}\right) , \qquad \text{for } k = 0, \ldots, m-1 \, .$$

   This estimation may be done using the Fast Fourier Transform (direct or inverse).

2. Fast simulation of Q^*Z: by decomposing Q^*Z into real and imaginary parts, simulating of $W = Q^*Z$ amounts to the two following substeps:

   $\longrightarrow$ generate U, V two independent normal variables, $\mathcal{N}(0,1)$, and write $W_0 = U$ and $W_{\frac{m}{2}} = V$.

   $\longrightarrow$ for $1 \leq j < \frac{m}{2}$, generate U_j, V_j two independent normal variables, $\mathcal{N}(0,1)$ and write

$$\begin{aligned} W_j &= \frac{1}{\sqrt{2}}(U_j + iV_j) \, , \\ W_{m-j} &= \frac{1}{\sqrt{2}}(U_j - iV_j) \, . \end{aligned}$$

3. Reconstruction of X: the last step consists in calculating

$$X\left(\frac{k}{N}\right) \;=\; \frac{1}{\sqrt{m}} \sum_{j=0}^{m-1} \sqrt{\lambda_j}\; W_j \;\exp\left(-2i\pi\,\frac{jk}{m}\right), \qquad \text{for } \; k = 0, \ldots, m-1 \; .$$

using again the FFT. We get a sample of a fBm, denoted $\widetilde{B}_H$ by evaluating cumulated sums of the vector $\left\{X(0), X(\frac{1}{N}), \ldots, X(\frac{N-1}{N})\right\}$ and setting moreover $\widetilde{B}_H(0) = 0$.

The method of Wood and Chan is exact for simulating a fGn, has a complexity of $N\log(N)$, and in a practical point of view is fast even for large values of N.

3.7 About the approximation of a fBm via the cumulated sums of a fGn

Let $\widetilde{B}_H$ denote the vector defined by:

$$\begin{aligned} \widetilde{B}_H(0) &= 0 \\ \widetilde{B}_H(i/N) &= \sum_{k=0}^{i} \widetilde{X}(k/N), \quad i = 1, \ldots, N-1 \; , \end{aligned}$$

where $\widetilde{X}$ denotes a sample of a fGn generated at times $i/N, i = 0, \ldots, N-1$. Recall that, $\mathbb{E}\left(\widetilde{X}(i/N)\widetilde{X}(j/N)\right) = \gamma\left((j-i)/N\right),\ i, j = 0, \ldots, N-1$. So, it is easy to see that

$$\mathbb{E}\left(\widetilde{B}_H\,(i/N)\,\widetilde{B}_H\,(j/N)\right) \;=\; \sum_{k=0}^{i}\sum_{l=0}^{j} \gamma\left((j-i)/N\right), \quad i, j = 1, \ldots, N-1.$$

Thus, one can discuss the approximation of a sample of a fBm via the cumulated sums of a sample of a fGn by estimating the relative error done on the second order structure, via the function E defined on $\{0, \ldots, N-1\}^2$ by

$$E(i,j) \;=\; \begin{cases} 0 & \text{if } i \text{ or } j = 0 \\ \left|\left(\Gamma(i,j) - \sum_{k=0}^{i}\sum_{l=0}^{j}\gamma(j-i)\right)/\,\Gamma(i,j)\right| & \text{otherwise.} \end{cases}$$

Figure Fig.3 displays the function E computed for different values of H. It is clear that such an approximation involves negligible mistakes on the covariance structure.

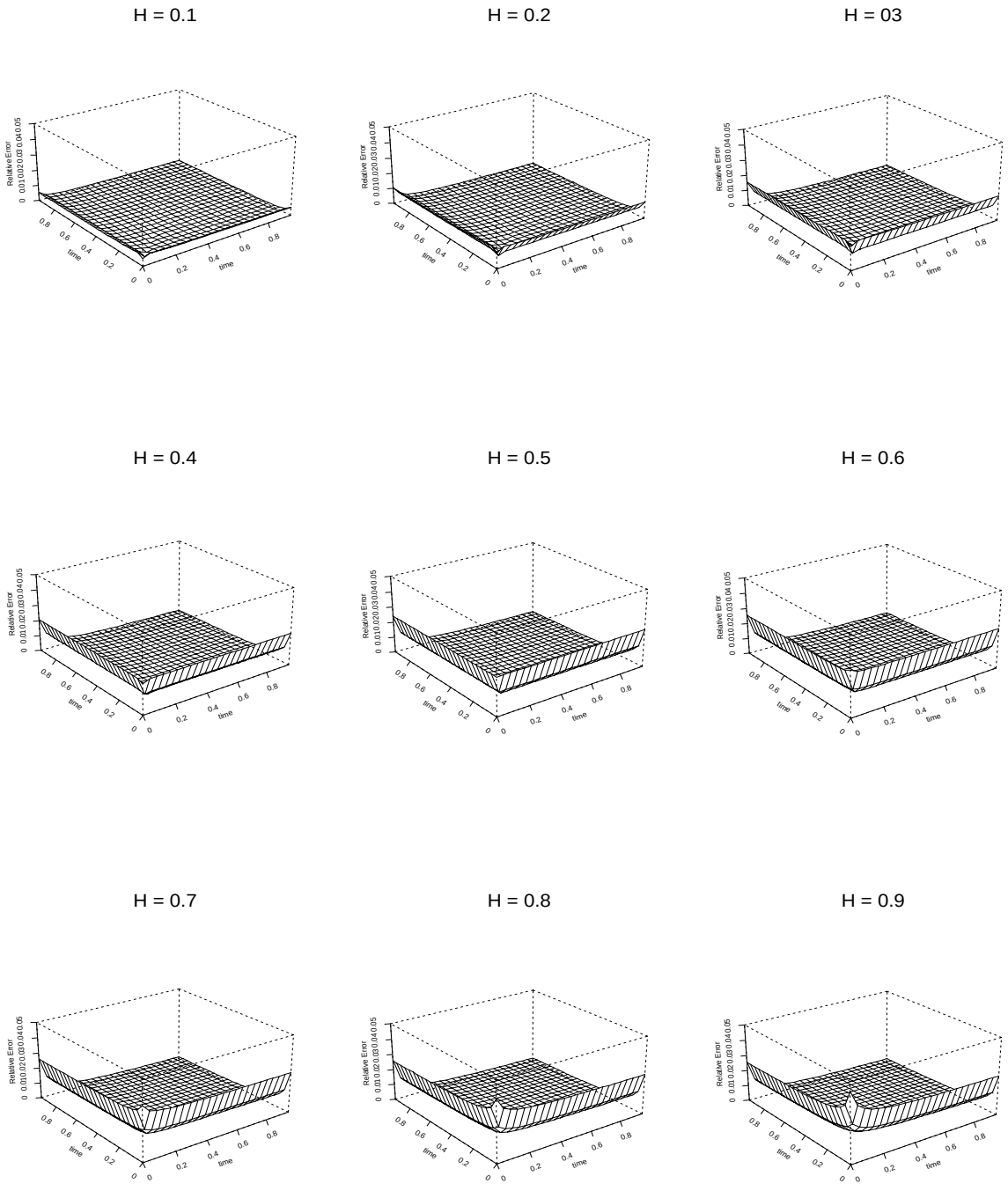


Figure 3: Relative Error on the covariance structure of a fBm approximated via the cumulated sums of a fGn, for different values of H.

4 Identification of the fractional Brownian motion

4.1 Statement of the problem

The irregularity's analysis of data modelled by a fBm, the study of its spectral behavior, and any forecasting problem based on fBm imply the necessity to estimate the Hurst parameter. In this Section, we briefly describe the main parametric methods to estimate the self-similarity parameter. We distinguish four approaches:

1. Spectral methods: log-periodogram, a variant of Lobato and Robinson's method.
2. Maximum likelihood: Whittle's estimator.
3. Time-scale methods: wavelet decomposition of the fBm.
4. Temporal methods: number of level crossings, discrete variations.

This list is not exhaustive (the R/S method or correlogram's approach has not been considered here, see *e.g.* Beran [6]) but presents the different ways of tackling the identification that have been discussed in the literature recently. Later on, we address the same problem when observing a sample $(B_{H,C})$ of a non-standard fBm of length N at times i/N for $i = 0, \ldots, N-1$; (X) will denote the sample of the increments of $(B_{H,C})$. Unless otherwise stated, the scale coefficient C is supposed to be unknown.

4.2 Spectral methods

4.2.1 Log-periodogram

This approach consists in exploiting, on the one hand, the spectral signature of the fGn, $f(\lambda) \sim c_f|\lambda|^{1-2H}$, as $|\lambda| \to 0$, and on the other hand the fact that the periodogram defined by

$$I_N(\lambda) = \frac{1}{2\pi N}\left|\sum_{t=0}^{N-1} X(t)\, e^{-it\lambda}\right|^2, \qquad \text{for } \lambda = \lambda_{k,N} = \frac{2\pi k}{N},$$

is an asymptotical unbiased estimator of the spectral density. One immediately notices that

$$\log \mathbb{E}\big(I_N(\lambda)\big) \simeq \log c_f + (1-2H)\log(|\lambda|),$$

pointing out the linearity in H of $\log \mathbb{E}(I_N(\lambda))$ in a neighborhood of 0. Let $1 \le m_1 < m_2 \le N^* = [N - 1/2]$. Define $\widehat{\alpha}_N$ the estimator deduced from the linear regression of $\big\{\log(I_N(\lambda_k))\big\}_{m_1\le k\le m_2}$ on $\big\{\log(\lambda_k)\big\}_{m_1\le k\le m_2}$, one gets an estimator of the self-similarity parameter by the equation

$$\widehat{H}_N(m_1, m_2) = \frac{1}{2}(1 - \widehat{\alpha}_N).$$

From a theoretical point of view, if m_1 and m_2 satisfy $\sqrt{m_2}\log(m_2)/m_1 \; + \; m_1\log^2(N)/m_2 \; \to \; 0$, Geweke and Porter-Hudak [13] prove the asymptotic normality for $\widehat{H}_N(m_1, m_2)$:

$$\frac{\sqrt{m_2}}{\log(N)} \left(\widehat{H}_N(m_1, m_2) - H \right) \; \xrightarrow{d} \; \mathcal{N}\left(0, \frac{\pi^2}{6}\right) . \tag{12}$$

4.2.2 Variant of Lobato and Robinson

By denoting $F(\lambda) = \int_0^\lambda f(\theta)d\theta$, Lobato and Robinson [20] remark that there exists, in a neighborhood of 0, a log-linear relation between two values of $F(\lambda)$. Let $q \in]0, 1[$, one gets immediately, $\frac{F(q\lambda)}{F(\lambda)} \sim q^{2-2H}$, as $|\lambda| \to 0$. By estimating $F(\lambda_k)$ by

$$\widehat{F}(\lambda_k) \; = \; \frac{2\pi}{N} \sum_{j=1}^{[N\lambda_{k,N}/2\pi]} I(\lambda_j), \qquad \text{for } \; k = m_1, \ldots, m_2 \; ,$$

one deduces an estimator of H:

$$\widehat{H}_N(q, m_1, m_2) \; = \; 1 \; - \; \frac{1}{2 \; \log(q)} \; \log\left\{ \; \frac{\widehat{F}(q\lambda_{m_2})}{\widehat{F}(\lambda_{m_2})} \; \right\} \; .$$

For $1/2 < H < 3/4$, Lobato and Robinson exhibit the optimal value of q via simulations. Let us mention that if one chooses m_1 and m_2 as previously, one may obtain an asymptotic normality result similar to (12).

4.3 Maximum likelihood: Whittle's estimator

Applied to our framework, the method consists in maximizing the log-likelihood of the fGn with parameter $\theta = \left(\frac{\sigma_\epsilon^2}{2\pi}, \; H \right)$, where σ_ϵ^2 is the innovation variance, given by:

$$L_N(x, \theta) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|det(G)| - \frac{1}{2}x^t G^{-1} x \; ,$$

where $G = G(\theta)$ is the covariance matrix of the fGn sample. The computational implementation points out two numerical problems: firstly the estimations of G^{-1} and $\log|det(G)|$ are particularly slow and expensive, and secondly G is extremely ill-conditioned. To get around these problems an approximation of the likelihood has been proposed, see *e.g.* Beran [6], giving rise to a new estimator called Whittle's estimator and explicitly given by:

$$\widehat{H}_N \; = \; \arg\min_H \sum_{j=1}^{N^*} \frac{I_N(\lambda_{j,N})}{f(\; \lambda_{j,N}, (1, H) \;)} \; ,$$

where $f(., (1, H))$ denotes the spectral density with parameters $(1, H)$ of the fGn, and $I_N(\lambda)$ the empirical periodogram. It is well-known, see *e.g.* Beran [6], that $\widehat{H}_N$ tends almost surely to H, and verifies the following asymptotic normality result:

$$\sqrt{N} \left(\widehat{H}_N - H \right) \; \xrightarrow{d} \; \mathcal{N}\left(0, \frac{2}{D} \right) , \qquad \text{with} \quad D = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left\{ \frac{\partial}{\partial H} \log f(x, \theta) \right\}^2 dx \; .$$

Despite these performances, the Whittle's estimator suffers from being slowly executable, biased for finite samples, and very sensitive to perturbation with an additive white noise.

4.4 Time-scale method: wavelet decomposition.

Let $\{\psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j}t - k),\ j = 1, \ldots, J, k \in \mathbb{Z}\}$ be the family of wavelet basis functions, generated from the mother wavelet ψ_0, itself defined via a multiresolution analysis of L^2, see [9]. We denote $< B_{H,C}, \psi_{j,k} >$ the coefficients of the discrete wavelet transform. Two reasons can explain the use of wavelet decomposition for the fBm identification:

1. Self-similarity of the wavelet coefficients: the self-similarity of the fBm ensures that the variance of the wavelet coefficients can be written

$$\mathbb{E}\big(< B_{H,C}, \psi_{j,k} >^2 \big) = \frac{1}{2} 2^{j(2H+1)} \frac{C^2}{N^{2H}} \int_{\mathbb{R}^2} |u - v|^{2H} \psi(u)\psi(v) du dv = K_H\ 2^{j(2H+1)} ,$$

   where $< \cdot, \cdot >$ denotes the inner product of $L^2(\mathbb{R})$.

2. The wavelet transform decorrelates the sample of the fBm. Indeed, if M denotes the number of vanishing moments of the mother wavelet, Flandrin and *al* [1], for example, show that

$$\mathbb{E}\big(< B_{H,C},\ \psi_{j,k} > < B_{H,C},\ \psi_{j',k'} > \big) \ = \ \mathcal{O}\big(|2^j k - 2^{j'} k'|^{2(H-M)} \big) .$$

Similarly to spectral methods, one notices the linearity in H of $\log_2 \mathbb{E}(< B_{H,C}, \psi_{j,k} >^2)$,

$$\log_2\ \mathbb{E}\big(< B_{H,C}, \psi_{j,k} >^2 \big) \ = \ j(2H + 1) \ + \ \log_2(K_H) .$$

One estimates $\mathbb{E}(< B_{H,C}, \psi_{j,k} >^2)$ by the empirical moment of order 2,

$$\mu_j \ = \ 2^{-j} \sum_{k=0}^{2^j - 1} < B_{H,C}, \psi_{j,k} >^2 .$$

Then, an estimator of H is deduced from a linear regression of $\big\{ \log_2(\mu_j) \big\}_{j_1 \le j \le j_2}$ on $\{j\}_{j_1 \le j \le j_2}$, where $[j_1, j_2]$ represents the used resolutions. Abry and *al* [3] have recently improved this method. Let ξ be the noise vector defined by

$$\xi_j \ = \ \log_2(\mu_j) - (2H + 1)j - \log_2(K_H) , \quad \text{for } j = j_1, \ldots, j_2$$

the above cited authors show that

$$\mathbb{E}(\ \xi_j\ ) \ = \ \frac{\Psi(2^{j-1})}{\log 2} \ - \ (j - 1) \quad \text{and} \quad \mathrm{Var}(\ \xi_j\ ) \ = \ \frac{\zeta(2, 2^{j-1})}{(\log 2)^2} ,$$

where Ψ is the Digamma function defined by $\Psi(t) = \frac{\Gamma'(t)}{\Gamma(t)}$, and $\zeta(2, t)$ is the generalized Riemann Zeta function. A new estimator of H is then deduced from a linear regression of $\big\{ \log_2(\mu_j) - \mathbb{E}(\xi_j) \big\}_{j_1 \le j \le j_2}$ on $\{j\}_{j_1 \le j \le j_2}$, weighted by $\big\{ \mathrm{Var}(\xi_j) \big\}_{j_1 \le j \le j_2}$.

In practice, the S-plus library `Wavethresh()` has been used to estimate the wavelet coefficients of the fBm.

4.5 Temporal methods

4.5.1 Statistic related to level 0 crossings of fGn

The smooth characteristic of the fGn covariance function ensures the convergence of the local time (see Azaïs [5]). Feuerverger and *al* [11] are based on this remark to estimate the Hölder exponent of a non differentiable Gaussian process, by counting the number of crossings of a given level. We present here a simplified version of this estimator. Let us define the mean number of 0 crossings of the fGn's sample (X) by the statistic

$$S_N \;=\; \frac{1}{N-1}\sum_{i=0}^{N-2} \mathbb{1}\!\left(X(i/N)X(i+1/N) < 0 \right) ,$$

where $\mathbb{1}(.)$ denotes the indicator function. The ergodicity of the increments and an integral calculus show that: $S_N \xrightarrow{a.s.} \theta/\pi$,

$$\text{where} \qquad \theta = \begin{cases} \arctan((1-r^2)^{1/2}/r) & \text{if } r > 0 \\ \pi/2 + \arctan(-r/(1-r^2)^{1/2}) & \text{otherwise,} \end{cases} \qquad \text{and} \quad r \;=\; 2^{2H-1} - 1 \; .$$

Assuming that one knows $\epsilon = \mathrm{sgn}(H - \frac{1}{2})$, one deduces immediately the following estimator

$$\widehat{H}_N \;=\; \frac{1}{2}\left\{ 1 + \log_2\left(1 + \epsilon |\cos(\pi S_N)| \right) \right\} .$$

Let us mention that for $0 < H < 3/4$, $\widehat{H}_N$ is asymptotically Gaussian with a rate of convergence $1/\sqrt{N}$, see Ho and Sun [15].

4.5.2 Discrete variations of the fBm

This method is relatively recent: the first results are due to Istas and Lang [16], on the one hand, and Kent and *al* [18], on the other hand. It relies on a specific filtering of the sample of a fBm that is designed to destroy long-range dependence of observations. In [7], we have generalized these results: the convergence of the k-th absolute moment of discrete variations, defined by

$$S_N(k,a) \;=\; \frac{1}{N-\ell}\sum_{i=\ell}^{N-1} \left| V^a(i/N) \right|^k , \qquad \text{for } k > 0$$

is studied. The parameter a denotes a filter of length $\ell+1$ and of order $p \geq 1$, verifying $\sum_{q=0}^{\ell} a_q \, q^r = 0$, for $r = 0, \ldots, p-1$, and (V^a) denotes the vector derived from filtering the vector $(B_{H,C})$ with the vector a,

$$V^a\left(\frac{i}{N}\right) \;=\; \sum_{q=0}^{\ell} a_q \; B_{H,C}\left(\frac{i-q}{N}\right) , \qquad\qquad \forall\, i \in \{\ell, \ldots, N-1\} \; .$$

We provide two classes of estimators of H: the first one, assuming that the scale coefficient is known, is convergent at a rate $1/\sqrt{N}\log(N)$, for $0 < H < 1$; the second one, without any assumption on C

is convergent at a rate $1/\sqrt{N}$.

Scale parameter known:

Without loss of generality, one can assume that $C = 1$. The definition of our estimators proceeds from ergodicity and self-similarity of fBm increments. Let π_H^a be the covariance function of (V^a). >From properties of a, one has

$$\pi_H^a(j) \;=\; \mathbb{E}\Big(\; V^a\Big(\frac{i}{N}\Big)V^a\Big(\frac{i+j}{N}\Big)\;\Big) \;=\; -\frac{1}{2}\,\frac{1}{N^{2H}}\sum_{q,r=0}^{\ell} a_q a_r\;|j+q-r|^{2H}\;. \tag{13}$$

Moreover by denoting $g_{k,a,N}(t) \;=\; \pi_t^a(0)^{\frac{k}{2}}\;E_k$, where $E_k = \mathbb{E}(|Y|^k)$ and $Y \rightsquigarrow \mathcal{N}(0,1)$, one shows that $\mathbb{E}(S_N(k,a)) \;=\; g_{k,a,N}(H)$ which allows us to define a first class of estimators based on a non-linear regression

$$\widehat{H}_N(k,a) \;=\; g_{k,a,N}^{-1}(\;S_N(k,a)\;)\;.$$

We prove ([7]) that if one chooses $p > H + 1/4$, this class is well-defined, converges almost surely to H, and verifies a central limit theorem for $0 < H < 1$, with a rate of convergence $1/\sqrt{N}\log(N)$. If $p = 1$ (resp. $p \geq 2$) the results are available for $0 < H < 3/4$ (resp. for $0 < H < 1$). It is also proved in [7] that the asymptotic constant of the estimators' variance is minimal for $k = 2$. In that case, one obtains the simple asymptotic result

$$\sqrt{N}\log(N)\Big(\;\widehat{H}_N(2,a)\;-\;H\;\Big)\;\xrightarrow{\;d\;}\;\mathcal{N}\left(\;0,\frac{1}{2}\;\sum_{i\in\mathbb{Z}}\rho_H^a(i)^2\;\right)\;,\quad\text{where}\;\;\rho_H^a(i)=\frac{\pi_H^a(i)}{\pi_H^a(0)}\;.$$

This asymptotic behavior will be further used to explore the quality of different simulation methods (see Section 6).

Scale parameter unknown:

Through spectral methods or wavelet methods, we have noticed that the use of a log-linear regression allowed us to exhibit estimators that are independent of C. We may apply a similar reasoning in this framework. Let us define the sequence of filters $(a^m)_{1\leq m\leq M}$ by

$$a_i^m \;=\left\{\begin{array}{ll} a_j & \text{if } i = jm \\ 0 & \text{otherwise}\end{array}\right.\;,\qquad \text{for } i = 0,\ldots,m\ell+1\;.$$

One immediately sees that

$$\mathbb{E}(S_N(k,a^m)) = \;m^{Hk}\;\mathbb{E}(S_N(k,a))\;.$$

By estimating $\mathbb{E}(S_N(k,a^m))$ by $S_N(k,a^m)$, an estimator of H can be deduced from a simple linear regression of $\big\{\log S_N(k,a^m)\big\}_{1\leq m\leq M}$ on $\big\{k\log(m)\big\}_{1\leq m\leq M}$. This procedure allows us to get a class of estimators denoted by $\widetilde{H}_N^{ols}(k,a,M)$ for which we have proved ([7]) the almost sure convergence and the asymptotic normality (with a rate of convergence in $1/\sqrt{N}$) if $p > H + 1/4$. Let us mention that

we have also considered a linear regression weighted by the diagonal weight matrix, of size $M \times M$, $\widehat{G}_k$ defined for $m = 1, \ldots, M$ by

$$\widehat{g}_{mm} \;=\; \sum_{j\geq 1} (c_{2j}^k)^2 \; (2j)! \sum_{i\in\mathbb{Z}} \rho_{\widetilde{H}^{ols}}^{a^m}(i)^{2j} \;, \quad \text{with } c_{2j}^k \;=\; \frac{1}{(2j)!} \prod_{q=0}^{j-1} (k-2q).$$

The consistency, the asymptotic normality, and the optimality for $k = 2$ hold again.

5 Examples of simulations

We intend to illustrate the implemented methods through simulations. For a sample size $N = 1000$, we generate 50 paths of standard fractional Brownian motion with parameter $H = 0.9$, discretized uniformly on $[0, 1[$, via methods of Mandelbrot, Sellan and *al*, Choleski, Levinson and Wood-Chan. The S-plus functions related to these methods are respectively `mvnFBM()`, `waveFBM()`, `cholFBM()`, `levFBM()`, `circFBM()`, see Section 7. For each path, we compute the estimators of H, via the methods explained in Section 4, and for which scripts are presented hereafter. Some Boxplots, Fig.4, Fig.5 and Fig.6, illustrate the results. A high value of the Hurst parameter was used to exhibit several weak points for the various simulation and estimation methods, discussed in previous Sections.

6 Quality of generators

In this section, our aim is to explore the quality of different simulation methods. Such a study has been already undertaken by Jennane and *al* [17]. They consider various simulators of the fBm, and give explicitly three procedures for testing the normality, the stationarity and the self-similarity of fGn. The two first procedures do not allow an exploration of the quality of generators: indeed Tables 3 and 4 of [17] show that four simulators among the five considered pass their tests successfully. Due to these omnibus tests, we have not envisaged to test the normality and the stationarity. And instead of using the proposed procedure to check the increments self-similarity, we orientate ourself towards another approach based on the theoretical asymptotic behavior of a *parametric* estimation method. Recall that if $\widehat{H}_N(2, a)$ denotes the estimator obtained by discrete variations of *standard* fractional Brownian motion (method explained p.16), we have the asymptotic result: if a denotes a filter of order $p > H + 1/4$ then

$$\sqrt{N}\log(N)\left(\,\widehat{H}_N(2,a) - H\,\right) \;\xrightarrow{d}\; \mathcal{N}(0, \frac{1}{2}\sum_{i\in\mathbb{Z}} \rho_H^a(i)^2)\;.$$

where ρ_H^a is given by (13). Let us describe how we use this property to extract an efficient method of simulation. For different values of the Hurst parameter, and for two filters (one is related to the increments, $a = Inc1 = (1, -1)$, the other is related to a Daubechies wavelet of order 4, $a = Db4 = (0.4829629, -0.8365163, 0.2241439, 0.1294095)$):

1. Simulate 200 paths of a standard fractional Brownian motion for each simulation method.

2. For $i = 1, \ldots, 200$:

   - Evaluate the estimator of the self-similarity parameter by discrete variations and denote it by $\widehat{H}_i(a)$.
   - Build a confidence interval of level $\alpha = 0.05$ related to the Gaussian distribution

   $$I(\sigma^2, H, a) \;=\; \left]H - u_\alpha \frac{\sigma}{\sqrt{N}\log(N)} \;;\; H + u_\alpha \frac{\sigma}{\sqrt{N}\log(N)}\right[\,,$$

   where $u_\alpha = \Phi^{-1}(\alpha/2) \approx 1.96$, and σ^2 is the asymptotic constant depending on H and the filter used, and given by Tab.1.

Hurst parameter	H=0.1	H=0.3	H=0.5	H=0.7	H=0.9
$a = Inc1$	0.6820765	0.5625909	0.5	0.7854074	$\times$
$a = Db4$	0.7790751	0.7396438	0.6388889	0.5922214	0.5661291

Table 1: Estimates of the asymptotic constant $\sigma^2 = \frac{1}{2}\sum_{i\in\mathbb{Z}} \rho_H^a(i)^2$ for different values of H and for filters $Inc1$ and $Db4$

3. Estimate the percentage test success by the statistic:

$$S(H, a) \;=\; \frac{1}{200}\sum_{i=1}^{200} \mathbb{1}\left(\widehat{H}_i(a) \in I(\sigma^2, H, a)\right) .$$

The results are presented in Tab.2. For $H > 3/4$, the test has not been evaluated for the filter $Inc1$ since the asymptotic behavior for $\widehat{H}_N(2, Inc1)$ is available only for $0 < H < 3/4$. One can notice that the methods of Mandelbrot and *al* and Levinson are not efficient for low and high values of the Hurst parameter. Concerning the method of Choleski and the one developed by Wood and Chan, the results are quite satisfactory since the level 95% is almost always reached. It appears also clearly that the estimators deduced from samples simulated by the method of Sellan and *al* are strongly biased. This study points out that the most stable simulation method are the Choleski's method and the one based on circulant matrix.

As a general conclusion, since Subsection 3.7 illustrates the excellent approximation of a sample of a fBm via the cumulated sums of a fGn, we advice to use the method of Wood and Chan [29], for its rapidity (even for large N) for simulating a fBm.

Hurst parameter		H=0.1	H=0.3	H=0.5	H=0.7	H=0.9
`mvnFBM()`	$Inc1$	72.0 %	89.0 %	97.0 %	83.0 %	×
	$Db4$	92.0 %	93.0 %	96.0 %	3.0 %	0.0 %
`waveFBM()`	$Inc1$	69 %	2.5 %	0 %	0 %	×
	$Db4$	87.0 %	22.0 %	20.0 %	7.5 %	0.0 %
`cholFBM()`	$Inc1$	97.5 %	94.5 %	93.0 %	91.0 %	×
	$Db4$	93.0 %	96.5 %	94.5 %	% 97.5 %	93.0 %
`levFBM()`	$Inc1$	50.5 %	79.5 %	97.0 %	63.5 %	×
	$Db4$	85.0 %	93.5 %	99.0 %	89.0 %	0.0 %
`circFBM()`	$Inc1$	94 %	93.5 %	92.5 %	92.0 %	×
	$Db4$	94.5 %	96.5 %	96.5 %	97.5 %	100 %

Table 2: Results of percentage test success for various methods of simulation.

7 S-plus scripts

In this Section, we present the S-plus scripts implementing the simulation and identification methods for the fBm, i.e. the methods described in Section 3 and 4. The Table Tab.7 summarizes this study: for each method its S-plus function is associated, with the arguments it needs, and eventually the subroutines related to. We advice the user to fix precisely the arguments of each function: for example, to generate a standard fBm of length N with parameter $H = 0.8$, by the Wood-Chan's method, one writes `circFBM(n=1000, H=0.8)`. By default `plotfBm=1`, the resulting path will be drawn. The function `waveST()` estimating H by a wavelet method has been written using the functions from the S-plus library `Wavethresh()`, available on the web at the following address `http://www.stat.cmu.edu/S/`. This library allows only the decomposition of a signal of length a power of 2. For a signal of any length N, one mirrors the data to the smallest power of two larger than N.

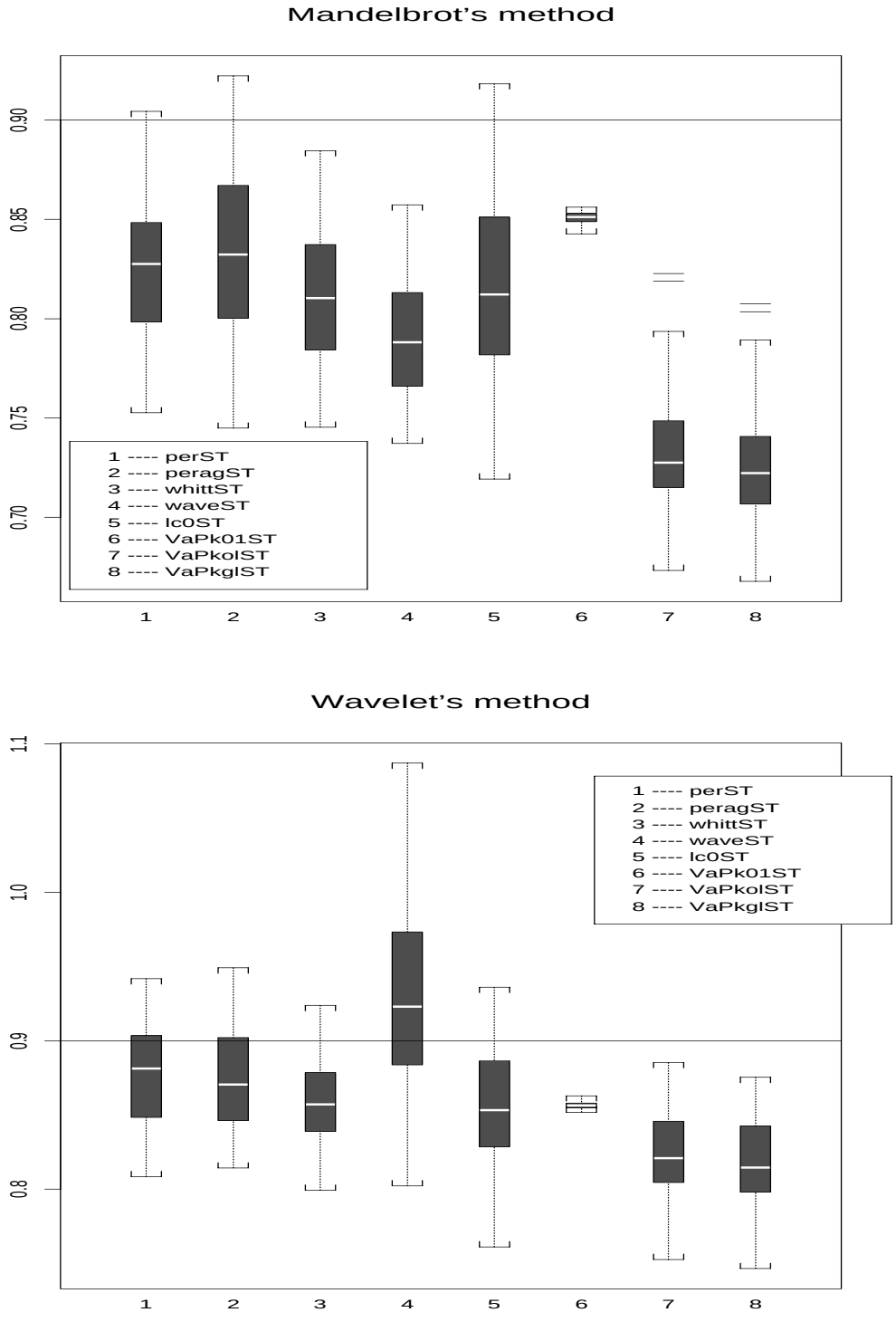


Figure 4: Boxplots of estimators of H for 50 paths simulated respectively by Mandelbrot's method and by wavelet synthesis.

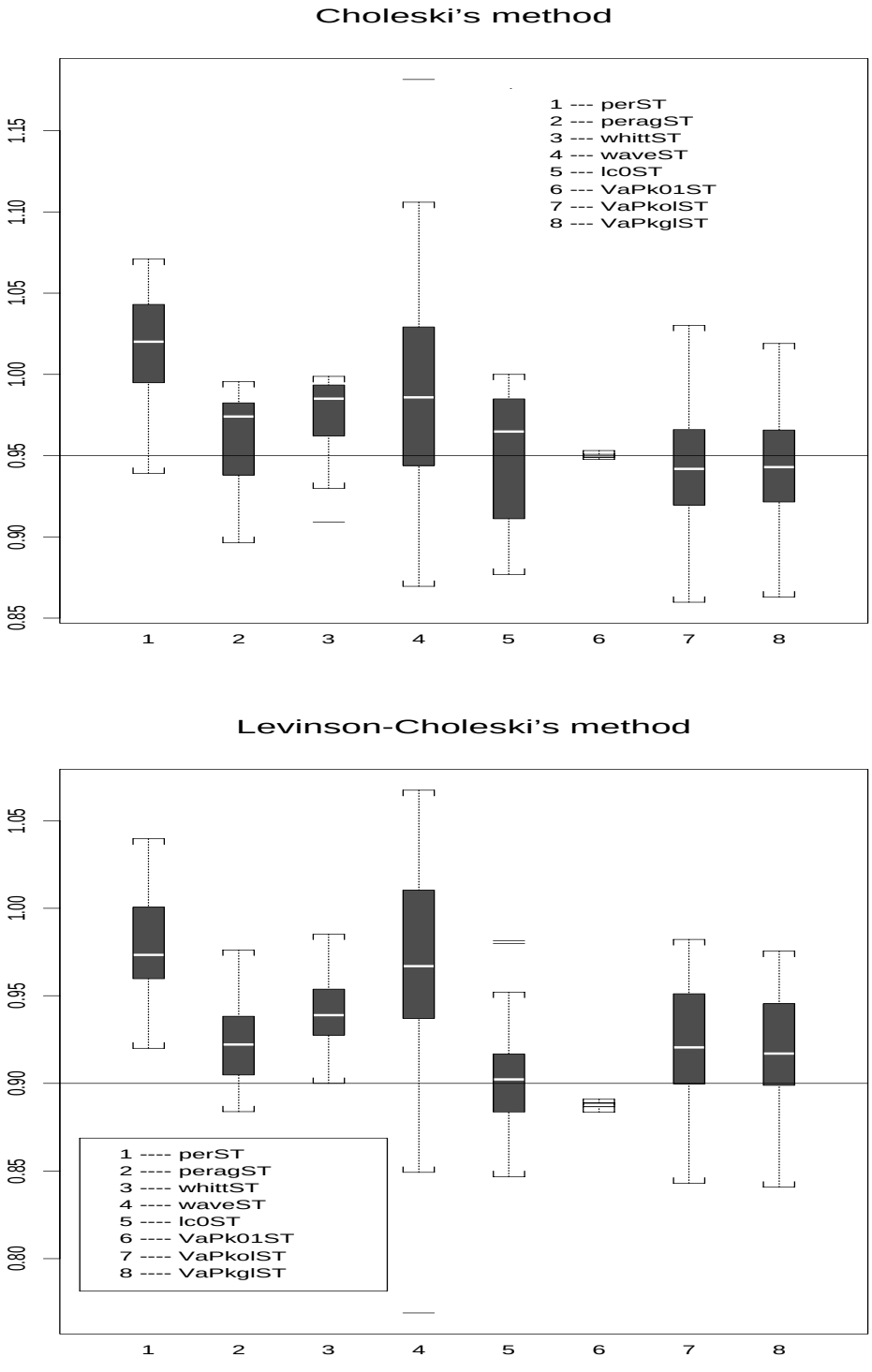


Figure 5: Boxplots of estimators of H for 50 paths simulated respectively by Choleski's and Levinson's method.

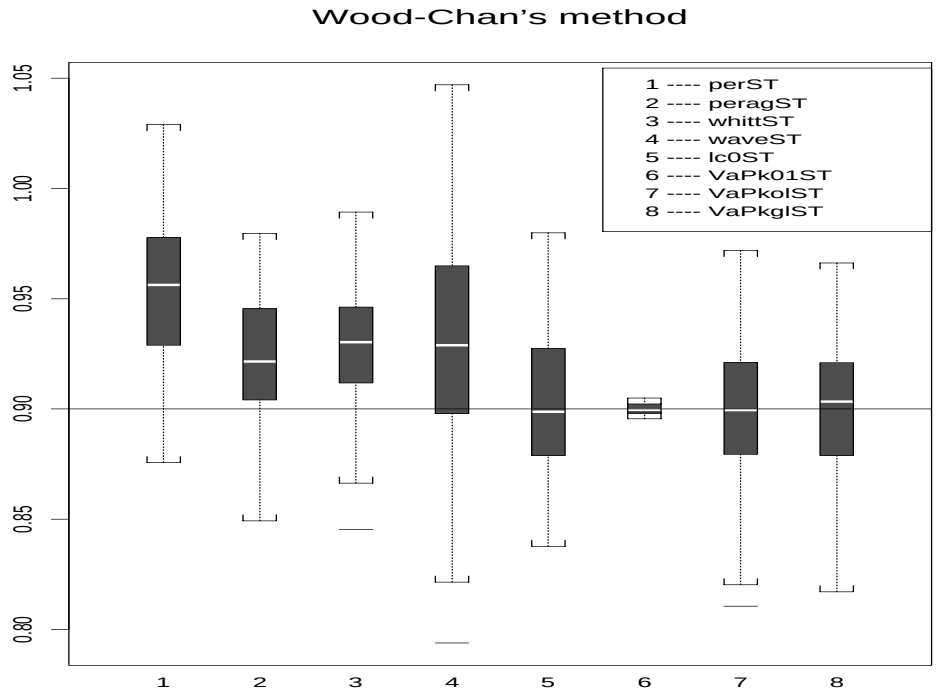
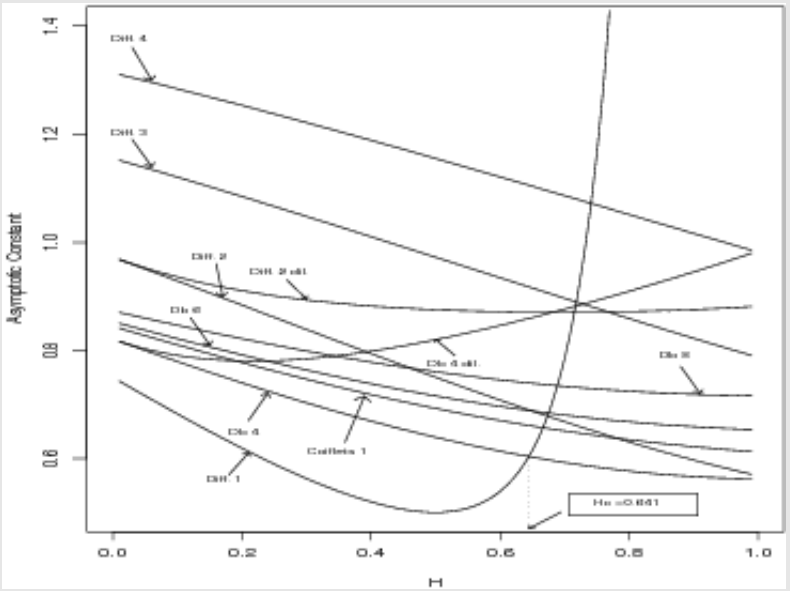


Figure 6: Boxplots of estimators of H for 50 paths simulated by Wood-Chan's method.

	S-plus name	Object
Generators of fBm	`mvnFBM(n= ,H= ,plotfBm= )`	Mandelbrot's method
	`waveFBM(n= ,H= ,J= ,plotfBm= )` subroutine: `convol(x,y)`	wavelet synthesis: fractional integration of a MRA
	`cholFBM(n= ,H= ,plotfBm= )`	Choleski decomposition of the covariance matrix.
	`levFBM(n= ,H= ,plotfBm= )`	Levinson's algorithm for Toeplitz matrices
	`circFBM(n= ,H= ,plotfBm= )`	method of circulant matrix
Estimators of H	`perST(fBm= ,m1= ,m2= ,llplot= )`	log-periodogram
	`peraggST(fBm= ,q= ,m1= ,m2= )`	variant of Lobato and Robinson
	`whittST(fBm= ,Hprel= )` subroutine: `spdFGN(Htry= ,n= )`	Whittle's estimator
	`waveST(fBm= ,j1= ,j2= ,llplot= )`	wavelet decomposition
	`lc0ST(fBm= ,sign= )`	level 0 crossings of fGn
	`VaPkstST(fBm= ,k= ,a= ,Hprel= )` subroutine: `piaH(a= ,H= ,i= )`	k-th absolute empirical moment of discrete variations of standard fBm
	`VaPkolST(fBm= ,k= ,a= ,M= ,llplot= )` subroutine: `piaH(a= ,H= ,i= )`	k-th absolute empirical moment of discrete variations of fBm: ordinary least squares
	`VaPkglST(fBm= ,k= ,a= ,M= ,llplot= )` subroutine: `piaH(a= ,H= ,i= )` `rhoadil(a,H,j,m1,m2)`	k-th absolute empirical moment of discrete variations of fBm: generalized least squares

Table 3: Summary of synthesis and analysis methods implemented over the software S-plus for fractional Brownian motion.



[Coe01]
J.-F. Coeurjolly. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Infer. Stoch. Process.*, 4(2):199-227, 2001.

Estimating the Parameters of a Fractional Brownian Motion by discrete variations of its sample paths

Jean-François Coeurjolly (jean-francois.coeurjolly@imag.fr)
IMAG–LMC, University Joseph Fourier, B. P. 53, 38041 Grenoble Cedex 09, France

31/01/2001

Abstract. This paper develops a class of consistent estimators of the parameters of a fractional Brownian motion based on the asymptotic behavior of the k-th absolute moment of discrete variations of its sampled paths over a discrete grid of the interval $[0,1]$. We derive explicit convergence rates for these types of estimators, valid through the whole range $0 < H < 1$ of the self-similarity parameter. We also establish the asymptotic normality of our estimators. The effectiveness of our procedure is investigated in a simulation study.

Keywords: Fractional Brownian motion, Fractional Gaussian noise, discrete variations, consistency, self-similarity.

Abbreviations: fBm – Fractional Brownian Motion; lssG – locally self-similar Gaussian process; fGn – Fractional Gaussian Noise; MLE – Maximum Likelihood Estimator; DWT – Discrete Wavelet Transform.

AMS (1991) Subject Classification: Primary 62M10, Secondary 62G05

1. Introduction

Many naturally occurring phenomena can be effectively modeled using self-similar processes. For such processes observations that are far apart (in time or space) are correlated too strongly indicating the presence of a long-range dependence. As a result self-similar processes have been used to successfully model data exhibiting long-range dependence and arising in a variety of different scientific fields, including hydrology (McLeod et *al.*, 1978), biology (Collins and De Luca, 2000), medicine (Kuklinski et *al.*, 1989), economics (Granger, 1966) or traffic network (Willinger, 1995). The empirical presence of long-memory in such series is found in a local version of the power spectrum which behaves, as $|\lambda|^{1-2H}$, as $\lambda \to 0$, where $H \in]1/2, 1[$ is the long-memory parameter. Beran (1994) gives a good review of historical and statistical aspects of self-similar and long-memory processes.

Among the simplest models that display long-range dependence, one can consider the fractional Brownian motion (fBm), introduced in the statistics community by Mandelbrot *and al.* (1968). Consider the process, null at the

origin, defined for real $t \geq 0$ by the stochastic integral

$$B_{H,C}(t) \;=\; C\,V_H^{1/2} \int_{\mathbb{R}} f_t(s)\,dB(s) \qquad (1)$$

$$\text{with} \qquad f_t(s) = \frac{1}{\Gamma(H+1/2)} \left\{ |t-s|^{H-\frac{1}{2}} \mathbb{1}_{]-\infty,t]}(s) - |s|^{H-1/2} \mathbb{1}_{]-\infty,0]}(s) \right\},$$

with $B_{H,C}(0) = 0$ and $V_H = \Gamma(2H+1)\sin(\pi H)$; Γ denotes the Gamma function and B a standard Brownian motion.The fractional Brownian motion of index H ($0 < H < 1$) and scale parameter C is the Gaussian process $\{B_{H,C}(t)\,,\, t \geq 0\}$ with mean 0, stationary and self-similar increments such that

$$\mathbb{E}\Big(\left\{ B_{H,C}(t) - B_{H,C}(s) \right\}^2 \Big) \;=\; C^2\,|t-s|^{2H} \quad \forall\, s,t \in \mathbb{R}^+.$$

The index H characterizes the self-similar behavior of the process and many methods have been proposed in the literature for estimating H, some of which are described in the recent monograph of Beran (1994). Among these the most well known are variance and covariance based methods (rescaled range R/S, variogram, log-periodogram) and maximum likelihood based methods (Whittle estimator).

Recently, some new approaches have been considered which are either based on the discrete variations of observed sample paths or on the behavior of a wavelet transform of the sample path across scales. Both methods rely on a specific filtering of a discrete observation of one sample path that is designed to destroy the long-range dependence of the observations. This is also the approach that we are going to consider for our estimation procedure. These recent methods were designed not only for the analysis of fractional Brownian motion but also for a larger class of stochastic processes, namely the class of Gaussian processes that are locally self-similar at the origin.

A zero-mean Gaussian process $\{X(t),\, t \geq 0\}$, with stationary increments, is said to be locally self-similar at 0, and is denoted by lssG, if its semi-variance function $v(t)$ defined by

$$v(t) = \frac{1}{2}\mathbb{E}\big(\; (X(s+t) - X(s))^2 \;\big)\;,$$

satisfies, as $t \to 0$, the following property

$$v^{2D}(t) \;=\; v^{2D}(0) \;+\; (-1)^D C|t|^{2H} \;+\; o(|t|^{2H}), \qquad (2)$$

with $0 < H < 1$, and D denoting the largest integer such that v is $2D$-times differentiable. The local self-similarity, say at a point y, is developed by Benassi et *al.* (1998), and means that the tangent process at y tends in distribution to a self-similar process. In the Gaussian framework, the local self-similarity at 0 is equivalent to (2).

Let $\{V^a(t), t \geq 0\}$ be the process derived by filtering $\{X(t), t \geq 0\}$ with a filter a, of length $\ell+1$ and order $p \geq 1$, such that $\sum_{q=0}^{\ell} a_q\, q^r = 0$, for $r = 0, \ldots, p-1$. Let $\{X(t),\, t \geq 0\}$ be observed at times $\left\{ 0, \frac{1}{N}, \ldots, \frac{N-1}{N} \right\}$, and let us define the variations of some function $\mathbb{F}$ of $\{X(t),\, t \geq 0\}$, say the $\mathbb{F}$-variations of X, by

$$V_N(\mathbb{F}, a) \;=\; \frac{1}{N-\ell} \sum_{i=\ell}^{N-1} \mathbb{F}\left(V^a\left(\frac{i}{N}\right)\right).$$

Many recent estimation procedures are based on such $\mathbb{F}$-variations, with the particular choice

$$\mathbb{F}(t) = \mathbb{H}^k(t) = \frac{1}{\mathbb{E}(|V^a(0)|^k)}\, |t|^k \;-\; 1.$$

For example, in the fBm case, Bardet (1997), Poggi and Viano (1998), using $\mathbb{F} = \mathbb{H}^2$ or Higuchi (1988) using $\mathbb{F} = \mathbb{H}^1$, and relying upon a regression of the $\mathbb{F}$-variations of several filters of order 1, derive estimators of the self-similarity index H that are consistent at a rate $1/\sqrt{N}$, for $H < 3/4$. For the lssG case, Istas and Lang (1997) and Kent and Wood (1997), study the asymptotic behavior of quadratic variations for $p \geq 2$, and derive estimators of H that are consistent at a rate $1/\sqrt{N}$, for all H. Finally, for a standard fBm model ($C = 1$), Peltier and Lévy-Véhel (1994) use a fractal approach for the identification problem and obtain a class of estimators of H that are consistent at a rate $1/\sqrt{N}\log(N)$, for $H < 3/4$, by controlling the behavior of the k-th absolute moment of the increments of a standard fBm ($\mathbb{F} = \mathbb{H}^k$, for $k \in \mathbb{N}^*$). The nature of Peltier's estimator, which is based on the fact that the scaling factor C is known, is such that deriving a more concise rate in the case of an unknown value for C is a particularly awkward task.

The main objective of this paper is to clarify and extend the above results by studying the $\mathbb{H}^k$ –variations (for $k \in \mathbb{R}^{+*}$) for a standard fBm ($C = 1$) and also for a nonstandard fBm process, using filters of arbitrary order. Our main contribution is to exhibit, for each model, the optimal k, to obtain consistency rates valid for any value of H within $]0,1[$ and to derive a statistic for testing the mean of a noncentered general fBm or lssG process.

The paper is organized as follows. Section 2 introduces some notations, defines a statistic called k-variations which consists in the k-th absolute moment of normalized discrete variations of a fBm, and proves some useful results concerning its convergences. We apply these results to the identification problem in Section 3. We distinguish the cases C known and C unknown, and explicit two classes of estimators. When C is known, we obtain estimators convergent at a rate $1/\sqrt{N}\log(N)$, $\forall H \in]0,1[$, and when C is unknown, we derive estimators convergent at a rate $1/\sqrt{N}$, $\forall H \in]0,1[$. For both classes of

estimators, we prove that the asymptotic distribution is Gaussian, and that the optimal value for k is 2. In Section 4, we establish a central limit theorem for a free statistic, allowing us to test its mean. Simulation results are given in the last section. Proofs of results in previous sections are presented in the Appendix.

2. k-variations of the fBm

The aim of this section is to introduce some notations and to derive the asymptotic results of convergence for the k-th absolute moment of the empirical normalized discrete variations of the fBm. Hereafter, $B_{H,C}$ will denote a path of a fBm, with parameters $(H,C) \in]0,1[\times\mathbb{R}^{+*}$, observed at times $i/N, i = 0,\ldots,N-1$. Moreover, given two integers ℓ and p, we shall denote by a, a filter of length $\ell+1$, and of order $p \geq 1$, that is a $\ell+1$-dimensional vector with real components a_j, $j = 0,\ldots,\ell$ with the following properties: for all indices $0 \leq r < p$,

$$\sum_{j=0}^{\ell} j^r a_j = 0 \quad \text{and} \quad \sum_{j=0}^{\ell} j^p a_j \neq 0.$$

To any such filter a, we associate the $\{B_H(t),\, t \geq 0\}$ time series filtered by a, say (V^a), which is the time series defined by

$$V^a\left(\frac{i}{N}\right) \;=\; \sum_{q=0}^{\ell} a_q\, B_H\left(\frac{i-q}{N}\right), \qquad \text{for } i = \ell,\ldots N-1.$$

Let us note E_k the k-th absolute moment of a standard Gaussian variable, explicitly given for $k > 0$, by: $E_k = 2^{k/2}\Gamma(k+1/2)/\Gamma(1/2)$. Now define the statistic $V_N(k,a)$, called k-variations, by:

$$V_N(k,a) \;=\; \frac{1}{N-\ell} \sum_{i=\ell}^{N-1} \left\{ \frac{|V^a(i/N)|^k}{\mathbb{E}\left(|V^a(i/N)|^k\right)} - 1 \right\}. \tag{3}$$

Let π_H^a denote the covariance function of the series (V^a). This series is clearly stationary, and the self-similarity of the fBm implies that for $j \in \mathbb{Z}$

$$\pi_H^a(j) \;=\; \mathbb{E}(V^a(i/N)V^a((i+j)/N)) \;=\; -\frac{C^2}{2} \sum_{q,r=0}^{\ell} a_q a_r |q-r+j|^{2H}. \tag{4}$$

The pertinence of using discrete filters is revealed by the following Lemma, see Istas and Lang (1997):

LEMMA 1. *Let a be a filter of order $p \geq 1$. Then, for j large enough,*

$$\pi_H^a(j) \sim \kappa j^{2H-2p}$$

where $\kappa = -\frac{1}{2}\sum_{q,r=0}^{\ell} \beta_{2p,H} \frac{(q-r)^{2p}}{(2p)!}$ *with* $\beta_{2p,H} = 2H(2H-1)\ldots(2H-2p+1)$.

The above lemma, whose proof relies upon a Taylor-expansion of π_H^a, closely relates the decay in correlation with the order of the filter. From now on, we will note

$$\rho_H^a(j) = \frac{\pi_H^a(j)}{\pi_H^a(0)}. \tag{5}$$

Using the works of Doob (1953, p.492), of Breuer et *al.* (1983), and from the expansion in Hermite polynomials of the function $\mathbb{H}^k(t) = |t|^k/E_k - 1$, whose coefficients denoted c_{2j}^k are explicited in Section 6.1, we obtain the following result

PROPOSITION 1. *Let a be a filter of order $p \geq 1$ and let k be a positive real number, then as $N \to +\infty$ we have*

(i)

$$V_N(k,a) \xrightarrow{a.s.} 0. \tag{6}$$

(ii) If $p > H + 1/4$,

$$\sqrt{N}\, V_N(k,a) \xrightarrow{d} \mathcal{N}\left(0, A_1(H,k,a)\right), \tag{7}$$

where

$$A_1(H,k,a) = \sum_{j\geq 1} (c_{2j}^k)^2\, (2j)! \sum_{i\in\mathbb{Z}} \rho_H^a(i)^{2j} \quad \textit{and} \quad c_{2j}^k = \frac{1}{(2j)!} \prod_{q=0}^{j-1} (k - 2q). \tag{8}$$

The condition $p > H + 1/4$ of (ii) results from the square summability of ρ_H^a, resulting itself from Theorem 1 of Breuer et *al.* (Breuer and Major, 1983), Thus, if $p = 1$ the convergence in distribution is valid for $H \in]0,3/4[$, whereas it is valid for $H \in]0,1[$ as soon as $p \geq 2$.

Corollary 1 answers to the following question: what happens if we allow k to vary with N? This question is motivated by the fact that choosing $k = N^{-\alpha}$ implies, when $N \to +\infty$, that

$$c_{2j}^k \sim N^{-\alpha}\theta_{2j} \quad \text{with} \quad \theta_{2j} = (-1)^{j-1}2^{j-1}\frac{(j-1)!}{(2j)!}.$$

COROLLARY 1. *Let a be a filter of order $p > H + 1/4$ and let α be a positive real number, then, as $N \to +\infty$, we have*

$$N^{\alpha+\frac{1}{2}}\, V_N(\, N^{-\alpha}, a\,) \;\xrightarrow{d}\; \mathcal{N}\big(\, 0, A_1(H,a)\,\big)\;,$$

with

$$A_1(H,a) = \sum_{j\geq 1} (\theta_{2j})^2\, (2j)!\, \sum_{i\in\mathbb{Z}} \rho_H^a(i)^{2j} \quad \textit{and} \quad \theta_{2j} = (-1)^{j-1} \frac{2^{j-1}\,(j-1)!}{(2j)!}\;.$$

We can now apply these results to the identification problem. We will distinguish the cases C known, and C unknown.

3. Identification of the fBm

3.1. Case of the standard fBm

In this section, we will suppose $C = 1$. Define the k-th empirical absolute moment of the discrete variations of the fBm by

$$S_N(k,a) \;=\; \frac{1}{N-\ell} \sum_{i=\ell}^{N-1} |V^a(i/N)|^k\;. \tag{9}$$

From the stationarity of (V^a) and the introduced notations, we have:

$$\mathbb{E}(S_N(k,a)) \;=\; \frac{1}{N^{Hk}}\, \{\pi_H^a(0)\}^{k/2}\, E_k.$$

Estimation of the self-similarity parameter H, using a regression method based on the k-variations on several filters, has been suggested and studied by in a number of recent papers, (see *e.g.* Istas and Lang (1997), Kent and Wood (1997), Poggi and Viano (1998)). They prove that the benefit of using estimators based on k-variations is that the estimators are independent of $\mathbb{E}(V^a(0)^2)$. For this reason, a local self-similarity around 0, suffices for estimating the Hölder index. On the other hand, for the fBm model, the self-similarity is global, $\mathbb{E}S_N(k,a)$ is explicitly obtained and it turns out that all the information on H is retained by using only one filter. Estimating $\mathbb{E}S_N(k,a)$ by $S_N(k,a)$, we may construct a first class of estimators. More precisely, for any filter a, and for all real $k > 0$, set

$$\widehat{H}_N(k,a) \;=\; g_{k,a,N}^{-1}(\, S_N(k,a)\,), \tag{10}$$

where we define for an integer $N \geq 1$, a filter a, and a real k strictly positive, the function $g_{k,a,N}$ on $]0,1[$ by

$$g_{k,a,N}(t) \;=\; \frac{1}{N^{tk}}\, \{\pi_t^a(0)\}^{k/2} E_k.$$

Estimators $\widehat{H}_N(\cdot,\cdot)$ are well defined since $g_{k,a,N}$ is invertible. Indeed, by computing the derivative of $g_{k,a,N}$, one sees that $g_{k,a,N}$ is monotically decreasing as soon as the sample size N satisfies the following inequality:

$$N \;>\; \mathrm{Max}_{t\in]0,1[}\;\; \exp\left\{ \frac{\sum_{|q-r|>2} a_q a_r \;\log(|q-r|)\;|q-r|^{2t}}{\sum_{q,r} a_q a_r\;|q-r|^{2t}} \right\}. \qquad (\mathcal{H})$$

For the usual filters used in practice (increments of a given order, Daubechies wavelet filters, coiflets, ... (see e.g. Daubechies (1992)) the condition $(\mathcal{H})$ is fulfilled whatever $N \in \mathbb{N}^*$. The case $p = 1$, for which the process V^a is a standard fractional Gaussian noise, has been studied by Peltier and Lévy-Véhel (1994). In particular, these authors obtain an explicit analytical expression for the estimators, which is not possible anymore when $p \geq 2$. Let us precise that numerically $\widehat{H}_N(k,a)$ is obtained by minimization of $g_{k,a,N}(t) - S_N(k,a)$ on $]0,1[$. Proposition 1 allows us to prove the following result concerning the convergences of $\widehat{H}_N(k,a)$.

PROPOSITION 2. *Let a be a filter of order $p \geq 1$ and let k be a positive real number, then, as $N \to +\infty$, we have*

(i)

$$\widehat{H}_N(k,a) \;\xrightarrow{a.s.}\; H\;. \qquad (11)$$

(ii) If $p > H + 1/4$,

$$\sqrt{N}\log(N)\Big(\;\widehat{H}_N(k,a)\;-\;H\;\Big) \;\xrightarrow{\mathcal{L}}\; \mathcal{N}\left(0, \frac{A_1(H,k,a)}{k^2}\right), \qquad (12)$$

where $A_1(H,k,a)$ is given by (8).

The computation of the exact asymptotic variance allows us to compare estimators. And we may wonder if there exists an optimal value of k that minimizes the asymptotic variance of $\widehat{H}_N(k,a)$.

COROLLARY 2. *For any filter a of order $p \geq 1$, the asymptotic variance of $\widehat{H}_N(k,a)$ is minimal for $k = 2$.*

Proof. Let us note $(c_{2j}^k)' = \frac{1}{k^2}(c_{2j}^k)^2(2j)!$, where c_{2j}^k is the $2j$-th coefficient in Hermite expansion of the function $\mathbb{H}^k(t) = |t|^k/E_k - 1$. Section 6.1 explicits this computation and shows that $c_{2j}^k = \frac{1}{(2j)!}\prod_{q=0}^{j-1}(k-2q)$. We have

$$\frac{1}{k^2}A_1(H,k,a) = \sum_{j\geq 1}(c_{2j}^k)' \sum_{i\in\mathbb{Z}} \rho_H^a(i)^{2j} \;\geq\; (c_2^k)' \sum_{i\in\mathbb{Z}} \rho_H^a(i)^2 = \frac{1}{2}\sum_{i\in\mathbb{Z}} \rho_H^a(i)^2.$$

It is therefore sufficient to note that $(c_{2j}^2)' = 0$ for $j > 1$, i.e. that $\frac{1}{4}A_1(H,2,a) = \frac{1}{2}\sum_{j\in\mathbb{Z}} \rho_H^a(j)^2$.

Therefore, for any given filter, $\widehat{H}_N(2,a)$ is the optimal estimator, in the sense that it minimizes the variance, and the following limit holds:

$$\sqrt{N}\log(N)\left(\widehat{H}_N(2,a) - H\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{2}\sum_{i\in\mathbb{Z}} \rho_H^a(i)^2\right). \qquad (13)$$

Even if it assumes the knowledge of the discretization step and the scaling coefficient C, the class of estimators $\widehat{H}_N(k,a)$ is particularly interesting since it allows us to derive an asymptotic test based on (13) for judging the quality of various simulators of the fBm, see Coeurjolly (2000).

3.2. Identification of non standard fBm

Let us assume that the process is sampled on the time grid $\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\}$. When trying to estimate the self-similarity parameter from real data under the assumption that the observed times series is a realization of a fractional Brownian motion two problems may arise: firstly, the series is not necessarily sampled on [0,1], and secondly the initial variance $\sigma^2 = \mathbb{E}([\, B_{H,C}(1) - B_{H,C}(0)\,]^2)$, is not equal to 1. This means that there exists a scaling coefficient C, such that the data vector $\mathbf{D}$, of size N, can be written as:

$$D(i) = C \cdot B_H(i/N), \qquad \text{for } i = 0, \ldots, N-1,$$

where $\{B_H(0), \ldots, B_H(\frac{N-1}{N})\}$ is a sample of a standard fractional Brownian motion on [0,1]. In order to obtain estimators of the self-similarity index that are independent of the scaling constant C, we are going to use a regression method. For a given integer $m \geq 1$, let a^m be the filter defined by:

$$a_i^m = \begin{cases} a_j & \text{if } i = jm \\ 0 & \text{otherwise} \end{cases} \qquad \text{for } i = 0, \ldots, m\ell + 1,$$

and the kth empirical moment of the corresponding discrete variations, given by:

$$S_N(k, a^m) = \frac{1}{N - m\ell} \sum_{i=m\ell}^{N-1} \left| V^{a^m}\left(\frac{i}{N}\right)\right|^k.$$

Hereafter we will follow arguments similar to those of Kent and Wood (1997), who have studied the case $k = 2$; noting that

$$\mathbb{E}S_N(k, a^m) = \frac{\pi_H^{a^m}(0)^{\frac{k}{2}}}{N^{Hk}} E_k = m^{Hk}\, \mathbb{E}S_N(k,a),$$

we observe that $\log \mathbb{E}(S_N(k,a^m))$ is linear in H. It is therefore natural to consider M filters and to regress by ordinary least squares L_N on X_M, where $L_N = \left\{ \log S_N(k,a^m) \right\}_{1\le m\le M}$ and X_M is the $M\times 2$ design matrix defined by:

$$X_M = \begin{pmatrix} 0 & k\log(2) & \dots & k\log(M) \\ 1 & 1 & \dots & 1 \end{pmatrix}^t.$$

Such a procedure leads to a class of estimators of $\alpha = (H,\theta)^t$, which will be denoted by $\widetilde{\alpha}_N^{ols}(k,a,M) = \left(\widetilde{H}_N^{ols}(k,a,M), \widetilde{\theta}_N^{ols}(k,a,M)\right)^t$, θ being the following parameter

$$\theta = \log \mathbb{E}(S_N(k,a)) = k\log(C) - Hk\log(N) + \log\left(\pi_H^a(0)^{\frac{k}{2}}/E_k \right).$$

With the above notation, this class may be written as

$$\widetilde{\alpha}_N^{ols}(k,a,M) = (X_M^t X_M)^{-1} X_M^t L_N.$$

We can now state the following result concerning the M-dimensional convergence of the k-variations, and the asymptotic normality of the error term in the linear regression of L_N on X_M.

PROPOSITION 3. *Let $M>1$ be an integer and k a positive real number. Let also a be a filter of order $p\ge 1$. For all $H\in]0,1[$ we have, as $N\to+\infty$:*

(i)
$$\left(V_N(k,a^1),\dots,V_N(k,a^M) \right) \xrightarrow{a.s.} (0,\dots,0). \quad (14)$$

(ii) If $p > H+1/4$,
$$\sqrt{N}\left(V_N(k,a^1),\dots,V_N(k,a^M) \right) \xrightarrow{d} \mathcal{N}(0,G_k). \quad (15)$$

(iii) If $p > H+1/4$,
$$\sqrt{N}\left(L_N - X_M\,\alpha \right) \xrightarrow{d} \mathcal{N}(0,G_k), \quad (16)$$

where G_k denotes the $M\times M$ matrix whose generic entries are given by

$$g_{mn} = \sum_{j\ge 1} (c_{2j}^k)^2\,(2j)! \sum_{i\in\mathbb{Z}} \rho_H^{a^m,a^n}(i)^{2j},$$

with

$$\rho_H^{a^m,a^n}(i) = \frac{\sum_{q=0}^{m\ell}\sum_{r=0}^{n\ell} a_q\,a_r\,|mq-nr+i|^{2H}}{m^H n^H \sum_{q,r=0}^{\ell} a_q a_r |q-r|^{2H}}.$$

The definition of θ explicitly leads to a family of estimators of C defined as follows:

$$\widetilde{C}_N^{ols}(k,a,M) = N^{\widetilde{H}^{ols}} \left(\pi_{\widetilde{H}^{ols}}^a(0)^{\frac12} E_k^{\frac1k} \right)^{-1} \exp\left(\frac{1}{k}\,\widetilde{\theta}_N^{ols}(k,a,M) \right), \quad (17)$$

where $\widetilde{H}^{ols} = \widetilde{H}_N^{ols}(k,a,M)$. We can now state the next proposition concerning the convergence of estimators of the self-similarity parameter and the scaling coefficient C.

PROPOSITION 4. *Let $M > 1$ be an integer and k a positive real number. Let also a be a filter of order $p \geq 1$. For all $H \in]0,1[$ we have, as $N \to +\infty$:*

$$(i) \qquad \left(\widetilde{H}_N^{ols}(k,a,M), \widetilde{C}_N^{ols}(k,a,M) \right) \xrightarrow{a.s.} (H,C)\,. \tag{18}$$

(ii) If $p > H + 1/4$,

$$\sqrt{N}\left(\widetilde{H}_N^{ols}(k,a,M) - H \right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{ols}^2(k)\right), \tag{19}$$

and

$$\frac{\sqrt{N}}{\log(N)} \frac{1}{C}\left(\widetilde{C}_N^{ols}(k,a,M) - C \right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{ols}^2(k)\right). \tag{20}$$

where the asymptotic variance $\sigma_{ols}^2(k)$, is defined by: $\sigma_{ols}^2(k) = \frac{A^t\, G_k'\, A}{\|A\|^4}$; $\|\cdot\|$ being the Euclidean norm and A the vector with components $A_m = \log(m) - \frac{1}{M}\sum_{m=1}^{M}\log(m)$ and G_k' being the $M \times M$ matrix whose entries are given by:

$$g'_{mn} = \sum_{j\geq 1} \frac{(c_{2j}^k)^2}{k^2} (2j)! \sum_{i\in\mathbb{Z}} \rho_H^{a^m,a^n}(i)^{2j} \quad . \tag{21}$$

In our procedure estimation, we have avoided estimating H by a nonlinear regression of $\{S_N(k,a^m)\}_{1\leq m\leq M}$ on $\{\mathbb{E}(S_N(k,a^m))\}_{1\leq m\leq M}$ by noting that $\log\mathbb{E}(S_N(k,a^m))$ is linear in H. Thus one may be worried that the nonlinearity due to this logarithmic transform could induce a nonnegligible bias. In the case $k = 2$, we refer the reader to Coeurjolly (1999), where we have proved that this bias behaves like $1/N$ and is thus negligible.

We now state a result similar to Corollary 2 for the asymptotic variances of $\widetilde{H}_N^{ols}(k,a,M)$ and $\widetilde{C}_N^{ols}(k,a,M)$:

COROLLARY 3. *For any filter a of order $p \geq 1$, the asymptotic constant $\sigma_{ols}^2(k)$ is minimal for $k = 2$.*

Proof. It suffices to note that $\forall m,n = 1,\ldots,M$ and $\forall k > 0$, we have

$$\begin{aligned}\left(G_k'\right)_{m,n} = \sum_{j\geq 1}\left(c_{2j}^k\right)' \sum_{i\in\mathbb{Z}} \rho_H^{a^m,a^n}(i)^{2j} \geq \left(c_2^k\right)' \sum_{i\in\mathbb{Z}} \rho_H^{a^m,a^n}(i)^2 &= \frac{1}{2}\sum_{i\in\mathbb{Z}} \rho_H^{a^m,a^n}(i)^2 \\ &= \left(G_2'\right)_{m,n},\end{aligned}$$

which is the result, since the vector A is independent of k.

Following Bardet (1997), we could advocate a weighted linear regression using an estimate $\widehat{G}_k = G_k(\widetilde{H}_N^{ols}(k,a,M))$ of the covariance matrix G_k. With our notation, the generalized least squares regression estimate is then given by

$$\widetilde{\alpha}_N^{gls}(k,a,M) = \left(X_M^t\, (\widehat{G}_k)^{-1}\, X_M \right) X_M^t\, L_N.$$

We therefore obtain the estimators $\widetilde{H}_N^{gls}(k,a,M)$ and $\widetilde{C}_N^{gls}(k,a,M)$. We refer the reader to Coeurjolly (1999), where this procedure has been theoretically and practically investigated. The results are not presented here, since we show that the weighted linear regression doesn't really improve the performances of estimators of H and C, whereas it has the drawback to be numerically costly.

4. Tests for a fractional Brownian motion

Statistical validation of a fractional Brownian motion model is a difficult question, that has been addressed, at our knowledge, by Beran (1992) and Bardet (1997). In his paper, Beran proposes to test the parametric form of a spectral density using Whittle's estimator. We do not consider this approach. Concerning Bardet's work, he investigated to test the self-similarity of the fBm. He presents a statistic based on the asymptotic behavior of the residuals from a linear regression of L_N on X_M weighted by $\widehat{G}_k$. To achieve this, one first computes the standardized distance of the observations to the regression line,

$$D_N(k,a,M) = \left\| L_N - X_M \, \widetilde{\alpha}_N^{gls}(k,a,M) \right\|^2_{\widehat{G}_k^{-1}}.$$

When $k=2$ and $a=(1,-1)$, it is shown in Bardet (1997), using Cochran's Lemma that,

$$D_N\big(\,2,(1,-1),M\,\big) \xrightarrow{d} \chi^2_{M-2}, \qquad \text{for} \quad 0<H<3/4. \tag{22}$$

In a first part, he proves the result (22) under fBm assumption, and in a second part for processes that are locally self-similar at 0. It allows to form the test (with X denoting the process)

$$H_0: \; X \text{ is a lssG} \qquad H_1: \; X \text{ is not a lssG}. \tag{23}$$

Under H_0, we have (22). Thus, one can test, for example, the fact that

$$\mathbb{E}\big(\,D_N(2,(1-1),M)\,\big) = M-2.$$

Such a testing procedure has been studied by fixing k and dilating a number of times the filter a. The introduction of the parameter k, suggests an equivalent testing procedure obtained this time by keeping the filter a fixed and letting k to vary. Let us therefore define, for a given filter a and a positive real r, the following statistic

$$P_N(k,a) = \frac{E_k^2}{E_{2k}} \frac{S_N(2k,a)}{S_N(k,a)^2}\,, \tag{24}$$

where $S_N(k,a)$ is given by (9). The next Proposition is related to the convergence of $P_N(k,a)$. In what follows $\widetilde{H}$ denotes a consistent estimator on H.

COROLLARY 4. *Let a be a filter of order $p \geq 1$, and k a positive real number. If the observed process is a fBm, we have, for all $H \in]0,1[$:*

$$(i) \qquad P_N(k,a) \xrightarrow{a.s.} 1 . \tag{25}$$

(ii) *If $p > H + 1/8$,*

$$\sqrt{N}\ \{P_N(k,a) - 1\}\ / A_2(\widetilde{H},k,a)^{\frac{1}{2}} \xrightarrow{d} \mathcal{N}(0,1) . \tag{26}$$

with

$$A_2(H,k,a) = \sum_{j\geq 2}\sum_{i\in\mathbb{Z}} (d_{2j}^k)^2\ (2j)!\ \rho_H^a(i)^{2j} , \tag{27}$$

and

$$d_{2j}^k = \frac{1}{(2j)!}\left\{ \prod_{q=0}^{j-1}(2k-2q) - 2\prod_{q=0}^{j-1}(k-2q) \right\} , \qquad j \geq 2. \tag{28}$$

Note that for the particular case $k = 2$ the asymptotic variance is

$$A_2(H,2,a) = \frac{8}{3}\sum_{i\in\mathbb{Z}} \rho_H^a(i)^4 .$$

Furthermore, if $p = 1$ (resp. $p \geq 2$) the results are valid for $0 < H < 7/8$ (resp. $0 < H < 1$). Letting k vary in terms of N, we have a result similar to the one stated in Corollary 1.

COROLLARY 5. *Let a be a filter of order $p > H + 1/8$, and let α be a positive real number. If the observed process is a fBm, we have, for all $H \in]0,1[$,*

$$N^{2\alpha+\frac{1}{2}}\ \{\ P_N(\ N^{-\alpha},a\ )\ -\ 1\ \}\ \xrightarrow{d}\ \mathcal{N}(\ 0,A_2(H,a))\ ,$$

where

$$A_2(H,a) = \sum_{j\geq 2} (\delta_{2j})^2\ (2j)!\sum_{i\in\mathbb{Z}} \rho_H^a(i)^{2j} , \tag{29}$$

and

$$\delta_{2j} = (-1)^j\ \frac{2^{j-1}(j-1)!}{(2j)!}\sum_{q=1}^{j-1}\frac{1}{q} .$$

These results have been extended to the case of Gaussian processes that are locally self-similar at 0, see Coeurjolly (2000), which allows as Bardet's test to extend the reject region for the test, i.e. to construct a test with hypothesis similar to (23). The simulations hereafter illustrate the behavior of the test for such processes.

5. Numerical computations and simulation experiments

In this section, we provide some numerical computations for the asymptotic variances involved in the various estimators we have studied. Due to Corollary 2 and 3, we evaluate the asymptotic variances of $\sqrt{N}\left(\widetilde{H}_N(k,a,M)-H\right)$ and $\sqrt{N}\log(N)\left(\widehat{H}_N(k,a)-H\right)$ for the value $k=2$. Then, we proceed to some Monte Carlo experiments to investigate the finite sample behavior of the methods of estimation that we have described.

5.1. Numerical determination of the asymptotic variance $\sqrt{N}\log(N)\,\widehat{H}_N(2,a)$

In order to proceed to the numerical approximations of the asymptotic constants, we will use hereafter the following notation:

Diff. i	:	denotes the filter of differences of order i (i zero moments).
Diff. 2 dil	:	denotes the filter of differences of order 2 dilated one time.
Db i	:	denotes a Daubechies' wavelet filter of order i.
Coiflets 1	:	denotes a Coiflet's wavelet filter of order 1 (2 zero moments).

Note that the definition of ρ_H^a ensures that the value of the asymptotic constant is independent of the normalization of the filter. One can see, either from the results displayed in Fig. 1, that when $p=1$, π_H^a is not square summable for values of $H>3/4$. Moreover, when $p\geq 2$, the calculations show that the asymptotic variance increases with the filter's order whatever the value of $H\in]0,1[$ is. Among the filters of order 2 that we have studied, Daubechies' filter $Db4$ is the best, and dilating such a filter doesn't improve its performance. We have computed the critical value H_c, above which the filter $a=(1,-1)$, is worse than $Db4$. This allows to derive the best filter (among the filters considered) for the estimator $\widehat{H}_N(2,a)$:

$$a_{opt}=\begin{cases}(1,-1) & \text{if } H\leq H_c\\ Db4 & \text{if } H\geq H_c\end{cases}\qquad \text{with } H_c\simeq 0.641.$$

Since a_{opt} depends on H, an a priori knowledge of H is needed.

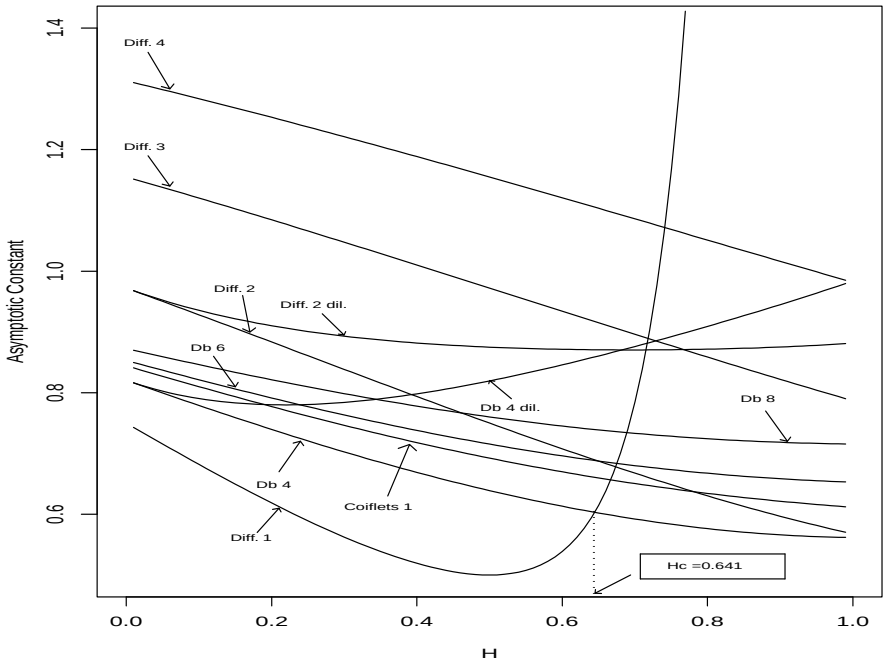


Figure 1. Asymptotic constant $\frac{1}{2}\sum_{i\in\mathbb{Z}} \rho_H^a(i)^2$ for several filters as a function of H

5.2. Numerical determination of the asymptotic variance $\sqrt{N}\left(\widetilde{H}_N(2,a,M)-H\right)$

We have also considered in a numerical study (not presented here), the behaviors of $\sigma^2_{ols}(2)$ in terms of a and M. Among the filters considered, when M is fixed, we note that $\sigma^2_{ols}(2)$ is minimal for the filter $a=(1,-1)$ within the range $]0,3/4[$, and is minimal among the filters of order $p\geq 2$, for $a=Db4, \forall H\in]0,1[$. For these two filters, we made M to vary. It appears that $\sigma^2_{ols}(2)$ is decreasing when M is increasing. However, in a simulation study, we have observed that a too high choice for the parameter M is decreasing the empirical variance but is introducing a small bias in the estimation of the parameters. We have chosen $M=5$ as a good compromise.

5.3. Monte Carlo simulation experiments

We present here some simulations results that illustrate the finite sample behavior of the estimates of the self-similarity parameter and of the scaling coefficient in the non standard case. To simulate a sample path from a fBm on $[0,1]$, we have used the method of Davies (1987), reconsidered by

Beran (1994), and further improved by Wood and Chan (1994), which consists in extracting the square root of the autocovariance matrix of a sampled fBm by embedding it into a circularly matrix which is easily diagonalizable. Among several other known methods for simulating sample paths of a fBm, the method described above has the advantage to be particularly fast and to provide an exact covariance structure. For all our simulations, we have used the same discretization grid $\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\}$. We shall denote Db$q$, the filter of a Daubechies' wavelet filter of order q ($q/2$ zero moments).

We made various simulations, studying the behaviors of the estimators $\widehat{H}_N(k,a)$, $\widetilde{H}_N^{ols}(k,a,M)$ and $\widetilde{C}_N^{ols}(k,a,M)$ in terms of k, of a and of M. In order to illustrate the improvements with respect to the estimators developed by Peltier and Lévy Véhel (1994), Kent and Wood (1997), Istas and Lang (1997), Poggi and Viano (1998), Bardet (1997), we only present here a simulation for the behaviors of $\widehat{H}_N(k,a)$ and $\widetilde{H}_N^{ols}(k,a,M)$ in terms of k. We refer the reader interested in more complete simulations to Coeurjolly (2000). We simulate 50 sample paths of a fBm of length $N = 1000$ and with parameters $C = 1$ and $H = 0.9$. We choose a high value of the self-similarity parameter to illustrate the efficiency of filters of order > 1 when $H > 3/4$. We select the filter $a = Db4$. This is pertinent according to Section 5.1 and 5.2. Moreover, for the estimator $\widetilde{H}_N^{ols}(k,a,M)$ the parameter M has bee fixed to 5. Figures Fig.2 and Fig.3 display boxplots of estimators $\widehat{H}_N(k,a)$ and $\widetilde{H}_N^{ols}(k,a,M)$ respectively. These two figures underline the optimality of the value $k = 2$. When $k > 2$, the estimator remains unbiased, but its variance increases quite rapidly with k. Moreover, choosing values of $k < 1$, doesn't improve the properties of the estimator. Note also that the distributions are well concentrated around the median since there are a small number of extreme values.

Finally, in order to compare rates of convergence of estimators $\widehat{H}_N(2,Db4)$ and $\widetilde{H}_N(2,Db4,5)$, we evaluate empirical mean squared errors (MSE) in terms of N, based on 100 sample paths of a standard fBm with parameter $H = 0.9$. It appears clearly that the rate of convergence of $\widehat{H}_N(2,Db4)$ is greater than the one of $\widetilde{H}_N(2,Db4,5)$.

Table I. MSE in terms of N for estimators $\widehat{H}_N(2,Db4)$ and $\widetilde{H}_N(2,Db4,5)$ based on 100 sample paths of a standard fBm with parameter $H = 0.9$.

Sample Size	50	100	500	1000	10.000
MSE$\left(\widehat{H}_N(2,db4)\right)$	$1.63\,10^{-4}$	$6.68\,10^{-5}$	$1.19\,10^{-5}$	$4.13\,10^{-6}$	$3.47\,10^{-7}$
MSE$\left(\widetilde{H}_N(2,db4,5)\right)$	$2.41\,10^{-2}$	$1.10\,10^{-2}$	$1.62\,10^{-3}$	$8.06\,10^{-4}$	$1.05\,10^{-4}$

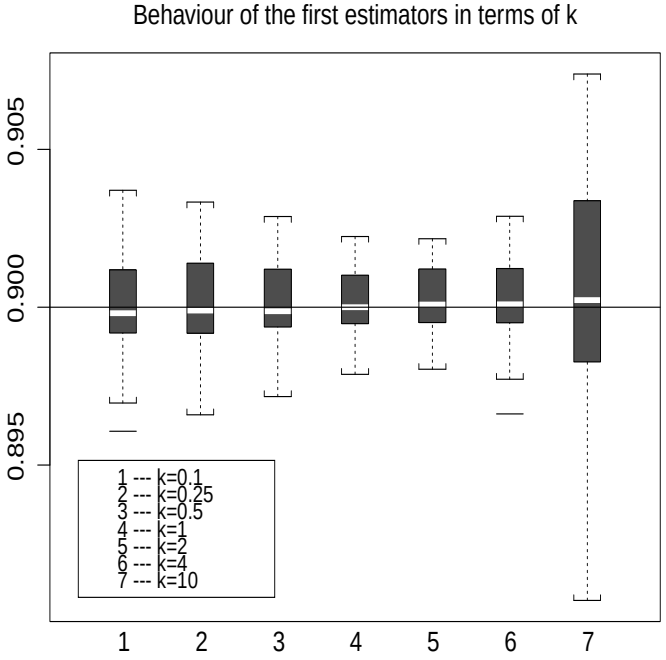


Figure 2. Boxplots of the estimates $\widehat{H}_N(k,a)$ using 50 Monte Carlo simulations of sample paths from a standard fBm with $H = 0.9$ as a function of k ($a = Db4$, $N = 1000$).

5.4. TESTING LSSG PROCESSES.

This section is devoted to the simulation study of our testing procedure for the mean of a lssG process, based on the $P_N(2,a)$ statistic. We have simulated R sample paths from various lssG processes; with each of them, we have derived the following α-level acceptance region

$$\mathrm{I}_{P_N(2,a)} = \left] P_N(2,a) - u_\alpha \frac{\sigma}{\sqrt{N}} \; ; \; P_N(2,a) + u_\alpha \frac{\sigma}{\sqrt{N}} \right[,$$

with $\sigma^2 = \frac{8}{3} \sum_{i\in\mathbb{Z}} \rho^a_{\widetilde{H}}(i)^4$ and where $u_\alpha = \Phi^{-1}(\frac{\alpha}{2})$, and $\widetilde{H}$ is an estimator of H. For accelerating the computations we have chosen $\widetilde{H} = \widetilde{H}^{ols}_N(2,Db4,5)$. To estimate the success rate of our test we have evaluated the following rate $T = \frac{1}{R}\sum_{i=1}^R \mathbf{I}\big(1 \in \mathrm{I}_{P_{i,N}(2,a)} \big)$. We have first applied the above procedure on lssG processes. Recall that a Gaussian process is said to be lssG if its semi-variance function satisfies

$$v^{2D}(t) = v^{2D}(0) + (-1)^D C|t|^{2H} + o(|t|^{2H}) ,$$

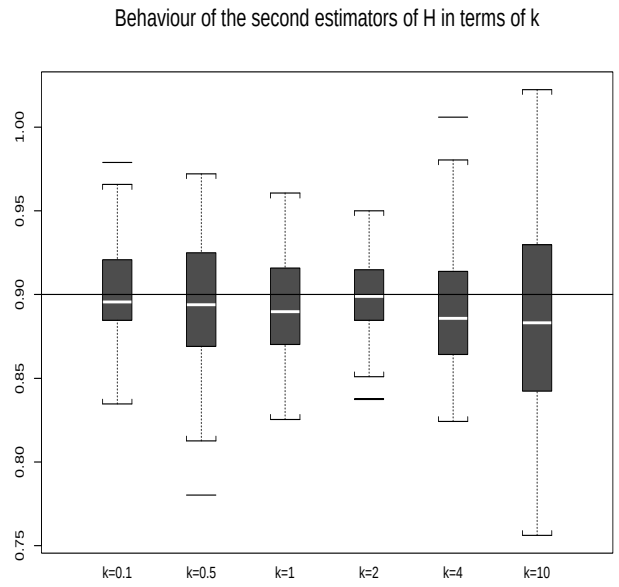


Figure 3. Boxplots of the estimates $\widetilde{H}_N^{ols}(k,a,M)$ using 50 Monte Carlo simulations of sample paths from a standard fBm with $H = 0.9$ as a function of k ($a = Db4$, $M = 5$, $N = 1000$).

where D is the largest integer such that v is $2D$-times differentiable.
To investigate the power of our testing procedure, we have also considered the following alternatives:

– Model 1:

$$X_t^1 = B_H(t) + f(t)\ ,\ f(t) = \gamma\Big(\mathbf{I}(\,t \in]0,1/2]\,) - \mathbf{I}(\,t \in]1/2,1[\,)\Big).$$

– Model 2:

$$X_t^2 = B_H(t) + \varepsilon(t)\ ,$$

for $t = 0,\ldots,N-1/N$. The vector $\{\varepsilon(0),\ldots,\varepsilon(N-1/N)\}$ is a vector of i.i.d. Gaussian variables with mean 0, and variance σ^2.

The results of the tests are summarized in Tab. II. For $\alpha = 0.05$, the test detects the local self-similarity at 0 quite well, for several lssG processes and for relatively small samples ($N = 256$). As for the alternatives, note that, when γ or σ^2 is zero, the process X_t^1 or X_t^2 is nothing else than a fractional Brownian motion. As for its power, as soon as we slightly increase the value of γ or σ, the test detects appropriately a non self-similarity at 0.

Table II. Rates of success for testing the mean on 500 Monte Carlo simulations of $P_N(2,a)$ (with a=Db4), for several lssG processes, and for the X_t^1 and X_t^2 processes.

		Sample size		
Process	Parameters	256	512	1024
fractional Brownian motion	$D=0,\ H=0.9$	95 %	95.6 %	96.2 %
fractional arima	$D=0,\ H=0.9$	94.4 %	97 %	95.8 %
Gaussian process with covariance: $\exp(-\lvert t\rvert^{1.8})$	$D=0,\ H=0.9$	95.2 %	96.8 %	95.2 %
$X_t^1 = (B_H)_t + f(t)$	$\gamma=0.004$	92 %	75 %	3 %
	$\gamma=0.008$	62 %	1.5 %	0 %
$X_t^2 = (B_H)_t + \varepsilon(t)$	$\sigma=0.05$	95 %	84 %	72 %
	$\sigma=0.01$	39 %	18 %	12 %

6. Conclusion

To conclude, we may summarize the results obtained in this work as follows. When the scaling coefficient of a fractional Brownian motion is known, and when the sample paths are discretized on the grid $\{0,\frac{1}{N},\ldots,\frac{N-1}{N}\}$, we have obtained estimates of the self-similarity parameter that are strongly consistent with an asymptotic rate of the order $1/\sqrt{N}\log(N)$ whatever the value of H is. The results of Proposition 2 have been used for testing the quality of various simulation methods of a fBm, see Coeurjolly (2000). However, for practical purposes it is unrealistic to assume that the scaling factor is known. When the scaling factor is unknown, we have introduced two estimators families of the self-similarity parameter that converge at a rate of the order $1/\sqrt{N}$, $\forall\, H\in[0,1]$. In Coeurjolly (2000), we have also considered the following dicretization $i/\Delta_N, i=0,\ldots,N-1$ with $\Delta_N\to+\infty$ as $N\to+\infty$ and we have obtained convergent estimators at a rate $1/\sqrt{N}\log(\Delta_N)$ when C is known, see remark p.24. When C is unknown , we proved that the estimators of H are independent of Δ_N.

We have also derived a testing procedure for testing an fBm assumption and that is convergent at a rate $1/\sqrt{N}$ for any H. The simulations have shown that the behavior of such a test is good. Moreover, the estimation and testing procedures developed in this work are easily implemented and computationally fast.

Appendix

6.1. PRELIMINARY RESULT

Let c_j^k denote the jth coefficient of the development in Hermite polynomials of $\mathbb{H}^k(t) = |t|^k/E_k - 1$: $c_j^k = \mathbb{E}\big(\ \mathbb{H}^k(Y)H_j(Y)\ \big)$, where Y is a standard normal variable. The jth Hermite polynomial H_j is defined by

$$H_j(x) \ = \ \sum_{p=0}^{[j/2]} (-1)^p \frac{j!}{(j-2p)!p!} 2^{-p}\, x^{j-2p},$$

where $[t]$ denotes the greatest integer lower than t.

LEMMA 2. *Let k be a positive real number, then we have*

$$c_{2j+1}^k \ = \ 0 \qquad\qquad \forall\ j \geq 0 \tag{30}$$

$$c_{2j}^k \ = \ \frac{1}{(2j)!} \prod_{i=0}^{j-1} (k-2i)\ , \qquad \forall\ j \geq 1. \tag{31}$$

Proof. (30) is obtained easily by noticing that $\mathbb{H}^k$ is even and that the polynomials $(H_{2j+1})_{j\geq 0}$ are odd. To prove (31), let us denote by P_j, the polynomial with real coefficients defined by

$$P_j(k) \ = \ \mathbb{E}\big(\ \mathbb{H}^k(Y)H_{2j}(Y)\ \big)\ .$$

It follows from the definition of Hermite polynomials that the degree of P_j is j. Indeed, for $j \geq 1$, we have

$$P_j(k) \ = \ \sum_{q=0}^{j} (-1)^q \frac{(2j)!}{(2j-2q)!\, q!\, 2^q}\ \frac{E_{k+2j-2q}}{E_k}\ ,$$

and it is clear that

$$E_{k+2j-2q} \ = \ (k+1)(k+3)\ldots(k+2j-2q-1)\, E_k, \qquad \text{if}\ \ j > q\ .$$

Now choose $n \in \{0,\ldots,j-1\}$. Noting that $\mathbb{H}^{2n}$ is $2j$ times differentiable in $L^2(\phi)$ (the space of square integrable functions with respect to the standard Gaussian measure), we have (see Neveu):

$$\mathbb{E}\big(\ \mathbb{H}^{2n}(Y)\, H_{2j}(Y)\ \big) \ = \ \mathbb{E}\big(\ (\mathbb{H}^{2n})^{(2j)}(Y)\ \big) \ = \ 0\ , \quad Y \stackrel{d}{=} \mathcal{N}(0,1)$$

Thus, the set $\{0,2,4,\ldots,2(j-1)\}$ defines the roots of P_j, and by unicity of the factorization it comes that $P_j(k) \ = \ \prod_{q=0}^{j-1}(k-2q)$.

6.2. k-VARIATIONS

In the following proofs, a denotes a filter of length $\ell+1$ and of order $p \geq 1$.

Proof of Lemma 1. Remember that π_H^a is defined by

$$\pi_H^a(j) \;=\; -\frac{1}{2}\sum_{q,r=0}^{\ell} a_q a_r \,|j+q-r|^{2H}.$$

Choosing $j \geq \ell$ implies

$$\pi_H^a(j) \;=\; -\frac{1}{2} j^{2H} \sum_{q,r} a_q a_r \left(1+\frac{q-r}{j}\right)^{2H}.$$

Thanks to a series expansion for each term of the sum, one has

$$\pi_H^a(j) \;=\; -\frac{1}{2} j^{2H} \sum_{q,r} a_q a_r \left\{ \sum_{s=0}^{2p-1} \frac{\beta_{s,H}}{s!}\frac{(q-r)^s}{j^s} \;+\; \frac{1}{j^{2p}} \sum_{s\geq 2p} \frac{\beta_{s,H}}{s!}\frac{(q-r)^s}{j^{s-2p}} \right\},$$

with $\beta_{0,H}=1$ and $\beta_{q,H} = 2H(2H-1)\ldots(2H-q+1)$, $q>0$, one gets

$$\pi_H^a(j) \;=\; -\frac{1}{2} j^{2H} \sum_{s=0}^{2p-1} \frac{\beta_{s,H}}{s!} \sum_{q,r} a_q a_r (q-r)^s \;-\frac{1}{2}\, j^{2H-2p} \sum_{q,r=0}^{\ell} \sum_{s\geq 2p} \frac{\beta_{s,H}}{s!}\frac{(q-r)^s}{j^{s-2p}}.$$

Because of the property of a, one of the terms $\sum_{q=0}^{\ell}(a_q q^s)\sum_{r=0}^{\ell}(a_r r^{2p-1-s})$ is zero for $s=0,\ldots,2p-1$. So, $\pi_H^a(j) = j^{2H-2p}\,L(j)$ with

$$L(j) = \sum_{q,r=0}^{\ell} \sum_{s\geq p} \frac{\beta_{s,H}}{s!}\,\frac{(q-r)^s}{j^{s-p}} \sim -\frac{1}{2}\sum_{q,r=0}^{\ell} \frac{\beta_{2p,H}}{(2p)!}(q-r)^{2p}, \text{ as } j\to+\infty.$$

Proof of Proposition 1.

(i) The idea is to use a well-known formula on Hermite polynomials, which can be seen as a particular case of the diagram formula (Taqqu (1975), Breuer and Major (1983)): let U and V be two standard Gaussian random variables with correlation coefficient ρ, then

$$\mathbb{E}\big(\, H_i(U)H_j(V)\,\big) \;=\; \delta_{i,j}\, i!\,\rho^i, \qquad \forall\; i,j\geq 1\,. \tag{32}$$

From (3) and (32), we have

$$\begin{aligned}
\mathbb{E}\big(\, V_N(k,a)^2\,\big) \;&=\; \frac{1}{(N-\ell)^2}\sum_{i,j=\ell}^{N-1} \mathbb{E}\Big(\; \mathbb{H}^k(Z^a(i))\mathbb{H}^k(Z^a(j))\;\Big) \\
&=\; \frac{1}{(N-\ell)^2}\sum_{q,r\geq 0} c_q^k\, c_r^k \sum_{i,j=\ell}^{N-1} \mathbb{E}\big(\, H_q(Z^a(i))H_r(Z^a(j))\,\big)\;,
\end{aligned}$$

where $\left(c_q^k, q \geq 1 \right)$ are the coefficients of the development in Hermite polynomials of $\mathbb{H}^k$. By Lemma 2, c_{2q+1}^k and c_0^k are zero, so we get

$$\mathbb{E}\left(V_N(k,a)^2 \right) = \frac{1}{(N-\ell)^2} \sum_{q\geq 1} \alpha_{2q}^k \sum_{i=\ell-N}^{N-\ell} (N-\ell-|i|) \rho_H^a(i)^{2q} ,$$

with $\alpha_{2q}^k = (c_{2q}^k)^2 (2q)!$. From Lemma 1 $\rho_H^a(i) = o(|i|^{2H-2p})$, so that if $4H-4p<1$, i.e. $p > H+1/4$ ($p \geq 2$ and $H \in]0,1[$ or $p=1$ and $H<3/4$), the series $\sum_{i\in\mathbb{Z}} \rho_H^a(i)^{2q}$ converges for each $q \geq 1$, and therefore

$$\mathbb{E}\left(V_N(k,a)^2 \right) = o\left(\frac{1}{N}\right) . \tag{33}$$

Now, if $p=1$ and $H=3/4$, we have:

$$\mathbb{E}\left(V_N(k,a)^2 \right) = o\left(\frac{1}{N} \sum_{|i|<N} \frac{1}{i} \right) = o\left(\frac{\log(N)}{N} \right) . \tag{34}$$

If $p=1$ and $H>3/4$, then

$$\mathbb{E}\left(V_N(k,a)^2 \right) = o\left(\frac{1}{N} \sum_{|i|<N} \frac{1}{i^{2-2H}} \right) = o\left(\frac{1}{N^{2-2H}} \right) . \tag{35}$$

The almost-sure convergence is ensured by (33), (34) or (35), according to the values of p and H, and by Theorem 6.2 of Doob (1953, p.492), establishing a condition under which almost sure convergence is equivalent to mean-squared convergence for the convergence of empirical means of discrete stationary processes.

(ii) Lemma 2 ensures that the Hermite rank of $\mathbb{H}^k$ is 2; furthermore, for $k>0$ and $p > H+\frac{1}{4}$, ρ_H^a is square summable. Consequently, from Theorem 1 of Breuer and Major (1983) we have

$$\sqrt{N}\, V_N(k,a) \xrightarrow{d} \mathcal{N}\left(0, A_1(H,k,a) \right) ,$$

where $A_1(H,k,a) = \lim_{N\to+\infty} \sigma_{N,k,a}^2$, $\sigma_{N,k,a}^2 = \mathbb{E}\left(\frac{1}{N} \left(\sum_{i=\ell}^{N-1} \mathbb{H}^k\left(Z^a(i) \right) \right)^2 \right)$, and from the proof of Proposition 1, one gets

$$\begin{aligned} \sigma_{N,k,a}^2 &= \frac{1}{N} \sum_{j\geq 1} (c_{2j}^k)^2 \, (2j)! \sum_{i=\ell-N}^{N-\ell} (N-\ell-|i|) \, \rho_H^a(i)^{2j} \\ &\overset{N\to+\infty}{\longrightarrow} \sum_{j\geq 1} (c_{2j}^k)^2 \, (2j)! \sum_{i\in\mathbb{Z}} \rho_H^a(i)^{2j} . \end{aligned}$$

Proof of Corollary 1. Let f be the function defined by the following development in Hermite polynomials

$$f(t) \;=\; \sum_{j=1}^{+\infty} \theta_{2j}\, H_{2j}(t)\;, \qquad \text{with}\;\; \theta_{2j} \;=\; (-1)^{j-1} \frac{2^{j-1}\,(j-1)!}{(2j)!}\;.$$

The Stirling formula justifies this development. Indeed,

$$(\theta_{2j})^2\,(2j)! \;\sim\; \frac{\sqrt{\pi}}{4}\,\frac{1}{j^{\frac{3}{2}}}\,, \tag{36}$$

so $\sum_{j\geq 1}\,(\theta_{2j})^2(2j)! < \infty$. Let X_N be the random variable defined by

$$X_N \;=\; \frac{1}{N-\ell}\,\sum_{i=\ell}^{N-1} f(Z^a(i))\;.$$

The function f being of Hermite rank 2, Theorem 1 of Breuer and Major (1983) ensures us, as soon as $p > H + \frac{1}{4}$, that as $N \to +\infty$

$$\sqrt{N}\,X_N \;\xrightarrow{d}\; \mathcal{N}\big(\,0, A_1(H,a)\,\big)\;,$$

where $A_1(H,a) \;=\; \sum_{j\geq 1}(\theta_{2j})^2\,(2j)!\,\sum_{i\in\mathbb{Z}} \rho_H^a(i)^{2j}$. Now, let us remark that $c_{2j}^{N^{-\alpha}} \sim \frac{1}{N^\alpha}\theta_{2j}$ as $N \to +\infty$. So, if we prove the convergence in probability of $X_N' = \sqrt{N}\,(X_N - N^\alpha V_N(N^{-\alpha},a))$ to 0, Slutsky's Theorem, *e.g.* Grimmet et *al.* (1992), will ensure, that, as $N \to +\infty$, $N^{\alpha+1/2}V_N(N^{-\alpha},a) \xrightarrow{d} \mathcal{N}(0, A_1(H,a))$. We have,

$$X_N' \;=\; \frac{\sqrt{N}}{N-\ell}\,\sum_{i=\ell}^{N-1}\,\sum_{j\geq 2} \Theta_{2j,N}\,H_{2j}(Z^a(i)) \quad \text{with} \quad \Theta_{2j,N} = \;\theta_{2j} - N^\alpha\,c_{2j}^{N^{-\alpha}}.$$

We obtain for the variance of X_N'

$$\mathbb{E}\big(\;{X_N'}^2\;\big) \;=\; \sum_{j\geq 2} \frac{N}{(N-\ell)^2}\,(\Theta_{2j,N})^2(2j)!\,\sum_{i=\ell-N}^{N-\ell}(N-\ell-|i|)\;\rho_H^a(i)^{2j}.$$

Since $p > H + 1/4$, we have for all $j \geq 2$, $\sum_{|i|<N-\ell}\,(N-\ell-|i|)\;\rho_H^a(i)^{2j} = \mathcal{O}(N)$. Then, from Lemma 2 one can prove that

$$\Theta_{2j,N} = \theta_{2j} - N^\alpha c_{2j}^{N^{-\alpha}} \sim \frac{1}{N^{2\alpha}}(-\theta_{2j})\,\sum_{q=1}^{j-1}\frac{1}{2q},\ \text{as}\ N \to +\infty.$$

Define $\theta_{2j}' = -\theta_{2j}\sum_{q=1}^{j-1}\frac{1}{q}$, we have from (36), $(\theta_{2j}')^2(2j)! \sim \frac{\sqrt{\pi}}{16}\frac{\log(j)^2}{j^{3/2}}$, which allows us to finally obtain, as $N \to +\infty$

$$\mathbb{E}\big(\;{X_N'}^2\;\big) \;=\; \mathcal{O}\left(\frac{1}{N^{4\alpha}}\sum_{j\geq 2}\frac{\log(j)^2}{j^{3/2}}\right) = \mathcal{O}\,(N^{-4\alpha}).$$

And thus X'_N converges in probability to 0.

6.3. IDENTIFICATION OF THE FBM

Proof of Proposition 2.

(i) Note that

$$(\,1+V_N(k,a)\,) = N^{k(H-\widehat{H}_N(k,a))}\left(\frac{\pi^a_{\widehat{H}_N(k,a)}(0)}{\pi^a_H(0)}\right)^{\frac{k}{2}}. \tag{37}$$

From the almost sure convergence of $V_N(k,a)$ to 0, and the continuity of the logarithm in 1, we have

$$k\,\log(N)(H-\widehat{H}_N(k,a)\,)+\frac{k}{2}\log\left(\frac{\pi^a_{\widehat{H}_N(k,a)}(0)}{\pi^a_H(0)}\right)\xrightarrow{a.s.}0\,,$$

so,

$$(\,H-\widehat{H}_N(k,a)\,)+\frac{1}{2\,\log(N)}\log\left(\frac{\pi^a_{\widehat{H}_N(k,a)}(0)}{\pi^a_H(0)}\right)\xrightarrow{a.s.}0\,. \tag{38}$$

Again from the almost sure convergence of $V_N(k,a)$ one has: $N^{Hk}S_N(k,a)\overset{a.s.}{\to}\{\pi^a_H(0)\}^{k/2}E_k$. Thus, there exists $N_0\in\mathbb{N}^*$, such that $\forall N\geq N_0$ one has almost surely

$$S_N(k,a)\leq\frac{2}{N^{Hk}}\{\pi^a_H(0)\}^{k/2}E_k.$$

Now assume $H\in]\varepsilon,1[$ for $\varepsilon>0$, then there exists $N_1\in\mathbb{N}^*$ such that $\forall N\geq N_1$ one has almost surely

$$0<S_N(k,a)<\frac{1}{N^{\varepsilon k}}\{\pi^a_\varepsilon(0)\}^{k/2}E_k=g_{k,a,N}(\varepsilon).$$

So, there exists $N_1\in\mathbb{N}^*$ such that $\forall N\geq N_1$ one has almost surely

$$\widehat{H}_N(k,a)=g^{-1}_{k,a,N}(S_N(k,a))\in]\varepsilon,1[,\quad\forall\varepsilon>0.$$

This equation and (38) ensure the almost sure convergence of $\widehat{H}_N(k,a)$.

(ii) The almost sure convergence of $\widehat{H}_N(k,a)$ and the differentiability of $\pi^a_\cdot(0)$ prove that

$$\frac{\pi^a_{\widehat{H}_N(k,a)}(0)}{\pi^a_H(0)}-1=\frac{\sum^{\ell}_{q,r=0,\ |q-r|\neq0}2a_qa_r\log(|q-r|)|q-r|^{2H}}{\sum^{\ell}_{q,r=0}a_qa_r|q-r|^{2H}}\left(\widehat{H}_N(k,a)-H\right)(1+o(1))\,,$$

holds almost surely. There exists κ_H such that one obtains almost surely

$$\frac{k}{2}\log\left(\frac{\pi^a_{\widehat{H}_N(k,a)}(0)}{\pi^a_H(0)}\right) \;=\; \kappa_H\left(\widehat{H}_N(k,a)-H\right)(1+o(1))\,. \tag{39}$$

Equations (37) and (39) imply that almost surely

$$\begin{aligned}\log(1+V_N(k,a)) &= k\log(N)\left(H-\widehat{H}_N(k,a)\right)\left(1+\frac{\kappa_H}{\log(N)}(1+o(1))\right)\\ &= k\log(N)\left(H-\widehat{H}_N(k,a)\right)(1+o(1))\,. \end{aligned}\tag{40}$$

Moreover,

$$\log(\,1+V_N(k,a)\,) \;=\; V_N(k,a)\,(1+o(1))\,, \tag{41}$$

holds almost surely. Equations (40) and (41) ensure that almost surely one gets

$$V_N(k,a) \;=\; k\log(N)\left(H-\widehat{H}_N(k,a)\right)(1+o(1))\,. \tag{42}$$

From Slutsky's Theorem, *e.g.* Grimmet et *al.* (1992), we thus deduce that $k\sqrt{N}\log(N)\left(\widehat{H}_N(k,a)-H\right)$ tends in distribution to the same limit as $\sqrt{N}V_N(k,a)$.

Remark: assume observing the fBm at times i/Δ_N for $i=0,\ldots,N-1$ with $\Delta_N\to+\infty$ as $N\to+\infty$. Define

$$\begin{aligned} S_{\Delta_N}(k,a) &= \frac{1}{N-\ell}\sum_{i=\ell}^{N-1}|V^a(i/\Delta_N)|^k\\ V_{\Delta_N}(k,a) &= \frac{1}{N-\ell}\sum_{i=\ell}^{N-1}\left|\frac{V^a(i/\Delta_N)}{\mathbb{E}(V^a(i/\Delta_N)^2)^{1/2}}\right|^k,\\ \text{and}\quad \widehat{H}_{\Delta_N}(k,a) &= g_{k,a,\Delta_N}\left(S_{\Delta_N}(k,a)\right), \end{aligned}$$

where

$$g_{k,a,\Delta_N}(t) \;=\; \frac{\pi^a_t(0)}{\Delta_N^{tk}}E_k, \qquad \text{for } t\in]0,1[.$$

We leave the reader to check that $V_{\Delta_N}(k,a)\stackrel{d}{=}V_N(k,a)$. Then, since $\Delta_N\to+\infty$ as $N\to+\infty$, the almost sure convergence of $\widehat{H}_{\Delta_N}(k,a)$ can be proved as the same way as the one of $\widehat{H}_N(k,a)$.
Then, one can obtain a result similar to (42), i.e.

$$V_N(k,a) \;=\; k\log(\Delta_N)\left(H-\widehat{H}_{\Delta_N}(k,a)\right)(1+o(1))\,.$$

And again from Slutsky's theorem, we conclude, as soon as $\Delta_N \to +\infty$ as $N \to +\infty$, that $k\sqrt{N}\log(\Delta_N)\left(H - \widehat{H}_{\Delta_N}(k,a)\right)$ tends in distribution to the same limit as $\sqrt{N}V_N(k,a)$.

Proof of Proposition 3 .

(i) is obvious from Proposition 1.

(ii) The asymptotic normality of the k-variations vector is obtained as in Istas and Lang (1997), when $k = 2$. The generic entries of the asymptotic covariance matrix are evaluated from

$$\lim_{N\to+\infty} N\,\mathbb{E}\big(\ V_N(k,a^m)\,V_N(k,a^n)\ \big)\,, \qquad \text{for } m,n \in \{1,\dots,M\}\ .$$

We have:

$$\begin{aligned} & N\,\mathbb{E}\big(\ V_N(k,a^m)\,V_N(k,a^n)\ \big) \\ &= \frac{N}{(N-m\ell)(N-n\ell)} \sum_{q=m\ell}^{N-1}\sum_{r=n\ell}^{N-1} \mathbb{E}\big(\ \mathbb{H}^k(Z^{a^m}(q))\,\mathbb{H}^k(Z^{a^n}(r))\ \big) \\ &\overset{N\to+\infty}{\longrightarrow} \sum_{j\geq 1} (c_{2j}^k)^2\,(2j)!\sum_{i\in\mathbb{Z}} \rho_H^{a^m,a^n}(i)^{2j}\ . \end{aligned}$$

(iii) It suffices to note that almost surely we have

$$\log S_N(k,a^m) \;-\; Hk\log(m) \;-\; \log\mathbb{E}(S_N(k,a)) \;=\; V_N(k,a^m)\big(1+o(1)\big). \quad (43)$$

Proof of Proposition 4.

(i) Let ξ_N be the vector defined by $\xi_N = L_N - X_M\alpha$. Theory of linear regression shows that

$$\begin{aligned} \widetilde{H}_N^{ols}(k,a,M) \;-\; H \;&=\; \frac{A^t}{k\,\|A\|^2}\,L_N\ , \\ \text{and} \qquad \widetilde{\theta}_N^{ols}(k,a,M) \;-\; \theta \;&=\; \frac{(A')^t}{k\,\|A'\|^2}\,L_N\ , \end{aligned}$$

where A and A' are the vectors defined for $m = 1,\dots,M$ by

$$\begin{aligned} A_m \;&=\; \log(m) \;-\; \frac{1}{M}\sum_{m=1}^{M}\log(m)\ , \\ A'_m \;&=\; \frac{1}{M}\sum_{m=1}^{M}\log(m)^2 \;-\; \log(m)\,\frac{1}{M}\sum_{m=1}^{M}\log(m)\ , \end{aligned}$$

The result (i) of Proposition 3 implies that $\xi_N \xrightarrow{p.s.} (0,0,\ldots,0)^t$. The convergence of $\widetilde{H}_N^{ols}$ and $\widetilde{\theta}_N^{ols}$ therefore is ensured. From Theorem 6.2 of Doob (1953, p.492), one can prove that $\log(N)V_N(k,a) \xrightarrow{a.s.} 0$, which allows us to obtain

$$\log(N)\ (\ \widetilde{H}_N^{ols}(k,a,M) - H\ )\ \xrightarrow{a.s.}\ 0\ ,$$

which ensures from (17) the almost sure convergence of $\widetilde{C}_N^{ols}$ to C.

(ii) For $p > H + 1/4$, $\widetilde{H}_N^{ols}(k,a,M)$ being defined as a linear combination of random variables that are asymptotically Gaussian, is asymptotically Gaussian. Furthermore,

$$\begin{aligned}\mathbb{E}(\ \{\ \sqrt{N}\ (\widetilde{H}_N^{ols}(k,a,M) - H)\ \}^2\ ) \quad &= \quad \frac{A^t}{k^2\ \|A\|^4}\ \mathbb{E}(\ N\ \xi_N \xi_N^t\ )\ A \\ &\xrightarrow{N\to+\infty}\ \frac{1}{\|A\|^4}\ A^t\ \frac{G_k}{k^2}\ A\ .\end{aligned}$$

In the same way, one easily shows that

$$\sqrt{N}(\ \widetilde{\theta}_N^{ols}(k,a,M) - \theta\ )\ \xrightarrow{d}\ \mathcal{N}\ (\ 0, \frac{(A')^t\ G_k'\ A'}{\|A\|^4}\ )\ . \tag{44}$$

We will notice that

$$\log(\ \widetilde{C}_N^{ols}(k,a,M)\ ) - \log(C)\ =\ \frac{1}{C}(\ \widetilde{C}_N^{ols}(k,a,M) - C\ )\ (\ 1 + o(1)\ )\ ,$$

and also that

$$\begin{aligned}\log(\ \widetilde{C}_N^{ols}(k,a,M)\ ) - \log(C) &= (\ \widetilde{H}^{ols} - H\ )\log(N) \\ &\quad + \frac{1}{2}\log\left(\ \frac{\pi_H^a(0)}{\pi_{\widetilde{H}^{ols}}^a(0)}\ \right) + \frac{1}{k}(\ \widetilde{\theta}_N^{ols}(k,a,M) - \theta\ )\ .\end{aligned}$$

From Eq.(44) and Proposition 4 (i), we have

$$\frac{\sqrt{N}}{\log(N)}\ \left\{\ \frac{1}{2}\log\left(\ \pi_H^a(0)\ /\ \pi_{\widetilde{H}^{ols}}^a(0)\ \right) + \frac{1}{k}(\ \widetilde{\theta}_N^{ols}(k,a,M) - \theta\ )\ \right\}\ \xrightarrow{d}\ 0\ ,$$

which implies that $\frac{\sqrt{N}}{\log(N)}\ \frac{1}{C}\ (\widetilde{C}_N^{ols}(k,a,M) - C)$ tends in distribution to the same limit as $\sqrt{N}(\ \widetilde{H}_N^{ols}(k,a,M) - H\ )$.

6.4. TESTS FOR A FRACTIONAL BROWNIAN MOTION

Proof of Corollary 4.

(i) It suffices to note that

$$P_N(k,a) \;=\; \frac{V_N(2k,a)+1}{(\,V_N(k,a)+1\,)^2}\,, \tag{45}$$

and to use the almost sure convergence of $V_N(k,a)$ established by Proposition 1.

(ii) From Eq.(45), we can write

$$\begin{aligned} P_N(k,a) \;-\; 1 \;&=\; (\,V_N(k,a)+1\,)^{-2}\,\{\,V_N(2k,a)+1-(V_N(k,a)+1)^2\,\} \\ &=\; (\,V_N(k,a)+1\,)^{-2}\{\,V_N(2k,a)-2V_N(k,a)-V_N(k,a)^2\,\}\,. \end{aligned}$$

From Proposition 1, $\sqrt{N}\big(P_N(k,a)-1\big)$ tends in distribution to the same limit as $\sqrt{N}\big(V_N(2k,a)-2V_N(k,a)\big)$. Moreover,

$$\begin{aligned} V_N(2k,a) \;-\; 2\,V_N(k,a) \;&=\; \frac{1}{N-\ell}\sum_{i=\ell}^{N-1}\Big\{\;\mathbb{H}^{2k}\;-2\mathbb{H}^k\;\Big\}\,(Z^a(i)) \\ &=\; \frac{1}{N-\ell}\sum_{i=\ell}^{N-1} d_{2j}^k\,H_{2j}(Z^a(i))\,, \end{aligned}$$

where the coefficients of the development in Hermite polynomials of $\mathbb{H}^{2k}-2\mathbb{H}^k$ are defined by:

$$d_{2j}^k \;=\; \frac{1}{(2j)!}\left\{\;\prod_{q=0}^{j-1}(2k-2q)-2\prod_{q=0}^{j-1}(k-2q)\;\right\}\,. \tag{46}$$

The Hermite rank of $\mathbb{H}^{2k}-2\mathbb{H}^k$ is equal to 4. Therefore, if $p>H+\frac{1}{8}$, $(\rho_H^a)^4$ is summable, and from Theorem 1 of Breuer and Major (1983) there exists $A_2(H,k,a)$ such that:

$$\sqrt{N}\Big(\;V_N(2k,a)\;-2V_N(k,a)\;\Big)\;\xrightarrow{\;d\;}\;\mathcal{N}\,(\,0,A_2(H,k,a)\,)\,. \tag{47}$$

To evaluate $A_2(H,k,a)$, one will note that

$$\mathbb{E}\Big(N\big\{V_N(2k,a)-2V_N(k,a)\big\}^2\Big)=\frac{1}{N}\sum_{j\geq 2}\sum_{i=\ell-N}^{N-\ell}(N-\ell-|i|)(d_{2j}^k)^2(2j)!\rho_H^a(i)^{2j}.$$

Thus, we get

$$A_2(H,k,a) \;=\; \sum_{j\geq 2}\sum_{i\in\mathbb{Z}}\,(d_{2j}^k)^2\;(2j)!\;\rho_H^a(i)^{2j}\,.$$

We have the result since $\widetilde{H} = \widetilde{H}_N^{ols}(k,a,M)$ or $\widetilde{H}_N^{gls}(k,a,M)$ is a consistent estimator of H.

Proof of Corollary 5. The proof parallels the proof of Corollary 1. Let F be the function defined by the following development in Hermite polynomials

$$F(t) = \sum_{j\geq 2} \delta_{2j} H_{2j}(t), \qquad \text{with } \delta_{2j} = (-1)^j \frac{2^{j-1}(j-1)!}{(2j)!} \sum_{q=1}^{j-1} \frac{1}{q}.$$

The Stirling formula justifies this development. Indeed,

$$(\delta_{2j})^2 (2j)! \sim \frac{\sqrt{\pi}}{4} \frac{\log(j)^2}{j^{\frac{3}{2}}},$$

so, $\sum_{j\geq 2} (\delta_{2j})^2 (2j)! < \infty$. Let Y_N be the random variable defined by

$$Y_N = \frac{1}{N-\ell} \sum_{i=\ell}^{N-1} F(Z^a(i)).$$

The function F being of Hermite rank 4, Theorem 1 of Breuer and Major Breuer and Major (1983) ensures us, as soon as $p > H + 1/8$, that as $N \to +\infty$

$$\sqrt{N} Y_N \xrightarrow{d} \mathcal{N}\left(0, \sigma^2(H,a)\right), \text{ where } \sigma^2(H,a) = \sum_{j\geq 2} (\delta_{2j})^2 (2j)! \sum_{i\in\mathbb{Z}} \rho_H^a(i)^{2j}.$$

Now, let us remark that $d_{2j}^{N^{-\alpha}} \sim \frac{\delta_{2j}}{N^{2\alpha}}$ as $N \to +\infty$. So, if we prove the convergence in probability of $Y_N' = \sqrt{N}\left(Y_N - N^{2\alpha}\{V_N(2N^{-\alpha},a) - 2V_N(N^{-\alpha},a)\}\right)$ to 0, Slutsky's Theorem, *e.g.* Grimmet et *al.* (1992), will ensure that, as $N \to +\infty$, $N^{2\alpha+1/2}\{V_N(2N^{-\alpha},a) - 2V_N(N^{-\alpha},a)\} \xrightarrow{d} \mathcal{N}(0,\sigma^2(H,a))$. We have,

$$Y_N' = \frac{\sqrt{N}}{N-\ell} \sum_{i=\ell}^{N-1} \sum_{j\geq 3} \Delta_{2j,N} H_{2j}(Z^a(i)) \quad \text{with} \quad \Delta_{2j,N} = \delta_{2j} - N^{2\alpha} d_{2j}^{N^{-\alpha}}.$$

We obtain for the variance of Y_N'

$$\mathbb{E}\left(Y_N'^2\right) = \frac{N}{(N-\ell)^2} \sum_{j\geq 3} (\Delta_{2j,N})^2 (2j)! \sum_{|i|<N-\ell} (N-\ell-|i|)\rho_H^a(i)^{2j}.$$

Since $p > H + 1/8$, we have for all $j \geq 3$, $\sum_{|i|<N-\ell}(N-\ell-|i|)\rho_H^a(i)^{2j} = o(N)$. Then, from (28) one can prove that

$$\Delta_{2j,N} = \delta_{2j} - N^{2\alpha} d_{2j}^{N^{-\alpha}} \sim \frac{1}{N^{3\alpha}} \delta_{2j}'.$$

where we define $\delta'_{2j} = 6(-1)^{j-1}\frac{2^{j-1}(j-1)!}{(2j)!}\sum_{q=1}^{j-2}\sum_{r=q+1}^{j-1}\frac{1}{(2r)(2s)}$. From Stirling's formula,

$$(\delta'_{2j})^2(2j)! \sim \frac{9\sqrt{\pi}}{4}\frac{1}{j^{3/2}}\left\{\sum_{r=1}^{j-2}\sum_{s=r+1}^{j-1}\frac{1}{rs}\right\}^2.$$

There exists $J \in \mathbb{N}^*$ such that $\forall\, j \geq J$ one has

$$\begin{aligned}(\delta'_{2j})^2(2j)! &= \frac{1}{j^{3/2}}O\left(\left\{\sum_{q=1}^{j-2}\frac{log(r)}{r}\right\}^2\right) = \frac{1}{j^{3/2}}O\left(\left\{\sum_{q=1}^{j-2}\frac{1}{r^{3/8}}\right\}\right)\\ &= \frac{1}{j^{3/2}}O\left(j^{1/4}\right) = O\left(j^{-5/4}\right).\end{aligned}$$

We finally obtain, as $N \to +\infty$

$$\mathbb{E}\left(Y_N'^2\right) = O\left(\frac{1}{N^{6\alpha}}\sum_{j\geq 3}\frac{1}{j^{5/4}}\right) = O(N^{-4\alpha}).$$

Thus, Y'_N converges in probability to 0.

Acknowledgements

I would like to thank A. Antoniadis, my thesis advisor, J. Istas and the referee for their comments and suggestions.

References

Abry, P., Gonçalvès, P., and Flandrin, P.: Wavelets, spectrum analysis and $1/f$ processes. *Wavelets and Statistics*. Lecture Notes in Statistics, Eds A. Antoniadis and G. Oppenheim, **103**, 15–29. Springer-Verlag, New York, 1995.

Abry, P. and Veitch, D.: Wavelet Analysis of Long-Range-Dependent Traffic, *IEEE Transactions on Information Theory* **14** (1998), 2-15.

Antoniadis, A., Berruyer, J. and Carmona, R.: *Régression non linéaire et applications*, Economica, Paris ,1992.

Bardet, J.M.: Test d'auto-similarité pour des processus Gaussiens. Dimension fractale et dimension de corrélation, Thesis, University of Paris Sud, 1997.

Benassi, A., Jaffard, S. and Roux, D.: Gaussian processes and Pseudodifferential Elliptic operators *Revista Mathematica Iberoamericana* **13** (1998), 19-89.

Beran, J.: A goodness of fit test for time series with long range dependence, *J. Roy. Statist. Soc.* B, **54**, 3, (1992), 749-760.

Beran, J.: *Statistics for long memory processes*, Chapman and Hall, London, 1994.

Breuer, P. and Major, P.: Central limit theorems for non-linear functionals of Gaussian fields, *J. Multiv. Anal.* **3** (1983), 425-441.

Coeurjolly, J.F.: Identification du mouvement brownien fractionnaire par variations discrètes, *Technical report* 1016-M-LMC, Univ. J. Fourier, Grenoble (1999).

Coeurjolly, J-F.: Inférence statistique pour les mouvements browniens fractionnaire et multifractionnaire, *Phd Thesis*, Univ. J.Fourier, Grenoble (2000).

Coeurjolly, J-F.: Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study, *Journal of Statistical Software* **5** (2000).

Collins, J.J., De Luca, C.J.: Upright, correlated random walks: A statistical-biomechanics approach to the human postural control system, *Chaos* **5**, 1, (1994), 57-63.

Davies, B.: Tests for Hurst effect, *Biometrika* **74** (1987), 95-101.

Daubechies, I: *Ten lectures on wavelets*, SIAM, Philadelphia, 1992.

Doob, J.L.: *Stochastic processes*, Wiley Classics Library, USA, 1953.

Granger, C.W.J.: The typical spectral shape of an economic variable, *Econometrica* **34** (1966), 150-161.

Grimmet, G.R., Stirzaker D.R.: *Probability and Random Processes*, Clarendon Press, Oxford (1992).

Guyon, X. and Leon, J.: Convergence en loi des H-variations d'un processus Gaussien stationnaire sur $\mathbb{R}$, *Ann. Inst. H. Poincaré* **25** (1989), 265-282.

Higuchi, T.: Approach to an irregular time series on the basis of the fractal theory *Physica D* **31** (1988), 277-283.

Istas, J. et Lang, G.: Quadratic variations and estimation of the Hölder index of a Gaussian process, *Ann. Inst. Henri Poincaré* **33** (1997), 407-436.

Kent, J.T. and Wood, A.T.A.: Estimating the fractal dimension of a locally self-similar Gaussian process using increments, *Journal of the Royal Statistical Society Series B* **59** (1997), 679-700.

Kuklinski, W.S., Chandra, K., Ruttimann, U.E., and Webber, R.L.: Application of Fractal Texture Analysis to *Segmentation of Dental Radiographs SPIE, Medical Imaging III: Image Processing* **1092** (1989), 111-117.

McLeod, A.I., Hipel, K.W.: Preservation of the Rescaled Adjusted Range: a reassessment of the Hurst exponent *Water Ressources Research* **14**, 3 (1978), 491-508.

Mandelbrot, B., Van Ness and Wallis, J.: Fractional Brownian motions, fractional noises and applications, *SIAM review* **10** (1968), 422-437.

Neveu, J.: Processus aléatoires Gaussiens, Séminaire de Mathématiques Supérieures, Cours Montréal, p.155, 1968.

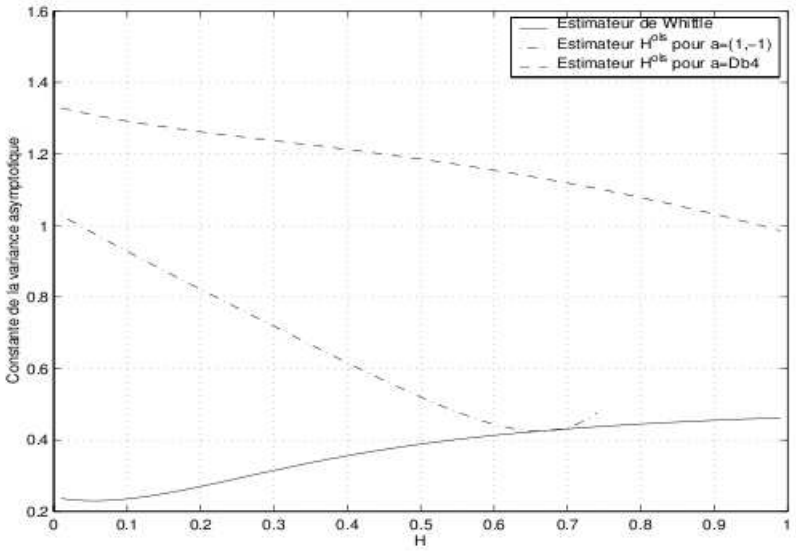
Peltier, R.F. and Lévy-Véhel, J.: A new method for estimating the parameter of fractional Brownian motion, *Research report*, INRIA Rocqencourt, 1994.

Poggi, J.M. and Viano, M.C.: An estimate of the fractal index using multiscale aggregates, *Journal of Time Series Analysis* **19** (1998), 221-233.

Taqqu, M.S.: Weak convergence to fractional Brownian motion and to the Rosenblatt process, *Z. Wahrsch. Verw. Gebiete* **40** (1975), 203-238.

Willinger, W., Taqqu, M.S., Leland, W.E., and Wilson, D.V.: Self-similarity in high speed packet traffic: analysis and modelisation of ethernet traffic measurements *Statistical Science* **10** (1995), 67-85.

Wood, A. and Chan, G.: Simulation of stationary Gaussian processes in $[0,1]^d$, *Journal of computational and graphical statistics* **3** (1994), 409-432.



[CI01]
J.-F. Coeurjolly and J. Istas. Cramèr-Rao bounds for fractional Brownian motions. *Stat. Probab. Lett.*, 53(4):435-447, 2001.

ELSEVIER

Statistics & Probability Letters 53 (2001) 435–447

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Cramèr–Rao bounds for fractional Brownian motions

Coeurjolly Jean-François[a,*], Istas Jacques[b]

[a] *LMC-IMAG, BP 53-F-38041 Grenoble Cedex 9, France*
[b] *LabSad, BP 47, F-38040 Grenoble Cedex 9, France*

Received September 1999; received in revised form August 2000

Abstract

We obtain Cramèr–Rao bounds for parameters estimators of fractional Brownian motions. We point out the differences of behavior whether these processes are standard or not. The key-point of this study relies upon a linear algebra result we prove, exhibiting bounds for elements of inverse of localized matrices.

Keywords: Cramèr–Rao bounds; Fractional Brownian motion; Invertibility of localized matrices

1. Introduction

Many naturally occurring phenomena can be effectively modeled using self-similar, or locally self-similar processes, see for instance Lévy-Véhel et al. (1997) for various examples, (Rao, 1991) for image analysis, (Frisch, 1995) for turbulence theory. In the Gaussian framework, the only self-similar processes are the fractional Brownian motions (FBM), first introduced in a theoretical setup by Kolmogorov (1940), and popularized by Mandelbrot and Van Ness (1968). The FBM $\{B_{H,C}(t),\ t \geqslant 0\}$, of fractional parameter H, $0 < H < 1$, and scaling parameter C, $C > 0$, is the only Gaussian centered self-similar process with stationary increments such that:

$$\mathbf{E}(B_{H,C}(t) - B_{H,C}(s))^2 = C^2|t-s|^{2H} \quad \forall s,t \in \mathbb{R}^+.$$

In this paper, we call standard fractional Brownian motion (SFBM), the FBM with scaling parameter $C \equiv 1$, and we denote by $\{B_H(t),\ t \geqslant 0\}$ the SFBM.

The identification of the parameters of FBM and SFBM from the observations of discretization of a unique sample path is an interesting question. Assume that one observes an FBM at times i/Δ_N, $i = 0,\ldots,N-1$. The maximum likelihood methods (Beran, 1994), wavelets methods (Abry et al., 1995) and generalized quadratic variations methods (Istas and Lang, 1994, 1997; Benassi et al., 1988; Coeurjolly, 1999) lead to consistent estimators of H with $\sqrt{N}$-rate of convergence with Gaussian limit distribution for every H, $0 < H < 1$. For

* Corresponding author.
E-mail addresses: jean-francois.coeurjolly@imag.fr (J.-F. Coeurjolly), jacques.istas@upmf-grenoble.fr (J. Istas).

0167-7152/01/$ - see front matter
PII: S0167-7152(00)00197-8

the SFBM, Peltier and Lévy-Véhel (1994), for a uniform discretization (i.e. $\Delta_N = N$) and for $0 < H < 3/4$ obtain consistent estimators with $\sqrt{N}\log(N)$-rate of convergence. Coeurjolly (1999) improves this result for $0 < H < 1$, and obtains, for any discretization step $\Delta_N \to +\infty$, consistent estimators with $\sqrt{N}\log(\Delta_N)$-rate of convergence.

In order to test the efficiency of an estimator, one usually compares its mean squared error with the Cramèr–Rao bound (CRB) of the statistical model. Let us recall some standard notions about CRB. Let $x = (x_1, \ldots, x_N)$ be a random vector whose probability distribution depends on an unknown parameter θ. Under suitable smoothness assumptions (e.g. Dacunha-Castelle and Duflo, 1994, p. 178), one obtains that every unbiased estimator $\hat{\theta}$ satisfies the following inequality:

$$\mathbf{E}(\hat{\theta} - \theta)^2 \geqslant I_N(\theta)^{-1} \equiv \mathrm{CRB}_N(\theta), \tag{1}$$

where $I_N(\theta)$ is the Fisher information matrix defined by

$$I_N(\theta) = \left\{ \mathbf{E}\left(\frac{\partial}{\partial\theta_i} \log L(x,\theta) \frac{\partial}{\partial\theta_j} \log L(x,\theta) \right) \right\}_{1 \leqslant i,\, j \leqslant k} \tag{2}$$

where $L(x,.)$ is the likelihood of the model. A similar inequality for biased estimators, is also available (e.g. Dacunha-Castelle and Duflo, 1994, p. 178).

Our statistical models are given by the observations of an FBM or SFBM at times i/Δ_N, $i = 0, \ldots, N-1$. The parameters of interest are H and C for FBM, H for SFBM. The aim of this paper is to give the asymptotical rates of the CRB for these two models associated with FBM and SFBM. The results are given in Section 2. Section 3 presents an extension of a result of Jaffard (1990), where fine estimations of the inverse of an infinite matrix whose terms decay polynomially far away from the diagonal are given. We apply this result in Section 4 to prove theorems presented in Section 1.

These results are partially known. Dahlhaus (1989) exhibits the Cramèr–Rao bounds of parameters estimators for various self-similar processes. In particular, the increments process of the FBM, called the fractional Gaussian noise (FGN) is studied. The key-point relies upon the fact that for this model, one controls $\mathrm{Tr}\left((\mathrm{d}/\mathrm{d}H)\, G_H \cdot G_H^{-1}\right)$, where G_H denotes the covariance matrix of the FGN. Our approach is different: on the one hand, we discuss the cases C known and C unknown. On the other hand, instead of studying the increments process of the FBM, we exhibit a linear transform of our statistical model for which we control the coefficients behaviors of the covariance matrix, its inverse and the one built as the derivative w.r.t. H.

2. Main results

We now state the main results of the paper, firstly for SFBM and secondly for FBM. Recall that $\mathrm{CRB}_N(.)$ denotes the Cramèr–Rao bounds of a parametric model, associated with a random vector of size N. From now on, we use the following notations for x_N and y_N, two real sequences: $x_N \sim y_N$ if the limit $\lim_{N\to+\infty} x_N/y_N$ exists and equals 1. Furthermore, $x_N = \mathcal{O}(y_N)$ (resp. $x_N \asymp y_N$) if there exists constant $c > 0$ (resp. $c_1, c_2 > 0$), such that $|x_N| \leqslant c|y_N|$ (resp. $c_1|y_N| \leqslant |x_N| \leqslant c_2|y_N|$). For matrices, we note $A_N \asymp B_N$ if for every i, j, one has $(A_N)_{i,j} \asymp (B_N)_{i,j}$.

Theorem 1 (SFBM). *Let $0 < H < 1$ be the fractional index of an SFBM observed at times i/Δ_N, $i = 0, \ldots, N-1$. Then the CRB verifies that, as $N \to +\infty$, $\Delta_N \to +\infty$:*

$$\mathrm{CRB}_N(H) \sim \frac{1}{2} \frac{1}{N \log^2(\Delta_N)}. \tag{3}$$

Theorem 2 (FBM). *Let $0 < H < 1$ and $0 < C$ be the fractional index and scaling parameter of an FBM observed at times i/Δ_N, $i = 0, \ldots, N-1$. Then the CRB verifies that, as $N \to +\infty, \Delta_N \to +\infty$:*

$$\mathrm{CRB}_N(H, C^2) \asymp \begin{pmatrix} 1/N & \log(\Delta_N)/N \\ \log(\Delta_N)/N & \log^2(\Delta_N)/N \end{pmatrix}. \tag{4}$$

Remark. The optimal rates of convergence for unbiased estimators of fractional parameter H are therefore $\sqrt{N}\log(\Delta_N)$ if the scaling parameter C is known and $\sqrt{N}$ if the scaling parameter C is unknown. These rates correspond to the rates of convergence of the estimators given in the introduction (see Abry et al., 1995; Beran, 1994; Coeurjolly, 1999; Istas and Lang, 1994,1997; Benassi et al., 1988; Peltier and Lévy-Véhel, 1994).

3. Inverse of finite matrices with terms decaying polynomially far away from the diagonal

Let us start with the definition of a polynomially localized matrix:

Definition 1. For given $\alpha > 0$, $c > 0$, let $(Q_{\alpha,N})_{N \geqslant 1}$ be the sequence of subsets of $N \times N$ matrices such that

$$A \in Q_{\alpha,N} \quad \Leftrightarrow \quad |(A)_{k,k'}| \leqslant c(1 + |k' - k|)^{-\alpha}, \quad \forall k, k' = 1, \ldots, N.$$

Clearly, from the definition, c and α are independent of N. The set $Q_{\alpha,N}$ is a subset of matrices with polynomial decay far away from the diagonal.

Lemma 1. *$Q_{\alpha,N}$ is an algebra for $\alpha > 1$.*

The proof of this fundamental result is omitted since it is similar to the one of Proposition 1 of Jaffard (1990).

Theorem 3. *Let $\alpha \in]1, 3/2[$, and let $(A_N)_{N \geqslant 1}$ be a sequence of $N \times N$ symmetric definite positive matrices such that $A_N \in Q_{\alpha,N}$. Let A_∞ be the operator defined, if the following limit exists, for $(i, j) \in \mathbb{Z}^2$ by*

$$(A_\infty)_{i,j} = \lim_{N \to +\infty} (A_N)_{i,j}.$$

Assume the following assumptions to be fullfilled:

(A_1) $(A_\infty)_{i,i+j} = a(j)$ *with* $a(j) \sim c|j|^{-\alpha}$ *as* $|j| \to +\infty$

(A_2) $\hat{a}(\lambda) > 0$ *for* $\lambda \neq 0$

(A_3) $\displaystyle\sum_{i,j=1}^{N} \{(A_\infty)_{i,j} - (A_N)_{i,j}\}^2 = o(1)$ *as* $N \to +\infty$

then $A_N^{-1} \in Q_{\alpha,N}$.

We have denoted by $\hat{a}$ the Fourier transform of the sequence $a \in \ell^2(\mathbb{Z})$.

Jaffard (1990) studies infinite matrices whose terms decay polynomially far away from the diagonal: let Q_α, for a real $\alpha > 0$, denote the set of matrices, M, such that $M_{i,j} = \mathcal{O}(|j - i|^{-\alpha})$, $\forall (i, j) \in \mathbb{Z}^2$. Jaffard (1990) proves that if $M \in Q_\alpha$, with $\alpha > 1$, is invertible as operator $\ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ then $M^{-1} \in Q_\alpha$. Our result is finer since M satisfies the condition $M_{i,j} \sim c|j - i|^{-\alpha} (c > 0)$, we manage to replace the ℓ^2-inversibility assumption by the condition $\alpha \in]1, \frac{3}{2}[$ and Assumption (A_2), which allows us to apply the result to the case of sequences of symmetric definite positive matrices.

Proof. From now on, $(A_N)_{N\geqslant 1}$ is a sequence of $N\times N$ matrices satisfying the assumptions of Theorem 3 and c a generic constant strictly positive. And for the sake of simplicity, we use the same notations as in Jaffard (1990).

$$\|A_N\| = \sup_{X\in\mathbb{R}^N, X\neq 0} \frac{\|A_N X\|_{\ell^2}}{\|X\|_{\ell^2}} = \sup_{X\in\mathbb{R}^N, X\neq 0} \frac{X^t A_N X}{X^t X}.$$

$$\|A_N\|_\alpha = \sup_{k,j} |(A_N)_{j,k}|(1+|k-j|)^\beta, \quad \beta > 0.$$

$$\|A_N\|_{\ell^p} = \sup_j \left[\sum_k |(A_N)_{j,k}|^p\right]^{1/p}.$$

Moreover, for a sequence $u=(u_j)_{j\in\mathbb{Z}}$, we will denote $\|u\|_{\ell^p} = \{\sum_{j\in\mathbb{Z}} |u_j|^p\}^{1/p}$, for $p\geqslant 1$. Define $B_N = I - (A_N/\|A_N\|)$. The first lemma gives an upper bound for $\|B_N\|$.

Lemma 2. *There exists* $0<r<1$, *r independent of N, such that* $\|B_N\|\leqslant r<1$.

Proof of Lemma 2.

- Let us start by proving that A_∞ is a continuous and invertible operator $\ell^2(\mathbb{Z})\to\ell^2(\mathbb{Z})$. Let $u\in\ell^2(\mathbb{Z})$, since $A_\infty u = a * u$ (where $*$ denotes the convolution product), from Young's inequality one has

  $$\|A_\infty u\|_{\ell^2} = \|a*u\|_{\ell^2} \leqslant \|a\|_{\ell^1}\|u\|_{\ell^2}.$$

  Since from Assumption (A_1) $a\in\ell^1$, A_∞ is continuous. Then, $\widehat{a*u} = \hat{a}\hat{u}$ and since $\hat{a}>0$ almost everywhere, $a*u=0$ implies $u=0$ so that A_∞ is injective. Now, note $v=a*u$ and suppose $v\in\ell^2$, from Parseval's equality and Hölder's inequality, it is easy to see that

  $$\|u\|_{\ell^2} \leqslant \|v\|_{\ell^2}\ \|\hat{a}^{-1}\|_{L^2(]-\pi,\pi[)}.$$

  We have $a(j)\sim c|j|^{-\alpha}$, as $|j|\to+\infty$, with $\alpha\in]1,3/2[$. From the result of Zygmund (1953), $\hat{a}^{-1}(\lambda)\sim|\lambda|^{1-\alpha}$ as $|\lambda|\to 0$, and thus from Assumption (A_2) $\hat{a}^{-1}\in L^2(]-\pi,\pi[)$. Finally, A_∞ is a continuous and invertible operator $\ell^2(\mathbb{Z})\to\ell^2(\mathbb{Z})$. As a consequence, A_∞^{-1} exists and is continuous.
- Little linear algebra leads to

  $$\|A_N\| = \lambda_M, \quad \text{and} \quad \|B_N\| = 1-\frac{\lambda_m}{\lambda_M}, \tag{5}$$

  where λ_M (resp. λ_m) is the largest (resp. smallest) eigenvalue of A_N. To prove Lemma 2, it is sufficient to bound the ratio λ_M/λ_m independently of N. Let X_M and X_m be the eigenvectors associated to λ_M and λ_m. Define the vector X_M^0 and the matrix A_N^0 by:

  $$\left(X_M^0\right)_i = \begin{cases} (X_M)_i & \text{if } 1\leqslant i\leqslant N. \\ 0 & \text{otherwise} \end{cases} \qquad \left(A_N^0\right)_{i,j} = \begin{cases} (A_N)_{i,j} & \text{if } 1\leqslant i,j\leqslant N. \\ 0 & \text{otherwise} \end{cases}$$

  One defines also the vector X_m^0 associated to X_m.
- Let us start to bound λ_M:

  $$\lambda_M = \frac{X_M^t A_N X_M}{X_M^t X_M} = \frac{(X_M^0)^t A_N^0 X_M^0}{(X_M^0)^t X_M^0} = \frac{(X_M^0)^t (A_N^0 - A_\infty) X_M^0}{(X_M^0)^t X_M^0} + \frac{(X_M^0)^t A_\infty X_M^0}{(X_M^0)^t X_M^0}.$$

From Cauchy-Schwarz inequality, one gets:

$$\frac{(X_M^0)^t(A_N^0 - A_\infty)X_M^0}{(X_M^0)^t X_M^0} \leqslant \left\{ \sum_{1\leqslant i,j\leqslant N} \{(A_N - A_\infty)_{i,j}\}^2 \right\}^{1/2} = o(1), \text{ from Assumption } (A_3).$$

Moreover,

$$\frac{(X_M^0)^t A_\infty X_M^0}{(X_M^0)^t X_M^0} \leqslant \sup_{X\in\ell^2, X\neq 0} \frac{X^t A_\infty X}{X^t X} = \|A_\infty\|.$$

So, there exists $N_0 \in \mathbb{N}^*$ such that $\forall N \geqslant N_0$ one has

$$\lambda_M \leqslant 2\|A_\infty\|. \tag{6}$$

- Now, let us bound $1/\lambda_m$. With the introduced notations, one has:

$$\lambda_m = \frac{X_m^t A_N X_m}{X_m^t X_m} = \frac{(X_m^0)^t A_N^0 X_m^0}{(X_m^0)^t X_m^0} = \frac{(X_m^0)^t (A_N^0 - A_\infty)X_m^0}{(X_m^0)^t X_m^0} + \frac{(X_m^0)^t A_\infty X_m^0}{(X_m^0)^t X_m^0}.$$

Again,

$$\frac{(X_m^0)^t(A_N^0 - A_\infty)X_m^0}{(X_m^0)^t X_m^0} = o(1), \text{ from Assumption } (A_3).$$

The operator A_∞ being symmetric definite positive, let us define $Y_m^0 = A_\infty^{1/2} X_M^0$. One gets

$$\frac{(X_m^0)^t A_\infty X_m^0}{(X_m^0)^t X_m^0} = \frac{(Y_m^0)^t Y_m^0}{(Y_m^0)^t A_\infty^{-1} Y_m^0} \geqslant \frac{1}{\|A_\infty^{-1}\|}.$$

So, there exists $N_1 \in \mathbb{N}^*$ such that $\forall N \geqslant N_1$ one has:

$$\frac{1}{\lambda_m} \leqslant 2\|A_\infty^{-1}\|. \tag{7}$$

The results (6) and (7) prove that $\forall N \geqslant \max\{N_0, N_1\}$ one has $\lambda_M/\lambda_m \leqslant 4\|A_\infty\|\|A_\infty^{-1}\|$. The first part of the proof implies that this ratio is finite which end the proof of Lemma 2.

Lemma 2 allows us to use the framework established by Jaffard: one has to prove that $A_N^{-1} \in Q_{\alpha,N}$ that is to say $\|A_N^{-1}\|_\alpha \leqslant c$. If one writes $A_N^{-1} = \sum_{n\geqslant 0}(B_N)^n$, it is therefore sufficient to bound the series $\sum_{n\geqslant 0}\|(B_N)^n\|_\alpha$ independently of N. The following lemmas are due to Jaffard (1990). We check that they are available under the assumptions of Theorem 3.

Lemma 3. *$\forall p \in]1,2]$, $\|B_N\|_{\ell^p} \leqslant c_p \|B_N\|_1^{2/p-1}\|B_N\|_{\ell^2}^{2-2/p}$, with c_p independent of N.*

Proof. Let $b_{jk} = (B_N)_{j,k}$. Let $T \in \{1,\dots,N\}$ (T is chosen later). Then,

$$\left(\sum_k |b_{jk}|^p\right)^{1/p} \leqslant \left(\sum_{|k-j|\leqslant T} |b_{jk}|^p\right)^{1/p} + \left(\sum_{|k-j|>T} \frac{\|B_N\|_1^p}{|k-j|^p}\right)^{1/p}. \tag{6}$$

From Hölder's inequality

$$\left(\sum_{i=-T}^{T}|a_i|^p\right)^{1/p} \leqslant (3T)^{1/p-1/2}\left(\sum_{i=-T}^{T}|a_i|^2\right)^{1/2}.$$

Moreover,

$$\sum_{k=T+1}^{N}\frac{1}{k^p} \leqslant \int_T^{+\infty} x^{-p}\,\mathrm{d}x = \frac{1}{p-1}T^{1-p}.$$

From upper bound (6),

$$\|B_N\|_{\ell^p} \leqslant (3T)^{1/p-1/2}\|B_N\|_{\ell^2} + 2(p-1)^{-1/p}T^{1/p-1}\|B_N\|_1.$$

Since $a \in \ell^1$, $\|B_N\|_1$ is bounded independently of N. Moreover, $(\|B_N\|_{\ell^2})_{N\geqslant 1}$ is an increasing sequence; therefore for N large enough, we can choose $T = [\|B_N\|_1^2/\|B_N\|_{\ell^2}^2]$. Then,

$$\|B_N\|_{\ell^p} \leqslant c_p\|B_N\|_1^{2/p-1}\|B_N\|_{\ell^2}^{2-2/p} \quad \text{with } c_p = 2\times\max(3^{1/p-1/2}, 2(p-1)^{-1/p}).$$

Lemma 4. *Let $\tilde{B}_N$ be the matrix $(\tilde{B}_N)_{j,k} = \tilde{b}_{jk} = (k-j)b_{jk}$. If p satisfies $1 < p < 2/(3-\alpha)$, and if M_1 and M_2 are two $N\times N$ matrices such that $\|M_1\|_{\ell^p} < +\infty$ and $\|M_2\|_{\ell^p} < +\infty$, then there exists c independent of N such that*

$$|(M_1\tilde{B}_N M_2)_{i,j}| \leqslant c\|B_N\|_\alpha\|M_1\|_{\ell^p}\|M_2\|_{\ell^p}.$$

Proof. Let us first prove that $\tilde{B}_N$ is bounded from $\ell^p \to \ell^r$, with r such that $1/p + 1/r = 1$. We have $|\tilde{b}_{jk}| \leqslant c|k-j|^{1-\alpha}$. Now, the sequence $(|k|^{1-\alpha})_{k\in\mathbb{Z}^*} \in \ell^q$ if $q > 1/(\alpha-1)$, and so from Young's inequality $\tilde{B}_N$ will be continuous with $\ell^p \to \ell^r$ if $1/r < 1/p + 1/q - 1$. The condition on p allows us to choose r such that $1/p + 1/r = 1$ (which will justify the use of Hölder's inequality hereafter).

Let (e_i) be the canonical basis of ℓ^2, $M_2e_j \in \ell^p$ so that $\|M_2e_j\|_{\ell^p} \leqslant \|M_2\|_{\ell^p}$. Therefore, $\tilde{B}_N M_2 e_j \in \ell^r$ and $\|\tilde{B}_N M_2 e_j\|_{\ell^r} \leqslant c\|B_N\|_\alpha\|M_2\|_{\ell^p}$. Since $M_1^t e_j \in \ell^r$, we deduce from Hölder's inequality that

$$|\langle M_1^t e_j, \tilde{B}_N M_2 e_j\rangle| \leqslant c\|B_N\|_\alpha\|M_1\|_{\ell^p}\|M_2\|_{\ell^p}.$$

Lemma 5. *There exists $\kappa = \kappa(R)$ independent of N such that for $n \geqslant 0$,*

$$\|(B_N)^n\|_1 \leqslant \kappa R^n, \quad \forall R > \|B_N\|.$$

Proof. Let $b_{n,jk}$ be the coefficient of order (j,k) of $(B_N)^n$, and $\tilde{b}_{n,jk} = (k-j)b_{n,jk}$. One has

$$b_{n,jk} = \sum_{i_1,\ldots,i_{n-1}} b_{ji_1}b_{i_1i_2}\ldots b_{i_{n-1}k}.$$

Since

$$|k-j| \leqslant |j-i_1| + \cdots + |i_{n-1}-k|,$$

we have

$$|\tilde{b}_{n,jk}| \leqslant \sum_{i_1}|\tilde{b}_{ji_1}||b_{i_1i_2}|\ldots|b_{i_{n-1}k}| + \cdots + \sum_{i_{n-1}}|b_{ji_1}||b_{i_1i_2}|\ldots|\tilde{b}_{i_{n-1}k}|.$$

Lemma 4 gives the following upper bound:

$$\|(B_N)^n\|_1 \leqslant 2c\|B_N\|_\alpha\|(B_N)^{n-1}\|_{\ell^p} + c\sum_{i=1}^{n-2}\{\|B_N\|_\alpha\|(B_N)^i\|_{\ell^p}\|(B_N)^{n-i-1}\|_{\ell^p}\}.$$

From Lemma 3, there exists c' independent of N such that

$$\|(B_N)^n\|_1 \leqslant c'\|B_N\|_\alpha \left\{ \|(B_N)^{n-1}\|_1^{2/p-1}\|B_N\|^{(n-1)(2-2/p)} + \sum_{i=1}^{n-2}(\|(B_N)^i\|_1\|(B_N)^{n-1-i}\|_1)^{2/p-1}\|B_N\|^{(n-1)(2-2/p)} \right\}. \tag{7}$$

Since $Q_{\alpha,N}$ is an algebra (see Jaffard, 1990 for a detailed proof), there exists $R_0 > 0$ such that, for every n, one has $\|(B_N)^n\|_1 \leqslant R_0^n$.

From (7), we obtain

$$\|(B_N)^n\|_1 \leqslant 2c'n\|B_N\|_\alpha R_0^{(n-1)(2/p-1)}\|B_N\|^{(n-1)(2-2/p)}.$$

The end of the proof is similar: if $R_1 > R_0^{2/p-1}\|B_N\|^{2-2/p}$, there exists $\kappa_1 = \kappa_1(R_1)$ such that $\|(B_N)^n\|_1 \leqslant \kappa_1 R_1^n$. Iterations of this proof and noting that $\|B_N\|$ is the unique fixed point of the application $x \mapsto x^{2/p-1}\|B_N\|^{2-2/p}$ lead to the proof.

Lemma 6. *There exists $\gamma > 0$ such that for every $R > \|B_N\|$, one has*

$$\|(B_N)^n\|_{1+\gamma} \leqslant \kappa R^n \quad \textit{with } \kappa = \kappa(R) \textit{ independent of } N.$$

Proof. The proof is similar to the one of Lemma 4 of Jaffard (1990). It relies upon the fact that one can prove a result similar to Lemma 4 for the matrix $\tilde{\tilde{B}}_N$ defined by $(\tilde{\tilde{B}}_N)_{j,k} = \|k-j|^{1+\gamma}b_{jk}$ (with $0 < \gamma < 1$) and then apply the scheme of the proof of Lemma 5 for the matrix with entires $|k-j|^{1+\gamma}b_{n,jk}$.

Lemma 7. *There exists $\gamma > 0$ such that $A_N^{-1} \in Q_{\gamma+1,N}$.*

Proof. Since from Lemma 3 $\|B_N\| \leqslant r < 1$, we can choose $R = r$ in Lemma 6. Then, we have $\sum_{n\geqslant 0}\|(B_N)^n\|_{1+\gamma} \leqslant \kappa/(1-r)$, which implies that $A_N^{-1} \in Q_{1+\gamma,N}$.

Let us complete the proof of Theorem 3. Note that

$$(k-j)(A_N^{-1})_{j,k} = -\sum_\ell\sum_m (A_N^{-1})_{k,\ell}(\ell-m)(A_N)_{\ell,m}(A_N^{-1})_{m,j}.$$

Now, there exists c independent of N, such that

$$|(A_N^{-1})_{k,\ell}| \leqslant c(1+|k-\ell|)^{-\gamma-1} \quad \text{and} \quad |(\ell-m)(A_N)_{\ell,m}| \leqslant c(1+|\ell-m|)^{1-\alpha}.$$

From Lemma 1, it can be proved that if $M \in Q_{\varepsilon,N}$ and $M' \in Q_{\beta,N}$ with $\varepsilon < 1$, $\beta > 1$ then $(MM') \in Q_{\varepsilon,N}$. Thus, there exists $c > 0$ independent of N such that

$$|(k-j)(A_N^{-1})_{j,k}| \leqslant c(1+|k-j|)^{1-\alpha},$$

and Theorem 3 is proved.

4. Proofs of Theorems 1 and 2

4.1. Preliminary computations

Let us start with some notations. $B_H = \{B_H(0), \ldots, B_H(N - 1/\Delta_N)\}$ will denote a sample of an SFBM. Let c be the vector defined for $d \in]-\frac{1}{2}, \frac{1}{2}[$ and for $k \geqslant 0$ by

$$c_k = (-1)^k \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(k-d+1)},$$

where Γ denotes the Gamma function. The sequence c represents a fractional filter of order d, i.e. the coefficients of the expansion in series of the function $(1-z)^d$. We refer to Beran (1994) for the proof that the Fourier transform of c, denoted as $\hat{c}$, is given by $\hat{c}(\lambda) = (1 - e^{i\lambda})^d$ for $\lambda \in]-\pi, \pi[$. Now, define $\Delta_d B_H$ the vector obtained by the partial fractional filtering of the increments vector of B_H, with the vector c:

$$\Delta_d B_H(i) = \sum_{k=0}^{N-1-i} c_k X_H\left(\frac{k+i}{\Delta_N}\right) \quad \text{for } i = 0, \ldots, N' - 1,$$

where $N' = [N(1 - 1/\log N)]$ ($[x]$ denoting the integer part of a real x), and where $X_H(0) = B_H(0)$ and $X_H(j/\Delta_N) = B_H(j/\Delta_N) - B_H(j-1/\Delta_N)$, for $j = 1, \ldots, N' - 1$. Let $\Sigma_{H,\Delta}$ denote the covariance matrix of $\Delta_d B_H$ defined by $\Sigma_{H,\Delta} = \mathbf{E}((\Delta_d B_H)^t(\Delta_d B_H))$. The self-similarity of the FBM suffices to prove that

$$(\Sigma_{H,\Delta})_{i,j} = \frac{1}{(\Delta_N)^{2H}} \sum_{(k,k')=(0,0)}^{(N-i,N-j)} c_k c_{k'} \gamma(k' - k + j - i), \quad i, j = 1, \ldots, N', \tag{8}$$

where γ denotes the covariance function of the increments process of the FBM given by $\gamma(\tau) = 1/2(|\tau - 1|^{2H} - 2|\tau|^{2H} + |\tau + 1|^{2H})$. Σ_H denotes the $(N' \times N')$ normalized matrix $\Sigma_{H,1}$, $|\Sigma_H|$ the determinant of the matrix Σ_H and $(\mathrm{d}/\mathrm{d}H)\Sigma_H$ the matrix with entries defined as the derivative w.r.t. H of the coefficients of Σ_H.

Lemma 8 shows the relevance of fractionally filtering the increments vector of B_H. This lemma proves that for a certain choice of the real d, the matrix Σ_H of size $(N' \times N')$ satisfies Assumptions (A_1) and (A_2) of Theorem 3. By defining the vector $\Delta_d B_H$ for $0 \leqslant i < N'$, one builds a vector for which each component is obtained through filtering at least $N/\log(N) \to +\infty$, observations of the increments vector of B_H, which is primordial to ensure Assumption (A_3) of Theorem 3 that would not have been satisfied if $N' = N$. To sum up, Theorem 3 allows us to obtain fine estimations for all $H \in]0, 1[$ for the coefficients of Σ_H^{-1}. If we had considered the increments process of the FBM, the assumptions of Theorem 3 would have been fulfilled only for $H \in]\frac{1}{4}, \frac{1}{2}[$.

Let us add that due to their equivalence, we will abuse notations by denoting with $\mathrm{CRB}_N(H)$ the Cramèr–Rao bound of H based either on a sample of SFBM of length N or on the sample $\Delta_d B_H$ since $N' \sim N$. Without loss of generality, the proof becomes more readable.

Finally, we introduce the following notations:

$$X_{N'} = X_{N'}(H, \Delta_N) \equiv ((\Delta_N)^H \Delta_d B_H)^t (\Sigma_H)^{-1} ((\Delta_N)^H \Delta_d B_H), \tag{9}$$

$$Y_{N'} = Y_{N'}(H, \Delta_N) \equiv ((\Delta_N)^H \Delta_d B_H)^t \frac{\mathrm{d}}{\mathrm{d}H}(\Sigma_H)^{-1} ((\Delta_N)^H \Delta_d B_H) \tag{10}$$

and

$$u_{N'} = \mathrm{Var}(X_{N'}), \qquad v_{N'} = \mathrm{Var}(Y_{N'}) \quad \text{and} \quad c_{N'} = \mathrm{Cov}(X_{N'}, Y_{N'}). \tag{11}$$

We now state an auxiliary result, concerning the behaviors of $v_{N'}$ and $c_{N'}$, proved in Section 4.3.

Proposition 1. *As $N \to +\infty$, $\Delta_N \to +\infty$, one has*

$$v_{N'} = \mathcal{O}(N), \qquad c_{N'} = \mathcal{O}(N), \qquad w_{N'} = u_{N'} v_{N'} - c_{N'}^2 > 0 \quad \textit{and} \quad w_{N'} = \mathcal{O}(N^2).$$

4.2. Proofs of Theorems 1 and 2

Proof of Theorem 1 (SFBM).

The log-likelihood of the discrete second differences of the SFBM is given by

$$\ell_{N'}(\Delta_d B_H, H) = -\frac{N'}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_{H,\Delta}| - \frac{1}{2}(\Delta_d B_H)^t(\Sigma_{H,\Delta})^{-1}(\Delta_d B_H). \tag{12}$$

The score, $S(\Delta_d B_H, H)$, defined as the derivative of the likelihood with respect to parameter H, is given by

$$S(\Delta_d B_H, H) = -\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}H}\log|\Sigma_{H,\Delta}| - \frac{1}{2}(\Delta_d B_H)^t\frac{\mathrm{d}}{\mathrm{d}H}(\Sigma_{H,\Delta})^{-1}(\Delta_d B_H). \tag{13}$$

Using the definition of the density of vector $\Delta_d B_H$,

$$(2\pi)^{-N'/2}|\Sigma_{H,\Delta}|^{-1/2}\int_{\mathbb{R}^N}\exp(-\frac{1}{2}x^t(\Sigma_{H,\Delta})^{-1}x)\,\mathrm{d}x = 1. \tag{14}$$

The function $\exp(-\frac{1}{2}x^t\Sigma_H^{-1}x)$ is differentiable with respect to H, and the derivative is $f(x) = -x^t(\mathrm{d}/\mathrm{d}H)\Sigma_H^{-1}x\exp(-\frac{1}{2}x^t\Sigma_H^{-1}x)$. Let $c = \max_{j,k}(\mathrm{d}/\mathrm{d}H)(\Sigma_H^{-1})_{j,k}$, we have clearly $|f(x)| \leqslant c(x^tx)\exp(-\frac{1}{2}x^t\Sigma_H^{-1}x)x \in L^1(\mathbb{R}^{N'})$. Thus differentiating (14) with respect to H, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}H}\log|\Sigma_{H,\Delta}| = -\mathbf{E}\left(\frac{\mathrm{d}}{\mathrm{d}H}(\Delta_d B_H)^t(\Sigma_{H,\Delta})^{-1}\Delta_d B_H\right),$$

so that (13) can be written as

$$S(\Delta_d B_H, H) = \frac{1}{2}\left\{\mathbf{E}((\Delta_d B_H)^t\frac{\mathrm{d}}{\mathrm{d}H}(\Sigma_{H,\Delta})^{-1}(\Delta_d B_H)) - (\Delta_d B_H)^t\frac{\mathrm{d}}{\mathrm{d}H}(\Sigma_{H,\Delta})^{-1}(\Delta_d B_H)\right\}. \tag{15}$$

Now,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}H}(\Sigma_{H,\Delta})^{-1} &= \frac{\mathrm{d}}{\mathrm{d}H}((\Delta_N)^{2H}(\Sigma_H)^{-1})\\ &= 2(\Delta_N)^{2H}\log(\Delta_N)(\Sigma_H)^{-1} + (\Delta_N)^{2H}\frac{\mathrm{d}}{\mathrm{d}H}(\Sigma_H)^{-1}.\end{aligned}$$

From (9) and (10), (15) becomes

$$S(\Delta_d B_H, H) = \log(\Delta_N)\{\mathbf{E}(X_{N'}) - X_{N'}\} + \tfrac{1}{2}\{\mathbf{E}(Y_{N'}) - Y_{N'}\}. \tag{16}$$

Using notations (11), the Fisher information matrix is therefore given by (see (2))

$$\begin{aligned}I(\Delta_d B_H, H) &= \mathbf{E}(S(\Delta_d B_H, H)^2) = \mathrm{Var}(S(\Delta_d B, H))\\ &= u_{N'}\log^2(\Delta_N) + \tfrac{1}{4}v_{N'} + c_{N'}\log(\Delta_N).\end{aligned} \tag{17}$$

One can easily check that the distribution of $X_{N'}$ is a chi-square distribution with N' degrees of freedom, so that $u_{N'} = 2N'$. This and Proposition 1 imply that

$$I(\Delta_d B_H, H) = 2N'\log^2(\Delta_N) + \mathcal{O}(N) + \mathcal{O}(N\log(\Delta_N)),$$

and Theorem 1 is proved ($N' \sim N$).

Proof of Theorem 2 (FBM).

The score depends now on parameters H and C:

$$S(\Delta_d B_{H,C}, H, C^2) \overset{\mathcal{L}}{=} \begin{pmatrix}\log(\Delta_N)\{\mathbf{E}(X_{N'}) - X_{N'}\} + \frac{1}{2}\{\mathbf{E}(Y_{N'}) - Y_{N'}\}\\ \frac{1}{2C^2}\{X_{N'} - \mathbf{E}(X_{N'})\}\end{pmatrix}.$$

The Fisher information matrix is given by

$$I(\Delta_d B_{H,C}, H, C^2) = \begin{pmatrix} u_{N'}\log^2(\Delta_N) + \frac{1}{4}v_{N'} + c_{N'}\log(\Delta_N) & -\dfrac{1}{2C^2}u_{N'}\log(\Delta_N) - \frac{1}{4C^2}c_{N'} \\ -\dfrac{1}{2C^2}u_{N'}\log(\Delta_N) - \frac{1}{4C^2}c_{N'} & \dfrac{1}{4C^4}u_{N'} \end{pmatrix}.$$

From Proposition 1, $|I(\Delta_d B_{H,C}, H, C^2)| \neq 0$ and $I(\Delta_d B_{H,C}, H, C^2)$ is invertible. One has

$$I(\Delta_d B_{H,C}, H, C^2)^{-1} = (u_{N'}v_{N'} - c_{N'}^2)^{-1} \times \begin{pmatrix} 4u_{N'} & 4C^2(2u_{N'}\log(\Delta_N) + c_{N'}) \\ 4C^2(2u_{N'}\log(\Delta_N) + c_{N'}) & 4C^4(4u_{N'}\log^2(\Delta_N) + v_{N'} + 4c_{N'}\log(\Delta_N)) \end{pmatrix}.$$

Using Proposition 1, we obtain a lower bound for the CRB. But the results in Istas and Lang (1997) provide an estimator of parameters (H, C) whose mean square error behaves like (4). The mean square error of this estimator gives an upperbound of the Cramèr–Rao bound. Therefore, this remark ends the proof of Theorem 2.

4.3. Proof of Proposition 1

Proposition 1 relies upon the following lemma exhibiting the coefficients behavior of $(\mathrm{d}/\mathrm{d}H)\Sigma_H$ and Σ_H^{-1}.

Lemma 8.
(i) *There exists* $\alpha \in]1, \frac{3}{2}[$ *such that* Σ_H *fulfills the assumptions of Theorem* 3.
(ii) *There exists* $\beta > 1$ *such that* $(\mathrm{d}/\mathrm{d}H)(\Sigma_H) \in Q_{\beta,N'}$.

Proof. $\kappa(H)$ will denote a generic constant depending on H.
(i) For $0 < \varepsilon < \frac{1}{8}$, define

$$d = \begin{cases} H - 1/4 - \varepsilon & \text{if } H \in]0, 3/4 - \varepsilon], \\ 1/2 - \varepsilon & \text{if } H \in]3/4 - \varepsilon, 1 - \varepsilon[. \end{cases} \tag{18}$$

- Define also the process $(Y(t))_{t\in\mathbb{Z}}$ obtained by filtering the increments process of the FBM with the filter c of fractional order d:

$$Y(i) = \sum_{k=0}^{+\infty} c_k X\left(\frac{k-i}{\Delta_N}\right), \quad i \in \mathbb{Z}.$$

Since $d \in]-\frac{1}{2}, \frac{1}{2}[$, the spectral density of Y, denoted by $\hat{\sigma}$ is explicitly given by (see Beran, 1994, p. 61)

$$\hat{\sigma}(\lambda) = 2^d(1 - \cos(\lambda))^d \hat{f}(\lambda) \quad \text{for } \lambda \in]-\pi, \pi[,$$

where $\hat{f}$ denotes the spectral density of the increments process of the FBM. Let us recall that $\hat{f}(\lambda) \sim \kappa(H)|\lambda|^{1-2H}$ as $\lambda \to 0$, and so $\hat{\sigma}(\lambda) \sim \kappa(H)|\lambda|^{1-2H+2d}$ as $\lambda \to 0$. The real d has been chosen such that $2d + 2 - 2H \in]1, \frac{3}{2}[$ for $H \in]0, 1 - \varepsilon[$, thus from Zygmund (1953) $\sigma(j) \sim \kappa(H)|j|^{2H-2d-2}$ as $|j| \to +\infty$. We have thus proved the existence of $\alpha \in]1, \frac{3}{2}[$ such that for all $H \in]0, 1[, \sigma(j) \sim \kappa(H)|j|^{-\alpha}$ as $|j| \to +\infty$.

By definition,

$$\sigma(j) = \sum_{(k,k')=(0,0)}^{(+\infty,+\infty)} c_k c_{k'} \gamma(k'-k+j),$$

where γ denotes the autocovariance function of the FBM. This means that the limit $\lim_{N'\to+\infty}(\Sigma_H)_{i,i+j}$ exists and is equal to $\sigma(j)$, which fulfills Assumption (A_1) of Theorem 3.

- Since the spectral density of the increments process of the FBM, $\hat{f}(\lambda)$ is strictly positive for $\lambda \neq 0$, $\hat{\sigma} > 0$ for $\lambda \neq 0$, and Assumption (A_2) is satisfied.
- Let us denote as Σ_∞ the infinite matrix defined for $(i,j) \in \mathbb{Z}^2$ by $(\Sigma_\infty)_{i,j} = \sigma(j-i)$. We have to prove that $S_N = \sum_{1\leqslant i,j\leqslant N'}\{(\Sigma_\infty)_{i,j} - (\Sigma_H)_{i,j}\}^2 = o(1)$. It is well known that $|\gamma(\tau)| \sim \kappa|\tau|^{2H-2}$, as $|\tau| \to +\infty$. From Zygmund (1953), one has $|c_\tau| \sim \kappa|\tau|^{-d-1}$, as $|\tau| \to +\infty$. Thus, one can obtain the following upperbound for S_N:

$$S_N \leqslant \kappa N' \sum_{|i|\leqslant N'} \left\{ \sum_{k,k'=N''}^{+\infty} |k|^{-d-1}|k'|^{-d-1}|k'-k+i|^{2H-2} \right\}^2,$$

$$\text{with } N'' = N - N' \ (N'' \to +\infty \text{ as } N \to +\infty),$$

$$\leqslant N' \sum_{k_1,\ldots,k_4\geqslant N''} |k_1|^{-d-1}\ldots|k_4|^{-d-1} \left\{ \sum_{|i|\leqslant N'} |k_1-k_3+i|^{2H-2}|k_2-k_4+i|^{2H-2} \right\}.$$

Based on the proof of Lemma 1, one can write

$$\sum_{|i|\leqslant N'} |k_1-k_3+i|^{2H-2}|k_2-k_4+i|^{2H-2} \leqslant \kappa x_{N'}|k_1+k_2-k_3-k_4|^{2H-2},$$

with

$$x_{N'} = \begin{cases} 1 & \text{if } H < \frac{1}{2}, \\ \log(N') & \text{if } H = \frac{1}{2}, \\ (N')^{2H-1} & \text{if } H > \frac{1}{2}. \end{cases}$$

Thus, we have the following upperbound $S_N \leqslant \kappa N' x_{N'} \sum_{k_1,\ldots,k_4\geqslant N''} F(k_1,\ldots,k_4)$ where $F(k_1,\ldots,k_4) = |k_1|^{-d-1}\ldots|k_4|^{-d-1}|k_1+k_2-k_3-k_4|^{2H-2}$. Define for $\eta > 0$

$$E_1 = \{k_1,\ldots,k_4 \geqslant N'', 0 < (N'')^{1+\eta}x_{N'}|k_1+k_2-k_3-k_4|^{2H+2+\eta} < |k_1k_2k_3k_4|^{d+1}\},$$

$$E_2 = \{k_1,\ldots,k_4 \geqslant N'', (N'')^{1+\eta}x_{N'}|k_1+k_2-k_3-k_4|^{2H+2+\eta} \geqslant |k_1k_2k_3k_4|^{d+1}\}.$$

Then

$$\sum_{(k_1,\ldots,k_4)\in E_1} F(k_1,\ldots,k_4) \leqslant (N'')^{-1-\eta}x_{N'}^{-1} \sum_{k_1,\ldots,k_4\geqslant N''} |k_1+k_2-k_3-k_4|^{-4-\eta}$$

$$\leqslant (N'')^{-1-\eta}x_{N'}^{-1} \tag{19}$$

and

$$\sum_{(k_1,\ldots,k_4)\in E_2} F(k_1,\ldots,k_4) \leqslant (N'')^{[(1+\eta)(2-2H)]/[2+2H+\eta]}x_{N'}^{[2-2H]/[2H+2+\eta]} \left\{ \sum_{k\geqslant N''} |k|^{-(d+1)(1+2-2H/2+2H+\eta)} \right\}^4.$$

$$\leqslant (N'')^{[(1+\eta)(2-2H)]/[2+2H+\eta]}x_{N'}^{[2-2H]/[2H+2+\eta]}(N'')^{4-[4(d+1)(4+\eta)]/[2H+2+\eta]}. \tag{20}$$

Inequalities (19), (20) and the definitions of N' and N'' imply finally that

$$S_N = L(N)\{\mathcal{O}(N^{-\eta}) + \mathcal{O}(x_N^{(4+\eta)/(2+2H+\eta)} N^{\{5-[4(4+\eta)(d+1)-(1+\eta)(2-2H)]/(2+2H+\eta)\}})\}, \tag{21}$$

where L is a slowly varying function. We leave the reader to check that the right-hand-side term of (21) tends to 0, as $N \to +\infty$ if

$$\frac{8\varepsilon + 2\eta + 2\varepsilon\eta}{4+3\eta} < H < 1 - \varepsilon - \frac{\varepsilon\eta}{4}, \tag{22}$$

where ε is the real parameter introduced to define d, see (18). Eq. (22) being available for all $\eta > 0$, $\varepsilon > 0$, one obtains that for all $H \in]0,1[$, $S_N \to 0$ as $N \to +\infty$, and Assumption (A_3) is fulfilled.

(ii) It can be easily checked that $(\mathrm{d}/\mathrm{d}H)\hat{\sigma}(\lambda) \sim \kappa(H)|\lambda|^{1-2H+2d} L(1/\lambda)$ as $|\lambda| \to 0$ where L denotes a slowly varying function. From Zygmund, one gets $(\mathrm{d}/\mathrm{d}H)\sigma(j) \sim \kappa(H)|j|^{2H-2d-2} L(j)$ as $|j| \to +\infty$. As a conclusion, there exists $\beta > 1$ such that $(\mathrm{d}/\mathrm{d}H)\sigma(j) = \mathcal{O}(|j|^{-\beta})$ and finally that $(\mathrm{d}/\mathrm{d}H)\Sigma_H \in Q_{\beta,N'}$.

Proof of Proposition 1. Little algebra and Isserlis formula for Gaussian variables, (Isserlis, 1918), lead to:

$$v_{N'} = 2\,\mathrm{Tr}\left(\left(\frac{\mathrm{d}}{\mathrm{d}H}\Sigma_H\right)\Sigma_H^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}H}\Sigma_H\right)\Sigma_H^{-1}\right),$$

$$c_{N'} = 2\,\mathrm{Tr}\left(\left(\frac{\mathrm{d}}{\mathrm{d}H}\Sigma_H\right)\Sigma_H^{-1}\right).$$

Since $Q_{\alpha,N'}$ is an algebra for $\alpha > 1$ (Lemma 1), Lemma 8 implies the existence of $\beta > 1$ such that $(\mathrm{d}/\mathrm{d}H)\Sigma_H . \Sigma_H^{-1} \in Q_{\beta,N'}$. Therefore, $v_{N'} = \mathcal{O}(N') = \mathcal{O}(N)$ and $c_{N'} = \mathcal{O}(N') = \mathcal{O}(N)$. Finally, let us denote for the sake of simplicity $M = (\mathrm{d}/\mathrm{d}H)\Sigma_H . \Sigma_H^{-1}$. Using again Isserlis formula (Isserlis, 1918), one has

$$w_{N'} \equiv u_{N'}v_{N'} - c_{N'}^2 = 4N'\sum_{i,j}(M)_{i,j}^2 - 4\left\{\sum_i (M)_{i,i}\right\}^2$$

$$\geqslant 4N'\sum_{i\neq j}(M)_{i,j}^2 + 4N'\left\{\sum_i (M)_{i,i}^2 - \left(\sum_i (M)_{i,i}\right)^2\right\} > 0. \tag{23}$$

Clearly, $w_{N'} \leqslant u_{N'}v_{N'} = \mathcal{O}(N^2)$, and Proposition 1 is proved.

References

Abry, P., Gonçalvès, P., Flandrin, P. 1995. Wavelets, spectrum analysis and 1/f processes. In: Antoniadis, A., Oppenheim, G. (Eds.), Wavelets and Statistics, Lectures Note in Statistics, Vol. 103, Springer, New York, pp. 15–29.

Beran, J., 1994. Statistics for Long Memory Processes. Chapman Hall, London.

Benassi, A., Cohen, S., Istas, J., Jaffard, S., 1988. Identification of filtered white noises. Stochastic Process. Appl. 75, 31–49.

Coeurjolly, J.F., 1999. Identification du mouvement brownien fractionnaire par variations discrètes. Rapport de Recherche LMC/IMAG 1016-M, 1–33.

Dacunha-Castelle, D., Duflo, M., 1994. Probabilités et statistiques, Vol. 1. Masson, Paris.

Dahlhaus, R., 1989. Efficient parameter estimation for self-similar processes. Ann. Stat. 17, 1749–1766.

Frisch, U., 1995. Turbulence: The legacy of A.N. Kolmogorov. Cambridge University Press, Cambridge.

Istas, J., Lang, G., 1994. Variations quadratiques et estimation de l'exposant de Hölder local d'un processus gaussien. C.R. Acad. Sci. Paris 319, I, 201–206.

Istas, J., Lang, G., 1997. Quadratic variations and estimation of the local Hölder index of a Gaussian process. Ann. Inst. Henri Poincaré 33, 407–436.

Isserlis, L., 1918. On a formula for the product moment coefficient of any order normal frequency distribution in any number of variables. Biometrika 12, 134–139.

Jaffard, S., 1990. Propriétés des matrices "bien" localisées près de leur diagonale et quelques applications. Ann. Inst. Henri Poincaré 7 (5), 461–476.

Kolmogorov, A.N., 1940. Winersche spiralen und einige andere interessante kurven in Hilbertschen Raum. S.R. (Doklady) Acad. Sci. USSR (N.S.) 26, 115–118.

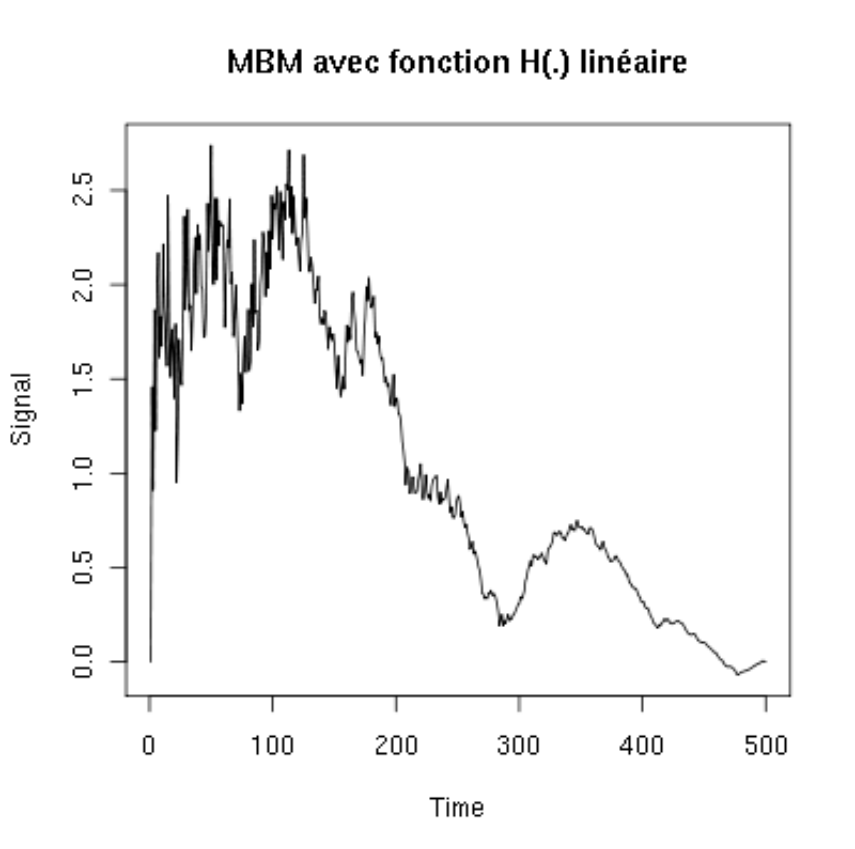
Lévy-Véhel, J., Lutton, E., Tricot, C (Eds.), 1997. Fractals in Engineering. Springer, Berlin.

Mandelbrot, B., Van Ness, J.W., 1968. Fractional Brownian motions, fractional noises and applications. SIAM Rev. 10, 422–437.

Peltier, R.F., Lévy-Véhel, J., 1994. A new method for estimating the parameter of fractional Brownian motion. Rapport de recherche de l'INRIA 2396, 1–40.

Rao, A.R., 1991. A Taxonomy for Texture Description and Identification. Springer, Berlin.

Zygmund, A., 1953. Trigonometric Series. Cambridge University Press, London, (Chapter V, 2).



[Coe05]
J.-F. Coeurjolly. Identification of multifractional Brownian motion, *Bernoulli*, 11(6):987-1008, 2005. (Erratum 12(2):381, 2006).

Bernoulli **11**(6), 2005, 987–1008

Identification of multifractional Brownian motion

JEAN-FRANÇOIS COEURJOLLY

Labsad, BSHM, UPMF, Avenue Centrale, B.P. 47, 38040 Grenoble Cedex 9, France.
E-mail: jean-francois.coeurjolly@upmf-grenoble.fr.

We develop a method for estimating the Hurst function of a multifractional Brownian motion, which is an extension of the fractional Brownian motion in the sense that the path regularity can now vary with time. This method is based on a local estimation of the second-order moment of a unique discretized filtered path. The effectiveness of our procedure is investigated in a short simulation study.

Keywords: filtering; fractional Brownian motion; functional estimation; multifractional Brownian motion

1. Introduction

Since the pioneering work of Mandelbrot and Van Ness (1968), self-similar processes and, in particular, fractional Brownian motion have been widely used to model data that exhibit long-range dependence and scaling phenomena. However, in certain situations occurring either in the field of turbulence (Frisch 1999) or in biomechanics (Collins and De Luca 1994), a more flexible model is necessary in order to control the dependence structure locally and to allow the path regularity to vary with time.

With such a perspective, a stochastic model leading to an important extension of fractional Brownian motion has recently been developed. This model, called the multifractional Brownian motion, was obtained in two different ways. Both involve replacing the Hurst parameter H by a function of time within the two main stochastic integral representations of fractional Brownian motion. The first representation is a mean average approach and was proposed by Peltier and Lévy Véhel (1995). This leads to the process denoted by $(W_1(t))_{t\geqslant 0}$. The second one is a spectral approach introduced by Benassi *et al.* (1998). This process is denoted by $(W_2(t))_{t\geqslant 0}$. These processes are defined as follows:

$$W_1(t) = C\{\pi K(2H(t))\}^{1/2}\int_{\mathbb{R}} f_t(s)\mathrm{d}B_1(s), \tag{1}$$

with

1350–7265

$$f_t(s) = \frac{1}{\Gamma(H(t)+1/2)}\left\{|t-s|^{H(t)-1/2}\mathbb{1}_{]-\infty,t]}(s) - |s|^{H(t)-1/2}\mathbb{1}_{]-\infty,0]}(s)\right\},$$

$$W_2(t) = C\{K(2H(t))/2\}^{1/2}\int_{\mathbb{R}} \frac{(\exp(\mathrm{i}t\lambda)-1)}{|\lambda|^{H(t)+1/2}}\mathrm{d}B_2(\lambda), \tag{2}$$

where C is a positive constant, B_1 and B_2 are two Brownian motions, and K is the function defined on $]0, 2[$ by $K(\alpha) = \Gamma(\alpha+1)\sin(\alpha\pi/2)/\pi$. The processes W_1 and W_2 are well defined (i.e. square-integrable) if the function $H(\cdot)$ is Hölderian of order $0 < \eta \leqslant 1$ on $[0, 1]$ (the set of such functions is denoted by $\mathcal{C}^\eta([0, 1])$). Under these conditions, Cohen (1999) proved the equality in distribution of both processes normalized in such a way that

$$\mathbb{E}(W_1(t)^2) = \mathbb{E}(W_2(t)^2) = C^2|t|^{2H(t)}.$$

From now on, a multifractional Brownian motion with Hurst function $H(\cdot)$ and scaling parameter C, defined by (1) or (2), is denoted by $(W(t))_{t\geqslant 0}$. As a Gaussian model, this process can be defined as the only centred Gaussian process, zero at origin and with covariance function defined for $H \in \mathcal{C}^\eta([0, 1])$ and $s, t \in [0, 1]$ by

$$\mathbb{E}(W(t)W(s)) = \frac{C^2}{2}\, g(H(t), H(s))\left\{|t|^{H(t)+H(s)} + |s|^{H(t)+H(s)} - |t-s|^{H(t)+H(s)}\right\}, \tag{3}$$

where g is given by

$$g(H(t), H(s)) = K(H(t)+H(s))^{-1}\{K(2H(t))K(2H(s))\}^{1/2}. \tag{4}$$

The covariance function can easily be obtained using the representation theorem for $|u|^\alpha$ (see, for example, von Bahr and Esseen 1965):

$$|u|^\alpha = K(\alpha)\int_{\mathbb{R}} \frac{1-\cos(\lambda u)}{|\lambda|^{\alpha+1}}d\lambda,\ \forall u \in \mathbb{R},\ 0 < \alpha < 2.$$

Multifractional Brownian motion leads to a more flexible model since it satisfies our fractional Brownian motion extension conditions.

The main objective of this paper is to develop and study an estimation procedure for multifractional Brownian motion. This problem was partially examined by Benassi *et al.* (1998), where an estimator of a continuously differentiable Hurst function is derived and its consistency is proved. We seek to extend and complete their work by considering Hölderian Hurst functions (of arbitrary order $\eta > 0$) and by establishing limit theorems associated with the functional estimators. These results constitute our main contribution and allow us to construct confidence intervals, confidence bands and parametric tests.

Let us formulate the estimation problem. The identification of such a model is a difficult task since the increment process of a multifractional Brownian motion is no longer stationary, no longer a self-similar process, and its path regularity explicitly varies with time. However, several nice properties of fractional Brownian motion still hold locally for multifractional Brownian motion. Indeed, assuming that $H \in \mathcal{C}^\eta([0, 1])$ and is such that $\sup_t H(t) < \min(1, \eta)$, Benassi *et al.* (1998) proved that

$$\lim_{\epsilon \to 0^+} \left(\frac{W(t+\epsilon u) - W(t)}{\epsilon^{2H(t)}} \right)_{u \in \mathbb{R}^+} \stackrel{d}{=} C\big(B_{H(t)}(u)\big)_{u \in \mathbb{R}^+}, \tag{5}$$

where $B_{H(t)}(u)$ denotes a standard fractional Brownian motion with parameter $H(t)$ defined on $\mathbb{R}^+$. To estimate the Hurst function $H(\cdot)$ of a multifractional Brownian motion, this result suggests the local adaptation of global methods used to identify a fractional Brownian motion. The method we propose is a local version of the quadratic variations method studied by Istas and Lang (1997), Kent and Wood (1997) and Coeurjolly (2001). It involves first filtering the observations of a self-similar (or locally self-similar at 0) stationary Gaussian process to weaken the dependence of the observations, and then estimating the empirical second-order moment of the filtered series. For the fractional Brownian motion case, this method exhibits nice properties: it produces estimators having rate of convergence that achieve Cramér–Rao bounds (for fractional Brownian motion parameters) (Coeurjolly and Istas 2001) and it is computationaly fast, numerically stable and behaves efficiently with respect to the maximum likelihood for small sample sizes (Coeurjolly 2000, p. 35).

The rest of this paper is organized as follows. Section 2 introduces some notation and defines the local H_2-variations statistic. We prove some convergence results and apply them to the identification problem in Section 3. We derive estimators of Hölderian Hurst functions, and prove their consistency and asymptotical normality. When C is assumed to be known the estimators derived in Coeurjolly (2000) had higher convergence rate, but such an assumption seems quite unrealistic. Our method is a local one and as such depends on a neighbourhood size whose choice is discussed in Section 4, where a procedure to estimate the optimal neighbourhood is proposed and studied. A simulation study is conducted in Section 5 to explore the qualities of the estimators. Finally, proofs of different results are presented in Section 6.

2. Local H_2-variations of a multifractional Brownian motion

In this section we introduce some notation and derive convergence results for the local second-order moment of the discrete variations of a multifractional Brownian motion. Later on, only a path $(W_t)_{t\in[0,1]}$ of a multifractional Brownian motion with Hurst function $H(\cdot)$ and scaling coefficient C is considered, and our statistical model corresponds to its discretized version $(\mathbf{W}) = (W(i/N))_{i=1,\ldots,N}$. The Hurst function $H(\cdot)$ is assumed to be a Hölderian function defined on $[0, 1]$, of order $0 < \eta \leqslant 1$, and such that $\sup_t H(t) < \min(1, \eta)$. Denote by a a filter of length $\ell + 1$ and of order $p \geqslant 1$, that is, a vector with real components such that

$$\sum_{q=0}^{\ell} a_q q^{\mathrm{i}} = 0, \qquad \text{for } i = 0, \ldots, p-1,$$

$$\sum_{q=0}^{\ell} a_q q^{p} \neq 0.$$

Let $(\mathbf{V}^a)$ be the series obtained by filtering $(\mathbf{W})$ with a, that is,

$$V^a\left(\frac{j}{N}\right) = \sum_{q=0}^{\ell} a_q W\left(\frac{j-q}{N}\right), \qquad \text{for } j = \ell+1, \ldots, N-1.$$

For example, when $a = (1, -1)$, $(\mathbf{V}^a)$ represents the increments $(\mathbf{W})$, and when $a = (1, -2, 1)$, $(\mathbf{V}^a)$ represents the second-order differences of $(\mathbf{W})$. As Lemma 1 reveals, filtering the discretized path of a multifractional Brownian motion allows the series to be made locally stationary and the dependence structure between observations to be destroyed locally. Now, denote by $V_{N,\varepsilon}(t, a)$ the following random variable:

$$V_{N,\varepsilon}(t, a) = \frac{1}{\upsilon_N(t)} \sum_{j \in \mathcal{V}_{N,\varepsilon}(t)} \left\{ \frac{V^a(j/N)^2}{\mathbb{E}(V^a(j/N)^2)} - 1 \right\}, \tag{6}$$

where $\mathcal{V}_{N,\varepsilon}(t)$ denotes a neighbourhood of t, defined, for a parameter $\varepsilon > 0$, by $\mathcal{V}_{N,\varepsilon}(t) = \{j = \ell+1, \ldots, N, |j/N - t| \leqslant \varepsilon\}$, and where $\upsilon_N(t) = \#\mathcal{V}_{N,\varepsilon}(t)$. From now on, we define ε as a function of N, say ε_N, such that $\varepsilon = \varepsilon_N \to 0$ and $N\varepsilon_N \to +\infty$ as $N \to +\infty$. The neigbourhood $\mathcal{V}_{N,\varepsilon}(t)$ is sure to contain asymptotically an infinite number of points and to be of length asymptotically zero. More precisely, we suppose the specific form below for ε_N:

$$\varepsilon = \varepsilon_N = \kappa N^{-\alpha} \log(N)^{\beta}, \qquad \text{with } \kappa > 0,\ 0 < \alpha < 1,\ \beta \in \mathbb{R}. \tag{7}$$

Remark. The statistics $V_{N,\varepsilon}(t, a)$ can be seen as the local H_2-variations of a certain Gaussian process (H_2 being the second Hermite polynomial defined by $H_2(t) = t^2 - 1$).

We can now state convergence results for the local H_2-variations of a discretized path of a multifractional Brownian motion (almost surely and in distribution for the topology of Skorohod).

Proposition 1. (i) *Let $t \in [0, 1]$, let a be a filter of order $p \geqslant 1$, and let ε_N be of the form* (7). *Then, as $N \to +\infty$, we have almost surely*

$$V_{N,\varepsilon}(t, a) \to 0. \tag{8}$$

(ii) *Let a be a filter of order $p > \overline{H} + 1/4$, where $\overline{H} = \sup_t H(t)$, and let ε_N be of the form* (7). *Then, as $N \to +\infty$, the following convergence in distribution on* $]0, 1[$ *holds:*

$$\sqrt{2N\varepsilon_N}\, V_{N,\varepsilon}(\cdot, a) \longrightarrow \mathbb{G}, \tag{9}$$

where $\mathbb{G} = \{\mathbb{G}(t),\ t \in]0, 1[\}$ is a centred Gaussian process with covariance function given for $s, t \in]0, 1[$ by

$$\operatorname{cov}(\mathbb{G}(s), \mathbb{G}(t)) = 2 \sum_{k \in \mathbb{Z}} \frac{\pi^a_{H(s)/2+H(t)/2}(k)^2}{\pi^a_{H(s)}(0)\pi^a_{H(t)}(0)}. \tag{10}$$

Moreover, the function $\pi^a_{\cdot}(k)$, defined for $H \in]0, 1[$, is given by

$$\pi_H^a(k) = -\frac{1}{2}\sum_{q,q'=0}^{\ell} a_q a_{q'} |q - q' + k|^{2H}. \tag{11}$$

These two results are local versions of the ones obtained for the fractional Brownian motion case (Coeurjolly 2001, Proposition 1). Note that a filter of order at least 2 ensures asymptotic normality for all the values of the function $H(\cdot)$. For a filter of order 1 (i.e. the filter $a = (1, -1)$), this convergence is available if and only if $0 < \overline{H} < 3/4$.

We now turn our attention to the identification of a multifractional Brownian motion and, in particular, the estimation of the Hurst function using a method of moments.

3. Applications to the Hurst function estimation

Let us introduce, for $m \geqslant 1$, the filter defined, for $i = 0, \ldots, m\ell$, by

$$a_i^m = \begin{cases} a_j, & \text{if } i = jm, \\ 0, & \text{otherwise,} \end{cases}$$

which is nothing more than the filter a dilated m times. Define

$$S_{N,\varepsilon}(t, a^m) = \frac{1}{v_N(t)} \sum_{j \in \mathcal{V}_{N,\varepsilon}(t)} V^{a^m}\left(\frac{j}{N}\right)^2, \qquad \text{for } t \in [0, 1]. \tag{12}$$

The interest of the sequence $(a^m)_{m \geqslant 1}$ relies on the fact that $\pi_H^{a^m}(0) = m^{2H}\pi_H^a(0)$. By virtue of Lemma 1, we have

$$\begin{aligned} \mathbb{E}(S_{N,\varepsilon}(t, a^m)) &= \frac{C \times \pi_{H(t)}^{a^m}(0)}{N^{2H(t)}} + \mathcal{O}\big(\varepsilon_N^\eta \log(N)\big) \\ &= m^{2H(t)} \times \frac{C \times \pi_{H(t)}^{a}(0)}{N^{2H(t)}} + \mathcal{O}\big(\varepsilon_N^\eta \log(N)\big), \end{aligned}$$

which can be restated as

$$\log \mathbb{E}(S_{N,\varepsilon}(t, a^m)) \sim 2H(t)\log(m) + \log(C \times g_{N,a}(H(t))), \qquad \text{as } N \to +\infty.$$

Let $M \geqslant 2$ be an integer. The above relation suggests estimating $H(t)$ by a simple local linear regression of $L_{N,\varepsilon}(t, a, M) = (\log(S_{N,\varepsilon}(t, a^m)))_{m=1,\ldots,M}$ on $(\log(m))_{m=1,\ldots,M}$. We obtain a class of estimators, defined for $t \in]0, 1[$ by

$$\hat{H}_{N,\varepsilon}(t, a, M) = \frac{A^t}{2\|A\|^2} L_{N,\varepsilon}(t, a, M), \tag{13}$$

where A is the vector defined for $m = 1, \ldots, M$ by $A_m = \log(m) - M^{-1}\sum_{m=1}^{M} \log(m)$. This class is a local version of the one obtained for estimating the Hurst parameter of a non-standard fractional Brownian motion (Coeurjolly 2001). Note that the functional estimator of $H(\cdot)$ is clearly independent of the value of C.

Proposition 2. *Let a be a filter of order $p > \overline{H} + 1/4$, where $\overline{H} = \sup_t H(t)$, $M \geqslant 2$ an integer, and assume ε_N is of the form* (7).

(i) *Then, as $N \to +\infty$, we have, for all $t \in [0, 1]$,*

$$\operatorname{bias}\big(\hat{H}_{N,\varepsilon}(t, a, M)\big) = \mathcal{O}\big(\varepsilon_N^{\eta} \log(N)\big), \qquad \operatorname{var}\big(\hat{H}_{N,\varepsilon}(t, a, M)\big) = \mathcal{O}\left(\frac{1}{N\varepsilon_N}\right)$$

$$\text{and, almost surely, } \hat{H}_{N,\varepsilon}(t, a, M) \to H(t). \tag{14}$$

(ii) *Assume ε_N is of the form* (7) *with $\alpha \geqslant 1/(2\eta + 1)$* and *$\beta < 0$. On $]0, 1[$, the following convergence in distribution holds:*

$$\sqrt{2N\varepsilon_N}\big(\hat{H}_{N,\varepsilon}(\cdot, a, M) - H(\cdot)\big) \to \mathbb{G}', \tag{15}$$

where $\mathbb{G}' = \{\mathbb{G}'(t),\ t \in]0, 1[\}$ is a centred Gaussian process with covariance function given for $t, t' \in]0, 1[$ by:

$$\operatorname{cov}(\mathbb{G}'(t), \mathbb{G}'(t')) = \frac{1}{4\|A\|^4} A^t \Sigma\left(\frac{H(t)}{2} + \frac{H(t')}{2},\ H(t),\ H(t')\right) A, \tag{16}$$

with $\Sigma(H_1, H_2, H_3)$ the $M \times M$ matrix whose (m, n)th entry is

$$\big(\Sigma(H_1, H_2, H_3)\big)_{m,n} = 2 \sum_{j \in \mathbb{Z}} \frac{\pi_{H_1}^{a^m, a^n}(j)^2}{\pi_{H_2}^{a^m}(0)\pi_{H_3}^{a^n}(0)}, \quad m, n = 1, \ldots, M.$$

where

$$\pi_H^{a^m, a^n}(j) = \sum_{q=0}^{m\ell} \sum_{q'=0}^{n\ell} a_q a'_q |mq - nq' + j|^{2H}. \tag{17}$$

Remarks.

- Benassi *et al.* (1998) proved the consistency of $\hat{H}_{N,\varepsilon}$ (for the particular filter $(1, -2, 1)$) under the condition that $\varepsilon_N = \mathcal{O}\big(N^{-\alpha} \log(N)^{\beta}\big)$ with $0 < \alpha < 1/2$.
- The condition $\alpha \geqslant 1/(2\eta + 1)$ and $\beta < 0$ in (ii) ensures that $\varepsilon_N^{2\eta} \log(N)^2 = o(N\varepsilon_N)$.
- To estimate $H(t)$ we could have performed a weighted linear regression of $L_{N,\varepsilon}(t, a, M)$ on $(\log(m))_{m=1,\ldots,M}$. In this case, it is easy to derive a result similar to Proposition 2 by locally adapting the corresponding result in Coeurjolly (2000, Proposition 2.5). We have not done this here, because we believe that the gain is too small with respect to the computational cost involved in the estimation of the covariance matrix Σ.

4. Optimal neighbourhood

We now analyse the asymptotic behaviour of the mean integrated squared error (MISE) of $\hat{H}_{N,\varepsilon}(\cdot) - H(\cdot)$. Such a criterion is widely used in functional estimation. Thanks to Proposition 2, it is clear that the MISE (depending on ε_N) has the following behaviour:

$$\mathrm{MISE}(\varepsilon_N) = \mathbb{E}\left(\int_0^1 \left\{\hat{H}_{N,\varepsilon}(t) - H(t)\right\}^2 \mathrm{d}t\right) = \mathcal{O}\left(\varepsilon_N^{2\eta}\log(N)^2\right) + \mathcal{O}\left(\frac{1}{N\varepsilon_N}\right).$$

Considering a discretized version of the MISE, say

$$R_N(\varepsilon) = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{H}_{N,\varepsilon}\left(\frac{i}{N}\right) - H\left(\frac{i}{N}\right)\right)^2,$$

it is immediately evident that the asymptotic behaviour of $\mathbb{E}(R_N(\varepsilon))$ is the same as that of the MISE. Let $\varepsilon_N^\star = \mathrm{argmin}_{\varepsilon_\mathrm{N}} E(R_n(\varepsilon_n))$. From (7), $\varepsilon_n^\star = K_n^{\star -\alpha^\star}\log(n)^{\beta^\star}$ and one can easily see that

$$\alpha^\star = \frac{1}{2\eta+1},\ \beta^\star = -\frac{2}{2\eta+1},\ \text{and MISE}(\varepsilon_N^\star) = \mathcal{O}\left(N^{-2\eta/(2\eta+1)}\log(N)^{2/(2\eta+1)}\right).$$

Define also

$$R'_N(\varepsilon) = \frac{1}{N}\sum_{i=1}^{N}\hat{H}_{N,\varepsilon}\left(\frac{i}{N}\right)^2 - \frac{2}{N}\sum_{i=1}^{N}\hat{H}_{N,\varepsilon}\left(\frac{i}{N}\right)H\left(\frac{i}{N}\right).$$

Since the function $H(\cdot)$ is independent of ε_N, we have

$$\operatorname*{argmin}_{\varepsilon_N}\mathbb{E}(R'_N(\varepsilon_N)) = \operatorname*{argmin}_{\varepsilon_N}\mathbb{E}(R_N(\varepsilon_N)) = \mathcal{O}\left(N^{-2\eta/(2\eta+1)}\log(N)^{2/(2\eta+1)}\right).$$

The above asymptotic result suggests a natural procedure for estimating the optimal neighbourhood. We propose estimating $R'_N(\varepsilon)$ by

$$\hat{R}'_N(\varepsilon) = \frac{1}{N}\sum_{i=1}^{N}\hat{H}_{N,\varepsilon}\left(\frac{i}{N}\right)^2 - \frac{2}{N}\sum_{i=1}^{N}\hat{H}_{N,\varepsilon}\left(\frac{i}{N}\right)\tilde{H}_{N,\varepsilon^2}\left(\frac{i}{N}\right)$$

where $\tilde{H}_{N,\varepsilon^2}(i/N)$ is defined by

$$\tilde{H}_{N,\varepsilon_N^2}(i/N) = \frac{1}{\#\mathcal{V}_{N,\varepsilon^2}(i/N)}\sum_{j\in\mathcal{V}_{N,\varepsilon^2}(i/N)}\hat{H}_{N,\varepsilon^2}(j/N) \tag{18}$$

and represents the average of $\hat{H}_{N,\varepsilon^2}(\cdot)$, in a neighbourhood of i/N of size (of the order of) $N\varepsilon_N^2$, which is the functional estimation of $H(\cdot)$ calculated with a neighbourhood of the same size. Now write

$$\hat{\varepsilon}_N^\star = \operatorname*{argmin}_{\varepsilon_N\in E}\hat{R}'_N(\varepsilon_N),$$

where the set E is defined by

$$E = \left\{(\varepsilon_N)_{N\geqslant 1},\ \varepsilon_N = \kappa N^{-\alpha}\log(N)^\beta,\ \text{with}\ \frac{1}{2\eta+1}\leqslant\alpha<1,\ \text{and}\ \beta<-\frac{1}{6}\right\}. \tag{19}$$

The set E is assumed to be discrete and to contain at most n^ρ elements, for some $\rho>0$.

The following result proves the almost sure convergence of $\hat{H}_{N,\hat{\varepsilon}_N^\star}$ towards $H(t)$ and proves its optimality in the sense that the average of the empirical risk calculated with $\hat{\varepsilon}_N^\star$ is equivalent to $\mathbb{E}(R_N(\hat{\varepsilon}_N^\star))$.

Proposition 3. (i) *For all* $t \in [0, 1]$, *we have almost surely, as* $N \to +\infty$,

$$\hat{H}_{N,\hat{\varepsilon}^\star_N} \to H(t). \tag{20}$$

(ii) *As* $N \to +\infty$,

$$\frac{\mathbb{E}\big(R_N(\hat{\varepsilon}^\star_N)\big)}{\mathbb{E}\big(R_N(\varepsilon^\star_N)\big)} \to 1, \qquad \text{i.e. } \mathbb{E}\big(R_N(\hat{\varepsilon}^\star_N)\big) = \mathcal{O}\Big(N^{-2\eta/(2\eta+1)} \log(N)^{2/(2\eta+1)}\Big). \tag{21}$$

5. A simulation study

To generate a sample path of a standard multifractional Brownian motion discretized at times i/N, $i = 1, \ldots, N$, with covariance matrix $C_{H(\cdot)}$, one can simply extract the square root of $C_{H(\cdot)}$. Then, one generates $Z \rightsquigarrow \mathcal{N}(0, I_N)$ and estimates $\mathbf{W} := C^{1/2}_{H(\cdot)} Z$. This method (which is exact in theory) is sufficiently fast for reasonable sample size N (which is the case here since we chose $N = 2500$). For larger N, it becomes expensive in time and memory, and is numerically unstable. We consider two types of Hurst function: a linear function and a logistic one:

$$H_1(t) = 0.1 + (0.9 - 0.1)t, \tag{22}$$

$$H_2(t) = 0.3 + 0.3/(1 + \exp(-100(t - 0.7))). \tag{23}$$

We generate $R = 50$ series of length $N = 2500$. A Daubechies filter of order 4, $a = Db4$ (with two zero moments) is used to define $\hat{H}_{N,\varepsilon}$. We fix the number of filters to $M = 5$. Let us concentrate first on the optimal neighbourhood. For the sake of simplicity, we choose ε such that $N\varepsilon$ is an integer. Define $h = N\varepsilon$, $h^\star = N\varepsilon^\star$ and $\hat{h}^\star = N\hat{\varepsilon}^\star$. Table 1 summarizes the different estimates of $h^\star$, $\hat{h}^\star$, $\hat{E}(R_N(\varepsilon^\star))$ and $\hat{E}(\hat{R}'_N(\hat{\varepsilon}^\star))$.

Then we applied the previous procedure for each of the $R = 50$ paths to the estimation of the Hurst function. Figure 1 displays the empirical distribution of the functional estimator together with the true function and the estimated confidence bands (at a confidence level $1 - \alpha = 0.95$).

6. Proofs

6.1. Local quadratic variations

Before analysing the asymptotic behaviour of $V_{N,\varepsilon}(t, a)$, we need the following lemma concerning the correlation structure of $(\mathbf{V}^a)$.

Lemma 1. *Let* a *be a filter of order* $p \geqslant 1$, *let* t, $t' \in [0, 1]$, *let* $j \in \mathcal{V}_{N,\varepsilon}(t)$, $j' \in \mathcal{V}_{N,\varepsilon}(t')$ *and let* $\varepsilon = \varepsilon_N$ *be of the form* (7).

Table 1. Theoretical and estimated optimal empirical windows and associated risks ($\hat{\mathbb{E}}$ denotes the empirical mean based on the 50 paths generated)

	$h^\star$	$\widehat{\mathbb{E}}(\hat{h}^\star)$	$\hat{E}(R_N(\varepsilon^\star))$	$\hat{E}(R_N(\hat{\varepsilon}^\star))$
$H(t)$ linear	203	239.76	3.88×10^{-3}	3.631×10^{-3}
$H(t)$ logistic	192	226	4.17×10^{-3}	3.91×10^{-3}

(i) *We have, as* $N \to +\infty$,

$$\mathbb{E}\left(V^a\left(\frac{j}{N}\right) V^a\left(\frac{j'}{N}\right)\right) = \frac{C^2}{N^{H(t)+H(t')}} \pi^a_{H(t)/2+H(t')/2}(j'-j) \times \left\{1 + \mathcal{O}(\varepsilon_N^\eta \log(N))\right\}, \tag{24}$$

where

$$\pi^a_H(k) = -\frac{1}{2} \sum_{q,q'=0}^{\ell} a_q a_{q'} |q - q' + k|^{2H}.$$

(ii) *Define* $Z(j) = \dfrac{V^a(j/n)}{\mathbb{E}(V^a(j/n)^2)^{1/2}}$. *We have*

$$\mathbb{E}(Z(j)Z(j')) = \frac{\pi^a_{H(t)/2+H(t')/2}(j'-j)}{\left\{\pi^a_{H(t)}(0)\pi^a_{H(t')}(0)\right\}^{1/2}} \times \left\{1 + \mathcal{O}(\varepsilon_N^\eta \log(N))\right\}. \tag{25}$$

(iii) *Moreover, as* $k \to +\infty$, *we have* $\pi^a_H(k) = \mathcal{O}(|k|^{2H-2p})$, $\forall H \in]0, 1[$.

Proof. (i) To compute the covariance function, the stochastic representation of a multifractional Brownian motion is used. For the sake of simplicity, let us denote $C'(t) = C \times K(2H(t))^{1/2}/\sqrt{2}$. From (2) and the change of variables $\lambda = Nu$, we obtain

$$\begin{aligned}
&\mathbb{E}\left(V^a\left(\frac{j}{N}\right) V^a\left(\frac{j'}{N}\right)\right) \\
&\quad = \int_{\mathbb{R}} \sum_{q,q'} a_q a_{q'} \left(e^{i(j-q)u} - 1\right)\left(e^{-i(j'-q')u} - 1\right) \frac{C'((j-q)/N)C'((j'-q')/N)}{|Nu|^{H((j-q)/N)+H((j'-q')/N)}} N\,du \\
&\quad = A + B,
\end{aligned}$$

where

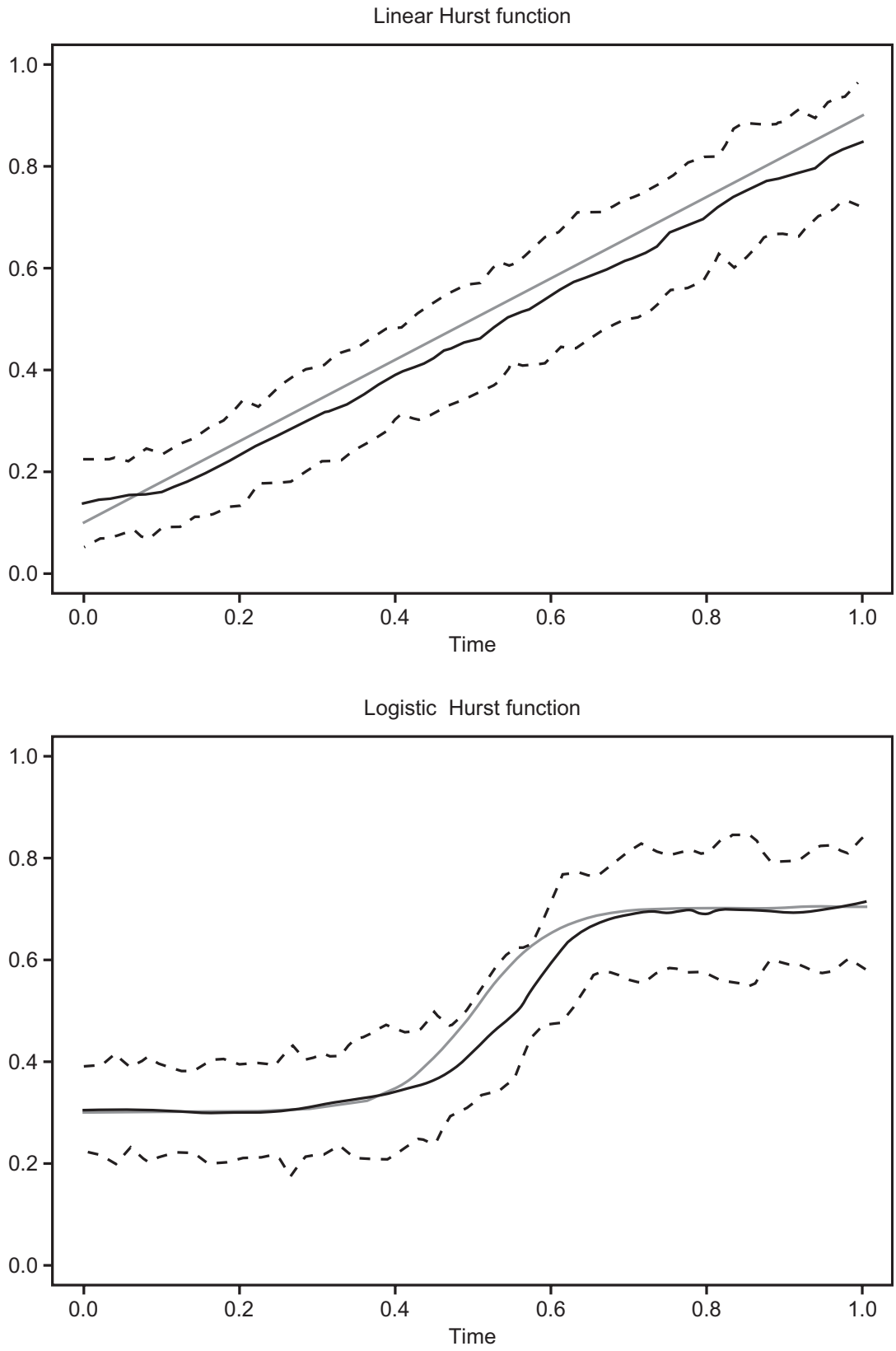


Figure 1. Empirical distributions for the functional estimators of two Hurst functions defined by (22) and (23), and theoretical discretized confidence bands to the level $1 - \alpha = 0.95$.

$$A = \int_{\mathbb{R}} \sum_{q,q'=0}^{\ell} a_q a_{q'} \left(\mathrm{e}^{\mathrm{i}(j-q)u} - 1\right)\left(\mathrm{e}^{-\mathrm{i}(j'-q')u} - 1\right) \frac{C'(t)C'(t')}{N^{H(t)+H(t')}} \frac{\mathrm{d}u}{|u|^{H(t)+H(t')+1}},$$

$$B = \int_{\mathbb{R}} \sum_{q,q'=0}^{\ell} a_q a_{q'} \left(\mathrm{e}^{\mathrm{i}(j-q)u} - 1\right)\left(\mathrm{e}^{-\mathrm{i}(j'-q')u} - 1\right) \frac{|N|^{-H(t)-H(t')}}{|u|^{H(t)+H(t')+1}}$$
$$\times \left\{ \frac{C'((j-q)/N)C'((j'-q')/N)}{|Nu|^{H((j-q)/N)+H((j'-q')/N)-H(t)-H(t')}} - C'(t)C'(t') \right\} \mathrm{d}u.$$

Since the filter is of order at least 1,

$$\sum_{q,q'=0}^{\ell} a_q a_{q'} \left(\mathrm{e}^{\mathrm{i}(j-q)u} - 1\right)\left(\mathrm{e}^{-\mathrm{i}(j'-q')u} - 1\right) = \sum_{q,q'=0}^{\ell} \mathrm{e}^{\mathrm{i}(j-j'+q'-q)u} a_q a_{q'},$$

which allows us to rewrite A as

$$A = N^{-H(t)-H(t')} C'(t)C'(t') \sum_{q,q'=0}^{\ell} a_q a_{q'} \int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}(j-j'+q'-q)u}}{|u|^{H(t)+H(t')+1}} \mathrm{d}u$$

$$= -\tfrac{1}{2} N^{-H(t)-H(t')} C^2 \sum_{q,q'=0}^{\ell} a_q a_{q'} |q - q' + j' - j|^{H(t)+H(t')}$$

$$= \frac{C^2}{N^{H(t)+H(t')}} \pi^a_{H(t)/2+H(t')/2}(j' - j).$$

Since $H \in \mathcal{C}^\eta([0, 1])$,

$$|Nu|^{H((j-q)/N)-H(t)+H((j'-q')/N)-H(t')} = 1 + \mathcal{O}\left(\varepsilon_N^\eta \log(|Nu|)\right), \tag{26}$$

$$C'\left(\frac{j-q}{N}\right) = C'(t) + \mathcal{O}(\varepsilon_N^\eta) \quad \text{and} \quad C'\left(\frac{j'-q'}{N}\right) = C'(t') + \mathcal{O}(\varepsilon_N^\eta). \tag{27}$$

Equations (26) and (27) enable the following upper bound to be obtained for B:

$$B \leqslant \frac{C^2}{N^{H(t)+H(t')}} \pi^a_{H(t)/2+H(t')/2}(j' - j) \mathcal{O}\left(\varepsilon_N^\eta \log(N)\right)$$

$$+ C'(t)C'(t') \sum_{q,q'=0}^{\ell} a_q a_{q'} \left(\int_{\mathbb{R}} \frac{\mathrm{e}^{\mathrm{i}(j-j'+q'-q)u}}{|u|^{H(t)+H(t')+1}} \log(|u|)\mathrm{d}u \right) \times \mathcal{O}(\varepsilon_N^\eta).$$

In a neighbourhood of 0,

$$\sum_{q,q'} a_q a_{q'} \frac{\mathrm{e}^{\mathrm{i}(j-j'+q'-q)u}}{|u|^{H(t)+H(t')+1}} \log(|u|) = o\left(|u|^{2p-H(t)-H(t')} \log(|u|)\right).$$

Thus we can conclude that

$$\mathbb{E}\left(V^a\left(\frac{j}{N}\right)V^a\left(\frac{j'}{N}\right)\right) = \frac{C^2}{N^{H(t)+H(t')}}\pi^a_{H(t)/2+H(t')/2}(j'-j)\left\{1+\mathcal{O}(\varepsilon_N^\eta \log(N))\right\}.$$

(ii) The proof is trivial.

(iii) We refer the reader to Coeurjolly (2001, Lemma 1) for the proof of the asymptotic expansion of $\pi^a_H(k)$. □

Proof of Proposition 1(i). Let us define

$$Z(j) = \frac{V^a(j/N)}{\mathbb{E}(V^a(j/N)^2)^{1/2}}$$

and let H_2 the second Hermite polynomial defined by $H_2(u) = u^2 - 1$. We obtain

$$\mathbb{E}\left(V_{N,\varepsilon}(t,a)^2\right) = \frac{1}{v_N(t)^2}\sum_{j,j'\in\mathcal{V}_{N,\varepsilon}(t)}\mathbb{E}(H_2(Z(j))H_2(Z(j'))),$$

From Lemma 1(ii),

$$\begin{aligned}\mathbb{E}(H_2(Z(j))H_2(Z(j'))) &= 2\mathbb{E}(Z(j)Z(j'))^2\\ &= 2\frac{\pi^a_{H(t)}(j'-j)^2}{\pi^a_{H(t)}(0)^2}\left\{1+\mathcal{O}\left(\varepsilon_N^\eta\log(N)\right)\right\}, \qquad \text{as } N\to+\infty.\end{aligned}$$

Thus, as $N \to +\infty$,

$$\begin{aligned}\mathbb{E}\left(V_{N,\varepsilon}(t,a)^2\right) &\sim \frac{2}{v_N(t)^2}\sum_{j,j'}\frac{\pi^a_{H(t)}(j'-j)^2}{\pi^a_{H(t)}(0)^2}\\ &\sim \frac{2}{v_N(t)^2}\sum_{|j|\le v_N(t)}(v_N(t)-|j|)\frac{\pi^a_{H(t)}(j)^2}{\pi^a_{H(t)}(0)^2}\\ &= \mathcal{O}\left(\frac{1}{v_N(t)}\sum_{|j|\le v_N(t)}\frac{\pi^a_{H(t)}(j)^2}{\pi^a_{H(t)}(0)^2}\right).\end{aligned}$$

Lemma 1 gives the upper bound $\pi^a_{H(t)}(j) \le \mathcal{O}(|j|^{4H(t)-4p})$. Thus,

$$\mathbb{E}\left(V_{N,\varepsilon}(t,a)^2\right) = \begin{cases}\mathcal{O}\left(\dfrac{1}{v_N(t)}\right), & \text{if } p > H(t)+1/4 \text{ (i.e. if } p\ge 2 \text{ or } p=1 \text{ and } H(t)<3/4),\\ \mathcal{O}\left(\dfrac{\log(v_N(t))}{v_N(t)}\right), & \text{if } p=1 \text{ and } H(t)=3/4,\\ \mathcal{O}\left(\dfrac{1}{v_N(t)^{4-4H(t)}}\right), & \text{if } p=1 \text{ and } H(t)>3/4.\end{cases} \tag{28}$$

Therefore, for all $p \geqslant 1$, and for all $H(t) \in]0, 1[$, there exists $\alpha > 1$ such that $\mathbb{E}(V_{N,\varepsilon}(t, a)^2) = \mathcal{O}(v_N(t)^{-\alpha})$. We adapt a result of Doob (1953, p. 492) giving a condition for which the empirical mean of a stationary (centred discretized) process tends almost surely to 0. Let $\overline{V}_{v_N(t)} = V_{N,\varepsilon}(t, a)$.

Let $\alpha' \in \mathbb{R}$ and $m \in \mathbb{N}$ be such that $\alpha\alpha' > 1$ and $v_N(t) > m^{\alpha'}$. We have $\mathbb{E}(\overline{V}^2_{v_N(t)}) = \mathcal{O}(m^{-\alpha\alpha'})$. If $(w_m)_{m \geqslant 1}$ denotes the sequence of integers defined by $w_m = [m^{\alpha'}] + 1$, then, for all $\varepsilon > 0$,

$$\mathbb{P}\big(|\overline{V}_{w_m}| > \varepsilon\big) \leqslant \frac{K^2}{\varepsilon^2} \frac{1}{m^{\alpha\alpha'}}.$$

From the Borel–Cantelli lemma, we have, as $m \to +\infty$, $\overline{V}_{w_m} \to 0$ almost surely. Furthermore,

$$\begin{aligned}
\mathbb{E}\left(\max_{w_m \leqslant v_N(t) \leqslant w_{m+1}} \left| \overline{V}_{v_N(t)} - \frac{w_m}{v_N(t)} \overline{V}_{w_m} \right|^2 \right) &\leqslant \frac{1}{|w_m|^2} \mathbb{E}\left(\left(\sum_{j=w_m}^{w_{m+1}} H_2(Z(j)) \right)^2 \right) \\
&= \frac{2}{|w_m|^2} \sum_{|j| \leqslant w_{m+1} - w_m} (w_{m+1} - w_m - |j|) \frac{\pi^a_{H(t)}(j' - j)^2}{\pi^a_{H(t)}(0)^2} \\
&\leqslant \frac{K}{w_m^2} \frac{(w_{m+1} - w_m)^2}{w_m^2} \leqslant \frac{K'}{m^2}.
\end{aligned}$$

Therefore, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\max_{w_m \leqslant v_N(t) \leqslant w_{m+1}} \left| \overline{V}_{v_N(t)} - \frac{w_m}{v_N(t)} \overline{V}_{w_m} \right| > \varepsilon \right) \leqslant \frac{K'^2}{\varepsilon^2} \frac{1}{m^2},$$

and by Borel–Cantelli lemma

$$\overline{V}_{v_N(t)} - \frac{w_m}{v_N(t)} \overline{V}_{w_m} \to 0, \qquad \text{almost surely as } m \to +\infty.$$

Thus, we have the almost sure convergence of $\overline{V}_{v_N(t)}$ since $\overline{V}_{n_m} \to 0$ almost surely and since $v_N(t) > m^{\alpha'}$.

Finally, from previous computations, note that if $p > H(t) + 1/4$, then

$$\mathbb{E}\big(V_{N,\varepsilon}(t, a)^2\big) \sim \frac{2}{v_N(t)} \sum_{j \in \mathbb{Z}} \frac{\pi^a_{H(t)}(j' - j)^2}{\pi^a_{H(t)}(0)^2}, \qquad \text{as } N \to +\infty.$$

□

Lemma 2. (i) *If $p > H(t) + 1/4$, the following convergence in distribution holds:*

$$\sqrt{v_N(t)} V_{N,\varepsilon}(t, a) \xrightarrow{d} \mathcal{N}\big(0, \sigma^2(H(t), a)\big), \tag{29}$$

with

$$\sigma^2(H(t),a)=2\sum_{j\in\mathbb{Z}}\frac{\pi^a_{H(t)}(j)^2}{\pi^a_{H(t)}(0)^2}. \tag{30}$$

(ii) *Let $d>1$ and let $t_1,\ldots,t_d\in[0,1]$. Then if $p>\overline{H}+1/4$, with $\overline{H}=\sup_t H(t)$, we have*

$$\left(\sqrt{v_N(t_1)}V_{N,\varepsilon}(t_1,a),\ldots,\sqrt{v_N(t_d)}V_{N,\varepsilon}(t_d,a)\right)^{\mathrm{T}}\xrightarrow{d}(\mathbb{G}(t_1),\ldots,\mathbb{G}(t_d))^{\mathrm{T}}, \tag{31}$$

where $(\mathbb{G}(t_1),\ldots,\mathbb{G}(t_d))^{\mathrm{T}}$ is a centred Gaussian vector, such that, for all $i,j\in\{1,\ldots,d\}$,

$$\operatorname{cov}\big(\mathbb{G}(t_i),\mathbb{G}(t_j)\big)=2\sum_{k\in\mathbb{Z}}\frac{\pi^a_{H(t_i)/2+H(t_j)/2}(k)^2}{\pi^a_{H(t_i)}(0)\pi^a_{H(t_j)}(0)}.$$

Proof. (i) Recall that $Z(j)$ denotes the random variable $V^a(j/N)/\mathbb{E}(V^a(j/N)^2)^{1/2}$. From Theorem 1 of Breuer and Major (1983, p. 429) adapted to non-stationary Gaussian vectors, the necessary condition to obtain an asymptotic normality result for $V_{N,\varepsilon}(t,a)$ is the squared summability of $\mathbb{E}((j')Z(j+j'))$, for all $j'\in\mathbb{Z}$. But

$$\sum_{j\in\mathbb{Z}}\mathbb{E}(Z(j')Z(j'+j))^2\sim\sum_{j\in\mathbb{Z}}\mathcal{O}\big(|j|^{4H(t)-4p}\big),\qquad\text{as } N\to+\infty.$$

The result is obtained using the fact that $p>H(t)+1/4$.

(ii) We treat the case $d=2$, since the case $d>2$ can easily be derived. Define for $\lambda,\mu\in\mathbb{R}$ and $t_1,t_2\in]0,1[$, the random variable $T_{N,\varepsilon}(\lambda,\mu)=\lambda V_{N,\varepsilon}(t_1,a)+\mu V_{N,\varepsilon}(t_2,a)$. Note that

$$\begin{aligned}T_{N,\varepsilon}(\lambda,\mu)&=\frac{1}{v_N(t_1)}\sum_{j_1\in\mathcal{V}_{N,\varepsilon}(t_1)}\lambda H_2(Z(j_1))+\frac{1}{v_N(t_2)}\sum_{j_2\in\mathcal{V}_{N,\varepsilon}(t_2)}\mu H_2(Z(j_2))\\&\sim\frac{1}{v_N}\left(\sum_{j_1\in\mathcal{V}_{N,\varepsilon}(t_1)}\lambda v(t_1)H_2(Z(j_1))+\sum_{j_2\in\mathcal{V}_{N,\varepsilon}(t_2)}\mu v(t_2)H_2(Z(j_2))\right)\end{aligned}$$

where $v(t)=\lim_{N\to+\infty}v_N(t_1)/v_N$. Now rewrite $T_{N,\varepsilon}(\lambda,\mu)$ as a simple sum:

$$T_{N,\varepsilon}(\lambda,\mu)=\frac{1}{v_N}\sum_{j=1}^{v_N(t_1)+v_N(t_2)}g_j(Z(j^*)),$$

where

$$j^*=\begin{cases}(\mathcal{V}_{N,\varepsilon}(t_1))_j, & \text{if } 1\leqslant j\leqslant v_N(t_1),\\(\mathcal{V}_{N,\varepsilon}(t_2))_{j-v_N(t_1)}, & \text{if } v_N(t_1)<j\leqslant v_N(t_1)+v_N(t_2),\end{cases}$$

and

$$g_j(\cdot)=\begin{cases}\lambda v(t_1)H_2(\cdot), & \text{if } 1\leqslant j\leqslant v_N(t_1)\\\mu v(t_2)H_2(\cdot) & \text{if } v_N(t_1)<j\leqslant v_N(t_1)+v_N(t_2).\end{cases}$$

The function g_j clearly has Hermite rank 2. Moreover, for $j_1^*, j_2^* \in \{1, \ldots, v_N(t_1) + v_N(t_2)\}$ ($j_1^* \leqslant j_2^*$), it follows from Lemma 1(ii) that

$$\mathbb{E}\big(Z(j_1^*)Z(j_2^*)\big) \underset{N\to+\infty}{\longrightarrow} \begin{cases} \dfrac{\pi^a_{H(t_1)}(j_2^* - j_1^*)}{\pi^a_{H(t_1)}(0)} = \mathcal{O}\big(|j_2^* - j_1^*|^{2H(t_1)-2p}\big), & \text{if } j_1^*, j_2^* \leqslant v_N(t_1), \\ \dfrac{\pi^a_{H(t_2)}(j_2^* - j_1^*)}{\pi^a_{H(t_2)}(0)} = \mathcal{O}\big(|j_2^* - j_1^*|^{H(t_1)+H(t_2)-2p}\big), & \text{if } j_1^*, j_2^* > v_N(t_1), \\ \dfrac{\pi^a_{H(t_1)/2+H(t_2)/2}(j_2^* - j_1^*)}{\Big\{\pi^a_{H(t_1)}(0)\pi^a_{H(t_2)}(0)\Big\}^{1/2}} = \mathcal{O}\big(|j_2^* - j_1^*|^{H(t_1)+H(t_2)-2p}\big), & \text{otherwise.} \end{cases}$$

Thus, for all $j' \in \{1, \ldots, v_N(t_1) + v_N(t_2)\}$, and since $p > \overline{H} + 1/4$, we obtain

$$\sum_{j=1}^{v_N(t_1)+v_N(t_2)} \mathbb{E}\big(Z(j^*)Z(j'^{,*})\big)^2 = \mathcal{O}(1).$$

From Theorem 1 of Breuer and Major (1983, p. 429) adapted to non-stationary Gaussian vectors, there exists $\sigma^2(t_1, t_2)$ such that, for all $\lambda, \mu \in \mathbb{R}$, $\sqrt{N\varepsilon}T_{N,\varepsilon}(\lambda, \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(t_1, t_2))$. As a conclusion, the vector $(\sqrt{v_N(t_1)}V_{N,\varepsilon}(t_1, a), \sqrt{v_N(t_2)}V_{N,\varepsilon}(t_2))^{\mathrm{T}}$ is asymptotically Gaussian. Finally, from previous computations,

$$\operatorname{cov}\Big(\sqrt{v_N(t_1)}V_{N,\varepsilon}(t_1, a), \sqrt{v_N(t_2)}V_{N,\varepsilon}(t_2, a)\Big) \to 2\sum_{j\in\mathbb{Z}} \frac{\pi^a_{H(t_1)/2+H(t_2)/2}(j)^2}{\pi^a_{H(t_1)}(0)\pi^a_{H(t_2)}(0)}. \tag{32}$$

□

To obtain the convergence in distribution of $V_{N,\varepsilon}(\cdot, a)$ for the topology of Skorohod, we need the following inequality, ensuring a tightness criterion.

Lemma 3. *Let a be a filter of order $p > \overline{H} + 1/4$ where $\overline{H} = \sup_t H(t)$, let r be an odd integer greater than 4 and let $t, t' \in I_{N,\varepsilon_N} = [\ell/N + \varepsilon_N, (N-1)/N - \varepsilon_N]$. For $t \in I_{N,\varepsilon_N}$, let $v_N(t) = 2N\varepsilon_N$. Then*

$$\mathbb{E}\Big((2N\varepsilon_N)^{r/2}(V_{N,\varepsilon}(t, a) - V_{N,\varepsilon}(t', a))^r\Big) = \mathcal{O}(|t - t'|^{r\eta}). \tag{33}$$

Proof. Let $v_N = 2N\varepsilon_N$. Let $r \geqslant 4$ be an integer and let $t^* = [N(t' - t)]$. Then

$$\mathbb{E}\Big(v_N^{r/2}(V_{N,\varepsilon}(t, a) - V_{N,\varepsilon}(t', a))^r\Big) = \frac{1}{v_N^{r/2}}\mathbb{E}\left\{\sum_{j\in\mathcal{V}_{N,\varepsilon}(t)} H_2(Z(j)) - H_2(Z(j + t^*))\right\}^r$$

$$= \frac{1}{v_N^{r/2}}\sum_{j_1,\ldots,j_r}\sum_{q=0}^{r}(-1)^q \mathrm{C}_r^q \mathbb{E}\big(H_2(Z(j_1 + t^*)) \ldots H_2(Z(j_q + t^*))H_2(Z(j_{q+1})) \ldots H_2(Z(j_r))\big).$$

From the diagram formula (see, for example, Taqqu 1975),

$$\mathbb{E}\big(H_2(Z(j_1+t^*))\dots H_2(Z(j_q+t^*))H_2(Z(j_{q+1}))\dots H_2(Z(j_r))\big) = T_1 + T_2,$$

where T_1 (T_2) represents the terms obtained by the product of covariances (terms obtained by the product of covariances to the power 2).

Up to permutations on indices, each term of T_1 can be rewritten as

$$T^1_{j_1,\dots,j_r} = \mathbb{E}\big(Z(j_1+t^*)Z(j_2+t^*)\big)\dots\mathbb{E}\big(Z(j_{q-1}+t^*)Z(j_q+t^*)\big)\dots\mathbb{E}(Z(j_{r-1})Z(j_r)). \tag{34}$$

From Lemma 1, there exists $K>0$, $N_1\in\mathbb{N}^*$, such that, for all $N\geqslant N_1$,

$$\begin{aligned}
&\frac{1}{v_N^{r/2}}\sum_{j_1,\dots,j_r} T^1_{j_1,\dots,j_r}\\
&\quad\leqslant \frac{1}{v_N^{r/2}}\sum_{j_1,\dots,j_r}\pi^a_{H(t')}(j_2-j_1)\dots\pi^a_{H(t')}(j_q-j_{q-1})\pi^a_{H(t)/2+H(t')/2}(j_{q+1}-j_q)\dots\\
&\quad\dots\pi^a_{H(t)}(j_r-j_{r-1})\pi^a_{H(t)/2+H(t')/2}(j_1-j_r)\\
&\quad\leqslant \frac{K}{v_N^{r/2}}\sum_{j_2,\dots,j_r}\left\{\sum_{j_1}\pi^a_{H(t)/2+H(t')/2}(j_1-j_r)\pi^a_{H(t')}(j_2-j_1)\right\}\dots\pi^a_{H(t)}(j_r-j_{r-1}).
\end{aligned}\tag{35}$$

Let A_1, A_2 and A_3 be the covariance matrices related to the operators $\pi^a_{H(t)}$, $\pi^a_{H(t')}$ and $\pi^a_{H(t)/2+H(t')/2}$, and let Q_α be the set of squared matrices with terms satisfying

$$|(A)_{j,k}|\leqslant c(1+|k-j|)^{-\alpha},\qquad \alpha>0,\ c>0.$$

It is clear that A_1, A_2 and $A_3\in Q_{2p-2\overline{H}}$, where $\overline{H}=\sup_t H(t)$. Moreover, Jaffard (1990) has proved that Q_α is an algebra, for $\alpha>1$. Thus,

$$A_iA_j\in Q_{2p-2\overline{H}},\qquad \forall i,j=1,2,3.$$

Iterating this argument leads to the existence of a matrix $B\in Q_{2p-2\overline{H}}$ such that

$$\frac{1}{v_N^{r/2}}\sum_{j_1,\dots,j_r}T^1_{j_1,\dots,j_r}\leqslant\frac{C_1}{v_N^{r/2}}\sum_{j_r}(B)_{j_rj_r}\leqslant\frac{C_2}{v_N^{r/2-1}}.$$

Consequently,

$$\frac{1}{v_N^{r/2}}\sum_{j_1,\dots,j_r}\sum_{q=0}^{r}(-1)^q\mathrm{C}_r^q\times T_1\to 0,\qquad \text{as } (N,\varepsilon)\to(+\infty,0). \tag{36}$$

Turning now to the diagram formula and Lemma 1(ii), there exists $K>0$, $N_2\in\mathbb{N}^*$ such that, for all $N\geqslant N_2$,

$$\frac{1}{\upsilon_N^{r/2}}\sum_{q=0}^{r}(-1)^q \mathrm{C}_r^q \sum_{j_1,\ldots,j_r} T_2 \leqslant K\sum_{q=0}^{r}(-1)^q C_r^q$$
$$\times \sum_{\substack{i=0,\ldots,\min(q,r-q)\\ q-i \text{ pair}}} \left\{ \mathrm{A}_{\max(q,r-q)}^{|r-2q|} S\left(\frac{H(t)}{2}+\frac{H(t')}{2}\right)^i \mathrm{N}_{r-q-i} S(H(t))^{(r-q-i)/2} \mathrm{N}_{q-i} S(H(t'))^{(q-i)/2} \right\}. \tag{37}$$

where

$$S(H)=\sum_{j\in\mathbb{Z}} \pi_H^a(j)^2 \quad \text{and} \quad \mathrm{N}_\alpha = (\alpha-1)\times(\alpha-3)\times\ldots\times 3\times 1.$$

Using (36) and (37), we verify that

$$\mathbb{E}\left(\upsilon_N^{r/2}(V_{N,\varepsilon}(t,a)-V_{N,\varepsilon}(t',a))^r\right) = \mathcal{O}\left\{\left(S(H(t))-2S\left(\frac{H(t)}{2}+\frac{H(t')}{2}\right)+S(H(t'))\right)^{r/2}\right\}. \tag{38}$$

Let

$$U(t,t')=S(H(t))-2S\left(\frac{H(t)}{2}+\frac{H(t')}{2}\right)+S(H(t')).$$

Using (11), we obtain:

$$\begin{aligned}
U(t,t') &= \frac{1}{4}\sum_{j\in\mathbb{Z}}\sum_{q_1,\ldots,q_4} a_{q_1}\ldots a_{q_4}(|q_1-q_2+j|\,|j|)^{H(t)+H(t')} \\
&\quad\times\left\{(|q_1-q_2+j|\,|q_3-q_4+j|)^{H(t)-H(t')}-1+(|q_1-q_2+j|\,|q_3-q_4+j|)^{H(t')-H(t)}-1\right\} \\
&= \frac{1}{4}\sum_{j\in\mathbb{Z}}\sum_{q_1,\ldots,q_4} a_{q_1}\ldots a_{q_4}(|q_1-q_2+j|\,|q_3-q_4+j|)^{H(t)+H(t')} \\
&\quad\times\left\{|t-t'|^{2\eta}\log(|q_1-q_2+j|\,|q_3-q_4+j|)^2(1+o(1))\right\} \\
&= |t-t'|^{2\eta}\sum_{j\in\mathbb{Z}}\mathcal{O}\left(|j|^{2H(t)+2H(t')-4p}\log(1+|j|)\right).
\end{aligned}$$

The series converges if $p>\overline{H}+1/4$. Consequently, $U(t,t')=\mathcal{O}(|t-t'|^{2\eta})$. □

Proof of Proposition 1(ii). Let $r=2(1+[1/\eta])$. It follows from Lemma 3 that

$$\mathbb{E}\left((2N\varepsilon_N)^{r/2}(V_{N,\varepsilon}(t,a)-V_{N,\varepsilon}(t',a))^r\right)=\mathcal{O}(|t-t'|^{r\eta}).$$

By Lemma 2 and since $r\eta>1$, we obtain the convergence in distribution, for the topology of Skorohod, of $V_{N,\varepsilon}(\cdot,a)$ towards the Gaussian process $\mathbb{G}$ with covariance function defined by (10). □

6.2. Identification of multifractional Brownian motion

For ease of presentation, let

$$\xi_{N,\varepsilon}(t,\,a) = L_{N,\varepsilon}(t,\,a) - X_M\alpha(t) = \log\left(N^{2H(t)}\frac{S_{N,\varepsilon}(t,\,a)}{C^{\,2}\pi^{a}_{H(t)}(0)}\right).$$

Proof of Proposition 2(i). Note that

$$\hat{H}_{N,\varepsilon}(t,\,a,\,M) - H(t) = \frac{A^t}{2\|A\|^2}\left(\xi_{N,\varepsilon}(t,\,a^m)\right)_{m=1,\ldots,M}, \tag{39}$$

and that, almost surely

$$N^{2H(t)}\frac{S_{N,\varepsilon}(t,\,a^m)}{C^{\,2}\pi^{a^m}_{H(t)}(0)} - 1 = V_{N,\varepsilon}(t,\,a^m) + \mathcal{O}\big(\varepsilon_N^{\eta}\log(N)\big). \tag{40}$$

From Proposition 1(i), we have that almost surely

$$N^{2H(t)}\frac{S_{N,\varepsilon}(t,\,a^m)}{C^{\,2}\pi^{a^m}_{H(t)}(0)} \to 1;$$

therefore $\xi_{N,\varepsilon}(t,\,a^m) \xrightarrow{\text{a.s.}} 0$, which implies the almost sure convergence of $\hat{H}_{N,\varepsilon}(t,\,a,\,M)$ towards $H(t)$. Observe, moreover, that $\mathbb{E}(\xi_{N,\varepsilon}(t,\,a^m)) = \mathcal{O}(\varepsilon_N^{\eta}\log(N))$, so $\mathbb{E}\big(\hat{H}_{N,\varepsilon}(t,\,a,\,M) - H(t)\big) = \mathcal{O}(\varepsilon_N^{\eta}\log(N))$, then that $\operatorname{var}(\xi_{N,\varepsilon}(t,\,a^m)) = \mathcal{O}(v_N(t)^{-1})$, and so $\operatorname{var}\big(\hat{H}_{N,\varepsilon}(t,\,a,\,M)\big) = \mathcal{O}((N\varepsilon_N)^{-1})$. □

Before proving the convergence in distribution, we examine the finite-dimensional convergence of our estimators.

Lemma 4. *Let a be a filter of order $p > \overline{H} + 1/4$, $M \geqslant 2$ an integer and assume that ε_N is of the form* (7) *with $\alpha \geqslant 1/(2\eta+1)$ and $\beta < 0$. Let $d \geqslant 1$ and let $t_1, \ldots, t_d \in [0,\,1]$. Then, writing $B_{N,\hat{H}}(t) = \hat{H}_{N,\varepsilon}(t,\,a,\,M) - H(t)$, we have*

$$\left(\sqrt{v_N(t_1,\,a,\,M)}B_{N,\hat{H}}(t_1),\,\ldots,\,\sqrt{v_N(t_d,\,a,\,M)}B_{N,\hat{H}}(t_d)\right)^{\mathrm{T}} \xrightarrow{d} (\mathbb{G}'(t_1),\,\ldots,\,\mathbb{G}'(t_d))^{\mathrm{T}}, \tag{41}$$

where $(\mathbb{G}'(t_1),\,\ldots,\,\mathbb{G}'(t_d))^{\mathrm{T}}$ is a centred Gaussian vector such that, for all $i, j, \in \{1,\,\ldots,\,d\}$,

$$\operatorname{cov}\big(\mathbb{G}'(t_i),\,\mathbb{G}'(t_j)\big) = \frac{1}{4\|A\|^4}A^{\mathrm{T}}\Sigma\left(\frac{H(t_i)}{2}+\frac{H(t_j)}{2},\,H(t_i),\,H(t_j)\right)A,$$

*with $\Sigma(H_1,\,H_2,\,H_3)$ the $M\times M$ matrix whose $(m,\,n)$*th *entry is*

$$\left(\Sigma(H_1,\,H_2,\,H_3)\right)_{m,n} = 2\sum_{j\in\mathbb{Z}}\frac{\pi^{a^m,a^n}_{H_1}(j)^2}{\pi^{a^m}_{H_2}(0)\pi^{a^n}_{H_3}(0)}, \qquad m,\,n = 1,\,\ldots,\,M,$$

with

$$\pi_H^{a^m,a^n}(j) = \sum_{q=0}^{m\ell}\sum_{q'=0}^{n\ell} a_q a'_q |mq - nq' + j|^{2H},$$

and where A is the vector defined for $m = 1, \ldots, M$ by $A_m = \log(m) - M^{-1}\sum_{m=1}^{M}\log(m)$.

Proof. Let us concentrate on the case $d = 1$. From (40), we have almost surely

$$\xi_{N,\varepsilon}(t, a^m) = \log\left(N^{2H(t)}\frac{S_{N,\varepsilon}(t, a^m)}{C^2\pi^{a^m}_{H(t)}(0)}\right) = V_{N,\varepsilon}(t, a^m)(1 + o(1)) + \mathcal{O}(\varepsilon_N^\eta \log(N)). \quad (42)$$

Since $\alpha \geqslant 1/(2\eta + 1)$ and $\beta < 0$, it follows from (42) that $\sqrt{v_N(t)}\xi_{N,\varepsilon}(t, a^m)$ tends in distribution to the same limit as the random variable $\sqrt{v_N(t, a^m)}V_{N,\varepsilon}(t, a^m)$ and from Coeurjolly (2001, Proposition 3) and Lemma 2(i) we obtain

$$(\sqrt{v_N(t)}(\xi_{N,\varepsilon}(t, a^m))_{m=1,\ldots,M})^{\mathrm{T}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(H(t), H(t), H(t))), \quad (43)$$

where $\Sigma(H(t), H(t), H(t))$ is the $M \times M$ matrix whose (m, n)th entry is

$$\begin{aligned}(\Sigma(H(t), H(t), H(t)))_{m,n} &= \lim_{N\to+\infty}\sqrt{v_N(t)}\sqrt{v_N(t)}\mathbb{E}(V_{N,\varepsilon}(t, a^m)V_{N,\varepsilon}(t, a^n)) \\ &= 2\sum_{j\in\mathbb{Z}}\frac{\pi^{a^m,a^n}_{H(t)}(j)^2}{\pi^{a^m}_{H(t)}(0)\pi^{a^n}_{H(t)}(0)}, \qquad m, n = 1, \ldots, M,\end{aligned}$$

with

$$\pi_H^{a^m,a^n}(j) = \sum_{q=0}^{m\ell}\sum_{q'=0}^{n\ell} a_q a'_q |mq - nq' + j|^{2H}.$$

The results (39) and (43) ensure that $\sqrt{v_N(t)}(\hat{H}_{N,\varepsilon}(t, a) - H(t))$ is asymptotically Gaussian.

The case $d > 1$ is easily deduced using Lemma 2(ii), which implies the finite-dimensional Gaussian convergence of $\xi_{N,\varepsilon}(\cdot, a)$. We end with the following computation for $t, t' \in [0, 1]$:

$$\begin{aligned}&\mathrm{cov}(\sqrt{v_N(t)}\{\hat{H}_{N,\varepsilon}(t, a, M) - H(t)\}, \sqrt{v_N(t')}\{\hat{H}_{N,\varepsilon}(t', a, M) - H(t')\}) \\ &\sim \frac{A^{\mathrm{T}}}{4\|A\|^4}\mathbb{E}(\sqrt{v_N(t)}\{V_{N,\varepsilon}(t, a^1), \ldots, V_{N,\varepsilon}(t, a^M)\}^{\mathrm{T}}\sqrt{v_N(t')}\{V_{N,\varepsilon}(t', a^1), \ldots, V_{N,\varepsilon}(t', a^M)\})A \\ &\to \frac{1}{4\|A\|^4}A^{\mathrm{T}}\Sigma\left(\frac{H(t)}{2} + \frac{H(t')}{2}, H(t), H(t')\right)A, \qquad \text{as } N \to +\infty.\end{aligned}$$

Proof of Proposition 2(ii). Let $r \geqslant 4$ be an integer, and let $t, t' \in]0, 1[$. By (42), we have almost surely

$$(\xi_{N,\varepsilon}(t, a^m) - \xi_{N,\varepsilon}(t', a^m)) = (V_{N,\varepsilon}(t, a^m) - V_{N,\varepsilon}(t', a^m))(1 + o(1)) + \mathcal{O}(\varepsilon_N^\eta \log(N)). \quad (44)$$

Thus, if ε_N is such that $\alpha \geqslant 1/(2\eta + 1)$ and $\beta < 0$, we obtain

$$\mathbb{E}((2N\varepsilon_N)^{r/2}\{\xi_{N,\varepsilon}(t,\, a^m) - \xi_{N,\varepsilon}(t',\, a^m)\}^r) = \mathcal{O}(\mathbb{E}((2N\varepsilon_N)^{r/2}\{V_{N,\varepsilon}(t,\, a^m) - V_{N,\varepsilon}(t',\, a^m)\}^r)).$$

Choosing r large enough, we obtain, using Lemma 3, the convergence in distribution on the range $]0,\, 1[$ of $\sqrt{2N\varepsilon_N}\xi_{N,\varepsilon}(\cdot,\, a^m)$, and then of $\sqrt{2N\varepsilon_N}(\hat{H}_{N,\varepsilon}(\cdot,\, a,\, M) - H(\cdot))$ using (39). □

Proof of Proposition 3. (i) In fact it is sufficient to prove that, for all $t \in [0,\, 1]$, almost surely

$$\sup_{\varepsilon_N \in E} |\hat{H}_{N,\varepsilon}(t)| \to 0, \qquad \text{as } N \to +\infty.$$

Let $\lambda > 0$ and $k \geqslant 1$ be an integer. We have, from (19) and from Chebyshev's inequality,

$$\mathbb{P}\left(\sup_{\varepsilon_N \in E}\left|\hat{H}_{N,\varepsilon}(t) - H(t)\right| \geqslant \lambda\right) \leqslant n^\rho \sup_{\varepsilon_N \in E}\mathbb{P}\left(\left|\hat{H}_{N,\varepsilon}(t) - H(t)\right| \geqslant \lambda\right)$$

$$\leqslant n^\rho \frac{1}{\rho^{2k}} \sup_{\varepsilon_N \in E}\mathbb{E}\left(\left(\hat{H}_{N,\varepsilon}(t) - H(t)\right)^{2k}\right). \tag{45}$$

Using (44) and the fact that $\alpha \geqslant 1/(2\eta + 1)$ and $\beta < 0$, we have

$$\mathbb{E}\left(\left(\hat{H}_{N,\varepsilon}(t) - H(t)\right)^{2k}\right) = \mathcal{O}\left(\frac{1}{(N\varepsilon_N)^k}\right).$$

Choosing k sufficiently large leads to the summability of $\sum_n \mathbb{P}(\sup_{\varepsilon_N \in E}|\hat{H}_{N,\varepsilon}(t) - H(t)| \geqslant \lambda)$ and to the result, using the Borel–Cantelli lemma.

(ii) Since $\hat{\varepsilon}^\star_N = \operatorname{argmin}_{\varepsilon_N \in E}\hat{R}'_N(\varepsilon_N)$ and since $\varepsilon^\star_N \in E$, we have almost surely $\hat{R}'_N(\hat{\varepsilon}^\star_N) \leqslant \hat{R}'_N(\varepsilon^\star_N)$ and so $\mathbb{E}(\hat{R}'_N(\hat{\varepsilon}^\star_N)) \leqslant \mathbb{E}(\hat{R}'_N(\varepsilon^\star_N))$. Now

$$\frac{|\mathbb{E}(R_N(\hat{\varepsilon}^\star_N)) - \mathbb{E}(R_N(\varepsilon^\star_N))|}{\mathbb{E}(R_N(\varepsilon^\star_N))} = \frac{\mathbb{E}(R_N(\hat{\varepsilon}^\star_N)) - \mathbb{E}(R_N(\varepsilon^\star_N))}{\mathbb{E}(R_N(\varepsilon^\star_N))} = \frac{\mathbb{E}(R'_N(\hat{\varepsilon}^\star_N)) - \mathbb{E}(R'_N(\varepsilon^\star_N))}{\mathbb{E}(R_N(\varepsilon^\star_N))}$$

$$\leqslant \frac{\mathbb{E}(R'_N(\hat{\varepsilon}^\star_N)) - \mathbb{E}(R'_N(\varepsilon^\star_N)) + \mathbb{E}(\hat{R}'_N(\varepsilon^\star_N)) - \mathbb{E}(\hat{R}'_N(\hat{\varepsilon}^\star_N))}{\mathbb{E}(R_N(\varepsilon^\star_N))},$$

since $\mathbb{E}(\hat{R}'_N(\varepsilon^\star_N)) - \mathbb{E}(\hat{R}'_N(\varepsilon^\star_N) \geqslant 0$. Finally, we have

$$\frac{|\mathbb{E}(R_N(\hat{\varepsilon}^\star_N)) - \mathbb{E}(R_N(\varepsilon^\star_N))|}{\mathbb{E}(R_N(\varepsilon^\star_N))} \leqslant 2\sup_{\varepsilon_N \in E}\frac{\left|\mathbb{E}\left(R'_N(\varepsilon_N) - \hat{R}'_N(\varepsilon_N)\right)\right|}{\mathbb{E}(R_N(\varepsilon^\star_N))}. \tag{46}$$

For $\varepsilon_N \in E$, there exists $N_0 \in \mathbb{N}$ such that, for all $N \geqslant N_0$,

$$\mathbb{E}\left(R'_N(\varepsilon_N) - \hat{R}'_N(\varepsilon_N)\right) = \frac{2}{N}\sum_{i=0}^{N-1}\mathbb{E}\left(\left(\tilde{H}_{N,\varepsilon^2}\left(\frac{i}{N}\right) - H\left(\frac{i}{N}\right)\right)\hat{H}_{N,\varepsilon}\left(\frac{i}{N}\right)\right)$$

$$\leqslant \frac{4}{N}\sum_{i=0}^{N-1}\mathbb{E}\left(\tilde{H}_{N,\varepsilon^2}\left(\frac{i}{N}\right) - H\left(\frac{i}{N}\right)\right). \tag{47}$$

Moreover,

$$\mathbb{E}\left(\tilde{H}_{N,\varepsilon^2}\left(\frac{i}{N}\right) - H\left(\frac{i}{N}\right)\right) = \frac{1}{\#\mathcal{V}_{N,\varepsilon^2}(i/N)} \sum_{j\in\mathcal{V}_{N,\varepsilon^2}(i/N)} \mathbb{E}\left(\hat{H}_{N,\varepsilon^2}\left(\frac{j}{N}\right) - H\left(\frac{j}{N}\right)\right)$$

$$+ \frac{1}{\#\mathcal{V}_{N,\varepsilon^2}(i/N)} \sum_{j\in\mathcal{V}_{N,\varepsilon^2}(i/N)} \left(H\left(\frac{j}{N}\right) - H\left(\frac{i}{N}\right)\right)$$

$$= \mathcal{O}\left(\varepsilon_N^{2\eta}\log(N)\right) + \mathcal{O}\left(\varepsilon_N^{2\eta}\right) = \mathcal{O}\left(\varepsilon_N^{2\eta}\log(N)\right). \tag{48}$$

Combining (47) and (48), we obtain

$$\frac{\mathbb{E}\left(R'_N(\varepsilon_N) - \hat{R}'_N(\varepsilon_N)\right)}{\mathbb{E}(R_N(\varepsilon_N^\star))} = \mathcal{O}\left(N^{2\eta/(2\eta+1)-2\alpha\eta}\log(N)^{1+2\beta\eta-2/(2\eta+1)}\right).$$

Since $0 < \eta \leqslant 1$, the proof is achieved using the definition of the set E (see (19), and (46)). □

Acknowledgement

This work was financially supported by a grant from IMAG project AMOA. The author thanks Professor Anestis Antoniadis for helpful comments and the editor and a referee for their suggestions and guidance.

References

Benassi, A., Cohen, S. and Istas, J. (1998) Identifying the multifractional function of a Gaussian process. *Statist. Probab. Lett.*, **39**, 337–345.

Breuer, P. and Major, P. (1983) Central limit theorems for non-linear functionals of Gaussian fields. *J. Multivariate Anal.*, **13**, 425–441.

Coeurjolly, J.-F. (2000) Inférence statistique pour les mouvements browniens fractionnaires et multi-fractionnaires. Doctoral thesis, Université de Grenoble.

Coeurjolly, J.-F. (2001) Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Statist. Inference Stochastic Process*., **4**, 199–227.

Coeurjolly, J.-F. and Istas, J. (2001). Cramér–Rao bounds for fractional Brownian motions. *Statist. Probab. Lett.*, **53**, 435–448.

Cohen, S. (1999) From self-similarity to local self-similarity: the estimation problem. In M. Dekking, J. Lévy Véhel, E. Lutton, and C. Tricot (eds), *Fractals: Theory and Applications in Engineering*, pp. 3–16. London: Springer-Verlag.

Collins, J.J. and De Luca, C.J. (1994) Upright, correlated random walks: A statistical-biomechanics approach to the human postural control system. *Chaos*, **5**(1), 57–63.

Doob, J.L. (1953) *Stochastic Processes*. New York: Wiley.

Frisch, U. (1997) *Turbulence. The Legacy of A.N. Kolmogorov.* Cambridge: Cambridge University Press.

Istas, J. and Lang, G. (1997) Quadratic variations and estimation of the Hölder index of a Gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.*, **33**, 407–436.

Jaffard, S. (1990) Propriétés ds matrices 'bien localisées' près de leur diagonale et quelques applications. *Ann. Inst. H. Poincaré Anal. Non Lineaire*, **7**, 461–476.

Kent. J.T. and Wood, A.T.A. (1997) Estimating the fractal dimension of a locally self-similar Gaussian process using increments. *J. Roy. Statist. Soc. Ser. B*, **59**, 679–700.

Mandelbrot, B. and Van Ness, J. (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, **10**, 422–437.

Peltier, R.-F. and Lévy Véhel, J. (1995) Multifractional Brownian motion: definition and preliminary results. INRIA Research Report No. 2645.

Taqqu, M.S. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **40**, 203–238.

Von Bahr, B. and Esseen, C.G. (1965) Inequalities for the rth moment of a sum of random variables, $1 \leqslant r \leqslant 2$. *Ann. Math. Statist.*, **36**, 299–303.

Received April 2003 and revised November 2004

Erratum : Identification of Multifractional Brownian Motion

COEURJOLLY Jean-François

E-mail: Jean-Francois.Coeurjolly@upmf-grenoble.fr.

Labsad, BSHM, UPMF, Avenue Centrale B.P. 47 - 38040 Grenoble Cedex 9, FRANCE.

Date : 13th January 2006.

An error occurs in the proof of Lemma 2 (Coeurjolly (2005), p.1001). The following has to be read

$$\mathbb{E}\left(Z(j_1^*)Z(j_2^*)\right) \overset{N\to+\infty}{\longrightarrow} \begin{cases} \frac{\pi^a_{H(t_1)}(j_2^*-j_1^*)}{\pi^a_{H(t_1)}(0)} = \mathcal{O}\left(|j_2^*-j_1^*|^{2H(t_1)-2p}\right) & \text{if } j_1^*, j_2^* \leq v_N(t_1) \\ \frac{\pi^a_{H(t_2)}(j_2^*-j_1^*)}{\pi^a_{H(t_2)}(0)} = \mathcal{O}\left(|j_2^*-j_1^*|^{H(t_1)+H(t_2)-2p}\right) & \text{if } j_1^*, j_2^* > v_N(t_1) \\ 0 & \text{else.} \end{cases}$$

For the third situation, it was previously written

$$\frac{\pi^a_{\frac{H(t_1)}{2}+\frac{H(t_2)}{2}}(j_2^*-j_1^*)}{\left\{\pi^a_{H(t_1)}(0)\pi^a_{H(t_2)}(0)\right\}^{1/2}} = \mathcal{O}\left(|j_2^*-j_1^*|^{H(t_1)+H(t_2)-2p}\right).$$

But for n sufficiently large, t_1 and t_2 are sufficiently separated in the sense that the neighborhoods $\mathcal{V}_{N,\varepsilon}(t_1)$ and $\mathcal{V}_{N,\varepsilon}(t_2)$ do not overlap, which then implies that this term tends to zero.

The consequence of this error is that the correct statement of Proposition 1 (ii) is the following: the finite-dimensional laws of the process $\left\{\sqrt{2N\varepsilon_N}V_{N,\varepsilon}(t,a), t\in]0,1[\right\}$ converges, when $N\to+\infty$, towards those of a centered Gaussian $\{\mathbb{G}(t), t\in]0,1[\}$ with covariance function defined by

$$Cov\left(\mathbb{G}(s),\mathbb{G}(t)\right) = \begin{cases} 2\sum_{k\in\mathbb{Z}} \frac{\pi^a_{H(t)}(k)^2}{\pi^a_{H(t)}(0)^2} & \text{if } s=t \\ 0 & \text{if } s\neq t. \end{cases}$$

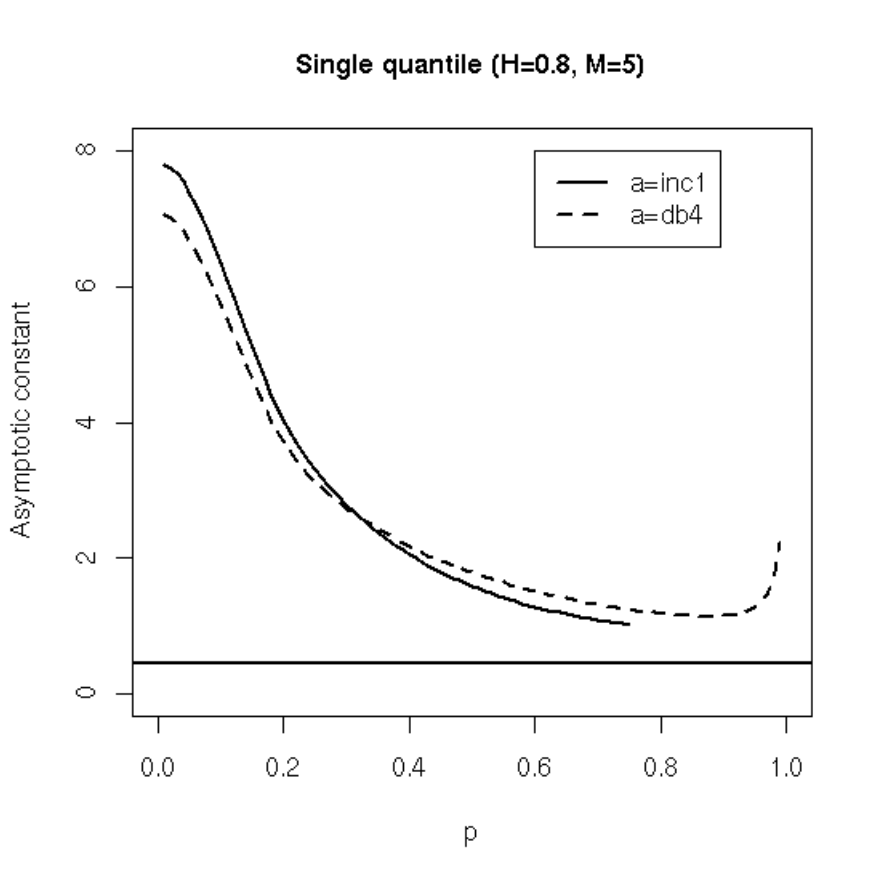
A similar remark applies to Proposition 2.

I am sincerely very grateful to A. Begyn (University of Toulouse III), who indicated me this mistake. This error has been corrected in his paper (Begyn (2005)) that generalizes this work.

References

A. Begyn (2005), Functional limit theorems for generalized quadratic variations of Gaussian processes, preprint available at `http://www.lsp-ups-tlse.fr/Fp/Begyn/`.

J.-F. Coeurjolly (2005), Identification of multifractional Brownian motion, Bernoulli **11**(6), 987-1008.



[Coe08a]
J.-F. Coeurjolly. Bahadur representation of sample quantiles for functional of Gaussian dependent sequences under a minimal assumption. *Stat. Probab. Lett.*, 78(15):2485-2489, 2008.

Statistics and Probability Letters 78 (2008) 2485–2489

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Bahadur representation of sample quantiles for functional of Gaussian dependent sequences under a minimal assumption

Jean-François Coeurjolly

SAGAG, Department of Statistics, Grenoble, 1251 Av. Centrale BP 47, 38040 GRENOBLE Cedex 09, France

ARTICLE INFO

Article history:
Received 9 March 2007
Received in revised form 29 August 2007
Accepted 14 February 2008
Available online 20 March 2008

ABSTRACT

We obtain a Bahadur representation for sample quantiles of a nonlinear functional of Gaussian sequences with correlation function decreasing as $k^{-\alpha}$ for some $\alpha > 0$. This representation is derived under a minimal assumption.

1. Introduction

We consider the problem of obtaining a Bahadur representation of sample quantiles in a certain dependence context. Before stating in what a Bahadur representation consists, let us specify some general notation. Given some random variable Y, $F(\cdot) = F_Y(\cdot)$ is referred to as the cumulative distribution function of Y, $\xi(p) = \xi_Y(p)$ for some $0 < p < 1$ as the quantile of order p. If $F(\cdot)$ is absolutely continuous with respect to Lebesgue measure, the probability density function is denoted by $f(\cdot) = f_Y(\cdot)$. On the basis of the observation of a vector $\boldsymbol{Y} = (Y(1), \ldots, Y(n))$ of n random variables distributed as Y, the sample cumulative distribution function and the sample quantile of order p are respectively denoted by $\widehat{F}_Y(\cdot; \boldsymbol{Y})$ and $\widehat{\xi}_Y(p; \boldsymbol{Y})$ or simply by $\widehat{F}(\cdot)$ and $\widehat{\xi}(p)$.

Let $\boldsymbol{Y} = (Y(1), \ldots, Y(n))$ be a vector of n i.i.d. random variables such that $F''(\xi(p))$ exists and is bounded in a neighborhood of $\xi(p)$ and such that $F'(\xi(p)) > 0$. Bahadur (1966) proved that as $n \to +\infty$,

$$\widehat{\xi}(p) - \xi(p) = \frac{p - \widehat{F}(p)}{f(\xi(p))} + r_n,$$

with $r_n = \mathcal{O}_{a.s.}\left(n^{-3/4}\log(n)^{3/4}\right)$ where a sequence of random variables U_n is said to be $\mathcal{O}_{a.s.}(v_n)$ if U_n/v_n is almost surely bounded. Kiefer (1967) obtained the exact rate $n^{-3/4}\log\log(n)^{3/4}$. Under an assumption on $F(\cdot)$ which is quite similar to the one made by Bahadur, extensions of above results to dependent random variables have been pursued in Sen (1972) for ϕ-mixing variables, in Yoshihara (1995) for strongly mixing variables, and recently in Wu (2005) for short-range and long-range dependent linear processes, following works of Hesse (1990) and Ho and Hsing (1996). Finally, such a representation has been obtained by Coeurjolly (2008) for a nonlinear functional of Gaussian sequences with correlation function decreasing as $k^{-\alpha}$ for some $\alpha > 0$.

Ghosh (1971) proposed in the i.i.d. case a much simpler proof of Bahadur's result which suffices for many statistical applications. He established under a weaker assumption on $F(\cdot)$ ($F'(\cdot)$ exists and is bounded in a neighborhood of $\xi(p)$ and $f(\xi(p)) > 0$) that the remainder term satisfies $r_n = o_{\mathbb{P}}(n^{-1/2})$, which means that $n^{1/2}r_n$ tends to 0 in probability. This result is sufficient for example to establish a central limit theorem for the sample quantile. Our goal is to extend Ghosh's result to a nonlinear functional of Gaussian sequences with correlation function decreasing as $k^{-\alpha}$. The Bahadur representation is presented in Section 2 and is applied to a central limit theorem for the sample quantile. Proofs are deferred to Section 3.

E-mail address: Jean-Francois.Coeurjolly@upmf-grenoble.fr.

0167-7152/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2008.02.037

2. Main result

Let $\{Y(i)\}_{i=1}^{+\infty}$ be a stationary (centered) Gaussian process with variance 1, and correlation function $\rho(\cdot)$ such that, as $i \to +\infty$

$$|\rho(i)| \sim i^{-\alpha} \tag{1}$$

for some $\alpha > 0$.

Let us recall some background on Hermite polynomials: the Hermite polynomials form an orthogonal system for the Gaussian measure and are in particular such that $\mathbf{E}\left(H_j(Y)H_k(Y)\right) = j!\,\delta_{j,k}$, where Y is referred to a standard Gaussian variable. For some measurable function $g(\cdot)$ defined on $\mathbb{R}$ such that $\mathbf{E}(g(Y)^2) < +\infty$, the following expansion holds:

$$g(t) = \sum_{j\geq\tau} \frac{c_j}{j!} H_j(t) \quad \text{with } c_j = \mathbf{E}\left(g(Y)H_j(Y)\right),$$

where the integer τ defined by $\tau = \inf\{j \geq 0, c_j \neq 0\}$, is called the Hermite rank of the function g. Note that this integer plays an important role. For example, it is related to the correlation of $g(Y_1)$ and $g(Y_2)$, for Y_1 and Y_2 two standard Gaussian variables with correlation ρ, since $\mathbf{E}(g(Y_1)g(Y_2)) = \sum_{k\geq\tau} \frac{(c_k)^2}{k!}\rho^k = \mathcal{O}\left(\rho^\tau\right)$.

Our result is based on the assumption that $F'_{g(Y)}(\cdot)$ exists and is bounded in a neighborhood of $\xi(p)$. This is achieved if the function $g(\cdot)$ satisfies the following assumption (see e.g. Dacunha-Castelle and Duflo (1982), p. 33).

Assumption $\boldsymbol{A(\xi(p))}$: there exist U_i, $i = 1, \ldots, L$, disjoint open sets such that U_i contains a unique solution to the equation $g(t) = \xi_{g(Y)}(p)$, such that $F'_{g(Y)}(\xi(p)) > 0$ and such that g is a $\mathcal{C}^1$-diffeomorphism on $\cup_{i=1}^L U_i$.

Note that this assumption allows us to obtain

$$F'_{g(Y)}(\xi_{g(Y)}(p)) = f_{g(Y)}(\xi_{g(Y)}(p)) = \sum_{i=1}^{L} \frac{\phi(g_i^{-1}(t))}{g'(g_i^{-1}(t))},$$

where $g_i(\cdot)$ is the restriction of $g(\cdot)$ on U_i and where $\phi(\cdot)$ is referred to the probability density function of a standard Gaussian variable.

Now, define, for some real u, the function $h_u(\cdot)$ by

$$h_u(t) = \mathbf{1}_{\{g(t)\leq u\}}(t) - F_{g(Y)}(u). \tag{2}$$

We denote by $\tau(u)$ the Hermite rank of $h_u(\cdot)$. For the sake of simplicity, we set $\tau_p = \tau(\xi_{g(Y)}(p))$. For some function $g(\cdot)$ satisfying Assumption $\boldsymbol{A(\xi(p))}$, we define

$$\overline{\tau}_p = \inf_{\gamma\in\cup_{i=1}^L g(U_i)} \tau(\gamma), \tag{3}$$

that is the minimal Hermite rank of $h_u(\cdot)$ for u in a neighborhood of $\xi_{g(Y)}(p)$. Denote also by $c_j(u)$ the j-th Hermite coefficient of the function $h_u(\cdot)$.

Theorem 1. *Under Assumption* $\boldsymbol{A(\xi(p))}$*, the following result holds as* $n \to +\infty$*:*

$$\widehat{\xi}\left(p; \boldsymbol{g(Y)}\right) - \xi_{g(Y)}(p) = \frac{p - \widehat{F}\left(\xi_{g(Y)}(p); \boldsymbol{g(Y)}\right)}{f_{g(Y)}(\xi_{g(Y)}(p))} + o_{\mathbb{P}}\left(r_n(\alpha, \overline{\tau}_p)\right), \tag{4}$$

where $\boldsymbol{g(Y)} = (g(Y(1)), \ldots, g(Y(n)))$*, for* $i = 1, \ldots, n$*, and where the sequence* $\left(r_n(\alpha, \overline{\tau}_p)\right)_{n\geq 1}$ *is defined by*

$$r_n(\alpha, \overline{\tau}_p) = \begin{cases} n^{-1/2} & \text{if } \alpha\overline{\tau}_p > 1, \\ n^{-1/2}\log(n)^{1/2} & \text{if } \alpha\overline{\tau}_p = 1, \\ n^{-\alpha\overline{\tau}_p/2} & \text{if } \alpha\overline{\tau}_p < 1. \end{cases} \tag{5}$$

Remark 1. The sequence $r_n(\alpha, \overline{\tau}_p)$ is related to the short-range or long-range dependent behaviour of the sequence $h_u(Y(1)), \ldots, h_u(Y(n))$ for u in a neighborhood of $\xi(p)$. More precisely, it corresponds to the asymptotic behaviour of the sequence

$$\left(\frac{1}{n}\sum_{|i|<n} \rho(i)^{\overline{\tau}_p}\right)^{1/2}.$$

Corollary 2. *Under Assumption* $\boldsymbol{A(\xi(p))}$*, then the following convergences in distribution hold as* $n \to +\infty$*:*

(i) *if* $\alpha\overline{\tau}_p > 1$

$$\sqrt{n}\left(\widehat{\xi}\left(p; \boldsymbol{g(Y)}\right) - \xi_{g(Y)}(p)\right) \xrightarrow{d} \mathcal{N}(0, \sigma_p^2), \tag{6}$$

where

$$\sigma_p^2 = \frac{1}{f(p)^2}\sum_{i\in\mathbb{Z}}\sum_{j\geq\overline{\tau}_p}\frac{c_j(p)^2}{j!}\rho(i)^j \quad \text{with } f(p) = f_{g(Y)}(\xi_{g(Y)}(p)) \text{ and } c_j(p) = c_j(\xi_{g(Y)}(p)),$$

(ii) *if* $\alpha\overline{\tau}_p < 1$

$$n^{\alpha\overline{\tau}_p/2}\left(\widehat{\xi}\,(p;\boldsymbol{g}(\boldsymbol{Y})) - \xi_{g(Y)}(p)\right) \xrightarrow{d} \frac{c_{\overline{\tau}_p}(p)}{\overline{\tau}_p! f(p)} Z_{\overline{\tau}_p}, \tag{7}$$

where

$$Z_{\overline{\tau}_p} = K(\overline{\tau}_p,\alpha)\int_{\mathbb{R}^{\overline{\tau}_p}}{}' \frac{\exp(i(\lambda_1+\cdots+\lambda_{\overline{\tau}_p}))-1}{i(\lambda_1+\cdots+\lambda_{\overline{\tau}_p})}\prod_{j=1}^{\overline{\tau}_p}|\lambda_j|^{(\alpha-1)/2}\widetilde{B}(d\lambda_j)$$

and

$$K(\overline{\tau}_p,\alpha) = \left(\frac{(1-\alpha\overline{\tau}_p/2)(1-\alpha\overline{\tau}_p)}{\overline{\tau}_p!\,(2\Gamma(\alpha)\sin(\pi(1-\alpha)/2))^{\overline{\tau}_p}}\right)^{1/2}.$$

The measure $\widetilde{B}$ *is a Gaussian complex measure and the symbol* $\int'$ *means that the domain of integration excludes the hyperdiagonals* $\{\lambda_i = \pm\lambda_j, i\neq j\}$.

The proof of this result is omitted since it is a direct application of Theorem 1 and general limit theorems adapted to nonlinear functionals of Gaussian sequences; e.g. Breuer and Major (1983) and Dehling and Taqqu (1989).

3. Proofs

3.1. Auxiliary lemma

Lemma 3. *For every* $j\geq 1$ *and for all positive sequence* $(u_n)_{n\geq 1}$ *such that* $u_n\to 0$, *as* $n\to+\infty$, *we have, under Assumption* $\boldsymbol{A}(\xi(\boldsymbol{p}))$,

$$I = \int_{\mathbb{R}} H_j(t)\phi(t)\mathbf{1}_{\{|g(t)-\xi_{g(Y)}(p)|\leq u_n\}}\mathrm{d}t \sim u_n\kappa_j, \tag{8}$$

where κ_j *is defined, for every* $j\geq 1$, *by*

$$\kappa_j = \begin{cases} -2\displaystyle\sum_{i=1}^{L}\frac{\phi'\left(g_i^{-1}(\xi(p))\right)}{g'\left(g_i^{-1}(\xi(p))\right)} & \text{if } j=1,\\ 2(-1)^j\displaystyle\sum_{i=1}^{L}\frac{\phi^{(j-2)}\left(g_i^{-1}(\xi(p))\right)}{g'\left(g_i^{-1}(\xi(p))\right)} & \text{if } j>1. \end{cases} \tag{9}$$

Proof. Under Assumption $\boldsymbol{A}(\boldsymbol{\xi}(\boldsymbol{p}))$, there exists $n_0\in\mathbb{N}$ such that for all $n\geq n_0$,

$$I = \sum_{i=1}^{L} I_i \quad \text{with } I_i = \int_{U_i} H_j(t)\phi(t)\mathbf{1}_{\{\xi(p)-u_n\leq g(t)\leq\xi(p)+u_n\}}\mathrm{d}t. \tag{10}$$

Assume without loss of generality that the restriction of $g(\cdot)$ on U_i (denoted by $g_i(\cdot)$) is an increasing function; we have

$$\begin{aligned} I_i &= \int_{U_i} H_j(t)\phi(t)\mathbf{1}_{\{\xi(p)-u_n\leq g(t)\leq\xi(p)+u_n\}}\mathrm{d}t\\ &= \int_{g_i^{-1}(\xi(p)-u_n)}^{g_i^{-1}(\xi(p)+u_n)} H_j(t)\phi(t)\mathrm{d}t\\ &= \begin{cases}\phi(m_{i,n})-\phi(M_{i,n}) = (m_{i,n}-M_{i,n}) & \text{if } j=1\\ (-1)^j\left(\phi^{(j-1)}(M_{i,n})-\phi^{(j-1)}(m_{i,n})\right) & \text{if } j>1,\end{cases} \end{aligned}$$

where $M_{i,n} = g_i^{-1}(\xi(p)+u_n)$ and $m_{i,n} = g_i^{-1}(\xi(p)-u_n)$. Then, there exists $\omega_{n,i,j}\in[m_{i,n},M_{i,n}]$ for every $j\geq 1$ such that

$$I_i = \begin{cases}(m_{i,n}-M_{i,n})\phi^{(1)}(\omega_{n,i,1}) & \text{if } j=1\\ (-1)^j\,(M_{i,n}-m_{i,n})\,\phi^{(j-2)}(\omega_{n,i,j}) & \text{if } j>1.\end{cases}$$

Under Assumption $\boldsymbol{A(\xi(p))}$, we have, as $n \to +\infty$,

$$\omega_{n,i,j} \sim g_i^{-1}(\xi(p)) \quad \text{and} \quad M_{i,n} - m_{i,n} \sim 2u_n \frac{1}{g'(g_i^{-1}(\xi(p)))},$$

which ends the proof. ■

3.2. Proof of Theorem 1

For the sake of simplicity, we set $\widehat{\xi}(p) = \widehat{\xi}(p; \boldsymbol{g}(\boldsymbol{Y}))$, $\xi(p) = \xi_{g(Y)}(p)$, $\widehat{F}(\cdot) = \widehat{F}(\cdot; \boldsymbol{g}(\boldsymbol{Y}))$, $F(\cdot) = F_{g(Y)}(\cdot)$, $f(\cdot) = f_{g(Y)}(\cdot)$ and $r_n = r_n(\alpha, \overline{\tau}_p)$. Define

$$V_n = r_n^{-1}\left(\widehat{\xi}(p) - \xi(p)\right) \quad \text{and} \quad W_n = r_n^{-1}\left(\frac{p - \widehat{F}(p)}{f(p)}\right).$$

The result is established if $V_n - W_n \xrightarrow{\mathbb{P}} 0$ as $n \to +\infty$. It suffices to prove that V_n and W_n satisfy the conditions of Lemma 1 of Ghosh (1971):

- *Condition* (a): for all $\delta > 0$, there exists $\varepsilon = \varepsilon(\delta)$ such that $\mathbb{P}(|W_n| > \varepsilon) < \delta$.
- *Condition* (b): for all $y \in \mathbb{R}$ and for all $\varepsilon > 0$

$$\lim_{n\to+\infty} \mathbb{P}(V_n \le y, W_n \ge k + \varepsilon) \quad \text{and} \quad \lim_{n\to+\infty} \mathbb{P}(V_n \ge y + \varepsilon, W_n \ge k).$$

Condition (a): From the Bienaymé–Tchebyshev inequality it is sufficient to prove that $\mathbf{E}W_n^2 = \mathcal{O}(1)$. Rewrite $W_n = \frac{r_n^{-1}}{n}\sum_{i=1}^n h_{\xi(p)}(Y(i))$. Let c_j (for some $j \ge 0$) denote the j-th Hermite coefficient of $h_{\xi(p)}(\cdot)$. Since $h_{\xi(p)}(\cdot)$ has at least Hermite rank $\overline{\tau}_p$, then

$$\begin{aligned}
\mathbf{E}W_n^2 &= \frac{r_n^{-2}}{n^2}\sum_{i_1,i_2=1}^n \mathbf{E}\left(h_{\xi(p)}(Y(i_1))\, h_{\xi(p)}(Y(i_2))\right) \\
&= \frac{r_n^{-2}}{n^2}\sum_{i_1,i_2=1}^n \sum_{j_1,j_2\ge\tau_p} c_{j_1}c_{j_2}\mathbf{E}\left(H_{j_1}(Y(i_1))\, H_{j_2}(Y(i_2))\right) \\
&= \frac{r_n^{-2}}{n^2}\sum_{i_1,i_2=1}^n \sum_{j\ge\overline{\tau}_p} \frac{(c_j)^2}{(j)!}\,\rho(i_2 - i_1)^j \\
&= \mathcal{O}\left(r_n^{-2} \times \frac{1}{n}\sum_{|i|<n} \rho(i)^{\overline{\tau}_p}\right) = \mathcal{O}(1),
\end{aligned}$$

from Remark 1.

Condition (b): Let $y \in \mathbb{R}$; we have

$$\begin{aligned}
\{V_n \le y\} &= \left\{\widehat{\xi}(p) \le y \times r_n + \xi(p)\right\} \\
&= \left\{p \le \widehat{F}(y \times r_n + \xi(p))\right\} = \{Z_n \le y_n\},
\end{aligned} \tag{11}$$

with

$$Z_n = \frac{r_n^{-1}}{f(\xi(p))}\left(F(y \times r_n + \xi(p)) - \widehat{F}\left(\frac{y}{\sqrt{r_n}} + \xi(p)\right)\right)$$

and

$$y_n = \frac{r_n^{-1}}{f(\xi(p))}\left(F(y \times r_n + \xi(p)) - p\right).$$

Under Assumption $\boldsymbol{A(\xi(p))}$, we have $y_n \to y$, as $n \to +\infty$. Now, prove that $Z_n - W_n \xrightarrow{\mathbb{P}} 0$. Without loss of generality, assume $y > 0$. Then, we have

$$\begin{aligned}
W_n - Z_n &= \frac{r_n^{-1}}{f(\xi(p))}\left(\widehat{F}(y \times r_n + \xi(p)) - F(y \times r_n + \xi(p)) - \widehat{F}(\xi(p)) + F(\xi(p))\right) \\
&= \frac{r_n^{-1}}{n}\frac{1}{f(\xi(p))}\sum_{i=1}^n h_{\xi(p),n}(Y(i))
\end{aligned}$$

where $h_{\xi(p),n}(\cdot)$ is the function defined for $t \in \mathbb{R}$ by

$$h_{\xi(p),n}(t) = \mathbf{1}_{\{\xi(p)\le g(t)\le\xi(p)+y\times r_n\}}(t) - \mathbb{P}(\xi(p) \le g(Y) \le \xi(p) + y \times r_n).$$

For n sufficiently large, the function $h_{\xi(p),n}(\cdot)$ has Hermite rank $\overline{\tau}_p$. Denote by $c_{j,n}$ the j-th Hermite coefficient of $h_{\xi(p),n}(\cdot)$. From Lemma 3, there exists a sequence $(\kappa_j)_{j\geq\overline{\tau}_p}$ such that, as $n\to+\infty$,

$$c_{j,n}\sim\kappa_j\times r_n.$$

Since, for all $n\geq 1$, $\mathbf{E}(h_n(Y)^2)=\sum_{j\geq\overline{\tau}_p}(c_{j,n})^2/j!<+\infty$, it is clear that the sequence $(\kappa_j)_{j\geq\overline{\tau}_p}$ is such that $\sum_{j\geq\overline{\tau}_p}(\kappa_j)^2/j!<+\infty$. By denoting as λ a positive constant, we get, as $n\to+\infty$,

$$\begin{aligned}
\mathbf{E}(W_n-Z_n)^2 &= \frac{r_n^{-2}}{n^2}\frac{1}{f(\xi(p))^2}\sum_{i_1,i_2=1}^{n}\mathbf{E}\left(h_{\xi(p),n}\left(Y(i_1)\right)h_{\xi(p),n}\left(Y(i_2)\right)\right)\\
&= \frac{r_n^{-2}}{n^2}\frac{1}{f(\xi(p))^2}\sum_{i_1,i_2=1}^{n}\sum_{j_1,j_2\geq\overline{\tau}_p}c_{j_1,n}c_{j_2,n}\mathbf{E}\left(H_{j_1}\left(Y(i_1)\right)H_{j_2}\left(Y(i_2)\right)\right)\\
&= \frac{r_n^{-2}}{n^2}\frac{1}{f(\xi(p))^2}\sum_{i_1,i_2=1}^{n}\sum_{j\geq\overline{\tau}_p}\frac{c_{j,n}^2}{j!}\rho(i_2-i_1)^j\\
&\leq \lambda\frac{r_n^{-2}}{n}\sum_{j\geq\overline{\tau}_p}\frac{(\kappa_j)^2}{j!}r_n^2\sum_{|i|<n}\rho(i)^j=\mathcal{O}\left(\frac{1}{n}\sum_{|i|<n}\rho(i)^{\overline{\tau}_p}\right)=\mathcal{O}(r_n^2),
\end{aligned}$$

from Remark 1. Therefore, W_n-Z_n converges to 0 in probability, as $n\to+\infty$. Thus, for all $\varepsilon>0$, we have, as $n\to+\infty$,

$$\mathbb{P}\left(V_n\leq y,W_n\geq y+\varepsilon\right)=\mathbb{P}\left(Z_n\leq y_n,W_n\geq y+\varepsilon\right)\to 0.$$

Following the sketch of this proof, we also have $\mathbb{P}\left(V_n\geq y+\varepsilon,W_n\leq y\right)\to 0$, ensuring condition (b). Therefore, W_n-Z_n converges to 0 in probability, as $n\to+\infty$. Thus, for all $\varepsilon>0$, we have, as $n\to+\infty$,

$$\mathbb{P}\left(V_n\leq y,W_n\geq y+\varepsilon\right)=\mathbb{P}\left(Z_n\leq y_n,W_n\geq y+\varepsilon\right)\to 0.$$

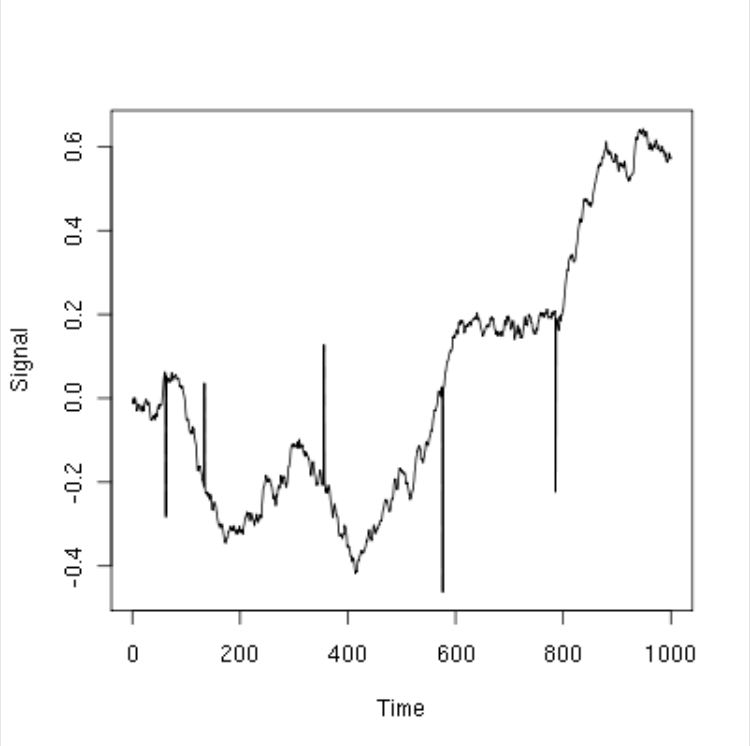
Following the sketch of this proof, we also have $\mathbb{P}\left(V_n\geq y+\varepsilon,W_n\leq y\right)\to 0$, ensuring condition (b).

Acknowledgement

The author would like to thank the referee for comments improving the statement of Corollary 2.

References

Bahadur, R.R., 1966. A note on quantiles in large samples. Ann. Math. Statist. 37, 577–580.
Breuer, P., Major, P., 1983. Central limit theorems for non-linear functionals of Gaussian fields. J. Multivariate Ann. 13, 425–441.
Coeurjolly, J.-F., 2008. Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. Ann. Statist. (in press). E-print arxiv: http://fr.arxiv.org/abs/math.ST/0506290, pp. 1–44.
Dacunha-Castelle, D., Duflo, M., 1982. Exercices de probabilitès statistiques. Tome 1 Probles temps fixe. In: Collection Mathathiques Appliqus pour la marise, Masson, Paris.
Dehling, H., Taqqu, M.S., 1989. The empirical process of some long-range dependent sequences with an application to U-statistics. Ann. Statist. 17, 1767–1783.
Ghosh, J.K., 1971. A new proof of the Bahadur Representation of Quantiles and an application. Ann. Math. Statist. 42, 1957–1961.
Hesse, C.H., 1990. A Bahadur-type representation for empirical quantiles of a large class of stationary, possibly infinite-variance, linear processes. Ann. Statist. 18, 1188–1202.
Ho, H.C., Hsing, T., 1996. On the asymptotic expansion of the empirical process of long-memory moving averages. Ann. Statist. 24, 992–1024.
Kiefer, J., 1967. On Bahadur's representation of sample quantiles. Ann. Math. Statist. 38, 1323–1342.
Sen, P.K., 1972. On the Bahadur representation of sample quantiles for sequences of ϕ-mixing random variables. J. Multivariate Anal. 2, 77–95.
Yoshihara, K.-I., 1995. The Bahadur representation of sample quantiles for sequences of strongly mixing random variables. Statist. Probab. Lett. 24, 299–304.
Wu, W.-B., 2005. On the Bahadur representation of sample quantiles for dependent sequences. Ann. Statist. 33, 1934–1963.



[Coe08b]
J.-F. Coeurjolly. Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. *Annals of Statistics*, 36(3):1404-1434, 2008.

The Annals of Statistics
2008, Vol. 36, No. 3, 1404–1434
DOI: 10.1214/009053607000000587

HURST EXPONENT ESTIMATION OF LOCALLY SELF-SIMILAR GAUSSIAN PROCESSES USING SAMPLE QUANTILES

BY JEAN-FRANÇOIS COEURJOLLY[1]

University of Grenoble 2

This paper is devoted to the introduction of a new class of consistent estimators of the fractal dimension of locally self-similar Gaussian processes. These estimators are based on convex combinations of sample quantiles of discrete variations of a sample path over a discrete grid of the interval $[0, 1]$. We derive the almost sure convergence and the asymptotic normality for these estimators. The key-ingredient is a Bahadur representation for sample quantiles of nonlinear functions of Gaussian sequences with correlation function decreasing as $k^{-\alpha}L(k)$ for some $\alpha > 0$ and some slowly varying function $L(\cdot)$.

1. Introduction. Many naturally occurring phenomena can be effectively modeled using self-similar processes. Among the simplest models, one can consider the fractional Brownian motion introduced in the statistics community by Mandelbrot and Van Ness [22]. Fractional Brownian motion can be defined as the only centered Gaussian process, denoted by $(X(t))_{t\in\mathbb{R}}$, with stationary increments and with variance function $v(\cdot)$, given by $v(t) = \sigma^2|t|^{2H}$, for all $t \in \mathbb{R}$. The fractional Brownian motion is an H-self-similar process, that is for all $c > 0$, $(X(ct))_{t\in\mathbb{R}} \overset{d}{=} c^H(X(t))_{t\in\mathbb{R}}$ (where $\overset{d}{=}$ means equal in finite-dimensional distributions) with autocovariance function behaving like $\mathcal{O}(|k|^{2H-2})$ as $|k| \to +\infty$. So the discretized increments of the fractional Brownian motion (called the fractional Gaussian noise) constitute a short-range dependent process, when $H < 1/2$, and a long-range dependent process, when $H > 1/2$. The index H also characterizes the path regularity since the fractal dimension of the fractional Brownian motion is equal to $D = 2 - H$. According to the context (long-range dependent processes, self-similar processes, ...), a very large variety of estimators of the parameter H has been investigated. The reader is referred to Beran [6], Coeurjolly [8] or Bardet et al. [5] for an overview of this problem. Among the most often used estimators we have: methods based on the variogram, on the log-periodogram, for example, Geweke and Porter-Hudak [15] in the context of long-range dependent processes,

Received June 2005; revised February 2007.
[1]Supported by a grant from IMAG Project AMOA.
AMS 2000 subject classifications. Primary 60G18; secondary 62G30.
Key words and phrases. Locally self-similar Gaussian process, fractional Brownian motion, Hurst exponent estimation, Bahadur representation of sample quantiles.

maximum likelihood estimator (and Whittle estimator) when the model is parametric, for example, fractional Gaussian noise, methods based on the wavelet decomposition, for example, Flandrin [14] or Stoev et al. [26] and the references therein, and on discrete filtering studied by Kent and Wood [19], Istas and Lang [18] and Coeurjolly [9]. We are mainly interested in the last one, which has several similarities with the wavelet decomposition method. Following Constantine and Hall [11], Kent and Wood [19], Istas and Lang [18], in the case when the process is observed at times i/n for $i = 1, \dots, n$, this method is adapted to a larger class than the fractional Brownian motion, namely, the class of centered Gaussian processes with stationary increments that are locally self-similar (at zero). A process $(X(t))_{t\in\mathbb{R}}$ is said to be locally self-similar (at zero) if its variance function, denoted by $v(\cdot)$, satisfies

$$v(t) = \mathbf{E}(X(t)^2) = \sigma^2|t|^{2H}\big(1 + r(t)\big) \qquad \text{with } r(t) = o(1) \text{ as } |t| \to 0, \tag{1}$$

for some $0 < H < 1$. An estimator of H is derived by using the stationarity of the increments and the local behavior of the variance function. When observing the process at regular subdivisions, the stationarity of the increments is crucial since the method based on discrete filtering (and the one based on the wavelet decomposition) essentially uses the fact that the variance of the increments can be estimated by the sample moment of order 2. We do not believe that this framework could be valid for the estimation of the Hurst exponent of Riemann–Liouville's process, for example, Alòs, Mazet and Nualart [1] which is an H-self-similar centered Gaussian process but with increments satisfying only some kind of local stationarity, see Remark 2 for more details.

Let us be more specific on the construction of the wavelet decomposition method (see, e.g., Flandrin [14]): the authors noticed that the variance of the wavelet coefficient at a scale say j behaves like $2^{j(2H-1)}$. An estimator of H is then derived by regressing the logarithm of sample moment of order 2 at each scale against $\log(j)$ for various scales. This procedure exhibits good properties since it is also proved that the more vanishing moments the wavelet has the observations are more decorrelated. And so asymptotic results are quite easy to obtain. However, Stoev et al. [26] illustrate the fact that this kind of estimator is very sensitive to additive outliers and to nonstationary artefacts. Therefore, they mainly propose to replace at each scale, the sample moment of order 2, by the sample median of the squared coefficients. This procedure, for which the authors assert that no theoretical result is available, is clearly more robust.

The main objective of this paper is to extend the procedure proposed by Stoev et al. [26] by deriving semiparametric estimators of the parameter H, using discrete filtering methods, for the class of processes defined by (1). The procedure is extended in the sense that we consider either convex combinations of sample quantiles or trimmed-means. Moreover, we provide convergence results. The key-ingredient is a Bahadur representation of sample quantiles obtained in a certain

dependence framework. Let $\mathbf{Y} = (Y(1), \ldots, Y(n))$ be a vector of n i.i.d. random variables with cumulative distribution function F, as well denote by $\xi(p)$ and $\widehat{\xi}(p)$ the quantile, respectively, the sample quantile of order p. By assuming that $F'(\xi(p)) > 0$ and $F''(\xi(p))$ exists, Bahadur proved that as $n \to +\infty$,

$$\widehat{\xi}(p) - \xi(p) = \frac{p - \widehat{F}(p)}{f(\xi(p))} + r_n,$$

with $r_n = \mathcal{O}_{a.s.}(n^{-3/4}\log(n)^{3/4})$. Using a law of iterated logarithm's type result, Kiefer [20] obtained the exact rate $n^{-3/4}\log\log(n)^{3/4}$. Extensions of the above results to dependent random variables have been pursued in Sen [24] for ϕ-mixing variables, in Yoshihara [29] for strongly mixing variables, and recently in Wu [30] for short-range and long-range dependent linear processes, following works of Hesse [16] and Ho and Hsing [17]. Our contribution is to provide a Bahadur representation for sample quantiles in another context that is for nonlinear functions of Gaussian processes with correlation function decreasing as $k^{-\alpha}L(k)$ for some $\alpha > 0$ and some slowly varying function $L(\cdot)$. The bounds for r_n are obtained under the same assumption as those used by Bahadur [4].

The paper is organized as follows. In Section 2, we give some basic notations and some background on discrete filtering. In Section 3, we derive semiparametric estimators of the parameter H, when a single sample path of a process defined by (1) is observed over a discrete grid of the interval [0, 1]. Section 4 presents the main results: Bahadur representations and asymptotic results for our estimators. In Section 5 some numerical computations are presented to compare the theoretical asymptotic variance of our estimators and a simulation study is also given. In particular, we illustrate the relative efficiency with respect to Whittle estimator and the fact that such estimators are more robust than classical ones. Finally, proofs of different results are presented in Section 6.

2. Some notations and some background on discrete filtering. Given some random variable Y, $F_Y(\cdot)$ denotes the cumulative distribution function of Y and $\xi_Y(p)$ the quantile of order p, $0 < p < 1$. If $F_Y(\cdot)$ is absolutely continuous with respect to Lebesgue measure, the probability density function is denoted by $f_Y(\cdot)$. The cumulative distribution (resp. probability density) function of a standard Gaussian variable is denoted by $\Phi(\cdot)$ [resp. $\phi(\cdot)$]. Based on the observation of a vector $\mathbf{Y} = (Y(1), \ldots, Y(n))$ of n random variables distributed as Y, the sample cumulative distribution function and the sample quantile of order p are respectively denoted by $\widehat{F}_Y(\cdot; \mathbf{Y})$ and $\widehat{\xi}_Y(p; \mathbf{Y})$ or simply by $\widehat{F}(\cdot; \mathbf{Y})$ and $\widehat{\xi}(p; \mathbf{Y})$. Finally, for some measurable function $g(\cdot)$, we denote by $\mathbf{g}(\mathbf{Y})$ the vector of length n with real components $g(Y(i))$, for $i = 1, \ldots, n$.

A sequence of real numbers u_n is said to be $\mathcal{O}(v_n)$ [resp. $o(v_n)$] for an other sequence of real numbers v_n, if u_n/v_n is bounded (resp. converges to 0 as $n \to +\infty$). A sequence of random variables U_n is said to be $\mathcal{O}_{a.s.}(v_n)$ [resp. $o_{a.s.}(v_n)$] if U_n/v_n is almost surely bounded (resp. if U_n/v_n converges toward 0 with probability 1).

The statistical model corresponds to a discretized version $\mathbf{X} = (X(i/n))_{i=1,\ldots,n}$ of a locally self-similar Gaussian process defined by (1).

One of the ideas of our method is to construct some estimators by using some properties of the variance of the increments of $\mathbf{X}$ or the variance of the increments of order 2 of $\mathbf{X}$. While considering the increments of $\mathbf{X}$ is conventional since the associated sequence is stationary, considering the increments of order 2 (or of a higher order) could be stranger. However, the main interest relies upon the fact that the observations of the latter resulting sequences are less correlated than those of the simple increments' sequence. All these vectors can actually be seen as special discrete filtering of the vector $\mathbf{X}$. Let us now specify some general background on discrete filtering and its consequence on the correlation structure. The vector $\mathbf{a}$ is a filter of length $\ell + 1$ and of order $\nu \geq 1$ with real components if

$$\sum_{q=0}^{\ell} q^j a_q = 0 \qquad \text{for } j = 0, \ldots, \nu - 1 \quad \text{and} \quad \sum_{q=0}^{\ell} q^\nu a_q \neq 0.$$

For example, $\mathbf{a} = (1, -1)$ [resp. $\mathbf{a} = (1, -2, 1)$] is a filter with order 1 (resp. 2). Let $\mathbf{X}^{\mathbf{a}}$ be the series obtained by filtering $\mathbf{X}$ with $\mathbf{a}$, then:

$$X^{\mathbf{a}}\left(\frac{i}{n}\right) = \sum_{q=0}^{\ell} a_q X\left(\frac{i-q}{n}\right) \qquad \text{for } i \geq \ell + 1.$$

Applying in turn the filter $\mathbf{a} = (1, -1)$ and $\mathbf{a} = (1, -2, 1)$ leads to the increments of $\mathbf{X}$, respectively the increments of $\mathbf{X}$ of order 2. One may also consider other filters such as Daubechies wavelet filters, for example, Daubechies [13].

The following assumption is needed by different results presented hereafter:

ASSUMPTION $A_1(k)$. For $i = 1, \ldots, k$

$$v^{(i)}(t) = \sigma^2 \beta(i) |t|^{2H-i} + o(|t|^{2H-i})$$

with $\beta(i) = 2H(2H-1)\cdots(2H-i+1)$ (where $k \geq 1$ is an integer).

This assumption assures that the variance function $v(\cdot)$ is sufficiently smooth around 0. It allows us to assert that the correlation structure of a locally self-similar discretized and filtered Gaussian process can be compared to the one of the fractional Brownian motion. This is announced more precisely in the following lemma.

LEMMA 1 (e.g., Kent and Wood [19]). *Let* $\mathbf{a}$ *and* $\mathbf{a}'$ *be two filters of length* $\ell + 1$ *and* $\ell' + 1$, *of order* ν *and* $\nu' \geq 1$. *Then we have*

$$\begin{aligned} \mathbf{E}\left(X^{\mathbf{a}}\left(\frac{i}{n}\right) X^{\mathbf{a}'}\left(\frac{i+j}{n}\right)\right) &= \frac{-\sigma^2}{2} \sum_{q,q'=0}^{\ell} a_q a'_{q'} v\left(\frac{q - q' + j}{n}\right) \\ &= \gamma_n^{\mathbf{a},\mathbf{a}'}(j)\left(1 + \delta_n^{\mathbf{a},\mathbf{a}'}(j)\right), \end{aligned} \tag{2}$$

with

$$\gamma_n^{\mathbf{a},\mathbf{a}'}(j) = \frac{\sigma^2}{n^{2H}}\gamma^{\mathbf{a},\mathbf{a}'}(j), \qquad \gamma^{\mathbf{a},\mathbf{a}'}(j) = -\tfrac{1}{2}\sum_{q,q'=0}^{\ell} a_q a'_{q'}|q-q'+j|^{2H} \tag{3}$$

and

$$\delta_n^{\mathbf{a},\mathbf{a}'}(j) = \frac{\sum_{q,q'} a_q a_{q'}|q-q'+j|^{2H} \times r((q-q'+j)/n)}{\gamma^{\mathbf{a},\mathbf{a}'}(j)}. \tag{4}$$

Moreover, as $|j| \to +\infty$,

$$\gamma^{\mathbf{a},\mathbf{a}'}(j) = \mathcal{O}\left(\frac{1}{|j|^{2H-\nu-\nu'}}\right). \tag{5}$$

Finally, under Assumption $A_1(\nu+\nu')$, *as* $n \to +\infty$

$$\delta_n^{\mathbf{a},\mathbf{a}'}(j) = o(1). \tag{6}$$

REMARK 1. In the case of the fractional Brownian motion the sequence δ_n is equal to 0, whereas it converges toward 0 for more general locally self-similar Gaussian processes, such as the Gaussian processes with stationary increments and with variance function $v(t) = 1 - \exp(-|t|^{2H})$ or $v(t) = \log(1+|t|^{2H})$ for which Assumption $A_1(k)$ is satisfied (for every $k \geq 1$).

REMARK 2. The stationarity of the increments and the local self-similarity required on the process $X(\cdot)$ are important, if the process is observed at times i/n for $i = 1, \ldots, n$. The crucial result of Lemma 1 is that the variance function of the filtered series behaves asymptotically as $\gamma_n^{\mathbf{a}}(0)$. It seems to be difficult to relax the constraint of stationarity. Consider for example the Riemann–Liouville's process, for example, Alòs, Mazet and Nualart [1]. This process is a Gaussian process which is H-self similar Gaussian but with increments satisfying only some kind of local stationarity. Following the computations of Lim [21], the variance of the increments' series of the Riemann–Liouville's process is equal to

$$\mathbf{E}\left(\left(X\left(\frac{i+1}{n}\right) - X\left(\frac{i}{n}\right)\right)^2\right) = \frac{1}{n^{2H}}\frac{1}{\Gamma(H+1/2)^2}\left\{I + \frac{1}{2H}\right\},$$

with $I = \int_0^i ((1+u)^{H-1/2} - u^{H-1/2})^3\,du + \int_0^{i/n} u^{2H-1}\,du$. This integral cannot be asymptotically independent of time. Note that this could be the case if the process is observed at irregular subdivisions. This question has not been investigated.

Define $\mathbf{Y}^{\mathbf{a}}$ as the normalized vector $\mathbf{X}^{\mathbf{a}}$ with variance 1. The covariance between $Y^{\mathbf{a}}(i/n)$ and $Y^{\mathbf{a}'}(i+j/n)$ is denoted by $\rho_n^{\mathbf{a},\mathbf{a}'}(j)$. Under Assumption $A_1(\nu+\nu')$, the following equivalence holds as $n \to +\infty$

$$\rho_n^{\mathbf{a},\mathbf{a}'}(j) \sim \rho^{\mathbf{a},\mathbf{a}'}(j) = \frac{\gamma^{\mathbf{a},\mathbf{a}'}(j)}{\sqrt{\gamma^{\mathbf{a},\mathbf{a}}(0)\gamma^{\mathbf{a}',\mathbf{a}'}(0)}}. \tag{7}$$

When $\mathbf{a} = \mathbf{a}'$, we set, for the sake of simplicity $\gamma_n^{\mathbf{a}}(\cdot) = \gamma_n^{\mathbf{a},\mathbf{a}}(\cdot)$, $\delta_n^{\mathbf{a}}(\cdot) = \delta_n^{\mathbf{a},\mathbf{a}}(\cdot)$, $\rho_n^{\mathbf{a}}(\cdot) = \rho_n^{\mathbf{a},\mathbf{a}}(\cdot)$, $\gamma^{\mathbf{a},\mathbf{a}}(\cdot) = \gamma^{\mathbf{a}}(\cdot)$ and $\rho^{\mathbf{a}}(\cdot) = \rho^{\mathbf{a},\mathbf{a}}(\cdot)$.

3. New estimators of H.

3.1. *Estimators based on a convex combination of sample quantiles.* Let $(\mathbf{p}, \mathbf{c}) = (p_k, c_k)_{k=1,\dots,K} \in ((0,1) \times \mathbb{R}^+)^K$ for an integer $1 \le K < +\infty$. Define the following statistics based on a convex combination of sample quantiles:

$$\widehat{\xi}(\mathbf{p}, \mathbf{c}; \mathbf{X}^{\mathbf{a}}) = \sum_{k=1}^{K} c_k \widehat{\xi}(p_k; \mathbf{X}^{\mathbf{a}}), \tag{8}$$

where $c_k, k = 1, \dots, K$ are positive real numbers such that $\sum_{k=1}^{K} c_k = 1$. For example, this corresponds to the sample median when $K = 1, \mathbf{p} = 1/2, \mathbf{c} = 1$, to a mean of quartiles when $K = 2, \mathbf{p} = (1/4, 3/4), \mathbf{c} = (1/2, 1/2)$. Consider the following computation: from Lemma 1, we have, as $n \to +\infty$,

$$\widehat{\xi}(\mathbf{p}, \mathbf{c}; \mathbf{X}^{\mathbf{a}}) \sim \frac{\sigma^2}{n^{2H}} \gamma^{\mathbf{a}}(0) \widehat{\xi}(\mathbf{p}, \mathbf{c}; \mathbf{Y}^{\mathbf{a}}).$$

REMARK 3. It may be expected that $\widehat{\xi}(\mathbf{p}, \mathbf{c}; \mathbf{Y}^{\mathbf{a}})$ converges toward a constant as $n \to +\infty$. In itself, this result is not interesting, since two parameters remain unknown: σ^2 and H and thus, it is impossible to derive an estimator of H.

Remark 3 suggests that we have to use at least two filters. Among all available filters, let us consider the sequence $(\mathbf{a}^m)_{m \ge 1}$ defined by

$$a_i^m = \begin{cases} a_j, & \text{if } i = jm, \\ 0, & \text{otherwise,} \end{cases} \qquad \text{for } i = 0, \dots, m\ell,$$

which is none other than the filter $\mathbf{a}$ dilated m times. For example, if the filter $\mathbf{a} = \mathbf{a}^1$ corresponds to the filter $(1, -2, 1)$, then $\mathbf{a}^2 = (1, 0, -2, 0, 1)$, $\mathbf{a}^3 = (1, 0, 0, -2, 0, 0, 1), \dots$. As noted by Kent and Wood [19] or Istas and Lang [18], the filter $\mathbf{a}^m$, of length $m\ell + 1$, is of order ν and has the following interesting property:

$$\gamma^{\mathbf{a}^m}(0) = m^{2H} \gamma^{\mathbf{a}}(0). \tag{9}$$

From Lemma 1, this simply means that $\mathbf{E}(\mathbf{X}^{\mathbf{a}^m}(i/n)^2) = m^{2H} \mathbf{E}(\mathbf{X}^{\mathbf{a}}(i/n)^2)$, exhibiting some kind of self-similarity property of the filtered coefficients. As specified in the Introduction, the same property can be pointed out in the context of wavelet decomposition.

Our methods, that exploit the nice property (9), are based on a convex combination of sample quantiles $\widehat{\xi}(\mathbf{p}, \mathbf{c}; \mathbf{g}(\mathbf{X}^{\mathbf{a}^m}))$ for two positive functions $g(\cdot)$: $g(\cdot) = |\cdot|^\alpha$ for $\alpha > 0$ and $g(\cdot) = \log|\cdot|$. For the sake of conciseness of the paper, we only

present version of estimators with $g(\cdot) = |\cdot|^\alpha$ and refer the reader to Coeurjolly [10] for more details. For such functions $g(\cdot)$ we manage, by using some property established in Lemma 1, to define some very simple estimators of the Hurst exponent through a simple linear regression. Other choices of the function $g(\cdot)$ have not been investigated in this paper. At this stage, let us specify that our methods extend the one proposed by Stoev et al. [26]; indeed they only consider the statistic $\widehat{\xi}(\mathbf{p}, \mathbf{c}; \mathbf{g}(\mathbf{X}^{\mathbf{a}^m}))$ for $\mathbf{p} = 1/2$, $\mathbf{c} = 1$, $g(\cdot) = (\cdot)^2$, that is the sample median of the squared coefficients. From (3) and (9), we have

$$\begin{aligned}\widehat{\xi}(\mathbf{p}, \mathbf{c}; |\mathbf{X}^{\mathbf{a}^m}|^{\boldsymbol{\alpha}}) &= \mathbf{E}\big((X^{\mathbf{a}^m}(1/n))^2\big)^{\alpha/2}\widehat{\xi}(\mathbf{p}, \mathbf{c}; |\mathbf{Y}^{\mathbf{a}^m}|^{\boldsymbol{\alpha}}) \\ &= m^{\alpha H}\frac{\sigma^\alpha}{n^{\alpha H}}\gamma^{\mathbf{a}}(0)^{\alpha/2}\big(1 + \delta_n^{a^m}(0)\big)^{\alpha/2}\widehat{\xi}(\mathbf{p}, \mathbf{c}; |\mathbf{Y}^{\mathbf{a}^m}|^{\boldsymbol{\alpha}}).\end{aligned} \tag{10}$$

Denote by $\kappa_H = n^{-2H}\sigma^2\gamma^{\mathbf{a}}(0)$. Equation (10) can be rewritten as

$$\log\widehat{\xi}(\mathbf{p}, \mathbf{c}; |\mathbf{X}^{\mathbf{a}^m}|^{\boldsymbol{\alpha}}) = \alpha H\log(m) + \log\big(\kappa_H^{\alpha/2}\xi_{|Y|^\alpha}(\mathbf{p}, \mathbf{c})\big) + \varepsilon_m^\alpha, \tag{11}$$

with the random variables ε_m^α defined by

$$\varepsilon_m^\alpha = \log\left(\frac{\widehat{\xi}(\mathbf{p}, \mathbf{c}; |\mathbf{Y}^{\mathbf{a}^m}|^{\boldsymbol{\alpha}})}{\xi_{|Y|^\alpha}(\mathbf{p}, \mathbf{c})}\right) + \frac{\alpha}{2}\log\big(1 + \delta_n^{a^m}(0)\big), \tag{12}$$

where, for some random variable Z, $\xi_Z(\mathbf{p}, \mathbf{c}) = \sum_{k=1}^K c_k\xi_Z(p_k)$. We decide to rewrite equation (10) as (11), since we expect that ε_m^α converges (almost surely) toward 0 as $n \to +\infty$. From Remark 3, an estimator of H can be defined through a linear regression of $(\log\widehat{\xi}(\mathbf{p}, \mathbf{c}; |\mathbf{X}^{\mathbf{a}^m}|^\alpha))_{m=1,\dots,M}$ on $(\log m)_{m=1,\dots,M}$ for some $M \geq 2$. This estimator is denoted by $\widehat{H}^\alpha$. By denoting $\mathbf{A}$ the vector of length M with components $A_m = \log m - \frac{1}{M}\sum_{m=1}^M \log(m)$, $m = 1, \dots, M$, we have explicitly from (11) and the definition of least squares estimates (see, e.g., Antoniadis et al. [2]):

$$\widehat{H}^\alpha = \frac{\mathbf{A}^T}{\alpha\|\mathbf{A}\|^2}(\log\widehat{\xi}(\mathbf{p}, \mathbf{c}; |\mathbf{X}^{\mathbf{a}^m}|^{\boldsymbol{\alpha}}))_{m=1,\dots,M}, \tag{13}$$

where $\|\mathbf{z}\|$ for some vector $\mathbf{z}$ of length d denotes the norm defined by $(\sum_{i=1}^d z_i^2)^{1/2}$.

We can point out that $\widehat{H}^\alpha$ is independent of the scaling coefficient σ^2.

3.2. *Estimators based on trimmed means.* Let $0 < \beta_1 \leq \beta_2 < 1$ and $\boldsymbol{\beta} = (\beta_1, \beta_2)$, denote by $\overline{\mathbf{g}(\mathbf{X}^{\mathbf{a}})}^{(\boldsymbol{\beta})}$ the $\boldsymbol{\beta}$-trimmed mean of the vector $\mathbf{g}(\mathbf{X}^{\mathbf{a}})$ given by

$$\overline{\mathbf{g}(\mathbf{X}^{\mathbf{a}})}^{(\boldsymbol{\beta})} = \frac{1}{n - [n\beta_2] - [n\beta_1]}\sum_{[n\beta_1]+1}^{n-[n\beta_2]}(\mathbf{g}(\mathbf{X}^{\mathbf{a}}))_{(i),n},$$

where $(\mathbf{g}(\mathbf{X^a}))_{(1),n} \le (\mathbf{g}(\mathbf{X^a}))_{(2),n} \le \cdots \le (\mathbf{g}(\mathbf{X^a}))_{(2),n}$ are the order statistics of $(\mathbf{g}(\mathbf{X^a}))_1, \ldots, (\mathbf{g}(\mathbf{X^a}))_n$. It is well known that $(\mathbf{g}(\mathbf{X^a}))_{(i),n} = \widehat{\xi}(\frac{i}{n}; \mathbf{g}(\mathbf{X^a}))$. Hence, by following the ideas of the previous section, one may obtain

$$\log\big(\overline{|\mathbf{X^{a^m}}|^{\boldsymbol{\alpha}}}^{(\boldsymbol{\beta})}\big) = \alpha H \log(m) + \log\big(\kappa_H^{\alpha/2}\overline{|Y|^{\alpha}}^{(\boldsymbol{\beta})}\big) + \varepsilon_m^{\alpha,tm}, \tag{14}$$

with

$$\varepsilon_m^{\alpha,tm} = \overline{|\mathbf{Y^{a^m}}|^{\boldsymbol{\alpha}}}^{(\boldsymbol{\beta})} - \overline{|Y|^{\alpha}}^{(\boldsymbol{\beta})} + \frac{\alpha}{2}\log\big(1 + \delta_n^{a^m}(0)\big), \tag{15}$$

where for some random variable Z, $\overline{Z}^{(\boldsymbol{\beta})}$ is referring to

$$\overline{Z}^{(\boldsymbol{\beta})} = \frac{1}{1-\beta_2-\beta_1}\int_{\beta_1}^{1-\beta_2} \xi_Z(p)\,dp. \tag{16}$$

As in the previous section, an estimator of H, denoted by $\widehat{H}^{\alpha,tm}$, is derived through a log-linear regression

$$\widehat{H}^{\alpha,tm} = \frac{\mathbf{A}^T}{\alpha\|\mathbf{A}\|^2}\big(\overline{|\mathbf{X^{a^m}}|^{\boldsymbol{\alpha}}}^{(\boldsymbol{\beta})}\big)_{m=1,\ldots,M}. \tag{17}$$

REMARK 4. The estimator referred to the "estimator based on the quadratic variations" in the simulation study and studied with the same formalizm by Coeurjolly [9] corresponds to the estimator $\widehat{H}^{\alpha,tm}$ with $\alpha = 2$, $\beta_1 = \beta_2 = 0$.

4. Main results. To simplify the presentation of different results, consider the two following assumptions on different parameters involved in the estimation procedures.

ASSUMPTION $A_2(p, c)$. $\mathbf{a}$ is a filter of order $\nu \ge 1$, α is a positive real number, $\mathbf{p}$ (resp. $\mathbf{c}$) is a vector of length K (for some $1 \le K < +\infty$) such that $0 < p_k < 1$ (resp. $c_k > 0$ and $\sum_{k=1}^K c_k = 1$), M is an integer ≥ 2.

ASSUMPTION $A_3(\beta)$. $\mathbf{a}$ is a filter of order $\nu \ge 1$, α is a positive real number, $\beta = (\beta_1, \beta_2)$ is such that $0 < \beta_1 \le \beta_2 < 1$, M is an integer ≥ 2.

Since $A^T(\log(m))_{m=1,\ldots,M} = \|A\|^2$ and $A^T\mathbf{1} = 0$ [where $\mathbf{1} = (1)_{m=1,\ldots,M}$], we have

$$\widehat{H}^{\alpha} - H = \frac{\mathbf{A}^T}{\alpha\|\mathbf{A}\|^2}\boldsymbol{\varepsilon}^{\boldsymbol{\alpha}} \quad \text{and} \quad \widehat{H}^{\alpha,tm} - H = \frac{\mathbf{A}^T}{\alpha\|\mathbf{A}\|^2}\boldsymbol{\varepsilon}^{\boldsymbol{\alpha},\mathbf{tm}}, \tag{18}$$

where $\boldsymbol{\varepsilon}^{\boldsymbol{\alpha}} = (\varepsilon_m^{\alpha})_{m=1,\ldots,M}$ and $\boldsymbol{\varepsilon}^{\boldsymbol{\alpha},\mathbf{tm}} = (\varepsilon_m^{\alpha,tm})_{m=1,\ldots,M}$. Hence, in order to study the convergence of different estimators, it is sufficient to obtain some convergence results of sample quantiles $\widehat{\xi}(p, \mathbf{g}(\mathbf{Y^a}))$ for some function $g(\cdot)$ and some filter $\mathbf{a}$. Therefore, we first establish a Bahadur representation of sample quantiles for some

nonlinear function of Gaussian sequences with correlation function decreasing as $k^{-\alpha}$, for some $\alpha > 0$. In fact, the existing literature on nonlinear function of Gaussian sequences (e.g., Taqqu [27]) allows us to slightly extend this framework by considering correlation function decreasing as $k^{-\alpha}L(k)$, for some slowly varying function $L(\cdot)$.

4.1. *Bahadur representation of sample quantiles.* Let us recall some important definitions on Hermite polynomials. The jth Hermite polynomial (for $j \geq 0$) is defined for $t \in \mathbb{R}$ by

$$H_j(t) = \frac{(-1)^j}{\phi(t)} \frac{d^j \phi(t)}{dt^j}. \tag{19}$$

The Hermite polynomials form an orthogonal system for the Gaussian measure. More precisely, we have $\mathbf{E}(H_j(Y)H_k(Y)) = j!\delta_{j,k}$. For a measurable function $g(\cdot)$ defined on $\mathbb{R}$ for which $\mathbf{E}(g(Y)^2) < +\infty$, the following expansion holds:

$$g(t) = \sum_{j \geq \tau} \frac{c_j}{j!} H_j(t) \qquad \text{with } c_j = \mathbf{E}(g(Y)H_j(Y)),$$

where the integer τ defined by $\tau = \inf\{j \geq 0, c_j \neq 0\}$, is called the Hermite rank of the function g. Note that this integer plays an important role. For example, it is related to the correlation of $g(Y_1)$ and $g(Y_2)$ (for Y_1 and Y_2 two standard Gaussian variables with correlation ρ) since $\mathbf{E}(g(Y_1)g(Y_2)) = \sum_{k \geq \tau} \frac{(c_k)^2}{k!}\rho^k \leq \rho^\tau \|g\|_{L^2(d\phi)}$.

In order to obtain a Bahadur representation (see, e.g., Serfling [25]), we have to ensure that $F'_{g(Y)}(\xi(p)) > 0$ and $F''_{g(Y)}(\cdot)$ exists and is bounded in a neighborhood of $\xi(p)$. This is achieved if the function $g(\cdot)$ satisfies the following assumption (see, e.g., Dacunha-Castelle and Duflo [12], page 33).

ASSUMPTION $A_4(\xi(p))$. There exist U_i, $i = 1, \ldots, L$, disjoint open sets such that U_i contains a unique solution to the equation $g(t) = \xi_{g(Y)}(p)$, such that $F'_{g(Y)}(\xi(p)) > 0$ and such that g is a $\mathcal{C}^2$-diffeomorphism on $\bigcup_{i=1}^L U_i$.

Note that under this assumption

$$F'_{g(Y)}\big(\xi_{g(Y)}(p)\big) = f_{g(Y)}\big(\xi_{g(Y)}(p)\big) = \sum_{i=1}^{L} \frac{\phi(g_i^{-1}(\xi(p)))}{g'(g_i^{-1}(\xi(p)))},$$

where $g_i(\cdot)$ is the restriction of $g(\cdot)$ on U_i. Now, define, for some real u, the function $h_u(\cdot)$ by

$$h_u(t) = \mathbf{1}_{\{g(t) \leq u\}}(t) - F_{g(Y)}(u). \tag{20}$$

We denote by $\tau(u)$ the Hermite rank of $h_u(\cdot)$. For the sake of simplicity, we set $\tau_p = \tau(\xi_{g(Y)}(p))$. For some function $g(\cdot)$ satisfying Assumption $A_4(\xi(p))$, we denote by

$$\overline{\tau}_p = \inf_{\gamma \in \bigcup_{i=1}^L g(U_i)} \tau(\gamma), \tag{21}$$

that is the minimal Hermite rank of $h_u(\cdot)$ for u in a neighborhood of $\xi_{g(Y)}(p)$.

THEOREM 2. *Let $\{Y(i)\}_{i=1}^{+\infty}$ be a stationary (centered) Gaussian process with variance 1, and correlation function $\rho(\cdot)$ such that, as $i \to +\infty$*

$$|\rho(i)| \sim L(i)i^{-\alpha}, \tag{22}$$

for some $\alpha > 0$ and some slowly varying function at infinity $L(s), s \geq 0$. Then, under Assumption $A_4(\xi(p))$, we have almost surely, as $n \to +\infty$

$$\widehat{\xi}(p; \mathbf{g}(\mathbf{Y})) - \xi_{g(Y)}(p) = \frac{p - \widehat{F}(\xi_{g(Y)}(p); \mathbf{g}(\mathbf{Y}))}{f_{g(Y)}(\xi_{g(Y)}(p))} + \mathcal{O}_{a.s.}(r_n(\alpha, \overline{\tau}_p)), \tag{23}$$

the sequence $(r_n(\alpha, \overline{\tau}_p))_{n\geq 1}$ being defined by

$$r_n(\alpha, \overline{\tau}_p) = \begin{cases} n^{-3/4}\log(n)^{3/4}, & \text{if } \alpha\overline{\tau}_p > 1, \\ n^{-3/4}\log(n)^{3/4}L_{\overline{\tau}_p}(n)^{3/4}, & \text{if } \alpha\overline{\tau}_p = 1, \\ n^{-1/2-\alpha\overline{\tau}_p/4}\log(n)^{\overline{\tau}_p/4+1/2}L(n)^{\overline{\tau}_p/4}, & \text{if } 2/3 < \alpha\overline{\tau}_p < 1, \\ n^{-\alpha\overline{\tau}_p}\log(n)^{\overline{\tau}_p}L(n)^{\overline{\tau}_p}, & \text{if } 0 < \alpha\overline{\tau}_p \leq 2/3, \end{cases} \tag{24}$$

where for some $\tau \geq 1$, $L_\tau(n) = \sum_{|i|\leq n} |\rho(i)|^\tau$.

Note that if $L(\cdot)$ is an increasing function, $L_\tau(n) = \mathcal{O}(\log(n)L(n)^\tau)$.

REMARK 5. Without giving any details here, let us say that the behavior of the sequence $r_n(\cdot,\cdot)$ is related to the characteristic (short-range or long-range dependence) of the process $\{h_u(Y(i))\}_{i=1}^{+\infty}$ for u in a neighborhood of $\xi_{g(Y)}(p)$. In the case $\alpha\overline{\tau}_p > 1$, corresponding to short-range dependent processes, the result is similar to the one proved by Bahadur, see, for example, Serfling [25], in the i.i.d. case. For short-range dependent linear processes, using a law of iterated logarithm's type result Wu [30] obtained a sharper bound, that is $n^{-3/4}\log\log(n)^{3/4}$. This bound is obtained under the assumption that $F'(\cdot)$ and $F''(\cdot)$ exist and are uniformly bounded. For long-range dependent processes ($\alpha\overline{\tau}_p \leq 1$), we can observe that the rate of convergence is always lower than $n^{-3/4}\log(n)^{3/4}$ and that the dominant term $n^{-3/4}$ is obtained when $\alpha\overline{\tau}_p \to 1$.

We now propose a uniform Bahadur type representation of sample quantiles. Such a representation has an application in the study of trimmed-mean. For $0 < p_0 \leq p_1 < 1$ consider the following assumption which extends Assumption $A_4(\xi(p))$.

ASSUMPTION $A_5(p_0, p_1)$. There exists U_i, $i = 1, \ldots, L$, disjoint open sets such that U_i contains a solution to the equation $g(t) = \xi_{g(Y)}(p)$ for all $p_0 \leq$

$p \le p_1$, such that $F'_{g(Y)}(\xi(p)) > 0$ for all $p_0 \le p \le p_1$ and such that g is a $\mathcal{C}^2$-diffeomorphism on $\bigcup_{i=1}^{L} U_i$.

Under the previous assumption, define

$$\tau_{p_0,p_1} = \inf_{\gamma \in \bigcup_{i=1}^{L} g(U_i)} \tau(\gamma). \tag{25}$$

THEOREM 3. *Under the conditions of Theorem 2 and Assumption $A_5(p_0, p_1)$, we have almost surely, as $n \to +\infty$*

$$\begin{aligned} &\sup_{p_0 \le p \le p_1} \left| \widehat{\xi}(p; \mathbf{g}(\mathbf{Y})) - \xi_{g(Y)}(p) - \frac{p - \widehat{F}(\xi_{g(Y)}(p); \mathbf{g}(\mathbf{Y}))}{f_{g(Y)}(\xi_{g(Y)}(p))} \right| \\ &\qquad = \mathcal{O}_{a.s.}(r_n(\alpha, \tau_{p_0,p_1})). \end{aligned} \tag{26}$$

REMARK 6. To obtain convergence results of estimators of H, some results are needed concerning sample quantiles of the form $\widehat{\xi}(p; \mathbf{g}(\mathbf{Y}^{\mathbf{a}^{\mathbf{m}}}))$, with $g(\cdot) = |\cdot|$. Lemma 13 asserts that the Hermite rank τ_p of the function $h_{\xi_{g(Y)}(p)}(\cdot)$ with $g(\cdot) = |\cdot|$, is equal to 2 for all $0 < p < 1$. Moreover, for all $0 < p < 1$ and for all $0 < p_0 \le p_1 < 1$, Assumptions $A_4(\xi(p))$ and $A_5(p_0, p_1)$ are satisfied, and we have $\overline{\tau}_p = \tau_{p_0,p_1} = 2$. Since from Lemma 1, the correlation function of $\mathbf{Y}^{\mathbf{a}^{\mathbf{m}}}$ satisfies (22) with $\alpha = 2\nu - 2H$ and $L(\cdot) = 1$, by applying Theorem 2, the sequence $r_n(\cdot, \cdot)$ is then given by

$$r_n(2\nu - 2H, 2) = n^{-3/4} \log(n)^{3/4}, \qquad \text{if } \nu \ge 2 \tag{27}$$

and for $\nu = 1$

$$r_n(2 - 2H, 2) = \begin{cases} n^{-3/4} \log(n)^{3/4}, & \text{if } 0 < H < 3/4, \\ n^{-3/4} \log(n)^{3/2}, & \text{if } H = 3/4, \\ n^{-1/2-(1-H)} \log(n), & \text{if } 3/4 < H < 5/6, \\ n^{-2(2-2H)} \log(n)^2, & \text{if } 5/6 \le H < 1. \end{cases} \tag{28}$$

4.2. *Convergence results of estimators of H.* In order to specify convergence results, we make the following assumption concerning the remainder term of the variance function $v(\cdot)$.

ASSUMPTION $A_6(\eta)$. There exists $\eta > 0$ such that $v(t) = \sigma^2 |t|^{2H}(1 + \mathcal{O}(|t|^{\eta}))$, as $|t| \to 0$.

The first result concentrates itself on estimators $\widehat{H}^{\alpha}$ based on a convex combination of sample quantiles.

THEOREM 4. *Under Assumptions $A_1(2\nu)$, $A_2(p, c)$ and $A_6(\eta)$:*

(i) *we have almost surely, as* $n \to +\infty$,

$$\widehat{H}^{\alpha} - H = \begin{cases} \mathcal{O}(n^{-\eta}) + \mathcal{O}_{a.s.}(n^{-1/2}\log(n)), & \text{if } \nu > H + \frac{1}{4}, \\ \mathcal{O}(n^{-\eta}) + \mathcal{O}_{a.s.}(n^{-1/2}\log(n)^{3/2}), & \text{if } \nu = 1,\ H = \frac{3}{4}, \\ \mathcal{O}(n^{-\eta}) + \mathcal{O}_{a.s.}\big(n^{-2(1-H)}\log(n)\big), & \text{if } \nu = 1,\ \frac{3}{4} < H < 1. \end{cases} \tag{29}$$

(ii) *The mean squared errors (MSE) of* $\widehat{H}^{\alpha}$ *satisfies*

$$MSE(\widehat{H}^{\alpha} - H) = \mathcal{O}\big(v_n(2\nu - 2H)\big) + \mathcal{O}\big(r_n(2\nu - 2H, 2)^2\big) + \mathcal{O}(n^{-2\eta}). \tag{30}$$

The sequence $r_n(2\nu - 2H, 2)$ *is given by* (27) *and* (28) *and the sequence* $v_n(\cdot)$ *is defined by*

$$v_n(2\nu - 2H) = \begin{cases} n^{-1}, & \text{if } \nu > H + \frac{1}{4}, \\ n^{-1}\log(n), & \text{if } \nu = 1,\ H = \frac{3}{4}, \\ n^{-4(1-H)}, & \text{if } \nu = 1,\ \frac{3}{4} < H < 1. \end{cases} \tag{31}$$

(iii) *If the filter* $\mathbf{a}$ *is such that* $\nu > H + 1/4$, *and if* $\eta > 1/2$, *then we have the following convergence in distribution, as* $n \to +\infty$,

$$\sqrt{n}(\widehat{H}^{\alpha} - H) \longrightarrow \mathcal{N}(0, \sigma_{\alpha}^2), \tag{32}$$

where σ_{α}^2 *is defined for* $\alpha \geq 0$ *by*

$$\sigma_{\alpha}^2 = \sum_{i \in \mathbb{Z}} \sum_{j \geq 1} \frac{1}{(2j)!} \left(\sum_{k=1}^{K} \frac{H_{2j-1}(q_k) c_k}{q_k} \pi_k^{\alpha} \right)^2 \mathbf{B}^T \underline{\mathbf{R}}(i, j) \mathbf{B}. \tag{33}$$

The vector $\mathbf{B}$ *is defined by* $\mathbf{B} = \frac{\mathbf{A}^T}{\|\mathbf{A}\|^2}$, *and the real numbers* q_k *and* π_k^{α} *are defined by*

$$q_k = \Phi^{-1}\left(\frac{1 + p_k}{2}\right) \quad \text{and} \quad \pi_k^{\alpha} = \frac{(q_k)^{\alpha}}{\sum_{j=1}^{K} c_j (q_j)^{\alpha}}. \tag{34}$$

Finally, the matrix $\underline{\mathbf{R}}(i, j)$, *defined for* $i \in \mathbb{Z}$ *and* $j \geq 1$, *is a* $M \times M$ *matrix whose* (m_1, m_2) *entry is*

$$(\underline{\mathbf{R}}(i, j))_{m_1, m_2} = \rho^{\mathbf{a^{m_1}}, \mathbf{a^{m_2}}}(i)^{2j}, \tag{35}$$

where $\rho^{\mathbf{a^{m_1}}, \mathbf{a^{m_2}}}(\cdot)$ *is the correlation function defined by* (7).

REMARK 7. The expression of the variance σ_{α}^2 given by (33) could appear to be very complicated. However, given some vectors $\mathbf{p}$ and $\mathbf{c}$ and some integer M, it does not take unreasonable effort to compute it for each value of H by truncating the two series. This issue is investigated in Section 5 to compare the different parameters.

REMARK 8. Let us discuss the result (30). The first term, $\mathcal{O}(v_n)$, is due to the variance of the sample cumulative distribution function. The second term, $\mathcal{O}(r_n^2)$ is due to the departure of $\widehat{\xi}(p) - \xi(p)$ from $\widehat{F}(\xi(p)) - p$. We leave the reader to check that

$$\mathcal{O}\big(r_n(2\nu - 2H, 2)^2\big) + \mathcal{O}\big(v_n(2\nu - 2H)\big) = \begin{cases} \mathcal{O}\big(v_n(2\nu - 2H)\big), & \text{if } \nu \geq H + \frac{1}{4}, \\ \mathcal{O}\big(r_n(2\nu - 2H, 2)^2\big), & \text{if } \nu < H + \frac{1}{4}. \end{cases}$$

Finally, the third one, $\mathcal{O}(n^{-2\eta})$ is a bias term due to the misspecification of the variance function $v(\cdot)$ around 0.

REMARK 9. If $K = 1$, we have, for every $\alpha > 0$,

$$\sigma_\alpha^2 = \sigma_0^2 = \sum_{i\in\mathbb{Z}} \sum_{j\geq 1} \frac{H_{2j-1}(q)^2}{q^2(2j)!} \mathbf{B}^T \underline{\mathbf{R}}(i, j)\mathbf{B}.$$

Assume Assumption $A_6(\eta)$ with $\eta > 1/2$ which allows to neglict the bias term with respect to the variance one. The result (32) is proved by using some general central limit theorem obtained in this dependence context by Arcones [3], which is available as soon as $\rho^{\mathbf{a}}(\cdot)^2$ is summable. Therefore, if only Assumption $A_1(2)$ is assumed, the filter $\mathbf{a}$ cannot exceed 1 [and then correspond to $\mathbf{a} = (1, -1)$] and, due to (5), the result (32) is valid only for $0 < H < 3/4$. As a practical point of view, one observes that for such a filter and large values of H, the estimators have very big variance. Note that if Assumption $A_1(2\nu)$ can be assumed for $\nu > 1$, then the asymptotic normality is valid for all the values of H.

The following theorem presents the analog results obtained for the estimator $\widehat{H}^{\alpha,tm}$ based on trimmed-means.

THEOREM 5. *Under Assumptions $A_1(2\nu)$, $A_3(\beta)$ and $A_6(\eta)$, properties* (i) *and* (ii) *of Theorem 4 hold for the estimator $\widehat{H}^{\alpha,tm}$ with the same rates of convergences.*

(iii) *if the filter* $\mathbf{a}$ *is such that $\nu > H + 1/4$ and if $\eta > 1/2$, then, under the notations of Theorem 4, we have the following convergence in distribution, as $n \to +\infty$*

$$\sqrt{n}(\widehat{H}^{\alpha,tm} - H) \longrightarrow \mathcal{N}(0, \sigma_{\alpha,tm}^2), \tag{36}$$

where $\sigma_{\alpha,tm}^2$ is defined for $\alpha \geq 0$ by

$$\sigma_{\alpha,tm}^2 = \sum_{i\in\mathbb{Z}} \sum_{j\geq 1} \frac{1}{(2j)!} \left(\frac{\int_{\beta_1}^{1-\beta_2} H_{2j-1}(q) q^{\alpha-1}\, dp}{\int_{\beta_1}^{1-\beta_2} q^\alpha\, dp} \right)^2 \mathbf{B}^T \underline{\mathbf{R}}(i, j)\mathbf{B}, \tag{37}$$

with $q = \Phi^{-1}(\frac{1+p}{2})$.

5. Numerical computation and simulations.

5.1. *Asymptotic constants σ_α^2 and $\sigma_{\alpha,tm}^2$.* In order to compare the different estimators, we intend to compute the asymptotic constants σ_α^2 and $\sigma_{\alpha,tm}^2$ defined by (33) and (37) for various set of parameters $(\mathbf{a}, \mathbf{p}, \mathbf{c}, \boldsymbol{\beta}, M)$. For this work, both series defining σ_α^2 and $\sigma_{\alpha,tm}^2$ are truncated ($|i| \leq 200$, $j \leq 150$). Figure 2 illustrates a part of this work. We can propose the following general remarks:

- Among all filters tested, the best one seems to be

$$\mathbf{a}^\star = \begin{cases} inc1, & \text{if } 0 < H < 3/4, \\ db4, & \text{otherwise,} \end{cases}$$

  where $inc1$ and $db4$ respectively denote the filter $(1, -1)$ and the Daubechies wavelet filter with two zero moments explicitly given by

$$db4 = (0.4829629, -0.8365763, 0.22414386, 0.12940952).$$

- Choice of M: increasing M seems to reduce the asymptotic constant σ_α^2. Obviously, a too large M increases the bias since $\widehat{\xi}(\mathbf{p}, \mathbf{c}; \mathbf{g}(\mathbf{X}^{\mathbf{a}^\mathbf{M}}))$ or $\overline{\mathbf{g}(\mathbf{X}^{\mathbf{a}^\mathbf{M}})}^{(\boldsymbol{\beta})}$ are estimated with $N - M\ell$ observations. We recommend setting it to the value 5.
- We did not manage [theoretically and numerically since series defining (33) and (37) are truncated] to determine the optimal value of α. However, for examples considered, it should be near the value 2.
- Again, this is quite difficult to know theoretically and numerically which choice of $\mathbf{p}$ is optimal. What we observed is that, for fixed parameters $\mathbf{a}$, M and α, the asymptotic constants are very close to each other.
- Choice of p in the case of a single quantile (see Figure 2): the optimal p seems to be near the value 90%. However, $p = 1/2$, corresponding to the estimator based on the median, leads to good results.
- Choice of $\beta_1 = \beta_2 = \beta$ for the estimators based on trimmed-means (see Figure 2): obviously the constant grows with β but we can point out that estimators based on 10%-trimmed-means are very competitive with the ones obtained by quadratic variations ($\beta = 0$).

5.2. *Simulation.* A short simulation study is proposed in Table 1 and Figure 1 for $n = 1000$ and $H = 0.8$. We consider two locally self-similar Gaussian processes whose variance functions are in turn $v(t) = |t|^{2H}$ (fractional Brownian motion) and $v(t) = 1 - \exp(-|t|^{2H})$. To generate sample paths discretized over a grid $[0, 1]$, we use the method of circulant matrix (see Wood and Chan [28]), which is particularly fast, even for large sample sizes. Various versions of estimators are considered and compared with classical ones, that is the one based on quadratic variations, Coeurjolly [9], and the Whittle estimator, Beran [6]. In order

TABLE 1
Mean and standard deviations for $n = 1{,}000$ and $H = 0.8$ using 500 Monte Carlo simulations of sample paths of processes with variance function $v(\cdot) = |\cdot|^{2H}$, respectively, $v(\cdot) = 1 - \exp(-|\cdot|^{2H})$ (first table) and contaminated versions (second table); see (38)

Estimators	$\boldsymbol{v(\cdot) = \lvert\cdot\rvert^{2H}}$	$\boldsymbol{v(\cdot) = 1 - \exp(-\lvert\cdot\rvert^{2H})}$
Noncontaminated sample paths		
$\mathbf{p} = 1/2$, $\mathbf{c} = 1$ (median)	0.796 (0.042)	0.801 (0.042)
$\mathbf{p} = 0.9$, $\mathbf{c} = 1$	0.797 (0.035)	0.798 (0.036)
$\mathbf{p} = (1/4, 3/4)$, $\mathbf{c} = (1/2, 1/2)$, $g(\cdot) = \lvert\cdot\rvert^2$	0.795 (0.036)	0.800 (0.037)
10%-trimmed mean, $g(\cdot) = \lvert\cdot\rvert^2$	0.797 (0.03)	0.799 (0.034)
Quadratic variations method	0.802 (0.032)	0.798 (0.032)
Whittle estimator	0.805 (0.024)	0.806 (0.024)
Contaminated sample paths		
$\mathbf{p} = 1/2$, $\mathbf{c} = 1$ (median)	0.798 (0.047)	0.803 (0.045)
$\mathbf{p} = 0.9$, $\mathbf{c} = 1$	0.793 (0.033)	0.789 (0.032)
$\mathbf{p} = (1/4, 3/4)$, $\mathbf{c} = (1/2, 1/2)$, $g(\cdot) = \lvert\cdot\rvert^2$	0.797 (0.040)	0.796 (0.037)
10%-trimmed mean, $g(\cdot) = \lvert\cdot\rvert^2$	0.792 (0.037)	0.797 (0.033)
Quadratic variations method	0.329 (0.162)	0.353 (0.149)
Whittle estimator	0.519 (0.106)	0.510 (0.100)

to illustrate the robustness of our estimators, we also applied them to contaminated version of sample path processes. We obtain a new sample path discretized at times

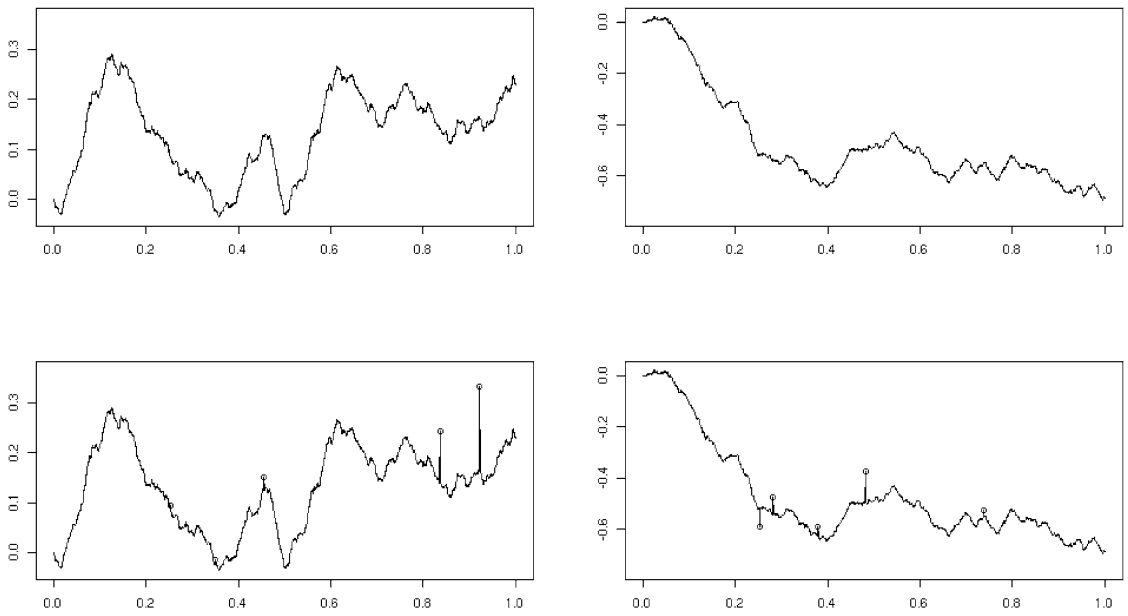


FIG. 1. *Two examples for the sample paths of noncontaminated (top) and contaminated processes with variance function $v(\cdot) = |\cdot|^{2H}$ (left), respectively, $v(\cdot) = 1 - \exp(-|\cdot|^{2H})$ (right); see (38).*

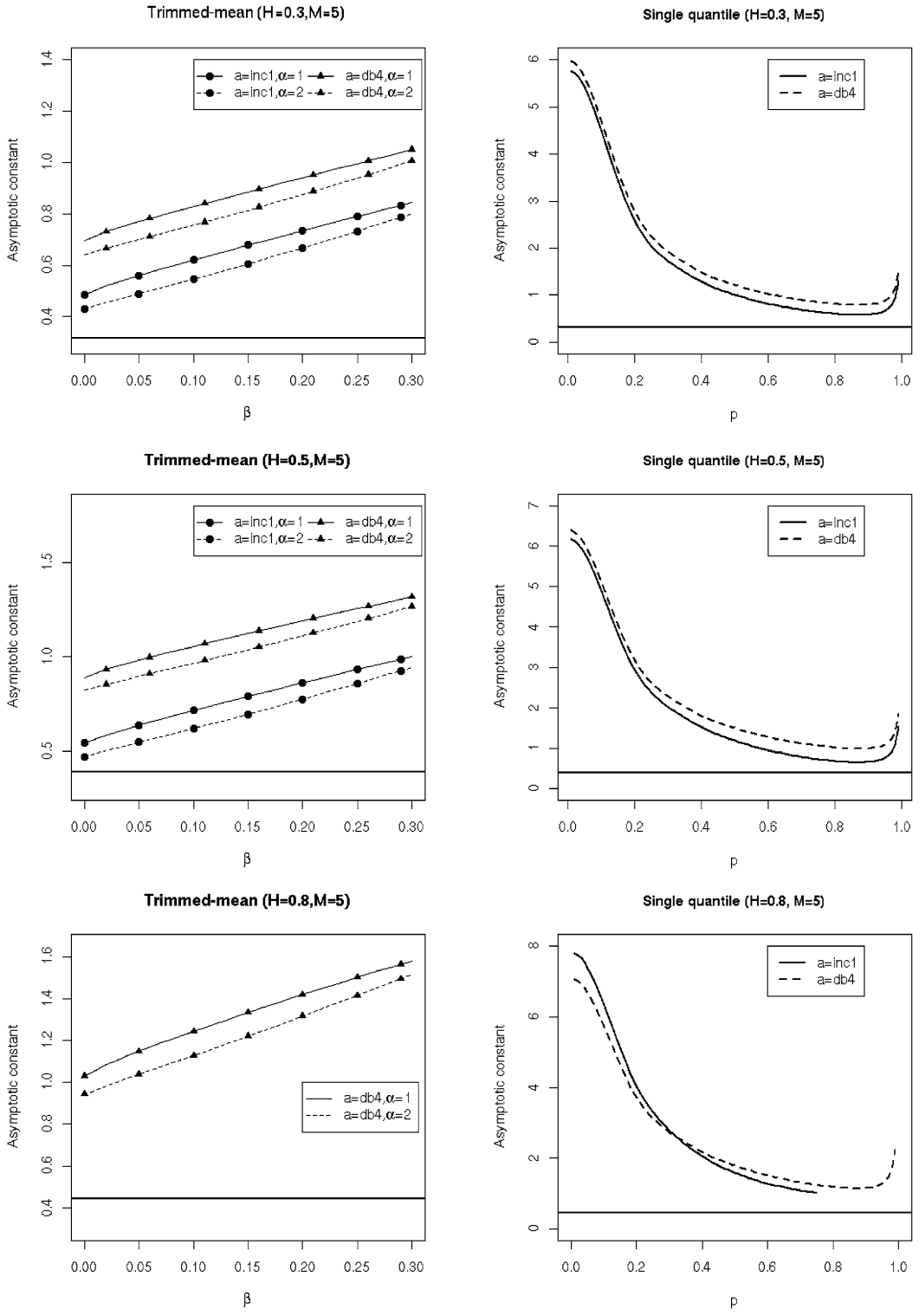


FIG. 2. *Left*: $\sigma^2_{\alpha,tm}$ *in terms of* $\boldsymbol{\beta}$; *Right*: σ^2_{α} *for estimators based on a single quantile in terms of* p. *Three values of the parameter* H *are considered*: 0.3 (*top*), 0.5 (*middle*), 0.8 (*bottom*). *The parameter* M *is fixed to* $M = 5$. *The constant line corresponds to the asymptotic variance of the Whittle's estimator.*

i/n and denoted by $X^C(i/n)$ for $i = 1, \ldots, n$ through the following model:

$$X^C(i/n) = X(i/n) + U(i)V(i), \tag{38}$$

where $U(i), i = 1, \ldots, n$, are Bernoulli independent variables $\mathcal{B}(0.005)$, and $V(i)$, $i = 1, \ldots, n$, are independent centered Gaussian variables with variance $\sigma_C^2(i)$ such that the signal noise ratio at time i/n is equal to 20 dB. As a general conclusion of Table 1, one can say that all versions of our estimators are very competitive with classical ones when the processes are observed without contamination and they seem to be particularly robust to additive outliers. Both bias and variance are approximately unchanged. This is clearly not the case for classical estimators. Indeed, concerning quadratic variations' method, the estimation procedure is based on the estimation of $\mathbf{E}((X^{\mathbf{a}^m}(1/n))^2)$ by sample mean of order 2 of $(\mathbf{X}^{\mathbf{a}^m})^2$ (Coeurjolly [9]) that is particularly sensitive to additive outliers. Bad results of Whittle estimator can be explained by the fact that maximum likelihood methods are also nonrobust methods.

6. Proofs. We denote by $\|\cdot\|_{L^2(d\phi)}$ (resp. $\|\cdot\|_{\ell^q}$) the norm defined by $\|h\|_{L^2(d\phi)} = E(h(Y)^2)^{1/2}$ for some measurable function $h(\cdot)$ [resp. $(\sum_{i\in\mathbb{Z}}|u_i|^q)^{1/2}$ for some sequence $(u_i)_{i\in\mathbb{Z}}$]. In order to simplify the presentation of proofs, we use the notations $F(\cdot)$, $\xi(\cdot)$, $f(\cdot)$, $\widehat{F}(\cdot)$ and $\widehat{\xi}(\cdot)$, instead of $F_{g(Y)}(\cdot)$, $\xi_{g(Y)}(\cdot)$, $f_{g(Y)}(\cdot)$, $\widehat{F}_{g(Y)}(\cdot;\mathbf{g}(\mathbf{Y}))$ and $\widehat{\xi}_{g(Y)}(\cdot;\mathbf{g}(\mathbf{Y}))$, respectively. For some real x, $[x]$ denotes the integer part of x. Finally, λ denotes a generic positive constant.

6.1. *Sketch of the proof of Theorem 2.* We give here a brief explanation of the strategy to prove Theorem 2. This proof follows exactly the one proposed by Serfling [25] in the i.i.d. case. One starts by writing

$$\frac{p - \widehat{F}(\xi(p))}{f(\xi(p))} - \big(\widehat{\xi}(p) - \xi(p)\big) = A(p) + B(p) + C(p),$$

with

$$A(p) = \frac{p - \widehat{F}(\widehat{\xi}(p))}{f(\xi(p))}, \tag{39}$$

$$B(p) = \frac{\widehat{F}(\widehat{\xi}(p)) - \widehat{F}(\xi(p)) - (F(\widehat{\xi}(p)) - F(\xi(p)))}{f(\xi(p))}, \tag{40}$$

$$C(p) = \frac{F(\widehat{\xi}(p)) - F(\xi(p))}{f(\xi(p))} - \big(\widehat{\xi}(p) - \xi(p)\big). \tag{41}$$

From the definition of sample quantile, we have almost surely, see, for example, Serfling [25], $A(p) = \mathcal{O}_{a.s.}(n^{-1})$. Now, in order to control the term $C(p)$, Taylor's theorem is used and a control of $\widehat{\xi}(p) - \xi(p)$ is needed. The latter one is done by Lemma 9 which exhibits the sequence $\varepsilon_n(\alpha, \tau_p)$ such that $\widehat{\xi}(p) - \xi(p) =$

$\mathcal{O}_{a.s.}(\varepsilon_n(\alpha,\tau_p))$. Then, in order to control $B(p)$ it is sufficient to control the random variable

$$S_n(\xi(p),\varepsilon_n(\alpha,\tau_p)) = \sup_{|x|\le\varepsilon_n(\alpha,\tau_p)} \big|\Delta\big(\xi(p)+x\big)-\Delta(\xi(p))\big|,$$

with $\Delta(\cdot)=\widehat{F}(\cdot)-F(\cdot)$. This result is detailed in Lemma 10. In order to specify the rate explicited by Theorem 2, we present and prove Lemmas 9 and 10. Some preliminary results, given by Lemma 6, Corollary 7 and Lemma 8, are needed. Among other things, Lemma 6 and Corollary 7 propose some inequalities for controlling the sample mean of nonlinear function of Gaussian sequences with correlation function satisfying (22).

6.2. *Auxiliary lemmas for the proof of Theorem 2.*

LEMMA 6. *Let $\{Y(i)\}_{i=1}^{+\infty}$ a Gaussian stationary process with variance 1 and correlation function $\rho(\cdot)$ such that, as $i\to+\infty$, $|\rho(i)|\sim L(i)i^{-\alpha}$, for some $\alpha>0$ and some slowly varying function at infinity $L(\cdot)$. Let $h(\cdot)\in L^2(d\phi)$ and denote by τ its Hermite rank. Define*

$$\overline{Y}_n=\frac{1}{n}\sum_{i=1}^{n}h(Y(i)).$$

Then, for all $\gamma>0$, there exists a positive constant $\kappa_\gamma=\kappa_\gamma(\alpha,\tau)$, such that

$$\mathbb{P}(|\overline{Y}_n|\ge\kappa_\gamma y_n)=\mathcal{O}(n^{-\gamma}), \tag{42}$$

with

$$y_n=y_n(\alpha,\tau)=\begin{cases} n^{-1/2}\log(n)^{1/2}, & \text{if } \alpha\tau>1,\\ n^{-1/2}\log(n)^{1/2}L_\tau(n)^{1/2}, & \text{if } \alpha\tau=1,\\ n^{-\alpha\tau/2}\log(n)^{\tau/2}L(n)^{\tau/2}, & \text{if } 0<\alpha\tau<1,\end{cases} \tag{43}$$

where $L_\tau(n)=\sum_{|i|\le n}|\rho(i)|^\tau$. In the case $\alpha\tau=1$, we assume that for all $j>\tau$, the limit, $\lim_{n\to+\infty}L_\tau(n)^{-1}\sum_{|i|\le n}|\rho(i)|^j$ exists.

PROOF. Let $(y_n)_{n\ge1}$ be the sequence defined by (43). The proof is splitted into three parts according to the value of $\alpha\tau$.

Case $\alpha\tau<1$. From Chebyshev's inequality, we have for all $q\ge1$

$$\mathbb{P}(|\overline{Y}_n|\ge\kappa_\gamma y_n)\le\frac{1}{\kappa_\gamma^{2q}y_n^{2q}}\mathbf{E}((\overline{Y}_n)^{2q}).$$

From Theorem 1 of Breuer and Major [7] and in particular equation (2.6), we have, as $n\to+\infty$

$$\mathbf{E}((\overline{Y}_n)^{2q})\sim\frac{(2q)!}{2^q q!}\frac{1}{n^q}\sigma^{2q},\qquad \text{with } \sigma^2=\sum_{i\in\mathbb{Z}}\sum_{j\ge\tau}\frac{(c_j)^2}{j!}\rho(i)^j, \tag{44}$$

where c_j denotes the jth Hermite coefficient of $h(\cdot)$. Note that $\sigma^2 \leq \|h\|^2_{L^2(d\phi)} \times \|\rho\|^2_{\ell^\tau}$. Thus, for n large enough, we have

$$\mathbb{P}(|\overline{Y}_n| \geq \kappa_\gamma y_n) \leq \frac{\lambda}{n^q y_n^{2q}} \frac{(2q)!}{2^q q!} \big(\|h\|^2_{L^2(d\phi)} \|\rho\|^2_{\ell^\tau} \kappa_\gamma^{-2}\big)^q. \tag{45}$$

From Stirling's formula, we have as $q \to +\infty$

$$\frac{(2q)!}{2^q q!} \sim \sqrt{2} q^q (2e^{-1})^q. \tag{46}$$

From (43) by choosing $q = [\log(n)]$, (45) becomes

$$\mathbb{P}(|\overline{Y}_n| \geq \kappa_\gamma y_n) \leq \lambda \big(2e^{-1} \|h\|^2_{L^2(d\phi)} \|\rho\|^2_{\ell^\tau} \kappa_\gamma^{-2}\big)^{\log(n)} = \mathcal{O}(n^{-\gamma}),$$

if $\kappa_\gamma^2 > 2\|h\|^2_{L^2(d\phi)} \|\rho\|^2_{\ell^\tau} \exp(\gamma - 1)$.

Case $\alpha\tau = 1$. Using the proof of Theorem $1'$ of Breuer and Major [7], we can prove that for all $q \geq 1$

$$\begin{aligned} \mathbf{E}((n^{1/2} L_\tau(n)^{-1/2} \overline{Y}_n)^{2q}) &\leq \lambda \frac{2q!}{2^q q!} \mathbf{E}((n^{1/2} L_\tau(n)^{-1/2} \overline{Y}_n)^2)^q \\ &\leq \lambda \frac{2q!}{2^q q!} \left(\sum_{j \geq \tau} \frac{(c_j)^2}{j!} \lim_{n \to +\infty} L_\tau(n)^{-1} \sum_{|i| \leq n} |\rho(i)|^j \right)^q \\ &\leq \lambda \frac{2q!}{2^q q!} \|h\|^{2q}_{L^2(d\phi)}. \end{aligned} \tag{47}$$

Then from Chebyshev's inequality, we have for all $q \geq 1$

$$\mathbb{P}(|\overline{Y}_n| \geq \kappa_\gamma y_n) \leq \lambda \frac{L_\tau(n)^q}{n^q y_n^{2q}} \frac{2q!}{2^q q!} \big(\|h\|^2_{L^2(d\phi)} \kappa_\gamma^{-2}\big)^q.$$

From (43) by choosing $q = [\log(n)]$, we obtain

$$\mathbb{P}(|\overline{Y}_n| \geq \kappa_\gamma y_n) \leq \lambda \big(2e^{-1} \|h\|^2_{L^2(d\phi)} \kappa_\gamma^{-2}\big)^{\log(n)} = \mathcal{O}(n^{-\gamma}),$$

if $\kappa_\gamma^2 > 2\|h\|^2_{L^2(d\phi)} \times \exp(\gamma - 1)$.

Case $\alpha\tau < 1$. Denote by k_α the lowest integer satisfying $k_\alpha \alpha > 1$, that is $k_\alpha = [1/\alpha] + 1$, and for $j \geq \tau$ denote by Z_j the following random variable:

$$Z_j = \frac{1}{n} \sum_{i=1}^n \frac{c_j}{j!} H_j(Y(i)).$$

Denote by $\kappa_{1,\gamma}$ and $\kappa_{2,\gamma}$ two positive constants such that $\kappa_\gamma = \max(\kappa_{1,\gamma}, \kappa_{2,\gamma})$. From the triangle inequality,

$$\mathbb{P}(|\overline{Y}_n| \geq \kappa_\gamma y_n) \leq \mathbb{P}\left(\left| \overline{Y}_n - \sum_{j=\tau}^{k_\alpha - 1} Z_j \right| \geq \kappa_{1,\gamma} y_n \right) + \sum_{j=\tau}^{k_\alpha - 1} \mathbb{P}(|Z_j| \geq \kappa_{2,\gamma} y_n). \tag{48}$$

Since

$$\overline{Y}_n - \sum_{j=\tau}^{k_\alpha - 1} Z_j = \frac{1}{n}\sum_{i=1}^{n}\sum_{j \geq k_\alpha} \frac{c_j}{j!} H_j(Y(i)) = \frac{1}{n}\sum_{i=1}^{n} h'(Y(i)),$$

where $h'(\cdot)$ is a function with Hermite rank k_α. Applying Lemma 6 in the case $\alpha\tau > 1$, it follows that, for all $\gamma > 0$, there exists a constant $\kappa_{1,\gamma}$ such that, for n large enough

$$\mathbb{P}\left(\left|\overline{Y}_n - \sum_{j=\tau}^{k_\alpha - 1} Z_j\right| \geq \kappa_{1,\gamma} y_n\right) = \mathcal{O}(n^{-\gamma}). \tag{49}$$

Now, let $\tau \leq j < k_\alpha$ and $q \geq 1$, from Theorem 3 of Taqqu [27], we have

$$\begin{aligned} &\mathbb{P}(|Z_j| \geq \kappa_{2,\gamma} y_n) \\ &\quad \leq \frac{1}{\kappa_{2,\gamma}^{2q} y_n^{2q}} \left(\frac{c_j}{j!}\right)^{2q} n^{-2q} \mathbf{E}\left(\sum_{i_1,\ldots,i_{2q}} H_j(Y(i_1)) \cdots H_j(Y(i_{2q}))\right) \\ &\quad \leq \lambda \frac{L(n)^{jq}}{n^{\alpha j q} y_n^{2q}} \left(\frac{c_j}{j!}\kappa_{2,\gamma}^{-1}\right)^{2q} \mu_{2q}, \end{aligned} \tag{50}$$

where μ_{2q} is a constant such that $\mu_{2q} \leq (\frac{2}{1-\alpha j})^q \mathbf{E}(H_j(Y)^{2q})$. It is also proved in Taqqu [27], page 228, that $\mathbf{E}(H_j(Y)^{2q}) \sim (2jq)!/(2^{jq}(jq)!)$, as $q \to +\infty$. Thus, from Stirling's formula, we obtain as $q \to +\infty$

$$\begin{aligned} &\mathbb{P}(|Z_j| \geq y_n) \\ &\quad \leq \lambda \frac{L(n)^{(j-\tau)q}}{n^{\alpha(j-\tau)q}} \log(n)^{-\tau q} q^{jq} \left(\frac{2}{1-\alpha j}\left(\frac{c_j}{j!}\right)^2 \left(\frac{2j}{e}\right)^j \kappa_{2,\gamma}^{-1}\right)^q. \end{aligned}$$

By choosing $q = [\log(n)]$, we finally obtain, as $n \to +\infty$

$$\begin{aligned} \sum_{j=\tau}^{k_\alpha - 1} \mathbb{P}(|Z_j| \geq \kappa_{2,\gamma} y_n) &\leq \lambda \left(\frac{2}{1-\alpha\tau}\left(\frac{c_\tau}{\tau!}\right)^2 \left(\frac{2\tau}{e}\right)^\tau \kappa_{2,\gamma}^{-2}\right)^{\log(n)} \\ &= \mathcal{O}(n^{-\gamma}), \end{aligned} \tag{51}$$

if $\kappa_{2,\gamma}^2 > \frac{2}{1-\alpha\tau}(\frac{c_\tau}{\tau!})^2 (2\tau)^\tau \exp(\gamma - \tau)$. From (48), we get the result by combining (49) and (51). □

COROLLARY 7. *Under conditions of Lemma 6, for all* $\alpha > 0$, $j \geq 1$ *and* $\gamma > 0$, *there exists* $q = q(\gamma) \geq 1$ *and* $\zeta_\gamma > 0$ *such that*

$$\mathbf{E}\left(\left\{\frac{1}{n}\sum_{i=1}^{n} H_j(Y(i))\right\}^{2q}\right) \leq \zeta_\gamma n^{-\gamma}. \tag{52}$$

PROOF. (44), (47) and (50) imply that there exists $\lambda = \lambda(q) > 0$ such that for all $q \geq 1$, we have

$$\begin{aligned}\mathbf{E}\left(\left\{\frac{1}{n}\sum_{i=1}^{n} H_j(Y(i))\right\}^{2q}\right) &\leq \lambda(q) n^{-q} \\ &= \lambda(q) \times \begin{cases} n^{-q}, & \text{if } \alpha j > 1, \\ L_{\tau_p}(n) n^{-q}, & \text{if } \alpha j = 1, \\ L(n)^{\alpha j q} n^{-\alpha j q}, & \text{if } \alpha j < 1, \end{cases} \\ &= \mathcal{O}(n^{-\gamma}).\end{aligned} \tag{53}$$

Indeed, it is sufficient to choose q such that, $q > \gamma$ if $\alpha j \geq 1$ and $q > \gamma/\alpha j$ if $\alpha j < 1$. □

LEMMA 8. *Let* $0 < p < 1$, *denote by* $g(\cdot)$ *a function satisfying Assumption* $A_4(\xi(p))$ *and by* $(x_n)_{n\geq 1}$ *a sequence with real components, such that* $x_n \to 0$, *as* $n \to +\infty$. *Then, for all* $j \geq 1$, *there exists a positive constant* $d_j = d_j(\xi(p)) < +\infty$ *such that, for* n *large enough*

$$\left|c_j\left(\xi(p) + x_n\right) - c_j(\xi(p))\right| \leq d_j |x_n|. \tag{54}$$

PROOF. Let $j \geq 1$, under Assumption $A_4(\xi(p))$, for n large enough, $\xi(p) + x_n \in \bigcup_{i=1}^{L} g(U_i)$. Thus, for n large enough,

$$\begin{aligned} &c_j\left(\xi(p) + x_n\right) - c_j(\xi(p)) \\ &\quad = \int_{\mathbb{R}} \left(h_{\xi(p)+x_n}(t) - h_{\xi(p)}(t)\right) H_j(t)\phi(t)\,dt \\ &\quad = \sum_{i=1}^{L} \int_{U_i} \left(\mathbf{1}_{g_i(t)\leq \xi(p)+x_n} - \mathbf{1}_{g_i(t)\leq \xi(p)}\right) H_j(t)\phi(t)\,dt \\ &\quad = \sum_{i=1}^{L} \int_{m_{i,n}}^{M_{i,n}} (-1)^j \phi^{(j)}(t)\,dt, \\ &\quad = \begin{cases} \displaystyle\sum_{i=1}^{L} -\left(\phi(M_{i,n}) - \phi(m_{i,n})\right), & \text{if } j = 1, \\ \displaystyle\sum_{i=1}^{L} (-1)^j \left(\phi^{(j-1)}(M_{i,n}) - \phi^{(j-1)}(m_{i,n})\right), & \text{if } j > 1, \end{cases} \end{aligned}$$

where $g_i(\cdot)$ is the restriction of $g(\cdot)$ to U_i, and where $m_{i,n}$ (resp. $M_{i,n}$) is the minimum (resp. maximum) between $g_i^{-1}(\xi(p)+x_n)$ and $g_i^{-1}(\xi(p))$. We leave the reader to check that there exists a positive constant d_j, such that, for n large enough

$$\begin{aligned}&\bigl|c_j\bigl(\xi(p)+x_n\bigr)-c_j\bigl(\xi(p)\bigr)\bigr|\\&\quad\le d_j|x_n|\times\begin{cases}\displaystyle\sum_{i=1}^{L}\bigl|\phi^{(j)}\bigl(g_i^{(-1)}(u)\bigr)\bigl(g_i^{(-1)}\bigr)'(u)\bigr|, & \text{if } j=1,2,\\ \displaystyle\sum_{i=1}^{L}\bigl|\phi^{(j-2)}\bigl(g_i^{(-1)}(u)\bigr)\bigl(g_i^{(-1)}\bigr)'(u)\bigr|, & \text{if } j>2,\end{cases}\end{aligned}$$

which is the desired result. □

LEMMA 9. *Under conditions of Theorem 2, there exists a constant denoted by* $\kappa_\varepsilon=\kappa_\varepsilon(\alpha,\tau_p)$, *such that, we have almost surely, as* $n\to+\infty$,

$$\bigl|\widehat{\xi}(p;\mathbf{g}(\mathbf{Y}))-\xi_{g(Y)}(p)\bigr|\le\varepsilon_n, \tag{55}$$

where $\varepsilon_n=\varepsilon_n(\alpha,\tau(\xi(p)))=\kappa_\varepsilon y_n(\alpha,\tau(\xi(p)),\ y_n(\cdot,\cdot)$ *being defined by* (43).

PROOF. We have

$$\mathbb{P}\bigl(|\widehat{\xi}(p)-\xi(p)|\ge\varepsilon_n\bigr)=\mathbb{P}\bigl(\widehat{\xi}(p)\le\xi(p)-\varepsilon_n\bigr)+\mathbb{P}\bigl(\widehat{\xi}(p)\ge\xi(p)+\varepsilon_n\bigr). \tag{56}$$

Using Lemma 1.1.4(iii) of Serfling [25], we have

$$\mathbb{P}\bigl(\widehat{\xi}(p)\le\xi(p)-\varepsilon_n\bigr)\le\mathbb{P}\bigl(\widehat{F}\bigl(\xi(p)-\varepsilon_n\bigr)\ge p\bigr). \tag{57}$$

Under Assumption $A_4(\xi(p))$, for n large enough

$$p-F\bigl(\xi(p)-\varepsilon_n\bigr)=f(\xi(p))\varepsilon_n+o(\varepsilon_n)\ge\frac{f(\xi(p))}{2}\varepsilon_n.$$

Consequently, for n large enough and from (57)

$$\mathbb{P}\bigl(\widehat{\xi}(p)\le\xi(p)-\varepsilon_n\bigr)\le\mathbb{P}\biggl(\widehat{F}\bigl(\xi(p)-\varepsilon_n\bigr)-F\bigl(\xi(p)-\varepsilon_n\bigr)\ge\frac{f(\xi(p))}{2}\varepsilon_n\biggr). \tag{58}$$

Define $\tau_{p,n}=\tau(\xi(p)-\varepsilon_n)$, from Lemma 8, we have for n large enough

$$\widehat{F}\bigl(\xi(p)-\varepsilon_n\bigr)-F\bigl(\xi(p)-\varepsilon_n\bigr)\ge2\bigl(\widehat{F}(\xi(p))-F(\xi(p))\bigr)+2\varepsilon_n\sum_{j\in J_n}Z_{n,j}, \tag{59}$$

where

$$J_n=\begin{cases}\{\tau_p<j\le\tau_{p,n}\}, & \text{if } \tau_{p,n}>\tau_p,\\ \varnothing, & \text{if } \tau_{p,n}=\tau_p,\\ \{\tau_{p,n}\le j<\tau_p\}, & \text{if } \tau_{p,n}<\tau_p.\end{cases}\quad\text{and}\quad Z_{n,j}=\frac{1}{n}\sum_{i=1}^{n}\frac{d_j}{j!}H_j(Y(i)).$$

Now, define $c_\varepsilon = \kappa_\varepsilon f(\xi(p))/4$. Let $\gamma > 0$, (52) implies that there exists $q \geq 1$ such that, for n large enough

$$\begin{aligned}\mathbb{P}\left(|2\varepsilon_n Z_n| \geq \frac{f(\xi(p))}{2}\varepsilon_n\right) &\leq \sum_{j\in J_n} \mathbb{P}(|Z_{n,j}| > c_\varepsilon) \\ &\leq \sum_{j\in J_n} \frac{1}{c_\varepsilon^{2q}} \mathbf{E}(Z_{n,j}^{2q}) = \mathcal{O}(n^{-\gamma}).\end{aligned} \tag{60}$$

Let us fix $\gamma = 2$. From (58), (59) and (60) and from Lemma 6 [applied to the function $h_{\xi(p)}(\cdot)$], we obtain

$$\mathbb{P}\big(\widehat{\xi}(p) \leq \xi(p) - \varepsilon_n\big) \leq \mathbb{P}\big(|\widehat{F}(\xi(p)) - F(\xi(p))| \geq c_\varepsilon \varepsilon_n\big) + \mathcal{O}(n^{-2}) = \mathcal{O}(n^{-2}),$$

if $c_\varepsilon > \kappa_2$ that is if $\kappa_\varepsilon > 4/f(\xi(p))\kappa_2$.

Let us now focus on the second right-hand term of (56). Following the sketch of this proof, we may also obtain, for n large enough

$$\mathbb{P}\big(\widehat{\xi}(p) \geq \xi(p) + \varepsilon_n\big) = \mathcal{O}(n^{-2}),$$

if $\kappa_\varepsilon > 4/f(\xi(p))\kappa_2$. Thus, for n large enough $\mathbb{P}(|\widehat{\xi}(p) - \xi(p)| \geq \varepsilon_n) = \mathcal{O}(n^{-2})$, which leads to the result thanks to Borel–Cantelli's lemma. □

The following lemma is an analogous result obtained by Bahadur in the i.i.d. framework; see Lemma E, page 97, of Serfling [25].

LEMMA 10. *Under conditions of Theorem 2, denote by $\Delta(z)$ for $z \in \mathbb{R}$ the random variable, $\Delta(z) = \widehat{F}(z; \mathbf{g}(\mathbf{Y})) - F_{g(Y)}(z)$. Then, we have almost surely, as $n \to +\infty$*

$$\begin{aligned}S_n\big(\xi_{g(Y)}(p), \varepsilon_n(\alpha, \tau_p)\big) &= \sup_{|x|\leq\varepsilon_n} \big|\Delta\big(\xi_{g(Y)}(p) + x\big) - \Delta\big(\xi_{g(Y)}(p)\big)\big| \\ &= \mathcal{O}_{a.s.}(r_n(\alpha, \overline{\tau}_p)),\end{aligned} \tag{61}$$

where $\varepsilon_n = \varepsilon_n(\alpha, \tau_p)$ is defined by (55) and $r_n(\alpha, \overline{\tau}_p)$ is defined by (24).

PROOF. Put $\varepsilon_n = \varepsilon_n(\alpha, \tau_p)$ and $r_n = r_n(\alpha, \overline{\tau}_p)$. Denote by $(\beta_n)_{n\geq 1}$ and $(\eta_{b,n})_{n\geq 1}$ the following two sequences

$$\beta_n = [n^{3/4}\varepsilon_n] \quad \text{and} \quad \eta_{b,n} = \xi(p) + \varepsilon_n \frac{b}{\beta_n},$$

for $b = -\beta_n, \ldots, \beta_n$. Using the monotonicity of $F(\cdot)$ and $\widehat{F}(\cdot)$, we have

$$S_n(\xi(p), \varepsilon_n) \leq \max_{-\beta_n\leq b\leq\beta_n} |M_{b,n}| + G_n, \tag{62}$$

where $M_{b,n} = \Delta(\eta_{b,n}) - \Delta(\xi(p))$ and $G_n = \max_{-\beta_n \le b \le \beta_n - 1}(F(\eta_{b+1,n}) - F(\eta_{b,n}))$. Under Assumption $A_4(\xi(p))$, we have for n large enough

$$G_n \le (\eta_{b+1,n} - \eta_{b,n}) \times \sup_{|x| \le \varepsilon_n} f\big(\xi(p) + x\big) = \mathcal{O}(n^{-3/4}). \tag{63}$$

The proof is finished if one can prove that for all $\gamma > 0$ (in particular $\gamma = 2$) and for all b, there exists κ'_γ such that

$$\mathbb{P}(|M_{b,n}| \ge \kappa'_\gamma r_n) = \mathcal{O}(n^{-\gamma}). \tag{64}$$

Indeed, since $\beta_n = \mathcal{O}(n^{1/2+\delta})$ for all $\delta > 0$, if (64) is true, then we have

$$\begin{aligned}\mathbb{P}\Big(\max_{-\beta_n \le b \le \beta_n} |M_{b,n}| \ge \kappa'_2 r_n(\alpha, \tau_p)\Big) &\le (2\beta_n + 1) \times \max_{-\beta_n \le b \le \beta_n} \mathbb{P}(|M_{b,n}| \ge \kappa'_2 r_n) \\ &= \mathcal{O}(n^{-3/2+\delta}).\end{aligned}$$

Thus, from Borel–Cantelli's lemma, we have almost surely

$$\max_{-\beta_n \le b \le \beta_n} |M_{b,n}| = \mathcal{O}_{a.s.}(r_n).$$

And so, from (62) and (63).

$$S_n(\xi(p), \varepsilon_n) = \mathcal{O}_{a.s.}(r_n) + \mathcal{O}(n^{-3/4}) = \mathcal{O}_{a.s.}(r_n), \tag{65}$$

which is the stated result.

So, the rest of the proof is devoted to prove (64). For the sake of simplicity, denote by $h'_n(\cdot)$ the function $h_{\eta_{b,n}}(\cdot) - h_{\xi(p)}(\cdot)$. For n large enough, the Hermite rank of $h'_n(\cdot)$ is at least equal to $\overline{\tau}_p$, that is defined by (21). In the sequel, we need the following bound for $\|h'_n\|^2_{L^2(d\phi)}$

$$\|h'_n\|^2_{L^2(d\phi)} = \mathbf{E}(h'_n(Y)^2) = \omega_n(1 - \omega_n)$$
$$\text{with } \omega_n = \big|F_{g(Y)}(\eta_{b,n}) - F_{g(Y)}(\xi(p))\big|.$$

As previously, we have $\omega_n = \mathcal{O}(\varepsilon_n)$ and so, there exists $\zeta > 0$, such that

$$\|h'_n\|^2_{L^2(d\phi)} \le \zeta \varepsilon_n. \tag{66}$$

From now on, in order to simplify the proof, we use the following upper-bound:

$$\varepsilon_n = \varepsilon_n(\alpha, \tau_p) \le \varepsilon_n(\alpha, \overline{\tau}_p),$$

and with a slight abuse, we still denote $\varepsilon_n = \varepsilon_n(\alpha, \overline{\tau}_p)$. Note also, that from Lemma 8, the jth Hermite coefficient, for some $j \ge \overline{\tau}_p$, is given by $c_j(\eta_{b,n}) - c_j(\xi(p))$. And there exists a positive constant $d_j = d_j(\xi(p))$ such that for n large enough

$$|c_j(\eta_{b,n}) - c_j(\xi(p)| \le d_j \varepsilon_n \frac{|b|}{\beta_n} \le d_j \varepsilon_n. \tag{67}$$

We now proceed like in the proof of Lemma 6.

Case $\alpha\overline{\tau}_p > 1$: Using Theorem 1 of Breuer and Major [7] and (45), we can obtain for all $q \geq 1$

$$\mathbb{P}(|M_{b,n}| \geq \kappa'_\gamma r_n) \leq \lambda \frac{1}{n^q r_n^{2q}} \frac{(2q)!}{2^q q!} \frac{1}{(\kappa'_\gamma)^{2q}} \|h'_n\|^{2q}_{L^2(d\phi)} \|\rho\|^{2q}_{\ell^{\overline{\tau}_p}}. \tag{68}$$

As $q \to +\infty$, we get

$$\mathbb{P}(|M_{b,n}| \geq \kappa'_\gamma r_n) \leq \lambda \frac{\varepsilon_n^q}{n^q r_n^{2q}} q^q \left(2\zeta e^{-1} \|\rho\|^2_{\ell^{\overline{\tau}_p}} \frac{1}{(\kappa'_\gamma)^2}\right)^q.$$

From (24), (43) (with $\tau = \overline{\tau}_p$) and by choosing $q = [\log(n)]$, we have

$$\mathbb{P}(|M_{b,n}| \geq \kappa'_\gamma r_n) \leq \lambda \left(2\zeta \kappa_\varepsilon e^{-1} \|\rho\|^2_{\ell^{\overline{\tau}_p}} \frac{1}{(\kappa'_\gamma)^2}\right)^{\log(n)} = \mathcal{O}(n^{-\gamma}), \tag{69}$$

if ${\kappa'_\gamma}^2 > 2\zeta\kappa_\varepsilon \|\rho\|^2_{\ell^{\tau_p}} \exp(\gamma - 1)$.

Case $\alpha\overline{\tau}_p = 1$ from (47), we can obtain for all $q \geq 1$

$$\begin{aligned}\mathbf{E}(M_{b,n}^{2q}) &\leq \lambda \frac{(2q)!}{2^q q!} \frac{L_{\overline{\tau}_p}(n)^q}{n^q} \|h'_n\|^{2q}_{L^2(d\phi)} \leq \lambda \zeta^q \frac{(2q)!}{2^q q!} \frac{L_{\overline{\tau}_p}(n)^q \varepsilon_n^q}{n^q} \\ &\leq \lambda \frac{L_{\overline{\tau}_p}(n)^q \varepsilon_n^q}{n^q} (2\zeta e^{-1})^q q^q.\end{aligned}$$

From (24), (43) (with $\tau = \overline{\tau}_p$), by choosing $q = [\log(n)]$, we have

$$\begin{aligned}\mathbb{P}(|M_{b,n}| \geq \kappa'_\gamma r_n) &\leq \frac{1}{{\kappa'_\gamma}^{2q} r_n^{2q}} \mathbf{E}(M_{b,n}^{2q}) \\ &\leq \lambda \left(2\zeta \kappa_\varepsilon e^{-1} \frac{d^2_{\tau_p}}{\tau_p!} \frac{1}{{\kappa'_\gamma}^2}\right)^{\log(n)} = \mathcal{O}(n^{-\gamma}),\end{aligned}$$

if ${\kappa'_\gamma}^2 > 2\zeta\kappa_\varepsilon d^2_{\tau_p}/\tau_p! \exp(\gamma - 1)$.

Case $\alpha\overline{\tau}_p < 1$: Denote by $(r_{1,n})_{n\geq 1}$ and by $(r_{2,n})_{n\geq 1}$ the following two sequences

$$\begin{aligned}r_{1,n} &= n^{-1/2-\alpha\overline{\tau}_p/4} \log(n)^{\overline{\tau}_p/4+1/2} L(n)^{\overline{\tau}_p/4} \quad \text{and} \\ r_{2,n} &= n^{-\alpha\overline{\tau}_p} \log(n)^{\overline{\tau}_p} L(n)^{\overline{\tau}_p}.\end{aligned} \tag{70}$$

Note that $\max(r_{1,n}, r_{2,n})$ is equal to $r_{1,n}$, when $2/3 < \alpha\overline{\tau}_p < 1$ and to $r_{2,n}$, when $0 < \alpha\overline{\tau}_p \leq 2/3$. So, in order to obtain (64) in the case $0 < \alpha\overline{\tau}_p < 1$, it is sufficient to prove that there exists κ'_γ such that, for n large enough

$$\mathbb{P}\big(|M_{b,n}| \geq \kappa'_\gamma \max(r_{1,n}, r_{2,n})\big) = \mathcal{O}(n^{-\gamma}).$$

Denote by k_α the integer $[1/\alpha]+1$ for which $\alpha k_\alpha > 1$, and by $Z_{j,n}$ for $\overline{\tau}_p \le j < k_\alpha$ the random variable defined by

$$Z_{j,n} = \frac{1}{n}\sum_{i=1}^{n} \frac{c_j(\eta_{b,n}) - c_j(\xi(p))}{j!} H_j(Y(i)).$$

From the triangle inequality, we have

$$\begin{aligned}(71)\qquad &\mathbb{P}\big(|M_{b,n}| \ge \kappa'_\gamma \max(r_{1,n}, r_{2,n})\big)\\ &\le \mathbb{P}\left(\left|M_{b,n} - \sum_{j=\overline{\tau}_p}^{k_\alpha-1} Z_{j,n}\right| \ge \kappa'_\gamma r_{1,n}\right) + \sum_{j=\overline{\tau}_p}^{k_\alpha-1} \mathbb{P}(|Z_{j,n}| \ge \kappa'_\gamma r_{2,n}).\end{aligned}$$

Since,

$$M_{b,n} - \sum_{j=\tau_p}^{k_\alpha-1} Z_{j,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\ge k_\alpha} \frac{c_j(\eta_{b,n}) - c_j(\xi(p))}{j!} H_j(Y(i)) = \frac{1}{n}\sum_{i=1}^{n} h''_n(Y(i)),$$

where $h''_n(\cdot)$ is a function with Hermite rank k_α, such that $\alpha k_\alpha > 1$, we have from (68)

$$(72)\qquad \mathbb{P}\left(\left|M_{b,n} - \sum_{j=\overline{\tau}_p}^{k_\alpha-1} Z_{j,n}\right| \ge \kappa'_\gamma r_{1,n}\right) \le \lambda \frac{1}{n^q r_{1,n}^{2q}} \|h'_n\|^{2q}_{L^2(d\phi)} \frac{(2q)!}{2^q q!} \frac{1}{{\kappa'_\gamma}^{2q}} \|\rho\|^{2q}_{\ell^{k_\alpha}}$$

for all $q \ge 1$. From (66), we obtain, as $q \to +\infty$,

$$\mathbb{P}\left(\left|M_{b,n} - \sum_{j=\tau_p}^{k_\alpha-1} Z_{j,n}\right| \ge \kappa'_\gamma r_{1,n}\right) \le \lambda \frac{\varepsilon_n^q}{n^q r_{1,n}^{2q}} q^q (2\zeta e^{-1} \|\rho\|^2_{\ell^{k_\alpha}} {\kappa'_\gamma}^{-2})^q.$$

From (43) (with $\tau = \overline{\tau}_p$), (70) and by choosing $q = [\log(n)]$, we obtain

$$\begin{aligned}(73)\qquad \mathbb{P}\left(\left|M_{b,n} - \sum_{j=\tau_p}^{k_\alpha-1} Z_{j,n}\right| \ge \kappa'_\gamma r_{1,n}\right) &\le \lambda (2\zeta e^{-1} \|\rho\|^2_{\ell^{k_\alpha}} \kappa_\varepsilon {\kappa'_\gamma}^{-2})^{\log(n)}\\ &= \mathcal{O}(n^{-\gamma}),\end{aligned}$$

if ${\kappa'_\gamma}^2 > \kappa'_{1,\gamma} = 2\zeta \|\rho\|^2_{\ell^{k_\alpha}} \kappa_\varepsilon \exp(\gamma - 1)$. Now, concerning the last term of (71), from (50), we can prove, for all $\overline{\tau}_p \le j < k_\alpha$,

$$\mathbb{P}(Z_{j,n} \ge \kappa'_\gamma r_{2,n}) \le \lambda \frac{L(n)^{jq}}{n^{\alpha jq} r_{2,n}^{2q}} \frac{1}{{\kappa'_\gamma}^{2q}} \left(\frac{c_j(\eta_{b,n}) - c_j(\xi(p))}{j!}\right)^{2q} \mu_{2q},$$

where μ_{2q} is a constant such that, as $q \to +\infty$,

$$\mu_{2q} \le \lambda \left(\frac{2}{1-\alpha j}\right)^q \frac{(2jq)!}{2^{jq}(jq)!}.$$

From (67), we have, as $q \to +\infty$,

$$\mathbb{P}(Z_{j,n} \geq \kappa'_\gamma r_{2,n}) \leq \lambda \frac{\varepsilon_n^{2q} L(n)^{jq}}{n^{\alpha jq} r_{2,n}^{2q}} q^{jq} \left(\frac{2}{1-\alpha j} \left(\frac{2j}{e} \right)^j d_j^2 {\kappa'_\gamma}^{-2} \right)^{2q}.$$

From (24), (43) (with $\tau = \overline{\tau}_p$) by choosing $q = [\log(n)]$, we have, as $n \to +\infty$,

$$\mathbb{P}(Z_{j,n} \geq \kappa'_\gamma r_{2,n}) \leq \lambda \left(\frac{\log(n) L(n)}{n^\alpha} \right)^{(j-\overline{\tau}_p)q} \left(\frac{2}{1-\alpha j} \left(\frac{2j}{e} \right)^j d_j^2 \kappa_\varepsilon^2 {\kappa'_\gamma}^{-2} \right)^q.$$

Consequently, as $n \to +\infty$, we finally obtain

$$\begin{aligned} \sum_{j=\overline{\tau}_p}^{k_\alpha - 1} \mathbb{P}(Z_{j,n} \geq \kappa'_\gamma r_{2,n}) &\leq \lambda \left(\frac{2}{1-\alpha\overline{\tau}} \left(\frac{2\overline{\tau}}{e} \right)^{\overline{\tau}} d_{\overline{\tau}}^2 \kappa_\varepsilon^2 {\kappa'_\gamma}^{-2} \right)^{\log(n)} \\ &= \mathcal{O}(n^{-\gamma}), \end{aligned} \tag{74}$$

if ${\kappa'_\gamma}^2 > \kappa'^2_{2,\gamma} = \frac{2}{1-\alpha\overline{\tau}} (\frac{2\overline{\tau}}{e})^{\overline{\tau}} d_{\overline{\tau}}^2 \kappa_\varepsilon^2 \exp(\gamma - \overline{\tau})$. Let us choose κ'_γ such that ${\kappa'_\gamma}^2 > \max(\kappa'^2_{1,\gamma}, \kappa'^2_{2,\gamma})$. Then, by combining (73) and (74), we deduce from (71) that, for every $\gamma > 0$,

$$\mathbb{P}\big(|M_{b,n}| \geq \kappa'_\gamma \max(r_{1,n}, r_{2,n})\big) = \mathcal{O}(n^{-\gamma}),$$

and so (64) is proved. □

6.3. *Proof of Theorem 2.* Let us detail the proof presented in Section 6.1. We have

$$\frac{p - \widehat{F}(\xi(p))}{f(\xi(p))} - \big(\widehat{\xi}(p) - \xi(p)\big) = A(p) + B(p) + C(p)$$

with $A(p)$, $B(p)$ and $C(p)$, respectively, defined by (39), (40) and (41). Under Assumption $A_4(\xi(p))$, from Lemma 9 and Taylor's theorem, we have almost surely, as $n \to +\infty$,

$$C(p) \leq \sup_{|x| \leq \varepsilon_n(\alpha,\tau_p)} F''_{g(Y)}\big(\xi(p) + x\big)\big(\widehat{\xi}(p) - \xi(p)\big)^2 = \mathcal{O}_{a.s.}(\varepsilon_n(\alpha,\tau_p)^2).$$

From the definition of sample quantile, we have almost surely, see, for example, Serfling [25], $A(p) = \mathcal{O}_{a.s.}(n^{-1})$. Now, by combining Lemma 9 and Lemma 10, we have almost surely $B(p) = \mathcal{O}_{a.s.}(r_n(\alpha, \overline{\tau}_p))$. Thus, we finally obtain

$$\begin{aligned} \widehat{\xi}(p) - \xi(p) = \frac{p - \widehat{F}(\xi(p))}{f(\xi(p))} &+ \mathcal{O}_{a.s.}(n^{-1}) \\ &+ \mathcal{O}_{a.s.}(r_n(\alpha, \overline{\tau}_p)) + \mathcal{O}_{a.s.}(\varepsilon_n(\alpha,\tau_p)^2), \end{aligned}$$

which leads to the result by noticing that $\varepsilon_n(\alpha,\tau_p)^2 = \mathcal{O}(r_n(\alpha,\overline{\tau}_p))$.

6.4. *Auxiliary lemmas for the proof of Theorem 4.* Let $0 < p_0 \leq p_1 < 1$.

LEMMA 11. *Under conditions of Theorem 3, there exists a constant denoted by* $\theta = \theta(\alpha, \tau_{p_0,p_1})$ *such that, we have almost surely, as* $n \to +\infty$,

$$T = \sup_{p_0 \leq p \leq p_1} \left|\widehat{\xi}(p; \mathbf{g}(\mathbf{Y})) - \xi_{g(Y)}(p)\right| \leq \varepsilon_n(\alpha, \tau_{p_0,p_1}), \tag{75}$$

where $\varepsilon_n = \varepsilon_n(\alpha, \tau_{p_0,p_1}) = \theta y_n(\alpha, \tau_{p_0,p_1})$ *and* y_n *is given by* (50).

The following result is an extension of Lemma 10 and Theorem 4.2 obtained by Sen and Ghosh [23].

LEMMA 12. *Under assumptions of Theorem 3 and following Lemma 10, we have almost surely, as* $n \to +\infty$,

$$S_n^\star = \sup_{\substack{x,y \in [\xi(p_0), \xi(p_1)] \\ |x-y| \leq \varepsilon_n(\alpha, \tau_{p_0,p_1})}} |\Delta(x) - \Delta(y)| = \mathcal{O}_{a.s.}(r_n(\alpha, \tau_{p_0,p_1})) \tag{76}$$

where τ_{p_0,p_1} *is defined by* (25).

Proofs of Lemmas 11 and 12 are omitted since they are essentially based on the same arguments of proofs of Lemmas 9 and 10, see Coeurjolly [10].

6.5. *Proof of Theorem 3.* We follow the proof of Theorem 2. Let $p \in [p_0, p_1]$ and let $\varepsilon_n = \varepsilon_n(\alpha, \tau_{p_0,p_1})$, then

$$\frac{p - \widehat{F}(\xi(p))}{f(\xi(p))} - \big(\widehat{\xi}(p) - \xi(p)\big) = A(p) + B(p) + C(p),$$

where $A(p), B(p)$ and $C(p)$ are respectively defined by (39), (40) and (41). Similarly to the proof of Theorem 2, one may prove that $\sup_{p_0 \leq p \leq p_1} A(p) = \mathcal{O}_{a.s.}(n^{-1})$. Under Assumption $A_5(p_0, p_1)$, $C(p) \leq (\sup_{|x| \leq \varepsilon_n(\alpha,\tau_p)} F''(x + \xi(p)))\frac{(\widehat{\xi}(p) - \xi(p))^2}{f(\xi(p))}$. Therefore, for n large enough, $C(p) \leq \lambda(\sup_{p_0 \leq p \leq p_1}(\widehat{\xi}(p) - \xi(p)))^2$. And from Lemma 11, this leads to

$$\sup_{p_0 \leq p \leq p_1} C(p) = \mathcal{O}_{a.s.}(\varepsilon_n(\alpha, \tau_{p_0,p_1})^2).$$

In addition, using Lemma 12, one also has $\sup_{p_0 \leq p \leq p_1} B(p) = \mathcal{O}_{a.s.}(r_n(\alpha, \tau_{p_0,p_1}))$, which ends the proof.

6.6. *Auxiliary lemma for the proof of Theorem 4.*

LEMMA 13. *Consider for* $0 < p < 1$ *the function* $h_p(\cdot)$, *given by*

$$h_p(t) = \mathbf{1}_{\{|t| \le \xi_{|Y|}(p)\}}(t) - p, \tag{77}$$

that is the function $h_{\xi_{g(Y)}(p)}(\cdot)$ *with* $g(\cdot) = |\cdot|$. *Then by denoting* $c_j^{h_p}$ *the* j*th Hermite coefficient of* $h_p(\cdot)$, *we have for all* $j \ge 1$

$$c_0^{h_p} = c_{2j+1}^{h_p} = 0 \quad \textit{and} \quad c_{2j}^{h_p} = -2H_{2j-1}(q)\phi(q), \tag{78}$$

where $q = \xi_{|Y|}(p) = \Phi^{-1}(\frac{1+p}{2})$.

PROOF. Since $\mathbb{P}(|Y| \le q) = p$ and $h_p(\cdot)$ is even, we have $c_0^{h_p} = c_{2j+1}^{h_p} = 0$, for all $j \ge 1$. Now, (19) implies

$$\begin{aligned} c_{2j}^{h_p} &= \int_{\mathbb{R}} h_p(t) H_{2j}(t)\phi(t)\,dt = 2 \times \int_0^q H_{2j}(t)\phi(t)\,dt \\ &= 2 \times \left[\phi^{(2j-1)}(t)\right]_0^q = 2 \times \left[-H_{2j-1}(t)\phi(t)\right]_0^q \\ &= -2H_{2j-1}(q)\phi(q). \end{aligned}$$

□

REMARK 10. Let $g(\cdot) = \widetilde{g}(|\cdot|)$, where $\widetilde{g}(\cdot)$ is a strictly increasing function on $\mathbb{R}^+$, then for all $0 < p < 1$, we have

$$\xi_{|Y|}(p) = \widetilde{g}^{-1}\big(\xi_{g(Y)}(p)\big).$$

Consequently, the functions $h_{\xi_{g(Y)}(p)}(\cdot)$ for $g(\cdot) = |\cdot|$, $g(\cdot) = |\cdot|^\alpha$ and $g(\cdot) = \log|\cdot|$ are strictly identical. And so, their Hermite decomposition is given by (78) and their Hermite rank is equal to 2.

6.7. *Proofs of Theorems 4 and 5.* Once Lemma 13, Theorems 2 and 3 are established, the proofs of Theorem 4 and Theorem 5 are semiroutine. They are essentially based on the application of a general central limit theorem obtained by Arcones [3] for nonlinear functional of Gaussian vector fields. The reader is referred to Coeurjolly [10] for details.

Acknowledgments. The author is very grateful to Anestis Antoniadis and Rémy Drouilhet for helpful comments and to Kinga Sipos for a careful reading of the present paper.

REFERENCES

[1] ALÒS, E., MAZET, O. and NUALART, D. (1999). Stochastic calculus with respect to fractional Brownian motion with Hurst parameter lesser than $\frac{1}{2}$. *Stoch. Proc. Appl.* **86** 121–139. MR1741199

[2] ANTONIADIS, A., BERRUYER, J. and CARMONA, R. (1992). *Régression non linéaire et applications*. Editions Economica, Paris.

[3] ARCONES, M. A. (1994). Limit theorems for nonlinear functionals of stationary Gaussian field of vectors. *Ann. Probab.* **22** 2242–2274. MR1331224

[4] BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580. MR0189095

[5] BARDET, J.-M., LANG, G., OPPENHEIM, G., PHILIPPE, A., STOEV, S. and TAQQU, M. (2003). Semi-parametric estimation of the long-range dependence parameter: A survey. In *Theory and Applications of Long-Range Dependence* (P. Doukhan, G. Oppenheim and M. S. Taqqu, eds.) 557–577. Birkhäuser, Boston. MR1956041

[6] BERAN, J. (1994). *Statistics for Long Memory Processes.* Chapman and Hall, London. MR1304490

[7] BREUER, P. and MAJOR, P. (1983). Central limit theorems for non-linear functionals of Gaussian fields. *J. Multivariate Anal.* **13** 425–441. MR0716933

[8] COEURJOLLY, J.-F. (2000). Simulation and identification of the fractional Brownian motion: A bibliographical and comparative study. *J. Statist. Soft.* **5** 1–53.

[9] COEURJOLLY, J.-F. (2001). Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Statist. Inference Stoch. Process.* **4** 199–227. MR1856174

[10] COEURJOLLY, J.-F. (2007). Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. Available at http://fr.arxiv.org/abs/math.ST/0506290.

[11] CONSTANTINE, A. G. and HALL, P. (1994). Characterizing surface smoothness via estimation of effective fractal dimension. *J. Roy. Statist. Soc. Ser. B* **56** 97–113. MR1257799

[12] DACUNHA-CASTELLE, D. and DUFLO, M. (1982). *Exercices de probabilités et statistiques. Tome 1. Problèmes à temps fixe.* Collection Mathématiques Appliquées pour la Maîtrise, Masson, Paris. MR0680579

[13] DAUBECHIES, I. (1998). *Ten Lectures on Wavelets*. SIAM, Philadelphia. MR1162107

[14] FLANDRIN, P. (1992). Wavelet analysis of fractional Brownian motion. *IEE Trans. Inform. Theory* **38** 910–917. MR1162229

[15] GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long-memory time-series models. *J. Time Ser. Anal.* **4** 221–238. MR0738585

[16] HESSE, C. H. (1990). A Bahadur-type representation for empirical quantiles of a large class of stationary, possibly infinite-variance, linear processes. *Ann. Statist.* **18** 1188–1202. MR1062705

[17] HO, H. C. and HSING, T. (1996). On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992–1024. MR1401834

[18] ISTAS, J. and LANG, G. (1997). Quadratic variations and estimation of the Hölder index of a Gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.* **33** 407–436. MR1465796

[19] KENT, J. T. and WOOD, A. T. A. (1997). Estimating the fractal dimension of a locally self-similar Gaussian process using increments. *J. Roy. Statist. Soc. Ser. B* **59** 679–700. MR1452033

[20] KIEFER, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38** 1323–1342. MR0217844

[21] LIM, S. C. (2001). Fractional Brownian motion and multifractional of Riemann–Liouville type. *J. Phys. A: Math. Gen.* **34** 1301–1310. MR1819933

[22] MANDELBROT, B. and VAN NESS, J. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437. MR0242239

[23] SEN, P. K. and GHOSH, M. (1971). On bounded length sequential confidence intervals based on one-sample rank order statistics. *Ann. Math. Statist.* **42** 189–203. MR0279939

[24] SEN, P. K. (1972). On the Bahadur representation of sample quantiles for sequences of ϕ-mixing random variables. *J. Multivariate Anal.* **2** 77–95. MR0303581

[25] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

[26] STOEV, S., TAQQU, M., PARK, C., MICHAILIDIS, G. and MARRON, J. S. (2006). LASS: A tool for the local analysis of self-similarity. *Comput. Statist. Data Anal.* **50** 2447–2471. MR2225578

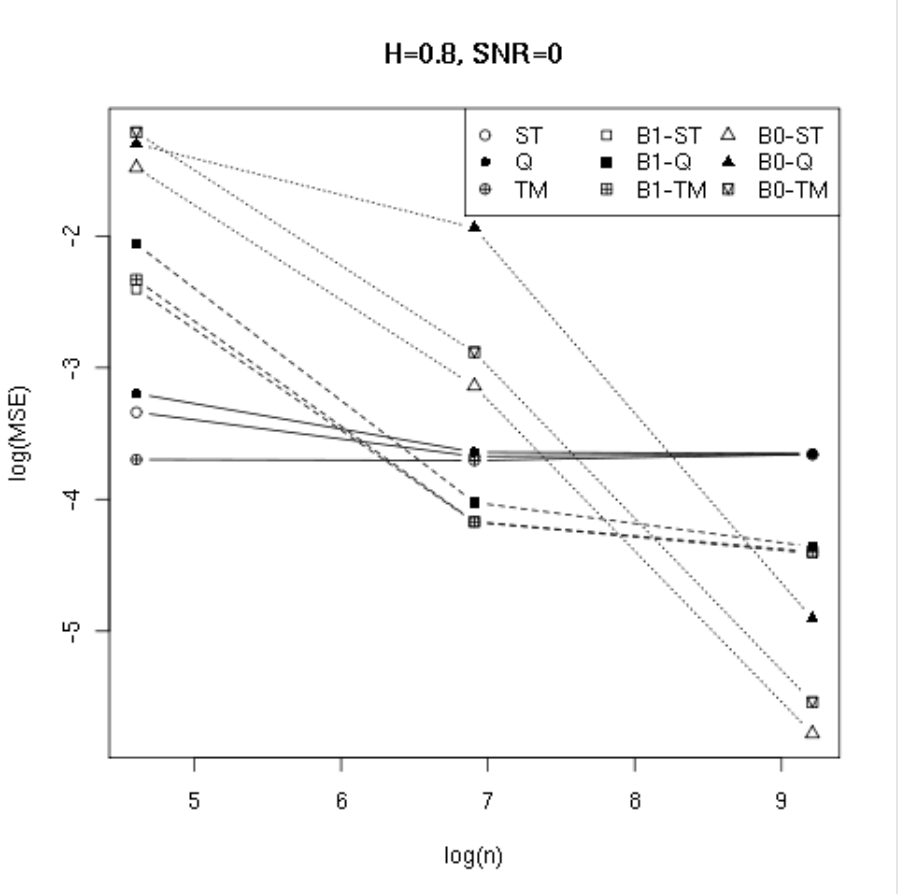
[27] TAQQU, M. S. (1977). Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Z. Wahrsch. Verw. Gebiete* **40** 203–238. MR0471045

[28] WOOD, A. T. A. and CHAN, G. (1994). Simulation of stationary Gaussian processes in $[0, 1]^d$. *J. Comput. Graph. Statist.* **3** 409–432. MR1323050

[29] YOSHIHARA, K.-I. (1995). The Bahadur representation of sample quantiles for sequences of strongly mixing random variables. *Stat. Probab. Lett.* **24** 299–304. MR1353887

[30] WU, W.-B. (2005). On the Bahadur representation of sample quantiles for dependent sequences. *Ann. Statist.* **33** 1934–1963. MR2166566

LJK, SAGAG TEAM, UNIVERSITÉ GRENOBLE 2
1251 AV. CENTRALE BP 47
38040 GRENOBLE CEDEX 09
FRANCE
E-MAIL: Jean-Francois.Coeurjolly@upmf-grenoble.fr



[AC10]
S. Achard and J.-F. Coeurjolly. Discrete variations for the fractional Brownian motion in presence of outliers and/or an additive noise. *Statistics Surveys*, 4:117-147, 2010.

Statistics Surveys
Vol. 4 (2010) 117–147
ISSN: 1935-7516
DOI: 10.1214/09-SS059

Discrete variations of the fractional Brownian motion in the presence of outliers and an additive noise*

Sophie Achard

GIPSA-lab, CNRS, France
e-mail: Sophie.Achard@gipsa-lab.inpg.fr

and

Jean-François Coeurjolly

GIPSA-lab, Grenoble University, France
and LJK, Grenoble University, France
e-mail: Jean-Francois.Coeurjolly@upmf-grenoble.fr

Abstract: This paper gives an overview of the problem of estimating the Hurst parameter of a fractional Brownian motion when the data are observed with outliers and/or with an additive noise by using methods based on discrete variations. We show that the classical estimation procedure based on the log-linearity of the variogram of dilated series is made more robust to outliers and/or an additive noise by considering sample quantiles and trimmed means of the squared series or differences of empirical variances. These different procedures are compared and discussed through a large simulation study and are implemented in the `R` package `dvfBm`.

AMS 2000 subject classifications: Primary 60G15, 62F10; secondary 62F35.
Keywords and phrases: Fractional Brownian motion, Hurst exponent estimation, discrete variations, robustness, outliers.

Received November 2009.

Contents

*This paper was accepted by Claudia Klüppelberg, Associate Editor for the Bernoulli.

1. Introduction

Since the pioneering work of Mandelbrot and Ness (1968), the fractional Brownian motion (fBm) has become widely popular in a theoretical context as well as in a practical one for modelling self-similar phenomena. Fractional Brownian motion can be defined as the only centered Gaussian process, denoted by $(B_H(t))_{t\in\mathbb{R}}$, with stationary increments and with variance function $v(\cdot)$, given by $v(t) = C^2|t|^{2H}$ for all $t \in \mathbb{R}$. The parameter $H \in (0,1)$ (resp. $C > 0$) is referred to as the Hurst parameter (resp. the scaling coefficient). In particular, the case $H = 1/2$ corresponds to the standard Brownian motion. In general, the fractional Brownian motion is an H-self-similar process, that is for all $\delta > 0$, $(B_H(\delta t))_{t\in\mathbb{R}} \overset{d}{=} \delta^H (B_H(t))_{t\in\mathbb{R}}$ (where $\overset{d}{=}$ means equal in finite-dimensional distributions) with autocovariance function behaving like $O(|k|^{2H-2})$ as $|k| \to +\infty$. Thus, the discretized increments of the fractional Brownian motion (called the fractional Gaussian noise) constitute a short-range dependent process, when $H < 1/2$, and a long-range dependent process, when $H > 1/2$. The index H characterizes also the path regularity since the fractal dimension of the fractional Brownian motion is equal to $D = 2 - H$. General references on self-similar processes and long-memory processes are given in Beran (1994) or Doukhan et al. (2003).

As the Hurst parameter H governs the fractal dimension of the fractional Brownian motion, its regularity and the long-memory behavior of its increments, the estimation of H is a very important (and quite difficult) task which has led to a very vast literature. We refer the interested reader to Coeurjolly (2000a), to the book of Doukhan et al. (2003) and the references therein and to the excellent paper of Fäy et al. (2009) which focuses on long-memory processes.
The present paper hightlights one class of these methods, namely the method based on discrete variations, which has known great developments these last years. This method originates simultaneously from works of Kent and Wood (1997) and Istas and Lang (1997) in the context of locally self-similar Gaussian processes and more deeply in Coeurjolly (2001) in the case of the fractional Brownian motion. These ideas have then been used/extended in many other

situations: *e.g.* Cohen and Istas (2002) for more general local self-similar processes, Coeurjolly (2005) for the multifractional Brownian motion, Coeurjolly (2008) for a more robust estimate, Richard and Biermé (2008) for anisotropic Gaussian random fields, Istas (2007) in the context of spherical Brownian motion, Brouste et al. (2007) for more general Gaussian random fields ...

This paper focuses on fBm-type processes by using discrete variations type procedures. Consider first, $\mathbf{B}_H$ a sample path of a fractional Brownian motion discretized at times $i = 1, \ldots, n$ and with parameters H, C. Let a be a vector with real components representing a filter and $\mathbf{B}_H^a$ the filtered series. For example, when $a = (1, -1)$ (resp. $(1, -2, 1)$), $\mathbf{B}_H^a$ corresponds to the increments (resp. the increments of the increments) of $\mathbf{B}_H$ (this is presented in Section 2). Moreover, let a^m be the filter a dilated m times (for example $(1, -2, 1)^2 = (1, 0, -2, 0, 1)$), the classical estimation procedure is based on the following property

$$Var\left(B_H^{a^m}(i)\right) = m^{2H} \times \gamma_{H,C} \Leftrightarrow \log\left(Var\left(B_H^{a^m}(i)\right)\right) = 2H\log(m) + \log(\gamma_{H,C}),$$

where $\gamma_{H,C}$ is a constant independent of m. It is now sufficient for different values of m to estimate the variance by its empirical version and to estimate H (actually $2H$) through a simple log-linear regression. This procedure has many advantages: it is extremely simple to implement, computationnally fast (it does not need a large number of dilated filters). In addition, the definition of the estimate is independent of the scaling coefficient and invariant of the discretization step. From a theoretical point of view (see *e.g.* Coeurjolly (2001)), this estimate is consistent and follows a central limit theorem if $p = 1$ and $H < 3/4$ and for any H if $p \geq 2$ where p is the order of the filter (1 for $a = (1, -1)$ and 2 for $a = (1, -2, 1), \ldots$). This is proved by the fact that the correlation function of the filtered series decays as $|k|^{2H-2p}$.

The aim of this paper is to show that, when the data are contaminated by outliers and/or by an additive noise, it is still possible to adapt the previous method in order to take into account the possible contaminations and to keep its principal properties: estimation of H without estimating any other parameters, simple and computationnally fast. The Sections 2 and 3 give a survey on this topic: we show how when replacing the empirical variance by sample quantiles or trimmed means of the squared series, it is possible to define an estimate more robust to outliers. This problem has been already considered by Coeurjolly (2008). We also demonstrate that if the data are composed of a fractional Brownian motion plus a standard Brownian motion or standard Gaussian variables, it is still possible to define an estimate of H by considering differences of empirical variances. Finally, we also show that it is possible to combine these different procedures. We propose consistency results (depending on the model) proved in appendix. In Section 5, we have conducted a large simulation study where pure and contaminated sample paths of fractional Brownian motions are considered. The different estimation procedures and parameters are compared and discussed. Finally, this paper is accompanied with a `R` package named `dvfBm` available on the `R CRAN` (`http://cran.r-project.org/`)

2. Discrete variations of the fractional Brownian motion

2.1. Some general notation

Let $\mathbf{X} = (X(1), \ldots, X(n))$ be a sample of a stochastic process (with stationary increments and finite variance) at times $i = 1, \ldots, n$. Define a as a filter of length $\ell + 1$ with order $p \geq 1$, that is a vector with $\ell + 1$ real components satisfying

$$\sum_{q=0}^{\ell} q^j a_q = 0 \text{ for } j = 0, \ldots, p-1 \text{ and } \sum_{q=0}^{\ell} q^p a_q \neq 0. \tag{1}$$

For instance, we shall consider the following filters:

- Increments 1: $a = i1 = (-1, 1)$,
- Increments 2: $a = i2 = (1, -2, 1)$,
- Daublets 4: $a = d4 = (-0.09150635, -0.15849365, 0.59150635, -0.34150635)$
- ...

We refer the reader to Percival and Walden (2000) or Daubechies (2006) for details on Daubechies wavelet filters and extensions. We define also the vector $\mathbf{X}^a$ as the vector $\mathbf{X}$ filtered with a and given for $i = \ell + 1, \ldots, n$ by

$$X^a(i) := \sum_{q=0}^{\ell} a_q X(i-q).$$

$\widetilde{\mathbf{X}}^a$ is the normalized vector of $\mathbf{X}^a$ defined by

$$\widetilde{X}^a(i) = \frac{X^a(i)}{\mathbf{E}\left(X^a(i)^2\right)^{1/2}} = \frac{X^a(i)}{\mathbf{E}\left(X^a(1)^2\right)^{1/2}},$$

due to the stationarity of the increments of $\mathbf{X}$. Let us also denote for a function $g(\cdot)$ the vector $\mathbf{g}(\mathbf{X}) = (g(X(1)), \ldots, g(X(n)))$. Moreover, $\overline{\mathbf{X}}$, $\widehat{\xi}(p, \mathbf{X})$ (for some $0 < p < 1$) and $\overline{\mathbf{X}}^{(\beta)}$ (for some vector $\beta = (\beta_1, \beta_2)$ satisfying $0 < \beta_1, \beta_2 < 1/2$) will respectively denote the empirical mean of $\mathbf{X}$, the sample quantile of $\mathbf{X}$ and the $\beta-$trimmed mean of $\mathbf{X}$ defined by

$$\overline{\mathbf{X}}^{(\beta)} = \frac{1}{n - [n\beta_2] - [n\beta_1]} \sum_{i=[n\beta_1]+1}^{n-[n\beta_2]} (\mathbf{X}^a)_{(i),n},$$

where $[\cdot]$ denotes the integer part and where $(\mathbf{X})_{(1),n} \leq (\mathbf{X})_{(2),n} \leq \cdots \leq (\mathbf{X})_{(n),n}$ are the order statistics of $X(1), \ldots, X(n)$. Finally, in the sequel Z will denote a random variable following a standard Gaussian distribution.

2.2. Applications to the fractional Brownian motion

Let $\mathbf{B}_H$ be a discretized sample path of a fractional Brownian motion at times $i = 1, \ldots, n$ with Hurst parameter $H \in (0, 1)$ and with scaling coefficient $C >$

0. Denote by $\mathbf{B}_H^a$ and $\widetilde{\mathbf{B}}_H^a$ its filtered version and normalized filtered version. The covariance and correlation functions of $\mathbf{B}_H^a$ are given for $i,j \in \mathbb{Z}$ by (see Coeurjolly (2001))

$$\mathbf{E}\left(B_H^a(j)B_H^a(i+j)\right) = C^2 \times \pi_H^a(i) \text{ with } \pi_H^a(i) = -\frac{1}{2}\sum_{q,r=0}^{\ell} a_q a_r |q-r+i|^{2H}$$

and

$$\mathbf{E}\left(\widetilde{B}_H^a(j)\widetilde{B}_H^a(i+j)\right) = \frac{\pi_H^a(i)}{\pi_H^a(0)}.$$

Note that the last expression is clearly independent of C. The interest of filtering a discretized sample of a fractional Brownian motion is revealed by the fact that the action of filtering destroys the correlation of the increments. Indeed, it was proved (see *e.g.* Coeurjolly (2001)) that $\rho_H^a(i) \sim k_H|i|^{2H-2p}$, as $|i| \to +\infty$.

Let us now explain how one can estimate the parameter H (independently of C). First, consider the collection of dilated filters $(a^m)_{m\geq 1}$ of a filter a. Recall that a^m is the filter of length $m\ell+1$ with order p and is defined for $i = 0,\ldots,m\ell$ by

$$a_i^m = \begin{cases} a_{i/m} & \text{if } i/m \text{ is an integer} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

As a typical example, if $a := a^1 = (1,-2,1)$, then $a^2 := (1,0,-2,0,1)$. It is shown that

$$\pi_H^{a^m}(0) = -\frac{1}{2}\sum_{q,r=0}^{m\ell} a_q^m a_r^m |q-r|^{2H} = -\frac{1}{2}\sum_{q,r=0}^{\ell} a_q a_r |mq-mr|^{2H} = m^{2H}\pi_H^a(0).$$

Now, consider the empirical mean of the squared filtered coefficients denoted by $\overline{(\mathbf{B}_H^{a^m})^2}$. Since,

$$\mathbf{E}\left(\overline{(\mathbf{B}_H^{a^m})^2}\right) = C^2\pi_H^{a^m}(0) = m^{2H}C^2\pi_H^a(0),$$

one may obtain, by denoting $\gamma = \gamma_{H,C} := C^2\pi_H^a(0)$ (which is independent of m), the following simple linear regression model

$$\log\left(\overline{(\mathbf{B}_H^{a^m})^2}\right) = 2H\log(m) + \log(\gamma) + \underbrace{\log\left(\frac{\overline{(\mathbf{B}_H^{a^m})^2}}{\mathbf{E}\left(B_H^{a^m}(1)^2\right)}\right)}_{:=\varepsilon_m^{ST}}. \tag{3}$$

The ordinary least squares estimate associated to the regression model (3) is then given by:

$$\widehat{H}^{ST} = \frac{\mathbf{A}^T}{2\|\mathbf{A}\|^2}\left(\log\left(\overline{(\mathbf{B}_H^{a^m})^2}\right)\right)_{m=M_1,\ldots,M_2}, \tag{4}$$

where $A_m = \log(m) - \frac{1}{M_2-M_1+1}\sum_{m=M_1}^{M_2}\log(m)$ and $1 \leq M_1 \leq M_2$.

3. Discrete variations of contaminated sample paths of the fractional Brownian motion

3.1. Robustness to outliers

As noted by Coeurjolly (2008) and Shen et al. (2007), the standard procedure (which is very close to a wavelet procedure) may be particularly affected by outliers. The aim of this section is to propose alternative procedures based on sample quantiles or trimmed means.

3.1.1. Using sample quantiles

Let us denote by $(\mathbf{p},\mathbf{c}) = (p_k,c_k)_{k=1,\dots,K} \in ((0,1)\times\mathbb{R}^+)^K$ for an integer $1 \le K < +\infty$. Let us also define the following statistics based on a convex combination of sample quantiles:

$$\widehat{\xi}(\mathbf{p},\mathbf{c},\mathbf{B}_H^a) = \sum_{k=1}^{K} c_k\, \widehat{\xi}(p_k,\mathbf{B}_H^a), \tag{5}$$

where $c_k, k=1,\dots,K$ are positive real numbers such that $\sum_{k=1}^K c_k = 1$. For example, this corresponds to the sample median when $K=1,\mathbf{p}=1/2,\mathbf{c}=1$, to a mean of quartiles when $K=2,\mathbf{p}=(1/4,3/4),\mathbf{c}=(1/2,1/2)$. The estimation procedure is based on the following remark

$$\widehat{\xi}(\mathbf{p},\mathbf{c},(\mathbf{B}_H^a)^2) = \mathbf{E}\left(B_H^a(1)^2\right) \times \widehat{\xi}(\mathbf{p},\mathbf{c},(\widetilde{\mathbf{B}}_H^a)^2).$$

It may be expected (see Proposition 2) that, as $n\to+\infty$, $\widehat{\xi}(\mathbf{p},\mathbf{c},(\widetilde{\mathbf{B}}_H^a)^2)$ converges almost surely to $\xi_{Z^2}(\mathbf{p},\mathbf{c})$ where $\xi_{Z^2}(\mathbf{p},\mathbf{c}) = \sum_{k=1}^K c_k \xi_{Z^2}(p_k)$ and where $\xi_{Z^2}(p)$ denotes the theoretical quantile of order p of a $\chi^2(1)$ distribution. Therefore, by using the collection of dilated filters we may write

$$\log\left(\widehat{\xi}\left(\mathbf{p},\mathbf{c},\overline{(\mathbf{B}_H^{a^m})^2}\right)\right) = 2H\log(m) + \log(\gamma\times\xi_{Z^2}(\mathbf{p},\mathbf{c})) + \underbrace{\log\left(\frac{\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\widetilde{\mathbf{B}}_H^{a^m}\right)^2\right)}{\xi_{Z^2}(\mathbf{p},\mathbf{c})}\right)}_{:=\varepsilon_m^Q} \tag{6}$$

Again, the regression model (6) allows us to define a simple estimator of H as the ordinary least squares estimator defined by

$$\widehat{H}^Q = \frac{\mathbf{A}^T}{2\|\mathbf{A}\|^2}\left(\log\left(\widehat{\xi}\left(\mathbf{p},\mathbf{c},\overline{(\mathbf{B}_H^{a^m})^2}\right)\right)\right)_{m=M_1,\dots,M_2}. \tag{7}$$

3.1.2. Using trimmed means

Let us replace the convex combination of sample quantiles by $\beta-$trimmed means. By using the notation presented in Section 2.1 and the previous ideas, we have

$$\overline{\left(\mathbf{B}_H^{a^m}\right)^2}^{(\beta)} = m^{2H}\ \gamma\ \times \overline{\left(\widetilde{\mathbf{B}}_H^{a^m}\right)^2}^{(\beta)}.$$

Then, we may write the following simple linear regression model

$$\log\left(\overline{\left(\mathbf{B}_H^{a^m}\right)^2}^{(\beta)}\right) = 2H\log(m) + \log(\gamma \times \overline{Z^2}^{(\beta)}) + \underbrace{\log\left(\frac{\overline{\left(\widetilde{\mathbf{B}}_H^{a^m}\right)^2}^{(\beta)}}{\overline{Z^2}^{(\beta)}}\right)}_{:=\varepsilon_m^{TM}} \tag{8}$$

where

$$\overline{Z^2}^{(\beta)} = \frac{1}{1-\beta_2-\beta_1}\int_{\beta_1}^{1-\beta_2} \xi_{Z^2}(p)dp.$$

Since it is again expected that ε_m^{TM} converges almost surely towards 0 as $n \to +\infty$, we can define the following estimator

$$\widehat{H}^{TM} = \frac{\mathbf{A}^T}{2\|\mathbf{A}\|^2}\ \left(\log\left(\overline{\left(\mathbf{B}_H^{a^m}\right)^2}^{(\beta)}\right)\right)_{m=M_1,\ldots,M_2}. \tag{9}$$

3.2. Robustness to an additive noise

This section is aimed at defining alternatives to the standard procedure when the discretized sample path of the fractional Brownian motion is corrupted by an additive noise. One may distinguish two types of models:

- the fractional Gaussian noise is contaminated by an additive Gaussian white noise which means that the fractional Brownian motion is contaminated by an additive Brownian motion. The following equation summarizes this model denoted in the sequel by $B0$: one assumes observing

  $$X(i) = B_H(i)\ +\ \sigma\ B^{(0)}(i),$$

  where $H \neq 1/2$, $\sigma > 0$ and where $B^{(0)}(i)$ for $i = 1,\ldots,n$ is a standard Brownian motion.
- the fractional Brownian motion is contaminated by an additive Gaussian white noise. The following equation summarizes this model denoted in the sequel by $B1$: one assumes observing

  $$X(i) = B_H(i)\ +\ \sigma\ B^{(1)}(i),$$

  where $\sigma > 0$ and where $B^{(1)}(i)$ for $i = 1,\ldots,n$ are i.i.d. standard Gaussian variables.

The aim of this section is to propose an estimator of H that would be independent of C and σ, easily and quickly computable. This problem (in particular for the model $B0$) has already been undertaken by several authors: Shen et al. (2007), Baykut et al. (2007) in a wavelet context, and Coeurjolly (2000b).

3.2.1. Model B0: $X(t) = B_H(t) + \sigma B^{(0)}(t)$

Let us see how the standard procedure is affected by this contamination: since $B^{(0)}$ is a fractional Brownian motion with Hurst parameter $H = 1/2$, the variance of the filtered series of $\mathbf{X}$ is

$$\mathbf{E}\left(X^a(i)^2\right) = C^2\ \pi_H^a(0)\ +\ \sigma^2\ \pi_{1/2}^a(0),$$

which leads to

$$\mathbf{E}\left(\overline{\left(\mathbf{X}^{a^m}\right)^2}\right) = m^{2H}\gamma + m\sigma^2\pi_{1/2}^a(0).$$

Let us define $Y^{a^m}(i) = \frac{X^{a^m}(i)}{\sqrt{m}}$, then the estimation procedure is based on the following idea which is valid as soons as $H \neq 1/2$

$$\begin{aligned}\mathbf{E}\left(\overline{\left(\mathbf{Y}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{Y}^{a^m}\right)^2}\right) &= \left((2m)^{2H-1} - m^{2H-1}\right)\gamma \\ &\quad + \frac{2m}{2m}\sigma^2\pi_{1/2}^a(0) - \frac{m}{m}\sigma^2\pi_{1/2}^a(0) \\ &= m^{2H-1}\left(2^{2H-1}-1\right)\gamma.\end{aligned}$$

Now, let us consider the following regression model:

$$\log\left(\left|\overline{\left(\mathbf{Y}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{Y}^{a^m}\right)^2}\right|\right) = (2H-1)\log(m) + \log\left(\left(\left|2^{2H-1}-1\right|\right)\gamma\right) + \varepsilon_m^{B0-ST} \tag{10}$$

with

$$\varepsilon_m^{B0-ST} = \log\left(\left|\overline{\left(\mathbf{Y}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{Y}^{a^m}\right)^2}\right| \ /\ \left|\mathbf{E}\left(\overline{\left(\mathbf{Y}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{Y}^{a^m}\right)^2}\right)\right|\right),$$

which is aimed at converging towards 0. The corresponding ordinary least squares estimate is denoted by $\widehat{H}^{B0}$ and is defined by

$$\widehat{H}^{B0-ST} = \frac{1}{2} + \frac{\mathbf{A}^T}{2\|\mathbf{A}\|^2}\ \left(\log\left(\left|\overline{\left(\mathbf{Y}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{Y}^{a^m}\right)^2}\right|\right)\right)_{m=M_1,\ldots,M_2}. \tag{11}$$

This method will be denoted in the following by B0-ST. Similarly to Section 3.1, one may define two new methods denoted by B0-Q and B0-TM when the sample variance is replaced by either a convex combination of sample quantiles of the squared filtered series or by a $\beta-$trimmed mean. The two new estimators are naturally denoted by $\widehat{H}^{B0-Q}$ and $\widehat{H}^{B0-TM}$.

3.2.2. Model B1: $X(t) = B_H(t) + \sigma B^{(1)}(t)$

Let us see how the standard procedure is affected by this contamination:

$$\begin{aligned}\mathbf{E}\left(X^a(i)^2\right) &= C^2\,\pi_H^a(0) \;+\; \sigma^2 \sum_{q,r=0}^{\ell} a_q a_r \mathbf{E}\left(B^{(1)}(j-q)B^{(1)}(i+j-r)\right),\\ &= \gamma + \sigma^2 \sum_{q,r=0}^{\ell} a_q a_r \delta_{q,r}\\ &= \gamma + \sigma^2 |a|^2,\end{aligned}$$

with $|a|^2 = \sum_{q=0}^{\ell} a_q^2$. Since $|a^m|^2 = |a|^2$, this leads to

$$\mathbf{E}\left(\overline{\left(\mathbf{X}^{a^m}\right)^2}\right) = m^{2H}\gamma + \sigma^2|a|^2.$$

Therefore by using the same idea as the previous section, one may obtain the following regression model

$$\log\left(\left|\overline{\left(\mathbf{X}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{X}^{a^m}\right)^2}\right|\right) = 2H\log(m) + \log\left(\left(2^{2H}-1\right)\gamma\right) + \varepsilon_m^{B1-ST} \quad (12)$$

with

$$\varepsilon_m^{B1-ST} = \log\left(\left|\overline{\left(\mathbf{X}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{X}^{a^m}\right)^2}\right| \;/\; \mathbf{E}\left(\overline{\left(\mathbf{X}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{X}^{a^m}\right)^2}\right)\right),$$

which is aimed at converging towards 0. The corresponding ordinary least squares estimate is denoted by $\widehat{H}^{B1-ST}$ and is defined by

$$\widehat{H}^{B1-ST} = \frac{\mathbf{A}^T}{2\|\mathbf{A}\|^2}\left(\log\left(\left|\overline{\left(\mathbf{X}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{X}^{a^m}\right)^2}\right|\right)\right)_{m=M_1,\ldots,M_2}. \quad (13)$$

This method will be denoted in the following by B1-ST. Similarly, one may define two new methods denoted by B1-Q and B1-TM leading to two other estimators denoted by $\widehat{H}^{B1-Q}$ and $\widehat{H}^{B1-TM}$.

4. Summary and general result

In Sections 2 and 3, we have defined estimators of the self-similarity index based on different ideas. These estimators are referenced by Equations (4), (7), (9), (11) and (13). All these estimators exploit the property of self-similarity of the dilated-filtered initial series. They have several common points: quickly computable, definition of an estimator which is independent of the scaling coefficient and of σ^2 (in the case of an additive noise). They all are obtained by

TABLE 1
Summary of the different Hurst parameter estimation methods based on discrete variations in the presence of outliers and/or an additive noise. The first column references the name of the method while the second one defines the regressor vector used in the definition of the estimator (see (14). The vector $\mathbf{X}$ *denotes the vector of initial data*

Method	$(U^{\bullet}_{M_1,M_2})_m$
ST	$\log\left((\mathbf{X}^{a^m})^2\right)$
Q	$\log\left(\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{X}^{a^m}\right)^2\right)\right)$
TM	$\log\left(\overline{(\mathbf{X}^{a^m})^2}^{(\beta)}\right)$
B0-ST	$\log\left(\left\|\overline{\left(\mathbf{X}^{a^{2m}}\right)^2}/(2m)-\overline{(\mathbf{X}^{a^m})^2}/m\right\|\right)$
B0-Q	$\log\left(\left\|\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{X}^{a^{2m}}\right)^2\right)/(2m)-\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{X}^{a^m}\right)^2\right)/(m)\right|\right)$
B0-TM	$\log\left(\left\|\overline{\left(\mathbf{X}^{a^{2m}}\right)^2}^{(\beta)}/(2m)-\overline{(\mathbf{X}^{a^m})^2}^{(\beta)}/m\right\|\right)$
B1-ST	$\log\left(\left\|\overline{\left(\mathbf{X}^{a^{2m}}\right)^2}-\overline{(\mathbf{X}^{a^m})^2}\right\|\right)$
B1-Q	$\log\left(\left\|\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{X}^{a^{2m}}\right)^2\right)-\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{X}^{a^m}\right)^2\right)\right|\right)$
B1-TM	$\log\left(\left\|\overline{\left(\mathbf{X}^{a^{2m}}\right)^2}^{(\beta)}-\overline{(\mathbf{X}^{a^m})^2}^{(\beta)}\right\|\right)$

an ordinary least squares procedure and may be summarized by the following equation:

$$\widehat{H}^{\bullet}=\frac{\mathbf{A}^T}{2\|\mathbf{A}\|^2}U^{\bullet}_{M_1,M_2}+\theta^{\bullet},\ \text{with}\ \theta^{\bullet}=\begin{cases}1/2 & \text{if } \bullet=\text{B0-ST, B0-Q, B0-TM,}\\ 0 & \text{otherwise.}\end{cases}, \tag{14}$$

and where $U^{\bullet}_{M_1,M_2}$ is summarized in Table 1.

The next result, Proposition 1, is proved in Section A.

Proposition 1 *The following convergences hold almost surely as* $n\to+\infty$

$$\widehat{H}^{\bullet}\longrightarrow H \quad \textit{with}\ \bullet=\begin{cases}\textit{ST, Q, TM,} & \\ \textit{B0-ST, B0-Q, B0-TM,} & \\ \textit{B1-ST, B1-Q, B1-TM} & \textit{when } \mathbf{X}=\mathbf{B}_H\\ \textit{B0-ST, B0-Q, B0-TM} & \textit{when } \mathbf{X}=\mathbf{B}_H+\sigma\mathbf{B}^{(0)}\\ \textit{B1-ST, B1-Q, B1-TM} & \textit{when } \mathbf{X}=\mathbf{B}_H+\sigma\mathbf{B}^{(1)}\end{cases} \tag{15}$$

Remark 1 *We can expect that each estimate follows a central limit theorem as soon as the order of the filter is sufficiently large. It has already been proved, if* $p>H+1/4$ *(which is always true as soon as* $p\geq 2$*) and when* $\mathbf{X}=\mathbf{B}_H$*, that the estimators* $\widehat{H}^{ST}$ *(see Proposition 4 of Coeurjolly (2001)),* $\widehat{H}^{Q}$ *and* $\widehat{H}^{TM}$

(see respectively Theorems 4 and 5 of Coeurjolly (2008)) follow a central limit theorem. Asymptotic variances are also computed in these two papers. The other methods $B0-\bullet$ and $B1-\bullet$ (for $\bullet = ST, Q, TM$) under the appropriate model need more attention and work. This question will be treated in a subsequent paper.

5. Simulation study and discussion

In order to study the performance of the different estimators $\widehat{H}^{\bullet}$ for $\bullet$ =ST, Q, TM, B0-ST, B0-Q, B0-TM, B1-ST, B1-Q, B1-TM, we first simulate sample paths of pure fractional Browninan motions to control the choice of the filters and their parameters. Secondly, we use three different types of contamination to test the robustness of the chosen estimators.

In the sequel, we will denote the estimators in three different classes, the classic one which corresponds to $\widehat{H}^{\bullet}$ for $\bullet$ =ST, Q, TM, the B0-class which corresponds to $\widehat{H}^{\bullet}$ for $\bullet$ =B0-ST, B0-Q, B0-TM, and finally the B1-class which corresponds to $\widehat{H}^{\bullet}$ for $\bullet$ =B1-ST, B1-Q, B1-TM.

In the following, $\mathbf{B_H} = (B_H(1), \ldots, B_H(n))$ is a sample path of a fractional Brownian motion with Hurst parameter H and with scaling coefficient C fixed to 1. Let us note that the variance of the increments that is the variance of the fGn is thus equal to 1 ($Var(B_H(i+1) - B_H(i)) = 1$).

For each simulation, we run 500 replications with time series of length 100, 1000 and 10000. We use specific values for $\widehat{H}^{Q}$, $\mathbf{p} = 1/2$ and $\mathbf{c} = 1$, this corresponds to the sample median. We specify $\beta_1 = \beta_2 = 10\%$ for $\widehat{H}^{TM}$.

5.1. Choice of filters and their parameters

Using simulations of sample paths of pure fractional Brownian motions, we test the convergence of the proposed estimators with five different filters and parameters. The filters $i1$, $i2$ and $i3$ correspond to the increments of order 1,2 and 3 respectively. The filters $d4$, $d6$ are the Daubechies wavelet filters of order 4 and 6 respectively. For each filter, M_1 was chosen equal to 1 and M_2 was chosen equal to 2 or 5.

The tables 2, 3 and 4 (postponed to Appendix B) present the results of simulations using the estimators defined in this paper. As previously shown in Coeurjolly (2008) and by exploring the columns corresponding to a length of 10000 points in the time series, the three classic estimators $\widehat{H}^{\bullet}$ for $\bullet$ = ST, Q, TM are asymptotically without bias and converge in mean square for all the choices of filters, and for all possible values of H. The same conclusions can be written for the B0-class and B1-class of estimators.

The choice of the filters is crucial in order to minimize the variance of the estimators. Based on the tables 2, 3 and 4, we decide to choose the filters $i1, i2$ and $d4$, with $M_1 = 1$ and $M_2 = 5$ for the other simulations.

For the specific choice of the filters, the variance of the estimators can be different: there are differences within one class and between the three classes. Inside the three classes, the estimators based on the standard scheme ($\widehat{H}^{\bullet}$ for $\bullet$ = ST, B0-ST, B1-ST), have lower variance than the other estimators based on the quantiles and trimmed means. Between the three classes, the classic estimators have less variance than the estimators based on model B0 and B1.

5.2. Robustness of the estimators to contaminated models

In this section, we explore the robustness of the estimators when the simulations are not simply pure fractional Browninan motions. We consider here three different models of contamination (some examples of several contaminated sample paths are given in Appendix B):

- Model AO (additive outliers): the fGn is contaminated by an additive outlier model (*e.g.* Beran (1994), p. 130)

$$X(i+1) - X(i) = U(i)(B_H(i+1) - B_H(i)) + \sigma(1 - U(i))Z(i),$$

  where $U(i)$ are independent Bernoulli random variables with parameter $p = 0.01$ and where $Z(i)$ are i.i.d. standard Gaussian random variables. This means that the expectation of the number of contaminated observations is $n \times 1\%$. The parameter σ is chosen such that the contaminated observation achieves a given Signal Noise Rate (SNR), that is such that

$$\begin{aligned} SNR &= 10 \log_{10}\left(\frac{Var(B_H(i+1) - B_H(i))}{Var(\sigma U(i))}\right) \\ &= 10 \log_{10}\left(\frac{1}{\sigma^2}\right) \Longleftrightarrow \sigma^2 = 10^{-SNR/10}. \end{aligned}$$

  The tables 5, 6, 7 (given in Appendix B) present the results for the estimators using SNR equal to 0, -10 and -20 respectively. We observe that there are no major differences when the SNR is equal to 0, but when the SNR is equal to -10 or -20, the bias is reduced for the estimators based on the quantiles and trimmed means. Especially, the bias of the standard estimators is increasing when the SNR is decreasing. In contrast, the estimators based on the quantiles and trimmed means are less affected (in terms of bias and variances) by the noise.
  On Figure 1, we show the mean squared errors (in short MSE) on a log-log plot. This clearly illustrates that the estimators based on the quantiles and trimmed means have the lowest values of mean square error.
- Model $B0$: the fGn is assumed to be contaminated by an additive Gaussian white noise, that is

$$X(i+1) - X(i) = B_H(i+1) - B_H(i) + \sigma B^{(1)}(i) \Longleftrightarrow X(i) = B_H(i) + \sigma B^{(0)}(i),$$

where $B^{(0)} = B(\cdot)$ is a standard Brownian motion and where $B^{(1)}(i) = B^{(0)}(i+1) - B^{(0)}(i)$. Hence, $B^{(1)}(i)$ are i.i.d. standard Gaussian variables. The parameter σ is chosen such that the increments of $X(\cdot)$ achieve a given Signal Noise Rate (SNR) such that

$$\begin{aligned} SNR &= 10\log_{10}\left(\frac{Var(B_H(i+1)-B_H(i))}{Var(\sigma B^{(1)}(i))}\right) \\ &= 10\log_{10}\left(\frac{1}{\sigma^2}\right) \Longleftrightarrow \sigma^2 = 10^{-SNR/10}. \end{aligned}$$

The tables 8, 9 (see Appendix B), present the results for the estimators using SNR equal to 0, 10 respectively. Under a contamination by model B0, the bias is increasing for all the estimators. When looking at the MSE, figure 2, only the MSE of the estimators $\widehat{H}^{\bullet}$ for $\bullet$ = B0-ST, B0-Q, B0-TM) seems to converge to 0. For the other class of estimators, the MSE does not seem to converge to 0. We always remark that the estimator $\widehat{H}^{B0-ST}$ is better than $\widehat{H}^{B0-TM}$ which is better than $\widehat{H}^{B0-Q}$.

- Model $B1$: the sample path of a fBm is assumed to be contaminated by an additive Gaussian white noise, that is

$$X(i) = B_H(i) + \sigma B^{(1)}(i)$$

$$\Longleftrightarrow X(i+1) - X(i) = B_H(i+1) - B_H(i) + \sigma(B^{(1)}(i+1) - B^{(1)}(i)).$$

where $B^{(1)}(i)$ are i.i.d. standard Gaussian random variables. Again, the parameter σ is chosen such that the increments of $X(\cdot)$ achieve a given Signal Noise Rate (SNR) such that

$$\begin{aligned} SNR &= 10\log_{10}\left(\frac{Var(B_H(i+1)-B_H(i))}{Var(\sigma(B^{(1)}(i+1)-B^{(1)}(i))}\right) \\ &= 10\log_{10}\left(\frac{1}{2\sigma^2}\right) \Longleftrightarrow \sigma^2 = \frac{10^{-SNR/10}}{2}. \end{aligned}$$

The tables 10, 11 (see Appendix B), present the results for the estimators using SNR equal to 0, 10 respectively. Under a contamination by model B1, the bias is increasing for all the estimators. When looking at the MSE, figure 3, only the MSE of the estimators $\widehat{H}^{\bullet}$ for $\bullet$ = B1-ST, B1-Q, B1-TM) seems to converge to 0. For the other class of estimators, the MSE does not seem to converge to 0. We always remark that the estimator $\widehat{H}^{B1-ST}$ is better than $\widehat{H}^{B1-TM}$ which is better than $\widehat{H}^{B1-Q}$.

5.3. General discussion and recommendations

In this paper, we review different estimation procedures of Hurst parameter, H of fractional Brownian motions, using dicrete variations of time series. Our aim was to provide estimators that were *quickly computable* and *independent* on

other parameter such as the scaling coefficient or parameters related to the contamination. This is done in this paper by strongly exploiting the self-similarity property of dilated discrete variations of the fractional Brownian motion. We then describe several methods of estimation of H:

1. the standard procedure based on the log-linearity of the variogram of dilated time series (ST).
2. robust alternatives to outliers using sample quantiles (Q) or trimmed means (TM).
3. robust alternatives to additive Gaussian white noise or to additive Brownian motion (methods $B0$ and $B1$).
4. robust alternatives to outliers and additive noise by combining these methods.

We also study, from a practical point of view, the robustness of the methods using three different models of contamination:

1. a model of additive outliers (AO).
2. a model of additive Gaussian white noise to the fBm (B0).
3. a model of additive Gaussian white noise to the fGn (B1).

Table 1 summarizes these different procedures. All these procedures are implemented in the `R` package `dvfbm`. This package provides the code to estimate the Hurst parameters described in the paper. It also provides the code to proceed to the contamination of the fractional Brownian motions.

Concerning the different internal parameters of the different methods, we recommend to use the methods based on quantiles with $\mathbf{p} = 1/2$ and $\mathbf{c} = 1$, the one based on trimmed means with $\beta_1 = \beta_2 = 10\%$ for $\widehat{H}^{TM}$, the general filter $a = d4$ corresponding to the wavelet Daubechies filter with two zero moments and $M = 5$ or 10 for the number of dilated filters.

In practice, for real data, we recommend to first observe the data for the presence of outliers. In this case, the use of estimations based on trimmed means (TM) (which seems to have slight better properties than the one based on quantiles (Q)) should be considered.

Concerning robustness to additive noise (models $B0$ or $B1$), we have shown that the appropriate procedures work well for $n \geq 10000$. We also recommend to use the standard method if we observe that the differences $|\widehat{H}^{ST} - \widehat{H}^{B0-ST}|$ and $|\widehat{H}^{ST} - \widehat{H}^{B1-ST}|$ are close to zero.

We plan in a future work to propose a procedure for choosing the best appropriate model of additive noise. We also plan to develop bootstrap methods in order to evaluate the variance of the estimators for real data.

As it is done in Coeurjolly (2008) for the methods Q and TM, we can expect that all these methods are appropriate for estimating the Hurst exponent of locally self-similar Gaussian process. Another research perspective could be to extend such methods to a non-Gaussian setting. The work of Chan and Wood (2004) may provide a thorough basis.

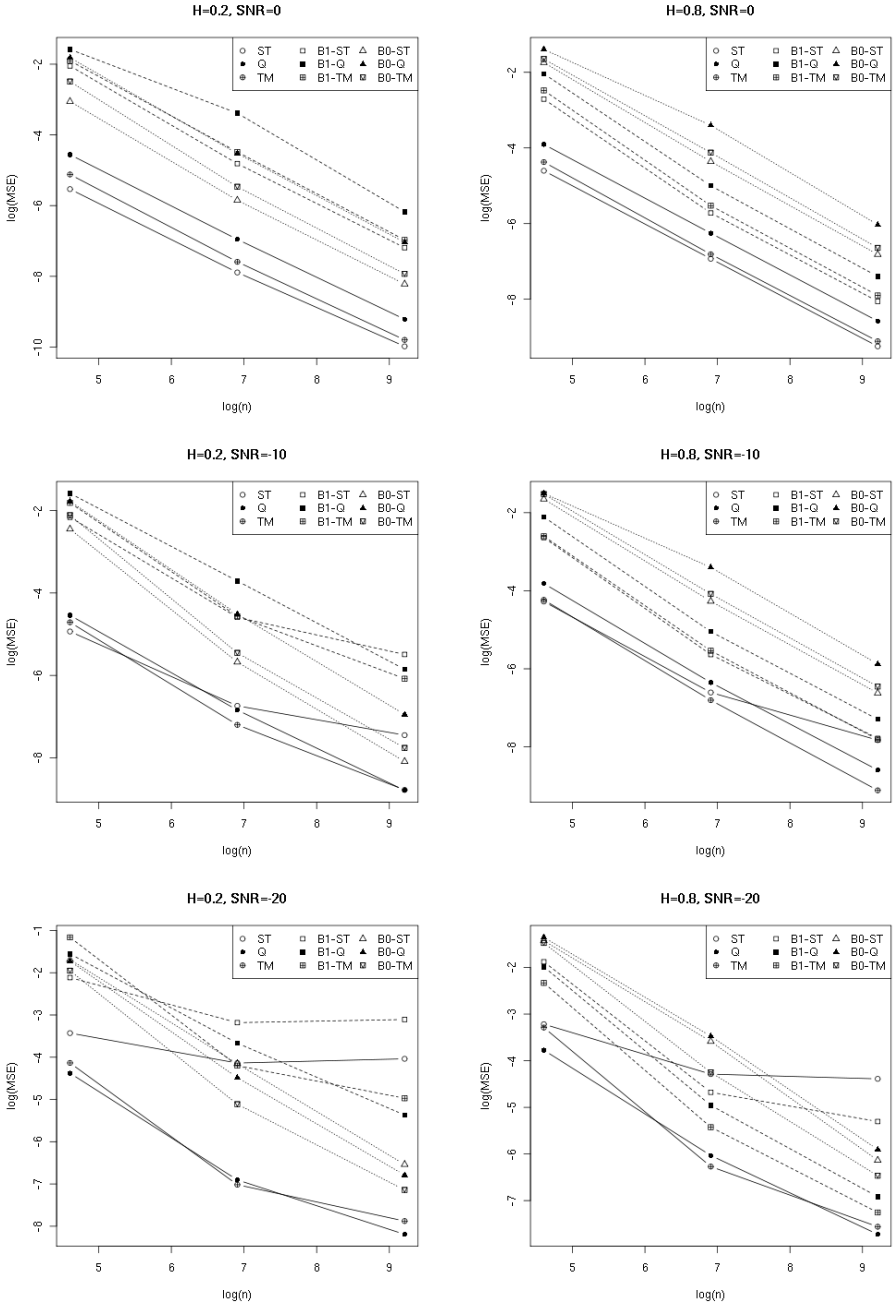


FIG 1. *Summary of Tables 5, 6 and 7 via plots of empirical MSE in terms of n (for* $n = 100, 1000, 10000$ *in log-log scales) for the AO model for the nine methods for* $SNR = 0, -10, -20$ *and with the optimal filter* $a_{opt} = i1$ *for* $H = 0.2$ *and* $a_{opt} = d4$ *for* $H = 0.8$*, with* $M_1 = 1$ *and* $M_2 = 5$.

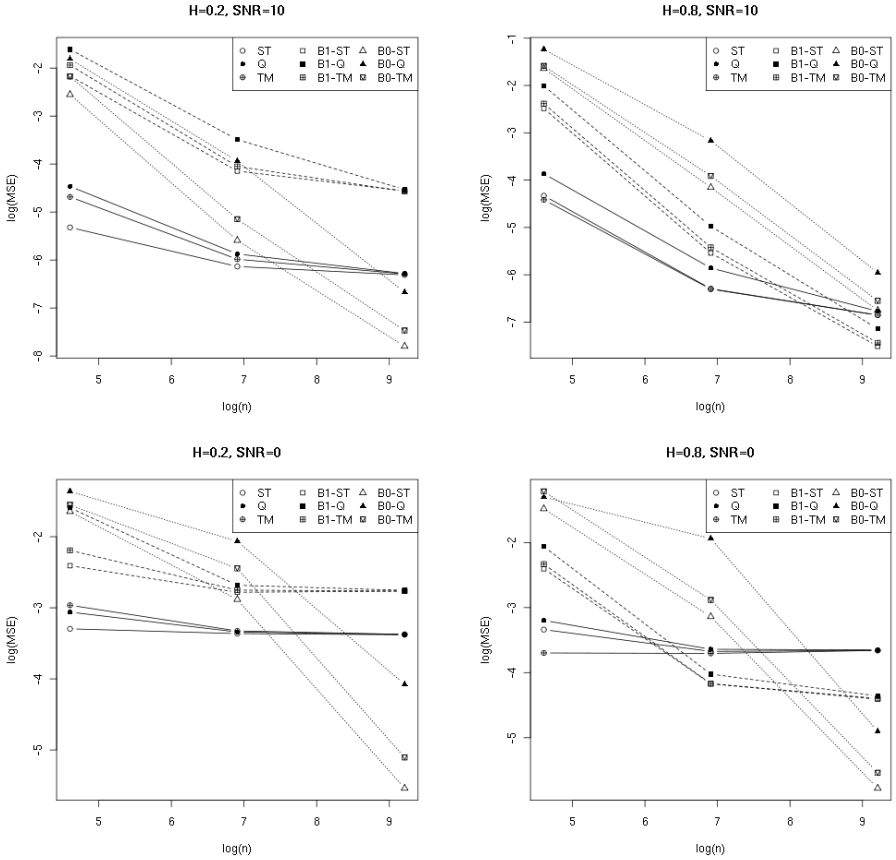


FIG 2. *Summary of Tables 10 and 11 via plots of empirical MSE in terms of n (for* $n = 100, 1000, 10000$ *in log-log scales) for the model B0 for the nine methods, for* $SNR = 10, 0$ *and with the optimal filter* $a_{opt} = i1$ *for* $H = 0.2$ *and* $a_{opt} = d4$ *for* $H = 0.8$ *with* $M_1 = 1$ *and* $M_2 = 5$.

Appendix A: Consistency of the different procedures

First of all, we leave the reader to verify that for all the methods presented in Sections 2 and 3, the variables $\varepsilon_m^\bullet$ have been defined in such a way that:

$$\widehat{H}^\bullet - H = \frac{\mathbf{A}^T}{2\|\mathbf{A}\|^2} (\varepsilon_m^\bullet)_{m=M_1,\dots,M_2}$$

Now, let us present some general result.

Proposition 2 *Under the previous notation, let* $\mathbf{X}$ *denote either* $\mathbf{B}_H$, $\mathbf{B}_H + \sigma\mathbf{B}^{(0)}$ *or* $\mathbf{B}_H + \sigma\mathbf{B}^{(1)}$, *then for any filter a of order* $p \geq 1$ *and for any* $H \in (0, 1)$,

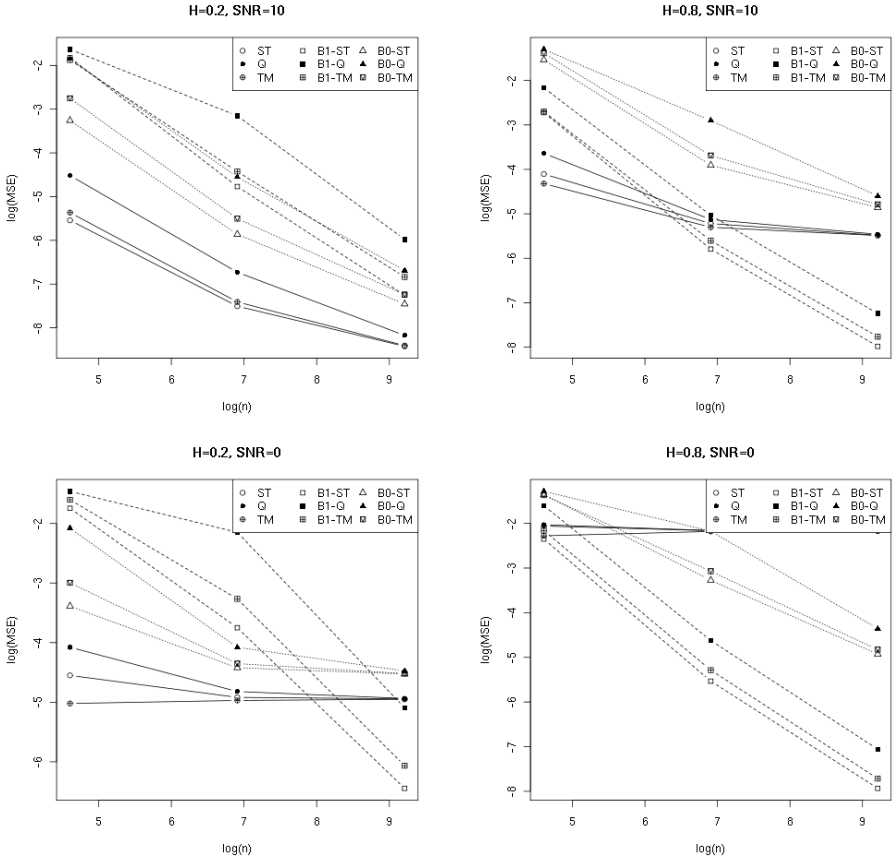


FIG 3. *Summary of Tables 8 and 9 via plots of empirical MSE in terms of n (for* $n = 100, 1000, 10000$ *in log-log scales) for the model B1 for the nine methods, for* $SNR = 10, 0$ *and with the optimal filter* $a_{opt} = i1$ *for* $H = 0.2$ *and* $a_{opt} = d4$ *for* $H = 0.8$ *with* $M_1 = 1$ *and* $M_2 = 5$.

we have the following almost sure convergences as $n \to +\infty$

$$\overline{(\mathbf{X}^a)^2} \longrightarrow \mathbf{E}\left(X^a(1)^2\right) \tag{16}$$

$$\widehat{\xi}\left(\mathbf{p}, \mathbf{c}, (\mathbf{X}^a)^2\right) \longrightarrow \mathbf{E}\left(X^a(1)^2\right) \times \xi_{Z^2}(\mathbf{p}, \mathbf{c}) \tag{17}$$

$$\overline{(\mathbf{X}^a)^2}^{(\beta)} \longrightarrow \mathbf{E}\left(X^a(1)^2\right) \times \overline{Z^2}^{(\beta)} \tag{18}$$

Proof.

- Model $\mathbf{X} = \mathbf{B}_H$: the proof of (16) can be found in Coeurjolly (2001) (Proposition 1) while the proofs of (17) and (18) can be found in Coeurjolly (2008) (Theorem 4 and 5).

- Model $\mathbf{X} = \mathbf{B}_H + \sigma\mathbf{B}^{(0)}$: for such a model the covariance function of $\mathbf{X}^a$ is given by

$$\mathbf{E}\left(X^a(j)X^a(i+j)\right) = C^2 \times \pi_H^a(i) + \sigma^2 \times \pi_{1/2}^a(i) \sim |i|^{2H-2p}, \text{ as } |i| \to +\infty,$$

since $\pi_{1/2}^a(i)$, which is nothing else than the covariance function of a filtered Brownian motion, vanishes for $|i| > \ell$. By following the proof of Proposition 1 of Coeurjolly (2001), we have to prove that for any filter and any $H \in (0,1)$,

$$\mathbf{E}\left(\overline{\left(\mathbf{X}^a\right)^2}\right) = o(1),$$

as $n \to +\infty$. The result (16) is then ensured by Theorem 6.2 of Doob (1953) (p. 492).
Now, since $\widetilde{\mathbf{X}}^a$ is a Gaussian sequence with correlation function decreasing hyperbolically, Theorem 2 of Coeurjolly (2008) (resp. Theorem 3 (with $g(\cdot) = (\cdot)^2$) ensures that a Bahadur representation (resp. an uniform Bahadur representation) holds for the sample quantile $\widehat{\xi}\left(p, \left(\widetilde{\mathbf{X}}^a\right)^2\right)$ for some $p \in (0,1)$ (resp. $\sup_{p_0 \le p \le p_1} \widehat{\xi}\left(p, \left(\widetilde{\mathbf{X}}^a\right)^2\right)$ for $0 < p_0 < p_1 < 1$. Then, following the proofs of Theorems 4 and 5 of Coeurjolly (2008) (devoted to the case $\mathbf{X} = \mathbf{B}_H$, we can obtain the results (17) and (18)).
- Model $\mathbf{X} = \mathbf{B}_H + \sigma\mathbf{B}^{(1)}$: the proof is quite similar to the previous one. The covariance function of $\mathbf{X}^a$ is given by

$$\mathbf{E}\left(X^a(j)X^a(i+j)\right) = C^2 \times \pi_H^a(i) \; + \; \sigma^2 \times \sum_{q,r=0}^{\ell} a_q a_r \delta_{q,r-i} \sim |i|^{2H-2p},$$

since the second term vanishes when $|i| > \ell$.

■

Now, let us prove Proposition 1.

Proof. The cases $\widehat{H}^\bullet$ for $\bullet$ = ST, Q, TM when $\mathbf{X} = \mathbf{B}_H$ have already been obtained in Coeurjolly (2001) and Coeurjolly (2008). Since a pure fractional Brownian is a fractional Brownian motion contaminated by an additive noise ($B^{(0)}$ or $(B^{(1)})$ with $\sigma = 0$, we just have to consider the two last cases of (15).

- Model $\mathbf{X} = \mathbf{B}_H + \sigma\mathbf{B}^{(0)}$: for any filter a of order $p \ge 1$, for any $H \in 0,1)$ and any $m \ge 1$ we have, from Proposition 2, as $n \to +\infty$

$$\begin{aligned}
\overline{\left(\mathbf{Y}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{Y}^{a^m}\right)^2} &= \frac{\overline{\left(\mathbf{X}^{a^{2m}}\right)^2}}{2m} - \frac{\overline{\left(\mathbf{X}^{a^m}\right)^2}}{m} \\
&\xrightarrow{a.s.} \frac{\mathbf{E}\left(\overline{\left(\mathbf{X}^{a^{2m}}\right)^2}\right)}{2m} - \frac{\mathbf{E}\left(\overline{\left(\mathbf{X}^{a^m}\right)^2}\right)}{m} \\
&= m^{2H-1}\left(2^{2H-1}-1\right)\gamma.
\end{aligned}$$

Therefore, as $n \to +\infty$

$$\begin{aligned}\left(\mathbf{U}_{M1,M2}^{B0-ST}\right)_m &:= \log\left(\left|\overline{\left(\mathbf{Y}^{a^{2m}}\right)^2} - \overline{\left(\mathbf{Y}^{a^m}\right)^2}\right|\right) \\ &\xrightarrow{a.s.} (2H-1)\log(m) + \log\left(\left|2^{2H-1}-1\right|\gamma\right)\end{aligned}$$

and so

$$\frac{\mathbf{A}^T\mathbf{U}_{m1,M_2}^{B0-ST}}{2\|\mathbf{A}\|^2} + \frac{1}{2} \xrightarrow{a.s.} (2H-1)\frac{\mathbf{A}^T\mathbf{A}}{2\|\mathbf{A}\|^2} + \frac{1}{2} = H,$$

since for $\mathbf{1} = (1,\ldots,1)^T$, $\mathbf{A}^T\mathbf{1} = 0$. The estimators $\widehat{H}^{B0-Q}$ and $\widehat{H}^{B0-TM}$ follow the same ideas. Consider the first one for example, we have from Proposition 2

$$\begin{aligned}&\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{Y}^{a^{2m}}\right)^2\right) - \widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{Y}^{a^m}\right)^2\right) \\ &= \widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{X}^{a^{2m}}\right)^2/(2m)\right) - \widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{X}^{a^m}\right)^2/m\right) \\ &= \frac{\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{X}^{a^{2m}}\right)^2\right)}{2m} - \frac{\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{Y}^{a^m}\right)^2\right)}{m} \\ &\xrightarrow{a.s.} \left(m^{2H-1}\left(2^{2H-1}-1\right)\gamma\right)\xi_{Z^2}(\mathbf{p},\mathbf{c}).\end{aligned}$$

Therefore, as $n \to +\infty$

$$\begin{aligned}\left(\mathbf{U}_{M1,M2}^{B0-Q}\right)_m &:= \log\left(\left|\widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{Y}^{a^{2m}}\right)^2\right) - \widehat{\xi}\left(\mathbf{p},\mathbf{c},\left(\mathbf{Y}^{a^m}\right)^2\right)\right|\right) \\ &\xrightarrow{a.s.} (2H-1)\log(m) + \log\left(\left|2^{2H-1}-1\right|\gamma \times \xi_{Z^2}(\mathbf{p},\mathbf{c})\right)\end{aligned}$$

and so

$$\frac{\mathbf{A}^T\mathbf{U}_{m1,M_2}^{B0-Q}}{2\|\mathbf{A}\|^2} + \frac{1}{2} \xrightarrow{a.s.} (2H-1)\frac{\mathbf{A}^T\mathbf{A}}{2\|\mathbf{A}\|^2} + \frac{1}{2} = H.$$

- Model $\mathbf{X} = \mathbf{B}_H + \sigma\mathbf{B}^{(1)}$: the proof is omitted since it follows exactly the same ideas as the previous model.

■

Appendix B: Details of the simulation results

In this section, we give, for a better reading of the results, the details of the different simulation results that have been discussed in Section 5 and that have been summarized through Figures 1, 2 and 3.

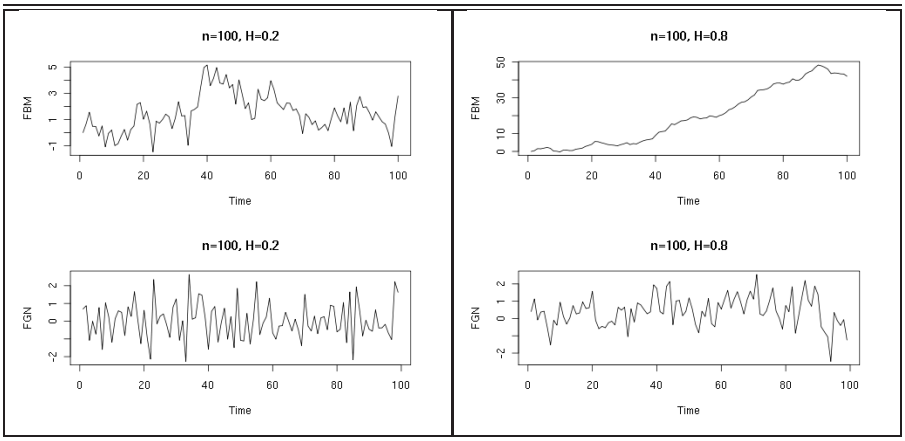


FIG 4. *Example of (pure) fractional Brownian motions with Hurst parameters* $H = 0.2$ *(left) and* $H = 0.8$ *(right) and length* $n = 100$.

TABLE 2

$m = 500$ *replications of a (pure) fractional Brownian motion for* $n = 100, 1000, 10000$ *and* $H = 0.2, 0.8$ *for different filters and different values of* $M2$ *for the methods ST, Q and TM*

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ST(i1,2)	0.195	0.095	0.199	0.029	0.200	0.009	0.779	0.062	0.796	0.024	0.800	0.010
ST(i2,2)	0.189	0.155	0.199	0.048	0.201	0.015	0.789	0.127	0.799	0.037	0.799	0.012
ST(d4,2)	0.195	0.120	0.199	0.036	0.201	0.011	0.786	0.109	0.799	0.033	0.800	0.011
ST(i3,2)	0.184	0.195	0.200	0.059	0.201	0.018	0.796	0.173	0.799	0.053	0.799	0.017
ST(d6,2)	0.195	0.132	0.198	0.039	0.201	0.013	0.788	0.139	0.800	0.040	0.799	0.013
ST(i1,5)	0.201	0.059	0.199	0.019	0.200	0.006	0.774	0.072	0.795	0.028	0.800	0.012
ST(i2,5)	0.202	0.078	0.199	0.025	0.200	0.008	0.784	0.103	0.799	0.031	0.800	0.010
ST(d4,5)	0.201	0.072	0.199	0.023	0.200	0.007	0.783	0.106	0.799	0.031	0.800	0.010
ST(i3,5)	0.201	0.090	0.199	0.028	0.200	0.009	0.787	0.119	0.800	0.037	0.800	0.011
ST(d6,5)	0.202	0.079	0.199	0.024	0.200	0.007	0.783	0.122	0.799	0.035	0.800	0.011
Q(i1,2)	0.187	0.213	0.201	0.064	0.202	0.021	0.809	0.151	0.800	0.050	0.801	0.017
Q(i2,2)	0.181	0.238	0.204	0.075	0.201	0.023	0.802	0.221	0.799	0.065	0.800	0.021
Q(d4,2)	0.197	0.218	0.205	0.068	0.200	0.021	0.799	0.210	0.798	0.062	0.799	0.020
Q(i3,2)	0.160	0.287	0.200	0.088	0.200	0.028	0.813	0.266	0.795	0.079	0.798	0.026
Q(d6,2)	0.182	0.229	0.202	0.070	0.201	0.021	0.821	0.229	0.799	0.072	0.801	0.023
Q(i1,5)	0.204	0.099	0.201	0.031	0.200	0.010	0.805	0.122	0.801	0.044	0.801	0.016
Q(i2,5)	0.208	0.111	0.201	0.037	0.200	0.011	0.801	0.141	0.801	0.041	0.799	0.014
Q(d4,5)	0.208	0.111	0.201	0.034	0.200	0.010	0.799	0.149	0.800	0.041	0.800	0.014
Q(i3,5)	0.212	0.122	0.201	0.039	0.200	0.012	0.804	0.159	0.801	0.047	0.800	0.015
Q(d6,5)	0.210	0.115	0.201	0.037	0.200	0.011	0.802	0.160	0.800	0.045	0.800	0.015
TM(i1,2)	0.191	0.116	0.199	0.037	0.201	0.011	0.791	0.073	0.799	0.029	0.800	0.011
TM(i2,2)	0.181	0.174	0.199	0.053	0.201	0.017	0.794	0.142	0.799	0.044	0.799	0.014
TM(d4,2)	0.185	0.141	0.199	0.042	0.200	0.013	0.791	0.123	0.799	0.038	0.799	0.013
TM(i3,2)	0.173	0.217	0.200	0.065	0.201	0.020	0.796	0.193	0.797	0.059	0.799	0.019
TM(d6,2)	0.182	0.152	0.199	0.046	0.201	0.014	0.786	0.156	0.799	0.047	0.800	0.015
TM(i1,5)	0.234	0.065	0.202	0.022	0.200	0.007	0.820	0.080	0.802	0.032	0.801	0.013
TM(i2,5)	0.234	0.085	0.202	0.028	0.200	0.008	0.824	0.111	0.802	0.034	0.800	0.011
TM(d4,5)	0.242	0.079	0.202	0.025	0.200	0.008	0.834	0.114	0.803	0.033	0.800	0.011
TM(i3,5)	0.241	0.096	0.202	0.031	0.201	0.009	0.833	0.129	0.803	0.039	0.800	0.012
TM(d6,5)	0.248	0.087	0.203	0.027	0.200	0.008	0.836	0.130	0.803	0.037	0.800	0.012

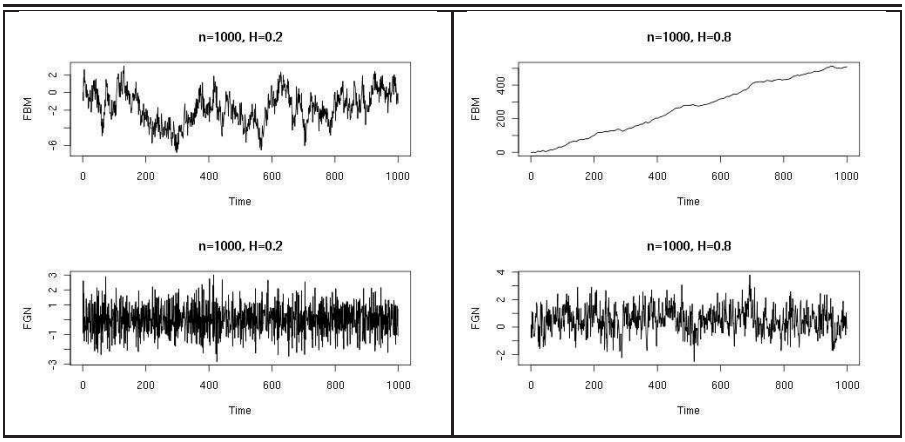


FIG 5. *Example of a (pure) fractional Brownian motions with Hurst parameters* $H = 0.2$ *(left) and* $H = 0.8$ *(right) and length* $n = 1000$.

TABLE 3

$m = 500$ *replications of a (pure) fractional Brownian motion for* $n = 100, 1000, 10000$ *and* $H = 0.2, 0.8$ *for different filters and different values of* $M2$ *for the methods B0-ST, B0-Q and B0-TM*

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
B0-ST(i1,2)	0.163	0.396	0.199	0.109	0.202	0.034	0.725	0.291	0.791	0.052	0.799	0.021
B0-ST(i2,2)	0.114	0.986	0.200	0.197	0.203	0.061	0.750	0.802	0.798	0.163	0.801	0.049
B0-ST(d4,2)	0.146	0.636	0.196	0.140	0.203	0.044	0.733	0.671	0.795	0.126	0.801	0.040
B0-ST(i3,2)	0.137	1.092	0.203	0.259	0.203	0.079	0.717	1.033	0.800	0.257	0.800	0.074
B0-ST(d6,2)	0.154	0.750	0.195	0.154	0.203	0.049	0.689	0.903	0.792	0.176	0.801	0.056
B0-ST(i1,5)	0.186	0.216	0.197	0.051	0.200	0.016	0.690	0.275	0.786	0.060	0.799	0.023
B0-ST(i2,5)	0.187	0.390	0.194	0.083	0.201	0.026	0.763	0.417	0.786	0.125	0.801	0.035
B0-ST(d4,5)	0.171	0.317	0.196	0.068	0.200	0.021	0.733	0.422	0.788	0.122	0.800	0.033
B0-ST(i3,5)	0.207	0.401	0.191	0.103	0.201	0.031	0.797	0.443	0.773	0.173	0.801	0.045
B0-ST(d6,5)	0.194	0.367	0.195	0.076	0.200	0.023	0.781	0.428	0.781	0.158	0.800	0.041
B0-Q(i1,2)	0.116	1.073	0.205	0.271	0.205	0.088	0.774	0.894	0.804	0.184	0.800	0.059
B0-Q(i2,2)	0.081	1.089	0.211	0.343	0.202	0.098	0.863	1.041	0.803	0.303	0.795	0.088
B0-Q(d4,2)	0.134	1.169	0.219	0.283	0.202	0.087	0.798	1.009	0.803	0.286	0.802	0.082
B0-Q(i3,2)	0.051	1.264	0.196	0.432	0.199	0.123	0.790	1.177	0.814	0.390	0.804	0.113
B0-Q(d6,2)	0.097	1.046	0.203	0.294	0.203	0.087	0.717	1.078	0.797	0.349	0.793	0.097
B0-Q(i1,5)	0.193	0.409	0.199	0.102	0.202	0.030	0.810	0.395	0.794	0.128	0.803	0.041
B0-Q(i2,5)	0.217	0.410	0.199	0.137	0.199	0.037	0.863	0.469	0.780	0.193	0.801	0.050
B0-Q(d4,5)	0.223	0.433	0.201	0.119	0.199	0.033	0.851	0.499	0.777	0.206	0.801	0.051
B0-Q(i3,5)	0.251	0.405	0.194	0.161	0.200	0.043	0.906	0.487	0.771	0.239	0.802	0.059
B0-Q(d6,5)	0.225	0.419	0.196	0.129	0.201	0.034	0.899	0.482	0.768	0.249	0.799	0.056
B0-TM(i1,2)	0.142	0.584	0.199	0.142	0.203	0.044	0.723	0.359	0.794	0.072	0.801	0.026
B0-TM(i2,2)	0.077	1.041	0.201	0.223	0.201	0.069	0.714	0.821	0.798	0.189	0.800	0.057
B0-TM(d4,2)	0.088	0.772	0.198	0.162	0.201	0.052	0.783	0.819	0.798	0.160	0.801	0.049
B0-TM(i3,2)	0.038	1.194	0.204	0.288	0.201	0.088	0.786	1.036	0.811	0.283	0.800	0.082
B0-TM(d6,2)	-0.005	0.893	0.193	0.180	0.202	0.057	0.825	0.888	0.802	0.203	0.799	0.065
B0-TM(i1,5)	0.163	0.282	0.195	0.064	0.200	0.019	0.724	0.324	0.794	0.074	0.801	0.028
B0-TM(i2,5)	0.173	0.385	0.192	0.096	0.201	0.029	0.809	0.423	0.789	0.141	0.801	0.040
B0-TM(d4,5)	0.164	0.368	0.196	0.079	0.200	0.024	0.777	0.474	0.788	0.141	0.800	0.038
B0-TM(i3,5)	0.211	0.381	0.191	0.117	0.201	0.034	0.850	0.464	0.772	0.203	0.801	0.049
B0-TM(d6,5)	0.175	0.405	0.195	0.088	0.200	0.026	0.853	0.458	0.783	0.170	0.800	0.045

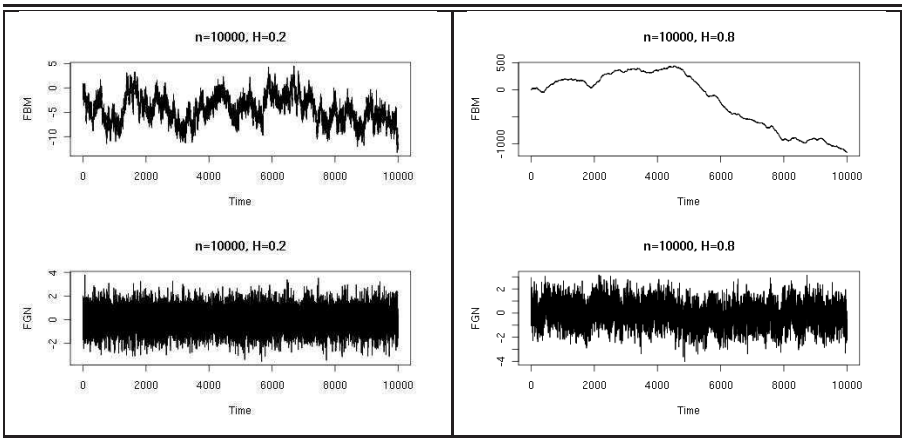


FIG 6. *Example of (pure) fractional Brownian motions with Hurst parameters $H = 0.2$ (left) and $H = 0.8$ (right) and length $n = 10000$.*

TABLE 4
$m = 500$ replications of a (pure) fractional Brownian motion for $n = 100, 1000, 10000$ and $H = 0.2, 0.8$ for different filters and different values of $M2$ for the methods B1-ST, B1-Q and B1-TM

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
B1-ST(i1,2)	0.248	0.794	0.198	0.173	0.197	0.052	0.765	0.110	0.794	0.035	0.800	0.015
B1-ST(i2,2)	0.230	1.118	0.196	0.318	0.196	0.092	0.772	0.252	0.800	0.075	0.800	0.023
B1-ST(d4,2)	0.202	1.004	0.202	0.223	0.195	0.066	0.772	0.229	0.799	0.062	0.800	0.020
B1-ST(i3,2)	0.241	1.221	0.185	0.451	0.196	0.120	0.763	0.375	0.802	0.110	0.800	0.033
B1-ST(d6,2)	0.204	1.045	0.204	0.244	0.195	0.073	0.766	0.310	0.798	0.082	0.800	0.026
B1-ST(i1,5)	0.144	0.402	0.199	0.086	0.199	0.028	0.740	0.155	0.792	0.041	0.799	0.018
B1-ST(i2,5)	0.192	0.444	0.199	0.139	0.197	0.042	0.753	0.239	0.796	0.060	0.801	0.018
B1-ST(d4,5)	0.163	0.450	0.199	0.113	0.198	0.035	0.743	0.249	0.797	0.061	0.800	0.018
B1-ST(i3,5)	0.221	0.440	0.194	0.183	0.195	0.050	0.721	0.321	0.794	0.071	0.801	0.022
B1-ST(d6,5)	0.196	0.472	0.198	0.127	0.198	0.038	0.702	0.351	0.796	0.070	0.801	0.021
B1-Q(i1,2)	0.249	1.170	0.199	0.451	0.194	0.133	0.801	0.297	0.803	0.085	0.800	0.029
B1-Q(i2,2)	0.311	1.237	0.175	0.644	0.197	0.150	0.786	0.466	0.804	0.127	0.798	0.038
B1-Q(d4,2)	0.314	1.194	0.167	0.492	0.196	0.133	0.783	0.440	0.803	0.121	0.801	0.037
B1-Q(i3,2)	0.362	1.257	0.201	0.758	0.201	0.189	0.767	0.648	0.808	0.155	0.801	0.049
B1-Q(d6,2)	0.364	1.131	0.199	0.534	0.196	0.132	0.754	0.506	0.801	0.138	0.797	0.043
B1-Q(i1,5)	0.231	0.437	0.191	0.190	0.196	0.048	0.768	0.277	0.802	0.070	0.802	0.026
B1-Q(i2,5)	0.233	0.480	0.188	0.235	0.198	0.059	0.783	0.307	0.798	0.081	0.801	0.025
B1-Q(d4,5)	0.259	0.466	0.189	0.200	0.200	0.052	0.779	0.342	0.797	0.087	0.801	0.026
B1-Q(i3,5)	0.243	0.484	0.191	0.289	0.197	0.067	0.784	0.363	0.798	0.091	0.802	0.028
B1-Q(d6,5)	0.307	0.463	0.188	0.233	0.198	0.053	0.770	0.389	0.797	0.092	0.800	0.028
B1-TM(i1,2)	0.238	0.918	0.196	0.224	0.195	0.066	0.776	0.139	0.798	0.043	0.801	0.017
B1-TM(i2,2)	0.231	1.264	0.191	0.370	0.199	0.104	0.767	0.286	0.800	0.085	0.800	0.026
B1-TM(d4,2)	0.276	1.105	0.199	0.259	0.199	0.078	0.793	0.267	0.800	0.074	0.800	0.023
B1-TM(i3,2)	0.297	1.193	0.177	0.571	0.199	0.134	0.788	0.407	0.807	0.119	0.800	0.036
B1-TM(d6,2)	0.362	1.091	0.209	0.284	0.198	0.085	0.808	0.343	0.803	0.093	0.800	0.030
B1-TM(i1,5)	0.179	0.390	0.200	0.104	0.199	0.032	0.769	0.167	0.798	0.048	0.801	0.020
B1-TM(i2,5)	0.232	0.455	0.197	0.165	0.197	0.046	0.783	0.249	0.799	0.065	0.801	0.020
B1-TM(d4,5)	0.218	0.466	0.197	0.129	0.198	0.040	0.770	0.269	0.798	0.068	0.800	0.020
B1-TM(i3,5)	0.222	0.460	0.189	0.229	0.196	0.055	0.752	0.331	0.796	0.077	0.801	0.024
B1-TM(d6,5)	0.239	0.495	0.196	0.148	0.198	0.042	0.755	0.349	0.798	0.076	0.800	0.023

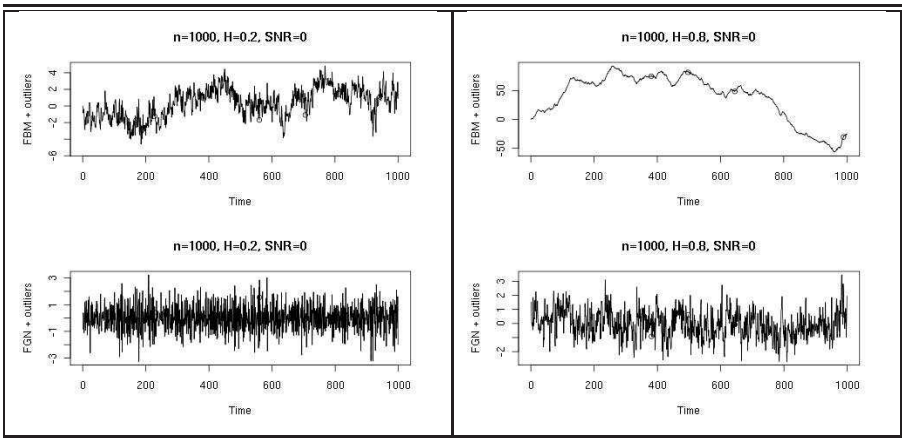


FIG 7. *Examples of contaminated fractional Brownian motions (model AO with a $SNR = 0$) with Hurst parameters $H = 0.2$ (left) and $H = 0.8$ (right) and length $n = 1000$.*

TABLE 5
$m = 500$ replications of a contaminated fractional Brownian motion (model AO with a $SNR = 0$) for $n = 100, 1000, 10000$ and $H = 0.2, 0.8$ for different filters and different values of $M2$

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ST(i1,5)	0.201	0.063	0.202	0.019	0.203	0.006	0.771	0.072	0.793	0.027	0.798	0.012
ST(i2,5)	0.200	0.081	0.201	0.025	0.202	0.008	0.786	0.098	0.796	0.031	0.798	0.010
ST(d4,5)	0.200	0.077	0.201	0.023	0.202	0.007	0.787	0.099	0.796	0.031	0.798	0.010
Q(i1,5)	0.209	0.102	0.203	0.031	0.202	0.010	0.792	0.123	0.798	0.042	0.800	0.017
Q(i2,5)	0.205	0.119	0.200	0.037	0.202	0.011	0.802	0.134	0.799	0.043	0.799	0.013
Q(d4,5)	0.211	0.110	0.201	0.035	0.202	0.011	0.808	0.142	0.799	0.044	0.799	0.014
TM(i1,5)	0.235	0.069	0.205	0.022	0.203	0.007	0.816	0.079	0.799	0.030	0.799	0.014
TM(i2,5)	0.232	0.086	0.203	0.028	0.202	0.009	0.825	0.104	0.800	0.033	0.799	0.010
TM(d4,5)	0.240	0.081	0.204	0.025	0.202	0.008	0.836	0.106	0.801	0.033	0.799	0.011
B0-ST(i1,5)	0.174	0.216	0.197	0.054	0.200	0.016	0.677	0.319	0.784	0.058	0.797	0.023
B0-ST(i2,5)	0.175	0.369	0.195	0.085	0.201	0.025	0.757	0.434	0.790	0.119	0.799	0.035
B0-ST(d4,5)	0.160	0.305	0.196	0.071	0.200	0.021	0.727	0.412	0.790	0.113	0.798	0.033
B0-Q(i1,5)	0.196	0.403	0.197	0.104	0.200	0.030	0.803	0.409	0.794	0.121	0.799	0.041
B0-Q(i2,5)	0.194	0.412	0.190	0.137	0.200	0.037	0.878	0.475	0.788	0.188	0.799	0.049
B0-Q(d4,5)	0.209	0.440	0.190	0.116	0.201	0.034	0.848	0.496	0.785	0.182	0.799	0.049
B0-TM(i1,5)	0.143	0.282	0.196	0.065	0.200	0.019	0.720	0.347	0.788	0.070	0.798	0.027
B0-TM(i2,5)	0.164	0.364	0.193	0.095	0.200	0.028	0.796	0.448	0.793	0.131	0.799	0.037
B0-TM(d4,5)	0.147	0.321	0.195	0.080	0.200	0.024	0.772	0.437	0.790	0.127	0.799	0.036
B1-ST(i1,5)	0.172	0.358	0.205	0.090	0.206	0.027	0.746	0.143	0.789	0.040	0.798	0.017
B1-ST(i2,5)	0.196	0.470	0.198	0.144	0.204	0.041	0.762	0.230	0.796	0.058	0.799	0.018
B1-ST(d4,5)	0.196	0.434	0.201	0.124	0.205	0.034	0.746	0.252	0.796	0.057	0.798	0.018
B1-Q(i1,5)	0.237	0.452	0.202	0.184	0.205	0.045	0.767	0.255	0.799	0.068	0.800	0.026
B1-Q(i2,5)	0.255	0.495	0.192	0.256	0.205	0.057	0.775	0.341	0.799	0.081	0.800	0.025
B1-Q(d4,5)	0.281	0.467	0.216	0.214	0.203	0.052	0.754	0.358	0.798	0.082	0.800	0.025
B1-TM(i1,5)	0.212	0.386	0.206	0.106	0.205	0.030	0.770	0.160	0.793	0.047	0.799	0.020
B1-TM(i2,5)	0.254	0.456	0.200	0.164	0.204	0.045	0.783	0.249	0.798	0.063	0.799	0.019
B1-TM(d4,5)	0.243	0.444	0.204	0.133	0.205	0.038	0.762	0.287	0.797	0.063	0.799	0.019

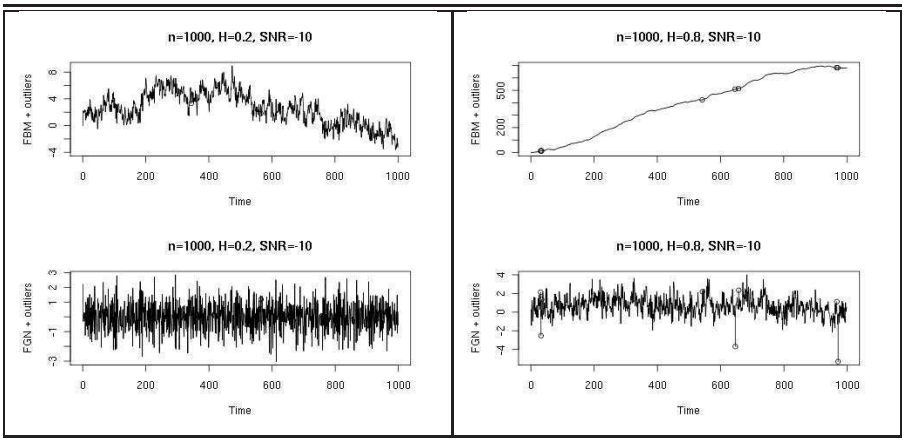


FIG 8. *Example of contaminated fractional Brownian motions (model AO with a $SNR = -10$) with Hurst parameters $H = 0.2$ (left) and $H = 0.8$ (right) and length $n = 1000$.*

TABLE 6
$m = 500$ replications of a contaminated fractional Brownian motion (model AO with a $SNR = -10$) for $n = 100, 1000, 10000$ and $H = 0.2, 0.8$ for different filters and different values of $M2$

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ST(i1,5)	0.237	0.076	0.224	0.025	0.223	0.008	0.747	0.081	0.786	0.031	0.790	0.013
ST(i2,5)	0.228	0.089	0.219	0.028	0.218	0.009	0.756	0.109	0.783	0.033	0.781	0.011
ST(d4,5)	0.232	0.087	0.220	0.026	0.219	0.008	0.757	0.110	0.784	0.033	0.783	0.010
Q(i1,5)	0.218	0.102	0.211	0.031	0.208	0.009	0.785	0.131	0.798	0.043	0.803	0.016
Q(i2,5)	0.222	0.117	0.212	0.037	0.211	0.011	0.812	0.136	0.805	0.040	0.804	0.013
Q(d4,5)	0.227	0.114	0.213	0.035	0.211	0.010	0.806	0.148	0.804	0.042	0.803	0.013
TM(i1,5)	0.259	0.074	0.215	0.023	0.210	0.007	0.811	0.084	0.802	0.033	0.802	0.013
TM(i2,5)	0.263	0.096	0.217	0.028	0.212	0.009	0.831	0.109	0.807	0.032	0.803	0.011
TM(d4,5)	0.275	0.094	0.218	0.027	0.213	0.008	0.837	0.115	0.806	0.033	0.802	0.010
B0-ST(i1,5)	0.153	0.290	0.201	0.059	0.201	0.018	0.678	0.298	0.782	0.065	0.798	0.024
B0-ST(i2,5)	0.201	0.394	0.199	0.093	0.202	0.028	0.759	0.435	0.791	0.123	0.801	0.037
B0-ST(d4,5)	0.199	0.348	0.198	0.079	0.202	0.023	0.718	0.430	0.788	0.118	0.801	0.036
B0-Q(i1,5)	0.172	0.410	0.197	0.105	0.196	0.031	0.784	0.415	0.792	0.129	0.797	0.039
B0-Q(i2,5)	0.215	0.432	0.195	0.131	0.198	0.039	0.857	0.453	0.782	0.181	0.794	0.051
B0-Q(d4,5)	0.202	0.439	0.192	0.122	0.197	0.035	0.824	0.470	0.774	0.181	0.794	0.053
B0-TM(i1,5)	0.103	0.334	0.194	0.065	0.195	0.020	0.708	0.323	0.785	0.077	0.795	0.028
B0-TM(i2,5)	0.185	0.394	0.195	0.098	0.198	0.030	0.786	0.447	0.783	0.132	0.791	0.038
B0-TM(d4,5)	0.185	0.353	0.193	0.083	0.197	0.025	0.732	0.462	0.781	0.129	0.791	0.039
B1-ST(i1,5)	0.255	0.335	0.253	0.086	0.258	0.027	0.720	0.168	0.785	0.045	0.794	0.018
B1-ST(i2,5)	0.258	0.424	0.237	0.136	0.244	0.040	0.725	0.259	0.790	0.059	0.791	0.018
B1-ST(d4,5)	0.207	0.429	0.244	0.110	0.247	0.034	0.718	0.255	0.789	0.059	0.792	0.019
B1-Q(i1,5)	0.296	0.443	0.228	0.154	0.229	0.045	0.761	0.252	0.798	0.071	0.800	0.025
B1-Q(i2,5)	0.335	0.454	0.225	0.215	0.233	0.055	0.759	0.341	0.801	0.079	0.799	0.025
B1-Q(d4,5)	0.322	0.435	0.233	0.184	0.237	0.051	0.760	0.346	0.796	0.080	0.799	0.026
B1-TM(i1,5)	0.308	0.389	0.238	0.095	0.237	0.030	0.751	0.176	0.793	0.050	0.798	0.020
B1-TM(i2,5)	0.305	0.434	0.236	0.150	0.238	0.044	0.758	0.272	0.798	0.062	0.797	0.019
B1-TM(d4,5)	0.273	0.443	0.242	0.120	0.241	0.037	0.755	0.268	0.796	0.063	0.797	0.020

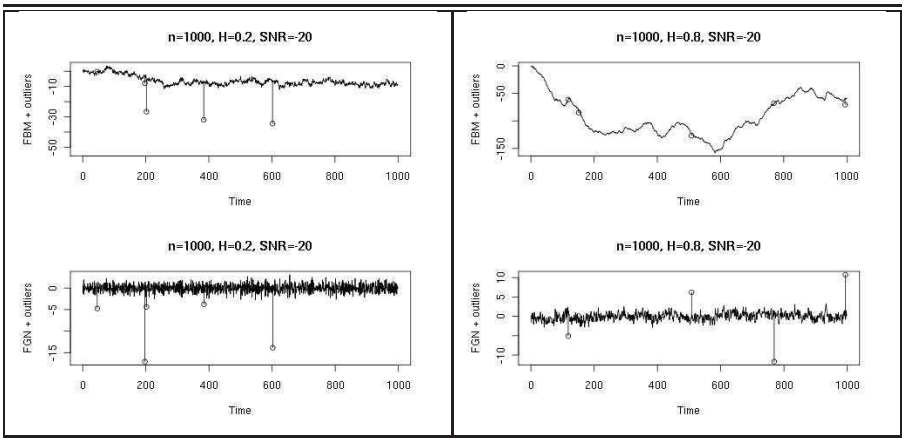


FIG 9. *Example of contaminated fractional Brownian motions (model AO with a $SNR = -20$) with Hurst parameters $H = 0.2$ (left) and $H = 0.8$ (right) and length $n = 1000$.*

TABLE 7
$m = 500$ replications of a contaminated fractional Brownian motion (model AO with a $SNR = -20$) for $n = 100, 1000, 10000$ and $H = 0.2, 0.8$ for different filters and different values of $M2$

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ST(i1,5)	0.332	0.122	0.314	0.055	0.332	0.019	0.680	0.114	0.731	0.049	0.729	0.018
ST(i2,5)	0.318	0.130	0.297	0.054	0.311	0.019	0.663	0.140	0.693	0.059	0.684	0.019
ST(d4,5)	0.315	0.144	0.301	0.054	0.316	0.019	0.665	0.147	0.699	0.059	0.690	0.019
Q(i1,5)	0.231	0.108	0.213	0.029	0.213	0.010	0.805	0.125	0.808	0.046	0.809	0.016
Q(i2,5)	0.252	0.121	0.221	0.036	0.223	0.011	0.850	0.139	0.817	0.044	0.816	0.013
Q(d4,5)	0.268	0.126	0.227	0.036	0.227	0.011	0.849	0.143	0.818	0.045	0.816	0.014
TM(i1,5)	0.288	0.090	0.221	0.022	0.218	0.007	0.843	0.083	0.813	0.035	0.811	0.013
TM(i2,5)	0.392	0.178	0.234	0.030	0.231	0.009	0.922	0.145	0.825	0.035	0.821	0.010
TM(d4,5)	0.418	0.196	0.240	0.030	0.237	0.009	0.922	0.150	0.825	0.036	0.820	0.011
B0-ST(i1,5)	0.262	0.425	0.187	0.125	0.199	0.038	0.698	0.342	0.790	0.068	0.799	0.024
B0-ST(i2,5)	0.349	0.460	0.174	0.194	0.198	0.058	0.770	0.498	0.794	0.192	0.800	0.052
B0-ST(d4,5)	0.399	0.491	0.173	0.183	0.199	0.051	0.745	0.486	0.794	0.166	0.801	0.046
B0-Q(i1,5)	0.176	0.421	0.182	0.105	0.191	0.032	0.786	0.417	0.803	0.132	0.817	0.039
B0-Q(i2,5)	0.260	0.456	0.159	0.166	0.180	0.041	0.874	0.491	0.821	0.163	0.825	0.049
B0-Q(d4,5)	0.354	0.466	0.155	0.171	0.175	0.038	0.822	0.507	0.805	0.176	0.818	0.049
B0-TM(i1,5)	0.239	0.375	0.175	0.073	0.181	0.021	0.759	0.314	0.806	0.079	0.814	0.025
B0-TM(i2,5)	0.469	0.550	0.129	0.167	0.161	0.036	0.841	0.467	0.823	0.121	0.824	0.037
B0-TM(d4,5)	0.481	0.573	0.119	0.160	0.153	0.033	0.757	0.476	0.808	0.119	0.813	0.037
B1-ST(i1,5)	0.348	0.314	0.379	0.098	0.410	0.024	0.668	0.180	0.756	0.051	0.759	0.020
B1-ST(i2,5)	0.322	0.408	0.361	0.123	0.391	0.035	0.627	0.326	0.732	0.073	0.728	0.023
B1-ST(d4,5)	0.308	0.390	0.364	0.124	0.396	0.032	0.613	0.343	0.736	0.072	0.733	0.022
B1-Q(i1,5)	0.317	0.445	0.255	0.150	0.254	0.042	0.774	0.253	0.809	0.075	0.814	0.025
B1-Q(i2,5)	0.398	0.486	0.293	0.215	0.289	0.049	0.806	0.325	0.824	0.078	0.821	0.026
B1-Q(d4,5)	0.455	0.477	0.304	0.158	0.305	0.043	0.769	0.369	0.819	0.082	0.817	0.026
B1-TM(i1,5)	0.530	0.451	0.276	0.097	0.278	0.028	0.793	0.179	0.809	0.052	0.812	0.018
B1-TM(i2,5)	0.627	0.584	0.326	0.141	0.327	0.040	0.830	0.321	0.825	0.064	0.823	0.020
B1-TM(d4,5)	0.577	0.572	0.341	0.127	0.343	0.036	0.787	0.311	0.817	0.064	0.817	0.021

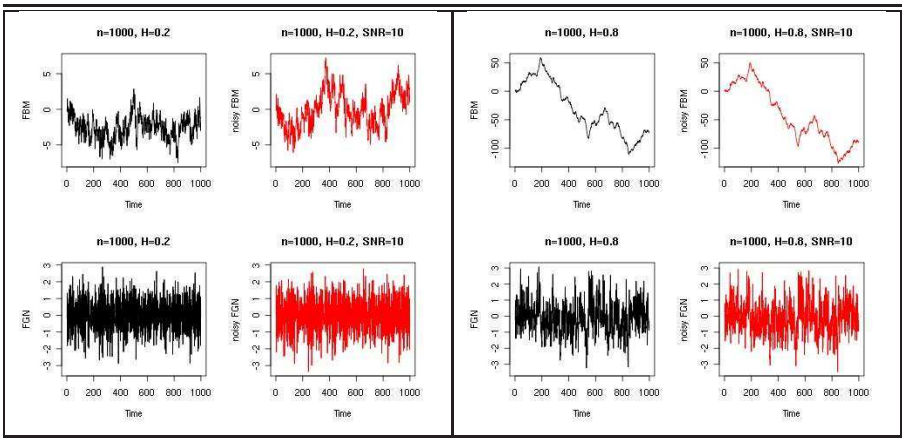


FIG 10. *Example of a contaminated fractional Brownian motion (model B0 with a $SNR = 10$) with Hurst parameters $H = 0.2$ (left) and $H = 0.8$ (right) and length $n = 1000$.*

TABLE 8
$m = 500$ replications of a contaminated fractional Brownian motion (model B0 with a $SNR = 10$) for $n = 100, 1000, 10000$ and $H = 0.2, 0.8$ for different filters and different values of $M2$

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ST(i1,5)	0.235	0.060	0.243	0.019	0.242	0.006	0.751	0.082	0.778	0.029	0.782	0.012
ST(i2,5)	0.226	0.080	0.232	0.023	0.233	0.008	0.752	0.102	0.767	0.031	0.766	0.009
ST(d4,5)	0.229	0.076	0.235	0.022	0.235	0.007	0.753	0.105	0.770	0.030	0.769	0.009
Q(i1,5)	0.241	0.099	0.242	0.032	0.242	0.009	0.773	0.131	0.782	0.044	0.783	0.015
Q(i2,5)	0.227	0.118	0.234	0.036	0.233	0.011	0.765	0.133	0.766	0.043	0.766	0.013
Q(d4,5)	0.229	0.110	0.235	0.035	0.235	0.010	0.771	0.142	0.769	0.043	0.769	0.013
TM(i1,5)	0.269	0.067	0.245	0.021	0.242	0.007	0.796	0.089	0.783	0.033	0.783	0.012
TM(i2,5)	0.258	0.086	0.235	0.026	0.233	0.009	0.789	0.107	0.770	0.034	0.766	0.010
TM(d4,5)	0.269	0.082	0.238	0.025	0.235	0.008	0.800	0.110	0.773	0.034	0.769	0.011
B0-ST(i1,5)	0.155	0.275	0.193	0.061	0.200	0.020	0.682	0.315	0.782	0.064	0.798	0.023
B0-ST(i2,5)	0.200	0.377	0.187	0.100	0.201	0.031	0.784	0.416	0.787	0.133	0.801	0.036
B0-ST(d4,5)	0.174	0.332	0.191	0.082	0.201	0.027	0.755	0.438	0.790	0.125	0.801	0.034
B0-Q(i1,5)	0.224	0.404	0.188	0.139	0.202	0.035	0.805	0.420	0.796	0.139	0.799	0.041
B0-Q(i2,5)	0.213	0.432	0.188	0.166	0.201	0.044	0.884	0.460	0.785	0.198	0.803	0.052
B0-Q(d4,5)	0.229	0.425	0.189	0.152	0.202	0.039	0.853	0.536	0.786	0.205	0.805	0.051
B0-TM(i1,5)	0.134	0.330	0.193	0.076	0.201	0.024	0.728	0.316	0.791	0.075	0.799	0.026
B0-TM(i2,5)	0.182	0.414	0.189	0.113	0.201	0.034	0.829	0.437	0.787	0.147	0.803	0.039
B0-TM(d4,5)	0.152	0.401	0.191	0.095	0.201	0.030	0.802	0.452	0.789	0.141	0.803	0.038
B1-ST(i1,5)	0.256	0.334	0.305	0.070	0.300	0.022	0.732	0.152	0.781	0.044	0.790	0.017
B1-ST(i2,5)	0.242	0.444	0.285	0.119	0.281	0.034	0.735	0.264	0.780	0.060	0.783	0.018
B1-ST(d4,5)	0.225	0.435	0.289	0.097	0.285	0.030	0.720	0.278	0.782	0.060	0.784	0.018
B1-Q(i1,5)	0.287	0.439	0.302	0.142	0.297	0.038	0.760	0.247	0.793	0.072	0.791	0.025
B1-Q(i2,5)	0.301	0.473	0.280	0.183	0.279	0.049	0.765	0.352	0.782	0.080	0.784	0.025
B1-Q(d4,5)	0.325	0.461	0.284	0.168	0.283	0.043	0.745	0.362	0.785	0.082	0.786	0.025
B1-TM(i1,5)	0.271	0.373	0.301	0.085	0.299	0.025	0.757	0.167	0.787	0.050	0.791	0.019
B1-TM(i2,5)	0.283	0.481	0.282	0.134	0.280	0.037	0.760	0.265	0.781	0.065	0.784	0.019
B1-TM(d4,5)	0.279	0.453	0.287	0.108	0.284	0.033	0.742	0.298	0.783	0.064	0.785	0.019

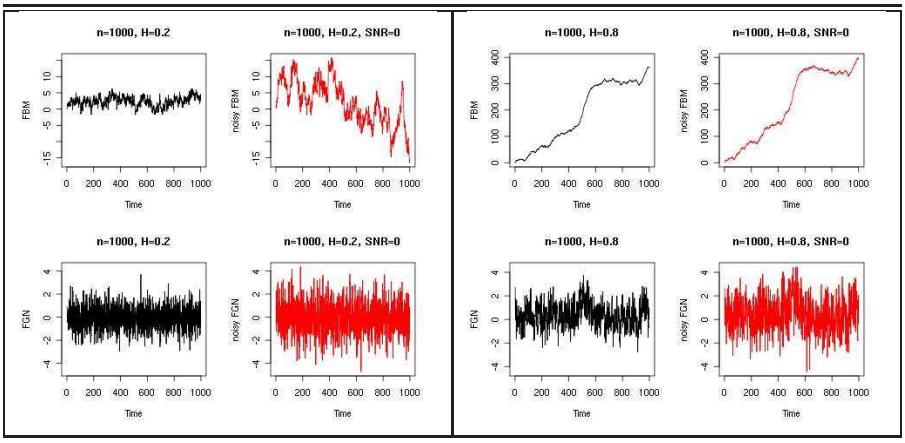


FIG 11. *Example of a contaminated fractional Brownian motion (model B0 with a $SNR = 0$) with Hurst parameters $H = 0.2$ (left) and $H = 0.8$ (right) and length $n = 1000$.*

TABLE 9

$m = 500$ replications of a contaminated fractional Brownian motion (model B0 with a $SNR = 0$) for $n = 100, 1000, 10000$ and $H = 0.2, 0.8$ for different filters and different values of $M2$

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ST(i1,5)	0.378	0.072	0.385	0.020	0.385	0.007	0.683	0.080	0.684	0.028	0.684	0.011
ST(i2,5)	0.358	0.094	0.365	0.027	0.363	0.009	0.634	0.101	0.636	0.030	0.633	0.010
ST(d4,5)	0.365	0.091	0.370	0.026	0.369	0.009	0.639	0.097	0.643	0.029	0.640	0.009
Q(i1,5)	0.387	0.109	0.385	0.033	0.385	0.011	0.699	0.132	0.686	0.044	0.685	0.015
Q(i2,5)	0.361	0.128	0.366	0.036	0.363	0.012	0.636	0.138	0.636	0.041	0.632	0.013
Q(d4,5)	0.366	0.127	0.371	0.037	0.369	0.012	0.653	0.139	0.644	0.042	0.639	0.013
TM(i1,5)	0.413	0.080	0.388	0.023	0.385	0.007	0.723	0.090	0.689	0.032	0.685	0.012
TM(i2,5)	0.391	0.102	0.368	0.029	0.363	0.010	0.668	0.108	0.639	0.032	0.633	0.011
TM(d4,5)	0.407	0.097	0.374	0.029	0.369	0.009	0.682	0.105	0.646	0.032	0.640	0.011
B0-ST(i1,5)	0.319	0.423	0.161	0.233	0.191	0.062	0.727	0.378	0.779	0.082	0.796	0.026
B0-ST(i2,5)	0.365	0.451	0.159	0.288	0.185	0.093	0.794	0.505	0.787	0.236	0.799	0.063
B0-ST(d4,5)	0.401	0.456	0.143	0.332	0.187	0.083	0.809	0.477	0.780	0.207	0.800	0.056
B0-Q(i1,5)	0.405	0.463	0.194	0.356	0.178	0.128	0.845	0.479	0.795	0.209	0.798	0.051
B0-Q(i2,5)	0.429	0.457	0.204	0.361	0.172	0.140	0.841	0.513	0.809	0.393	0.799	0.093
B0-Q(d4,5)	0.443	0.462	0.192	0.368	0.172	0.140	0.857	0.518	0.810	0.379	0.799	0.086
B0-TM(i1,5)	0.345	0.437	0.142	0.288	0.190	0.077	0.811	0.377	0.789	0.101	0.798	0.032
B0-TM(i2,5)	0.373	0.469	0.150	0.328	0.182	0.098	0.821	0.515	0.799	0.286	0.799	0.070
B0-TM(d4,5)	0.422	0.435	0.142	0.348	0.183	0.092	0.881	0.539	0.789	0.236	0.800	0.063
B1-ST(i1,5)	0.412	0.212	0.444	0.051	0.450	0.017	0.692	0.167	0.721	0.046	0.727	0.017
B1-ST(i2,5)	0.402	0.358	0.429	0.089	0.437	0.028	0.633	0.284	0.685	0.061	0.684	0.020
B1-ST(d4,5)	0.388	0.340	0.431	0.079	0.441	0.024	0.645	0.258	0.691	0.059	0.691	0.018
B1-Q(i1,5)	0.434	0.386	0.445	0.091	0.451	0.029	0.725	0.287	0.730	0.075	0.729	0.025
B1-Q(i2,5)	0.469	0.457	0.431	0.128	0.439	0.036	0.670	0.350	0.691	0.090	0.683	0.027
B1-Q(d4,5)	0.459	0.443	0.437	0.116	0.442	0.033	0.682	0.337	0.699	0.087	0.690	0.026
B1-TM(i1,5)	0.428	0.244	0.445	0.062	0.450	0.020	0.721	0.182	0.727	0.053	0.728	0.019
B1-TM(i2,5)	0.442	0.388	0.430	0.099	0.438	0.029	0.663	0.274	0.689	0.068	0.684	0.021
B1-TM(d4,5)	0.411	0.379	0.432	0.089	0.442	0.026	0.667	0.282	0.694	0.066	0.691	0.021

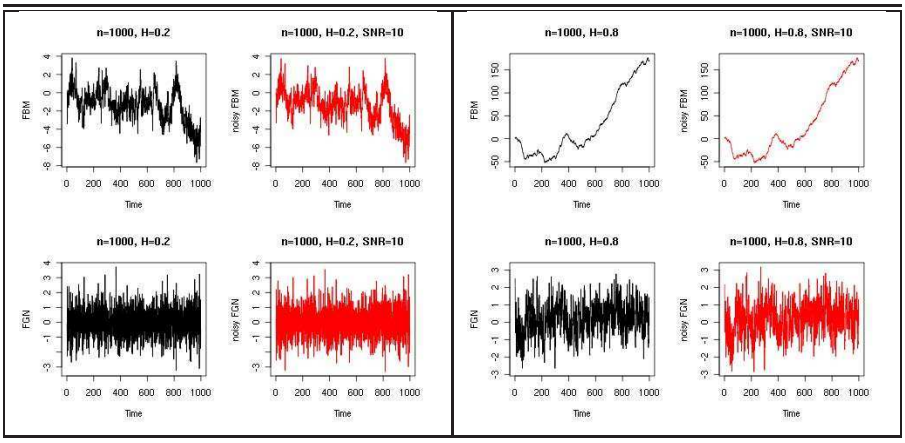


FIG 12. *Example of a contaminated fractional Brownian motion (model B1 with a $SNR = 10$) with Hurst parameters $H = 0.2$ (left) and $H = 0.8$ (right) and length $n = 1000$.*

TABLE 10
$m = 500$ replications of a contaminated fractional Brownian motion (model B1 with a $SNR = 10$) for $n = 100, 1000, 10000$ and $H = 0.2, 0.8$ for different filters and different values of $M2$

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ST(i1,5)	0.178	0.059	0.186	0.019	0.186	0.006	0.749	0.076	0.767	0.032	0.772	0.013
ST(i2,5)	0.180	0.076	0.184	0.024	0.185	0.008	0.711	0.100	0.720	0.031	0.723	0.010
ST(d4,5)	0.178	0.071	0.185	0.022	0.186	0.007	0.722	0.102	0.733	0.031	0.736	0.010
Q(i1,5)	0.180	0.103	0.185	0.031	0.186	0.009	0.778	0.129	0.772	0.049	0.773	0.017
Q(i2,5)	0.183	0.111	0.185	0.037	0.185	0.011	0.720	0.143	0.723	0.043	0.723	0.014
Q(d4,5)	0.182	0.106	0.185	0.035	0.185	0.010	0.732	0.147	0.735	0.041	0.736	0.013
TM(i1,5)	0.212	0.067	0.188	0.022	0.187	0.007	0.796	0.083	0.773	0.036	0.773	0.014
TM(i2,5)	0.210	0.084	0.187	0.027	0.185	0.008	0.746	0.109	0.724	0.034	0.723	0.011
TM(d4,5)	0.218	0.078	0.188	0.025	0.186	0.008	0.768	0.111	0.738	0.034	0.737	0.011
B0-ST(i1,5)	0.169	0.193	0.180	0.049	0.181	0.015	0.730	0.276	0.814	0.062	0.827	0.021
B0-ST(i2,5)	0.184	0.337	0.175	0.078	0.179	0.023	0.893	0.463	0.899	0.138	0.903	0.041
B0-ST(d4,5)	0.164	0.323	0.177	0.065	0.180	0.019	0.837	0.463	0.873	0.121	0.879	0.038
B0-Q(i1,5)	0.180	0.397	0.179	0.101	0.180	0.029	0.820	0.434	0.813	0.136	0.828	0.041
B0-Q(i2,5)	0.211	0.407	0.174	0.133	0.178	0.035	0.912	0.505	0.912	0.233	0.904	0.059
B0-Q(d4,5)	0.199	0.386	0.173	0.114	0.178	0.030	0.917	0.509	0.881	0.220	0.881	0.059
B0-TM(i1,5)	0.149	0.247	0.178	0.060	0.180	0.018	0.770	0.304	0.822	0.073	0.828	0.026
B0-TM(i2,5)	0.172	0.352	0.175	0.089	0.178	0.026	0.938	0.460	0.902	0.162	0.903	0.044
B0-TM(d4,5)	0.152	0.325	0.176	0.074	0.179	0.022	0.909	0.488	0.874	0.140	0.880	0.043
B1-ST(i1,5)	0.121	0.394	0.194	0.092	0.198	0.027	0.745	0.153	0.790	0.045	0.798	0.018
B1-ST(i2,5)	0.160	0.430	0.193	0.149	0.199	0.041	0.752	0.249	0.796	0.059	0.800	0.018
B1-ST(d4,5)	0.127	0.449	0.194	0.121	0.199	0.034	0.740	0.249	0.795	0.055	0.800	0.018
B1-Q(i1,5)	0.183	0.442	0.199	0.207	0.201	0.050	0.770	0.263	0.794	0.073	0.800	0.026
B1-Q(i2,5)	0.187	0.439	0.197	0.267	0.202	0.062	0.776	0.335	0.800	0.079	0.800	0.025
B1-Q(d4,5)	0.195	0.474	0.205	0.223	0.202	0.053	0.784	0.339	0.799	0.081	0.800	0.027
B1-TM(i1,5)	0.177	0.391	0.196	0.109	0.200	0.033	0.767	0.181	0.795	0.051	0.799	0.020
B1-TM(i2,5)	0.179	0.437	0.188	0.180	0.201	0.045	0.785	0.270	0.798	0.063	0.800	0.020
B1-TM(d4,5)	0.169	0.452	0.193	0.139	0.200	0.039	0.776	0.258	0.797	0.061	0.800	0.021

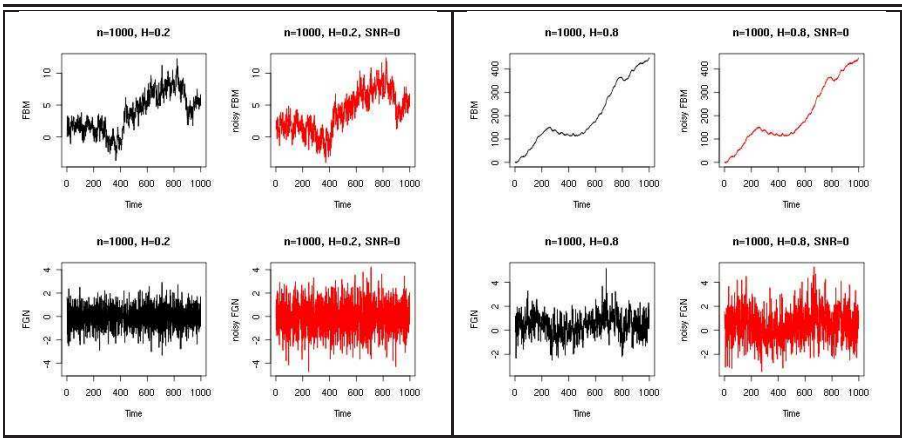


FIG 13. *Example of a contaminated fractional Brownian motion (model B1 with a $SNR = 0$) with Hurst parameters $H = 0.2$ (left) and $H = 0.8$ (right) and length $n = 1000$.*

TABLE 11
$m = 500$ replications of a contaminated fractional Brownian motion (model B1 with a $SNR = 0$) for $n = 100, 1000, 10000$ and $H = 0.2, 0.8$ for different filters and different values of $M2$

	$H = 0.2$						$H = 0.8$					
	$n = 100$		$n = 1000$		$n = 10000$		$n = 100$		$n = 1000$		$n = 10000$	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ST(i1,5)	0.112	0.053	0.116	0.018	0.116	0.005	0.575	0.096	0.603	0.041	0.607	0.017
ST(i2,5)	0.105	0.067	0.111	0.023	0.110	0.007	0.422	0.093	0.424	0.031	0.424	0.009
ST(d4,5)	0.108	0.062	0.113	0.021	0.112	0.006	0.456	0.096	0.462	0.031	0.462	0.009
Q(i1,5)	0.113	0.097	0.115	0.029	0.116	0.009	0.601	0.143	0.609	0.059	0.607	0.021
Q(i2,5)	0.112	0.106	0.112	0.033	0.111	0.011	0.426	0.123	0.423	0.041	0.424	0.012
Q(d4,5)	0.111	0.102	0.112	0.031	0.112	0.010	0.463	0.132	0.463	0.041	0.462	0.013
TM(i1,5)	0.146	0.061	0.119	0.020	0.116	0.006	0.619	0.106	0.609	0.047	0.607	0.019
TM(i2,5)	0.137	0.074	0.114	0.025	0.111	0.008	0.453	0.097	0.426	0.033	0.424	0.010
TM(d4,5)	0.149	0.068	0.116	0.023	0.113	0.007	0.496	0.102	0.466	0.033	0.462	0.010
B0-ST(i1,5)	0.080	0.139	0.099	0.042	0.097	0.013	0.996	0.487	1.413	0.336	1.728	0.327
B0-ST(i2,5)	0.059	0.260	0.094	0.062	0.092	0.019	0.535	0.452	0.572	0.231	0.714	0.098
B0-ST(d4,5)	0.064	0.219	0.096	0.051	0.093	0.016	0.643	0.483	0.711	0.173	0.728	0.046
B0-Q(i1,5)	0.068	0.328	0.096	0.077	0.096	0.025	0.884	0.504	1.271	0.344	1.588	0.347
B0-Q(i2,5)	0.080	0.363	0.093	0.095	0.092	0.029	0.612	0.510	0.527	0.327	0.670	0.113
B0-Q(d4,5)	0.081	0.339	0.094	0.082	0.093	0.026	0.703	0.518	0.669	0.311	0.736	0.093
B0-TM(i1,5)	0.066	0.179	0.098	0.049	0.097	0.015	0.967	0.533	1.399	0.368	1.698	0.340
B0-TM(i2,5)	0.048	0.306	0.092	0.069	0.092	0.021	0.578	0.476	0.568	0.237	0.702	0.103
B0-TM(d4,5)	0.052	0.238	0.095	0.058	0.093	0.018	0.677	0.490	0.706	0.194	0.731	0.057
B1-ST(i1,5)	0.191	0.418	0.196	0.153	0.200	0.040	0.736	0.146	0.793	0.044	0.800	0.017
B1-ST(i2,5)	0.191	0.463	0.201	0.283	0.199	0.062	0.778	0.329	0.805	0.071	0.800	0.022
B1-ST(d4,5)	0.215	0.471	0.194	0.219	0.199	0.050	0.747	0.304	0.801	0.063	0.800	0.019
B1-Q(i1,5)	0.194	0.482	0.213	0.341	0.205	0.078	0.805	0.279	0.805	0.076	0.803	0.026
B1-Q(i2,5)	0.176	0.458	0.187	0.378	0.202	0.099	0.822	0.420	0.810	0.111	0.802	0.032
B1-Q(d4,5)	0.177	0.410	0.203	0.349	0.204	0.086	0.826	0.448	0.805	0.099	0.801	0.029
B1-TM(i1,5)	0.235	0.448	0.199	0.195	0.201	0.048	0.769	0.172	0.799	0.054	0.801	0.020
B1-TM(i2,5)	0.215	0.462	0.207	0.303	0.200	0.071	0.826	0.352	0.806	0.080	0.801	0.024
B1-TM(d4,5)	0.215	0.453	0.195	0.241	0.200	0.058	0.786	0.339	0.801	0.071	0.801	0.021

References

S. Baykut, T. Akgül, and S. Ergintav. Estimation of spectral exponent parameter of 1/f process in additive white background noise. *EURASIP Journal on Advances in Signal Processing*, 2007(15):1–7, 2007.

J. Beran. *Statistics for long-memory processes*. Chapman & Hall/CRC, 1994. MR1304490

A. Brouste, J. Istas, and S. Lambert-Lacroix. On Fractional Gaussian Random Fields Simulations. *Journal of Statistical Software*, 23(1): 1–23, 2007.

G. Chan and A.T.A. Wood. Estimation of fractal dimension for a class of non-Gaussian stationary processes and fields. *Annals of Statistics*, 32(3):1222–1260, 2004. MR2065204

J.-F. Coeurjolly. Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *J. Stat. Softw.*, 5(7):1–53, November 2000a.

J.-F. Coeurjolly. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Infer. Stoch. Process.*, 4(2): 199–227, January 2001. MR1856174

J.-F. Coeurjolly. Identification of multifractional Brownian motion. *Bernoulli*, 11(6):987–1008, 2005. MR2188838

J.-F. Coeurjolly. Hurst exponent estimation of locally self-similar gaussian processes using sample quantiles. *Annals of Statistics*, 36(3):1404–1434, 2008. MR2418662

J.-F. Coeurjolly. *Inférence statistique pour les mouvements Browniens fractionnaire et multifractinonaire*. PhD thesis, Université Joseph Fourier, 2000b.

S. Cohen and J. Istas. An universal estimator of local self-similarity. *ESAIM: Probability and Statistics*, 2002.

I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and, Applied Mathematics*, 41 (7):909–996, 2006. MR0951745

J.L. Doob. *Stochastic Processes*. Wiley Classics Library, USA, 1953. MR0058896

P. Doukhan, G. Oppenheim, and M.S. Taqqu. *Theory and applications of long-range dependence*. Birkhauser, 2003. MR1956041

G. Fäy, E. Moulines, F. Roueff, and M.S. Taqqu. Estimators of long-memory: Fourier versus wavelets. *Journal of Econometrics*, 151(2):159–177, 2009. MR2559823

J. Istas. Quadratic variations of spherical fractional Brownian motions. *Stochastic Processes and their Applications*, 117 (4):476–486, 2007. MR2305382

J. Istas and G. Lang. Quadratic variations and estimation of the hölder index of a gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.*, 33: 407–436, 1997. MR1465796

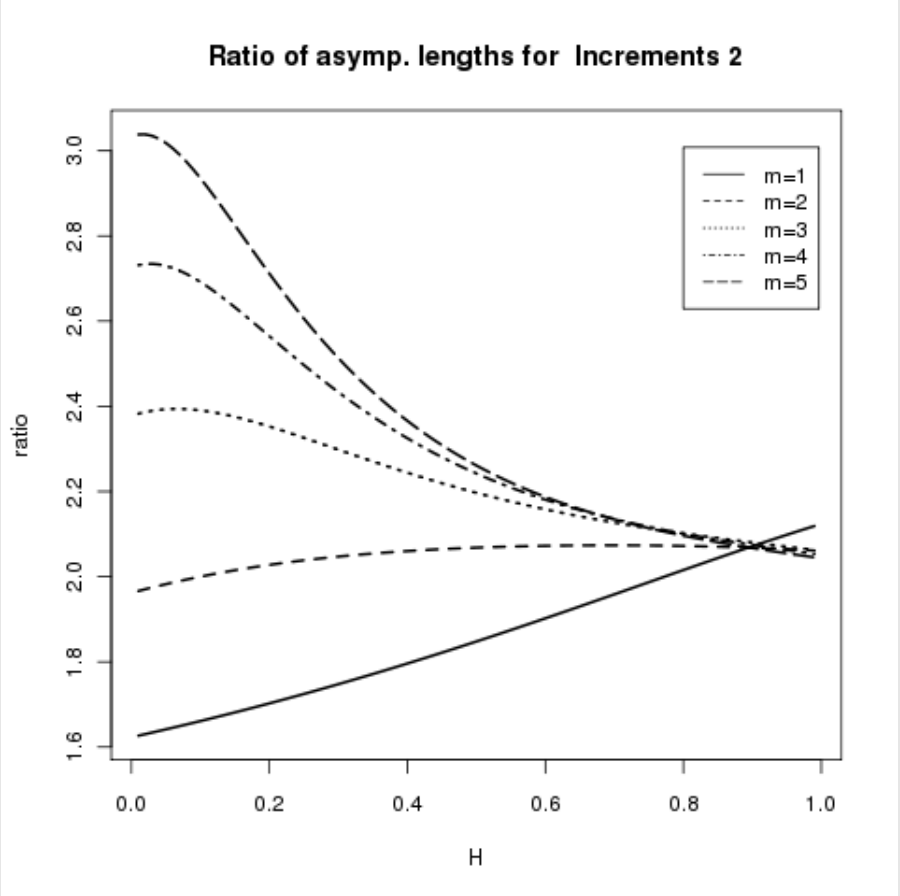
J.T. Kent and A.T.A. Wood. Estimating the fractal dimension of a locally self-similar gaussian process using increments. *J. Roy. Statist. Soc. Ser. B*, 59:679–700, 1997. MR1452033

B. Mandelbrot and J. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968. MR0242239

D. B. Percival and A. T. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, 2000. MR1770693

F. Richard and H. Biermé. Estimation of anisotropic gaussian fields through radon transform. *ESAIM Probab. Stat.*, 12(1):30–50, 2008. MR2367992

Haipeng Shen, Zhengyuan Zhu, and Thomas C. M. Lee. Robust estimation of the self-similarity parameter in network traffic using wavelet transform. *Signal Process.*, 87(9):2111–2124, 2007.



[BC10]
J.-C. Breton and J.-F. Coeurjolly. Confidence intervals for the Hurst parameter of a fractional Brownian motion based on concentration inequalities. *submitted for publication* (available at `http://arxiv.org/abs/0910.3088`, 2010.

Confidence intervals for the Hurst parameter of a fractional Brownian motion based on finite sample size

Jean-Christophe Breton* and Jean-François Coeurjolly†

June 1, 2010

Abstract

In this paper, we show how concentration inequalities for Gaussian quadratic form can be used to propose exact confidence intervals of the Hurst index parametrizing a fractional Brownian motion. Both cases where the scaling parameter of the fractional Brownian motion is known or unknown are investigated. These intervals are obtained by observing a single discretized sample path of a fractional Brownian motion and without any assumption on the parameter H.

Keywords: concentration inequalities, confidence intervals, fractional Brownian motion, Hurst parameter

Contents

*Laboratoire de Mathématiques, Images et Applications, Université de La Rochelle, France. Email: jcbreton@univ-lr.fr

†GIPSA-lab and Laboratory Jean Kuntzmann, Grenoble University, France. Email: Jean-Francois.Coeurjolly@upmf-grenoble.fr

1 Introduction

Since the pioneer work of Mandelbrot and Ness (1968), the fractional Brownian motion (fBm) has become widely popular as well as in a theoretical context as in applications. Fractional Brownian motion can be defined as the only centered Gaussian process, denoted by $(B_H(t))_{t\in\mathbb{R}}$, with stationary increments and with variance function $v(\cdot)$, given by $v(t) = C^2|t|^{2H}$ for all $t \in \mathbb{R}$. The parameter $H \in (0,1)$ (resp. $C > 0$) is referred to as the Hurst parameter (resp. the scaling coefficient). In particular, when $H = 1/2$, it is the standard Brownian motion. In general, the fractional Brownian motion is an H-self-similar process, that is for all $\delta > 0$, $(B_H(\delta t))_{t\in\mathbb{R}} \overset{d}{=} \delta^H (B_H(t))_{t\in\mathbb{R}}$ (where $\overset{d}{=}$ means equal in finite-dimensional distributions) with autocovariance function behaving like $O(|k|^{2H-2})$ as $|k| \to +\infty$. Thus, the discretized increments of the fractional Brownian motion (called the fractional Gaussian noise) constitute a short-range dependent process, when $H < 1/2$, and a long-range dependent process, when $H > 1/2$. The index H characterizes also the path regularity since the fractal dimension of the fractional Brownian motion is equal to $D = 2 - H$. General references on self-similar processes and long-memory processes are given in Beran (1994) or Doukhan et al. (2003).

The aim of this paper is to propose confidence intervals for the Hurst parameter based on a single observation of a discretized sample path of the interval $[0,1]$ of a fractional Brownian motion. To do so, the most popular strategy consists in using the *asymptotic normality* of some estimators of the Hurst parameter, see Coeurjolly (2000) for a survey on the estimation of the self-similarity or Shen et al. (2007) and Coeurjolly (2008) for more recent discussions in a robust context. Recently, a new strategy based on concentration inequalities for Gaussian processes obtained by Nourdin and Viens (2009) has been proposed by Breton et al. (2009). In this case, the confidence intervals are *non-asymptotic* and they appear to be very interesting when the sample size is moderate. Our contribution is to improve this direction both from a theoretical and practical point of view. In order to present our different contributions, let us first recall the confidence interval proposed by Breton et al. (2009).

Proposition 1 *Assume that one observes a fractional Brownian motion at times i/n for $i = 0,\dots,n+1$ with scaling coefficient $C = 1$ and with Hurst parameter satisfying $H \le H^\star$ for some known $H^\star \in (0,1)$. Fix $\alpha \in (0,1)$, then for all n large enough satisfying $q_n(\alpha) < (4 - 4^{H^\star})\sqrt{n}$, where $q_n(\alpha) := \frac{1}{2}\left(b(\alpha) + \sqrt{b(\alpha)^2 + 852\log\left(\frac{2}{\alpha}\right)}\right)$ with $b(\alpha) := \frac{71}{\sqrt{n}}\log\left(\frac{2}{\alpha}\right)$, we have*

$$\mathbb{P}\left(H \in \left[\max\left(0, \widetilde{H}_n^{inf}(q_n(\alpha))\right), \widetilde{H}_n^{sup}(q_n(\alpha))\right]\right) \ge 1 - \alpha, \tag{1}$$

where for $t > 0$

$$\begin{aligned} g_n\left(\widetilde{H}_n^{inf}(t)\right) &:= \frac{1}{2} - \frac{\log(S_n)}{2\log(n)} + \frac{\log\left(1 - \frac{t}{(4-4^{H^\star})\sqrt{n}}\right)}{2\log(n)} \\ g_n\left(\widetilde{H}_n^{sup}(t)\right) &:= \frac{1}{2} - \frac{\log(S_n)}{2\log(n)} + \frac{\log\left(1 + \frac{t}{(4-4^{H^\star})\sqrt{n}}\right)}{2\log(n)} \end{aligned}$$

where g_n is the function defined by $g_n(x) = x - \frac{\log(4-4^x)}{2\log(n)}$ and S_n is the following statistic

$$S_n := \frac{1}{n}\sum_{i=1}^{n}\left(B_H\left(\frac{i+1}{n}\right) - 2B_H\left(\frac{i}{n}\right) + B_H\left(\frac{i-1}{n}\right)\right)^2. \tag{2}$$

Let us give some general comments on this result. First, note that this procedure cannot be applied to a fractional Brownian motion whose scaling coefficient C is unknown. Secondly, important drawbacks of this procedure rely upon the assumptions made on $H^\star$ and n, which exclude the possibility to use this confidence interval when the sample size is small:

- Given α and $H^\star$, the following table presents the minimal value of the sample size n in order to ensure that $q_n(\alpha) < (4 - 4^{H^\star})\sqrt{n}$.

	$H^\star$								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\alpha = 1\%$	271	298	335	388	471	611	886	1592	4936
$\alpha = 5\%$	189	208	233	270	328	425	617	1108	3437
$\alpha = 10\%$	154	169	190	220	266	346	501	900	2791

- The following table exhibits the maximal value of $H^\star$, denoted by $\widetilde{H}^\star$, required in order to ensure $q_n(\alpha) < (4-4^{H^\star})\sqrt{n}$ in terms of α and n. Note that $\widetilde{H}^\star = \log\left(\max\left(1, 4 - q_n(\alpha)/\sqrt{n}\right)\right)/\log(4)$, which means that, given α and n, a confidence interval is only available for $H \in (0, \widetilde{H}^\star)$.

	n					
	50	100	200	500	10000	10000
$\alpha = 1\%$	0.00	0.00	0.00	0.53	0.93	0.93
$\alpha = 5\%$	0.00	0.00	0.17	0.65	0.94	0.94
$\alpha = 10\%$	0.00	0.00	0.34	0.70	0.95	0.95

We are now in position to specify our different contributions:

- We slightly improve the bounds of the concentration inequality obtained by Nourdin and Viens (2009), see Section 2 and Proposition 2 for more details. Note in particular that, in contrast to Nourdin and Viens (2009) and Breton et al. (2009), we are tracing the constant to optimize numerically our bounds.

- In the case where the scaling parameter C is known, we propose a new confidence interval without any preliminary assumption on the Hurst parameter H (in contrast to Breton et al. (2009)) and with a very slight condition on the sample size. For instance, in comparison to the previous tables, our confidence interval is computable as soon as $n \geq 3$. Furthermore, by using ideas similar in Coeurjolly (2001) for the problem of the estimation of the Hurst parameter, we also propose a confidence interval when the scaling parameter C is unknown. This new confidence interval has the nice property to be independent of C and independent of the discretization step. It is remarkable that, in the both cases (C known or unknown), the lengths of the confidence intervals we propose behave asymptotically like the ones derived in an asymptotic approach, that is they behave like $1/\sqrt{n}\log(n)$ when C is known and $1/\sqrt{n}$ when C is unknown.

- As suggested by the expression of the statistic in (2), the procedure described in Proposition 1 is based on the increments of order 2 of the discretized sample path of the fractional Brownian motion. Taking the increments of order 2 is a special case of filter to work with and it is known that discrete filtering has been proposed and used in an estimation context, see Istas and Lang (1997), Kent and Wood (1997) and Coeurjolly (2001). Recall that the main interest in filtering the fractional Browian motion is that the action of filtering changes the correlation so that, for instance, the increments of order 2 of the fractional

Brownian motion constitute a short-range dependent process (*i.e.* its correlation function is absolutely summable). Such a behaviour is required to obtain an efficient concentration inequality. In this paper, we propose to construct confidence intervals not only based on the increments of order 2 but on more general filters such as, for instance, increments of larger order or the Daubechies wavelet filters... Finally, let us also underline that a crucial step consists in obtaining an upper-bound of the supremum on the interval $(0,1)$ of the ℓ^1-norm of the correlation function of the discrete filtered series of the fractional Brownian motion. When considering the increments of order 2, Breton et al. (2009) have obtained the bound $17.75/(4-4^{H^\star})$. We have widely improved this point since we compute explicitly this supremum for a large class of filters (including increments of order 2). As an example, for the increments of order 2, this gives the explicit value $8/3$.

- Based on a large simulation study, we assess the efficiency of the different procedures that we propose and we compare them with ones based on an asymptotic scheme. We discuss and comment these results.

The rest of this paper is organized as follows. In Section 2, we give the concentration inequalities specially designed for our purposes. The filtering setting is introduced in Section 3 where the bounds for the ℓ^1-norm of the correlation function of the filtered series are also obtained. Our confidence intervals for the Hurst parameter are proposed and proved in Section 4, both when the scaling parameter is known or unknown. Our results are discussed and compared to the literature in Section 5. Finally, computations expliciting some bounds for some special filters are given in Appendix A.

2 Concentration inequalities

Proposition 1 above is based on concentration inequalities proposed by Nourdin and Viens (2009) (see Proposition 3) for smooth enough random variables with respect to Malliavin calculus (see Theorem 4.1-i)). By applying such inequalities to the random variables $\sqrt{n}V_n$ where $V_n = \frac{1}{n}\sum_{i=1}^n H_2(X_i)$, $H_2(t) = t^2-1$ is the second Hermite polynomial, and $X = \{X_i\}_{1\le i\le n}$ is a stationary Gaussian process with variance 1 and correlation function ρ, we obtain concentration inequalities for H_2-variations of stationary Gaussian processes. In the sequel, for a sequence $(u_i)_{i\in\mathbb{Z}}$, we set $\|u\|_{\ell^1_n} := \sum_{|i|\le n}|u_i|$.

Proposition 2 *Let $\kappa_n = 2\|\rho\|_{\ell^1_n}$. Then, for all $t>0$, we have:*

$$\mathbb{P}\left(\sqrt{n}V_n \ge t\right) \le \varphi_{r,n}(t;\kappa_n) := e^{-\frac{t\sqrt{n}}{\kappa_n}}\left(1+\frac{t}{\sqrt{n}}\right)^{\frac{n}{\kappa_n}} \tag{3}$$

$$\mathbb{P}\left(\sqrt{n}V_n \le -t\right) \le \varphi_{l,n}(t;\kappa_n) := e^{\frac{t\sqrt{n}}{\kappa_n}}\left(1-\frac{t}{\sqrt{n}}\right)^{\frac{n}{\kappa_n}}\mathbf{1}_{[0,\sqrt{n}]}(t). \tag{4}$$

Note that Proposition 2 can be applied to short-memory as well as to long-memory stationary Gaussian processes (as soon as n remains finite). In order to derive Proposition 2 below, we shall briefly use some notions of Malliavin calculus. We just recall the only necessary for our argument and we refer to Breton et al. (2009) and references therein for any further details. We stress that, once Proposition 2 is derived, only basic probability tools will be used. Without restriction, we assume the Gaussian random variables X_i have the form $X_i = X(h_i)$ where $X(\aleph) = \{X(h) : h\in\aleph\}$ is an isonormal Gaussian process over a real separable Hilbert space $\aleph$ and $\{h_i : i = 0,\dots,n\}$ is a finite subset of $\aleph$ verifying $\mathbb{E}[X(h_i)X(h_j)] = \rho(i-j) = \langle h_i, h_j\rangle_\aleph$. With

such a representation, V_n can be seen as a double Wiener-Itô integral with respect to X, *i.e.* $V_n = I_2\left(\frac{1}{n}\sum_{i=0}^n h_i \otimes h_i\right)$. In the sequel, to make easier the presentation, we rewrite Th. 4.1 of Nourdin and Viens (2009) only for such random variables, see Proposition 3. Actually, in order to optimize our forthcoming results, Proposition 3 is a slight improvement of Th. 4.1. Before, recall that multiple Wiener-Itô integrals $I_q(f)$ are well defined for $f \in \aleph^{\odot q}$, the qth symmetric tensor product of $\aleph$, $q \in \mathbb{N}\setminus\{0\}$; the Malliavin derivatives D transforms random variables (in its domain) into random elements with values in $\aleph$; multiple Wiener-Itô integrals are in the domain of D and we have $D_t(I_q(h)) = qI_{q-1}(h(\cdot,t))$. Recall also that the Hermite polynomials H_q are related to multiple Wiener-Itô integrals by $H_q(I_1(h)) = I_q(h^{\otimes q})$ when $\|h\|_\aleph = 1$; in particular, for $q=2$, we obtain $I_1(h)^2 - 1 = I_2(h^{\otimes 2})$.

Proposition 3 *Let $Z = I_2(f)$ satisfying*

$$\|DZ\|_\aleph^2 \leq aZ + b \tag{5}$$

for some constants $a \geq 0$ and $b > 0$. Then, for all $t > 0$

$$\begin{aligned}
\mathbb{P}(Z \geq t) &\leq \varphi_r(t;a,b) := e^{-\frac{2t}{a}}\left(1+\frac{at}{b}\right)^{\frac{2b}{a^2}} \\
\mathbb{P}(Z \leq -t) &\leq \varphi_l(t;a,b) := e^{\frac{2t}{a}}\left(1-\frac{at}{b}\right)^{\frac{2b}{a^2}}\mathbf{1}_{[0,b/a]}(t).
\end{aligned}$$

Proof: The proof is a slight improvement of the bounds in (Nourdin and Viens, 2009, Theorem 4.1) obtained by a careful reading of the proof (with the following correspondance with the notation therein: $g_Z(Z) = \frac{1}{2}\|DZ\|_\aleph^2$, $\alpha = a/2$ and $\beta = b/2$). Denoting by h the density of Z, the argument of (Nourdin and Viens, 2009, Theorem 4.1) is based on the following key formula (see (3.16) in Nourdin and Viens (2009))

$$\|DZ\|_\aleph^2 = \frac{2\int_Z^{+\infty} yh(y)dy}{h(Z)}. \tag{6}$$

For the sake of self-containess, we sketch the main steps of the argument. For any $A > 0$, define $m_A : [0,+\infty) \to \mathbb{R}$ by $m_A(\theta) = \mathbb{E}\left[e^{\theta Z}\mathbf{1}_{\{Z\leq A\}}\right]$. We have $m_A'(\theta) = \mathbb{E}\left[Ze^{\theta Z}\mathbf{1}_{\{Z\leq A\}}\right]$ and integration by part yields

$$\begin{aligned}
m_A'(\theta) &= \int_{-\infty}^A xe^{\theta x}h(x)dx \\
&\leq \theta\int_{-\infty}^A e^{\theta x}\left(\int_x^{+\infty} yh(y)dy\right)dx \qquad (7)\\
&\leq \frac{\theta}{2}\mathbb{E}\left[\|DZ\|_\aleph^2 e^{\theta Z}\mathbf{1}_{\{Z\leq A\}}\right]. \qquad (8)
\end{aligned}$$

where (7) comes from $\int_A^{+\infty} yh(y)dy \geq 0$ since $\mathbb{E}[Z] = 0$, and (8) comes from (6). Because of (5), we obtain for any $\theta \in (0, 2/a)$:

$$m_A'(\theta) \leq \frac{\theta b}{2-\theta a}m_A(\theta). \tag{9}$$

Solving (9), using $m_A(0) = \mathbb{P}(Z \leq A) \leq 1$ and applying Fatou's Lemma ($A \to +\infty$) yield the following bound for the Laplace transform and any $\theta \in (0, 2/a)$:

$$\mathbb{E}[e^{\theta Z}] \leq \exp\left(-\frac{b}{a}\theta - \frac{2b}{a^2}\ln\left(1-\frac{a\theta}{2}\right)\right).$$

The Chebychev inequality together with a standard minimization entail:

$$\mathbb{P}(Z \geq t) \leq \exp\left(\min_{\theta\in(0,2/a)}\left\{-\left(t+\frac{b}{a}\right)\theta - \frac{2b}{a^2}\ln\left(1-\frac{a\theta}{2}\right)\right\}\right)$$

The minimization is achieved in $\widetilde{\theta} = (2t)/(at+b)$ and gives the first bound in Proposition 3. Applying the same argument to $Y = -Z$, satisfying $\|DY\|_{\aleph}^2 \leq -aY + b$, we derive similarly the second bound. Note in particular that condition 5 implies that $Z \geq -b/a$ so that the left tail only makes sense for $t \in (-b/a, 0)$. □

Remark 1 *Nourdin and Viens (2009) have obtained the bounds*

$$\phi_l(t;a,b) = \exp\left(-\frac{t^2}{b}\right) \quad \text{and} \quad \phi_r(t;a,b) = \exp\left(-\frac{t^2}{at+b}\right).$$

Table 1 proposes a comparison of these bounds with ours through the comparisons of the values of their reciprocal functions since these quantities are of great interest for the considered problem. Observe that the most important differences occur when n is moderate. The example $a = 4/\sqrt{n}$ and $b = 4$ corresponds approximately to the choices of parameters that will be used in the next sections.

		$\alpha = 1\%$		$\alpha = 2.5\%$		$\alpha = 5\%$		$\alpha = 10\%$	
		$\varphi_l^{-1}(\alpha)$	$\varphi_r^{-1}(\alpha)$	$\varphi_l^{-1}(\alpha)$	$\varphi_r^{-1}(\alpha)$	$\varphi_l^{-1}(\alpha)$	$\varphi_r^{-1}(\alpha)$	$\varphi_l^{-1}(\alpha)$	$\varphi_r^{-1}(\alpha)$
$n = 50$	NV	6.0697	9.2102	5.4324	7.9062	4.8955	6.8751	4.2919	5.7878
	BC	4.4720	7.1547	4.1398	6.9040	3.8372	6.0847	3.4712	5.2008
$n = 100$	NV	6.0697	8.1851	5.4324	7.1048	4.8955	6.2383	4.2919	5.3107
	BC	4.9090	7.3551	4.4966	6.4575	4.1314	5.7249	3.7012	4.9267
$n = 500$	NV	6.0697	6.9492	5.4324	6.1322	4.8955	5.4606	4.2919	4.7235
	BC	5.5334	6.6309	5.0017	5.8810	4.5449	5.2591	4.0218	4.5708
$n = 1000$	NV	6.0697	6.6801	5.4324	5.9190	4.8955	5.2891	4.2919	4.5930
	BC	5.6877	6.4641	5.1259	5.7478	4.6462	5.1513	4.1000	4.4883
$n = 10000$	NV	6.0697	6.2567	5.4324	5.5819	4.8955	5.0168	4.2919	4.3850
	BC	5.9475	6.1931	5.3345	5.5312	4.8159	4.9757	4.2308	4.3536

Table 1: Computations of the quantities $\varphi_l^{-1}(\alpha)$ and $\varphi_r^{-1}(\alpha)$ for the bounds obtained by Nourdin and Viens (2009) (NV) and ours (BC) (see Remark 1 and Proposition 3) for different values of n and α and for the particular case where $a = 4/\sqrt{n}$ and $b = 4$.

Remark 2 *Note that $\varphi_r(\cdot;a,b)$ (resp. $\varphi_l(\cdot;a,b)$) is a bijective function from $(0,+\infty)$ (resp. $(0,b/a)$) to $(0,1)$. Obviously, the index l in φ_l (resp. r in φ_r) indicates we consider the left (resp. right) tails.*

We explain now how Proposition 2 derives from Proposition 3: standard Malliavin calculus shows that, for $Z = \sqrt{n}V_n$, $\|DZ\|_{\aleph}^2 = \frac{1}{n}\sum_{i,j=1}^n X(i)X(j)\rho(j-i)$, see Theorem 2.1 in Breton et al. (2009). The following lemma ensures that condition (5) in Proposition 3 holds true with $a = 2\kappa_n/\sqrt{n}$ and $b = 2\kappa_n$.

Lemma 4 *For $Z = \sqrt{n}V_n$, we have $\|DZ\|_{\aleph}^2 \leq \kappa_n\left(\frac{1}{\sqrt{n}}Z + 1\right)$.*

The proof of Lemma 4 is a very slight modification of the first part of the proof of Theorem 3.1 in Breton et al. (2009) to which we refer. Finally, Proposition 3 applies and entails Proposition 2.

3 Applications to quadratic variations of fractional Brownian motion

3.1 Notation

From now on, B_H stands for a fBm with Hurst parameter $H \in (0,1)$ and with scaling coefficient $C > 0$ and $\mathbf{B}_H$ is the vector of observations at times i/n for $i = 0, \dots, n-1$. We consider a filter a of length $\ell+1$ and order p, that is a vector with $\ell+1$ real components a_i, $0 \le i \le \ell$, satisfying

$$\sum_{q=0}^{\ell} q^j a_q = 0 \text{ for } j = 0, \dots, p-1 \text{ and } \sum_{q=0}^{\ell} q^p a_q \neq 0. \tag{10}$$

For instance, we shall consider the following filters: Increments 1 ($a = \{-1, 1\}$ with $\ell = 1$, $p = 1$), Increments 2 ($a = \{1, -2, 1\}$ with $\ell = 2$, $p = 2$), Daublets 4 ($a = \{-0.09150635,\ -0.15849365,\ 0.59150635,\ -0.34150635\}$ with $\ell = 3$, $p = 2$), Coiflets 6 ($a = \{-0.05142973,\ -0.23892973,\ 0.60285946,\ -0.27214054,\ -0.05142973,\ 0.01107027\}$ with $\ell = 5$, $p = 2$), see *e.g.* Daubechies (2006) and Percival and Walden (2000) for more details. Let $\mathbf{V}^a$ denote the vector $\mathbf{B}_H$ filtered with a and given for $i = \ell, \dots, n-1$ by

$$V^a\left(\frac{i}{n}\right) := \sum_{q=0}^{\ell} a_q B_H\left(\frac{i-q}{n}\right).$$

Let us denote by $\pi_H^a(\cdot)$ and $\rho_H^a(\cdot)$ the covariance and the correlation functions of the filtered series given by (see Coeurjolly (2001))

$$\mathbb{E}[V^a(k)V^a(k+j)] = C^2 \times \pi_H^a(j) \quad \text{with} \quad \pi_H^a(j) = -\frac{1}{2}\sum_{q,r=0}^{\ell} a_q a_r |q-r+j|^{2H} \tag{11}$$

and $\rho_H^a(\cdot) := \pi_H^a(\cdot)/\pi_H^a(0)$ which is independent of C. Finally, define S_n^a and V_n^a as

$$S_n^a := \frac{1}{n-\ell}\sum_{i=\ell}^{n-1} V^a\left(\frac{i}{n}\right)^2$$

and

$$V_n^a := \frac{n^{2H}}{C^2\pi_H^a(0)} S_n^a - 1 = \frac{1}{n-\ell}\sum_{i=\ell}^{n-1}\left(\frac{n^{2H}}{C^2\pi_H^a(0)} \times V^a\left(\frac{i}{n}\right)^2 - 1\right).$$

Note that $V_n^a \overset{d}{=} \frac{1}{n-\ell}\sum_{i=\ell}^{n-1} H_2(X_i^a)$ where $H_2(t) = t^2 - 1$ is the second Hermite polynomial and X^a is a stationary Gaussian process with variance 1 and with correlation function ρ_H^a. Observe that V_n^a, $n \ge 1$, satisfy a law of large number (LLN) and a central limit theorem (CLT)

$$V_n^a \to 0 \text{ a.s.}, \quad \sqrt{n} V_n^a \Rightarrow \mathcal{N}(0, \sigma_{H,a}^2) \tag{12}$$

with explicit variance $\sigma_{H,a}^2$, see Proposition 1 in Coeurjolly (2001), used to derive standard confidence interval for H. In contrast, our argument relies on concentration inequalities: applying Proposition 2 with these notation, we obtain fo all $s, t \ge 0$:

$$\mathbb{P}\left(-s \le \sqrt{n-\ell} V_n^a \le t\right) \ge 1 - \varphi_{r,n-\ell}(t; \kappa_{n,H}^a) - \varphi_{l,n-\ell}(s; \kappa_{n,H}^a) \tag{13}$$

where $\kappa^a_{n,H} = 2\sum_{|i|\le n} |\rho^a_H(i)|$. As previously explained, the action of filtering a discretized sample path of a fBm changes the correlations into summable correlations for the increments. More precisely, it is proved that, for some explicit k_H, $\rho^a_H(i) \sim k_H|i|^{2H-2p}$, see *e.g.* Coeurjolly (2001). Thus, $\rho^a_H(\cdot)$ is summable if $p > H + 1/2$, *i.e.* $\rho^a_H(\cdot)$ is summable for all $H \in (0,1)$ for $p \ge 2$ and only for $H \in (0, 1/2]$ if $p = 1$ (in the case $H = 1/2$, observe that $\rho^a_{1/2}(k) = 0$ for all $|k| \ge \ell$).

One of the aim is to obtain bounds in (13) independently of H and easily computable. Since $\varphi_{l,n}(t,\cdot)$ and $\varphi_{r,n}(t,\cdot)$ are non-decreasing, the bound (13) remains true with $\kappa^a := 2\sup_{H\in(0,\tau)} \|\rho^a_H\|_{\ell^1(\mathbb{Z})}$ replacing $\kappa_{n,H}$. Here, and in the sequel, we set $\tau = 1/2$ when $p = 1$ and $\tau = 1$ when $p \ge 2$. The following section will prove (among other things) that this quantity is finite.

3.2 Bounds of $\|\rho^a_H\|_{\ell^1(\mathbb{Z})}$ independent of H

In this section, we show that $\kappa^a = \sup_{H\in(0,\tau)} \kappa^a_H$ is finite for a large class of filters, including the collection of dilated filters $(a^m)_{m\ge1}$ of a filter a that will be used in the next section. Recall that a^m is the filter of length $m\ell + 1$ with same order p as a and defined for $i = 0, \dots, m\ell$ by

$$a^m_i = \begin{cases} a_{i/m} & \text{if } i/m \text{ is an integer} \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

As a typical example, if $a := a^1 = \{1, -2, 1\}$, then $a^2 := \{1, 0, -2, 0, 1\}$.

Since $\pi^a_H(0) \ne 0$, observe that, for a fixed $i \in \mathbb{Z}$, the functions $H \mapsto \pi^a_H(i)$ and $H \mapsto \rho^a_H(i)$ are continuous respectively on $[0,1]$ and on $(0,1)$. Moreover, since for any filter a,

$$\pi^a_0(0) = -\frac{1}{2}\sum_{q,r=0,q\ne r}^{\ell} a_q a_r = -\frac{1}{2}\sum_{q,r=0}^{\ell} a_q a_r + \frac{1}{2}\sum_{q=0}^{\ell} a_q^2 = \frac{1}{2}\sum_{q=0}^{\ell} a_q^2 > 0, \tag{15}$$

the function $H \mapsto \rho^a_H(i)$ is continuous in 0. In particular, this ensures that for $p = 1$, $\|\rho^a_\cdot\|_{\ell^1(\mathbb{Z})}$ is continuous on $[0, 1/2)$. Actually, this may be not continuous in $1/2$ but nevertheless $\kappa^a = 2\sup_{H\in[0,1/2]} \|\rho^a_H\|_{\ell^1(\mathbb{Z})} < +\infty$ for instance $\kappa^{\{-1,1\}} = 4$ and $\kappa^{\{-1,1\}^2} = 8$. We refer to Appendix A for the computation of the exact values and to Table 3 for the estimation of some other similar constants.

For any filter of order $p \ge 2$, observe that $\pi^a_1(i) = 0$ for all i. Let us consider the following assumption on the filter a, denoted $\mathbf{H}^a$:

$$\tau^a := \sum_{q,r=0}^{\ell} a_q a_r (q-r)^2 \log(|q-r|) \ne 0, \tag{16}$$

with the convention $0\log(0) = 0$. Tab. 2 below shows that Assumption $\mathbf{H^a}$ is satisfied for a large class of filters. Then, from the rule of l'Hospital,

$$\lim_{H\to1^-} \rho^a_H(i) = \frac{\sum_{q,r=0}^{\ell} a_q a_r (q-r+i)^2 \log(|q-r+i|)}{\sum_{q,r=0}^{\ell} a_q a_r (q-r)^2 \log(|q-r|)} < +\infty.$$

Therefore, under $\mathbf{H^a}$, $\rho^a_H(i)$ is a continuous function of $H \in [0,1]$. Actually, the same is true for the ℓ^1-norm of a filter of order $p \ge 2$ as stated in Proposition 5 below.

		m				
	a	1	2	3	4	5
$p=2$	Increments 2	5.55	22.18	49.91	88.72	138.63
	Daublets 4	0.62	2.47	5.56	9.89	15.45
	Coiflets 6	0.61	2.42	5.45	9.69	15.15
$p=3$	Increments 3	13.50	53.98	121.46	215.94	337.40
	Daublets 6	0.49	1.98	4.45	7.90	12.35
$p=4$	Increments 4	41.43	165.70	372.84	662.82	1035.66
	Daublets 8	0.45	1.81	4.08	7.25	11.32
	Symmlets 8	0.45	1.81	4.08	7.25	11.32
	Coiflets 12	0.45	1.79	4.03	7.16	11.19

Table 2: Computations of τ^{a^m} for different filters a and its dilatation a^m for $m = 1, \ldots, 5$.

Proposition 5 *Let a be a filter of order $p \geq 2$ satisfying $\mathbf{H}^a$ in (16). Then $\|\rho_H^a\|_{\ell^1(\mathbb{Z})}$ is a continuous function of $H \in [0,1]$.*

Proof: From (11), we have

$$\rho_H^a(j) = \frac{|j|^{2H}}{\sum_{q,r=0}^{\ell} a_q a_r |q-r|^{2H}} \sum_{q,r=0}^{\ell} a_q a_r \left|1 + \frac{q-r}{j}\right|^{2H}.$$

For $|j| \geq \ell + 1$, we have $q - r + j \geq 0$ for $0 \leq q, r \leq \ell$, so that:

$$\begin{aligned}
\rho_H^a(j) &= \frac{|j|^{2H}}{\sum_{q,r=0}^{\ell} a_q a_r |q-r|^{2H}} \sum_{q,r=0}^{\ell} a_q a_r \left(1 + \frac{q-r}{j}\right)^{2H} \\
&= \frac{|j|^{2H}}{\sum_{q,r=0}^{\ell} a_q a_r |q-r|^{2H}} \sum_{q,r=0}^{\ell} a_q a_r \sum_{k=0}^{+\infty} \frac{(2H)(2H-1)\ldots(2H-k+1)}{k!} \left(\frac{q-r}{j}\right)^k \\
&= \frac{|j|^{2H}}{\sum_{q,r=0}^{\ell} a_q a_r |q-r|^{2H}} \sum_{k=2p}^{+\infty} \frac{(2H)(2H-1)\ldots(2H-k+1)}{k! j^k} \sum_{q,r=0}^{\ell} a_q a_r (q-r)^k. \quad (17)
\end{aligned}$$

Observe that in (17), the outer sum starts at $k = 2p$. This is due to the property (10) of the filter a of order p which implies the following remark:

$$\begin{aligned}
\sum_{q,r=0}^{\ell} a_q a_r (q-r)^k &= \sum_{q,r=0}^{\ell} a_q a_r \sum_{i=0}^{k} \binom{k}{i} q^i (-r)^{k-i} \\
&= \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} \left(\sum_{q=0}^{\ell} a_q q^i \sum_{r=0}^{\ell} a_r r^{k-i}\right) \\
&= 0 \text{ if } k \leq 2p-1.
\end{aligned}$$

As a consequence, for $p \geq 2$, each summand in the outer sum (17) contains the factor $2H - 2$ in the product $(2H)(2H-1)\ldots(2H-k+1)$. Observe that under $\mathbf{H}^a$ in (16), the rule of l'Hospital ensures that the function $\theta_a(H) = (2-2H)/(\sum_{q \neq r} a_q a_r |q-r|^{2H})$ is bounded at $H = 1^-$. Since moreover this function is continuous in H, we derive, under $\mathbf{H}^a$, that $\|\theta_a\|_\infty := \sup_{H \in [0,1)} |\theta_a(H)| < +\infty$.

Now, from (17), we have

$$\begin{aligned}
& |\rho_H^a(j)| \\
= & \left| \theta_a(H)|j|^{2H-2p} \sum_{k=0}^{+\infty} \frac{(2H)(2H-1)(2H-3)\dots(2H-2p-k+1)}{(2p+k)!j^k} \sum_{q,r=0}^{\ell} a_q a_r (q-r)^{k+2p} \right| \\
\leq & |\theta_a(H)||j|^{2H-2p} \sum_{k=0}^{+\infty} \frac{(2p+k-1)!}{(2p+k)!j^k} \sum_{q,r=0}^{\ell} |a_q||a_r||q-r|^{k+2p} \\
\leq & \|\theta_a\|_\infty |j|^{2H-2p} \sum_{q,r=0}^{\ell} |a_q||a_r||q-r|^{2p} \sum_{k=0}^{+\infty} \frac{1}{(k+1)} \left(\frac{|q-r|}{\ell+1} \right)^k \\
\leq & C(a)|j|^{2H-2p} \qquad (18)
\end{aligned}$$

where

$$C(a) = \|\theta_a\|_\infty \sum_{q,r=0}^{\ell} |a_q||a_r||q-r|^{2p} \left(\frac{(\ell+1)\ln(\ell+1)}{\ell} \right) < +\infty.$$

When $p \geq 2$, the bound (18) ensures that the convergence of the series $\sum_{i\in\mathbb{Z}} |\rho_H^a(i)|$ is uniform in $H \in [0,1]$ and thus $H \mapsto \|\rho_H^a\|_{\ell^1(\mathbb{Z})}$ is continuous on $[0,1]$. □

Proposition 5 proves the following bound is finite for a filter a of order $p \geq 2$ satisfying $\mathbf{H^a}$:

$$\kappa^a = 2 \sup_{H\in(0,1)} \kappa_H^a = 2 \sup_{H\in(0,1)} \|\rho_H^a\|_{\ell^1(\mathbb{Z})} = 2 \sup_{H\in[0,1]} \|\rho_H^a\|_{\ell^1(\mathbb{Z})} < +\infty. \qquad (19)$$

As a consequence of this result, this means that the constant κ^a can be obtained by optimizing the function $H \mapsto \|\rho_H^a\|_{\ell^1(\mathbb{Z})}$ on the interval $[0,1]$. See Tab. 3 below for the computation of such constants for different typical filters.

For dilated increment-type filters, we manage to compute the exact value of $\|\rho_H^a\|_{\ell^1(\mathbb{Z})}$ (see Appendix A for more details)

$$\|\rho_H^a\|_{\ell^1(\mathbb{Z})} = 1 + \sum_{k=1}^{\ell-1} \frac{\left|\sum_{j=-\ell}^{\ell} \alpha_j |j+k|^{2H}\right|}{-\sum_{j=1}^{\ell} \alpha_j j^{2H}} + (-1)^{p+1}\epsilon(2H-1)\frac{\sum_{k=-\ell+1}^{\ell} \alpha_k S_{\ell+k-1}^H}{-\sum_{j=1}^{\ell} \alpha_j j^{2H}},$$

where $\alpha_j = \sum_{\substack{q,r=0 \\ q-r=j}}^{\ell} a_q a_r$, $\epsilon(2H-1) := sign(2H-1)$ and where $S_k^H = \sum_{j=0}^{k} j^{2H}$. For the dilated double increments filter $a = \{1,-2,1\}^m$ for example, this leads to $\kappa^{\{1,-2,1\}} = 2 \times 8/3 = 16/3$ and $\kappa^{\{1,-2,1\}^2} = 2 \times \left(2 + \frac{25\log(5) - 27\log(3)}{8\log(2)}\right) \simeq 7.813554$.

		m				
	a	1	2	3	4	5
$p=1$	Increments 1	2	4	6	8	10
$p=2$	Increments 2	2.667	3.907	5.745	7.565	9.376
	Daublets 4	2.250	4.356	6.641	8.906	11.162
	Coiflets 6	2.259	4.327	6.582	8.816	11.042
$p=3$	Increments 3	3.200	3.783	5.396	7.406	9.200
	Daublets 6	2.429	4.516	6.688	8.833	10.966
$p=4$	Increments 4	3.657	4.304	6.364	8.514	10.350
	Daublets 8	2.648	5.026	7.349	9.648	12.044
	Coiflets 12	2.701	5.112	7.459	9.775	12.229

Table 3: Computation of $\sup_{H\in I}\|\rho_H^{a^m}\|_{\ell^1}$ for different filters a and for $m=1,\dots,5$. Note that $I=[0,0.5]$ for $p=1$ and $I=[0,1]$ for $p>1$.

4 Confidence intervals of the Hurst parameter

For any $\alpha\in(0,1)$, denote by $q^a_{\bullet,n}(\alpha):=\left(\varphi_{\bullet,n}\right)^{-1}(\alpha;\kappa^a)$ for $\bullet=l,r$. In order to make easier the presentation, define also

$$x^a_{l,n-\ell}(\alpha):=1-\frac{q^a_{l,n-\ell}(\alpha)}{\sqrt{n-\ell}}\quad\text{and}\quad x^a_{r,n-\ell}(\alpha):=1+\frac{q^a_{r,n-\ell}(\alpha)}{\sqrt{n-\ell}}.$$

Note that Remark 2 above ensures that for any $\alpha\in(0,1)$ and for all $n>\ell$, $x^a_{l,n-\ell}(\alpha)>0$. For further reference, observe that for $\bullet=l,r$ and $n\to+\infty$:

$$q^a_{\bullet,n-\ell}(\alpha)\sim q^a(\alpha):=\sqrt{2\kappa^a\log(1/\alpha)}. \tag{20}$$

In the sequel, we restrict ourselves, to filters of order $p\geq 2$ which allows us to make no assumption on H. Taking a filter of order $p=1$ would have constrained us to assume that $H\leq 1/2$.

4.1 Scaling parameter C known

In this section, we assume, without loss of generality, that $C=1$. Our confidence interval in Proposition 6 below is expressed in terms of the reciprocal function of $g_n(x):=2x\log(n)-\log\left(\pi^a_x(0)\right)$, $x\in(0,1)$. In order to ensure that g_n is indeed invertible, we assume that

$$n\geq\exp\left(\sup_{x\in(0,1)}\frac{\sum_{q,r=0}^{\ell}a_qa_r\log(|q-r|)|q-r|^{2x}}{\sum_{q,r=0}^{\ell}a_qa_r|q-r|^{2x}}\right). \tag{21}$$

In this case, the function g_n is a strictly increasing bijection from $(0,1)$ to $\big(-\log(\pi^a_0(0)),+\infty\big)$. Moreover recall that a filter of length $\ell+1$ requires a sample size $n\geq\ell+1$. Obviously, condition (21) only makes sense if the filter a satisfies:

$$\sup_{x\in(0,1)}\frac{\sum_{q,r=0}^{\ell}a_qa_r\log(|q-r|)|q-r|^{2x}}{\sum_{q,r=0}^{\ell}a_qa_r|q-r|^{2x}}<+\infty.$$

		m				
	a	1	2	3	4	5
$p=2$	Increments 2	3	4	6	9	11
	Daublets 4	4	6	10	13	15
	Coiflets 6	6	11	15	21	26
$p=3$	Increments 3	4	6	10	13	15
	Daublets 6	6	11	15	21	26
$p=4$	Increments 4	4	9	13	17	21
	Daublets 8	7	15	22	29	36
	Symmlets 8	7	15	22	29	36
	Coiflets 12	12	23	34	44	56

Table 4: Minimal sample size n required to satisfy (21) for different dilated filters a^m of different orders p.

Since $\lim_{x\to 1^-} \sum_{q,r=0}^{\ell} a_q a_r |q-r|^{2x} = 0^-$ (we stress that this function vanishes with non-positive values of because it is continuous, negative in $x=0$, see (15), and does not vanish), the previous condition is equivalent to the more explicit following one

$$\sum_{q,r=0}^{\ell} a_q a_r \log(|q-r|)(q-r)^2 \geq 0. \tag{22}$$

Table 4 exhibits the minimal sample size n required to satisfy (21) for different filters a^m (for $m=1,\ldots,5$) with different order $p=2,3,4$. Obviously, condition (22) is in force for all these filters.

We state now our main result when the scaling parameter is known:

Proposition 6 *Let $\alpha \in (0,1)$ be fixed and a be a filter satisfying $\mathbf{H}^a$ in (16)*

1. *For $n \geq \ell+1$, we have:*

$$\mathbb{P}\Big(\log\left(x^a_{l,n-\ell}(\alpha/2)\right) - \log\left(S^a_n\right) \leq g_n(H) \leq \log\left(x^a_{r,n-\ell}(\alpha/2)\right) - \log\left(S^a_n\right) \Big) \geq 1-\alpha. \tag{23}$$

2. *Moreover if the filter a satisfies (22) and $n \geq \ell+1$ satisfies (21), we have:*

$$\mathbb{P}\left(H \in \left[\widetilde{H}^{\inf}_n(\alpha), \widetilde{H}^{\sup}_n(\alpha) \right] \right) \geq 1-\alpha, \tag{24}$$

where

$$\begin{aligned} \widetilde{H}^{\inf}_n(\alpha) &:= \max\left(0, g_n^{-1}\left(\log\left(x^a_{l,n-\ell}(\alpha/2)\right) - \log\left(S^a_n\right)\right)\right) \\ \widetilde{H}^{\sup}_n(\alpha) &:= \min\left(\tau, g_n^{-1}\left(\log\left(x^a_{r,n-\ell}(\alpha/2)\right) - \log\left(S^a_n\right)\right)\right). \end{aligned}$$

3. *As $n \to +\infty$, the proposed confidence interval in (24) satisfies almost surely*

$$\left[\widetilde{H}^{\inf}_n(\alpha), \widetilde{H}^{\sup}_n(\alpha) \right] \to \{H\}$$

and the length μ_n of the confidence interval satisfies

$$\mu_n \sim \frac{2q^a(\alpha/2)}{\sqrt{n}} \frac{1}{g_n'(H)} \sim \frac{q^a(\alpha/2)}{\sqrt{n}\log(n)},$$

where q^a is defined above in (20).

Remark 3 *Proposition 6 generalizes Proposition 1 derived from Breton et al. (2009). The scaling parameter is still assumed to be known. However, we do not need to know an upper-bound of H and our condition on n is much sharper than the one required in Proposition 1. As an example, for $a = (1, -2, 1)$, condition* (21) *is satisfied for all $n \geq 3$, whereas the minimal sample size allowing to derive a confidence interval from Proposition 1 is* 1108 *for $\alpha = 5\%$ and $H^\star = 0.8$.*

Proof: Consider the set

$$A := \left\{ -q^a_{l,n-\ell}(\alpha/2) \leq \sqrt{n-\ell} V_n^a \leq q^a_{r,n-\ell}(\alpha/2) \right\}.$$

The bound (13) entails $\mathbb{P}(A) \geq 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$. It is now sufficient to notice that

$$\begin{aligned} A &= \left\{ x^a_{l,n-\ell}(\alpha/2) \leq 1 + V_n^a \leq x^a_{r,n-\ell}(\alpha/2) \right\} \\ &= \left\{ x^a_{l,n-\ell}(\alpha/2) \leq \frac{n^{2H}}{\pi_H^a(0)} S_n^a \leq x^a_{r,n-\ell}(\alpha/2) \right\} \\ &= \left\{ \log\left(\frac{x^a_{l,n-\ell}(\alpha/2)}{S_n^a} \right) \leq g_n(H) \leq \log\left(\frac{x^a_{r,n-\ell}(\alpha/2)}{S_n^a} \right) \right\} \end{aligned}$$

which proves (23). Next, since under (21) and (22), g_n is an increasing bijection, (24) comes immediately from (23). Finally, from (20), we have

$$\log\left(x^a_{l,n-\ell}(\alpha/2)\right) \sim -\frac{q^a(\alpha/2)}{\sqrt{n}} \quad \text{and} \quad \log\left(x^a_{r,n-\ell}(\alpha/2)\right) \sim \frac{q^a(\alpha/2)}{\sqrt{n}}$$

as $n \to +\infty$. Moreover, since $1 + V_n^a = \frac{n^{2H}}{\pi_H^a(0)} S_n^a = S_n^a e^{g_n(H)}$, using the LLN in (12), we have almost surely

$$-\log\left(S_n^a\right) = -\log\left(1 + V_n^a\right) + g_n(H) = g_n(H) - V_n^a(1 + o(1)) \sim g_n(H).$$

It is proved in Coeurjolly (2001) (Proposition 1) that V_n^a converges almost surely towards 0 for any filter and for all $H \in (0, 1)$ which implies the almost sure convergence of the confidence interval and the asymptotic behavior of the length μ_n of the confidence interval. □

4.2 Scaling parameter C unknown

The idea to construct confidence intervals when the scaling coefficient C is unknown consists in using the collection of the dilated filters a^m defined in (14).

Let us first introduce some specific notation: let $M \geq 2$ and consider a vector $\mathbf{d} = (d_1, \ldots, d_M)^T$ with non zero real components such that $\sum_{i=1}^M d_i = 0$ and such that $\mathbf{d}^T \mathbf{L_M} > 0$, where $\mathbf{L_M} = (\log(m))_{m=1,\ldots,M}$. Denote by I^- and I^+ the subsets of $\{1, \ldots, M\}$ defined by

$$I^- = \{i \in \{1, \ldots, M\} : d_i < 0\} \quad \text{and} \quad I^+ = \{i \in \{1, \ldots, M\} : d_i > 0\}.$$

The following confidence interval is expressed in terms of $\mathbf{L_{S_n}} := \left(\log\left(S_n^{a^m}\right)\right)_{m=1,\ldots,M}$.

Proposition 7 *Let $\alpha \in (0,1)$ be fixed and denote by $\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\inf}}$ and $\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\sup}}$ the two following vectors with components*

$$\left(\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\inf}}\right)_m = \begin{cases} \log\left(x_{l,n-m\ell}^{a^m}(\alpha/2M)\right) & \text{if } m \in I^- \\ \log\left(x_{r,n-m\ell}^{a^m}(\alpha/2M)\right) & \text{if } m \in I^+ \end{cases}, \left(\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\sup}}\right)_m = \begin{cases} \log\left(x_{r,n-m\ell}^{a^m}(\alpha/2M)\right) & \text{if } m \in I^- \\ \log\left(x_{l,n-m\ell}^{a^m}(\alpha/2M)\right) & \text{if } m \in I^+. \end{cases}$$

1. *Let $n \geq M\ell + 1$. Then we have*

$$\mathbb{P}\left(H \in \left[\widetilde{H}_n^{\inf}(\alpha), \widetilde{H}_n^{\sup}(\alpha)\right]\right) \geq 1 - \alpha \tag{25}$$

where

$$\begin{aligned} \widetilde{H}_n^{\inf}(\alpha) &= \max\left(0, \frac{1}{2\mathbf{d}^T\mathbf{L}_\mathbf{M}}\left(\mathbf{d}^T\mathbf{L}_{\mathbf{S}_\mathbf{n}} - \mathbf{d}^T\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\inf}}\right)\right) \\ \widetilde{H}_n^{\sup}(\alpha) &= \min\left(1, \frac{1}{2\mathbf{d}^T\mathbf{L}_\mathbf{M}}\left(\mathbf{d}^T\mathbf{L}_{\mathbf{S}_\mathbf{n}} - \mathbf{d}^T\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\sup}}\right)\right). \end{aligned}$$

2. *As $n \to +\infty$, the proposed confidence interval in (25) satisfies almost surely*

$$\left[\widetilde{H}_n^{\inf}(\alpha), \widetilde{H}_n^{\sup}(\alpha)\right] \to \{H\}$$

and its length μ_n satisfies

$$\mu_n := \frac{\mathbf{d}^T\left(\mathbf{L}_{\mathbf{X}_\mathbf{n}^{\inf}} - \mathbf{L}_{\mathbf{X}_\mathbf{n}^{\sup}}\right)}{2\mathbf{d}^T\mathbf{L}_\mathbf{M}} \sim \frac{1}{\sqrt{n}} \frac{\mathbf{d}^T\mathbf{q}_\mathbf{M}(\alpha/2M)}{\mathbf{d}^T\mathbf{L}_\mathbf{M}}$$

where $\mathbf{q}_\mathbf{M}(\alpha/2M)$ is the vector of length M with components defined by

$$\left(\mathbf{q}_\mathbf{M}(\alpha/2M)\right)_m := \begin{cases} -q^{a^m}(\alpha/2M) & \text{if } m \in I^- \\ q^{a^m}(\alpha/2M) & \text{if } m \in I^+ \end{cases}$$

with q^{a^m} defined in (20).

Remark 4 *Proposition 7 generalizes Proposition 6 since this new confidence interval does not assume that the scaling parameter, C is known. More specifically, note that the definition of the interval does not depend on C. Note also, that if $\mathbf{B}_\mathbf{H}$ were not observed on $[0,1)$ but with a dilatation factor, then the confidence interval would remain unchanged.*

Proof: For $m = 1, \ldots, M$, we consider the following event

$$A_m := \left\{x_{l,n-m\ell}^{a^m}(\alpha/2M) \leq 1 + V_n^{a^m} \leq x_{r,n-m\ell}^{a^m}(\alpha/2M)\right\}.$$

The bounds (13) entails that $\mathbb{P}(A_m) \geq 1 - \frac{\alpha}{2M} - \frac{\alpha}{2M} = 1 - \frac{\alpha}{M}$. First, recall that

$$V_n^{a^m} = \frac{n^{2H}}{C^2\pi_H^{a^m}(0)} S_n^{a^m} - 1 = \gamma \times \frac{1}{m^{2H}} S_n^{a^m} - 1 \quad \text{with } \gamma := \gamma_{C,H,n} = \frac{n^{2H}}{C^2\pi_H^a(0)}.$$

The crucial point in the definition of the confidence interval relies on the fact that γ is independent of m. Second, note that for $m = 1, \ldots, M$:

A_m

$$
\begin{aligned}
= \quad & \left\{\log\left(x_{l,n-m\ell}^{a^m}(\alpha/2M)\right) \le \log\left(1+V_n^{a^m}\right) \le \log\left(x_{r,n-m\ell}^{a^m}(\alpha/2M)\right)\right\} \\
= \quad & \left\{\log\left(x_{l,n-m\ell}^{a^m}(\alpha/2M)\right) - \log(\gamma) \le \log\left(S_n^{a^m}\right) - 2H\log(m)\right. \\
& \qquad\qquad \left. \le \log\left((x_{r,n-m\ell}^{a^m}(\alpha/2M)\right) - \log(\gamma)\right\} \\
= \quad & \left\{\log\left(S_n^{a^m}\right) - \log\left(x_{r,n-m\ell}^{a^m}(\alpha/2M)\right) + \log(\gamma) \le 2H\log(m)\right. \\
& \qquad\qquad \left. \le \log\left(S_n^{a^m}\right) - \log\left((x_{l,n-m\ell}^{a^m}(\alpha/2M)\right) + \log(\gamma)\right\} \\
= \quad & \left\{d_m\left((\mathbf{L}_{\mathbf{S_n}})_m - (\mathbf{L}_{\mathbf{X_n^{\inf}}})_m + \log(\gamma)\right) \le 2d_m H(\mathbf{L_M})_m \le d_m\left((\mathbf{L}_{\mathbf{S_n}})_m - (\mathbf{L}_{\mathbf{X_n^{\sup}}})_m + \log(\gamma)\right)\right\}.
\end{aligned}
$$

Next, we consider the following event

$$
\begin{aligned}
B \quad := \quad & \left\{\mathbf{d}^T\mathbf{L}_{\mathbf{S_n}} - \mathbf{d}^T\mathbf{L}_{\mathbf{X_n^{\inf}}} + \mathbf{d}^T\mathbf{1}\log(\gamma) \le 2H\mathbf{d}^T\mathbf{L_M} \le \mathbf{d}^T\mathbf{L}_{\mathbf{S_n}} - \mathbf{d}^T\mathbf{L}_{\mathbf{X_n^{\sup}}} + \mathbf{d}^T\mathbf{1}\log(\gamma)\right\} \\
= \quad & \left\{\mathbf{d}^T\mathbf{L}_{\mathbf{S_n}} - \mathbf{d}^T\mathbf{L}_{\mathbf{X_n^{\inf}}} \le 2H\mathbf{d}^T\mathbf{L_M} \le \mathbf{d}^T\mathbf{L}_{\mathbf{S_n}} - \mathbf{d}^T\mathbf{L}_{\mathbf{X_n^{\sup}}}\right\} \\
= \quad & \left\{H \in \left[\widetilde{H}_n^{\inf}(\alpha), \widetilde{H}_n^{\sup}(\alpha)\right]\right\}
\end{aligned}
$$

where $\mathbf{1} = (1,\ldots,1)^T$. Since $A_1 \cap A_2 \cap \ldots \cap A_M \subset B$, setting $A^c = \Omega \setminus A$, we have

$$
\begin{aligned}
\mathbb{P}(B) \quad \ge \quad & \mathbb{P}(A_1 \cap \ldots \cap A_M) = 1 - \mathbb{P}\left((A_1 \cap \ldots \cap A_M)^c\right) = 1 - \mathbb{P}(A_1^c \cup \ldots \cup A_M^c) \\
\ge \quad & 1 - \sum_{m=1}^{M} \mathbb{P}(A_m^c) = \sum_{m=1}^{M} \mathbb{P}(A_m) - (M-1) \qquad (26) \\
\ge \quad & M\left(1 - \frac{\alpha}{M}\right) - (M-1) = 1 - \alpha,
\end{aligned}
$$

which ends the proof of (25). Next with the LLN in (12), as $n \to +\infty$, the following estimate holds almost surely

$$
\begin{aligned}
\log\left(S_n^{a^m}\right) \quad = \quad & 2H\log(m) - \log(\gamma) + \log\left(1+V_n^{a^m}\right) \\
= \quad & 2H\log(m) - \log(\gamma) + V_n^{a^m}(1+o(1)),
\end{aligned}
$$

and implies that almost surely, when $n \to +\infty$,

$$
\begin{aligned}
\mathbf{d}^T\mathbf{L}_{\mathbf{S_n}} \quad = \quad & 2H\mathbf{d}^T\mathbf{L_M} - \mathbf{d}^T\mathbf{1}\log(\gamma) + \mathbf{d}^T\left(V_n^{a^m}\right)_{m=1,\ldots,M}(1+o(1)) \\
= \quad & 2H\mathbf{d}^T\mathbf{L_M} + \mathbf{d}^T\left(V_n^{a^m}\right)_{m=1,\ldots,M}(1+o(1)) \\
\to \quad & 2H\mathbf{d}^T\mathbf{L_M}.
\end{aligned}
$$

From (20), one has also the following estimates as $n \to +\infty$:

$$
\left(\mathbf{L}_{\mathbf{X_n^{\inf}}}\right)_m \sim \frac{1}{\sqrt{n}} \times \left\{\begin{array}{ll} -q^{a^m}(\alpha/2M) & \text{if } m \in I^- \\ q^{a^m}(\alpha/2M) & \text{if } m \in I^+ \end{array}\right. , \quad \left(\mathbf{L}_{\mathbf{X_n^{\sup}}}\right)_m \sim \frac{1}{\sqrt{n}} \times \left\{\begin{array}{ll} q^{a^m}(\alpha/2M) & \text{if } m \in I^- \\ -q^{a^m}(\alpha/2M) & \text{if } m \in I^+. \end{array}\right.
$$

These different results imply the almost sure convergence of the confidence interval towards $\{H\}$. For the asymptotic of the length μ_n of the confidence interval, it is sufficient to note that $(\mathbf{L}_{\mathbf{X}^{\inf}} - \mathbf{L}_{\mathbf{X}^{\sup}}) \sim \frac{1}{\sqrt{n}}\mathbf{q}_{\mathbf{M}}(\alpha/2M)$. □

5 Simulations and discussion

5.1 Confidence intervals based on the central limit theorem

5.1.1 Methodology

There exists a very wide litterature on the estimation of the Hurst parameter, see *e.g.* Coeurjolly (2000) and references therein. For all of the available procedures, the confidence interval comes from a limit theorem so that it is of asymptotic very nature. In contrast, our confidence intervals in (24) and (25) are non-asymptotic since they are based on concentration inequalities. In order to compare our procedures, we choose to focus only on one of these procedures which has several similarities with this paper. These procedures are based on discrete filtering and are presented in detail in Coeurjolly (2001). For the sake of self-containess, we first summarize them:

- **Scaling parameter C known.** The procedure is based on the fact that almost surely $\frac{n^{2H}}{\pi_H^a(0)} S_n^a \to 1$, $n \to +\infty$. With the same function $g_n(x) = 2x\log(n) - \log(\pi_x^a(0))$ as the one used to derive the confidence interval in Proposition 6, this yields the estimator:
$$\widehat{H}_n^{std}(a) := g_n^{-1}(-\log(S_n^a)).$$
Note that the confidence interval (24) is very close to this estimator. In particular, the middle of the interval (24) behaves asymptotically as $\widehat{H}_n^{std}(a)$.

- **Scaling parameter C unknown.** The idea of Coeurjolly (2000) in this context is to use the following property of quadratic variations of dilated filters $\mathbb{E}[S_n^{a^m}] = m^{2H}\gamma$ with $\gamma := \frac{C^2\pi_H^a(0)}{n^{2H}}$ and the almost sure convergence of $S_n^{a^m}/\mathbb{E}[S_n^{a^m}]$ towards 1 for all m. The idea is then to estimate H via a simple linear regression of $\mathbf{L_{S_n}}$ on $2\mathbf{L_M}$ for M dilated filters. Here, the notation $\mathbf{L_{S_n}}$ and $\mathbf{L_M}$ are the same as the ones in Proposition 7. This leads to the estimator
$$\widehat{H}_n^{gen}(a, M) := \frac{\mathbf{A}^T\mathbf{L_{S_n}}}{2\|\mathbf{A}\|^2},$$
where $\mathbf{A} = \left(\log(m) - \frac{1}{M}\sum_{m=1}^M \log(m)\right)_{m=1,\ldots,M}$. There is again an analogy between this estimator and our confidence interval in Proposition 7. Indeed, with $\mathbf{d} = \mathbf{A}$, the interval in (25) rewrites
$$\left[\max\left(0, \frac{\mathbf{A}^T\left(\mathbf{L_{S_n}} - \mathbf{L_{X_n^{inf}}}\right)}{2\|\mathbf{A}\|^2}\right), \min\left(1, \frac{\mathbf{A}^T\left(\mathbf{L_{S_n}} - \mathbf{L_{X_n^{sup}}}\right)}{2\|\mathbf{A}\|^2}\right)\right],$$
since $\mathbf{d}^T\mathbf{L_M} = \mathbf{A}^T\mathbf{A} = \|A\|^2$. Again, the middle of this interval behaves asymptotically as $\widehat{H}_n^{gen}(a, M)$. In the particular case $M = 2$ the estimator $\widehat{H}_n^{gen}(a, 2)$ takes the simple following form
$$\widehat{H}_n^{gen}(a, 2) := \frac{1}{2\log 2}\log\left(\frac{S_n^{a^2}}{S_n^{a^1}}\right)$$
and the bounds of the interval in (25) rewrite as
$$\begin{aligned}
\widetilde{H}_n^{\inf}(\alpha) &:= \max\left(0, \frac{1}{2\log 2}\left(\log\left(\frac{S_n^{a^2}}{S_n^{a^1}}\right) - \log\left(\frac{x_{r,n-2\ell}^{a^2}(\alpha/4)}{x_{l,n-\ell}^{a^1}(\alpha/4)}\right)\right)\right) \\
\widetilde{H}_n^{\sup}(\alpha) &:= \min\left(1, \frac{1}{2\log 2}\left(\log\left(\frac{S_n^{a^2}}{S_n^{a^1}}\right) - \log\left(\frac{x_{l,n-2\ell}^{a^2}(\alpha/4)}{x_{r,n-\ell}^{a^1}(\alpha/4)}\right)\right)\right).
\end{aligned}$$

5.1.2 Asymptotic confidence intervals

We refer the reader to Coeurjolly (2001) where the following central limit theorems (CLT) are proved for $\widehat{H}_n^{std}(a)$ and $\widehat{H}_n^{gen}(a,M)$

$$\sqrt{n}\log(n)\frac{\widehat{H}_n^{std}(a)-H}{\sigma_{std}(\widehat{H}_n^{std})} \xrightarrow{d} \mathcal{N}(0,1), \quad n\to+\infty \tag{27}$$

where $\xrightarrow{d}$ stands for the convergence in distribution, $\mathcal{N}(0,1)$ is the normal standard distribution and $\sigma^2_{std}(H) := \frac{1}{2}\left\|\rho_H^a\right\|_{\ell^2(\mathbb{Z})}$, and

$$\sqrt{n}\frac{\widehat{H}_n^{gen}(a,M)-H}{\sigma_{std}(\widehat{H}_n^{gen},M)} \xrightarrow{d} \mathcal{N}(0,1), \quad n\to+\infty \tag{28}$$

where $\sigma^2_{gen}(H,M) := \frac{\mathbf{A}^T\underline{\mathbf{G}}\mathbf{A}}{4\|\mathbf{A}\|^4}$ where $\underline{\mathbf{G}}$ is the $(M\times M)$-matrix defined by $G_{m_1,m_2} = \left\|\rho_H^{a^{m_1},a^{m_2}}\right\|^2_{\ell^2(\mathbb{Z})}$ for $m_1,m_2=1,\ldots,M$, and for all $i\in\mathbb{Z}$

$$\rho_H^{a^{m_1},a^{m_2}}(i) = \frac{-\frac{1}{2}\sum_{q,r=0}^{\ell} a_a a_r |m_1 q - m_2 r + i|^{2H}}{\sqrt{\pi_H^{a^{m_1}}(0)\pi_H^{a^{m_2}}(0)}}.$$

Note that in the special case where $M=2$, the constant $\sigma^2_{gen}(H,2)$ takes the simple form

$$\sigma^2_{gen}(H,2) = \frac{1}{2(\log 2)^2}\left(\left\|\rho_H^{a^1}\right\|^2_{\ell^2(\mathbb{Z})} + \left\|\rho_H^{a^2}\right\|^2_{\ell^2(\mathbb{Z})} - 2\left\|\rho_H^{a^1,a^2}\right\|^2_{\ell^2(\mathbb{Z})}\right).$$

Thanks to the CLTs, (27) and (28) an asymptotic confidence interval to the level $1-\alpha$, $\alpha\in(0,1)$, can be easily constructed

$$IC_\bullet^{clt}(\alpha) = \left[\max\left(0, \widehat{H}_n^\bullet - \Phi^{-1}(1-\alpha/2)\times\frac{\widehat{\sigma^\bullet}}{v_n^\bullet}\right), \min\left(1, \widehat{H}_n^\bullet + \Phi^{-1}(1-\alpha/2)\times\frac{\widehat{\sigma^\bullet}}{v_n^\bullet}\right)\right] \tag{29}$$

where $\bullet = std, gen$, $v_n^{std} = \sqrt{n}\log(n)$, $v_n^{gen} = \sqrt{n}$ and Φ is the cumulative distribution function of a standard Gaussian random variable.

5.2 Comparisons of approaches

In the following tables, we compare, via Monte-Carlo experiments, the confidence intervals based on concentration inequalities (24), (25) and on central limit theorems (29). The fractional Brownian motions have been generated by using the circulant matrix method (*e.g.* Kent and Wood (1997), Coeurjolly (2000)). We have realized a very large simulation study. The "best" results (in terms of choices of the filters a, of the maximum dilatation factor M) are summarized in Table 5 for the standard fractional Brownian motion (*i.e.* $C=1$) and in Table 6 for the general one (*i.e.* C unknown).

In Figure 1, we also compare, in terms of H, the asymptotic lengths of the confidence intervals obtained by each approach.

		$H = 0.2$			$H = 0.5$			$H = 0.8$		
		Cover.	Length	$\widehat{H}$	Cover.	Length	$\widehat{H}$	Cover.	Length	$\widehat{H}$
$n = 50$	CI[i2]	100.0	0.2191	0.1875	100.0	0.2029	0.4832	100.0	0.1553	0.7824
	CLT[i2]	95.2	0.1330	0.2058	97.0	0.1227	0.5013	99.6	0.1125	0.8003
	CI[d4]	100.0	0.2086	0.1886	100.0	0.1941	0.4841	100.0	0.1482	0.7834
	CLT[d4]	94.6	0.1217	0.2050	97.2	0.1133	0.5004	99.2	0.1076	0.7999
$n = 100$	CI[i2]	100.0	0.1298	0.1936	100.0	0.1212	0.4946	100.0	0.0952	0.7931
	CLT[i2]	95.0	0.0800	0.2009	97.6	0.0737	0.5017	99.8	0.0676	0.8003
	CI[d4]	100.0	0.1224	0.1941	100.0	0.1149	0.4949	100.0	0.0902	0.7933
	CLT[d4]	95.6	0.0732	0.2005	96.4	0.0680	0.5012	99.6	0.0646	0.7997
$n = 500$	CI[i2]	99.6	0.0430	0.1994	100.0	0.0408	0.4988	99.8	0.0336	0.7988
	CLT[i2]	94.4	0.0265	0.2004	96.4	0.0244	0.4998	98.8	0.0224	0.7998
	CI[d4]	99.6	0.0402	0.1995	100.0	0.0383	0.4990	99.8	0.0316	0.7989
	CLT[d4]	95.4	0.0243	0.2003	96.0	0.0225	0.4999	98.4	0.0214	0.7998
$n = 1000$	CI[i2]	100.0	0.0274	0.1998	100.0	0.0262	0.4996	100.0	0.0219	0.7997
	CLT[i2]	96.6	0.0169	0.2003	97.6	0.0155	0.5000	99.2	0.0142	0.8001
	CI[d4]	100.0	0.0256	0.1998	100.0	0.0245	0.4996	100.0	0.0205	0.7998
	CLT[d4]	96.4	0.0154	0.2002	97.2	0.0143	0.5000	98.8	0.0136	0.8001
$n = 10000$	CI[i2]	99.8	0.0066	0.2000	100.0	0.0063	0.4999	100.0	0.0055	0.8000
	CLT[i2]	94.2	0.0040	0.2000	96.2	0.0037	0.5000	98.4	0.0034	0.8000
	CI[d4]	99.8	0.0061	0.2000	99.8	0.0059	0.5000	100.0	0.0051	0.8000
	CLT[d4]	94.4	0.0037	0.2000	95.0	0.0034	0.5000	98.2	0.0032	0.8000

Table 5: Monte-carlo experiments based on 500 replications of a fractional Brownian motion with Hurst parameter $H = 0.2, 0.5, 0.8$ and scaling coefficient $C = 1$ (assumed to be known) and for different values of the sample size n. The filters i2 and d4 denote respectively the filter of Increments of order 2 and the Daublets 4.

		$H=0.2$			$H=0.5$			$H=0.8$		
		Cover.	Length	$\widehat{H}$	Cover.	Length	$\widehat{H}$	Cover.	Length	$\widehat{H}$
$n=50$	CLT[i2,2]	95.4	0.5970	0.3225	92.2	0.6776	0.5064	97.2	0.5422	0.7062
	CI[i2,2]	100.0	1.0000	0.5000	100.0	1.0000	0.5000	100.0	1.0000	0.5000
	CLT[i2,5]	89.4	0.3706	0.2121	88.2	0.5083	0.4838	94.2	0.4595	0.7265
	CI[i2,5]	100.0	1.0000	0.5000	100.0	1.0000	0.5000	100.0	1.0000	0.5000
	CLT[d4,2]	98.0	0.4899	0.2685	92.2	0.5817	0.4966	94.4	0.4836	0.7228
	CI[d4,2]	100.0	1.0000	0.5000	100.0	1.0000	0.5000	100.0	1.0000	0.5000
	CLT[d4,5]	86.8	0.3477	0.2064	88.2	0.4848	0.4739	91.8	0.4564	0.7183
	CI[d4,5]	100.0	1.0000	0.5000	100.0	1.0000	0.5000	100.0	1.0000	0.5000
$n=100$	CLT[i2,2]	97.0	0.4689	0.2628	94.0	0.5232	0.4939	98.0	0.4143	0.7604
	CI[i2,2]	100.0	0.9997	0.4999	100.0	1.0000	0.5000	100.0	1.0000	0.5000
	CLT[i2,5]	92.4	0.2907	0.1999	91.2	0.3670	0.4911	91.0	0.3521	0.7682
	CI[i2,5]	100.0	0.9998	0.4999	100.0	0.9992	0.5004	100.0	0.9078	0.5461
	CLT[d4,2]	97.6	0.3865	0.2299	93.6	0.4259	0.4900	93.8	0.3704	0.7690
	CI[d4,2]	100.0	1.0000	0.5000	100.0	1.0000	0.5000	100.0	1.0000	0.5000
	CLT[d4,5]	90.2	0.2691	0.1965	89.4	0.3509	0.4882	90.4	0.3486	0.7655
	CI[d4,5]	100.0	1.0000	0.5000	100.0	0.9993	0.5003	100.0	0.9026	0.5487
$n=500$	CLT[i2,2]	95.8	0.2540	0.2057	92.8	0.2365	0.4997	94.0	0.2095	0.7983
	CI[i2,2]	100.0	0.6990	0.3495	100.0	0.9399	0.5028	100.0	0.6864	0.6568
	CLT[i2,5]	95.0	0.1363	0.2004	93.6	0.1657	0.4980	93.8	0.1712	0.7983
	CI[i2,5]	100.0	0.5772	0.2886	100.0	0.7113	0.5192	100.0	0.5361	0.7319
	CLT[d4,2]	95.2	0.1965	0.2032	93.8	0.1908	0.4987	94.2	0.1820	0.7982
	CI[d4,2]	100.0	0.7002	0.3501	100.0	0.9459	0.5048	100.0	0.6806	0.6597
	CLT[d4,5]	93.6	0.1250	0.1997	93.6	0.1586	0.4977	94.2	0.1700	0.7967
	CI[d4,5]	100.0	0.5972	0.2986	100.0	0.7272	0.5316	100.0	0.5329	0.7335
$n=1000$	CLT[i2,2]	95.4	0.1829	0.2019	93.8	0.1673	0.4988	94.4	0.1485	0.7988
	CI[i2,2]	100.0	0.5500	0.2750	100.0	0.6912	0.5015	100.0	0.5441	0.7279
	CLT[i2,5]	95.0	0.0963	0.1990	92.2	0.1173	0.4992	94.0	0.1211	0.7972
	CI[i2,5]	100.0	0.4596	0.2302	100.0	0.5022	0.5092	100.0	0.4434	0.7779
	CLT[d4,2]	94.6	0.1392	0.2009	93.2	0.1350	0.4981	93.8	0.1287	0.7979
	CI[d4,2]	100.0	0.5491	0.2745	100.0	0.6873	0.5026	100.0	0.5412	0.7294
	CLT[d4,5]	96.0	0.0884	0.1993	92.8	0.1123	0.4998	94.4	0.1203	0.7974
	CI[d4,5]	100.0	0.4725	0.2365	100.0	0.5130	0.5168	100.0	0.4419	0.7790
$n=10000$	CLT[i2,2]	95.0	0.0579	0.2001	95.2	0.0529	0.5010	95.4	0.0469	0.8007
	CI[i2,2]	100.0	0.2179	0.2004	100.0	0.2179	0.5012	100.0	0.2179	0.8009
	CLT[i2,5]	94.4	0.0305	0.2001	94.8	0.0371	0.5002	96.4	0.0383	0.8006
	CI[i2,5]	100.0	0.1594	0.2008	100.0	0.1594	0.5009	100.0	0.1594	0.8013
	CLT[d4,2]	95.0	0.0440	0.2001	95.2	0.0427	0.5006	95.6	0.0407	0.8007
	CI[d4,2]	100.0	0.2165	0.2006	100.0	0.2165	0.5011	100.0	0.2165	0.8011
	CLT[d4,5]	94.4	0.0280	0.2001	94.0	0.0355	0.5001	97.0	0.0381	0.8004
	CI[d4,5]	100.0	0.1633	0.2020	100.0	0.1633	0.5020	100.0	0.1633	0.8023

Table 6: Monte-carlo experiments based on 500 replications of a fractional Brownian motion with Hurst parameter $H = 0.2, 0.5, 0.8$ and scaling coefficient $C = 1$ (assumed to be unknown), for $M = 2, 5$ and for different values of the sample size. The filters i2 and d4 denote respectively the filter of Increments of order 2 and the Daublets 4. For these simulations the vector $\mathbf{d}$ has been fixed to the vector $\mathbf{A}$.

5.3 Discussion

We propose non-asymptotic confidence intervals for the Hurst parameter of a standard or non-standard fBm based on concentration inequalities. They are computable in particular for small sample size and several theoretical improvements are obtained:

- When the scaling parameter C is known, we have refined the confidence interval proposed in Breton et al. (2009): the upper bound $H \leq H^\star < 1$ is relaxed, the condition on the sample size n is sharper and our new confidence intervals are valid for a large class of filter a.
- As a by-product in our way to optimize the numeric bounds, we have slightly improved the bounds obtained by Nourdin and Viens (2009) in the general concentration inequality (see Proposition 2).
- The case where C is unknown has never been considered with concentration inequalities before Proposition 7.
- The asymptotic properties are similar to that of confidence intervals based on central limit theorems. More specifically, the length of the confidence intervals derived by concentration inequalities behaves asymptotically as the ones of confidence intervals based on central limit theorems, that is $1/(\sqrt{n}\log(n))$ when C is known and $1/\sqrt{n}$ when C is unknown.

The comparison with confidence interval based on CLT is contrasted: while the Monte-Carlo experiments are correct when C is known (in terms of coverage rate and of lengths of the confidence intervals), they are not good when C is unknown: the lengths equal often 1, *i.e.* the intervals correspond to $(0, 1)$, when the sample size is small and are about five times larger when n is large. In fact, the confidence intervals derived from concentration inequalities are too much "sympathetic": the coverage rate is rather far from $1 - \alpha$ (based on 500 replications, it is even often equal to 100%). From a statistical point of view, this is the main reason why the length of the confidence interval is sometimes much larger than the ones based on central limit theorems. From a mathematical point of view, this is due to the fact that, in Proposition 7, the dilatations of a filter are actually handled separately. As a consequence, the errors induced by each dilatation, and controled by the concentration inequalities (3)–(4), add up, see (26). This explains that the proposed confidence interval based on concentration inequalities are less performing in this case while, in comparison, multivariate CLT are used for standard confidence intervals. Improvements would require to use multivariate concentration inequalities, generalizing Proposition 2, which, at the moment, are not available. This is the aim of future research to obtain such improvements.

As a conclusion, this work is the first attempt to define *computable* confidence intervals for the Hurst parameter H of a standard and a non-standard fractional Brownian motion with another approach than the classical one based on central limit theorems (at the very exception of Breton et al. (2009) where the first non-asymptotic confidence intervals were derived for the standard fBM with a more theoretical motivation). We did not get around the question of the numerical performances via Monte-Carlo experiments. The conclusion is that, based on concentration inequalities, confidence intervals can be proposed for a large class of filters and without assumption on the Hurst parameter. The performances are comparable to the stantard confidence interval based on CLT when the scale parameter C is known, while the procedure is underperforming when C is unknown. This later case requires preliminary theoretical improvements for multivariate Gaussian quadratic forms that motivate our future studies.

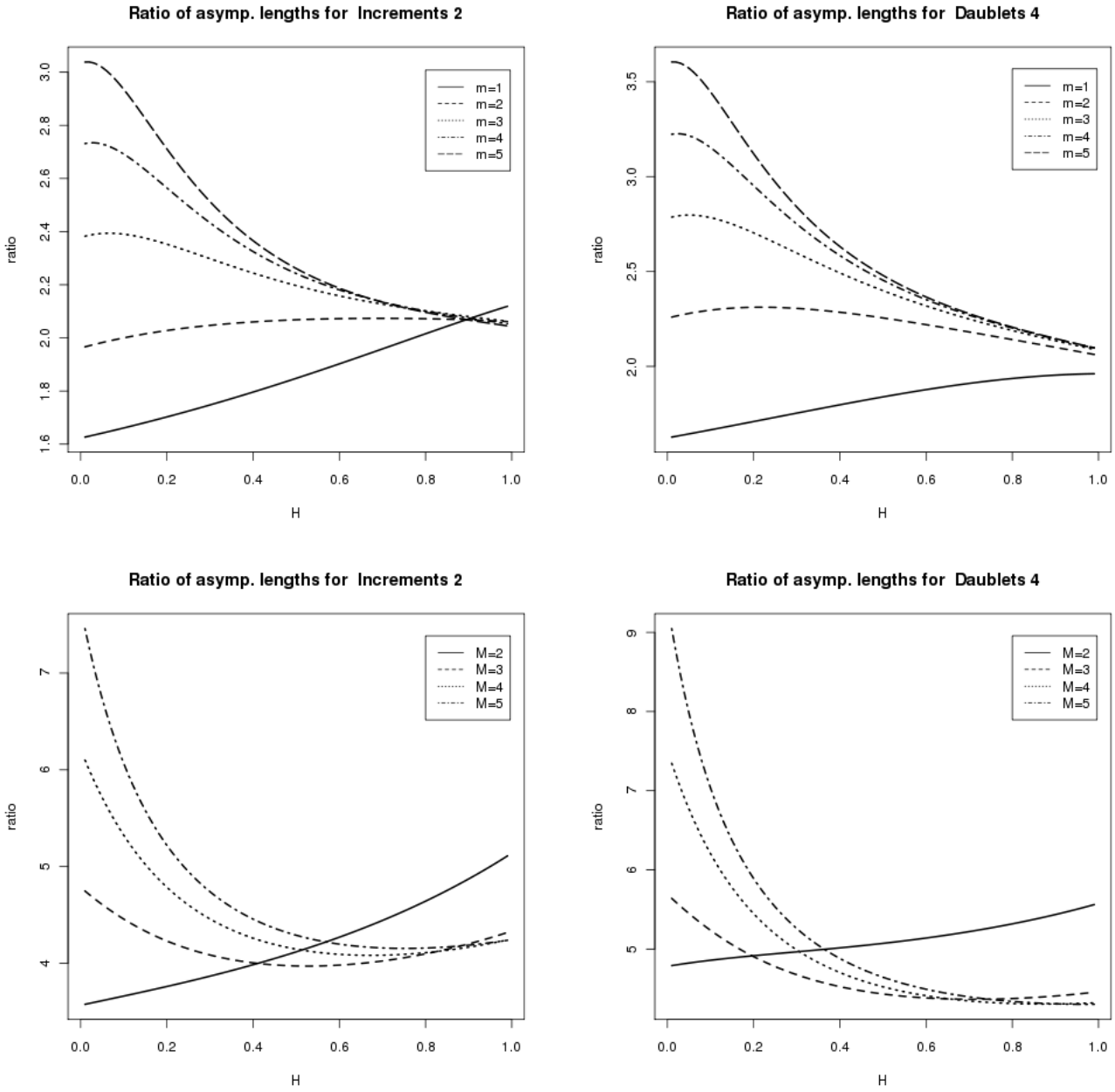


Figure 1: Ratio of asymptotic lengths of confidence intervals of procedures derived by concentration inequalities and central limit theorem when the scaling parameter C is known (top) and unknown (bottom). The confidence level equals $1 - \alpha = 95\%$. For the general procedure, the vector $\mathbf{d}$ has been fixed to $\mathbf{d} := \mathbf{L_M} - \overline{\mathbf{L_M}}$

A Exact computations of ℓ_1-norm for filtered fBm

In this section, we describe how explicit exact bound can be obtained for the correlation of a filtered fBm. Let a be a filter of order p and length ℓ. Its covariance function is given by

$$\pi_H^a(k) \;=\; -\frac{1}{2}\sum_{q,r=0}^{\ell} a_q a_r |q-r+k|^{2H} = -\frac{1}{2}\sum_{j=-\ell}^{\ell}\alpha_j |j+k|^{2H}$$

where $\alpha_j = \sum_{\substack{q,r=0\\ q-r=j}}^{\ell} a_q a_r$. Note that

- $\alpha_j = \alpha_{-j}$, in particular $\pi_H^a(0) = -\sum_{j=1}^{\ell}\alpha_j j^{2H}$;
- $\sum_{j=-\ell}^{\ell}\alpha_j = \sum_{q,r}^{\ell} a_q a_r = 0$,
- for all $h \le 2p-1$, we have

$$\begin{aligned}
\sum_{j=-\ell}^{\ell} j^h \alpha_j &= \sum_{j=-\ell}^{\ell} j^h \sum_{q-r=j} a_q a_r = \sum_{j=-\ell}^{\ell}\sum_{q-r=j}(q-r)^h a_q a_r = \sum_{q,r=0}^{\ell}(q-r)^h a_q a_r \\
&= \sum_{q,r=0}^{\ell}\sum_{k=0}^{h}\binom{h}{k} q^k (-r)^{h-k} a_q a_r \\
&= \sum_{k=0}^{h}\left((-1)^{h-k}\binom{h}{k}\left(\sum_{q=0}^{\ell} q^k a_q\right)\left(\sum_{q=0}^{\ell} r^{h-k} a_r\right)\right) \\
&= 0. \qquad (30)
\end{aligned}$$

- $\sum_{j\neq 0}\alpha_j = -\alpha_0 = -\sum_{q=0}^{\ell} a_q^2 < 0$, $\alpha_\ell = a_0 a_\ell$.

A crucial observation is that, at least for $|k|$ large enough, all the $\pi_H^a(k)$, and thus all the $\rho^a(k)$, have the same sign. Indeed, using (30), we have for $|k| \ge \ell$:

$$\begin{aligned}
\pi_H^a(k) &= -\frac{1}{2}\sum_{j=1}^{\ell}\alpha_j\left(|k+j|^{2H} + |k-j|^{2H} - 2|k|^{2H}\right) \\
&= -\frac{|k|^{2H}}{2}\sum_{j=1}^{\ell}\alpha_j\left((1+j/k)^{2H} + (1-j/k)^{2H} - 2\right) \\
&= -|k|^{2H}\sum_{i=p}^{+\infty}\left(\frac{(2H)(2H-1)\dots(2H-2i+1)}{(2i)!k^{2i}}\left(\sum_{j=1}^{\ell}\alpha_j j^{2i}\right)\right) \\
&\sim -|k|^{2H-2p}\frac{(2H)(2H-1)\dots(2H-2p+1)}{(2p)!}\left(\sum_{j=1}^{\ell}\alpha_j j^{2p}\right).
\end{aligned}$$

This observation allows to reduce the computation of the ℓ^1-norm $\|\rho_H^a\|_{\ell^1(\mathbb{Z})}$, which is an infinite sum with modulus, to an infinite sum of correlations but without modulus plus some finite sum (with modulus remaining). Essentially, it remains to compute the sum of correlation without modulus. This is done below. But observe first that if there exists some $k(H,a) \in \mathbb{N}$ so that the correlations $\rho_H^a(k)$ have all the same sign for $|k| \ge k(H,a)$ large enough. The value $k(H,a)$ is not known in general. However for some family of filters (including increment-type filters in and their dilatations $(in)^m$, $n, m \ge 1$), $k(H,a)$ is known and explicit computations are tractable:

Proposition 8 *For a dilated increment-type filter $a \in \{(in)^m : n, m \geq 1\}$, we have $k(H,a) = \ell$, i.e. the following property holds true:*

$$\text{for all } |j| \geq \ell,\ \pi_H^a(j) \text{ is of the same sign as } (-1)^{p+1}(2H-1). \tag{31}$$

Proof: Let $\theta_m(f)(x) = f(x+m) - 2f(x) + f(x-m)$. Observe that if f is a convex (resp. concave) function, then $\theta_m(f)(x) \geq 0$ (resp. $\theta_m(f)(x) \leq 0$). For the $i1$ filter, we have $\pi_H^{i1}(x) = \frac{1}{2}\theta_1(|x|^{2H})$, for the $i2$ filter, we have $\pi_H^{i2}(x) = -\frac{1}{2}\theta_1^{\circ 2}(|x|^{2H})$ and more generally for the m-dilatation of the in filter, we have $\pi_H^{(in)^m}(x) = \frac{(-1)^{n+1}}{2}\theta_m^{\circ n}(|x|^{2H})$.

Observe also that the function $|x|^{2H}$ and all its iterated derivatives $(|x|^{2H})^{(2p)}$ of even order are convex if $H \geq 1/2$, concave if $H \leq 1/2$. By an immediate induction on n, we show that the same holds true for all $\theta_m^{\circ n}(|x|^{2H})$. In particular for $|j| \geq \ell m$, we obtain that $\pi_H^{(in)^m}(j)$ is of the same sign as $(-1)^{n+1}(2H-1)$. □

Obviously, the property (31) does not hold true for any filter (consider for instance $\{1,-4,5,-2\}$). In order to make easier our following explicit computation to derive exact value for $\|\rho^a\|_{\ell^1(\mathbb{Z})}$, we consider a filter a satisfying (31) but we stress that for each particular filter the same strategy applies with some specific $k(H,a)$. First, for all $N \geq \ell$, we have:

$$\begin{aligned}
-2\sum_{j=\ell}^{N}\pi_H^a(j) &= \sum_{j=\ell}^{N}\sum_{k=-\ell}^{\ell}\alpha_k|j+k|^{2H} \\
&= \sum_{k=-\ell}^{\ell}\alpha_k\sum_{j=\ell}^{N}|j+k|^{2H} = \sum_{k=-\ell}^{\ell}\alpha_k\sum_{j=\ell+k}^{N+k}|j|^{2H} \\
&= \alpha_{-\ell}S_{N-\ell}^H + \sum_{k=-\ell+1}^{\ell}\alpha_k\left(S_{N+k}^H - S_{\ell+k-1}^H\right) \\
&= \alpha_{-\ell}S_{N-\ell}^H + \sum_{k=-\ell+1}^{\ell}\alpha_k\left(S_{N-l}^H + \sum_{j=N-\ell+1}^{N+k}|j|^{2H} - S_{\ell+k-1}^H\right) \\
&= \left(\sum_{k=-\ell}^{\ell}\alpha_k\right)S_{N-l}^H + \left(\sum_{k=-\ell+1}^{\ell}\alpha_k\sum_{j=N-\ell+1}^{N+k}|j|^{2H}\right) - \left(\sum_{k=-\ell+1}^{\ell}\alpha_k S_{\ell+k-1}^H\right) \\
&= x_N - \sum_{k=-\ell+1}^{\ell}\alpha_k S_{\ell+k-1}^H
\end{aligned}$$

where $S_k^H = \sum_{j=0}^{k} j^{2H}$ and

$$\begin{aligned}
x_N &= \sum_{j=N-\ell+1}^{N+\ell}\left(|j|^{2H}\sum_{k=j-N}^{\ell}\alpha_k\right) \\
&= |N+\ell|^{2H}\sum_{i=0}^{2\ell-1}\left(\left(1-\frac{i}{N+\ell}\right)^{2H}\sum_{k=\ell-i}^{\ell}\alpha_k\right) \\
&= |N+\ell|^{2H}\sum_{i=0}^{2\ell-1}\left(\left(1-\frac{2Hi}{N+\ell}+\frac{2H(2H-1)i^2}{2(N+\ell)^2}+O\left(\frac{1}{(N+\ell)^3}\right)\right)\sum_{k=\ell-i}^{\ell}\alpha_k\right)
\end{aligned} \tag{32}$$

But

$$\sum_{i=0}^{2\ell-1}\sum_{k=\ell-i}^{\ell}\alpha_k=\sum_{k=-\ell+1}^{\ell}(\ell+k)\alpha_k=\sum_{k=-\ell}^{\ell}(\ell+k)\alpha_k=0$$

and

$$\begin{aligned}\sum_{i=0}^{2\ell-1}\left(i\sum_{k=\ell-i}^{\ell}\alpha_k\right) &= \sum_{k=-\ell+1}^{\ell}\left(\alpha_k\sum_{i=\ell-k}^{2\ell-1}i\right)=\sum_{k=-\ell}^{\ell}\left(\alpha_k\sum_{i=\ell-k}^{2\ell-1}i\right)\\ &= \sum_{k=-\ell}^{\ell}\alpha_k\left(\frac{2\ell(2\ell-1)}{2}-\frac{(\ell-k)(\ell-k-1)}{2}\right)=0\end{aligned}$$

because of (30). We obtain $x_N=O\big((N+\ell)^{2H-2}\big)\to 0$, $N\to+\infty$. Actually, expanding $(1-i/(N+\ell))^{2H}$ to the $(2p-1)$-th order in (32), and since $\sum_{i=1}^{N}i^k$ is a polynomial in N of degree $k+1$, (30) shows that $x_N=O\big((N+\ell)^{2H-2p+1}\big)$. Finally with the property (31), we have:

$$2\sum_{j=\ell}^{+\infty}|\pi_H^a(j)|=(-1)^{p+1}\epsilon((2H-1)\sum_{k=-\ell+1}^{\ell}\alpha_k S_{\ell+k-1}^H$$

and

$$\begin{aligned}\|\rho_H^a\|_{\ell^1(\mathbb{Z})} &= 1+2\sum_{k=1}^{\ell-1}|\rho_H^a(k)|+2\sum_{k=\ell}^{+\infty}|\rho_H^a(k)|\\ &= 1+\sum_{k=1}^{\ell-1}\left|\frac{\sum_{j=-\ell}^{\ell}\alpha_j|j+k|^{2H}}{\sum_{j=1}^{\ell}\alpha_j j^{2H}}\right|+(-1)^{p+1}\epsilon(2H-1)\frac{\sum_{k=-\ell+1}^{\ell}\alpha_k S_{\ell+k-1}^H}{|\sum_{j=1}^{\ell}\alpha_j j^{2H}|}\\ &= 1+\sum_{k=1}^{\ell-1}\frac{\left|\sum_{j=-\ell}^{\ell}\alpha_j|j+k|^{2H}\right|}{-\sum_{j=1}^{\ell}\alpha_j j^{2H}}+(-1)^{p+1}\epsilon(2H-1)\frac{\sum_{k=-\ell+1}^{\ell}\alpha_k S_{\ell+k-1}^H}{-\sum_{j=1}^{\ell}\alpha_j j^{2H}},\quad(33)\end{aligned}$$

where we recall that $\epsilon(2H-1)=sign(2H-1)$. First, note that the modulus has been removed in the denominator of (33) according to the following observation:

$$\sum_{j=1}^{\ell}\alpha_j j^{2H}=\frac{1}{2}\sum_{j=-\ell}^{\ell}\alpha_j j^{2H}\to_{H\to 0}\frac{1}{2}\sum_{j\neq 0}\alpha_j=\sum_{j=-\ell}^{\ell}\alpha_j-\alpha_0=-\alpha_0<0.$$

Since we assume moreover $\pi_H^a(0)\neq 0$, this means that $\pi_H^a(0)>0$ and that $\left|\sum_{j=1}^{\ell}\alpha_j j^{2H}\right|=-\sum_{j=1}^{\ell}\alpha_j j^{2H}$.

Next, note that (33) is an explicit expression involving only finite sums and can be easily explicitely optimized for $H\in(0,1)$ for every given a satisfying $\mathbf{H}^a$. Note that, for $p\geq 2$, when $H\to 1$, right-hand side of (33) remains well defined. Observe first that since for any fixed k, $\lim_{H\to 1}S_k^H=S_k^1=\frac{k(k+1)(2k-1)}{6}$, we have using (30)

$$\lim_{H\to 1}\sum_{k=-\ell+1}^{\ell}\alpha_k S_{\ell+k-1}^H=\frac{1}{6}\sum_{k=-\ell+1}^{\ell}\alpha_k(\ell+k-1)(\ell+k)(2\ell+2k-1)=0.$$

The same holds true for $\sum_{j=-\ell}^{\ell}\alpha_j|j+k|^{2H}$ and $\sum_{j=1}^{\ell}\alpha_j j^{2H}$, but under $\mathbf{H}^a$ in (16), the rule of l'Hospital entails $\lim_{H\to 1^-}\|\rho_H^a\|_{\ell^1(\mathbb{Z})}$ exists and is finite. Since obviously, $\|\rho_H^a\|_{\ell^1(\mathbb{Z})}$ is a continuous function of $H\in[0,1)$, this ensures the continuity of $\|\rho_H^a\|_{\ell^1(\mathbb{Z})}$ on $[0,1]$ and the constant κ^a in our confidence interval is obtained by maximazing the explicit function in (33).

Dilated simple increments $(i1)^m=\{-1,1\}^m$**.** In this case, $\ell=m$, $p=1$, $\alpha_j=0$ for $1<j<m$ and $\alpha_0=2$, $\alpha_{\pm m}=-1$ so that (33) rewrites:

$$\left\|\rho_H^{\{-1,1\}^m}\right\|_{\ell^1(\mathbb{Z})}=1+\sum_{j=1}^{m-1}\frac{\left||j+m|^{2H}-2|j|^{2H}+|j-m|^{2H}\right|}{m^{2H}}+\frac{S_{2m-1}^H-2S_{m-1}^H}{m^{2H}}. \tag{34}$$

For instance for $m=1$, $\left\|\rho_H^{\{-1,1\}}\right\|_{\ell^1(\mathbb{Z})}=2$ and for $m=2$, $\left\|\rho_H^{\{-1,1\}}\right\|_{\ell^1(\mathbb{Z})}=2\frac{4^H+9^H-1}{4^H}$, so that $\kappa^{i1}=4$ and $\kappa^{(i1)^2}=8$ (recall that in this case, we optimize for $H\in(0,1/2]$).

In general, since the right-hand side of (34) is a continuous function of H, and since for all $k\geq 1$, $S_k^{1/2}=\frac{k(k+1)}{2}$, we have $\lim_{H\to(1/2)^-}\left\|\rho_H^{\{-1,1\}^m}\right\|_{\ell^1(\mathbb{Z})}=2m$ while $\left\|\rho_{1/2}^{\{-1,1\}^m}\right\|_{\ell^1(\mathbb{Z})}=m$, exhibiting a discontinuity of the ℓ^1-norm for the dilated i^1 filters.

Dilated double increments $(i2)^m=\{1,-2,1\}^m$**.** In this case, $\ell=2m$, $p=2$ and $\alpha_0=6$, $\alpha_{\pm m}=-4$, $\alpha_{\pm 2m}=1$, $\alpha_j=0$, $j\neq 0,\pm m,\pm 2m$, so that (33) rewrites:

$$\begin{aligned}\|\rho_H^{\{1,-2,1\}^m}\|_{\ell^1(\mathbb{Z})} &= 1+\sum_{k=1}^{2m-1}\frac{\left||k-2m|^{2H}-4|k-m|^{2H}+6|k|^{2H}-4|k+m|^{2H}+|k+2m|^{2H}\right|}{m^{2H}(4-4^H)}\\ &\quad+\epsilon(1-2H)\frac{-4S_{m-1}^H+6S_{2m-1}^H-4S_{3m-1}^H+S_{4m-1}^H}{m^{2H}(4-4^H)}.\end{aligned}$$

In order to obtain explicit values, we focus on the cases $m=1$ and $m=2$. First, for $m=1$, (33) reduces to

$$\|\rho_H^{i2}\|_{\ell^1(\mathbb{Z})} = \begin{cases}1+\frac{10-7\times 4^H+2\times 9^H}{4-4^H}, & H\leq 1/2\\ 2, & H\geq 1/2\end{cases}$$

and elementary computations entail:

$$\kappa^{i2}=2\times\lim_{H\to 0^+}\|\rho_H^{i2}\|_{\ell^1(\mathbb{Z})}=2\left(1+\frac{5}{3}\right)=\frac{16}{3}.$$

Next, for $m=2$, since

$$\begin{aligned}2\pi_H^{(i2)^2}(1) &= -2+3\times 9^H-25^H &&\geq 0 \quad \forall H\in(0,1)\\ 2\pi_H^{(i2)^2}(2) &= -7\times 4^H+4\times 16^H-36^H &&\leq 0 \quad \forall H\in(0,1)\\ 2\pi_H^{(i2)^2}(3) &= 3-6\times 9^H+4\times 25^H-49^H &&\leq 0 \quad \forall H\in(0,1)\end{aligned}$$

expression (33) reduces to

$$\left\|\rho_H^{(i2)^2}\right\|_{\ell^1(\mathbb{Z})} = \begin{cases}1+\frac{-6+10\times 4^H+12\times 9^H-7\times 16^H-8\times 25^H+2\times 36^H+2\times 49^H}{4^H(4-4^H)} & \text{for } H\leq 1/2\\ 1+\frac{-4+4\times 4^H+6\times 9^H-16^H-2\times 25^H}{4^H(4-4^H)} & \text{for } H\geq 1/2.\end{cases}$$

An elementary study of this function, together with the rule of l'Hospital, entails that

$$\kappa^{(i2)^2}=2\times\sup_{H\in[0,1]}\left\|\rho_H^{(i2)^2}\right\|_{\ell^1(\mathbb{Z})}=2\times\lim_{H\to 1-}\left\|\rho_H^{(i2)^2}\right\|_{\ell^1(\mathbb{Z})}=2\left(1+\frac{25\log(5)-27\log(3)}{8\log(2)}\right)\simeq 7.813554.$$

References

J. Beran. *Statistics for long-memory processes*. Chapman & Hall/CRC, 1994.

J.-C. Breton, I. Nourdin, and G. Peccati. Exact confidence intervals for the hurst parameter of a fractional brownian motion. *Electron. J. Statist.*, 3:416–425, 2009.

J.-F. Coeurjolly. Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *J. Stat. Softw.*, 5(7):1–53, November 2000.

J.-F. Coeurjolly. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Infer. Stoch. Process.*, 4(2):199–227, January 2001.

J.-F. Coeurjolly. Hurst exponent estimation of locally self-similar gaussian processes using sample quantiles. *J. Stat. Softw.*, 36(3):1404–1434, 2008.

I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on Pure and, Applied Mathematics*, 41(7):909–996, 2006.

P. Doukhan, G. Oppenheim, and M.S. Taqqu. *Theory and applications of long-range dependence*. Birkhauser, 2003.

J. Istas and G. Lang. Quadratic variations and estimation of the hölder index of a gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.*, 33:407–436, 1997.

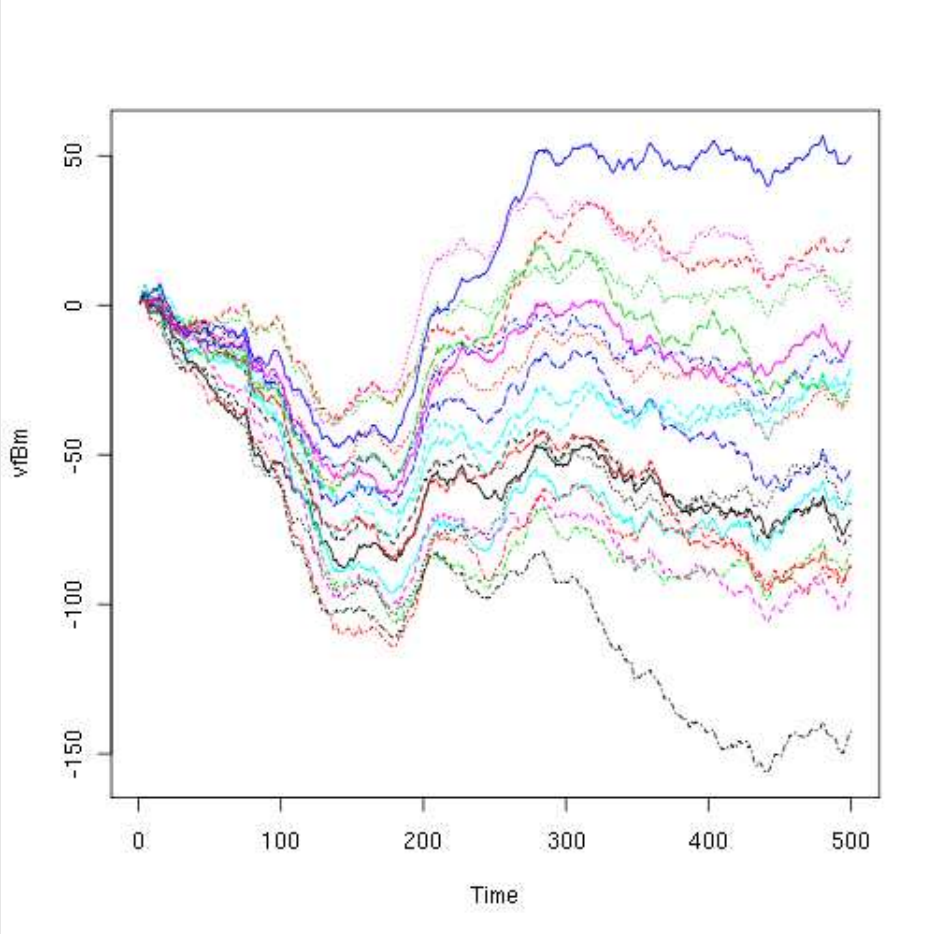
J.T. Kent and A.T.A. Wood. Estimating the fractal dimension of a locally self-similar gaussian process using increments. *J. Roy. Statist. Soc. Ser. B*, 59:679–700, 1997.

B. Mandelbrot and J. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968.

I. Nourdin and F.G. Viens. Density formula and concentration inequalities with malliavin calculus. *Elec. J. Probab.*, 14:2287–2309, 2009.

D. B. Percival and A. T. Walden. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, 2000.

Haipeng Shen, Zhengyuan Zhu, and Thomas C. M. Lee. Robust estimation of the self-similarity parameter in network traffic using wavelet transform. *Signal Process.*, 87(9):2111–2124, 2007.



[CAA10]
J.-F. Coeurjolly and P.O. Amblard and S. Achard. Normalized causal and well-balanced multivariate fractional Brownian motion. *submitted for publication* (available at `http://arxiv.org/abs/1007.2109`), 2010.

Normalized causal and well-balanced multivariate fractional Brownian motion

Jean-François Coeurjolly, Pierre-Olivier Amblard, Sophie Achard
GIPSAlab/CNRS UMR 5216/ BP46,
38402 Saint Martin d'Hères cedex, France
Jean-Francois.Coeurjolly@gipsa-lab.grenoble-inp.fr

July 14, 2010

Abstract

This paper is devoted to study some properties of an extension of the well-known fractional Brownian motion to the multivariate case. Following recent works from Lavancier *et. al.*, we study the covariance structure of the multivariate fractional Gaussian noise. We evaluate several parameters of the model that allow to control the correlation structure at lag zero between all the components of the multivariate process. We particularly focus on two cases for which we can relate characteristic parameters of the covariance function to parameters of the stochastic representation of the processes. These cases are the causal case, a direct multivariate generalization of Mandelbrot&Van Ness representation, and the well-balanced case which adds to the previous case an anti-causal filtering of a Brownian motion. The characterization of the covariance function is then used to study the multivariate fractional Gaussian noise, defined as the increment process of the multivariate fractional Brownian motion. We study the covariance structure as well as the spectral structure of this multivariate stationary process. We exhibit the intriguing facts that two fractional Gaussian noise may be long-range interdependent when only one is long-range dependent. We then perform a wavelet analysis of the multivariate fractional Brownian motion, and show that the wavelet analysis may destroy the long-range interdependence if the wavelet is properly chosen.

1 Introduction

Long-range dependence or memory is the accepted term to design long-range correlations in time series. It is defined as the non integrability of the correlation function due to a very slow decay at infinite lags. This slow decay is usually modeled as a power law $\tau^{\alpha-1}$ with an exponent α lower than one for long-range dependence. In the frequency domain, this corresponds to the divergence of the power spectrum at small frequencies, again with a power law $1/f^{\alpha}$. This type

of behavior may lead to dramatic difficulties when it comes to estimate some parameters from the long-range correlated data. Indeed, rates of convergence of usual estimates are much slower than the usual $1/\sqrt{N}$ rate found for classical mixing processes. Furthermore, this property is not rare at all and is found in many different fields.

In functional Magnetic Resonance Imaging (fMRI), measurements taken from one area of the brain are well modeled by a long-range dependent Gaussian process [3]. In order to study the flow of information using fMRI, neuroscientists have access to multiple correlated measurements. In network traffic monitoring, several measurements can be performed such as IP packet or bytes, and it is now well established that corresponding times series are long-range dependent [1]. Other examples may be found in economy, in biology, in physics, ... Thus, there is a need to develop models of multivariate long-range dependent processes. Models of discrete time series have already been studied recently as generalization of FARIMA models [2, 20].

Here, we will concentrate on a continuous time model recently introduced in [10, 16] as a generalization of fractional Brownian motion (fBm) to multivariate fractional Brownian motion (mfBm). The definition given in [10] concerns a wide scope generalization of the fBm where self-similarity becomes an operator self-similarity for the multivariate case. The authors establish stochastic integral representations of operator self-similar multivariate Gaussian processes with stationary increments, and study some of their properties. Lavancier *et al.* in [16] concentrate on the covariance structure of multivariate processes that are jointly self-similar, and possess stationary increments. Joint self-similarity can be viewed as a particular case of operator self-similarity when the operator is diagonal. In the particular case of Gaussian processes, Lavancier and co-workers then link their general findings with the representations established by Didier&Pipiras. Note that the work in these two papers have close connections with the work of Stoev&Taqqu [18] which concerns stochastic integral representations of the fBm with a time varying Hurst index.

The aim of the paper is to study further the multivariate fractional Brownian motion in the case of joint self-similarity based on the moving-average stochastic integral representation of [10]. In this case, a mfBm is a zero mean Gaussian process, with stationary increments, almost surely zero at time zero, that satisfies $x(\lambda t) = \lambda^H x(t)$ for all $t \in \mathbb{R}$, all $\lambda > 0$, where H is a diagonal matrix. In section 2 we will review some of its known properties. The process is evidently parameterized by the matrix parameter H but also by two matrices A_- and A_+ that control the correlation structure between the components of the process. The covariance structure is known since the work of Lavancier *et al.* However, the covariance depends on the parameters in such a way that it is difficult to generate sample paths with this covariance function. We thus propose another parameterization of the covariance function in order to ease the synthesis step of mfBm. This is done by linking the parameters (matrices A_+, A_-, H) to the correlation between the components at times 1. Even if the choice is arbitrary, it allows to control directly the correlation. Furthermore, we will not solve the problem for any matrices A_+, A_- but we will study two cases called the causal case for which $A_- = 0$ and the well-balanced case for which $A_+ = A_-$. A discussion concerning a more general case will be provided in section 5. The

parameterization we adopt is through the definition of a matrix A which needs to be positive-semidefinite. We give a necessary condition for semidefinite-positivity which can be used to invalidate the model. In section 3 we will concentrate on the increments of the mfBm. We will evaluate the correlation structure as well as the spectral density matrix of the process. In particular, we will exhibit long-range properties in the cross-correlation of different components or equivalently divergence at the zero frequency in the cross-spectral densities. We will also evaluate the coherence function between two components and relate it to the condition of non negativity mentioned above. We will continue the analysis of the process by analysing it through the lens of the wavelet transform. It is well-known now that fractality and long-range dependence may be adequately taken into account by wavelets. Section 4 is devoted to this analysis where we exhibit the ability of the wavelet analysis to reveal fractality (constant relative bandwidth filter bank) and to destroy long-range dependence if the wavelet is correctly chosen (nullity of the first moments). To conclude the paper, we will illustrate the process by depicting some sample paths, by discussing some points concerning more general models and by giving some ideas for further research.

2 Moving-average multivariate fractional Brownian motion

2.1 Model and properties

The fractional Brownian motion, as defined by Mandelbrot&Van Ness [17] is a causal linear transform of a Wiener process, with a kernel that respects self-similarity and which is parameterized by the self-similarity index $H \in (0,1)$. This transform can be generalized in several ways, including time-varying index and non causal integration [18], or operator self-similarity [10]. Here, we concentrate on particular cases of the latter, and study the multivariate fractional Brownian motion (mfBm) defined *via* an integration of the mixing of independent Wiener processes. This comes after the work of Didier&Pipiras in [10] when we restrict the operators involved to be diagonal matrices. Let $x(t)$ of dimension p be defined as

$$x(t) = \int (k_H(u,t)A_+ + l_H(u,t)A_-)dW(u) \tag{1}$$

where W is a vector of p independent standardized Wiener processes or Brownian motions, A_+ and A_- are $p \times p$ matrices of reals, H is a diagonal matrix of parameters $H_j \in (0,1), \forall j = 1,\ldots,p$, $k_H(u,t)$ is a matrix of kernels that reads $(t-u)_+^{H-1/2} - (-u)_+^{H-1/2}$ and $l_H(u,t)$ reads $(u-t)_+^{H-1/2} - (u)_+^{H-1/2}$. In this notation, $(a)_+ = \max(a,0)$ and t^H is understood as the exponential of a matrix $\exp(H\log(t))$. The terms $-(\pm u)_+^{H-1/2}$ insure that the mfBm is almost surely zero at time zero. As seen in the stochastic integral representation (1) of the mfBm, $x(t)$ is a multivariate non-stationary Gaussian process with stationary increments. Moreover, the components of $x(t)$ are correlated, and the structure of the correlation is inherited from the

presence of the mixing matrices A_+ and A_-. The correlation structure is sufficient to completely determine the process since it is Gaussian with zero mean (as a linear transform of a zero mean Gaussian process).

2.2 Covariances and cross-covariances

The following analysis relies heavily on the paper of Lavancier *et al.*, [16]. In this paper, the authors exhibit the general structure of the covariance of a zero mean multivariate self-similar process, that is a process that satisfies $x(\lambda t) = \lambda^H x(t)$ (in the sense of finite-dimensional distributions), where H is a diagonal matrix. As a particular case, the covariance of the mfBm is evaluated directly from the integral representation of Didier&Pipiras [10].

Let $r_{jk}(s,t) = E[x_j(s)x_k(t)]$ denote the cross-covariance of the components j and k of x. For the sake of simplicity, let $B_{jk} = B(H_j + .5, H_k + .5)$ where $B(x,y)$ is the beta function. Let $\sigma_j, j = 1,\dots,p$ be positive numbers, and $\rho_{jk}, j = 1,\dots,p, k > j$ be real numbers in $[-1,1]$. Among all possible models based on (1), our objective is to concentrate ourselves on those allowing us to parameterize the matrices A_+ and A_- only in terms of $\sigma_j = E[x_j(1)^2]$ and $\sigma_j\sigma_k\rho_{jk} = E[x_j(1)x_k(1)]$ (for $j,k = 1,\dots,p$). From proposition 3.1 of Lavancier *et al.*, [16], it consists in finding A_+ and A_- such that

$$\sigma_j\sigma_k\rho_{jk} = \frac{B_{jk}}{\sin(\pi(H_j+H_k))}\left(\mathcal{A}_{jk}+\mathcal{A}_{kj}\right) \tag{2}$$

$$\mathcal{A} = \cos(\pi H)A_+A_+^t + A_-A_-^t\cos(\pi H) - \sin(\pi H)A_+A_-^t\cos(\pi H) - \cos(\pi H)A_+A_-^t\sin(\pi H) \tag{3}$$

where $\cos(\pi H)$ and $\sin(\pi H)$ are diagonal matrices, with jth diagonal term defined as $\cos(\pi H_j)$ and $\sin(\pi H_j)$, and where this equation is valid only if $H_j + H_k \neq 1$. In general, equation (3) cannot be solved to determine explicitly A_+ and A_-. In this paper, we mainly focus on two particular cases: the causal case where $A_- = 0$ and the well-balanced case where $A_+ = A_-$. In the causal case, the integral representation is a direct generalization of the integral representation of Mandelbrot&Van Ness to the multivariate case. The well-balanced case by Stoev&Taqqu in one dimension [18], corresponds to $A_+ = A_-$. More general cases will be discussed in section 5. Note that in the well-balanced case, the existence of the integral representation is subjected to the restriction $H_i \neq 1/2, \forall i = 1,\dots,p$. This point is further discussed in section 2.3 below.

Theorem 2.1 of [16] states that two different cases have to be considered when evaluating the covariance, namely $H_j + H_k \neq 1$ and $H_j + H_k = 1$. We will show here that these two cases can be merged for the causal and the well-balanced mfBm. Before stating our first result let us define the matrix A:

Causal case $A_- = 0$:

$$A_{jj} = \frac{\sigma_j^2\sin(\pi H_j)}{B_{jj}} \tag{4}$$

$$A_{jk} = \begin{cases} \dfrac{\sigma_j \sigma_k \rho_{jk} \sin(\pi(H_j + H_k))}{(\cos(\pi H_j) + \cos(\pi H_k))B_{jk}} & \text{if } H_j + H_k \neq 1 \\ \dfrac{2\sigma_j \sigma_k \rho_{jk}}{(\sin(\pi H_j) + \sin(\pi H_k))B_{jk}} & \text{if } H_j + H_k = 1 \end{cases} \tag{5}$$

Well-balanced case $A_- = A_+$:

$$A_{jj} = \frac{\sigma_j^2 \sin(\pi H_j)}{2(1 - \sin(\pi H_j))B_{jj}} \tag{6}$$

$$A_{jk} = \begin{cases} \dfrac{\sigma_j \sigma_k \rho_{jk} \sin(\pi(H_j + H_k))}{2(\cos(\pi H_j) + \cos(\pi H_k) - \sin(\pi(H_j + H_k)))B_{jk}} & \text{if } H_j + H_k \neq 1 \\ \dfrac{\sigma_j \sigma_k \rho_{jk}}{(\sin(\pi H_j) + \sin(\pi H_k) - 2)B_{jk}} & \text{if } H_j + H_k = 1 \end{cases} \tag{7}$$

Note that the restriction that none Hurst parameters should be equal to $1/2$ in the well-balanced-case appears clearly in the matrix definition, since this case would lead to undefined entries. We can then state the following proposition:

Proposition 1 *if A is positive-semidefinite, the process $x(t)$ defined by*

$$x(t) = \int k_H(u,t) A_+ dW(u) \text{ in the causal case} \tag{8}$$

$$x(t) = \int (k_H(u,t) + l_H(u,t)) A_+ dW(u) \text{ in the well-balanced case} \tag{9}$$

where A_+ is a square root of A, i.e. $A = A_+ A_+^t$, is a vector of p correlated fBm of parameters $H_j, j = 1 \ldots, p$ and the parameterization of the matrix A is such that $r_{jj}(1,1) = E[x_j(1)^2] = \sigma_j^2$ and $r_{jk}(1,1) = E[x_j(1)x_k(1)] = \sigma_j \sigma_k \rho_{jk}$.

The proof of this proposition is immediate and is a direct use of theorem 2.1 and proposition 3.1 of [16] in the restricted cases considered here. Basically, the covariance is evaluated directly from the integral representation. The diagonal form of the kernel allows an easy evaluation. Note that the matrix A needs to be positive-semidefinite. This will be discussed later.

The covariance matrix of the process can then be parameterized as follows. We introduce the function $w_{jk}(t)$ for $t \in \mathbb{R}$, which is defined as

$$w_{jk}(t) = \begin{cases} c_{kj}(t)|t|^{H_j + H_k} & \text{if } j = k \text{ or } j \neq k \text{ and } H_j + H_k \neq 1 \\ |t| + f_{jk} t \log|t| & \text{if } j \neq k \text{ and } H_j + H_k = 1. \end{cases} \tag{10}$$

where $c_{jk}(t) = c_{jk}\mathbf{1}_{\mathbb{R}^+}(t) + c_{kj}\mathbf{1}_{\mathbb{R}^-}(t)$ and where
Causal case $A_- = 0$:

$$\begin{aligned} c_{jk} &= \frac{2\cos(\pi H_j)}{\cos(\pi H_j) + \cos(\pi H_k)} \quad (11) \\ f_{jk} &= \frac{2(H_k - H_j)}{B_{jk}(\sin(\pi H_j) + \sin(\pi H_k))} \\ &= \frac{2}{\pi \tan(\pi H_j)} = -\frac{2}{\pi \tan(\pi H_k)} = -f_{kj}. \quad (12) \end{aligned}$$

Well-balanced case $A_- = A_+$:

$$\begin{aligned} c_{jk} &= 1 \quad (13) \\ f_{jk} &= 0. \quad (14) \end{aligned}$$

Equipped with these definitions, we state the following:

Proposition 2 *For $(j,k) \in \{1,\dots,p\}^2$ and $(H_j, H_k) \in (0,1)^2$, the covariance between the jth and the kth component of a mfBm reads*

$$r_{jk}(s,t) = \frac{\sigma_j \sigma_k \rho_{jk}}{2}\Big\{ w_{kj}(s) + w_{jk}(t) - w_{jk}(t-s) \Big\}. \quad (15)$$

Once again, the proof of this result is a direct application of proposition 1 above and of theorem 2.1 in [16]. Several comments can now be made from this result.

- For $j = 1,\dots,p$, the j-th component $x_j(t)$ of $x(t)$ is a fractional Brownian motion, and we recover from (15) the well-known form of the covariance of a scalar fBm

$$r_{jj}(s,t) = \frac{\sigma_j^2}{2}\Big\{ |s|^{2H_j} + |t|^{2H_j} - |t-s|^{2H_j} \Big\}.$$

  where we have set $\rho_{jj} = 1$ of course.

- Then, note that A_{jj} can be obtained from A_{jk} when $j = k$ and $\rho_{jj} = 1$. Note also that when $H_j = H_k$, $c_{jk}(t) = c_{jk} = 1$. Thus, in this particular case, the cross-covariance function is proportional to the cross-covariance function of a fBm with Hurst parameter H_j.

- The limit of A_{jk} when $H_j + H_k \to 1$ is equal to the definition of A_{jk} when $H_j + H_k = 1$. This can be easily verified using elementary trigonometric identities. For example in the causal case, omitting σ's and ρ's, A_{jk} can be written as $\sin(\pi(H_j + H_k)/2)/\cos(\pi(H_j - H_k)/2)$ whereas A_{jj} writes $1/(\sin(\pi(H_j + H_k)/2).\cos(\pi(H_j - H_k)/2))$. Thus A_{jk} for $H_j + H_k = 1$ could have been defined by continuity.

- In the same spirit, the form of the covariance for $H_j + H_k \neq 1$ converges as $H_j + H_k \to 1$ to the form of the covariance obtained for $H_j + H_k = 1$. This is evident in the well-balanced case since for $H_j + H_k \neq 1$, $w_{jk}(t) = |t|^{H_j+H_k}$ and for $H_j + H_k = 1$, $w_{jk}(t) = |t|$. For the causal case, the proof of the assertion needs some more care. First it is easy to prove that $(H_j + H_k - 1)c_{kj}(t) \to f_{jk}\text{Sign}(t)$ when $H_j + H_k \to 1$. For this expand $\cos(\pi H_k) = -\cos(\pi(H_j + H_k - 1) - \pi H_j)$ as $\cos(\pi H_j) + \pi \sin(\pi H_j)(H_j + H_k - 1) + o(H_j + H_k - 1)$ and remember that $f_{jk} = -f_{kj}$. Then, let $\alpha = H_j + H_k$. Note that from eq. (11) we have $c_{jk}(t) + c_{kj}(t) = 2$. Then,

$$\begin{aligned}\frac{2r_{ij}(s,t)}{\sigma_j\sigma_k\rho_{jk}} &= c_{jk}(s)|s|^{\alpha-1} + c_{kj}(t)|t|^{\alpha-1} - c_{kj}(t-s)|t-s|^{\alpha-1} \\ &= (2 - c_{kj}(s))|s| + c_{kj}(t)|t| - c_{kj}(t-s)|t-s| \\ &+ (2 - c_{kj}(s))|s|^{\alpha-1} - |s|) + c_{kj}(t)|t|^{\alpha-1} - |t|) - c_{kj}(t-s)(|t-s|^{\alpha-1} - |t-s|)\end{aligned}$$

  It is easy to show that $(2 - c_{kj}(s))|s| + c_{kj}(t)|t| - c_{kj}(t-s)|t-s| = |s| + |t| - |t-s|$. Moreover, $c_{kj}(t)(|t|^{\alpha-1} - |t|) \to f_{jk}\text{Sign}(t)|t| \log|t| = f_{jk}t \log|t|$ which concludes the proof of the assertion. Thus, the case $H_j + H_k = 1$ can be defined by continuity from the case $H_j + H_k \neq 1$.

- In the well-balanced case $A_+ = A_-$, the covariance function takes the simple expression

$$r_{jk}(s,t) = \frac{\sigma_j\sigma_k\rho_{jk}}{2}(|s|^{H_j+H_k} + |t|^{H_j+H_k} - |t-s|^{H_j+H_k}).$$

  This result is due to the time reversibility ot the mfBm when $A_+ = A_-$, as observed by Didier&Pipiras in [10]. Time reversibility is clearly observed in the integral representation (1) when $A_+ = A_-$, but this condition is absolutely not necessary to insure time reversibility (see [10] for more details and a necessary and sufficient condition on the mixing matrices). Note finally that in the case $A_- = 0$, the process is not time reversible, and this is reflected in the more complicated structure of the covariance function.

2.3 On the validity of the stochastic representation

Didier&Pipiras give conditions for the existence of the representation. For the diagonal operator self-similarity considered here, the condition of existence of the time representation we use is that $H_i \neq 1/2, \forall i = 1, \ldots, p$.

However, in the causal case, this condition can be relaxed since representation given in eq. (8) is valid. Consider $B_{c,il}(t) = \int k_{H_i}(u,t)dW_l(u)$. Since H is diagonal, the kernel matrix k_H is also diagonal and the process may be written $x(t) = \sum_{l=1}^{p} A_{+,il}B_{c,il}(t)$. What happens to $B_{c,il}(t)$ when $H_i \to 1/2$? The kernel $k_{H_i}(t,u)$ converges to $\mathbf{1}_{\mathbb{R}^+}(t-u) - \mathbf{1}_{\mathbb{R}^+}(-u) = \mathbf{1}_{[0,t]}(u)$, the indicator function of the interval $[0,t]$. Thus, since $W_l(0) = 0$ almost surely, $B_{il}(t) = W_l(t)$ is a standard Wiener process.

In the well-balanced case, the previous analysis leads to a problem. If we introduce $B_{ac,il}(t) = \int l_{H_i}(u,t)dW_l(u)$, then $x(t) = \sum_{l=1}^{p} A_{+,il}(B_{c,il}(t) + B_{ac,il}(t))$. The kernel $l_{H_i}(t,u)$ converges to $\mathbf{1}_{\mathbb{R}^+}(u-t) - \mathbf{1}_{\mathbb{R}^+}(u) = -\mathbf{1}_{[0,t]}(u)$. Thus, $B_{ac,il}(t)$ converges to $-W_l(t)$, and thus $x(t) = 0$ almost surely. We recover this fact by evaluating the variance of the well-balanced process (involved in eq. (6)) which is equal to zero.

In all other cases, $A_{+,il}B_{c,il}(t) + A_{-,il}B_{ac,il}(t)$ converges to $(A_{+,il} - A_{-,il})W_l(t)$ as $H_i \to 1/2$, and if the ith line of A_+ is not equal to the ith line of A_- then $x(t)$ is well-defined.

However, it may be shown (see [10]) that for $H_i = 1/2$ the following stochastic representation holds for $x_i(t)$

$$x_i(t) = \sum_{l=1}^{p} \int \Big((\mathbf{1}_{\mathbb{R}^+}(t-u) - \mathbf{1}_{\mathbb{R}^+}(-u))A_{+,il} + \log\Big(\frac{|t-u|}{|u|}\Big)A_{-,il}\Big) dW_l(u)$$

This however introduces more special cases and we prefer to assume that $H_i \neq 1/2, \forall i = 1, \ldots, p$.

2.4 Semidefinite-positivity of A

The aim of this section is to examine the semidefinite-positivity condition of the matrix A defined by equation (5) in the causal case and by equation (7) in the well-balanced case. This condition is the main limitation of this model. Indeed, if the matrix A is not positive-semidefinite, it cannot be factorized into $A_+A_+^t$.

The first comment to be made is the fact that if $H_1 = \ldots = H_p = \widetilde{H}$, then A is positive-semidefinite. Indeed, it is easily verified that $A = c(\widetilde{H}) \times R(1,1)$ where $c(\widetilde{H}) = \sin(\pi\widetilde{H})/B(\widetilde{H}+.5, \widetilde{H}+.5)$ in the causal case, and $c(\widetilde{H}) = \sin(\pi\widetilde{H})/(2(1-\sin(\pi\widetilde{H}))B(\widetilde{H}+.5, \widetilde{H}+.5))$ in the well-balanced case. $R(1,1)$ is the covariance matrix of the mfBm at times (1,1). Hence, as the product between a positive constant and a positive-semidefinite matrix, A is positive-semidefinite. In this particular case, there is no limitation in the model: we can choose $\widetilde{H} \in (0,1)$ whatever $\rho_{jk} \in [-1,1]$ for all j,k.

In the general case, we could not find necessary and sufficient conditions to insure that A is positive-semidefinite. However, we establish the following necessary condition. Let $g(H_j, H_k)$ be defined as

Causal case $A_- = 0$:

$$g(H_j, H_k) = \Gamma(H_j + H_k + 1)\frac{\sin\left(\frac{\pi}{2}(H_j + H_k)\right)}{\cos\left(\frac{\pi}{2}(H_j - H_k)\right)}. \tag{16}$$

Well-balanced case $A_- = A_+$:

$$g(H_j, H_k) = \Gamma(H_j + H_k + 1)\sin\Big(\frac{\pi}{2}(H_j + H_k)\Big) \tag{17}$$

Proposition 3 *If A defined by equations (5) or (7) is positive-semidefinite, then*

$$\rho_{jk}^2 \frac{g(H_j, H_k)^2}{g(H_j, H_j)g(H_k, H_k)} \leq 1, \forall j \neq k \tag{18}$$

Proof. Let $z_{jk}^t = (0, 0, ..., z_j, 0, \ldots, 0, z_k, 0, \ldots, 0)$ be a vector whose all elements are zero except the jth and the kth. Since A is nonnegative, $z_{jk}^t A z_{jk} \geq 0$. Let B the 2×2 submatrix of A corresponding to the elements at the intersection of the jth and kth lines with the jth and kth columns. Then $z_{jk}^t A z_{jk} = (z_j z_k) B (z_j z_k)^t \geq 0$. Thus B is non negative and its determinant is positive. In the causal case, this determinant is given by

$$\sigma_j^2 \sigma_k^2 \frac{\sin(\pi H_j) \sin(\pi H_k)}{B_{jj} B_{kk}} - \sigma_j^2 \sigma_k^2 \rho_{jk}^2 \frac{\sin(\pi(H_j + H_k))^2}{(\cos(\pi H_j) + \cos(\pi H_k)) B_{jk}^2}$$

Factorize the positive quantity $\sigma_j^2 \sigma_k^2 \frac{\sin(\pi H_j) \sin(\pi H_k)}{B_{jj} B_{kk}}$, use $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$ and elementary trigonometric identities to get the result. The same of kind of simple calculations and noting the identity $(1 - \sin x)(1 - \sin y) = (\cos((x - y)/2) - \sin((x + y)/2))^2$ give the result in the well-balanced case. Since j, k, z_j, z_k are arbitrary, this ends the proof. ■

Even if this condition is only a necessary one, it gives a useful condition to be fulfilled by the parameters. Indeed, if the condition is violated we are ensured that the model is not defined.

For the 2 dimensional case, the condition is obviously necessary and sufficient. From (18), the condition depends only on H_j, H_k and $\rho = \rho_{jk}$. A plot is feasible to determine the range of possible parameters, see Fig. 1. For the causal and the well-balanced cases, we observe that the higher $|H_j - H_k|$ (resp. the lower), the lower the maximal possible correlation ρ (resp. the higher).

3 Increments of the multivariate fractional Brownian motion

This section aims at exploring the covariance structure and the spectral density matrix of the increments of size δ of the multivariate fractional Brownian motion. Let $\Delta_\delta x(t) = x(t+\delta) - x(t)$ denote the increments of the multivariate fractional Brownian motion of size δ and let $\Delta_\delta x_j(t)$ be its jth component.

3.1 Covariances and cross-covariances

Let $\gamma_{jk}(h, \delta) = E[\Delta_\delta x_j(t) \Delta_\delta x_k(t + h)]$ denote the cross-covariance of the increments of size δ of the components j and k. Expanding the expectation and using the covariance (15), we deduce

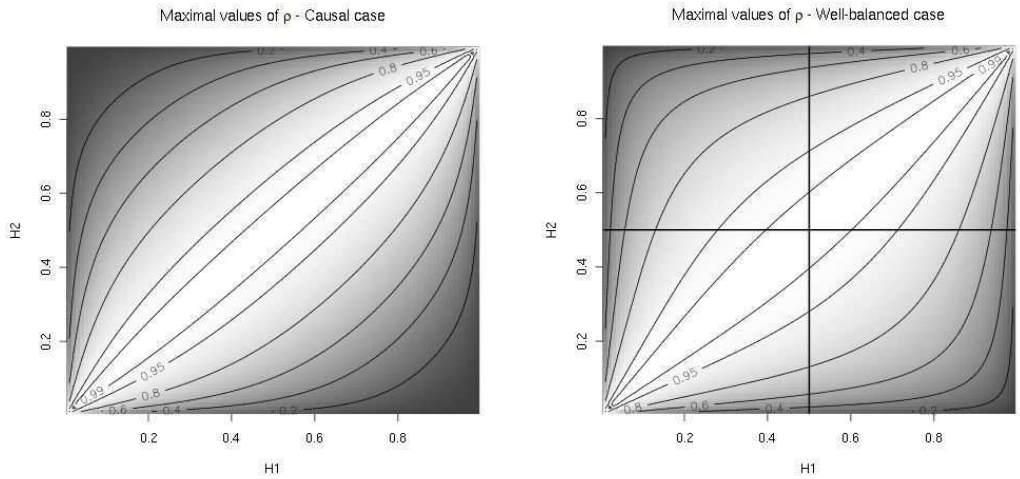


Figure 1: Maximal values of the absolute possible correlation parameter $|\rho_{12}|$ ensuring that the matrix A is positive-semidefinite, in terms of H_1 and H_2.

that $\gamma_{jk}(h,\delta)$ is given by

$$\gamma_{jk}(h,\delta) = \frac{\sigma_j\sigma_k\rho_{jk}}{2}\Big(w_{jk}(h-\delta) - 2w_{jk}(h) + w_{jk}(h+\delta)\Big). \tag{19}$$

We reproduce the definition, eq. (10), of w_{jk} for convenience

$$w_{jk}(t) = \begin{cases} c_{kj}(t)|t|^{H_j+H_k} & \text{if } j=k \text{ or } j\neq k \text{ and } H_j+H_k\neq 1 \\ |t| + f_{jk}t\log|t| & \text{if } j\neq k \text{ and } H_j+H_k = 1. \end{cases}$$

The first comment is that the result confirms that the increment process is a multivariate stationary random process. Stationarity is in the strict sense since the process is Gaussian.

In the well-balanced cased ($c_{jk}=1$ and $f_{jk}=0$), we observe that for all $H_j, H_k \neq 1/2$,

$$\gamma_{jk}(h,\delta) = \frac{\sigma_j\sigma_k\rho_{jk}}{2}\Big(|h-\delta|^{H_j+H_k} - 2|h|^{H_j+H_k} + |h+\delta|^{H_j+H_k}\Big). \tag{20}$$

Thefore, in this case, $\gamma_{jk}(h,\delta)$ is proportional to the covariance of a fractional gaussian noise with Hurst parameter $(H_j+H_k)/2$. In particular, it is a symmetric function with respect to h.

The causal case is different from the well-balanced case since, when $H_j+H_k\neq 1$, $c_{kj} = 2\cos(\pi H_k)/(\cos(\pi H_j)+\cos(\pi H_k)) \neq c_{jk}$. Note that when $H_k = 1/2$, we observe (since $c_{kj}=0$) that $\gamma_{jk}(h,\delta)=0$ for $h\geq\delta>0$. Let us also observe that, the case $H_j=H_k=1/2$ leading to $f_{jk}=0$ makes $\gamma_{jk}(h,\delta)$ proportional to the covariance of the increments of size δ of a Brownian motion. We now turn to the analysis of some of the properties of the covariance of the increments.

3.1.1 Long-memory type properties of the cross-covariance

For two functions f and g, we denote by $f \sim g$ when $\lim f(h)/g(h) = 1$, as $|h| \to +\infty$.

Proposition 4 *As $|h| \to +\infty$, we have for any $\delta > 0$*

$$\gamma_{jk}(h,\delta) \sim \frac{\sigma_j \sigma_k \rho_{jk}}{2} \delta^2 |h|^{H_j+H_k-2} \times \tau_{jk}(h), \tag{21}$$

where

$$\tau_{jk}(h) = \begin{cases} c_{kj}(h)(H_j+H_k)(H_j+H_k-1) & \text{if } j=k \text{ and } H_j \neq 1/2 \\ & \text{or } j \neq k \text{ and } H_j+H_k \neq 1 \\ f_{jk} \times Sign(h) & \text{if } H_j+H_k=1 \text{ and } H_j \neq 1/2. \end{cases}$$

The cases $j = k$ and $H_j = 1/2$ and $j \neq k$ and $H_j = H_k = 1/2$ are omitted since they correspond to the covariance of the increments of a Brownian motion and therefore in these cases, $\gamma_{jk}(h,\delta) = 0$ for $|h| \geq \delta$.

In Section 2.2, we mentioned that $c_{kj}(h)(H_j + H_k - 1) \sim f_{jk}\text{Sign}(h)$ as $H_j + H_k \to 1$, which makes the second case the limit of the first one.

Proof. Define $B(h) := w_{jk}(h-\delta) - 2w_{jk}(h) + w_{jk}(h+\delta)$. Without loss of generality, let us choose h such that $|h| \geq \delta$. For the first case, let $\alpha = H_j + H_k$. In this case, it is sufficient to note that $c_{kj}(h) = c_{kj}(h+\delta) = c_{kj}(h-\delta)$ and

$$B(h) \;=\; c_{kj}(h)|h|^\alpha \left(\left(1 - \frac{\delta}{h}\right)^\alpha - 2 + \left(1 + \frac{\delta}{h}\right)^\alpha \right) \sim c_{kj}(h)|h|^\alpha \alpha(\alpha-1)\delta^2 h^{-2}.$$

For the second case, for $|h| \geq \delta$, the expression of $B(h)$ reduces to

$$B(h) = f_{jk}\Big((h+\delta)\log\left(1+\frac{\delta}{h}\right) + (h-\delta)\log\left(1-\frac{\delta}{h}\right)\Big).$$

Using the expansion of $\log(1 \pm x)$ as $x \to 0$, we obtain $B(h) \sim f_{jk}\delta^2 h^{-1} = f_{jk}\delta^2|h|^{-1} \times \text{Sign}(h)$, which is the expected result. ■

At this point several interesting remarks may be done. First, setting $j = k$ and $\rho_{jj} = 1$ allows us to recover the well-known asymptotic behavior for the covariance of a monovariate fGn $\sigma_j^2 H_j(2H_j-1)\delta^2|h|^{2H_j-2}$ (see [17]). When $H_j + H_k = 1$ but $H_j \neq 1/2$, the increments of size δ are long-range interdependent since their cross-covariance is not summable. Note that in this case one fGn is long-range dependent and the other is necessarily not. When $H_j + H_k \neq 1$, the same conclusion may be drawn. If the two fGn are long-range dependent ($H_j > 1/2$ and $H_k > 1/2$), then necessarily they are long-range interdependent. Interestingly, two fGn can be long-range inderdependent when only one is long-range dependent.

3.1.2 Behavior of $\gamma_{jk}(.,\delta)$ for large h

Let $h \geq \delta$ and $\rho_{jk} \geq 0$. When $H_j + H_k \neq 1$

$$\gamma_{jk}(h,\delta) = \sigma_j \sigma_k \rho_{jk} c_{kj} \ \times \ \widetilde{\gamma}_{\frac{H_j+H_k}{2}}(h,\delta),$$

where $\widetilde{\gamma}_H(h,\delta)$ is the covariance function of a fGn (with size δ) with Hurst parameter H and with variance 1. Recall that for $h \geq \delta$, $\widetilde{\gamma}_H(\cdot,\delta)$ is a negative and increasing (resp. positive and decreasing) function when $H < 1/2$ (resp. $H > 1/2$). This corresponds to the behavior of $\gamma_{jk}(.,\delta)$ in the well-balanced case since $c_{kj} = 1$. In the causal case, we may derive the following statement (by studying the sign of c_{kj}) illustrated by Fig. 2:

$$\text{For } h \geq \delta, \ \ \gamma_{jk}(h,\delta) \begin{cases} \text{is negative and increasing} & \text{when } H_k < 1/2 \\ \text{is positive and decreasing} & \text{when } H_k > 1/2 \\ \text{equals zero} & \text{when } H_k = 1/2. \end{cases}$$

Let us underline that the study of the function $(h-\delta)\log(h-\delta) - 2h\log(h) + (h+\delta)\log(h+\delta)$ leads to the same conclusion when $H_j + H_k = 1$.

3.2 Spectral density and cross-spectral density

In all the following, the convention adopted for the Fourier transform $FT(f(t))$ of a function f is $F(\omega) = \int f(t)\exp(-i\omega t)dt$. Depending on the context, and this will be detailed, the transform will be understood in the L^1, L^2 or even in the generalized function sense. The inverse transform reads $f(t) = 1/(2\pi)\int F(\omega)\exp(i\omega t)d\omega$.

Even if for some values of $H_j + H_k$, the covariance $\gamma_{jk}(h,\delta)$ may be in L^1 or L^2, it is not the case for all values, and thus we evaluate the spectral density matrix in the generalized function (distribution) sense (see *e.g.* [14]).

Proposition 5 *(i) The Fourier transform of $\gamma_{jk}(h,\delta)$, denoted by $S_{jk}(\omega,\delta)$ is given for all j,k and for all H_j, H_k by*

$$\begin{aligned} S_{jk}(\omega,\delta) &= \sigma_j\sigma_k\rho_{jk}\frac{1-\cos(\omega\delta)}{|\omega|^{H_j+H_k+1}}\Gamma(H_j+H_k+1) \times \zeta_{jk}(\omega) \\ \zeta_{jk}(\omega) &= \begin{cases} -c_{kj}e^{-i\mathit{Sign}(\omega)\frac{\pi}{2}(H_j+H_k+1)} - c_{jk}e^{i\mathit{Sign}(\omega)\frac{\pi}{2}(H_j+H_k+1)} & \text{if } j=k \text{ or } j\neq k, H_j+H_k\neq 1 \\ 2 - i\pi f_{jk}\mathit{Sign}(\omega) & \text{if } j\neq k \text{ and } H_j+H_k = 1. \end{cases} \end{aligned} \quad (22)$$

In the causal and well-balanced cases, this reduces for all H_j, H_k to:
Causal case $A_- = 0$:

$$S_{jk}(\omega,\delta) = 2\sigma_j\sigma_k\rho_{jk}\ g(H_j,H_k)\ \frac{1-\cos(\omega\delta)}{|\omega|^{H_j+H_k+1}}e^{-i\mathit{Sign}(\omega)\frac{\pi}{2}(H_k-H_j)}. \quad (23)$$

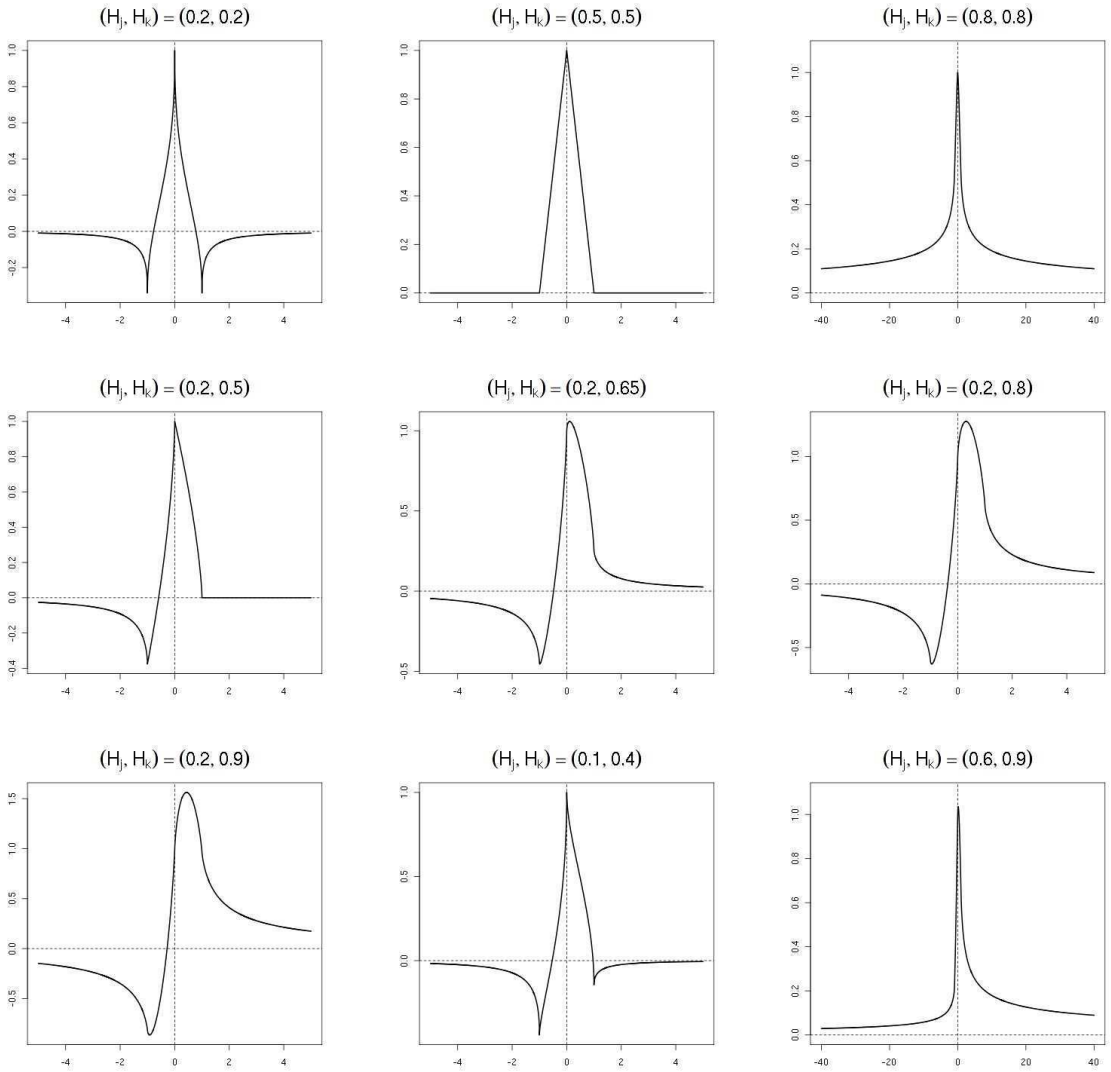


Figure 2: Examples of cross-covariance functions for the causal case for different parameters H_j, H_k. Without loss of generality, the parameters σ_j, σ_k and ρ_{jk} are fixed to 1. In the well-balanced, from (20), the top left (resp. right) plot corresponds to the covariance for all the values of H_j, H_k such that $H_j + H_k = 0.2$ (resp. 0.8). The top middle one has no sense in the well-balanced case since H_j and H_k must be different of $1/2$.

Well-balanced case $A_- = A_+$:

$$S_{jk}(\omega,\delta) = 2\sigma_j\sigma_k\rho_{jk}\; g(H_j,H_k)\; \frac{1-\cos(\omega\delta)}{|\omega|^{H_j+H_k+1}}, \tag{24}$$

where the function g is defined by (16) and (17).
(ii) For any fixed δ, we have for both cases, as $\omega \to 0$

$$|S_{jk}(\omega,\delta)| \sim \sigma_j\sigma_k\rho_{jk}\; g(H_j,H_k)\; \delta^2\; |\omega|^{1-H_j-H_k}. \tag{25}$$

(iii) Moreover, the coherence function between the two components j and k satisfies:

$$C_{jk}(\omega,\delta) := \frac{|S_{jk}(\omega,\delta)|^2}{S_{jj}(\omega,\delta)S_{kk}(\omega,\delta)} = \rho_{jk}^2 \times \frac{g(H_j,H_k)^2}{g(H_j,H_j)g(H_k,H_k)}. \tag{26}$$

Equations (23) and (24) are easily derived from (22) using elementary algebra and the definitions of c_{jk} and f_{jk}. Before turning to the proof, let us give some remarks concerning theses results:

- Unlike the covariance $\gamma_{jk}(h,\delta)$, note that the expression of $S_{jk}(\omega,\delta)$ is unchanged when $j \neq k$ and $H_j + H_k = 1$. For example in the causal case, it reduces to

$$\frac{2\sigma_j\sigma_k\rho_{jk}}{\sin(\pi H_j)}\frac{1-\cos(\omega\delta)}{|\omega|^2}e^{i\mathrm{Sign}(\omega)\pi(H_j-\frac{1}{2})}.$$

- When $j = k$ or when $H_j = H_k$, we recover the standard real spectral density function of a fGn. Actually, in the well-balanced case, $\forall j,k$, $S_{jk}(\omega,\delta)$ corresponds to the spectral density function of a fGn with Hurst parameter $\frac{H_j+H_k}{2}$.

- The analysis of the local behavior in a neighborhood of zero of the cross-spectral density (25) leads to the same remarks as the ones done in Section 3.1.1 concerning long-memory type properties.

- Finally, let us underline the fact that the coherence function is independent of the frequency. Furthermore, we recover the necessary condition of proposition 3. Indeed, the coherence must be lower than one, a condition satisfied if

$$\rho_{jk}^2 \times \frac{g(H_j,H_k)^2}{g(H_j,H_j)g(H_k,H_k)} \leq 1$$

Proof. We only concentrate on the cross-spectra since (25) and (26) are then easily derived. We denote by $FT(\cdot)$ the Fourier transform in the generalized function sense (see [14]). From (10),

$$\begin{aligned} S_{jk}(\omega,\delta) &:= FT(\gamma_{jk}(h,\delta)) \\ &= \frac{\sigma_j\sigma_j\rho_{jk}}{2}\left(FT(w_{jk}(h-\delta)) - 2FT(w_{jk}(h)) + FT(w_{jk}(h+\delta))\right) \\ &= \frac{\sigma_j\sigma_j\rho_{jk}}{2}(2\cos(\omega\delta)-2)FT(w_{jk}(h)). \end{aligned} \tag{27}$$

Let us now split the proof into the two different cases.
Case 1. $j=k$ or $j\neq k$ and $H_j+H_k\neq 1$.

Let $\alpha = H_j+H_k$. For $x\in\mathbb{R}$, consider the following generalized functions

$$x_+^\alpha = x^\alpha \mathbf{1}_{\mathbb{R}^+}(x) \quad \text{and} \quad x_-^\alpha = (-x)^\alpha \mathbf{1}_{\mathbb{R}^-}(x).$$

From the definition of $c_{kj}(h)$, the Fourier transform of $w_{jk}(h) = c_{kj}(h)|h|^\alpha$ writes $FT(c_{kj}(h)|h|^\alpha) = c_{kj}FT(h_+^\alpha) + c_{jk}FT(h_-^\alpha)$. The Fourier transforms of $h_\pm^\alpha$ exist in the generalized function sense (see [14]) and read

$$FT(h_+^\alpha) = \Gamma(\alpha+1)\left\{\omega_+^{-\alpha-1}e^{-i\frac{\pi}{2}(\alpha+1)} + \omega_-^{-\alpha-1}e^{i\frac{\pi}{2}(\alpha+1)}\right\} \tag{28}$$

$$FT(h_-^\alpha) = \Gamma(\alpha+1)\left\{\omega_+^{-\alpha-1}e^{i\frac{\pi}{2}(\alpha+1)} + \omega_-^{-\alpha-1}e^{-i\frac{\pi}{2}(\alpha+1)}\right\} \tag{29}$$

or alternatively

$$\begin{aligned} FT(h_+^\alpha) &= \Gamma(\alpha+1)e^{-i\mathrm{Sign}(\omega)\frac{\pi}{2}(\alpha+1)}|\omega|^{-\alpha-1} \\ FT(h_-^\alpha) &= \Gamma(\alpha+1)e^{i\mathrm{Sign}(\omega)\frac{\pi}{2}(\alpha+1)}|\omega|^{-\alpha-1}. \end{aligned}$$

Therefore, we obtain

$$FT(c_{kj}(h)|h|^\alpha) = \Gamma(\alpha+1)|\omega|^{-\alpha-1}\left(c_{kj}e^{-i\mathrm{Sign}(\omega)\frac{\pi}{2}(\alpha+1)} + c_{jk}e^{i\mathrm{Sign}(\omega)\frac{\pi}{2}(\alpha+1)}\right),$$

which, combined with (27), leads to the result.
Case 2. $j\neq k$ and $H_j+H_k=1$.

We have to concentrate on the Fourier transform of $|h|+f_{jk}h\log|h|$. For $\alpha>-1$, the Fourier transform of $|h|^\alpha$ equals $FT(|h|^\alpha) = -2\Gamma(\alpha+1)\sin(\pi\alpha/2)|\omega|^{-\alpha-1}$. Now, let us notice that $FT(|h|^\alpha\log|h|) = FT(\frac{d}{d\alpha}|h|^\alpha) = \frac{d}{d\alpha}FT(|h|^\alpha)$. Setting $\alpha=0$ in this last equation and the explicit derivation of $FT(|h|^\alpha)$ (with respect to α) leads to $FT(\log|h|) = -\pi|\omega|^{-1}$. In order to get $FT(h\log|h|)$, we just have to note that $FT(h\log|h|) = -\frac{1}{i}\times\frac{d}{d\omega}FT(\log|h|)$, leading to $FT(h\log|h|) = i\pi\mathrm{Sign}(\omega)|\omega|^{-2}$. We finally obtain

$$FT(|h|+f_{jk}h\log|h|) = |\omega|^{-2}\left(-2+i\pi f_{jk}\mathrm{Sign}(\omega)\right)$$

Combined with (27), the last equation leads to (22). ■

4 Wavelet Analysis

The use of wavelet analysis in the understanding of the monovariate fractional Brownian motion, and more generally for the study of fractal signals, goes back to the early works of Flandrin [12, 13], Tewfik [19], Wornell [22] to cite some but a few. It is now well accepted that wavelet analysis is the adequate analysis to extract information properly from fractal or multifractal signals. Several causes for this fact can be put forward, such as the adequation between $1/f$-like spectral densities of fractal signals and the constant relative bandwith filter bank underlying the wavelet analysis, ability to "kill" long-range dependence if the wavelet is correctly chosen.

The aim of this section is thus to analyse the multivariate fractional Brownian motion through the lens of the wavelet transform. We use the continuous wavelet transform here, but a similar analysis could be performed in the multiresolution framework using orthonormal wavelet bases. We will consider complex valued wavelets, not necessarily in the Hardy class, not necessarily with compact support. The hypothesis we impose on the wavelets will be detailed when needed.

4.1 Definition and stationarity

Let ψ be a complex wavelet function, let $a > 0$ and $b \in \mathbb{R}$ and consider $\psi_{ab}(.) = a^{-1/2}\psi((.-b)/a)$. Let

$$d^j_{a,b} \;:=\; \left\langle x_j \middle| \psi_{ab} \right\rangle_{L^2} = \int_{\mathbb{R}} x_j(t)\overline{\psi_{ab}}(t)dt \tag{30}$$

the wavelet transform of the jth component of a multivariate fractional Brownian motion. $\overline{\psi}$ denotes the complex conjugate of ψ. In this section, we assume that conditions [C1] and [C2(2)] are satisfied, where:

> [C1] Admissibility condition: $\psi(t) \in L^2$ and $|\widehat{\psi}(\omega)|^2/|\omega| \in L^1$, where $\widehat{\psi}$ is the Fourier transform of ψ.
>
> [C2(K)] $t^m\psi(t) \in L^1$ for $m = 0, 1, \ldots, K$.

Condition [C1] ensures that $\widehat{\psi}(0) = 0$ and that $\int_{\mathbb{R}} \psi(t)dt = 0$. We note, as [15], that under condition [C2(1)], the integral (30) is well-defined as a sample path integral and is a second-order random variable. This follows, since under [C2(1)] we have $\int_{\mathbb{R}} |s|^H|\psi_{ab}(s)|ds < +\infty, \forall H \in (0, 1)$.

The aim of this section is to focus on the correlation between the wavelet transforms (at different scales and different times) of two components j and k of the multivariate fractional Brownian motion. The wavelet transform is a random field. It is clearly zero mean and Gaussian. We have for $a_1, a_2 > 0$ and $b_1, b_2 \in \mathbb{R}$

$$E[d^j_{a_1,b_1}\overline{d^k_{a_2,b_2}}] = \int_{\mathbb{R}^2} r_{jk}(t_1, t_2)\overline{\psi_{a_1b_1}}(t_1)\psi_{a_2b_2}(t_2)dt_1dt_2.$$

Under [C1], and from (10) the last expression reduces to

$$E[d^j_{a_1,b_1}\overline{d^k_{a_2,b_2}}] = -\frac{\sigma_j\sigma_k\rho_{jk}}{2}\int_{\mathbb{R}^2} w_{jk}(t_2-t_1)\overline{\psi_{a_1b_1}}(t_1)\psi_{a_2b_2}(t_2)dt_1dt_2.$$

Let $\Gamma_\psi(v) := \int_{\mathbb{R}} \psi_{a_1b_1}(u)\overline{\psi_{a_2b_2}}(u+v)du$ be the correlation function between the two wavelets ψ_{a_1,b_1} and ψ_{a_2,b_2}. Then we have

$$E[d^j_{a_1,b_1}\overline{d^k_{a_2,b_2}}] = -\frac{\sigma_j\sigma_k\rho_{jk}}{2}\int_{\mathbb{R}} w_{jk}(v)\overline{\Gamma_\psi}(v)dv. \tag{31}$$

Note that [C2(2)] implies that for all the values of H_j and H_k, $\int_{\mathbb{R}} |w_{jk}(v)|\,|\Gamma_\psi(v)|\,dv < +\infty$. With two changes of variables, this may also be rewritten as

$$\begin{aligned} E[d^j_{a_1,b_1}\overline{d^k_{a_2,b_2}}] &= -\frac{\sigma_j\sigma_k\rho_{jk}}{2}\sqrt{a_1a_2}\times \\ &\quad \int_{\mathbb{R}^2} w_{jk}(a_2t_2-a_1t_1+b_2-b_1)\overline{\psi}(t_1)\psi(t_2)dt_1dt_2. \end{aligned} \tag{32}$$

If we interpret for fixed parameters a_1 and a_2, the quantity $E[d^j_{a_2,b_2}\overline{d^k_{a_2,b_2}}]$ as the cross-correlation between two signals, we observe that it depends only on the difference between the times at which it is evaluated (i.e. b_2-b_1). With the fact that the wavelet transform is a zero mean and Gaussian field, we conclude that $d^j_{a_1,\cdot}$ and $d^k_{a_2,\cdot}$ are jointly stationary signals. This is of course because the wavelet transform reveals the stationary increments property hidden in the fractional Brownian motion. The wavelet transform can be seen as a generalized derivative.

4.2 Self-similarity type property of the cross-wavelet transform

The variance of the wavelet transforms at similar scales for the fractional Brownian motion with Hurst parameter H exhibits some self-similarity. Indeed, it is proved in [12] for example that for all b

$$Var(d^j_{a,b}) = a^{2H+1}\times\left(-\frac{\sigma^2}{2}\int_{\mathbb{R}^2}|t_2-t_1|^{2H}\overline{\psi}(t_1)\psi(t_2)dt_1dt_2\right).$$

We note here that the same behavior holds for the cross-wavelet variance.

Proposition 6 *Under the assumptions [C1] and [C2(2)], let $b_1=b_2=b$ and fix $a_1=a_2=a>0$. Then,*

$$\begin{aligned} E[d^j_{a,b}\overline{d^k_{a,b}}] &= \frac{\sigma_j\sigma_k\rho_{jk}}{2}\,z_{jk}\,a^{H_j+H_k+1} \\ Corr[d^j_{a,b},d^k_{a,b}] &= \rho_{jk}\times\frac{z_{jk}}{\sqrt{z_{jj}z_{kk}}}, \end{aligned} \tag{33}$$

where $z_{jk} := -\int_{\mathbb{R}^2} w_{jk}(t_2-t_1)\overline{\psi}(t_1)\psi(t_2)dt_1dt_2$.

Proof. Consider Equation (32). When $H_j + H_k \neq 1$, it suffices to note that for $a > 0$, $c_{kj}(av) = c_{kj}(v)$ and therefore $w_{jk}(av) = a^{H_j+H_k} w_{jk}(v)$. Now, when $H_j + H_k = 1$, the result comes from Condition [C1] ensuring that $\int_{\mathbb{R}^2} f_{jk} \times (t_2 - t_1) \log(a) \overline{\psi}(t_1)\psi(t_2) dt_1 dt_2 = 0$. ■

Let us observe that the instantaneous cross-wavelet correlation is independent of the scale. This is the generalization of the fact that the coherence does not depend on the frequency.

4.3 Cross-correlation structure of the wavelet transform of the mfBm

For fixed scales, a_1, a_2, we now specify the behavior of the cross-wavelet covariance (or correlation) as $|b_2 - b_1| \to +\infty$. In particular, our aim is to exhibit the influence of the number of vanishing moments of the wavelet function on the asymptotic cross-wavelet covariance. Such a result needs the following assumption:

[C3] The wavelet function has $M \geq 1$ vanishing moments that is

$$\int_{\mathbb{R}} t^m \psi(t) dt = 0 \text{ for } m = 0, \ldots, M-1 \qquad \text{and} \qquad \int_{\mathbb{R}} t^M \psi(t) dt \neq 0.$$

We may now derive our result obtained as $|h| = |b_2 - b_1| \to +\infty$.

Proposition 7 *Under the assumptions [C1], [C2(2M+1)] and [C3], then as $|h| \to +\infty$, we have*

$$E[d^j_{a_1,b} \overline{d^k_{a_2,b+h}}] \sim -\frac{\sigma_j \sigma_k \rho_{jk}}{2} \kappa(\psi, M) |h|^{H_j+H_k-2M} \widetilde{\tau}_{jk}(h)$$

where $\kappa(\psi, M) := \binom{2M}{M} (a_1 a_2)^M \left| \int t^M \psi(t) dt \right|^2$ *and*

$$\widetilde{\tau}_{jk}(h) = \begin{cases} c_{kj}(h) \binom{H_j+H_k}{2M} & \text{if } j = k \text{ and } H_j \neq 1/2 \\ & \text{or } j \neq k \text{ and } H_j + H_k \neq 1 \\ \frac{f_{jk} \times Sign(h)}{2M(2M-1)} & \text{if } H_j + H_k = 1 \text{ and } H_j \neq 1/2. \end{cases}$$

Remark 1 *As for Proposition 4, we notice that the second case is the limit of the first one as $H_j + H_k \to 1$. Moreover, let us underline the importance of the number of vanishing moments for the wavelet. Similarly to the fractional Brownian motion, Proposition 7 asserts that the higher M, the less correlated the wavelet transforms of the components j and k of the multivariate fractional Brownian motion. This has many implications. In particular, this suggests that estimating the instantaneous cross-wavelet correlation at a scale a may be efficiently done by using the empirical correlation since at scale a, $d^j_{a,b}$ and $d^k_{a,b+h}$ are not too much correlated if M is large.*

The proof of this result is postponed until Appendix 6.

4.4 Cross-spectral density of the wavelet transform of the mfBm

In the case of fBm, the expression of the spectral density of the wavelet transform was provided by [12] and [13]. A rigorous proof of the existence of this spectral density in the L^1 sense was obtained by [15]. On the basis of this work, our ambition is, to provide the cross-spectral density between wavelet transforms (at different scales) of components j and k of the multivariate fractional Brownian motion.

Proposition 8 *Under Assumptions [C1], [C2(M)] and [C3] (with $M \geq 2$), we derive the following assertions.*
(i) The cross-spectral density of the wavelet transforms of two components j and k, that is the Fourier transform of the function $E[d^j_{a_1,b_1}\overline{d^k_{a_2,b_2}}]$ (in terms of $b_2 - b_1$) exists and is given by

$$\widetilde{S}_{jk}(\omega) \;=\; \frac{\sigma_j\sigma_k\rho_{jk}}{2}\frac{\Gamma(H_j+H_k+1)}{|\omega|^{H_j+H_k+1}}\sqrt{a_1a_2}\;\overline{\widehat{\psi}(-a_1\omega)}\widehat{\psi}(-a_2\omega)\times\zeta_{jk}(\omega) \tag{34}$$

where $\zeta_{jk}(\omega)$ is defined by equation (22). In the causal and well-balanced cases, this reduces for all H_j, H_k to:
Causal case $A_- = 0$:

$$\widetilde{S}_{jk}(\omega) := \frac{\sigma_j\sigma_k\rho_{jk}g(H_j,H_k)}{|\omega|^{H_j+H_k+1}}\sqrt{a_1a_2}\;e^{-iSign(\omega)\frac{\pi}{2}(H_k-H_j)}\overline{\widehat{\psi}(-a_1\omega)}\widehat{\psi}(-a_2\omega) \tag{35}$$

Well-balanced case $A_- = A_+$:

$$\widetilde{S}_{jk}(\omega) := \frac{\sigma_j\sigma_k\rho_{jk}g(H_j,H_k)}{|\omega|^{H_j+H_k+1}}\sqrt{a_1a_2}\;\overline{\widehat{\psi}(-a_1\omega)}\widehat{\psi}(-a_2\omega) \tag{36}$$

where the function g is defined by (16) and (17).
(ii) We have for both cases, as $\omega \to 0$

$$\left|\widetilde{S}_{jk}(\omega)\right| \sim \sigma_j\sigma_k\rho_{jk}g(H_j,H_k)(a_1a_2)^{M+1/2}|\widehat{\psi}^{(M)}(0)|^2|\omega|^{2M-1-\alpha}.$$

(iii) Moreover, the coherence function between the two components j and k satisfies:

$$\widetilde{C}_{jk}(\omega) := \frac{\left|\widetilde{S}_{jk}(\omega)\right|^2}{\widetilde{S}_{jj}(\omega)\widetilde{S}_{kk}(\omega)} = \rho^2_{jk}\times\frac{g(H_j,H_k)^2}{g(H_j,H_j)g(H_k,H_k)}\frac{\widehat{\psi}(-a_1\omega)\overline{\widehat{\psi}(-a_2\omega)}}{\overline{\widehat{\psi}(-a_1\omega)}\widehat{\psi}(-a_2\omega)}. \tag{37}$$

Similarly as proposition 5, equations (35) and (36) are obtained from (34) and using elementary algebra. The proof of this proposition is rejected in the appendices, see section 7. Note that the interpretation of the coherence in this case is difficult. It is complex valued, a property which is

not natural for a coherence. This comes from the fact that the quantities $\widetilde{S}_{jj}(\omega)$ are not power spectral densities but cross-spectral densities (cross-spectral density between two different scales of the wavelet transform of one signal). Thus, to interpret correctly the coherence, we should look at one scale only, in which case we recover the coherence evaluated in the usual spectral domain. And this result is logical since the usual coherence is independent of the frequency.

5 Discussion

To conclude the paper, we make some comments on synthesis, on a more general case and we give some ideas for future works.

5.0.1 On synthesis

The synthesis of monovariate fractional Brownian motion has found an elegant and efficient solution through the use of Wood&Chan method of simulation of Gaussian processes [21]. The method relies on the embedding of the correlation matrix of N regularly spaced samples of a fractional Gaussian noise into a larger circulant matrix. As a circulant matrix, the diagonalization is easy since it relies on the discrete Fourier transform. Furthermore, using the fast Fourier transform, it can be implemented in a very efficient way. Wood&Chan have generalized their technique for synthesizing multivariate Gaussian homogeneous random fields [5]. As a particular case, the simulation of multivariate Gaussian stationary signals can be performed. We have implemented their algorithm to generate the fractional Brownian motion, either in the well-balanced case and in the causal case. We show in figure (3) some examples of sample paths that may be generated using this algorithm. The parameters have been chosen not only to insure that proposition 3 is satisfied, but also to insure that the simulation provided by Chan&Wood is exact (see [5] and [8] for more details on the exactness of the simulation). The complexity of the algorithm dramatically increases with the number of samples needed and with the number of dimension required. This drawback may lead to the use of simpler but non exact simulation techniques, relying for example on the spectral matrix (see *e.g.* [6]).

5.1 A more general in-between case

When examining equations (2) and (3) we observe that the two cases we have studied leads to an easy solution. A more general case also leads to an easy solution. This case generalizes the causal and well-balanced cases and is obtained when matrices A_+ and A_- are proportional. If we write $A_- = \kappa A_+$, we recover evidently the causal case when $\kappa = 0$ and the well-balanced case when $\kappa = 1$. This case can be treated in the same spirit as the two others, and if we define

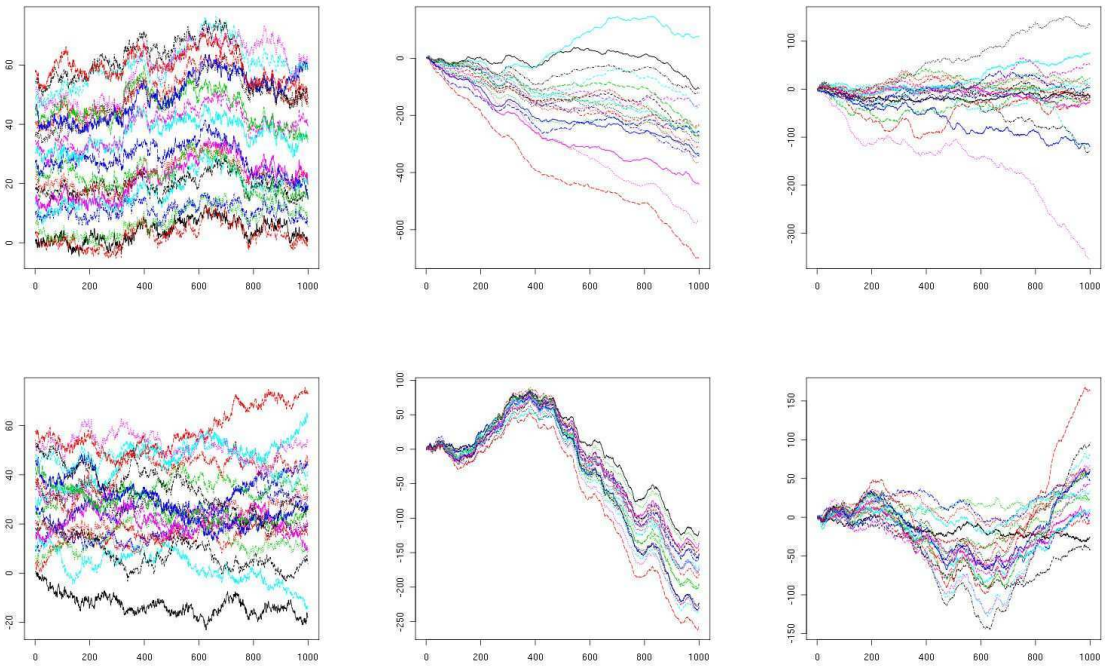


Figure 3: Examples of sample paths of length $n = 1000$ normalized causal (top) and well-balanced (bottom) multivariate fractional Brownian motion with $p = 20$ components. The Hurst parameters are equally spaced in $[0.3, 0.4]$ (left), $[0.8, 0.9]$ (middle) and $[0.4, 0.8]$ (right). The correlation parameters are all set to 0.7 (left, middle) and 0.3 (right). Note that the existence condition discussed in Proposition 3 is satisfied for these different choices of parameters. For convenience, the sample paths of the left column have been decentered.

matrix $A = (A_{jk})$ for $j, k = 1 \ldots, p$ by

$$A_{jk} = \begin{cases} \dfrac{\sigma_j \sigma_k \rho_{jk} \sin(\pi(H_j + H_k))}{B_{jk}\Big((\cos(\pi H_j) + \cos(\pi H_k))(1+\kappa^2) - 2\kappa \sin(\pi(H_j+H_k))\Big)} & \text{if } H_j + H_k \neq 1 \\ \dfrac{2\sigma_j \sigma_k \rho_{jk}}{B_{jk}\Big((\sin(\pi H_j) + \sin(\pi H_k))(1+\kappa^2) - 4\kappa\Big)} & \text{if } H_j + H_k = 1 \end{cases}$$

$$A_{jj} = \frac{\sigma_j^2 \sin(\pi H_j)}{B_{jj}(1 + \kappa^2 - 2\kappa \sin(\pi H_j))}$$

we will end up with the same parameterization of the covariance. From the expression of A_{jj} we can show that the model leads to $\sigma_{jj} = 0$ if and only if $H_j = 1/2$ and $\kappa = 1$. We thus recover the condition of existence in the well-balanced case. For this general case, there is another problem which appears for example if $H_j + H_k \neq 1$. We observe that the denominator of A_{jk} may cancel whenever

$$\kappa = \frac{\sin\left(\frac{\pi}{2}(H_j + H_k)\right) \pm \sqrt{-\cos(\pi H_j)\cos(\pi H_k)}}{\cos\left(\frac{\pi}{2}(H_j - H_k)\right)}$$

Of course this occurs if and only if the two cosine in the square root are of opposite sign, implying that H_j and H_k are not simultaneously in the same intervals $[0, 1/2), (1/2, 1]$. If there is a value $\kappa(H_j, H_k)$ that cancelled the denominator of A_{jk}, then $\rho_{jk} = 0$, and we cannot use the correlation at times 1 to normalize correctly the process. Anorther choice has to be done, such as $E[x_j(1)x_k(-1)]$.

5.2 Some future works

A first step in the future work is to adopt another more natural parameterization of the covariance function, in which function $w_{jk}(t)$ writes $\rho_{jk} + \text{Sign}(t)\eta_{jk}$, where ρ_{jk} has the same meaning of the correlation at times 1, and where η_{jk} is an antisymmetric parameter, linked to ρ_{jk}. This parameterization should lead to a more easy study of the condition of existence of the covariance function.

A next step in our work will be to tackle the problem of inferring the Hurst parameters from the observation of a sample path of either the multivariate fractional Brownian motion or the fractional Gaussian noise. The main question to answer concerns the comparison between p estimators designed for monovariate fBm (or fGn) and a multivariate estimator to be defined.

6 Appendix: Proof of Proposition 7

Proof.

Case 1. $j = k$ and $H_j \neq 1/2$ or $j \neq k$ and $H_j + H_k \neq 1$.

Let $D_h := \{(t_1, t_2) \in \mathbb{R}^2 : |a_2 t_2 - a_1 t_1| < \frac{|h|}{2}\}$, $\alpha := H_j + H_k$ and let us write $E[d^j_{a,b}\overline{d^k_{a,b+h}}] = -\frac{\sigma_j \sigma_k \rho_{jk}}{2}\sqrt{a_1 a_2} \times T$ with

$$T := \int_{\mathbb{R}^2} c_{kj}(a_2 t_2 - a_1 t_1 + h)|a_2 t_2 - a_1 t_1 + h|^\alpha \overline{\psi}(t_1)\psi(t_2)dt_1 dt_2 = T_1 + T_2,$$

and where T_1 (resp. T_2) corresponds to the integral on D_h (resp. $\mathbb{R}^2 \setminus D_h$). Let us first prove that $|h|^{2M-\alpha} T_2 \to 0$ as $|h| \to +\infty$. We have (since $2M - \alpha > 0$ and since $|c_{kj}(h)| \leq c^\vee := \max(|c_{jk}|, |c_{kj}|)$ for all h)

$$\begin{aligned} |h|^{2M-\alpha}|T_2| &\leq 2^{2M-\alpha} c^\vee \int_{\mathbb{R}^2 \setminus D_h} (a_2 t_2 - a_1 t_1)^{2M} \left|1 + \frac{h}{a_2 t_2 - a_1 t_1}\right|^\alpha |\psi(t_1)||\psi(t_2)|dt_1 dt_2 \\ &\leq 2^{2M-\alpha} 3^\alpha c^\vee \int_{\mathbb{R}^2 \setminus D_h} (a_2 t_2 - a_1 t_1)^{2M} |\psi(t_1)||\psi(t_2)|dt_1 dt_2. \end{aligned}$$

The result is then obtained by using assumption [C2(2M)] and the dominated convergence theorem. Now, within the domain D_h, one may use the series expansion of $(1+x)^\alpha$ (for $|x| < 1$).

$$\begin{aligned} T_1 &= |h|^\alpha \int_{D_h} c_{kj}(a_2 t_2 - a_1 t_1 + h)\left(1 + \frac{a_2 t_2 - a_1 t_1}{h}\right)^\alpha \overline{\psi}(t_1)\psi(t_2)dt_1 dt_2 \\ &= |h|^\alpha c_{kj}(h) \int_{D_h} \left(\sum_{\ell \geq 0} \binom{\alpha}{\ell} \left(\frac{a_2 t_2 - a_1 t_1}{h}\right)^\ell\right) \overline{\psi}(t_1)\psi(t_2)dt_1 dt_2, \end{aligned}$$

where $\binom{\alpha}{\ell}$ denotes the binomial coefficient $(\alpha)(\alpha-1)\ldots(\alpha-\ell+1)/\ell!$. Decompose T_1 into three terms (denoted by T_1', T_2' and T_3') corresponding to the $2M$ first terms of the series, the $(2M+1)$th ($\ell = 2M$) and the remainder terms. Then,

$$T_1' = |h|^\alpha c_{kj}(h) \sum_{\ell=0}^{2M-1} h^{-\ell} \binom{\alpha}{\ell} \int_{D_h} (a_2 t_2 - a_1 t_1)^\ell \overline{\psi}(t_1)\psi(t_2)dt_1 dt_2.$$

Under Assumption [C3], ψ has M vanishing moments and therefore the previous expression reduces to

$$T_1' = -|h|^\alpha c_{kj}(h) \sum_{\ell=0}^{2M-1} h^{-\ell} \binom{\alpha}{\ell} \int_{\mathbb{R}^2 \setminus D_h} (a_2 t_2 - a_1 t_1)^\ell \overline{\psi}(t_1)\psi(t_2)dt_1 dt_2.$$

Now,

$$|h|^{2M-\alpha}|T_1'| \leq c^\vee \sum_{\ell=0}^{2M-1} \left|\binom{\alpha}{\ell}\right| \int_{\mathbb{R}^2 \setminus D_h} 2^{2M-\ell} (a_2 t_2 - a_1 t_1)^{2M} |\psi(t_1)||\psi(t_2)|dt_1 dt_2.$$

Assumption [C2(2M)] and the dominated convergence theorem may be combined to prove that $|h|^{2M-\alpha}T_1' \to 0$ as $|h| \to +\infty$. The term T_2' is defined as

$$T_2' := |h|^{\alpha-2M} c_{kj}(h) \binom{\alpha}{2M} \int_{D_h} (a_2t_2 - a_1t_1)^{2M} \, \overline{\psi}(t_1)\psi(t_2)dt_1dt_2.$$

As previously, as $|h| \to +\infty$ we obtain

$$\begin{aligned} \frac{|h|^{2M-\alpha}T_2'}{c_{kj}(h)} &\to \binom{\alpha}{2M} \int_{\mathbb{R}^2} (a_2t_2 - a_1t_1)^{2M} \, \overline{\psi}(t_1)\psi(t_2)dt_1dt_2 \\ &= \binom{\alpha}{2M}\binom{2M}{M}(a_1a_2)^M \left| \int t^M \psi(t)dt \right|^2 = \binom{\alpha}{2M}\kappa(\psi, M). \end{aligned}$$

Since $T = T_1 + T_2 = T_1' + T_2' + T_3' + T_2$, the proof will be completed if we manage to prove that $|h|^{2M-\alpha}T_3' \to 0$. Let us write

$$\begin{aligned} |h|^{2M-\alpha}T_3' &= h^{2M} c_{kj}(h) \int_{D_h} \sum_{\ell \geq 2M+1} \binom{\alpha}{\ell} \left(\frac{a_2t_2 - a_1t_1}{h} \right)^{\ell} \overline{\psi}(t_1)\psi(t_2)dt_1dt_2 \\ &= \frac{c_{kj}(h)}{h} \int_{D_h} (a_2t_2 - a_1t_1)^{2M+1} \left(\sum_{\ell \geq 0} \binom{\alpha}{\ell + 2M + 1} \left(\frac{a_2t_2 - a_1t_1}{h} \right)^{\ell} \right) \overline{\psi}(t_1)\psi(t_2)dt_1dt_2. \end{aligned}$$

The binomial coefficient appearing in the last equation satisfies, with $\ell' = \ell + 2M + 1$

$$\begin{aligned} \left| \binom{\alpha}{\ell'} \right| &= \frac{|\alpha(\alpha-1)\cdots(\alpha-\ell'+1)|}{\ell'!} \\ &\leq \frac{2(2-\alpha)\cdots(\ell'-1-\alpha)}{\ell'!} \text{ since } \alpha \leq 2 \\ &\leq \frac{2(\ell'-1)!}{\ell'!} = \frac{2}{\ell'} \leq \frac{2}{\ell} \end{aligned}$$

Recall that in D_h we have $|a_2t_2 - a_1t_1|/|h| \leq 1/2$. The series in the previous integral then satisfies

$$\begin{aligned} \left| \sum_{\ell \geq 0} \binom{\alpha}{\ell + 2M + 1} \left(\frac{a_2t_2 - a_1t_1}{h} \right)^{\ell} \right| &\leq \left| \binom{\alpha}{2M+1} \right| + \sum_{\ell \geq 1} \left| \binom{\alpha}{\ell + 2M + 1} \right| \left| \frac{a_2t_2 - a_1t_1}{h} \right|^{\ell} \\ &\leq \frac{2}{2M+1} + \sum_{\ell \geq 1} \frac{2}{\ell} \, 2^{-\ell} \\ &= \frac{2}{2M+1} + 2\log(2) =: C_M \end{aligned}$$

Thus we obtain

$$|h|^{2M-\alpha}|T_3'| \;\leq\; \frac{C^M c^\vee}{|h|}\int_{\mathbb{R}^2}|a_2t_2-a_1t_1|^{2M+1}|\psi(t_1)||\psi(t_2)|dt_1dt_2$$

Since by Assumption [C2(2M+1)], $t^{2M+1}\psi(t)\in L^1$, we have $|h|^{2M-\alpha}|T_3'|=O(h^{-1})$, whence the result.

Case 2. $H_j+H_k=1$ and $H_j\neq 1/2$.

We take the same notation as previously. We first note that, under [C1], the term T can be rewritten as

$$T=\int_{\mathbb{R}^2}|a_2t_2-a_1t_1+h|+f_{jk}(a_2t_2-a_1t_1+h)\log\left|1+\frac{a_2t_2-a_1t_1}{h}\right|\overline{\psi}(t_1)\psi(t_2)dt_1dt_2.$$

We decompose T in T_1+T_2 (as done in case 1). The proof that $|h|^{2M-1}T_2\to 0$ as $|h|\to+\infty$ follows similar arguments as in the case 1 and is therefore omitted. Now, the term T_1 can be rewritten as

$$\begin{aligned}T_1 \;&=\; |h|\int_{D_h}\left(1+\frac{a_2t_2-a_1t_1}{h}\right)\overline{\psi}(t_1)\psi(t_2)dt_1dt_2\\ &\quad +f_{jk}h\int_{D_h}\left(1+\frac{a_2t_2-a_1t_1}{h}\right)\log\left(1+\frac{a_2t_2-a_1t_1}{h}\right)\overline{\psi}(t_1)\psi(t_2)dt_1dt_2.\end{aligned}$$

Denote by $\tilde{T}_1$ and $\tilde{T}_2$ these two integrals. Assumption [C1] leads to

$$\tilde{T}_1=-|h|\int_{\mathbb{R}^2\setminus D_h}\left(1+\frac{a_2t_2-a_1t_1}{h}\right)\overline{\psi}(t_1)\psi(t_2)dt_1dt_2.$$

Then, we assert that

$$\begin{aligned}|h|^{2M-1}|\tilde{T}_1| \;&\leq\; |h|^{2M-1}\int_{\mathbb{R}^2\setminus D_h}\left|a_2t_2-a_1t_1\right|\left|1+\frac{a_2t_2-a_1t_1}{h}\right||\psi(t_1)||\psi(t_2)|dt_1dt_2\\ &\leq\; 2^{2M-1}3\int_{\mathbb{R}^2\setminus D_h}\Big(a_2t_2-a_1t_1\Big)^{2M}|\psi(t_1)||\psi(t_2)|dt_1dt_2\to 0\end{aligned}$$

as $|h|\to+\infty$. For the term $\tilde{T}_2$, we may use the series expansion of $\log(1+x)$ (for $|x|<1$). We omit the details and leave the reader to prove that as $|h|\to+\infty$

$$\begin{aligned}\tilde{T}_2 \;&\sim\; f_{jk}h\int_{\mathbb{R}^2}\left(1+\frac{a_2t_2-a_1t_1}{h}\right)\left(\frac{(-1)^{2M}}{2M-1}\left(\frac{a_2t_2-a_1t_1}{h}\right)^{2M-1}\right.\\ &\qquad\left.\frac{(-1)^{2M+1}}{2M}\left(\frac{a_2t_2-a_1t_1}{h}\right)^{2M}\right)\overline{\psi}(t_1)\psi(t_2)dt_1dt_2\\ &\sim\; h^{1-2M}\times\frac{f_{jk}}{2M(2M-1)}\binom{2M}{M}(a_1a_2)^M\left|\int t^M\psi(t)dt\right|^2.\end{aligned}$$

Hence, $T \sim |h|^{1-2M} \times \frac{f_{jk} \times \text{Sign}(h)}{2M(2M-1)} \kappa(\psi, M)$.

In this proof, Fubini's theorem and interchanges of integrals and (in)finite sums are widely used. All of these are justified by the absolute convergence of the different series related to the expansions of $(1+x)^{-\alpha}$ or $\log(1+x)$ for $|x| < 1$ and Assumption [C2(2M+1)]. ■

7 Proof of Proposition 8

Before proving the proposition itself, we need to generalize a formula due to Von Bahr and Essen [4]. In 1965, they have obtained the following representation theorem for $|v|^\alpha$ for $\alpha \in (0,2)$:

$$|v|^\alpha = \frac{\Gamma(\alpha+1)\sin(\pi\alpha/2)}{\pi} \int_{\mathbb{R}} \frac{1-\cos(\omega v)}{|\omega|^{\alpha+1}} d\omega. \tag{38}$$

The following lemma provides a similar representation for v_+^α and v_-^α.

Lemma 9 *For any $\alpha \in (0,2)$, $\alpha \neq 1$, we have*

$$\begin{aligned} v_+^\alpha &= \frac{\Gamma(\alpha+1)}{2\pi} \int_{\mathbb{R}} \frac{\sin\left(\pi\frac{\alpha}{2}\right)(1-\cos(\omega v)) + \cos\left(\pi\frac{\alpha}{2}\right) Sign(\omega)(\sin(\omega v) - g(\omega v))}{|\omega|^{\alpha+1}} d\omega \\ &= \frac{\Gamma(\alpha+1)}{2\pi} \int_{\mathbb{R}} \frac{Re\left(e^{iSign(\omega)\frac{\pi}{2}(\alpha+1)}\left(e^{-i\omega v} - 1 + ig(\omega v)\right)\right)}{|\omega|^{\alpha+1}} d\omega \end{aligned}$$

and

$$\begin{aligned} v_-^\alpha &= \frac{\Gamma(\alpha+1)}{2\pi} \int_{\mathbb{R}} \frac{\sin\left(\pi\frac{\alpha}{2}\right)(1-\cos(\omega v)) - \cos\left(\pi\frac{\alpha}{2}\right) Sign(\omega)(\sin(\omega v) - g(\omega v))}{|\omega|^{\alpha+1}} d\omega \\ &= \frac{\Gamma(\alpha+1)}{2\pi} \int_{\mathbb{R}} \frac{Re\left(e^{-iSign(\omega)\frac{\pi}{2}(\alpha+1)}\left(e^{-i\omega v} - 1 + ig(\omega v)\right)\right)}{|\omega|^{\alpha+1}} d\omega \end{aligned}$$

where the function g equals zero when $\alpha \in (0,1)$ and is the identity function when $\alpha \in (1,2)$.

The second expressions for v_+^α and v_-^α are derived from the first one with elementary algebra.
Proof. Let us first observe that

$$v_+^\alpha = \frac{1}{2}\left(|v|^\alpha + \text{Sign}(v)|v|^\alpha\right) \text{ and } v_-^\alpha = \frac{1}{2}\left(|v|^\alpha - \text{Sign}(v)|v|^\alpha\right).$$

Therefore, the proof consists in giving a representation of $\text{Sign}(v)|v|^\alpha$. Let $\alpha \in (0,1)$, then from (38) and properties of the function Γ

$$\frac{1}{\alpha+1}|v|^{\alpha+1} = \frac{\Gamma(\alpha+1)}{\pi}\cos(\pi\alpha/2) \int_{\mathbb{R}} \frac{1-\cos(\omega v)}{|\omega|^{\alpha+2}} d\omega.$$

Since $\int_{\mathbb{R}} |\omega|^{-\alpha-1}|\sin(\omega v)| < +\infty$ for $\alpha \in (0,1)$, we can differentiate this integral with respect to v to obtain

$$\text{Sign}(v)|v|^{\alpha} = \frac{\Gamma(\alpha+1)}{\pi}\cos(\pi\alpha/2)\int_{\mathbb{R}} \frac{\text{Sign}(\omega)\sin(\omega v)}{|\omega|^{\alpha+1}}d\omega. \tag{39}$$

When $\alpha \in (1,2)$, then from (38) and properties of the function Γ

$$\alpha|v|^{\alpha-1} = \frac{\Gamma(\alpha+1)}{\pi}(-\cos(\pi\alpha/2))\int_{\mathbb{R}} \frac{1-\cos(\omega v)}{|\omega|^{\alpha}}d\omega.$$

Since $\int_{\mathbb{R}} |\omega|^{-\alpha-1}|\sin(\omega v) - \omega v|d\omega < +\infty$ for $\alpha \in (1,2)$, we can take the primitive of the last equation to get

$$\begin{aligned}\text{Sign}(v)|v|^{\alpha} &= \frac{\Gamma(\alpha+1)}{\pi}\cos(\pi\alpha/2)\int_{\mathbb{R}} \frac{\sin(\omega v)/\omega - v}{|\omega|^{\alpha}}d\omega \\ &= \frac{\Gamma(\alpha+1)}{\pi}\cos(\pi\alpha/2)\int_{\mathbb{R}} \frac{\text{Sign}(\omega)(\sin(\omega v) - \omega v)}{|\omega|^{\alpha+1}}d\omega,\end{aligned}$$

which ends the proof. ■

Let $\alpha \in (0,1)$, then by differentiating (39) with respect to α and taking the limit as $\alpha \to 1^-$, we may obtain

$$\text{Sign}(h)|h|\log|h| = h\log|h| = \lim_{\alpha\to 1^-} -\frac{1}{2}\int_{\mathbb{R}} \frac{\text{Sign}(\omega)\sin(\omega v)}{|\omega|^{\alpha+1}}d\omega. \tag{40}$$

We now turn to the proof of proposition 8.

Proof. (i) We recall that under [C1] and [C2(2)], Equation (31) holds, that is $E[d^j_{a_1,b_1}\overline{d^k_{a_2,b_2}}] = -\frac{\sigma_j\sigma_k\rho_{jk}}{2}T$ with $T := \int_{\mathbb{R}} w_{jk}(v)\overline{\Gamma_\psi}(v)dv$. Furthermore, note that the Fourier transforms of $\psi_{a,b}$ and $\Gamma_\psi(v)$ exist and are equal respectively to $\sqrt{a}\widehat{\psi}(a\omega)e^{-i\omega b}$ and to $\sqrt{a_1a_2}\widehat{\psi}(-a_1\omega)\overline{\widehat{\psi}(-a_2\omega)}e^{-i\omega(b_2-b_1)}$ which leads to

$$q(\omega) := \int_{\mathbb{R}} \overline{\Gamma_\psi}(v)e^{-i\omega v}dv = \sqrt{a_1a_2}\overline{\widehat{\psi}(a_1\omega)}\widehat{\psi}(a_2\omega)e^{-i\omega(b_2-b_1)}. \tag{41}$$

Now, let us split the proof into two cases.

Case 1. $j = k$ or $j \neq k$ and $H_j + H_k \neq 1$.

When $j = k$, at this step, the authors of [15] have used the representation of $|v|^{\alpha}$ obtained by [4]. We obtained a similar representation for the function v_+^{α} and v_-^{α} for $\alpha \in (0,2)$ and $\alpha \neq 1$ in lemma 9. For any α, let us set $S(\omega) := \text{Sign}(\omega)\frac{\pi}{2}(\alpha+1)$. We have by Fubini's theorem and under Assumption [C2(2M)] (with $M \geq 2$).

$$T = \frac{\Gamma(\alpha+1)}{2\pi}\int_{\mathbb{R}} |\omega|^{-\alpha-1}\int_{\mathbb{R}} \left(\frac{c_{kj}}{2}\left(e^{iS(\omega)}e^{-i\omega v} + e^{-iS(\omega)}e^{i\omega v}\right)\overline{\Gamma_\psi}(v)\right.$$

$$
\begin{aligned}
&\qquad + \frac{c_{jk}}{2}\left(e^{-iS(\omega)}e^{-i\omega v} + e^{iS(\omega)}e^{i\omega v}\right)\overline{\Gamma_\psi}(v)\Big)\,dvd\omega \\
&= \frac{\Gamma(\alpha+1)}{2\pi}\int_{\mathbb{R}} |\omega|^{-\alpha-1}\Big(\frac{c_{kj}}{2}\left(e^{iS(\omega)}q(\omega) + e^{-iS(\omega)}q(-\omega)\right) \\
&\qquad + \frac{c_{jk}}{2}\left(e^{-iS(\omega)}q(\omega) + e^{iS(\omega)}q(-\omega)\right)\Big)\,d\omega \\
&= \frac{\Gamma(\alpha+1)}{2\pi}\int_{\mathbb{R}} |\omega|^{-\alpha-1}\left(c_{kj}e^{iS(\omega)} + c_{jk}e^{-iS(\omega)}\right)q(\omega)d\omega. \qquad (42)
\end{aligned}
$$

Note that the condition $M \geq 2$ is required for $\alpha > 1$. For $\alpha < 1$, $M \geq 1$ is a sufficient condition. These conditions allow us to show that the contributions $\int (1 - ig(\omega v))\exp(\pm iS(\omega))\overline{\Gamma_\psi}(v)dv$ in this calculation are equal to zero.

Making the change of variable $\omega \leftrightarrow -\omega$ in the integral in (42) and using (41), we obtain

$$
T = \Gamma(\alpha+1)\sqrt{a_1a_2}\frac{1}{2\pi}\int_{\mathbb{R}} |\omega|^{-\alpha-1}\left(c_{kj}e^{-iS(\omega)} + c_{jk}e^{iS(\omega)}\right)\overline{\widehat{\psi}(-a_1\omega)}\widehat{\psi}(-a_2\omega)e^{i\omega(b_2-b_1)}d\omega.
$$

and therefore, reminding that $\zeta_{jk}(\omega) := -c_{kj}e^{-iS(\omega)} - c_{jk}e^{iS(\omega)}$, we have

$$
E[d^j_{a_1,b_1}\overline{d^k_{a_2,b_2}}] = \frac{\sigma_j\sigma_k\rho_{jk}}{2}\sqrt{a_1a_2}\frac{1}{2\pi}\int_{\mathbb{R}} \underbrace{|\omega|^{-\alpha-1}\zeta_{jk}(\omega)\overline{\widehat{\psi}(-a_1\omega)}\widehat{\psi}(-a_2\omega)}_{=:P(\omega)} e^{i\omega(b_2-b_1)}d\omega.
$$

By using Bochner's Theorem, the proof will be done, if one proves that the function $P(\omega)$ is integrable. Let us prove this last assertion. Under [C2(M)], $t^k\psi(t) \in L^1$ for $k = 0, \ldots, M$. Therefore, $\widehat{\psi}$ is a M times continuous and differentiable function. Using a Taylor expansion

$$
\widehat{\psi}(\omega) = \sum_{k=0}^{M-1}\omega^k\widehat{\psi}^{(k)}(\omega) + \omega^M\widehat{\psi}^{(M)}(\widetilde{\omega}) = \omega^M\widehat{\psi}^{(M)}(\widetilde{\omega}), \text{ with } \widetilde{\omega} \in [0\wedge\omega, 0\vee\omega],
$$

under [C2(M)]. And since $\psi^{(M)}$ is continuous at zero, $\widehat{\psi}(\omega) \sim \omega^M\widehat{\psi}^{(M)}(0)$ as $\omega \to 0$. Then as $\omega \to 0$:

$$
P(\omega) \sim \zeta_{jk}(\omega)|\omega|^{2M-1-\alpha}(a_1a_2)^M|\widehat{\psi}^{(M)}(0)|^2. \qquad (43)
$$

As a consequence, for $M \geq 2$, P is continuous at zero and $\lim_{\omega\to 0^\pm} P(\omega) = 0$. Therefore for $\varepsilon > 0$, P is integrable on the interval $[-\varepsilon, \varepsilon]$ as a continuous function on this interval. Finally (recall that $c^\vee := \max(|c_{jk}|, |c_{kj}|)$),

$$
\begin{aligned}
\int_{|\omega|\geq\varepsilon}|P(\omega)| &\leq 2c^\vee(a_1a_2)^{\alpha+1}\Big(\int_{|\omega|\geq a_1\varepsilon}\frac{|\widehat{\psi}(\omega)|^2}{|\omega|^{\alpha+1}}d\omega\Big)^{1/2}\Big(\int_{|\omega|\geq a_2\varepsilon}\frac{|\widehat{\psi}(\omega)|^2}{|\omega|^{\alpha+1}}d\omega\Big)^{1/2} \\
&\leq \frac{(a_1a_2)}{\varepsilon^{2\alpha}}\int_{\mathbb{R}}\frac{|\widehat{\psi}(\omega)|^2}{|\omega|}d\omega < +\infty
\end{aligned}
$$

under [C1]. Hence, $P(\omega) \in L^1$ and Bochner's Theorem may be applied.
Case 2. $j \neq k$ and $H_j + H_k = 1$.

We start with the representation of $v \log |v|$ given by (40).

$$\begin{aligned} w_{jk}(v) &= |v| + f_{jk} v \log |v| = \lim_{\alpha \to 1^-} |v|^\alpha + f_{jk} v \log |v| \\ &= \lim_{\alpha \to 1^-} \frac{1}{2\pi} \int_{\mathbb{R}} \frac{2(1-\cos(\omega v)) - \pi f_{jk} \mathrm{Sign}(\omega) \sin(\omega v)}{|\omega|^{\alpha+1}} d\omega. \end{aligned}$$

Now, we derive the computation of the term $T := \int_{\mathbb{R}} w_{jk}(v) \overline{\Gamma}_\psi(-v) dv$, similarly as the previous case. Using dominated convergence theorem and Fubini's Theorem,

$$\begin{aligned} T &= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\lim_{\alpha \to 1^-} \int_{\mathbb{R}} \frac{2(1-\cos(\omega v)) - \pi f_{jk} \mathrm{Sign}(\omega) \sin(\omega v)}{|\omega|^{\alpha+1}} d\omega \right) \overline{\Gamma}_\psi(v) dv \\ &= \frac{1}{2\pi} \lim_{\alpha \to 1^-} \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{2(1-\cos(\omega v)) - \pi f_{jk} \mathrm{Sign}(\omega) \sin(\omega v)}{|\omega|^{\alpha+1}} \overline{\Gamma}_\psi(v) dv \right) d\omega \\ &= \frac{1}{2\pi} \lim_{\alpha \to 1^-} \int_{\mathbb{R}} \frac{2 - i\pi f_{kj} \mathrm{Sign}(\omega)}{|\omega|^{\alpha+1}} q(\omega) d\omega \\ &= \frac{1}{2\pi} \lim_{\alpha \to 1^-} \int_{\mathbb{R}} |\omega|^{-\alpha-1} \left(2 - i\pi f_{kj} \mathrm{Sign}(\omega)\right) \overline{\widehat{\psi}(a_1\omega)} \widehat{\psi}(a_2\omega) e^{-i\omega(b_2-b_1)} d\omega. \end{aligned}$$

From (43), $|\omega|^{-\alpha-1} \overline{\widehat{\psi}(a_1\omega)} \widehat{\psi}(a_2\omega)$ is an integrable function for all $\alpha \in (0,2)$. Therefore, the integral and the limit may be interchanged. Making the change of variable $\omega \leftrightarrow -\omega$ we obtain

$$T = \sqrt{a_1 a_2} \frac{1}{2\pi} \int_{\mathbb{R}} \frac{2 - i\pi f_{jk} \mathrm{Sign}(\omega)}{\omega^2} \overline{\widehat{\psi}(-a_1\omega)} \widehat{\psi}(-a_2\omega) e^{i\omega(b_2-b_1)} d\omega.$$

and Bochner's Theorem can be applied.
(ii) is derived from (43). ■

References

[1] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi and D. Veitch. Multiscale nature of network traffic. *IEEE Signal. Proc. Mag.*, 19(3),pp 28–46, 2008.

[2] S. Achard, D. S. Bassett, A. Meyer-Lindenberg, and E. Bullmore. Fractal connectivity of long-memory networks. *Phys. Rev. E*, 77:036104, 2008.

[3] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *The Journal of Neuroscience*, 26(1):63–72, 2006.

[4] B. von Bahr and C.G. Esseen. Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *The Annals of Mathematical Statistics*, 36(1):299-303, 1965.

[5] G. Chan and A. Wood. Simulation of stationary gaussian vector fields. *Statistics and Computing*, 9(4):265–268, 1999.

[6] J. Chambers. The simulation of random vector time series with given spectrum. *Mathematical and Computer Modelling*, 22:1–6, 1995.

[7] J. F. Coeurjolly. Estimating the parameters of a fractional brownian motion by discrete variations of its sample paths. *Statistical Inference for stochastic processes*, 4(2):199–227, 2001.

[8] J. F. Coeurjolly, P. O. Amblard and S. Achard. On multivariate fractional Brownian motion and multivariate fractional Gaussian noise. *in Proceedings EUSIPCO*, Aalborg, Denmark, Aug. 2010.

[9] P. Craigmile. Simulating a class of stationary gaussian processes using the davies-harte algortihm, with application to long memory processes. *Journal of Time Series Analysis*, 24:505–510, Jan 2003.

[10] G. Didier and V. Pipiras. Integral representations of operator fractional brownian motion. *To appear in Bernouilli*, 2010.

[11] C. R. Dietrich and G. N. Newsam. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific and Statistical Computing*, 18:10881107., 1997.

[12] P. Flandrin On the spectrum of fractional Brownian motions. *IEEE Transactions on Information Theory*, 35(1):197–199, 1989.

[13] P. Flandrin Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, 38(2):910–917, 1992.

[14] Gel'fand, I. M. and Shilov, G. E. Generalized functions. Vol 1: Properties and Operations. Academic Press, 1964

[15] T. Kato and E. Masry On the spectral density of the wavelet transform of fractional Brownian motion. *Journal of Time Series Analysis*, 20(5):560-563, 1998.

[16] F. Lavancier, A. Philippe, and D. Surgailis. Covariance function of vector self-similar processes. *Statistics and Probability Letters*, 79 ,2415?2421, 2009

[17] B. Mandelbrot and J. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10(4):422–437, 1968.

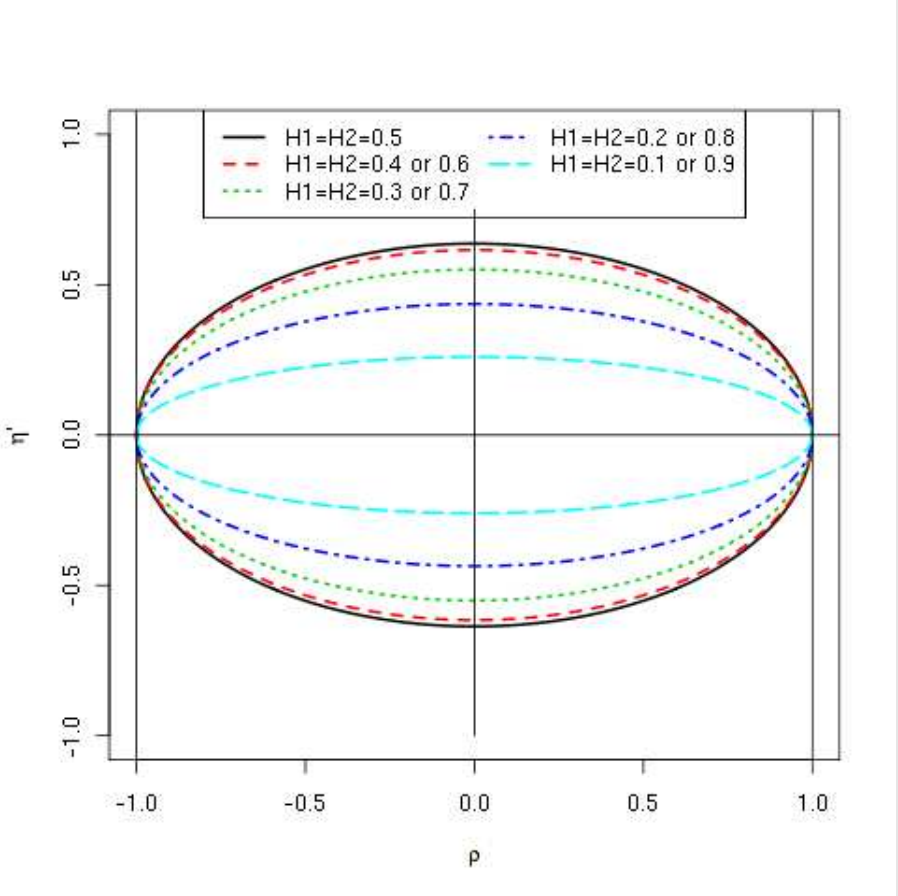
[18] S. Stoev and M. Taqqu. How rich is the class of multifractional brownian motions? *Stochastic Processes and their Applications*, 116:200–221, 2006.

[19] A. H.Tewfik and M. Kim, Correlation Structure of the Discrete Wavelet Coefficients of Fractional Brownian Motion. *IEEE Transactions on Information Theory*, 38(2):904–910., 1992.

[20] H. Wendt, A. Scherrer, P. Abry, and S. Achard. Testing fractal connectivity in multivariate long memory processes. *Proceedings of ICASSP*, page cdrom, 2009.

[21] A. Wood and G. Chan. Simulation of stationary gaussian processes in $[0, 1)]^d$. *Journal of computational and graphical statistics*, 3:409–432, 1994.

[22] G. W. Wornell, A Karhunen-Loeve-like Expansion for $1/f$ Processes via Wavelets *IEEE Trans. on Information Theory*, 36(4): 861–863,1990.



[ACLP10]
P.O. Amblard and J.-F. Coeurjolly and F. Lavancier and A. Philippe. Basic properties of the Multivariate Fractional Brownian Motion. *to appear in Bulletin de la Société Mathématique de France, Séminaires et Congrés* (available at `http://arxiv.org/abs/1007.0828`), 2010.

Basic properties of the Multivariate Fractional Brownian Motion

Pierre-Olivier Amblard[1], Jean-François Coeurjolly[1],
Frédéric Lavancier [2] and Anne Philippe[2]

[1] GIPSAlab/CNRS UMR 5216/ BP 46, 38402 Saint Martin d'Hères cedex, France
[2] Université de Nantes / Laboratoire de mathématiques Jean Leray/ 2 rue de la Houssinière 44322 Nantes, France

Abstract

This paper reviews and extends some recent results on the multivariate fractional Brownian motion (mfBm) and its increment process. A characterization of the mfBm through its covariance function is obtained. Similarly, the correlation and spectral analyses of the increments are investigated. On the other hand we show that (almost) all mfBm's may be reached as the limit of partial sums of (super)linear processes. Finally, an algorithm to perfectly simulate the mfBm is presented and illustrated by some simulations.
keywords : Self similarity ; Multivariate process ; Long-range dependence ; Superlinear process ; Increment process ; Limit theorem.

1 Introduction

The fractional Brownian motion is the unique Gaussian self-similar process with stationary increments. In the seminal paper of Mandelbrot and Van Ness[22], many properties of the fBm and its increments are developed (see also [30] for a review of the basic properties). Depending on the scaling factor (called Hurst parameter), the increment process may exhibit long-range dependence, and are commonly used in modeling physical phenomena. However in many fields of applications (e.g. neuroscience, economy, sociology, physics, etc), multivariate measurements are performed and they involve specific properties such as fractality, long-range dependence, self-similarity, etc. Examples can be found in economic time series (see [11], [14], [15]), genetic sequences [2], multipoint velocity measurements in turbulence, functional Magnetic Resonance Imaging of several regions of the brain [1].

It seems therefore natural to extend the fBm to a multivariate framework. Recently, this question has been investigated in [20, 19, 5]. The aim of this paper is to summarize and to complete some of these advances on the multivariate fractional Brownian motion and its increments. A multivariate extension of the fractional Brownian motion can be stated as follows :

Definition 1. *A Multivariate fractional Brownian motion (p-mfBm or mfBm) with parameter $H \in (0,1)^p$ is a p-multivariate process starting from $0 \in \mathbb{R}^p$ and satisfying the three following properties*

- *Gaussianity,*
- *Self-similarity with parameter $H \in (0,1)^p$,*
- *Stationarity of the increments.*

Here, self-similarity has to be understood as joint self-similarity. More formally, we use the following definition.

Definition 2. *A multivariate process* $(X(t) = (X_1(t), \cdots, X_p(t)))_{t\in\mathbb{R}}$ *is said self-similar if there exists a vector* $H = (H_1, \cdots, H_p) \in (0,1)^p$ *such that for any* $\lambda > 0$,

$$(X_1(\lambda t), \cdots, X_p(\lambda t)))_{t\in\mathbb{R}} \overset{\text{fidi}}{=} \left(\lambda^{H_1} X_1(t), \cdots, \lambda^{H_p} X_p(t)\right)_{t\in\mathbb{R}}, \tag{1}$$

where $\overset{\text{fidi}}{=}$ *denotes the equality of finite-dimensional distributions. The parameter* H *is called the self-similarity parameter.*

This definition can be viewed as a particular case of operator self-similar processes by taking diagonal operators (see [12, 16, 17, 21]).

Note that, as in the univariate case [18], the Lamperti transformation induces an isometry between the self-similar and the stationary multivariate processes. Indeed, from Definition 2, it is not difficult to check that $(Y(t))_{t\in\mathbb{R}}$ is a p-multivariate stationary process if and only if there exists $H \in (0,1)^p$ such that its Lamperti transformation $(t^{H_1} Y_1(\log(t)), \ldots, t^{H_p} Y_p(\log(t)))_{t\in\mathbb{R}}$ is a p-multivariate self-similar process.

The paper is organized as follows. In Section 2, we study the covariance structure of the mfBm and its increments. The cross-covariance and the cross-spectral density of the increments lead to interesting long-memory type properties. Section 3 contains the time domain as well as the spectral domain stochastic integral representations of the mfBm. Thanks to these results we obtain a characterization of the mfBm through its covariance matrix function. Section 4 is devoted to limit theorems, the mfBm is obtained as the limit of partial sums of linear processes. Finally, we discuss in Section 5 the problem of simulating sample paths of the mfBm. We propose a Wood and Chan algorithm [32] well adapted to generate multivariate stationary Gaussian random fields with prescribed covariance matrix function.

2 Dependence structure of the mfBm and of its increments

2.1 Covariance function of the mfBm

In this part, we present the form of the covariance matrix of the mfBm.

Firstly, as each component is a fractional brownian motion, the covariance function of the i-th component is well-known and we have

$$\mathbb{E}X_i(s)X_i(t) = \frac{\sigma_i^2}{2}\left\{|s|^{2H_i} + |t|^{2H_i} - |t-s|^{2H_i}\right\}. \tag{2}$$

with $\sigma_i^2 := \mathrm{var}(X_i(1))$. The cross covariances are given in the following proposition.

Proposition 3 (Lavancier *et al.* [20])**.** *The cross covariances of the mfBm satisfy the following representation, for all* $(i,j) \in \{1,\ldots,p\}^2$, $i \neq j$,

1. *If* $H_i + H_j \neq 1$, *there exists* $(\rho_{i,j}, \eta_{i,j}) \in [-1,1]\times\mathbb{R}$ *with* $\rho_{i,j} = \rho_{j,i} = \mathrm{corr}(X_i(1), X_j(1))$ *and* $\eta_{i,j} = -\eta_{j,i}$ *such that*

$$\begin{aligned}\mathbb{E}X_i(s)X_j(t) = \frac{\sigma_i\sigma_j}{2}\Big\{&(\rho_{i,j} + \eta_{i,j}\mathrm{sign}(s))|s|^{H_i+H_j} + (\rho_{i,j} - \eta_{i,j}\mathrm{sign}(t))|t|^{H_i+H_j} \\ &-(\rho_{i,j} - \eta_{i,j}\mathrm{sign}(t-s))|t-s|^{H_i+H_j}\Big\}. \end{aligned}\tag{3}$$

2. *If* $H_i + H_j = 1$, *there exists* $(\tilde\rho_{i,j}, \tilde\eta_{i,j}) \in [-1,1]\times\mathbb{R}$ *with* $\tilde\rho_{i,j} = \tilde\rho_{j,i} = \mathrm{corr}(X_i(1), X_j(1))$ *and* $\tilde\eta_{i,j} = -\tilde\eta_{j,i}$ *such that*

$$\mathbb{E}X_i(s)X_j(t) = \frac{\sigma_i\sigma_j}{2}\left\{\tilde\rho_{i,j}(|s| + |t| - |s-t|) + \tilde\eta_{i,j}(t\log|t| - s\log|s| - (t-s)\log|t-s|)\right\}. \tag{4}$$

Proof. Under some conditions of regularity, Lavancier `et al.` [20] actually prove that Proposition 3 is true for any L^2 self-similar multivariate process with stationary increments. The form of cross covariances is obtained as the unique solution of a functional equation. Formulae (3) and (4) correspond to expressions given in [20] after the following reparameterization : $\rho_{i,j} = (c_{i,j} + c_{j,i})/2$ and $\eta_{i,j} = (c_{i,j} - c_{j,i})/2$ where $c_{i,j}$ and $c_{j,i}$ arise in [20].

□

Remark 1. *Extending the definition of parameters $\rho_{i,j}, \tilde{\rho}_{i,j}, \eta_{i,j}, \tilde{\eta}_{i,j}$ to the case $i = j$, we have $\rho_{i,i} = \tilde{\rho}_{i,i} = 1$ and $\eta_{i,j} = \tilde{\eta}_{i,j} = 0$, so that (2) coincides with (3) and (4).*

Remark 2. *The constraints on coefficients $\rho_{i,j}, \tilde{\rho}_{i,j}, \eta_{i,j}, \tilde{\eta}_{i,j}$ are necessary but not sufficient conditions to ensure that the functions defined by (4) and (3) are covariance functions. This problem will be discussed in Section 3.4.*

Remark 3. *Note that coefficients $\rho_{i,j}, \tilde{\rho}_{i,j}, \eta_{i,j}, \tilde{\eta}_{i,j}$ depend on the parameters (H_i, H_j). Assuming the continuity of the cross covariances function with respect to the parameters (H_i, H_j), the expression (4) can be deduced from (3) by taking the limit as $H_i + H_j$ tends to 1, noting that $((s+1)^H - s^H - 1)/(1-H) \to s\log|s| - (s+1)\log|s+1|$ as $H \to 1$. We obtain the following relations between the coefficients : as $H_i + H_j \to 1$*

$$\rho_{i,j} \sim \tilde{\rho}_{i,j} \qquad \text{and} \qquad (1 - H_i - H_j)\eta_{i,j} \sim \tilde{\eta}_{i,j}.$$

This convergence result can suggest a reparametrisation of coefficients $\eta_{i,j}$ in $(1 - H_i - H_j)\eta_{i,j}$.

2.2 The increments process

This part aims at exploring the covariance structure of the increments of size δ of a multivariate fractional Brownian motion given by Definition 1. Let $\Delta_\delta X = (X(t+\delta) - X(t))_{t\in\mathbb{R}}$ denotes the increment process of the multivariate fractional Brownian motion of size δ and let $\Delta_\delta X_i$ be its i-th component.

Let $\gamma_{i,j}(h,\delta) = \mathbb{E}\Delta_\delta X_i(t)\Delta_\delta X_j(t+h)$ denotes the cross-covariance of the increments of size δ of the components i and j. Let us introduce the function $w_{i,j}(h)$ given by

$$w_{i,j}(h) = \begin{cases} (\rho_{i,j} - \eta_{i,j}\text{sign}(h))|h|^{H_i+H_j} & \text{if } H_i + H_j \neq 1, \\ \tilde{\rho}_{i,j}|h| + \tilde{\eta}_{i,j} h \log|h| & \text{if } H_i + H_j = 1. \end{cases} \tag{5}$$

Then from Proposition 3, we deduce that $\gamma_{i,j}(h,\delta)$ is given by

$$\gamma_{i,j}(h,\delta) = \frac{\sigma_i\sigma_j}{2}\Big(w_{i,j}(h-\delta) - 2w_{i,j}(h) + w_{i,j}(h+\delta)\Big). \tag{6}$$

Now, let us present the asymptotic behaviour of the cross-covariance function.

Proposition 4. *As $|h| \to +\infty$, we have for any $\delta > 0$*

$$\gamma_{i,j}(h,\delta) \sim \sigma_i\sigma_j\delta^2|h|^{H_i+H_j-2}\kappa_{i,j}(\text{sign}(h)), \tag{7}$$

with

$$\kappa_{i,j}(\text{sign}(h)) = \begin{cases} (\rho_{i,j} - \eta_{i,j}\text{sign}(h))(H_i+H_j)(H_i+H_j-1) & \text{if } H_i + H_j \neq 1, \\ \tilde{\eta}_{i,j}\text{sign}(h) & \text{if } H_i + H_j = 1. \end{cases} \tag{8}$$

Proof. Let $\alpha = H_i + H_j$. Let us choose h, such that $|h| \geq \delta$, which ensures that $\text{sign}(h-\delta) = \text{sign}(h) = \text{sign}(h+\delta)$. When $\alpha \neq 1$, this allows us to write

$$\gamma_{i,j}(h,\delta) = \frac{\sigma_i\sigma_j}{2}|h|^\alpha\,(\rho_{i,j} - \eta_{i,j}\text{sign}(h))\,B(h),$$

with $B(h) = \left(1 - \frac{\delta}{h}\right)^{\alpha} - 2 + \left(1 + \frac{\delta}{h}\right)^{\alpha} \sim \alpha(\alpha-1)\delta^2 h^{-2}$, as $|h| \to +\infty$. When $\alpha = 1$ and $|h| \geq \delta$, $\gamma_{i,j}(h,\delta)$ reduces to

$$\gamma_{i,j}(h,\delta) = \frac{\sigma_i \sigma_j}{2} \tilde{\eta}_{i,j} B(h) \text{ with } \quad B(h) = \left((h-\delta) \log\left(1 - \frac{\delta}{h}\right) + (h+\delta) \log\left(1 + \frac{\delta}{h}\right) \right).$$

Using the expansion of $\log(1 \pm x)$ as $x \to 0$ leads to $B(h) \sim \delta^2 |h|^{-1}$ as $|h| \to +\infty$, which implies the result. □

Proposition 4 and (6) lead to the following important remarks on the dependence structure. For $i \neq j$ and $H_i + H_j \neq 1$:

- If the two fractional Gaussian noises are short-range dependent (i.e. $H_i < 1/2$ and $H_j < 1/2$) then they are either short-range interdependent if $\rho_{i,j} \neq 0$ or $\eta_{i,j} \neq 0$, or independent if $\rho_{i,j} = \eta_{i,j} = 0$.
- If the two fractional Gaussian noises are long-range dependent (i.e. $H_i > 1/2$ and $H_j > 1/2$) then they are either long-range interdependent if $\rho_{i,j} \neq 0$ or $\eta_{i,j} \neq 0$, or independent if $\rho_{i,j} = \eta_{i,j} = 0$. This confirms the dichotomy principle observed in [12].
- In the other cases, the two fractional Gaussian noises can be short-range interdependent if $\rho_{i,j} \neq 0$ or $\eta_{i,j} \neq 0$ and $H_i + H_j < 1$, long-range interdependent if $\rho_{i,j} \neq 0$ or $\eta_{i,j} \neq 0$ and $H_i + H_j > 1$ or independent if $\rho_{i,j} = \eta_{i,j} = 0$.

Moreover, note that when $H_i + H_j = 1$, whatever the nature of the two fractional Gaussian noises (i.e. short-range or long-range dependent, or even independent), they are either long-range interdependent if $\tilde{\eta}_{i,j} \neq 0$ or independent if $\tilde{\eta}_{i,j} = 0$.

The following result characterizes the spectral nature of the increments of a mfBm.

Proposition 5 (Coeurjolly *et al.* [5])**.** *Let $S_{i,j}(\cdot,\delta)$ be the (cross)-spectral density of the increments of size δ of the components i and j, i.e. the Fourier transform of $\gamma_{i,j}(\cdot,\delta)$*

$$S_{i,j}(\omega,\delta) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\mathbf{i}h\omega} \gamma_{i,j}(h,\delta)\, \mathrm{d}h =: FT(\gamma_{i,j}(\cdot,\delta)).$$

(i) For all i, j and for all H_i, H_j, we have

$$S_{i,j}(\omega,\delta) = \frac{\sigma_i \sigma_j}{\pi} \Gamma(H_i + H_j + 1) \frac{1 - \cos(\omega\delta)}{|\omega|^{H_i+H_j+1}} \times \tau_{i,j}(\mathrm{sign}(\omega)), \tag{9}$$

where

$$\tau_{i,j}(\mathrm{sign}(\omega)) = \begin{cases} \rho_{i,j} \sin\left(\frac{\pi}{2}(H_i+H_j)\right) - \mathbf{i}\eta_{i,j}\mathrm{sign}(\omega) \cos\left(\frac{\pi}{2}(H_i+H_j)\right) & \text{if } H_i + H_j \neq 1, \\ \tilde{\rho}_{i,j} - \mathbf{i}\frac{\pi}{2}\tilde{\eta}_{i,j}\mathrm{sign}(\omega) & \text{if } H_i + H_j = 1. \end{cases} \tag{10}$$

(ii) For any fixed δ, when $H_i + H_j \neq 1$ then we have, as $\omega \to 0$,

$$\left|S_{i,j}(\omega,\delta)\right| \sim \frac{\sigma_i \sigma_j}{2\pi} \Gamma(H_i + H_j + 1)\delta^2 \frac{\left(\rho_{i,j}^2 \sin\left(\frac{\pi}{2}(H_i+H_j)\right)^2 + \eta_{i,j}^2 \cos\left(\frac{\pi}{2}(H_i+H_j)\right)^2\right)^{1/2}}{|\omega|^{H_i+H_j-1}}. \tag{11}$$

(iii) Moreover, when $H_i + H_j \neq 1$, the coherence function between the two components i and j satisfies, for all ω

$$\begin{aligned} C_{i,j}(\omega,\delta) &:= \frac{|S_{i,j}(\omega,\delta)|^2}{S_{i,i}(\omega,\delta) S_{j,j}(\omega,\delta)} \\ &= \frac{\Gamma(H_i+H_j+1)^2}{\Gamma(2H_i+1)\Gamma(2H_j+1)} \frac{\rho_{i,j}^2 \sin\left(\frac{\pi}{2}(H_i+H_j)\right)^2 + \eta_{i,j}^2 \cos\left(\frac{\pi}{2}(H_i+H_j)\right)^2}{\sin(\pi H_i)\sin(\pi H_j)}. \end{aligned} \tag{12}$$

(iv) When $H_i + H_j = 1$, (11) and (12) hold, replacing $\rho_{i,j}^2 \sin\left(\frac{\pi}{2}(H_i+H_j)\right)^2 + \eta_{i,j}^2 \cos\left(\frac{\pi}{2}(H_i+H_j)\right)^2$ by $\tilde{\rho}_{i,j}^2 + \frac{\pi^2}{4}\tilde{\eta}_{i,j}^2$.

Proof. The proof is essentially based on the fact that in the generalized function sense, for $\alpha > -1$,

$$\begin{aligned} FT(|h|^\alpha) &= -\frac{1}{\pi}\Gamma(\alpha+1)\sin\left(\frac{\pi}{2}\alpha\right)|\omega|^{-\alpha-1}, \\ FT(h_+^\alpha) &= \frac{1}{2\pi}\Gamma(\alpha+1)e^{-\mathbf{i}\mathrm{sign}(\omega)\frac{\pi}{2}(\alpha+1)}|\omega|^{-\alpha-1}, \\ FT(h_-^\alpha) &= \frac{1}{2\pi}\Gamma(\alpha+1)e^{\mathbf{i}\mathrm{sign}(\omega)\frac{\pi}{2}(\alpha+1)}|\omega|^{-\alpha-1}, \\ FT(h\log|h|) &= \mathbf{i}\frac{\mathrm{sign}(\omega)}{2\omega^2}. \end{aligned}$$

See [5] for more details. □

Remark 4. *From this proposition, we retrieve the same properties of dependence and interdependence of X_i and X_j as stated after Proposition 4.*

2.3 Time reversibility

A stochastic process is said to be time reversible if $X(t) = X(-t)$ for all t. As shows in [12], this is equivalent for zero-mean multivariate Gaussian stationary processes to $\mathbb{E}X_i(t)X_j(s) = \mathbb{E}X_i(s)X_j(t)$ for $s,t \in \mathbb{R}$ or that the cross covariance of the increments satisfies $\gamma_{i,j}(h,\delta) = \gamma_{i,j}(-h,\delta)$ for $h \in \mathbb{R}$. The following proposition characterizes this property.

Proposition 6. *A mfBm is time reversible if and only if $\eta_{i,j} = 0$ (or $\tilde{\eta}_{i,j} = 0$) for all $i,j = 1,\dots,p$.*

Proof. If $\eta_{i,j} = 0$ (or $\tilde{\eta}_{i,j} = 0$), $\gamma_{i,j}(h,\delta)$ is proportional to the covariance of a fractional Gaussian noise with Hurst parameter $(H_i + H_j)/2$ and is therefore symmetric. Let us prove the converse. Let $\alpha = H_i + H_j$, then

$$\gamma_{i,j}(h,\delta) - \gamma_{i,j}(-h,\delta) = \sigma_i\sigma_j \times \begin{cases} -\eta_{i,j}\left(\mathrm{sign}(h-\delta)|h-\delta|^\alpha + 2\mathrm{sign}(h)|h|^\alpha - \mathrm{sign}(h+\delta)|h+\delta|^\alpha\right) & \text{if } \alpha \neq 1, \\ \tilde{\eta}_{i,j}\left((h-\delta)\log|h-\delta| - 2h\log|h| + (h+\delta)\log|h+\delta|\right) & \text{if } \alpha = 1. \end{cases}$$

Assuming $\gamma_{i,j}(h,\delta) - \gamma_{i,j}(-h,\delta)$ equals zero for all h leads to $\eta_{i,j} = 0$ (or $\tilde{\eta}_{i,j} = 0$). □

Remark 5. *This result can also be viewed from a spectral point view. The time reversibility of a mfBm is equivalent to the fact that the spectral density matrix is real. Using (9), this implies $\eta_{i,j} = 0$ (or $\tilde{\eta}_{i,j} = 0$).*

3 Integral representation

3.1 Spectral representation

The following proposition contains the spectral representation of mfBm. This representation will be especially useful to obtain a condition easy to verify which ensures that the functions defined by (4) and (3) are covariance functions.

Theorem 7 (Didier and Pipiras, [12])**.** *Let $(X(t))_{t\in\mathbb{R}}$ be a mfBm with parameter $(H_1,\cdots,H_p) \in (0,1)^p$. Then there exists a $p\times p$ complex matrix A such that each component admits the following representation*

$$X_i(t) = \sum_{j=1}^{p}\int \frac{e^{\mathbf{i}tx}-1}{\mathbf{i}x}(A_{ik}x_+^{-H_i+1/2} + \bar{A}_{ik}x_-^{-H_i+1/2})\tilde{B}_j(\mathrm{d}x), \tag{13}$$

where for all $j = 1,\dots,p$, $\tilde{B}_j$ is a Gaussian complex measure such that $\tilde{B}_j = \tilde{B}_{j,1} + \mathbf{i}\tilde{B}_{j,2}$ with $\tilde{B}_{j,1}(x) = \tilde{B}_{j,1}(-x)$, $\tilde{B}_{j,2}(x) = -\tilde{B}_{j,2}(x)$, $\tilde{B}_{j,1}$ and $\tilde{B}_{j,2}$ are independent and $E(\tilde{B}_{j,i}(\mathrm{d}x)\tilde{B}_{j,i}(\mathrm{d}x)') = \mathrm{d}x$, $i = 1,2$.

Conversely, any p-multivariate process satisfying (13) is a mfBm process.

Proof. This representation is deduced from the general spectral representation of operator fractional Brownian motions obtained in [12]. By denoting $-\mathbb{H}+1/2 := \mathrm{diag}(-H_1+1/2,\cdots,-H_p+1/2)$ we have indeed

$$X(t) = \int \frac{e^{\mathbf{i}tx}-1}{\mathbf{i}x}(Ax_+^{-\mathbb{H}+1/2} + \bar{A}x_-^{-\mathbb{H}+1/2})\tilde{B}(\,\mathrm{d}x), \tag{14}$$

□

Any mfBm having representation (13) has a covariance function as in Proposition 3. The coefficients $\rho_{i,j}$, $\eta_{i,j}$, $\tilde{\rho}_{i,j}$ and $\tilde{\eta}_{i,j}$ involved in (3) and (4) satisfy

$$(AA^*)_{i,j} = \frac{\sigma_i\sigma_j}{2\pi}\Gamma(H_i+H_j+1)\tau_{i,j}(1), \tag{15}$$

where $\tau_{i,j}$ is given in (10) and where A^* is the transpose matrix of $\bar{A}$. This relation is obtained by identification of the spectral matrix of the increments deduced on one hand from (13) and provided on the other hand in Proposition 5.

Given (13), relation (15) provides easily the coefficients $\rho_{i,j}$, $\eta_{i,j}$, $\tilde{\rho}_{i,j}$ and $\tilde{\eta}_{i,j}$ which define the covariance function. The converse is more difficult to obtain. Given a covariance function as in Proposition 3, obtaining the explicit representation (13) requires to find a matrix A such that (15) holds. This choice is possible if and only if the matrix on the right hand side of (15) is positive semidefinite. Then a matrix A (which is not unique) may be deduced by the Cholesky decomposition. When $p=2$, an explicit solution is the matrix with entries, for $i,j=1,2$,

$$A_{i,j} = \lambda_{i,j}\left[\left(\rho_{i,j}\sin\left(\frac{\pi}{2}(H_i+H_j)\right) + \eta_{i,j}\sqrt{\frac{1-C_{i,j}}{C_{i,j}}}\cos\left(\frac{\pi}{2}(H_i+H_j)\right)\right)\right.$$
$$\left. +i\left(\rho_{i,j}\sqrt{\frac{1-C_{i,j}}{C_{i,j}}}\sin\left(\frac{\pi}{2}(H_i+H_j)\right) - \eta_{i,j}\cos\left(\frac{\pi}{2}(H_i+H_j)\right)\right)\right],$$

where $\lambda_{i,j} = \dfrac{\sigma_i}{2\sqrt{\pi}}\dfrac{\Gamma(H_i+H_j+1)}{\sqrt{\Gamma(2H_j+1)\sin(H_j\pi)}}$ and $C_{i,j}$ is given in (12), provided $H_1+H_2\neq 1$. When $H_1+H_2=1$, the same solution holds, replacing $\rho_{i,j}$ by $\tilde{\rho}_{i,j}$ and $\eta_{i,j}\cos\left(\frac{\pi}{2}(H_1+H_2)\right)$ by $-\frac{\pi}{2}\tilde{\eta}_{i,j}$.

3.2 Moving average representation

In the next proposition, we give an alternative characterization of the mfBm from an integral representation in the time domain (or moving average representation).

Theorem 8 (Didier and Pipiras, [12])**.** *Let $(X(t))_{t\in\mathbb{R}}$ be a mfBm with parameter $(H_1,\cdots,H_p)\in(0,1)^p$. Assume that for all $i\in\{1,...,p\}$, $H_i\neq 1/2$. Then there exist M^+, M^- two $p\times p$ real matrices such that each component admits the following representation*

$$X_i(t) = \sum_{j=1}^{p}\int_{\mathbb{R}} M_{i,j}^+\left((t-x)_+^{H_i-.5} - (-x)_+^{H_i-.5}\right) + M_{i,j}^-\left((t-x)_-^{H_i-.5} - (-x)_-^{H_i-.5}\right)W_j(\mathrm{d}x), \tag{16}$$

with $W(\mathrm{d}x) = (W_1(\mathrm{d}x),\cdots,W_p(\mathrm{d}x))$ is a Gaussian white noise with zero mean, independent components and covariance $\mathbb{E}W_i(\mathrm{d}x)W_j(\mathrm{d}x) = \delta_{i,j}\mathrm{d}x$.

Conversely, any p-multivariate process satisfying (16) *is a mfBm process.*

Proof. This representation is deduced from the general representation obtained in [12]. □

Remark 6. *When $H_i = 1/2$ for each $i\in\{1,...,p\}$, it is shown in [12] that each component of the mfBm admits the following representation :*

$$X_i(t) = \sum_{j=1}^{p}\int_{\mathbb{R}} M_{i,j}^+(\mathrm{sign}(t-x)-\mathrm{sign}(x)) + M_{i,j}^-\left(log|t-x| - log|x|\right)W_j(\mathrm{d}x).$$

Our conjecture is that this representation remains valid when $H_i = 1/2$ whatever the values of other parameters H_j, $j \neq i$.

Starting from the moving average representation (16), using results in [27], we can specify the coefficients $\rho_{i,j}$, $\eta_{i,j}$, $\tilde{\rho}_{i,j}$ and $\tilde{\eta}_{i,j}$ involved in the covariances (3) and (4) (see [20]). More precisely, let us denote

$$M^+(M^+)' = \left(\alpha_{i,j}^{++}\right), \quad M^-(M^-)' = \left(\alpha_{i,j}^{--}\right), \quad M^+(M^-)' = \left(\alpha_{i,j}^{+-}\right)$$

where M' is the transpose matrix of M. The variance of each component is equal to

$$\sigma_i^2 = \frac{B(H_i + .5, H_i + .5)}{\sin(H_i\pi)} \left\{\alpha_{i,i}^{++} + \alpha_{i,i}^{--} - 2\sin(H_i\pi)\alpha_{i,i}^{+-}\right\}.$$

Moreover, if $H_i + H_j \neq 1$ then

$$\sigma_i\sigma_j\rho_{i,j} = \frac{B(H_i + .5, H_j + .5)}{\sin((H_i + H_j)\pi)} \left\{(\alpha_{i,j}^{++} + \alpha_{i,j}^{--})(\cos(H_i\pi) + \cos(H_j\pi)) - (\alpha_{i,j}^{+-} + \alpha_{i,j}^{-+})\sin((H_i + H_j)\pi)\right\},$$

$$\sigma_i\sigma_j\eta_{i,j} = \frac{B(H_i + .5, H_j + .5)}{\sin((H_i + H_j)\pi)} \left\{(\alpha_{i,j}^{++} - \alpha_{i,j}^{--})(\cos(H_i\pi) - \cos(H_j\pi)) - (\alpha_{i,j}^{+-} - \alpha_{i,j}^{-+})\sin((H_i + H_j)\pi)\right\}.$$

If $H_i + H_j = 1$ then

$$\sigma_i\sigma_j\tilde{\rho}_{i,j} = B(H_i + .5, H_j + .5)\left\{\frac{\sin(H_i\pi) + \sin(H_j\pi)}{2}(\alpha_{i,j}^{++} + \alpha_{i,j}^{--}) - \alpha_{i,j}^{+-} - \alpha_{i,j}^{-+}\right\},$$
$$\sigma_i\sigma_j\tilde{\eta}_{i,j} = (H_j - H_i)(\alpha_{i,j}^{++} - \alpha_{i,j}^{--}).$$

Conversely, given a covariance function as in Proposition 3, if $H_i \neq 1/2$ for all i, one may find matrices M^+ and M^- such that (16) holds, provided the matrix on the right hand side of (15) is positive semidefinite. Indeed, in this case, a matrix A which solves (15) may be found by the Cholesky decomposition, then M^+ and M^- are deduced from relation (3.20) in [12]:

$$M^{\pm} = \sqrt{\frac{\pi}{2}}\left(D_1^{-1}A_1 \pm D_2^{-1}A_2\right),$$

where $A = A_1 + \mathbf{i}A_2$ and

$$D_1 = \operatorname{diag}\left(\sin(\pi H_1)\Gamma(H_1 + \frac{1}{2}), \ldots, \sin(\pi H_p)\Gamma(H_p + \frac{1}{2})\right),$$

$$D_2 = \operatorname{diag}\left(\cos(\pi H_1)\Gamma(H_1 + \frac{1}{2}), \ldots, \cos(\pi H_p)\Gamma(H_p + \frac{1}{2})\right).$$

3.3 Two particular examples

Let us focus on two particular examples which are quite natural: the causal mfBm ($M^- = 0$) and the well-balanced mfBm ($M^- = M^+$). In the causal case, the integral representation is a direct generalization of the integral representation of Mandelbrot and Van Ness [22] to the multivariate case. The well-balanced case is studied by Stoev and Taqqu in one dimension [27]. With the notation of the two previous sections, we note that the causal case (resp. well-balanced case) leads to $A_1 = \tan(\pi H)A_2$ (resp. $A_2 = 0$), where $\tan(\pi H) := \operatorname{diag}(\tan(\pi H_1), \ldots, \tan(\pi H_p))$. In these two cases, the covariance only depends on one parameter, for instance $\rho_{i,j}$ (or $\tilde{\rho}_{i,j}$). Indeed we easily deduce $\eta_{i,j}$ (or $\tilde{\eta}_{i,j}$) as follows :

- in the causal case : $M^- = 0$ or equivalently $A_1 = \tan(\pi H) A_2$.

$$\begin{aligned} \eta_{i,j} &= -\rho_{i,j}\ \tan(\frac{\pi}{2}(H_i+H_j))\tan(\frac{\pi}{2}(H_i-H_j)) \quad \text{if } H_i+H_j \neq 1, \\ \tilde{\eta}_{i,j} &= \tilde{\rho}_{i,j}\frac{2}{\pi\tan(\pi H_i)} \quad \text{if } H_i+H_j = 1. \end{aligned}$$

- in the well-balanced case : $M^- = M^+$ or equivalently $A_2 = 0$.

$$\begin{aligned} \eta_{i,j} &= 0 \quad \text{if } H_i+H_j \neq 1, \\ \tilde{\eta}_{i,j} &= 0 \quad \text{if } H_i+H_j = 1. \end{aligned}$$

Remark 7. *From Proposition 6, the well-balanced mfBm is therefore time reversible.*

3.4 Existence of the covariance of the mfBm

In this paragraph, we highlight some of the previous results in order to exhibit the sets of the possible parameters $(\rho_{i,j}, \eta_{i,j})$ or $(\tilde{\rho}_{i,j}, \tilde{\eta}_{i,j})$ ensuring the existence of the covariance of the mfBm.

For $i,j = 1,\ldots,p$, let us give $(H_i, H_j) \in (0,1)^2$, $(\sigma_i, \sigma_j) \in \mathbb{R}^+ \times \mathbb{R}^+$ and $(\rho_{i,j}\ \eta_{i,j}) \in [-1,1] \times \mathbb{R}$ with $\rho_{j,i} = \rho_{i,j}$ and $\eta_{j,i} = -\eta_{i,j}$ if $H_i + H_j \neq 1$, or $(\tilde{\rho}_{i,j},\ \tilde{\eta}_{i,j}) \in [-1,1] \times \mathbb{R}$ with $\tilde{\rho}_{j,i} = \tilde{\rho}_{i,j}$ and $\tilde{\eta}_{j,i} = -\tilde{\eta}_{i,j}$ if $H_i + H_j = 1$.

For this set of parameters, let us define the matrix $\Sigma(s,t) = (\Sigma_{i,j}(s,t))$ as follows : $\Sigma_{i,i}(s,t)$ is given by (2) and $\Sigma_{i,j}(s,t)$ is given by (3) when $H_i + H_j \neq 1$ and (4) when $H_i + H_j = 1$.

Proposition 9. *The matrix $\Sigma(s,t)$ is a covariance matrix function if and only if the Hermitian matrix $Q = (\Gamma(H_i+H_j+1)\tau_{i,j}(1))$ with $\tau_{i,j}$ defined in (10), is positive semidefinite. When $p = 2$, this condition reduces to $C_{1,2} \leq 1$ where $C_{1,2}$ is the coherence defined by (12).*

Proof. First, note that since $\rho_{j,i} = \rho_{i,j}$ and $\eta_{j,i} = -\eta_{i,j}$, Q is a Hermitian matrix. Now, if Q is positive semidefinite, then so is the matrix $(2\pi)^{-1}(\sigma_i\sigma_j Q_{i,j})$. Therefore there exists a matrix A satisfying (15). From Theorem 7, there exists a mfBm having $\Sigma(s,t)$ as covariance matrix function. Conversely, if $\Sigma(s,t)$ is a covariance matrix function of a mfBm then the representation (13) holds and by (15), the matrix Q is positive semidefinite.

When $p = 2$, the result comes from the fact that Q is positive semidefinite if and only if $\det(Q) \geq 0$ or equivalently $C_{1,2} \leq 1$. □

When $p = 2$, for fixed values of (H_1, H_2) the condition $C_{1,2} \leq 1$ means that the set of possible parameters $(\rho_{1,2}, \eta_{1,2})$ is the interior of an ellipse. These sets are represented in Figure 1 according to different values of H_1 and H_2. Note that, in order to compare the cases $H_1 + H_2 \neq 1$ and $H_1 + H_2 = 1$, we have reparameterized $\eta_{1,2}$ by $\eta'_{1,2} := \eta_{1,2}/(1 - H_1 - H_2)$. In such a way, the second ellipse becomes the limit of the first one as $H_1 + H_2 \to 1$ (see also Remark 3).

Let us underline that the maximum possible correlation between two fBm's is obtained when $\eta_{1,2} = 0$, i.e. when the 2-mfBm is time reversible according to Proposition 6.

Remark 8. *When $H_1 = \ldots = H_p = H \neq 1/2$, the matrix Q rewrites $Q_{i,j} = \Gamma(2H+1)(\sin(\pi H)\rho_{i,j} - \mathbf{i}\eta_{i,j}\cos(\pi H))$ and*

- *if the mfBm is time reversible, i.e. $\eta_{i,j} = 0$ (for $i,j = 1,\ldots,p$), then Q is a correlation matrix and is therefore positive-semidefinite for any $|\rho_{i,j}| \leq 1$,*
- *when $p = 2$, the set of possible values for $(\rho_{1,2}, \eta_{1,2})$ associated to H and $1-H$ are the same.*

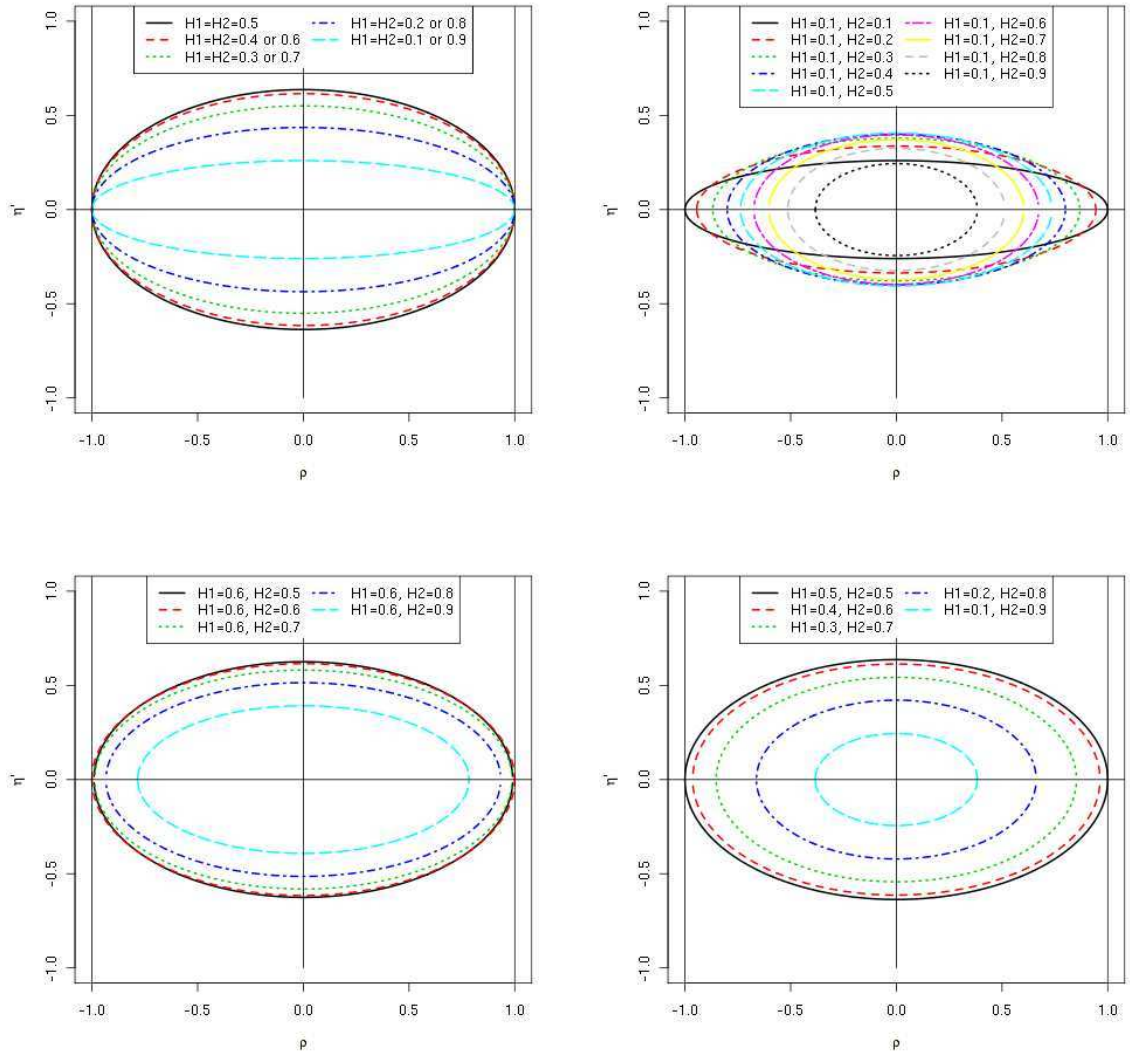


Figure 1: Various examples of possible values for $(\rho_{1,2}, \eta'_{1,2})$ with $\eta'_{1,2} := \eta_{1,2}/(1 - H_1 - H_2)$ when $H_1 + H_2 \neq 1$ and $(\tilde{\rho}_{1,2}, \tilde{\eta}_{1,2})$ when $H_1 + H_2 \neq 1$, ensuring that $\Sigma(s,t)$ is a covariance matrix function in the particular case $p = 2$.

In the particular case of the causal or the well-balanced mfBm, the matrix $\Sigma(s,t)$ can be expressed through the sole parameter $\rho_{i,j}$. The maximal possible correlation when $p = 2$ is given by

$$\rho_{1,2}^2 = \frac{\Gamma(2H_1+1)\Gamma(2H_2+1)}{\Gamma(H_1+H_2+1)^2}\frac{\sin(\pi H_1)\sin(\pi H_2)}{\sin(\frac{\pi}{2}(H_1+H_2))^2} \times \begin{cases} \cos(\frac{\pi}{2}(H_1-H_2))^2 & \text{in the causal case,} \\ 1 & \text{in the well-balanced case.} \end{cases}$$

Figure 2 represents $|\rho_{1,2}|$ with respect to (H_1, H_2).

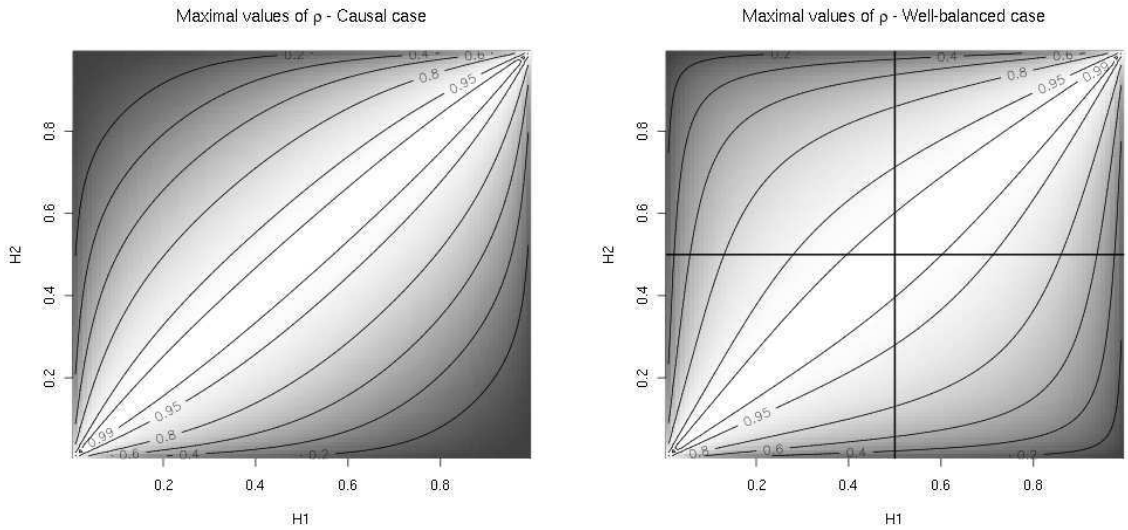


Figure 2: Maximal values of the absolute possible correlation parameter $|\rho_{1,2}|$ ensuring that $\Sigma(s,t)$ is a covariance matrix function in the case p = 2, in terms of H_1 and H_2 for the causal and well-balanced mfBm.

Figures 1 and 2 illustrate the main limitation of the mfBm model. Under self-similarity condition (1), it is not possible to construct arbitrary correlated fractional Brownian motions. For example, when $H_1 = 0.1$ and $H_2 = 0.8$, the correlation cannot exceed 0.514.

4 The mfBm as a limiting process.

A natural way to construct self-similar processes is through limits of stochastic processes. In dimension one, the result is due to Lamperti [18]. In [16], an extension to operator self-similar processes is given. A similar result for the mfBm is deduced and stated below. In the following, a p-multivariate process $(X(t))_{t\in\mathbb{R}}$ is said proper if, for each t, the law of $X(t)$ is not contained in a proper subspace of $\mathbb{R}^p$.

Theorem 10. *Let $(X(t))_{t\in\mathbb{R}}$ be a p-multivariate proper process, continuous in probability. If there exist a p-multivariate process $(Y(t))_{t\in\mathbb{R}}$ and p real functions $a_1, ..., a_p$ such that*

$$(a_1(n)Y_1(nt), \ldots, a_p(n)Y_p(nt)) \xrightarrow[fidi]{n\to\infty} X(t), \tag{17}$$

then the multivariate process $(X(t))$ is self-similar. Conversely, any multivariate self-similar process can be obtained as a such limit.

Proof. The proof is similar to Theorem 5 in [16]. Fix $k \in \mathbb{N}$ and $r > 0$. For each $T \in \mathbb{R}^k$ we denote $X(T) := (X(T_1), \ldots, X(T_k))$. Let $\mathcal{D}_{r,k}$ be the set of all invertible diagonal matrices α such that, for all $T \in \mathbb{R}^k$, $X(rT) = \alpha X(T)$.

Let us first show that $\mathcal{D}_{r,k}$ is not empty. According to (17), we have

$$\mathrm{diag}(a_1(n), \ldots, a_p(n))Y(nrT) \xrightarrow[d]{n\to\infty} X(rT),$$

and

$$\mathrm{diag}(a_1(rn), \ldots, a_p(rn))Y(nrT) \xrightarrow[d]{n\to\infty} X(T).$$

Since $(X(t))$ is proper, $\mathrm{diag}(a_1(n),\dots,a_p(n))$ and $\mathrm{diag}(a_1(rn),\dots,a_p(rn))$ are invertible for n large enough. Then, Theorem 2.3 in [31] ensures that α_n defined by

$$\alpha_n = \mathrm{diag}(a_1(n),\dots,a_p(n))\mathrm{diag}(a_1(nr),\dots,a_p(nr))^{-1}$$

has a limit in $\mathcal{D}_{r,k}$. Moreover if α is a limit of α_n then $X(rT) = \alpha X(T)$ and thus $\mathcal{D}_{r,k} \neq \emptyset$.

It is then straightforward to adapt Lemma 7.2-7.5 in [16] for the subgroup $\mathcal{D}_{r,k}$, which yields that for each r, $\cap_k \mathcal{D}_{r,k}$ is not empty. Therefore, for any fixed $r > 0$, there exists $\alpha \in \cap_k \mathcal{D}_{r,k}$ such that $(X(rt))$ and $(\alpha X(T))$ have the same finite dimensional distributions. Theorem 1 in [16] ensures that there exists $(H_1,\dots,H_p) \in (0,1)^p$ such that $\alpha = \mathrm{diag}(r^{H_1},\dots,r^{H_p})$. The converse is trivial.

□

As an illustration of Theorem 10, the mfBm can be obtained as the weak limit of partial sums of sum of linear processes (also called superlinear processes, see [33]). Some examples may be found in [8] and [19]. In Proposition 11 below, we give a general convergence result which allows to reach almost any mfBm from such partial sums. The unique restriction concerns the particular case when at least one of the Hurst parameters is equal to 1/2.

Let $(\epsilon_j(k))_{k\in Z}$, $j = 1,\dots,p$ be p independent i.i.d. sequences with zero mean and unit variance. Let us consider the superlinear processes

$$Z_i(t) = \sum_{j=1}^{p} \sum_{k\in\mathbb{Z}} \psi_{i,j}(t-k)\epsilon_j(k), \quad i = 1,\dots,p, \tag{18}$$

where $\psi_{i,j}(k)$ are real coefficients with $\sum_{k\in\mathbb{Z}} \psi_{i,j}^2(k) < \infty$.

Moreover, we assume that $\psi_{i,j}(k) = \psi_{i,j}^+(k) + \psi_{i,j}^-(k)$ where $\psi_{i,j}^+(k)$ satisfies one of the following conditions:

(i) $\psi_{i,j}^+(k) = \left(\alpha_{i,j}^+ + o(1)\right) k_+^{d_{i,j}^+ - 1}$ as $|k| \to \infty$, with $0 < d_{i,j}^+ < \frac{1}{2}$ and $\alpha_{i,j}^+ \neq 0$,

(ii) $\psi_{i,j}^+(k) = \left(\alpha_{i,j}^+ + o(1)\right) k_+^{d_{i,j}^+ - 1}$ as $|k| \to \infty$, with $-\frac{1}{2} < d_{i,j}^+ < 0$, $\sum_{k\in\mathbb{Z}} \psi_{i,j}^+(k) = 0$ and $\alpha_{i,j}^+ \neq 0$,

(iii) $\sum_{k\in\mathbb{Z}} \left|\psi_{i,j}^+(k)\right| < \infty$ and let $\alpha_{i,j}^+ := \sum_{k\in\mathbb{Z}} \psi_{i,j}^+(k) \neq 0$, $d_{i,j}^+ := 0$.

Similarly, $\psi_{i,j}^-(k)$ is assumed to satisfy (i), (ii) or (iii) where k_+, $d_{i,j}^+$ and $\alpha_{i,j}^+$ are replaced by k_-, $d_{i,j}^-$ and $\alpha_{i,j}^-$.

Proposition 11. *Let $d_i = max(d_{i1}^+, d_{i1}^-, \cdots, d_{ip}^+, d_{ip}^-)$, for $i = 1,\dots,p$. Consider the vector of partial sums, for $\tau \in \mathbb{R}$,*

$$S_n(\tau) = \left(n^{-d_1-(1/2)} \sum_{t=1}^{[n\tau]} Z_1(t), \cdots, n^{-d_p-(1/2)} \sum_{t=1}^{[n\tau]} Z_p(t) \right).$$

Then the finite dimensional distributions of $(S_n(\tau))_{\tau\in\mathbb{R}}$ converge in law towards a p-mfBm $(X(\tau))_{\tau\in\mathbb{R}}$.

- *When $d_i \neq 0$, $(X_i(\tau))_{\tau\in\mathbb{R}}$ is defined through the integral representation (16) where $M_{i,j}^+ = \alpha_{i,j}^+ d_i^{-1} \mathbf{1}_{d_{i,j}^+ = d_i}$ and $M_{i,j}^- = \alpha_{i,j}^- d_i^{-1} \mathbf{1}_{d_{i,j}^- = d_i}$.*
- *When $d_i = 0$, $X_i(\tau) = \sum_{j=1}^d (\alpha_{i,j}^+ \mathbf{1}_{d_{i,j}^+ = 0} + \alpha_{i,j}^- \mathbf{1}_{d_{i,j}^- = 0}) W_j(\tau)$, where W_j is a standard Brownian motion.*

Moreover, if for all $j = 1,\dots,p$, $E(\epsilon_j(0)^{2\alpha}) < \infty$ with $\alpha > 1 \vee (1 + 2d_{max})^{-1}$ where $d_{max} = max_i\{d_i\}$, then $S_n(.)$ converges towards the p-mfBm $X(.)$ in the Skorohod space $\mathcal{D}([0,1])$.

Sketch of proof. We focus on the convergence in law of $S_n(\tau)$ to $X(\tau)$, for a fixed τ in $\mathbb{R}$, the finite dimensional convergence is deduced in the same way. We set for simplicity $\tau = 1$.

According to the Cramér-Wold device, for any vector $(\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p$, we must show that $\lambda' S_n(1)$ converges in law to $\lambda' X(1)$. We may rewrite $\lambda' S_n(1)$ as a sum of discrete stochastic integrals (see [29] and [4]) :

$$\begin{aligned}
\lambda' S_n(1) &= \sum_{i=1}^{p} \lambda_i n^{-d_i-(1/2)} \sum_{t=1}^{n} Z_i(t) \\
&= \sum_{i=1}^{p} \lambda_i \sum_{j=1}^{p} \sum_{k\in\mathbb{Z}} n^{-d_i-(1/2)} \sum_{t=1}^{n} (\psi_{i,j}^{+}(t-k) + \psi_{i,j}^{-}(t-k))\epsilon_j(k) \\
&= \sum_{i=1}^{p} \lambda_i \sum_{j=1}^{p} \int_{\mathbb{R}} (f_{i,j,n}^{+}(x) + f_{i,j,n}^{-}(x)) W_{j,n}(dx), \qquad (19)
\end{aligned}$$

where the stochastic measures $W_{j,n}$, $j = 1, \ldots, p$ are defined on finite intervals C by

$$W_{j,n}(C) = n^{-1/2} \sum_{k/n\in C} \epsilon_j(k),$$

and $f_{i,j,n}^{+}$, $f_{i,j,n}^{-}$ are piecewise constant functions defined for all $k \in \mathbb{Z}$ and for $nx \in (k-1, k]$ by $f_{i,j,n}^{+}(x) = n^{-d_i} \sum_{t=1}^{n} \psi_{i,j}^{+}(t-k)$, respectively $f_{i,j,n}^{-}(x) = n^{-d_i} \sum_{t=1}^{n} \psi_{i,j}^{-}(t-k)$.

The following lemma states the convergence of a linear combination of discrete stochastic integrals as in (19). A function is said n-simple if it takes a finite number a nonzero constant values on intervals $(k/n, (k+1)/n]$, $k \in \mathbb{Z}$.

Lemma 12. *Let $(f_{1,n}, \cdots, f_{p,n})_{n\in\mathbb{N}}$ be a sequence of p n-simple functions in $L^2(\mathbb{R})$. If for any $j = 1, \ldots, p$, there exists $f_j \in L^2(\mathbb{R})$ such that $\int_{\mathbb{R}} |f_{j,n}(x) - f_j(x)|^2 dx \to 0$, then $\sum_{j=1}^{p} \int_{\mathbb{R}} f_{j,n}(x) W_{j,n}(dx)$ converges in law to $\sum_{j=1}^{p} \int_{\mathbb{R}} f_j(x) W_j(dx)$, where the W_j's are independent standard Gaussian random measures.*

When $p = 1$, this lemma is proved in [28]. The case $p = 2$ is considered in [4] and the extension to $p \geq 3$ is straightforward.

From Lemma 12 and (19), it remains to show that

$$\lim_{n\to\infty} \int_{\mathbb{R}} \left| f_{i,j,n}^{\pm}(x) - \frac{\alpha_{i,j}^{\pm}}{d_i} \left((1-x)_{\pm}^{d_i} - (-x)_{\pm}^{d_i} \right) \mathbf{1}_{d_{i,j}^{\pm} = d_i} \right|^2 dx = 0,$$

where we agree that $d_i^{-1}((1-x)_{\pm}^{d_i} - (-x)_{\pm}^{d_i}) = \mathbf{1}_{[0,1]}(x)$ when $d_i = 0$. Below, we only consider the pointwise convergence of $f_{i,j,n}^{\pm}(x)$, for $x \in \mathbb{R}$, when $d_{i,j}^{\pm} = d_i$. The convergence in L^2 is then deduced from the dominated convergence theorem (see [28], [29], [4] for details). It also follows easily that, when $d_{i,j}^{\pm} < d_i$, $\int_{\mathbb{R}} |f_{i,j,n}^{\pm}(x)|^2 dx \to 0$.

Under assumption (i), note that since $d_i > 0$, $(1-x)_{\pm}^{d_i} - (-x)_{\pm}^{d_i} = d_i \int_0^1 (t-x)_{\pm}^{d_i-1} dt$. Let us denote by $\lceil x \rceil$ the smallest integer not less than x. We have, for any $x \in \mathbb{R}$,

$$\begin{aligned}
f_{i,j,n}^{\pm}(x) &= n^{-d_i} \sum_{t=1}^{n} \psi_{i,j}^{\pm}(t - \lceil nx \rceil) \\
&= n^{-d_i} \int_0^n \psi_{i,j}^{\pm}(\lceil t \rceil - \lceil nx \rceil) dt \\
&= n^{-d_i} \int_0^n \left(\alpha_{i,j}^{\pm} + o(1) \right) (\lceil t \rceil - \lceil nx \rceil)_{\pm}^{d_i-1} dt \\
&= \int_0^1 \left(\alpha_{i,j}^{\pm} + o(1) \right) \left(\frac{\lceil nt \rceil - \lceil nx \rceil}{n} \right)_{\pm}^{d_i-1} dt \longrightarrow \alpha_{i,j}^{\pm} \int_0^1 (t-x)_{\pm}^{d_i-1} dt.
\end{aligned}$$

Under assumption (ii), $d_i < 0$. When $x \leq 0$, $(1-x)_+^{d_i} - (-x)_+^{d_i} = d_i \int_0^1 (t-x)^{d_i-1} dt$ and the convergence of $f_{i,j,n}^+(x)$ can be proved as above. When $x \geq 1$, $(1-x)_+^{d_i} - (-x)_+^{d_i} = 0 = f_{i,j,n}^+(x)$. When $0 \leq x \leq 1$, $(1-x)_+^{d_i} - (-x)_+^{d_i} = -d_i \int_1^{+\infty} (t-x)^{d_i-1} dt$ and, since $\sum_{k\in\mathbb{Z}} \psi_{i,j}^+(k) = 0$, we have

$$f_{i,j,n}^+(x) = n^{-d_i} \sum_{t=\lceil nx \rceil}^{n} \psi_{i,j}^+(t - \lceil nx \rceil) = n^{-d_i} \sum_{0}^{n-\lceil nx \rceil} \psi_{i,j}^+(t) = -n^{-d_i} \sum_{t>n-\lceil nx \rceil} \psi_{i,j}^+(t).$$

Therefore,

$$\begin{aligned} f_{i,j,n}^+(x) &= -n^{-d_i} \int_{n-\lceil nx \rceil}^{+\infty} \left(\alpha_{i,j}^+ + o(1)\right) \left(\lceil t \rceil\right)^{d_i-1} dt \\ &= -\int_{1-\frac{\lceil nx \rceil}{n}}^{+\infty} \left(\alpha_{i,j}^+ + o(1)\right) \left(\frac{\lceil nt \rceil}{n}\right)^{d_i-1} dt \longrightarrow -\alpha_{i,j}^+ \int_{1-x}^{+\infty} t^{d_i-1} dt = -\alpha_{i,j}^+ \int_1^{+\infty} (t-x)^{d_i-1} dt. \end{aligned}$$

This proves $f_{i,j,n}^+(x) \to d_i^{-1} \alpha_{i,j}^+ ((1-x)_+^{d_i} - (-x)_+^{d_i})$, for any $x \in \mathbb{R}$, under assumption (ii). The same scheme may be used to prove that $f_{i,j,n}^-(x) \to d_i^{-1} \alpha_{i,j}^- ((1-x)_-^{d_i} - (-x)_-^{d_i})$ under assumption (ii), noting that

$$(1-x)_-^{d_i} - (-x)_-^{d_i} = \begin{cases} 0 & \text{when} \quad x \leq 0, \\ -d_i \int_{-\infty}^0 (t-x)^{d_i-1} dt & \text{when} \quad 0 \leq x \leq 1, \\ d_i \int_0^1 (t-x)_-^{d_i-1} dt & \text{when} \quad x > 1. \end{cases}$$

Under assumption (iii),

$$f_{i,j,n}^{\pm}(x) = \sum_{t=1}^{n} \psi_{i,j}^{\pm}(t - \lceil nx \rceil) = \sum_{t=1-\lceil nx \rceil}^{n-\lceil nx \rceil} \psi_{i,j}^{\pm}(t).$$

Since $\sum_{t\in\mathbb{Z}} \psi_{i,j}^{\pm}(t) < \infty$, $f_{i,j,n}^{\pm}(x) \to 0$ for all $x \notin [0,1]$. When $x \in [0,1]$, we have $f_{i,j,n}^{\pm}(x) \to \alpha_{i,j}^{\pm}$.

Therefore, the first claim of the theorem is proved, i.e. the convergence in law of the finite dimensional distribution of $(S_n(\tau))_{\tau\in\mathbb{R}}$ to $(X(\tau))_{\tau\in\mathbb{R}}$. To extend this convergence to a functional convergence in $\mathcal{D}([0,1])$, it remains to show tightness of the sequence $(S_n(\tau))_{\tau\in[0,1]}$. This follows exactly from the same arguments as in the proof of Theorem 1.2 in [4]. □

5 Synthesis of the mfBm

5.1 Introduction

The exact simulation of the fractional Brownian motion has been a question of great interest in the nineties. This may be done by generating a sample path of a fractional Gaussian noise. An important step towards efficient simulation was obtained after the work of Wood and Chan [32] about the simulation of arbitrary stationary Gaussian sequences with prescribed covariance function. The technique relies upon the embedding of the covariance matrix into a circulant matrix, a square root of which is easily calculated using the discrete Fourier transform. This leads to a very efficient algorithm, both in terms of computation time and storage needs. Wood and Chan methods is an exact simulation method provided that the circulant matrix is definite positive, a property that is not always satisfied. However, for the fractional Gaussian noise, it can be proved that the circulant matrix is definite positive for all $H \in (0,1)$, see [9, 13].

In [7], Wood and Chan extended their method and provided a more general algorithm adapted to multivariate stationary Gaussian processes. The main characteristic of this method is that if a certain condition for a familiy of Hermitian matrices holds then the algorithm is exact in principal, i.e. the simulated data has the true covariance. We present hereafter the main ideas, briefly describe the algorithm and propose some examples.

Remark 9. *Other approaches could have been undertaken (see [3] for a review in the case $p = 1$). Approximate simulations can be done by discretizing the moving-average or spectral stochastic integrals (13) or (16). [6] also proposed an approximate method based on the spectral density matrix of the increments for synthesizing multivariate Gaussian time series. Thanks to Proposition 5, this could also be envisaged for the mfBm.*

5.2 Method and algorithm

For two arbitrary matrices $A = (A_{j,k})$ and B, we use $A \otimes B$ to denote the Kronecker product of A and B that is the block matrix $(A_{j,k}B)$.

Let $\Delta X := \Delta_1 X$ denotes the increments of size 1 ($\delta = 1$) of a mfBm. We have $\Delta X = (\Delta X(t))_{t\in\mathbb{R}} = ((\Delta X_1(t), \ldots, \Delta X_p(t))')_{t\in\mathbb{R}}$. The aim is to simulate a realization of a multivariate fractional Gaussian noise discretized at times $j = 1, \ldots, n$, that is a realization of $(\Delta X(1), \ldots, \Delta X(n))$. Then a realization of the discretized mfBm will be easily obtained.

We denote by $\Delta X^{(n)}$ the merged vector $\Delta X^{(n)} = (\Delta X(1)', \ldots, \Delta X(n)')'$ and by $\mathbb{G}$ its covariance matrix. $\mathbb{G}$ is the $np \times np$ Toeplitz block matrix $\mathbb{G} = (G(|i-j|))_{i,j=1,\ldots,n}$ where for $h = 0, \ldots, n-1$, $G(h)$ is the $p\times p$ matrix given by $G(h) := (\gamma_{j,k}(h))_{j,k=1,\ldots,p}$. The simulation problem can be viewed as the generation of a random vector following a $\mathcal{N}_{np}(0, \mathbb{G})$. This may be done by computing $\mathbb{G}^{1/2}$ but the complexity of such a procedure is $\mathcal{O}(pn^3)$. To overcome this numerical cost, the idea is to embed $\mathbb{G}$ into the block circulant matrix $C = circ\{C(j), j = 0, \ldots, m-1\}$, where m is a power of 2 greater than $2(n-1)$ and where each $C(j)$ is the $p \times p$ matrix defined by

$$C(j) = \begin{cases} G(j) & \text{if } 0 \le j < m/2 \\ \frac{1}{2}\big(G(j) + G(j)'\big) & \text{if } j = m/2 \\ G(j-m) & \text{if } m/2 < j \le m-1. \end{cases} \tag{20}$$

Such a definition ensures that C is a symmetric matrix with nested block circulant structure and that $\mathbb{G} = \{C(j), j = 0, \ldots, n-1\}$ is a submatrix of C. Therefore, the simulation of a $\mathcal{N}_{np}(0, \mathbb{G})$ may be achieved by taking the n "first" components of a vector $\mathcal{N}_{mp}(0, C)$, which is done by computing $C^{1/2}$. The last problem is more simple since one may exploit the circulant characteristic of C: there exist m Hermitian matrices $B(j)$ of size $p \times p$ such that the following decomposition holds

$$C = (J \otimes I_p)\ \mathrm{diag}(B(j), j = 0, \ldots, m-1)\ (J^* \otimes I_p), \tag{21}$$

where Q is the $m \times m$ unitary matrix defined for $j, k = 0, m-1$ by $J_{j,k} = e^{-2\mathbf{i}\pi jk/m}$. The computation of $C^{1/2}$ is much less expensive than the computation of $\mathbb{G}^{1/2}$ since, as in the one-dimensional case ($p = 1$), (21) will allow us to make use of the Fast Fourier Transform (FFT) which considerable reduces the complexity.

Now, the algorithm proposed by Wood and Chan may be described through the following steps. Let m be a power of 2 greater than $2(n-1)$.

Step 1. For $1 \le u \le v \le p$ calculate for $k = 0, \ldots, m-1$

$$B_{u,v}(k) = \sum_{j=0}^{m-1} C_{u,v}(j) e^{-2\mathbf{i}\pi jk/m}$$

where $C_{u,v}(j)$ if the element (u, v) of the matrix $C(j)$ defined by (20) and set $B_{vu}(k) = B_{u,v}(k)^*$.

Step 2. For each $j = 0, \ldots, m-1$ determine a unitary matrix $R(j)$ and real numbers $\xi_u(j)$ $(u = 1, \ldots, p)$ such that $B(j) = R(j)\ \mathrm{diag}(\xi_1(j), \ldots, \xi_p(j))\ R(j)^*$.

Step 3. Assume that the eigenvalues $\xi_1(j), \ldots, \xi_p(j)$ are non-negative (see Remark 11) and define $\widetilde{B}(j) = R(j)\ \mathrm{diag}(\sqrt{\xi_1(j)}, \ldots, \sqrt{\xi_p(j)})\ R(j)^*$.

Step 4. For $j = 0, \ldots, m/2$ generate independent vectors $U(j), V(j) \sim \mathcal{N}_p(0, I)$ and define

$$Z(j) = \frac{1}{\sqrt{2m}} \times \begin{cases} \sqrt{2}U(j) & \text{for } j = 0, \frac{m}{2} \\ U(j) + \mathbf{i}V(j) & \text{for } j = 1, \ldots, \frac{m}{2} - 1, \end{cases}$$

let $Z(m-j) = \overline{Z}(j)$ for $j = \frac{m}{2} + 1, \ldots, m-1$ and set $W(j) := \widetilde{B}(j)Z(j)$.

Step 5. For $u = 1, \ldots, p$ calculate for $k = 0, \ldots, m-1$

$$\Delta X_u(k) = \sum_{j=0}^{m-1} W_u(j) e^{-2\mathbf{i}\pi jk/m}$$

and return $\left\{\Delta X_u(k), 1 \leq u \leq p, k = 0, \ldots, n-1\right\}$.

Step 6. For $u = 1, \ldots, p$ take the cumulative sums ΔX_u to get the $u-th$ component X_u of a sample path of a mfBm.

Figure 3 gives some examples of sample paths of mfBm's simulated with this algorithm.

Remark 10. *Let us discuss the computation cost of the most expensive steps, that is steps 1, 2 and 5. Step 1 requires $\frac{p(p+1)}{2}$ applications of the FFT of signals of length m, Step 2 needs m diagonalisations of $p \times p$ Hermitian matrices and Step 5 requires p applications of the FFT of signals of length m. Therefore, the total cost, $\kappa(m,p)$ equals*

$$\kappa(m,p) = \mathcal{O}\left(\frac{p(p+1)}{2} m \log m\right) + \mathcal{O}(mp^3) + \mathcal{O}(pm \log m).$$

Remark 11. *The crucial point of the previous algorithm lies in the non-negativity of the eigenvalues $\xi_1(j), \ldots, \xi_p(j)$ for any $j = 0, \ldots, m-1$. In the one-dimensional case (when $p = 1$) Steps 2 and 3 disappear, and in Step 1, $B_{11}(k)$ corresponds to the $k-th$ eigenvalue of the circulant matrix C_{11} with first line defined by $C_{11}(j) = \gamma_{11}(j)$ for $0 \leq j \leq m/2$ and $\gamma(m-j)$ for $j = m/2+1, \ldots, m-1$. For the fractional Gaussian noise, it has been proved by Craigmile [9] for $H < 1/2$, and by Dietrich and Newsam [13] for $H > 1/2$ that such a matrix is semidefinite-positive for any m (and so for the first power of 2 greater than $2(n-1)$). In the more general case $p > 1$, the problem is much more complex: the quantities $B_{u,v}(k)$ are not necessarily real, and the establishment of a condition of positivity for the matrix $B_{u,v}(k)$ does not seem obvious. When the condition in Step 3 does not hold, Wood and Chan suggest to either increase the value of m and restart Steps 1,2 or to truncate the negative eigenvalues to zero which leads to an approximate procedure. These problems will be deserved in a separate paper. Let us assert that for the simulation examples presented in the next section, we have observed that this condition is satisfied for m equal to the first power of 2.*

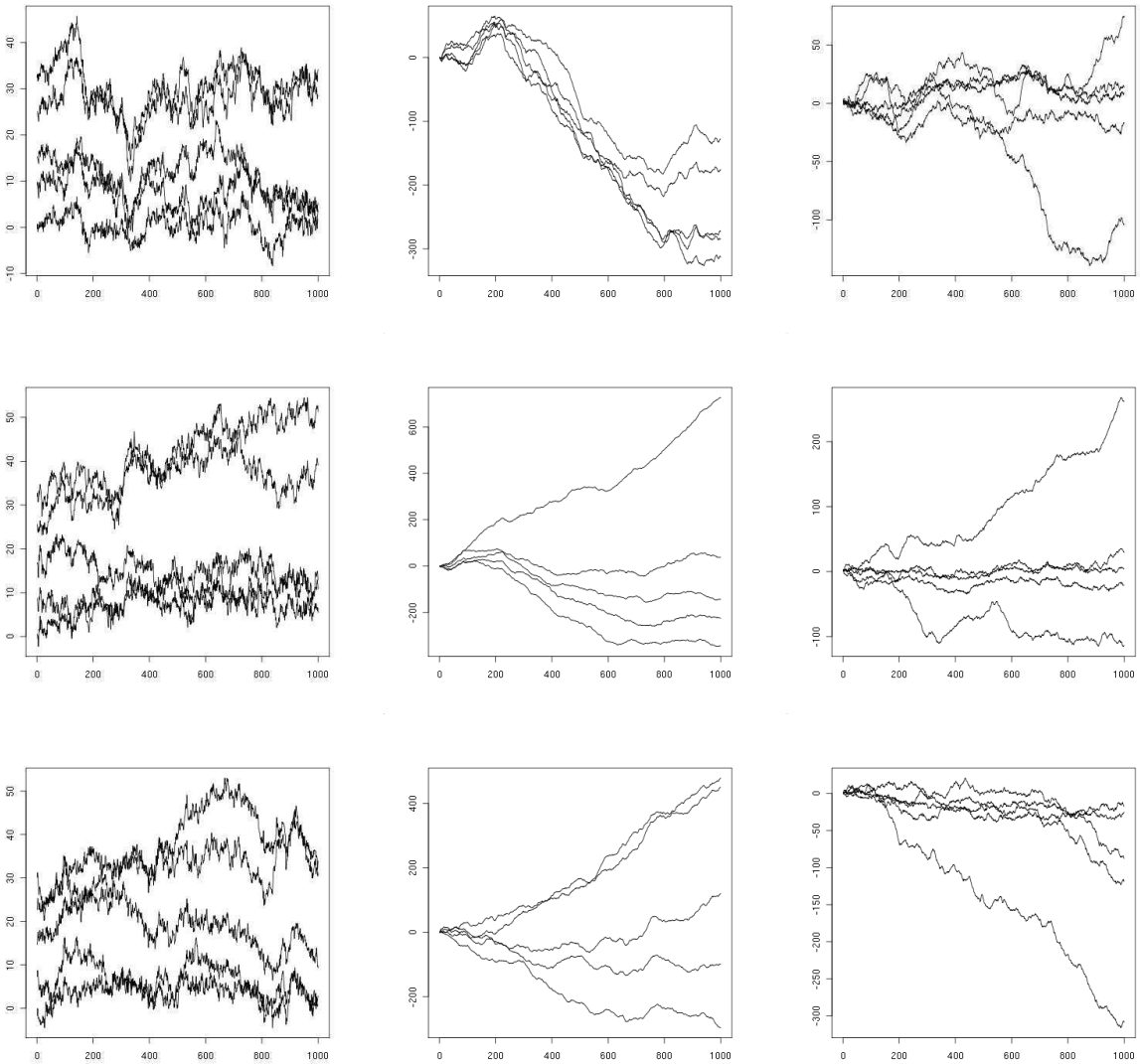


Figure 3: Examples of sample paths of length $n = 1000$ normalized causal (top), well-balanced (middle) and general (bottom) multivariate fractional Brownian motion with $p = 5$ components. The Hurst parameters are equally spaced in $[0.3, 0.4]$ (left), $[0.8, 0.9]$ (middle) and $[0.4, 0.8]$ (right). The correlation parameters $\rho_{i,j}$ are all set to 0.6 (left, middle) and 0.3 (right). For the general mfBm (bottom), the parameters $\eta_{i,j}$ are set to $0.15 \times (1 - H_i - H_j)$. Note that the existence condition discussed in Proposition 9 and the condition in Step 3 of the algorithm are satisfied for these different choices of parameters. For convenience, the sample paths of the left column have been decentered.

References

[1] Achard, S., Bassett, D. S., Meyer-Lindenberg, A. and Bullmore, E. (2008) Fractal connectivity of long-memory networks. *Phys. Rev. E*, 77:036104.

[2] Arianos, S. and Carbone, A. (2009) Cross-correlation of long-range correlated series. *Journal of Statistical Mechanics : Theory and Experiment*, page P03037.

[3] Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A. and Taqqu, M. (2003) Generators of long-range dependent processes : A survey. In: Doukhan, P., Oppenheim, G., Taqqu, M.S. (Eds.), *Theory and Applications of Long-Range Dependence: Theory and Applications*, pp. 557–578. Birkhäuser, Boston.

[4] Bružaitė, K. and Vaičiulis, M., (2005) Asymptotic independence of distant partial sums of linear process. Lithuanian Mathematical Journal 45, 387–404.

[5] Coeurjolly, J.-F., Amblard, P.O., Achard, S., (2010) Normalized causal and well-balanced multivariate fractional Brownian motion. Submitted.

[6] Chambers, J. (1995) The simulation of random vector time series with given spectrum. *Mathematical and Computer Modelling*, 22:1–6.

[7] Chan, G. and Wood, A. (1999) Simulation of stationary gaussian vector fields. *Statistics and Computing*, **9**(4), 265–268.

[8] Chung, C.-F. (2002) Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes. *Econometric Th.* **18**, 51–78.

[9] Craigmile P. (2003) Simulating a class of stationary gaussian processes using the davies-harte algortihm, with application to long memory processes. *Journal of Time Series Analysis*, **24**,505–510.

[10] Davidson, J. and de Jong, R.M. (2000) The functional central limit theorem and weak convergence to stochastic integrals. *Econometric Th.* **16**, 643–666.

[11] Davidson, J. and Hashimadze, N. (2008) Alternative frequency and time domain versions of fractional Brownian motion. *Econometric Th.* **24**, 256–293.

[12] Didier, G. and Pipiras, V. (2010) Integral representations of operator fractional Brownian motion. To appear in Bernoulli.

[13] Dietrich, C. R. and Newsam, G. N. (1997) Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific and Statistical Computing*, **18** 1088 - 1107.

[14] Fleming B.J.W., Yu D., Harrison R.G. and Jubb D. (2001) Wavelet-based detection of coherent structures and self-affinity in financial data. *The European Physical Journal B*, **20**(4), 543-546.

[15] Gil-Alana L.A. (2003) A fractional multivariate long memory model for the US and the Canadian real output. *Economics Letters*, **81**(3), 355-359.

[16] Hudson, W. and Mason, J. (1982) Operator-self-similar processes in a finite-dimensional space. *Trans. Amer. Math. Soc.* **273**, 281–297.

[17] Laha, R.G. and Rohatgi, V.K. (1981) Operator self-similar stochastic processes in R^d. *Stochastic Proces. Appl.* **12**, 73–84.

[18] Lamperti J. (1962) Semi-Stable Stochastic Processes. *Transactions of the American Mathematical Society*, **104**(1), pp. 62-78

[19] Lavancier, F., Philippe, A. and Surgailis, D. (2010) A two-sample test for comparison of long memory parameters. *Journal of Multivariate Analysis.* (To appear)

[20] Lavancier, F., Philippe, A. and Surgailis, D. (2009) Covariance function of vector self-similar process. *Statistics and Probability Letters.* **79**, 2415-2421.

[21] Maejima, M. and Mason, J. (1994) Operator-self-similar stable processes. *Stochastic Proces. Appl.* **54**, 139–163.

[22] Mandelbrot, B. and Van Ness, J. (1968) Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10(4):422–437

[23] Marinucci, D. and Robinson, P.M. (2000) Weak convergence of multivariate fractional processes. *Stochastic Process. Appl.* **86**, 103–120.

[24] Robinson, P.M. (2008) Multiple local Whittle estimation in stationary systems. *Ann. Statist.* **36**, 2508–2530.

[25] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non-Gaussian Random Processes.* Chapman and Hall, New York.

[26] Sato, K. (1991) Self-similar processes with independent increments. *Probab. Th. Rel. F.* **89**, 285–300.

[27] Stoev, S. and Taqqu, M.S. (2006) How rich is the class of multifractional Brownian motions? *Stochastic Process. Appl.* **11**, 200–221.

[28] Surgailis, D. (1982). Zones of attraction of self-similar multiple integrals. Lithuanian J. Math. **22** 185 201.

[29] Surgailis, D. (2003) Non CLTs: U-statistics, multinomial formula and approximations of multiple Itô-Wiener integrals. In: Doukhan, P., Oppenheim, G., Taqqu, M.S. (Eds.), *Theory and Applications of Long-Range Dependence: Theory and Applications*, pp. 129–142. Birkhäuser, Boston.

[30] Taqqu, M. (2003) Fractional Brownian motion and long-range dependence In: Doukhan, P., Oppenheim, G., Taqqu, M.S. (Eds.), *Theory and Applications of Long-Range Dependence: Theory and Applications*, pp. 5–38. Birkhäuser, Boston.

[31] Weissman, I. (1976) On convergence of types and processes in Euclidean space, *Probability Theory and Related Fields* **37**(1) pp 35-41.

[32] Wood, A. and Chan, G. (1994) Simulation of stationary gaussian processes in $[0,1]^d$. *Journal of computational and graphical statistics*, **3**, 409–432.

[33] Zhao, O. and Woodroofe, M. (2008) On martingale approximations, *Annals of Applied Probability*, **18**, 1831-1847.

Deuxième partie

Inférence pour les processus ponctuels spatiaux de Gibbs

Chapitre 1

Processus ponctuels spatiaux de Gibbs

Sommaire

Dans ce chapitre, nous nous proposons de présenter quelques notations et modèles de processus ponctuels spatiaux de Gibbs. Ces modèles ont constitué les motivations premières pour proposer ou compléter des résultats en termes d'estimation et de validation.

1.1 Introduction

Un processus ponctuel est une collection aléatoire de points localement finie dans l'espace. Un processus ponctuel est dit marqué si à chaque point de l'espace est associée une marque aléatoire. Le domaine des processus ponctuels constitue un domaine à part entière à l'interface de la physique statistique, des probabilités et de la statistique. Les données en résultant interviennent dans de nombreuses applications (foresterie, écologie, biologie, économie, physique, physique statistique, informatique graphique,. . .). Nous renvoyons entre autres le lecteur aux ouvrages [95], [96] or [67] ainsi qu'au livre plus récent de Guyon et Gaetan [55].

Dans ce domaine, le modèle de référence est le processus ponctuel de Poisson qui permet de modéliser unen configuration de points aléatoire dans l'espace sans aucune interaction entre ces points. En particulier, deux sous-configurations vivant dans deux sous-domaines disjoints sont indépendantes. Une manière d'introduire de la dépendance est de considérer la classe des modèles de Gibbs. De manière assez simple, dans un domaine borné de $\mathbb{R}^d$, un processus de Gibbs est défini via sa mesure de probabilité dont la densité par rapport à la mesure de Poisson est proportionnelle à $e^{-V(\varphi)}$, où $V(\varphi)$ correspond à la fonction énergie (appelée également dans la littérature Hamiltonien ou fonction d'interaction) d'une configuration de points φ. L'extension de la définition d'un modèle de Gibbs dans $\mathbb{R}^d$ est un problème plus complexe mais essentiel pour plusieurs raisons : par exemple pour étudier les phénomènes de percolation, de transition de phase mais surtout (en ce qui nous concerne) pour étudier les propriétés asymptotiques d'estimateurs (paramétriques) de la distribution d'un modèle de Gibbs observé dans une fenêtre dont le volume est destiné à croître vers $+\infty$.

La classe des modèles de Gibbs est extrêmement riche et souple. Après avoir posé quelques définitions et notations et défini de manière concise une mesure de Gibbs, nous proposons quelques exemples montrant un peu la diversité des modèles. Précisons que nos différentes contributions

en termes d'inférence s'inscrivent dans le cadre des processus stationnaires. Nous définissons donc dans ce document les modèles de Gibbs dans ce cadre.

1.2 Modèles de Gibbs

1.2.1 Quelques notations

Une région de $\mathbb{R}^d$ sera typiquement notée Λ et sera supposée être un borélien de mesure de Lebesgue positive. On notera $\Lambda \Subset \mathbb{R}^d$ si Λ est borné, Λ^c désignera le complémentaire de Λ dans $\mathbb{R}^d$. La notation $|\cdot|$ sera utilisée sans ambiguité pour différents objets. Pour un ensemble dénombrable, $\mathcal{J}$, $|\mathcal{J}|$ représente son cardinal; pour $\Lambda \Subset \mathbb{R}^d$, $|\Lambda|$ est le volume de Λ; pour $x \in \mathbb{R}^d$, $|x|$ correspond à la norme uniforme tandis que $\|x\|$ représente sa norme euclidienne.

L'espace $\mathbb{R}^d$ est muni de la tribu des boréliens et de la mesure de Lebesgue λ. Soit $\mathbb{M}$ un espace mesurable destiné à être l'espace des marques, munie d'une σ-algèbre $\mathcal{M}$ et d'une mesure de probabilité λ^{m}. L'espace d'état du processus ponctuel est noté $\mathbb{S} := \mathbb{R}^d \times \mathbb{M}$ et est muni de la mesure $\mu := \lambda \otimes \lambda^{\mathrm{m}}$. Nous noterons pour simplifier $x^m = (x, m)$ un élément de $\mathbb{S}$, i.e. un point marqué.

Une configuration est un sous-ensemble φ de $\mathbb{S}$ localement fini au sens où $\varphi_\Lambda := \varphi \cap (\Lambda \times \mathbb{M})$ est de cardinalité finie $N_\Lambda(\varphi) := |\varphi_\Lambda|$ pour tout $\Lambda \Subset \mathbb{R}^d$. L'espace des configurations $\Omega = \Omega(\mathbb{S})$ est muni de la σ-algèbre $\mathcal{F}$ générée par les variables de comptage $N_\Lambda(\varphi)$ avec $\Lambda \Subset \mathbb{R}^d$. Enfin, soit $T = (\tau_x)_{x\in\mathbb{R}^d}$ l'opérateur de translation, où $\tau_x : \Omega \to \Omega$ est la translation du vecteur $-x \in \mathbb{R}^d$. Pour simplifier on notera $\varphi \cup x^m := \varphi \cup \{x^m\}$ and $\varphi \setminus x^m := \varphi \setminus \{x^m\}$.

1.2.2 Définition d'un modèle de Gibbs marqué stationnaire

Un processus ponctuel marqué Φ est une variable aléatoire sur Ω, de mesure de probabilité P sur $(\Omega, \mathcal{F})$. L'exemple le plus célèbre est le processus de Poisson marqué π^z de mesure intensité $z\lambda \otimes \lambda^{\mathrm{m}}$ sur $\mathbb{S}$, avec $z > 0$ (voir par exemple [89] pour quelques définitions et propriétés). Pour $\Lambda \Subset \mathbb{R}^d$, notons π^z_Λ la mesure de probabilité marginale dans Λ du processus de Poisson d'intensité z. Sans perte de généralité l'intensité est fixée à 1 et on notera alors plus simplement π et π_Λ à la place de π^1 et π^1_Λ.

Soit $\theta \in \mathbb{R}^p$ (pour $p \geq 1$). Pour tout $\Lambda \Subset \mathbb{R}^d$, considérons la fonction paramétrique $V_\Lambda(.;\theta)$ définie sur Ω et à valeurs dans $\mathbb{R} \cup \{+\infty\}$. D'un point de vue physique, $V_\Lambda(\varphi;\theta)$ est l'energie φ_Λ dans Λ étant donnée la configuration extérieure φ_{Λ^c}. Nous nous focalisons sur les processus ponctuels stationnaires sur $\mathbb{R}^d$, i.e. de mesure de probabilité stationnaire (i.e. T-invariant). De ce fait, nous supposerons que $V_\Lambda(.;\theta))$ est T-invariante, i.e. $V_\Lambda(\tau_x\varphi;\theta) = V_\Lambda(\varphi;\theta)$ pour tout $x \in \mathbb{R}^d$. La famille d'énergies $(V_\Lambda(.;\theta))_{\Lambda\Subset\mathbb{R}^d}$ est dite compatible si pour tout $\Lambda \subset \Lambda' \Subset \mathbb{R}^d$, il existe une fonction mesurable $\psi_{\Lambda,\Lambda'}$ de Ω dans $\mathbb{R} \cup \{+\infty\}$ telle que

$$\forall \varphi \in \Omega, \quad V_{\Lambda'}(\varphi;\theta) = V_\Lambda(\varphi;\theta) + \psi_{\Lambda,\Lambda'}(\varphi_{\Lambda^c};\theta). \tag{1.1}$$

Un processus marqué de Gibbs stationnaire est alors caractérisé par sa mesure de Gibbs marquée définie comme suit (voir [101]).

Définition 17 *Une mesure de probabilité P_θ sur Ω est une mesure de Gibbs marquée stationnaire pour la famille compatible et T-invariante d'énergies $(V_\Lambda(.;\theta))_{\Lambda\Subset\mathbb{R}^d}$ si pour tout $\Lambda \Subset \mathbb{R}^d$, pour P_θ-a.e. configuration extérieure φ_{Λ^c}, la distribution de P_θ conditionellement à φ_{Λ^c} admet la densité conditionelle par rapport à π_Λ suivante :*

$$f_\Lambda(\varphi_\Lambda|\varphi_{\Lambda^c};\theta) = \frac{1}{Z_\Lambda(\varphi_{\Lambda^c};\theta)} e^{-V_\Lambda(\varphi;\theta)},$$

où $Z_\Lambda(\varphi_{\Lambda^c};\theta)$ est la constante de normalisation appelée également fonction de partition.

L'existence de mesures de Gibbs sur Ω satisfaisant ces spécifications conditionelles est un problème difficile, voir entre autres [103, 101, 18, 35, 38].

Nous supposerons dans ce mémoire que la famille d'énergie est héréditaire signifiant que pour tout $\Lambda \Subset \mathbb{R}^d$, $\varphi \in \Omega$ et $x^m \in \Lambda \times \mathbb{M}$,

$$V_\Lambda(\varphi;\theta) = +\infty \Rightarrow V_\Lambda(\varphi \cup x^m;\theta) = +\infty, \tag{1.2}$$

ou de manière équivalente pour tout $x^m \in \varphi_\Lambda$, $f_\Lambda(\varphi_\Lambda|\varphi_{\Lambda^c};\theta) > 0 \Rightarrow f_\Lambda(\varphi_\Lambda \setminus x^m|\varphi_{\Lambda^c};\theta) > 0$. Le cas de processus non-héréditaires peut également être traité (voir par exemple [35, 38] pour les problèmes d'existence).

Ne se souciant pas de problèmes d'existence ou d'unicité, l'hypothèse minimale que nous avons faite dans le cadre de nos contributions sur l'estimation et la validation est la suivante :

> **[Mod]** : nos données correspondent à la réalisation d'un processus marqué Φ de mesure de Gibbs marquée $P_{\theta^\star}$, où $\theta^\star \in \mathring{\mathbf{\Theta}}$, $\mathbf{\Theta}$ est un compact de $\mathbb{R}^p$ et, pour tout $\theta \in \mathbf{\Theta}$, il existe une mesure de Gibbs marquée stationnaire P_θ pour la famille d'énergies $(V_\Lambda(.;\theta))_{\Lambda \Subset \mathbb{R}^d}$ supposée compatible, T-invariante et héréditaire.

[Mod] assure l'existence d'au moins une mesure de Gibbs stationnaire. Quand la mesure de Gibbs n'est pas unique, on parle de transition de phase. Dans cette situation, l'ensemble des mesures de Gibbs est un simplexe de Choquet et toute mesure de Gibbs est un mélange de mesures de Gibbs ergodiques extrémales. Si la mesure de Gibbs est unique, elle est nécessairement ergodique (cf [57] pour plus de détails).

La plupart des méthodes d'estimation et de validation (ainsi que nos différentes hypothèses et preuves) utilisent abondamment le concept d'énergie locale définie comme étant l'énergie requise pour insérer un point marqué x^m dans une configuration φ et exprimée pour tout $\Lambda \ni x$ par

$$V\left(x^m|\varphi;\theta\right) := V_\Lambda(\varphi \cup x^m;\theta) - V_\Lambda(\varphi;\theta). \tag{1.3}$$

La compatibilité de la famille d'énergies (1.1) permet d'assurer que cette définition ne dépende pas de Λ. Pour terminer, citons un résultat d'existence basé sur ce concept, tout à fait simple et très facile à vérifier pour de nombreux exemples classiques, établi dans [18]. L'existence d'une mesure de Gibbs est assurée dès lors que les deux hypothèses suivantes le sont : pour toute marque m, toute configuration φ finie et tout $\theta \in \mathbf{\Theta}$,

> **[LS]** Locale stabilité : $\exists\ 0 < K < +\infty$ (indépendante de θ, m, φ) tel que $V\left(0^m|\varphi;\theta\right) \geq -K$.
>
> **[FR]** Portée finie : $\exists\ 0 < D < +\infty$ (indépendante de θ, m, φ) tel que $V\left(0^m|\varphi;\theta\right) = V\left(0^m|\varphi_{\mathcal{B}(0,D)};\theta\right)$.

1.2.3 Formule de Georgii-Nguyen-Zessin

Nous souhaitons ici présenter un ingrédient important qui intervient dans la plupart des preuves et qui est à la base de méthodes d'estimation (méthode de Takacs-Fiksel) et de validation (basée sur le concept de résidus). Cet ingrédient est connu sous le nom de formule de Georgii-Nguyen-Zessin (GNZ pour simplifier).

Théorème 18 (Formule GNZ [56]) *Sous l'hypothèse **[Mod]**, pour toute fonction $h(\cdot,\cdot;\theta) : \mathbb{S} \times \Omega \to \mathbb{R}$ telle que les quantités suivantes soient finies, alors*

$$E\left(\int_{\mathbb{R}^d \times \mathbb{M}} h\left(x^m, \Phi;\theta\right) e^{-V\left(x^m|\Phi;\theta^\star\right)} \mu(dx^m)\right) = E\left(\sum_{x^m \in \Phi} h\left(x^m, \Phi \setminus x^m;\theta\right)\right), \tag{1.4}$$

où E représente l'espérance par rapport à $P_{\theta^\star}$.

Pour les processus ponctuels stationnaires marqués de Gibbs, l'équation (1.4) se réduit plus simplement à

$$E\left(h\left(0^M,\Phi;\theta\right)e^{-V\left(0^M|\Phi;\theta^\star\right)}\right)=E\left(h\left(0^M,\Phi\setminus 0^M;\theta\right)\right) \tag{1.5}$$

où dans la suite de ce mémoire M désignera une variable aléatoire de mesure de probabilié λ^{m}.

1.3 Quelques exemples

Pour simplifier, nous considérons des exemples dans le cas $d=2$. La plupart des exemples ont une généralisation évidente pour les dimensions supérieures. Ce paragraphe n'a aucune vocation à présenter une liste exhaustive d'exemples mais simplement à présenter quelques modèles montrant la richesse des modèles de Gibbs.

Modèle de Strauss marqué

Le modèle de Strauss est certainement l'exemple le plus célèbre vu comme la première alternative à un processus de Poisson homogène (marqué). Nous le définissons ici pour deux marques $\mathbb{M}=\{1,2\}$ (pour simplifier). La famille d'énergie (V_Λ) est définie pour $\Lambda\Subset\mathbb{R}^2$ par :

$$V_\Lambda(\varphi;\theta)=\sum_{m=1}^{2}\theta_1^m|\varphi_\Lambda^m|+\sum_{1\leq m\leq m'\leq 2}\theta_2^{m,m'}\sum_{\substack{\{x^m,y^{m'}\}\in\mathcal{P}_2(\varphi)\\ \{x^m,y^{m'}\}\cap\Lambda\neq\emptyset}}\mathbf{1}_{[0,D^{m,m'}]}(\|y-x\|),$$

où $D^{m,m'}>0$ et où $|\varphi_\Lambda^m|$ représente le nombre de points dans Λ de la marque m. Ce modèle existe dans le cas inhibition i.e. $\theta_1^m\in\mathbb{R}$ et $\theta_2^{m,m'}>0$ pour tout $m,m'=1,2$, ou sans restriction sur $\theta_2^{m,m'}$ mais en supposant une hypothèse de type hard-core, voir par exemple [95] pour plus de détails. De manière alternative, l'énergie locale de ce modèle s'écrit

$$V\left(x^m|\varphi;\theta\right):=\theta_1^m+\sum_{m'=1}^{2}\theta_2^{m,m'}\sum_{y^{m'}\in\varphi_\Lambda}\mathbf{1}_{[0,D^{m,m'}]}(\|y-x\|).$$

La Figure 1.1 illustre des réalisations de ce modèle utilisant le package `R ebspat` développé par R. Drouilhet.

Modèle multi-Strauss sur un graphe particulier

Le modèle de Multi-Strauss (dans le cas non marqué, $\mathbb{M}=\{0\}$, pour simplifier) généralise le précédent en remplaçant la fonction d'interaction de paires égale à une fonction indicatrice par une fonction en escalier et le graphe complet $\mathcal{P}_2(\varphi)$ par un graphe plus structuré noté par la suite $\mathcal{G}_2(\varphi)$.

$$V_\Lambda(\varphi;\theta)=\theta_1|\varphi_\Lambda|+\sum_{k=1}^{K}\theta_2^k\sum_{\substack{\{x,y\}\in\mathcal{G}_2(\varphi)\\ \{x,y\}\cap\Lambda\neq\emptyset}}\mathbf{1}_{[D_{k-1},D_k]}(\|y-x\|),$$

où $K\geq 1$, $0=D_0<D_1<\ldots<D_K$. Ce modèle est très classique dans le cas où $G_2(\varphi)=\mathcal{P}_2(\varphi)$ voir par exemple [8] ou à nouveau [95]. Basé sur une idée initiale de Baddeley et Møller [6], Bertin, Billiot et Drouilhet ont étudié les problèmes d'existence pour des modèles où le graphe est remplacé par le Graphe de Delaunay (i.e. le graphe des sommets associé à la triangulation de Delaunay) ou le graphe des k-plus proches voisins par exemple [18] et les références associées. Ces modèles existent

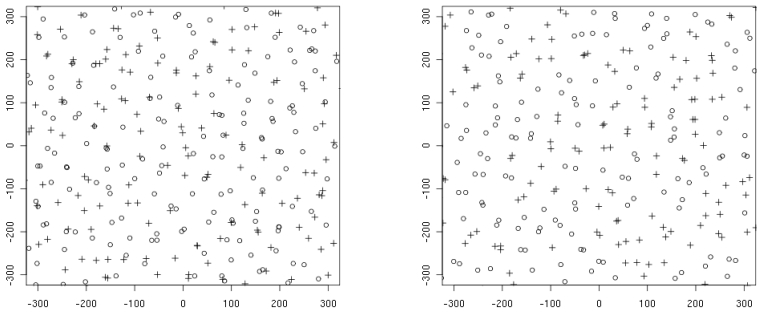


FIGURE 1.1 – Réalisations d'un modèle de Strauss marqué sur $[-300, 300]^2$. Les paramètres $D^{m,m'}$ sont tous fixés ici à $D = 40$ et $\theta_1^1 = \theta_1^2 = 2$. A gauche $\theta_2 = (2, 2, 2)$, à droite $(2, 2, 10)$.En particulier, on perçoit ici l'influence du paramètre $\theta_2^{1,2}$. Plus celui-ci est élevé, plus il est difficile d'obtenir des arêtes points-croix de longueur inférieure à D, ce qui crée des clusters de chaque marque.

sans aucune hypothèse sur les paramètres. Ceci a été montré dans [18] dans le cas des k-plus proches voisins et d'un graphe de Delaunay "légèrement modifié" et dans le travail plus récent et plus abouti [38] pour le graphe de Delaunay général. On notera que pour un graphe structuré, l'énergie locale ne s'écrit pas de manière aussi simple que pour le cas du graphe complet ; ceci étant dû au fait que pour $x \notin \varphi$ $Del_2(\varphi \cup x) \setminus Del_2(\varphi)$ ne correspond pas nécessairement à l'ensemble des arêtes partant de x (ce qui est le cas pour le graphe complet). Nous présentons dans la Figure 1.3 de telles réalisations pour le graphe complet et le graphe de Delaunay.

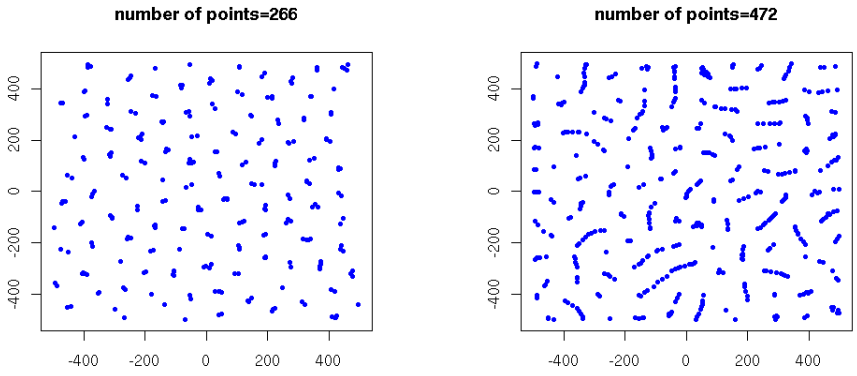


FIGURE 1.2 – Réalisations d'un modèle de multi-Strauss sur le graphe complet (à gauche) et sur le graphe de Delaunay (à droite) dans le domaine $[-400, 400]^2$, de paramètre $\theta = (1, 2, 4)$ et d'hyperparamètres $D_1 = 20, D_2 = 80$. Les paramètres font qu'il est relativement coûteux d'obtenir une réalisation ayant des arêtes de longueur entre 20 et 80. La structure de Delaunay force alors les points à se positionner le long de "fibres" tandis qu'elle crée plutôt des agrégats pour le graphe complet.

Interaction mesurée sur les cellules de Voronoï

Sur la même idée que précédemment d'autres modèles peuvent être obtenus en faisant interagir des structures plus complexes que des points. Par exemple, on peut faire interagir des triangles de Delaunay (au travers de la surface des triangles par exemple, du plus petit ou plus grand angle de chaque triangle). On peut également s'intéresser à définir des mosaïques de Voronoï en interaction. En biologie, depuis [65], l'utilisation des diagrammes de Voronoï est fréquente pour modéliser la configuration cellulaire d'un tissu. La modélisation est souvent réalisée en construisant des diagrammes de Voronoï dont les centres de cellules ont été générés par un processus ponctuel de Poisson. Dans différents travaux, la nécessité de contraindre le diagramme pour gagner en réalisme a été mise en évidence. Dans [48] (suivant un travail de Graner et Glazier [61]), l'interaction entre cellules épithéliales voisines est prise en compte au travers d'une énergie hamiltonienne. Cette dernière fait intervenir l'aire de chaque cellule de la mosaïque de Voronoï modélisant le tissu, mais aussi une interaction qui est fonction de la longueur de l'arête commune à deux cellules. Dans [46], la force d'adhésion entre deux cellules marquées par un biomarqueur d'un tissu est également modélisée à travers une énergie qui pénalise (entre autres) la longueur de la membrane, i.e. l'arête de Voronoï. A titre d'exemple, une version simplifiée de ce modèle sécrit énergétiquement

$$V_\Lambda(\varphi) = \theta_1 |\varphi_\Lambda| + \theta_2 \sum_{\substack{\{x,y\}\in Del_2(\varphi)\\ \{x,y\}\cap\Lambda\neq\emptyset}} |Vor(x)\cap Vor(y)|\, \mathbf{1}_{[0,D]}(||y-x||). \tag{1.6}$$

Notons que lorsque $D = +\infty$, le second terme correspond au périmètre de la mosaïque de Voronoï dans Λ. Ce type de modèles (faisant interagirir des cellules de Voronoï), a été étudié d'un point de vue existence dans [18, 38] et [35, 40]. Notons que dans ces derniers travaux des versions non-héréditaires très intéressantes sont proposées (pour interdire les cellules de Voronoï à être de surface plus grande qu'une valeur donnée par exemple).

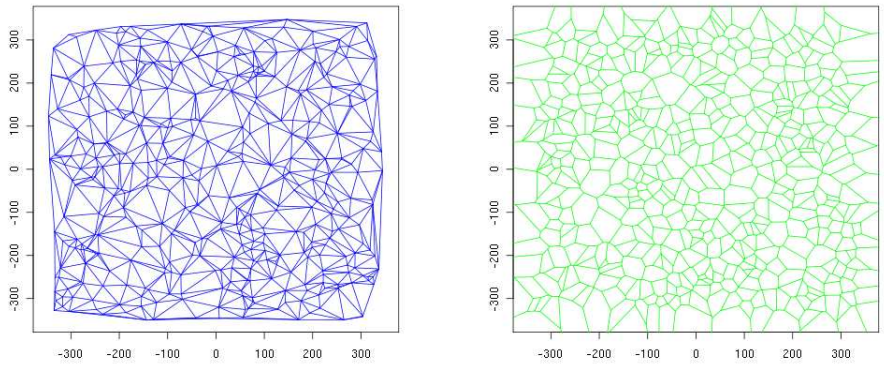


FIGURE 1.3 – Réalisation du modèle (1.6) dans le domaine $[-300, 300]^2$ avec $D = 80$ et $\theta = (7, 10)$ (réalisé avec `ebspat`). A gauche la triangulation de Delaunay résultante, à droite la mosaïque de Voronoï.

Modèle de Lennard-Jones

La plupart des modèles cités précédemment ont les caractéristiques **[LS]** (ou au moins une portée locale à la configuration, cf [38] pour une définition précise) et **[FR]** et appartiennent à

la famille exponentielle (i.e. la fonction énergie est linéaire en les paramètres). L'exemple le plus emblématique n'appartenant pas à cette classe de modèles est le modèle de Lennard-Jones prenant ses origines en physique statistique [81] pour modéliser des particules en interaction. Le processus ponctuel en résultant est un processus d'interaction de paires sur le graphe complet, non linéaire, non localement stable et de portée infinie. Définissons la famille d'énergies suivante pour $\Lambda \Subset \mathbb{R}^2$ par

$$V_\Lambda^{LJ}(\varphi;\theta) := \theta_1|\varphi_\Lambda| + H_\Lambda^{LJ}(\varphi;\theta) \text{ avec } H_\Lambda^{LJ}(\varphi;\theta) := \sum_{\substack{x_1\in\varphi_\Lambda \\ x_2\in\varphi_{\Lambda^c}}} g^{LJ}(||x_1-x_2||;\theta)$$

et

$$g^{LJ}(r;\theta) := 4\theta_2\left(\left(\frac{\theta_3}{r}\right)^{12} - \left(\frac{\theta_3}{r}\right)^{6}\right)\mathbf{1}_{[0,D]}(r).$$

où $\theta = (\theta_1,\theta_2,\theta_3) \in \mathbb{R}\times(\mathbb{R}^+)^2$. Lorsque $D = +\infty$, on obtient le modèle de Lennard-Jones. Lorsque $D < +\infty$ (restriction considérée pour montrer la normalité asymptotique de l'estimateur du maximum de pseudo-vraisemblance), on parle de modèle de Lennard-Jones à portée finie. Quelle que soit la valeur de D, Ruelle [103, 104] a montré l'existence d'une mesure de Gibbs (le résultat étant bien plus général car il prouve l'existence pour tous les potentiels superstables et réguliers inférieurement). Lorsque $D < +\infty$, ce résultat découle également du travail [38]. Voir Figure 1.4 pour quelques exemples.

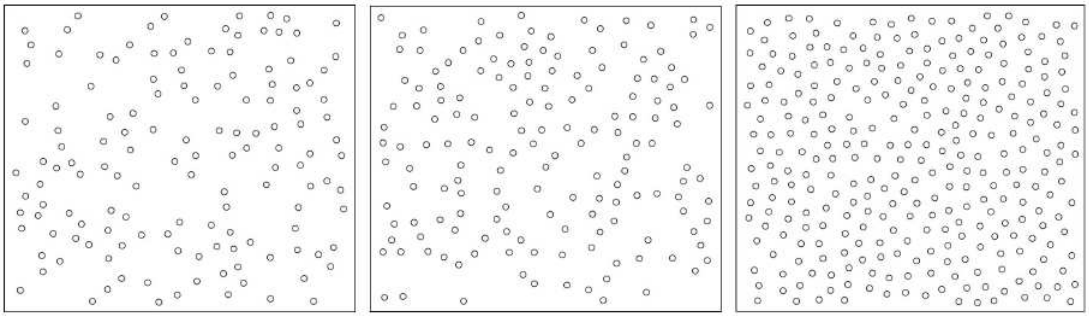


FIGURE 1.4 – Simulations du modèle de Lennard-Jones. L'intensité est constante pour ces trois réalisations et le paramètre θ_3 est fixé à 0.1. Le paramètre θ_2 vaut 0, à gauche, 0.1, au milieu, et 2, à droite. Ces figures réalisées avec le package `spatstat` sont tirées de [37].

Modèle Quermass

Le modèle Quermass a été introduit par [76]. Il constitue un processus ponctuel marqué destiné à modéliser des ensembles aléatoires dans $\mathbb{R}^2$. Il constitue une généralisation du modèle Booléen. Soit x^R un point marqué où x et $R > 0$ (i.e. la marque) représentent respectivement le centre et le rayon de la boule $\mathcal{B}(x,R)$. Pour une configuration finie φ, l'énergie est définie pour $(\theta_1,\theta_2,\theta_3,\theta_4) \in \mathbb{R}^4$ par :

$$V(\varphi;\theta) = \theta_1|\varphi| + \theta_2\,\mathcal{P}(\Gamma) + \theta_3\,\mathcal{A}(\Gamma) + \theta_4\,\mathcal{E}(\Gamma) \quad \text{where } \Gamma = \bigcup_{(x,R)\in\varphi}\mathcal{B}(x,R)$$

où $\mathcal{P}(\Gamma)$, $\mathcal{A}(\Gamma)$ et $\mathcal{E}(\Gamma)$ désignent respectivement le périmètre, la surface et la caractéristique d'Euler-Poincaré (i.e. le nombre de composantes connexes moins le nombre de trous) de l'ensemble Γ. L'extension de ce processus sur $\mathbb{R}^2$ nécessite de supposer que les rayons des boules sont

uniformément bornés sur $\mathbb{R}^2$. La famille d'énergies (V_Λ) est alors définie par

$$V_\Lambda(\varphi;\theta) = V\left(\varphi_{\Lambda\oplus B(0,2R_0)};\theta\right) - V\left(\varphi_{\Lambda\oplus B(0,2R_0)\setminus\Lambda};\theta\right).$$

Dereudre, dans [36], a montré l'existence de ce processus dans $\mathbb{R}^2$. Lorsque $\theta_2 = \theta_3 = \theta_4 = 0$, on retrouve le modèle Booléen (cf [112]) tandis que le processus connu sous le nom de "area process" (cf [10]) peut être obtenu en prenant $\theta_2 = \theta_4 = 0$. D'un point de vue inférentiel, la difficulté principale de ce modèle (et donc aussi l'intérêt !) tient du fait qu'en pratique on n'observe que la réalisation de l'ensemble aléatoire Γ. La position des graines et la valeur des marques sont totalement inconnues.

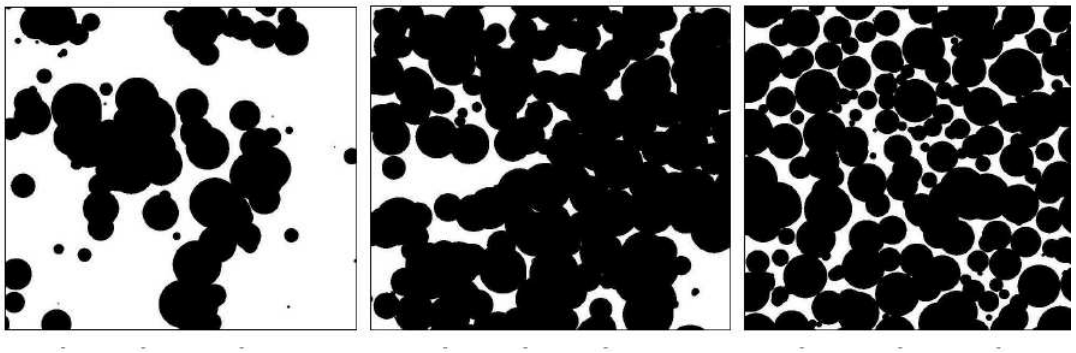


FIGURE 1.5 – Simulations du modèle Quermass sur la fenêtre $[0,30]^2$ pour une loi uniforme sur $[0,2]$ sur les rayons. Ces figures sont tirées de [37], elles-mêmes provenant de [94]. L'intensité est constante et les paramètres $(\theta_2,\theta_3,\theta_4)$ valent respectivement $(0,0.2,0)$ (à gauche), $(0,0,1)$ (au milieu) et $(-1,-1,0)$ (à droite).

Chapitre 2

Problèmes d'inférence considérés

Sommaire

Ce chapitre a pour objectif de résumer succinctement nos contributions d'ordre inférentiel au domaine de la statistique sur les processus ponctuels. La Section 2.1 discute des problèmes d'estimation tandis que la Section 2.2 décrit notre apport sur l'analyse des résidus et la description de tests d'adéquation.

De nombreux travaux se sont attachés à estimer la fonction énergie à partir d'une configuration de points généré par un procesus ponctuel marqué. Dans nos différents travaux, nous nous sommes intéressés à l'estimation paramétrique de la fonction énergie. Des approches non paramétriques existent et nous renvoyons le lecteur intéressé à [95]. En estimation paramétrique, l'approche la plus commune consiste à utiliser la fonction de vraisemblance. Le problème de cette méthode est qu'elle nécessite le calcul (ou l'approximation) de la constante de normalisation (qui dépend de θ). De récents développements en statistiques computationelles permettent néanmoins une utilisation relativement efficace de cette méthode (voir [17]). Sur un plan théorique, l'estimateur du maximum de vraisemblance souffre d'un manque de justifications. Seul [86] propose des résultats pour des configurations "sparse". Une autre idée dont l'origine remonte au travail [19] consiste à utiliser la fonction de pseudo-vraisemblance (basée sur la notion de densité conditionelle) qui permet d'éviter le calcul de la constante de normalisation (voir Section 2.1.1). Une autre approche consiste à s'appuyer sur la formule GNZ (1.5) pour proposer toute une classe d'estimateurs par minimum de contraste basé sur des versions empiriques des termes de gauche et droit de la formule (1.5). Cette méthode porte le nom de méthode de Takacs et en particulier englobe la méthode du maximum de pseudo-vraisemblance. Elle est présentée et étudiée dans la Section 2.1.2.

Une fois le modèle estimé, la question naturelle concerne la validation de modèles paramétriques. De ce point de vue la littérature est nettement moins importante. Très récemment, Baddeley et al. [9] ont proposé d'étendre la notion de résidus pour des processus ponctuels en dimension 1 ([53] ou [4]) aux processus spatiaux. En dimension 1, une mesure souvent utilisé consiste à construire la différence entre le nombre d'événements apparus dans un intervalle $[0, t]$ et le taux de hasard estimé paramétriquement et intégré de 0 à t. Le pendant naturel de ces deux quantités pour des processus ponctuels spatiaux est le nombre de points dans un domaine et l'intensité conditionelle de Papangelou (i.e. $e^{-V(x|\varphi;\widehat{\theta})}$) intégrée sur Λ. On s'aperçoit assez rapidement qu'à nouveau la formule (1.5) justifie cette mesure puisque, si $\widehat{\theta} = \theta^\star$, l'espérance de la différence précédente est nulle. Cette remarque a conduit les auteurs de [9] à proposer toute une série de mesures de diagnostics basées sur des fonctions tests h différentes. Ils proposent également une série de graphiques permettant une aide au diagnostic et une implémentation importante au sein du package `R spatstat`. Dans [7], Baddeley, Møller et Pakes fournissent quelques propriétés des mesures proposées et conjecturent entre autres que lorsque la fenêtre devient de plus en plus large, une loi des grands nombres et un

théorème central limite doivent pouvoir être obtenus. La Section 2.2 présente notre contribution en relation avec ce sujet.

2.1 Estimation paramétrique d'un modèle de Gibbs

2.1.1 Estimation par maximum de pseudo-vraisemblance

Suivant une idée de Besag [19], développée sur la grille, Jensen et Møller ont montré dans [73] que le contraste obtenu à partir des densités conditionnelles (calculé sur un domaine découpé en sous-cubes dont les domaines convergent vers l'infini) s'écrit pour une configuration φ et un domaine d'observation Λ

$$PL_\Lambda(\varphi;\theta) := \exp\left(-\int_{\Lambda\times\mathbb{M}} e^{-V(x^m|\varphi;\theta)}\mu(dx^m)\right)\prod_{x^m\in\varphi_\Lambda} e^{-V(x^m|\varphi\setminus x^m;\theta)}. \tag{2.1}$$

Cette fonction de contraste est appelée fonction de pseudo-vraisemblance. De manière évidente, on définit la fonction de log-pseudo-vraisemblance par

$$LPL_\Lambda(\varphi;\theta) = -\int_{\Lambda\times\mathbb{M}} e^{-V(x^m|\varphi;\theta)}\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda} V(x^m|\varphi\setminus x^m;\theta). \tag{2.2}$$

La (log-)pseudo-vraisemblance est un critère "calculable" pour toute valeur de θ. En effet, le premier terme se réduit au calcul d'une intégrale sur $\mathbb{R}^d$ qui peut être approché par une discrétisation tandis que le second est calculé de manière exacte puisqu'il ne dépend que des points de la configuration.

Dans le cadre de cette méthode, nous avons réalisé deux études complémentaires associées aux références [BCD08] et [CD10]. Ces résultats étendent les travaux existants sur la question de propriétés asymptotiques ([73], [72], [87, 88]). Nos contributions ne couvrent pas les modèles de Gibbs définis par des familles d'énergie non héréditaires. Cet aspect a été entrepris à partir de nos travaux dans [39].

Travail associé à [BCD08]

En collaboration avec Jean-Michel Billiot et Rémy Drouilhet, nous avons considéré des densités héréditaires issues de la famille exponentielle de modèles de Gibbs marqués (i.e. de telle sorte que la fonction énergie s'écrit linéairement en fonction des paramètres) et supposé la portée du processus ainsi que la stabilité de la fonction énergie locale (hypothèses notées dans le Chapitre précédent **[FR]** et **[LS]**). Notons $V(x^m|\varphi;\theta) = \theta'\mathbf{V}(x^m|\varphi)$ avec $\mathbf{V}(x^m|\varphi) = (V_1(x^m|\varphi),\dots,V_p(x^m|\varphi))'$ et désignons par **[Int]** l'hypothèse d'intégrabilité suivante

[Int] pour $i=1,\dots,p$, $\exists \kappa_i^{(sup)} \geq 0$, $k_i\in\mathbb{N}$: $\forall (m,\varphi)\in\mathbb{M}\times\Omega$, $V_i(0^m|\varphi) \leq \kappa_i^{(sup)}\,|\varphi_{\mathcal{B}(0,D)}|^{k_i}$.

Supposons que notre processus est observé sur $\Lambda_n \oplus D$, où D correspond à la portée et où, pour simplifier, $\Lambda_n = [-n,n]^d$. Nous avons établi le résultat suivant.

Proposition 19 *Sous les hypothèses **[Mod]** et **[Int]**,*
(i) sous l'hypothèse d'identifiabilité suivante
[Id-MPLE] $\forall\theta\neq\theta^\star,\quad P\left(V\left(0^M|\Phi;\theta\right)\neq V\left(0^M|\Phi;\theta^\star\right)\right)>0,$
l'estimateur du maximum de pseudo-vraisemblance, noté $\widehat{\theta}_{MPLE}(\Phi)$ converge presque sûrement vers $\theta^\star$.
(ii) Si $P_{\theta^\star}$ est ergodique, il existe une matrice symétrique positive notée $\mathbf{\Sigma}_{MPLE}$ telle que $|\Lambda_n|^{1/2}\left(\widehat{\theta}_{MPLE}(\Phi)-\theta^\star\right)\xrightarrow{d}\mathcal{N}(0,\mathbf{\Sigma}_{MPLE})$.
(iii) Sans supposer que $P_{\theta^\star}$ soit ergodique mais en supposant quelques conditions sur $\mathbf{\Sigma}_{MPLE}$, on peut définir un estimateur consistant $\widehat{\mathbf{\Sigma}}_{MPLE}^{-1/2}$ de $\mathbf{\Sigma}_{MPLE}^{-1/2}$ et ainsi obtenir un TCL normalisé.

La preuve de ce résultat réside essentiellement dans l'application conjointe de résultats sur les estimateurs de minimum de contraste (*e.g.* [63]) et sur un théorème ergodique établi par Nguyen et Zessin [97] pour (i) et sur un théoremè central limite établi par [72]. Notons que ce TCL n'est pas écrit sous des conditions de mélange mais pour des fonctionnelles de champs markoviens possédant une propriété de centrage conditionnel, ce qui permet (comme le montre (iii)) d'obtenir des résultats même en transition de phase. Enfin soulignons que dans [BCD08] nous donnons un critère pratique pour vérifier l'hypothèse **[Id-MPLE]**.

Travail associé à [CD10]

La Proposition 19 est extrêmement intéressante car elle est simple dans les hypothèses qu'elle contient et permet d'englober de nombreux modèles classiques (cf [8]) comme le modèle de Strauss marqué ou les multi-Strauss sur graphe complet et plus structuré. Les hypothèses **[Mod]** et **[Int]** sont également vérifiées pour de nombreux modèles avec interactions sur les triangles de Delaunay ou cellules de Voronoï. Ces modèles ne pouvaient pas être pris en compte par les résultats existants. En collaboration avec Rémy Drouilhet, nous nous sommes attachés, dans [CD10], à étendre ce précédent travail pour pouvoir inclure des modèles plus complexes tels que les modèles de Lennard-Jones. Rappelons que ce modèle héréditaire n'est plus localement stable, de portée infinie et la fonction énergie n'est plus linéaire en les paramètres (voir Section 1.3 pour plus de détails).

Sans rentrer dans les détails, précisons que nous avons établi sous la condition **[Mod]** (existence) et sous des conditions de régularité et d'intégrabilité de la fonction énergie, un résultat semblable à celui de la Proposition 19. Ces hypothèses sont plus complexes puisqu'elles s'expriment dans un cadre plus général, néanmoins nous avons montré (et c'est ici que réside la vraie originalité de ce travail à notre sens) qu'elle pouvaient être vérifiées pour certains modèles superstables et réguliers inférieurement (au sens de Ruelle [103, 104]) tels que le modèle de Lennard-Jones. En particulier, nous montrons que quelle que soit la valeur de D (définissant le modèle), finie ou non, l'estimateur converge presque sûrement vers $\theta^\star$. En revanche, un TCL (normalisé) n'est obtenu qu'à la condition $D < +\infty$. Ce travail vient compléter sur un plan théorique le travail de Goulard et *al.* [60] qui avaient proposé de modéliser des données issues de la foresterie par un modèle de Lennard-Jones et estimé ses paramètres par maximisation de la pseudo-vraisemblance.

Application en informatique graphique

Dans [HLT+09], en collaboration avec Thomas Hurtut, Pierre-Eric Landes, Joëlle Thollot (membre de l'équipe Artis), Yann Gousseau et Rémy Drouilhet, nous avons appliqué la modélisation via des processus ponctuels en informatique graphique. Le problème est un problème de rendu expressif. Un utilisateur dessine dans une fenêtre (relativement petite) un certain nombre d'objets géométriques et l'idée est de capturer de l'information sur ces différents motifs ainsi que leur arrangement spatial et de proposer un dessin sur une fenêtre plus large. Ce travail a été réalisé principalement en trois étapes (comme le montre la Figure 2.1). Tout d'abord les objets sont catégorisés et une série de descripteurs géométriques est obtenue (intensité du trait, orientation). Ensuite, nous avons modélisé les centres de ces objets par un processus ponctuel marqué (il s'agissait d'un modèle de multi-Strauss marqué sur le graphe complet), estimé les paramètres de ces modèles par maximisation de la pseudo-vraisemblance et simulé le modèle estimé sur une fenêtre plus large.

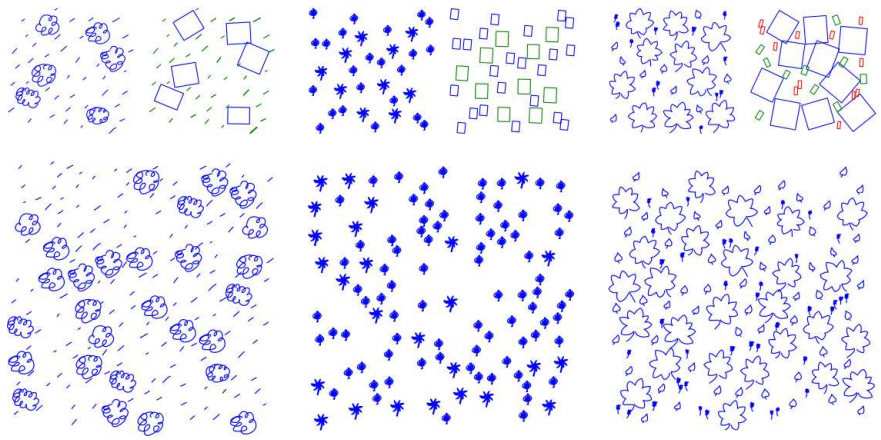


FIGURE 2.1 – Trois exemples d'application de la procédure envisagée : (a) En haut à gauche, le motif initial. (b) En haut à droite, l'étape de catégorisation. (c) En bas, simulation du modèle estimé.

2.1.2 Méthode de Takacs-Fiksel

Comme souligné en introduction de ce chapitre, la méthode de Takacs-Fiksel dérive de la formule (1.5). Soit $h(\cdot,\cdot;\theta) : \mathbb{S} \times \Omega \to \mathbb{R}$ et définissons pour tout $\varphi \in \Omega$, $\theta \in \boldsymbol{\Theta}$ et $\Lambda \Subset \mathbb{R}^d$

$$C_\Lambda(\varphi; h, \theta) := \int_{\Lambda \times \mathbb{M}} h(x^m, \varphi; \theta) e^{-V(x^m|\varphi;\theta)} \mu(dx^m) - \sum_{x^m \in \varphi_\Lambda} h(x^m, \varphi \setminus x^m; \theta) \tag{2.3}$$

Pour des choix appropriés de fonctionnelles h et un domaine Λ_n (régulier destiné à converger vers $\mathbb{R}^d$), il est possible d'appliquer un théorème ergodique et prouver que les deux termes de $|\Lambda_n|^{-1} C_{\Lambda_n}(\Phi; h, \theta)$ convergent respectivement vers les termes de gauche et droite de la formule (1.5). Ce principe définit la méthode de Takacs-Fiksel [49, 113, 50]. Donnons-nous K fonctions tests $h_k(\cdot,\cdot;\theta) : \mathbb{S} \times \Omega \to \mathbb{R}$ (pour $k = 1, \ldots, K$), l'estimateur en question est alors défini par

$$\widehat{\theta}_{TF}(\varphi) := \arg\min_{\theta \in \boldsymbol{\Theta}} \sum_{k=1}^{K} C_{\Lambda_n}(\varphi; h_k, \theta)^2. \tag{2.4}$$

En collaboration avec David Dereudre, Rémy Drouilhet et Frédéric Lavancier, nous nous sommes intéressés à cette méthode en étendant les travaux connus jusqu'alors [64, 21].

La méthode du maximum de pseudo-vraisemblance est un cas particulier puisque l'on peut montrer que lorsqu'on choisit $K = p$ et $h_k = \mathbf{V}_k^{(1)}$ (i.e. la dérivée par rapport à θ_k de la fonction énergie) les deux contrastes possèdent le même (unique) minimum.

Une première motivation pour étudier cette méthode est que des choix appropriés de fonctionnelles peuvent résoudre des problèmes numériques parfois compliqués et permettre d'obtenir rapidement une première estimation (qui peut, pourquoi pas, servir de point d'initialisation à une méthode plus évoluée). En effet, prenons, par exemple, la fonction $h_k(x^m, \varphi; \theta) := \mathbf{1}_{\mathcal{B}(0,r_k)}(\|x\|) e^{V(x^m|\varphi;\theta)}$ pour un $r_k > 0$. Si le processus n'est pas à noyau dur, le premier terme intégral se réduit à $|\mathcal{B}(0, r_k)|$ (pas besoin donc de discrétiser l'intégrale) tandis que le second terme ne va dépendre que des points de la configuration et peut être calculé très rapidement. Une deuxième motivation est liée au modèle Quermass. Rappelons que pour ce modèle, la fonction énergie dépend du nombre de points (non observés) et de caractéristiques géométriques de l'ensembe aléatoire formé par la réunion de boules. Møller et Helisova [94] ont proposé une approche MLE pour ce modèle en supposant que θ_1 soit connu. La pseudo-vraisemblance ne peut être appliquée ici car elle requiert la connaissance de

la position des points (au travers du terme de la somme). La méthode de Takacs-Fiksel permet de pouvoir estimer θ_1 de manière tout à fait astucieuse. En effet, il est possible de choisir une fonction test h telle que les deux termes de (2.3) soient calculables. Considérons la fonction test suivante

$$h_{per}(x^R, \varphi; \theta) = \mathcal{P}\left(\mathcal{C}(x, R) \cap \Gamma^c\right), \tag{2.5}$$

où $\mathcal{C}(x, R)$ est la sphère $\{y, |y - x| = R\}$. Pour toute configuration finie φ, on a

$$\sum_{x^R \in \varphi} h_{per}(x^R, \varphi \setminus x^R; \theta) = \mathcal{P}(\Gamma).$$

Ainsi le terme de la somme est calculable bien qu'aucun terme $h_{per}(x^R, \varphi \setminus x^R; \theta)$ ne le soit ! Le terme intégral ne pose quant à lui aucun problème puisqu'il met en jeu uniquement des quantités observables. Ainsi, si on suppose $(\theta_2, \theta_3, \theta_4)$ connus, θ_1 peut-être estimé par la méthode de Takacs-Fiksel. Autrement dit, la méthode de Takacs-Fiksel peut résoudre des problèmes d'inférence lorsque l'on n'observe pas les points. Nous renvoyons à notre papier soumis [CDDL10] pour d'autres exemples de fonctionnelles dans un cadre plus général.

Motivés par ces précédents points, nous avons dans [CDDL10] donné des conditions de régularité et d'intégrabilité très générales sur les fonctions tests (pouvant éventuellement dépendre de θ) et sur la fonction énergie locale (pas forcément linéaire en les paramètres) de telle sorte que sous l'hypothèse d'identifiabilité suivante

[Id-TF] $\sum_{k=1}^{K} E\left(h_k(0^M, \Phi; \theta)\left(e^{-V(0^M|\Phi;\theta)} - e^{-V(0^M|\Phi;\theta^\star)}\right)\right)^2 = 0 \quad \Longrightarrow \theta = \theta^\star,$

un résultat semblable à celui de la Proposition 19 puisse être obtenu. L'obtention de ce résultat utilise quasiment les mêmes outils que pour la preuve de la Proposition 19, à savoir théorème ergodique, estimateur de minimum de contraste et TCL. Précisons simplement que pour ce dernier, nous avons utilisé une version améliorée Théorème 23 qui nous permet de supposer uniquement que le domaine est un cube dont le volume tend vers l'infini. Comme nous le montrons, ce travail permet d'inclure de nombreux modèles et un large choix de fonctions tests.

Par ailleurs, et c'est là qu'à notre avis réside l'originalité de ce travail, nous discutons assez longuement l'hypothèse **[Id-TF]**. A la différence du contraste obtenu par pseudo-vraisemblance, le contraste obtenu par cette méthode n'est pas nécéssairement concave. Dans le cas, d'un modèle de Strauss nous illustrons ceci en Section 5.1 de [CDDL10]. Pour deux fonctionnelles particulières, nous montrons qu'il y a deux minima locaux. La question d'identifiabilité (et donc l'hypothèse **[Id-TF]**) est cruciale. L'hypothèse **[Id-TF]** n'est donc pas si simple à vérifier théoriquement. Dans [CDDL10], en se focalisant sur la famille exponentielle (i.e. $V(x^m|\varphi;\theta) = \theta' \mathbf{V}(x^m|\varphi)$), nous fournissons un critère plus pratique. Notons P_V la loi de $\mathbf{V}(0^M, \Phi)$ dans $\mathbb{R}^p$ et définissons la fonction Ψ_θ, pour tout $\theta \in \mathbf{\Theta}$ par

$$\begin{array}{rccc} \Psi_\theta : & \mathbb{R}^p & \longrightarrow & \mathbb{R}^K \\ & \mathbf{v} & \longmapsto & \begin{pmatrix} E\Big(h_1(0^M, \Phi; \theta)\Big|\mathbf{V}(0^M|\Phi) = \mathbf{v}\Big) \\ \vdots \\ E\Big(h_K(0^M, \Phi; \theta)\Big|\mathbf{V}(0^M|\Phi) = \mathbf{v}\Big) \end{pmatrix}. \end{array} \tag{2.6}$$

Proposition 20 *Si $K = p$ l'hypothèse suivante*

[Det($\neq$)] *Pour tout $\theta \in \mathbf{\Theta}$, $det(\mathbf{v}_1, \ldots, \mathbf{v}_p) det\big(\Psi_\theta(\mathbf{v}_1), \ldots, \Psi_\theta(\mathbf{v}_p)\big)$ n'est pas $(P_V)^{\otimes p}$-p.s. identiqement nul*

[Det($\geq$)] *Pour tout $\theta \in \mathbf{\Theta}$, il existe $\varepsilon = \pm 1$ tel que pour $(P_V)^{\otimes p}$-p.p. $(\mathbf{v}_1, \ldots, \mathbf{v}_p) \in (\mathbb{R}^p)^p$*

$$\varepsilon\, det(\mathbf{v}_1, \ldots, \mathbf{v}_p) det\big(\Psi_\theta(\mathbf{v}_1), \ldots, \Psi_\theta(\mathbf{v}_p)\big) \geq 0.$$

assure que la condition ***[Id-TF]*** *est satisfaite.*

Lorsque $\varepsilon = 1$ (respectivement $\varepsilon = -1$), **[Det(≥)]** signifie que Ψ_θ préserve le signe (resp. le signe opposé) du déterminant. Nous discutons quelques applications de ce résultat sur le modèle de Strauss et le modèle Quermass et donnons également une version de ce résultat lorsque le nombre K de fonctionnelles est supérieur strictement à p. Enfin, soulignons qu'en se basant sur le travail de Dereudre et Lavancier [39] proposant une alternative à la formule (1.5) dans le cas non héréditaire, nous avons étendu la méthode (et les résultats) pour inclure de tels modèles.

2.2 Validation de modèles par une analyse des résidus

Commençons par définir la notion de résidus proposée par [9]. Pour tout $\Lambda \Subset \mathbb{R}^d$, définissons les h-innovations (notés I_Λ) et les h-résidus (notés R_Λ et dépendant d'un estimateur $\widehat{\theta}$ de $\theta^\star$) par

$$\begin{aligned} I_\Lambda(\varphi; h, \theta^\star) &:= \int_{\Lambda\times\mathbb{M}} h(x^m, \varphi; \theta^\star)\, e^{-V(x^m|\varphi;\theta^\star)}\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda} h(x^m, \varphi\setminus x^m; \theta^\star) \\ R_\Lambda\left(\varphi; h, \widehat{\theta}\right) &:= \int_{\Lambda\times\mathbb{M}} h\left(x^m, \varphi; \widehat{\theta}\right) e^{-V\left(x^m|\varphi;\widehat{\theta}\right)}\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda} h\left(x^m, \varphi\setminus x^m; \widehat{\theta}\right). \end{aligned}$$

Sur un plan pratique, la dernière notion est évidemment la plus intéressante car elle fournit une mesure calculable. Les principaux exemples considérés par Baddeley et al. dans [9] (dans le cadre des processus ponctuels stationnaires) sont donnés par $h(x^m, \varphi; \theta) = 1$ pour les résidus bruts, $h(x^m, \varphi; \theta) = e^{V(x^m|\varphi;\theta)}$ pour les résidus inverses et $h(x^m, \varphi; \theta) = e^{V(x^m|\varphi;\theta)/2}$ pour les résidus de Pearson. Un autre exemple plus évolué consiste à considérer la fonction $h_r(x^m, \varphi; \theta) := \mathbf{1}_{[0,r]}(d(x^m, \varphi))\, e^{V(x^m|\varphi;\theta)}$, où $d(x^m, \varphi) = \inf_{y^m\in\varphi} \|y - x\|$. Ce choix conduit à

$$R_\Lambda(\varphi; h_r, \widehat{\theta}) = \int_{\Lambda\times\mathbb{M}} \mathbf{1}_{[0,r]}(d(x^m, \varphi))\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda} h_r(x^m, \varphi\setminus x^m; \widehat{\theta}). \tag{2.7}$$

Pour une large fenêtre $R(\varphi; h_h, \widehat{\theta})/|\Lambda|$ peut alors être vu comme une différence de deux estimateurs (l'un paramétrique et l'autre non paramétrique) de la fonction de vide F à la distance r. Rappelons que pour un processus marqué stationnaire (cf [95]), cette fonction est définie par $F(r) := P\left(d(0^M, \Phi) \le r\right)$.

En collaboration avec Frédéric Lavancier [CL10], nous avons complété significativement les travaux de [9] et [7], en proposant des résultats asymptotiques pour le processus des résidus et en dérivant des tests d'adéquation. En particulier, nous avons considéré deux cadres de travail différents et assez naturels :

- Contexte 1 : pour une fonction test h fixée, Λ_n est supposé être un cube divisé en un ensemble fini, $\mathcal{J}$, de sous-cubes (qui vont croître avec Λ_n), $\Lambda_n = \cup_{j\in\mathcal{J}}\Lambda_{j,n}$. On définit alors le vecteur $\mathbf{R}_1(\varphi; h) = \left(R_{\Lambda_{j,n}}(\varphi; h, \widehat{\theta})\right)_{j\in\mathcal{J}}$ constitué des h-résidus calculés sur chaque sous-cube.
- Contexte 2 : considérons $h_1, \ldots, h_s$, s différentes fonctions tests et définissons $\mathbf{R}_2(\varphi; \mathbf{h}) = \left(R_{\Lambda_n}(\varphi; h_j, \widehat{\theta})\right)_{j=1,\ldots,s}$ constitué des h_j-résidus calculés sur Λ_n.

Dans ces deux contextes, un estimateur de $\theta^\star$ est à utiliser. Nous avons supposé de manière assez naturelle que cet estimateur est calculé sur le même domaine initial Λ_n et avec les mêmes données (i.e. φ) qui sont utilisés pour le calcul des résidus. Le premier contexte est inspiré du test dit test des quadrats pour tester un Poisson homogène tandis que le second est inspiré par l'exemple de fonctions tests associées à la fonction de vide. Ils sont résumés dans la Figure 2.2.

En plus de l'hypothèse d'existence **[Mod]** et la portée du processus (hypothèse **[FR]**), nous avons considéré dans [CL10], des hypothèses d'intégrabilité et de régularité sur le modèle (défini

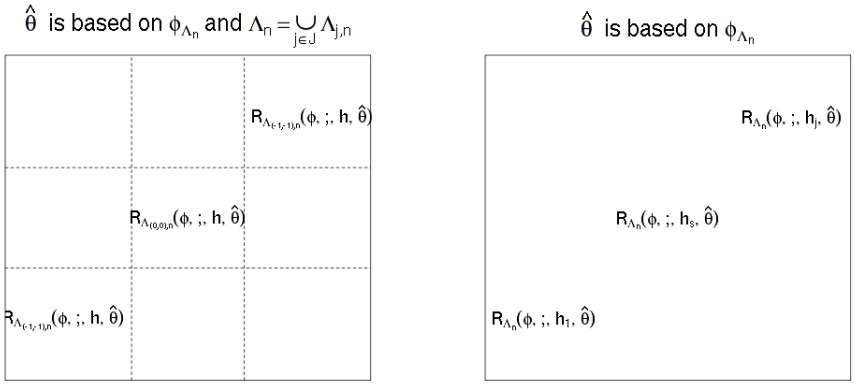


FIGURE 2.2 – Les deux contextes vus graphiquement.

par sa fonction énergie), sur la ou les fonctions tests et sur la nature de l'estimateur considéré (en gros consistance et normalité asymptotique de cet estimateur). Basés sur ces jeux d'hypothèses, nous avons alors pu obtenir le résultat suivant.

Proposition 21 *(i) Lorsque* $n \to +\infty$, $\mathbf{R}_1(\Phi;h)/|\Lambda_{0,n}|$ *et* $\mathbf{R}_2(\Phi;\mathbf{h})/|\Lambda_n|$ *convergent presque sûrement vers zéro.*
(ii) Si $P_{\theta^\star}$ *est ergodique, il existe* $\underline{\mathbf{\Sigma}}_1$ *et* $\underline{\mathbf{\Sigma}}_2$ *telles que* $|\Lambda_{0,n}|^{-1/2}\mathbf{R}_1(\Phi;h) \xrightarrow{d} \mathcal{N}(0,\underline{\mathbf{\Sigma}}_1)$ *et* $|\Lambda_n|^{-1/2}\mathbf{R}_2(\Phi;\mathbf{h}) \xrightarrow{d} \mathcal{N}(0,\underline{\mathbf{\Sigma}}_2)$.
(iii) Sans supposer que $P_{\theta^\star}$ *soit ergodique mais en supposant quelques conditions sur* $\underline{\mathbf{\Sigma}}_k$ *pour* $k=1,2$, *on peut définir un estimateur consistant* $\widehat{\underline{\mathbf{\Sigma}}}_k^{-1/2}$ *de* $\underline{\mathbf{\Sigma}}_k^{-1/2}$ *et ainsi obtenir un TCL normalisé.*

Nous renvoyons le lecteur aux Section 4 et 6 de [CL10] pour une discussion sur la caractéristique définie positive des matrices de covariances asymptotiques ainsi que la forme explicite de leurs estimations. Une conséquence du dernier résultat est que

$$|\Lambda_{0,n}|^{-1}\|\widehat{\underline{\mathbf{\Sigma}}}_1^{-1/2}\mathbf{R}_1(\Phi;h)\|^2 \xrightarrow{d} \chi^2_{|\mathcal{J}|} \quad \text{et} \quad |\Lambda_n|^{-1}\|\widehat{\underline{\mathbf{\Sigma}}}_2^{-1/2}\mathbf{R}_2(\Phi;\mathbf{h})\|^2 \xrightarrow{d} \chi^2_s. \tag{2.8}$$

Au contraire de la matrice $\underline{\mathbf{\Sigma}}_2$, la matrice $\underline{\mathbf{\Sigma}}_1$ possède une forme extrêmement particulière. En effet, elle peut s'écrire sous la forme $\lambda_{Inn}\,\underline{\mathbf{I}}_{|\mathcal{J}|} + |\mathcal{J}|^{-1}(\lambda_{Res}-\lambda_{Inn})\,\underline{\mathbf{J}}$ avec $\underline{\mathbf{J}} = \mathbf{e}\mathbf{e}'$ et $\mathbf{e} = (1,\ldots,1)'$. Les constantes λ_{Inn} et λ_{Res} dépendent de $P_{\theta^\star}$ (au travers de covariances), de $\theta^\star$ et de la fonction test considérée. Seule λ_{Res} dépend en plus de la nature de l'estimateur $\widehat{\theta}$ considéré.

Ceci nous amène à faire deux remarques. Premièrement, on peut montrer qu'on obtient en fonction de λ_{Inn} et λ_{Res} un forme explicite de $\underline{\mathbf{\Sigma}}_1^{-1/2}$ et donc de son estimation. Ensuite, cette décomposition de $\underline{\mathbf{\Sigma}}_1$ nous a incités à étudier le vecteur des résidus centrés pour le Contexte 1, et nous sommes parvenus au résultat suivant en notant $\overline{\mathbf{R}}_1(\varphi;h) = |\mathcal{J}|^{-1}\sum_{j\in\mathcal{J}} R_{\Lambda_{j,n}}\left(\varphi;h,\widehat{\theta}_n\right)$.

Corollaire 22 *Sous les hypothèses de la Proposition 21, en notant* $\widehat{\lambda}_{Inn}$ *un estimateur consistant de* λ_{Inn} *(supposé* >0*) alors lorsque* $n \to +\infty$

$$|\Lambda_{0,n}|^{-1}\widehat{\lambda}_{Inn}^{-1}\|\mathbf{R}_1(\Phi;h) - \overline{\mathbf{R}}_1(\Phi;h)\|^2 \xrightarrow{d} \chi^2_{|\mathcal{J}|-1}. \tag{2.9}$$

Les résultats (2.8) et (2.9) sont intéressants puisqu'ils nous permettent de pouvoir construire des tests d'adéquation (deux différents à partir de (2.8) et un autre à partir de (2.9)) pour rejeter un modèle donné (voir Section 5 de [CL10] pour plus de détails sur les différentes étapes). Soulignons que de ces trois tests, le test déduit de (2.9) est clairement le plus simple à mener car pour être appliqué il suffit (outre certaines hypothèses d'intégrabilité et de régularité) de vérifier que λ_{Inn} est une constante positive. Et ceci est relativement simple (voir Section 6 de [CL10] à ce sujet) essentiellement car elle ne dépend pas de la nature de l'estimateur considéré.

Nous souhaitons terminer ce paragraphe en précisant que pour obtenir l'ensemble de ces résultats, nous avons utilisé des théorèmes ergodiques, des techniques de δ-méthode et un théorème central limite multivarié dans un contexte non-stationnaire et pour des tableaux triangulaires généralisant le résultat de [72]. Cette nécessité d'étendre le travail de [72] est essentiellement due au Contexte 1.

Théorème 23 *Soit $X_{n,i}$, $n \in \mathbb{N}$, $i \in \mathbb{Z}^d$, un champ dans un espace mesurable S. Pour $n \in \mathbb{N}$, soit $\mathcal{K}_n \subset \mathbb{Z}^d$ et pour $k \in \mathcal{K}_n$, supposons $\mathbf{Z}_{n,k} = f_{n,k}\left(X_{n,k+i},\ i \in \mathcal{I}_0\right)$, où $\mathcal{I}_0 = \{i \in \mathbb{Z}^d,\ |i| \leq 1\}$ et $f_{n,k} : S^{\mathcal{I}_0} \to \mathbb{R}^p$. Soit $\mathbf{S}_n = \sum_{k \in \mathcal{K}_n} \mathbf{Z}_{n,k}$. Si*

(i) $c_3 := \sup_{n \in \mathbb{N}} \sup_{k \in \mathcal{K}_n} E|\mathbf{Z}_{n,k}|^3 < \infty$,

(ii) $\forall n \in \mathbb{N}$, $\forall k \in \mathcal{K}_n$, $E(\mathbf{Z}_{n,k} | X_{n,j},\ j \neq k) = 0$,

(iii) $|\mathcal{K}_n| \to +\infty$ *lorsque* $n \to \infty$,

(iv) Il existe une matrice symétrique $\underline{\mathbf{\Sigma}} \geq 0$ *telle que*

$$E \left\| |\mathcal{K}_n|^{-1} \sum_{k \in \mathcal{K}_n} \sum_{j \in \mathrm{B}_k(1) \cap \mathcal{K}_n} \mathbf{Z}_{n,k} \mathbf{Z}_{n,j}{}' - \underline{\mathbf{\Sigma}} \right\| \to 0,$$

alors $|\mathcal{K}_n|^{-1/2} \mathbf{S}_n \xrightarrow{d} \mathcal{N}(0, \underline{\mathbf{\Sigma}})$ *lorsque* $n \to \infty$.

Chapitre 3

Perspectives de recherche et travaux en cours

Sommaire

3.1 En lien direct avec le Chapitre précédent

Résultats de convergence du MLE

Afin de faire le tour sur les méthodes d'estimation paramétrique, il nous semble pertinent de démontrer des résultats plus complets pour l'estimateur du maximum de vraisemblance. Bien que très utilisée, cette méthode a été à notre connaissance étudiée uniquement par [86] sur un plan théorique. Il établit la consistance de l'estimateur du MLE uniquement pour une classe restreinte de processus de Gibbs. Elargir cette classe de manière à pouvoir (entre autres) inclure les exemples présentés dans le Chapitre 1 est une première perspective intéressante.

Autour de la méthode de TF

La méthode de Takacs-Fiksel est à double tranchant. Elle est riche et souple du fait du choix assez large de fonctionelles possibles. Mais se pose aussi la question du choix optimal de fonctionelles. Un futur travail pourrait alors consister à construire et étudier un estimateur agrégé, ou obtenu par une procédure de sélection de fonctions tests.

Les résultats asymptotiques obtenus suggèrent d'utiliser cet estimateur comme point de départ d'un algorithme plus fin. Citons par exemple, la première étape de l'algorithme de Newton Raphson (comme utilisé dans [66]) qui permet une approximation précise de l'estimateur du MLE en partant d'un estimateur consistant en une seule étape. Bien que les justifications soient manquantes dans le cadre des processus de Gibbs, il est bien connu que cette procédure conduit à un estimatuer efficace dans le cas i.i.d. (voir [80]). Une autre direction serait d'exploiter la propriété LAN (local asymptotic normality) du modèle de Gibbs. Celle-ci n'a été montrée par Mase [86] que pour des modèles restrictifs mais on peut espérer qu'elle reste vraie pour plus de modèles.

Autour des résidus

Dans ce cadre, une étude en simulation est indispensable pour valider l'étude théorique et l'intérêt pratique. Les méthodes existantes utilisant les résidus (cf [95]) sont basées sur des simulations sous l'hypothèse nulle pour construire la règle de décision. Une comparaison entre les deux approches est également nécessaire. Cette étude pourrait également être complétée par une étude de la fonction puissance. Ceci permettrait de pouvoir différencier un peu plus les trois types de tests

proposés.

Autour de la portée

Dans nos différentes contributions pour démontrer des résultats de convergence en loi, nous avons supposé la portée (hypothèse **[FR]**) du processus ponctuel. Ceci est "imposé" par le TCL (Théorème 23) que nous utilisons, ce qui est assez frustrant car l'existence d'une mesure ergodique et la consistance d'estimateurs et de résidus ne nécessitent pas nécessairement cette hypothèse. Par exemple, dans [38] il est montré que le modèle suivant (définie par la famille d'énergies)

$$V_\Lambda(\varphi;\theta) = \theta_1|\varphi_\Lambda| + \sum_{\substack{\{x,y\}\in Del_2(\varphi)\\ \{x,y\}\cap\Lambda\neq\emptyset}} \|y-x\|.$$

existe alors que nos résultats ne peuvent s'appliquer pour ce modèle. L'existence du modèle précédent provient du fait que le graphe de Delaunay assure que ce processus possède en quelque sorte une propriété de type portée pour chaque configuration. Rechercher un TCL multivarié qui pourrait prendre en compte ce type de portée et obtenir des résultats d'inférence constituent des objectifs tout à fait intéressants.

Vers la non stationnarité

L'ensemble des résultats (estimation/validation) ont été obtenus en supposant le processus ponctuel marqué stationnaire. Cette hypothèse n'est pas très réaliste pour un certain nombre d'applications d'une part et d'autre part la plupart des méthodes ont une extension en non stationnaire. Pouvoir décrire des résultats similaires en relâchant la contrainte de stationnarité est une perspective intéressante. En ce sens, l'utilisation du travail de Comets et Janzura [30] nous semble une bonne base de travail.

Transformation d'un processus ponctuel en un processus de Poisson

Dans un très récent travail [93], Møller et Berthelsen démontrent qu'il est possible de transformer un processus ponctuel de Gibbs en un processus de Poisson et proposent un algorithme efficace. Cet algorithme dépend évidemment de la densité du modèle initial et en particulier de l'énergie locale $V(x|\varphi;\theta^\star)$. Les auteurs proposent d'utiliser leur algorithme pour valider un modèle : (a) Envisager un modèle et estimer $\theta^\star$ à partir de φ. (b) Utiliser leur algorithme en remplaçant $\theta^\star$ par $\widehat{\theta}$ pour obtenir nouveau processus ponctuel ψ. (c) Si le modèle envisagé est correct, ψ est la réalisation d'une loi de Poisson (ce qui est très facile à tester à l'aide des fonctions classiques F ou L, cf [95]).

L'approche développée par Møller et Berthelsen est novatrice et prometteuse et il nous semble intéressant de travailler sur deux points distincts :

- *Validation.* Du fait de l'estimation de $\theta^\star$ par $\widehat{\theta}$, la loi de Ψ n'est en théorie pas une loi de Poisson. Comprendre l'écart entre l'estimation de la fonction F ou L basée sur ψ à celle d'un processus de Poisson est primordial pour obtenir des résultats asymptotiques.
- *Estimation.* On pourrait envisager d'estimer les paramètres d'un modèle de Gibbs en utilisant cet algorithme, i.e. en choisissant $\widehat{\theta}$ tel que la fonction de vide F soit la plus proche possible de celle d'un Poisson.

3.2 Projet IXXI : Sunspot

Nous venons d'obtenir (juin 2010) le financement pour un projet de deux ans IXXI Systèmes Complexes. Ce projet fédère trois laboratoires grenoblois : le Laboratoire Jean Kuntzmann (JF Coeurjolly, porteur du projet), le GIPSA-lab (Pierre-Olivier Amblard et Nicolas Le Bihan) et le laboratoire de Planétologie (Jean Lilenstein). Certaines des thématiques de ce projet de recherche

ont trait à la statistique spatiale et plus précisément aux processus ponctuels spatiaux d'où sa place dans ce Chaptitre.

L'objectif de ce projet concerne la modélisation du phénomène des taches solaires. Une tache solaire est une région de la surface du soleil marquée par une température inférieure à son environnement et par une intense activité magnétique inhibant la convection. C'est principalement la baisse de température de la tache relativement à son environnement qui la rend visible. La compréhension et l'analyse de ces taches solaires sont des problèmes extrêmement importants car ils sont des indices du niveau d'activité solaire (e.g. [70, 74]). L'observation de ces taches remonte aux temps anciens. Cependant on considère qu'elles sont à peu près fiables depuis 1874 et prétraitées d'une manière standard depuis 1976 (cf site web `http://solarscience.msfc.nasa.gov/greenwch.shtml`).

Ces données sont riches puisque l'on dispose d'une information temporelle (décompte quotidien des taches) et d'une information spatiale (position des taches à la surface du soleil). L'objectif de projet de recherche vise à proposer des outils et modèles adaptés à ce type de données en vue de proposer un modèle de processus ponctuel spatio-temporel sur la sphère. L'originalité de ce projet est clairement établie ici car à notre connaissance il n'existe aucune modélisation intégrant à la fois l'information temporelle et spatiale. La plupart des travaux se concentre sur l'aspect modélisation temporelle (e.g. [59]) ou sur la description spatiale (voir par exemple le travail pionnier de Maunder [90]).

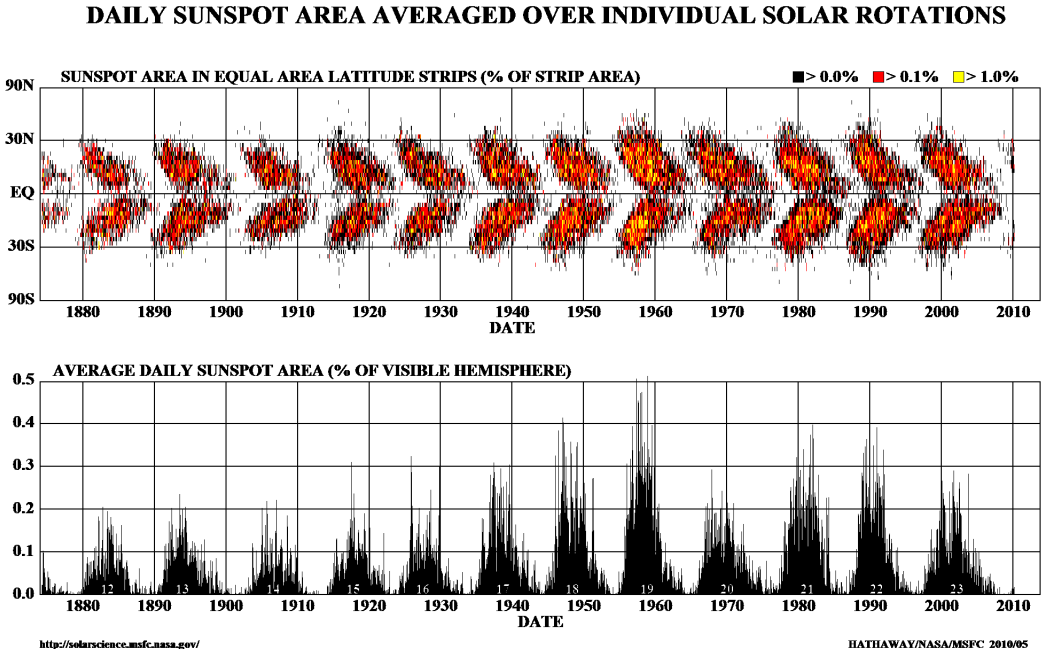
Les difficultés et verrous de ce projet sont intrinsèquement liés à la nature complexe et hétérogène des données :

- La série chronologique des nombres de taches solaires exhibe une périodicité liée au cycle solaire (de l'ordre de 10-11 ans), un caractère longue mémoire et certaines tendances non stationnaires.

- A tout instant, le décompte et le positionnement des taches solaires ne sont que partiels (face visible du soleil). De plus, une difficulté inhérente provient du fait que la modélisation et l'inférence doivent se faire avec des réalisations de variables aléatoires à valeurs sur une sphère (cas particulier d'une variété Riemanienne).

- Enfin l'arrangement spatial des taches solaires n'est pas homogène sur la sphère. Les taches sont plutôt concentrées autour de l'équateur (et de manière plutôt symétrique par rapport à celui-ci). Par ailleurs, plus l'activité solaire est faible plus les taches semblent localisées autour de l'équateur. Cette caractéristique spatiale est appelée "effet papillon".

Certains de ces aspects (caractère cyclique et arrangement inhomogène) sont illustrés par la figure suivante (source : David Hathaway, NASA Marshall Space Flight Center http ://solarscience.msfc.nasa.gov).



Dans le cadre du projet IXXI, nous avons distingué plusieurs axes de recherche. Ces pistes de réflexion sont clairement associées à certaines caractéristiques des données. Nos réflexions porteront sur le caractère méthodologique, l'obtention de résultats théoriques, un aspect computationnel important et une application des méthodes développées au problème considéré :

- *Aspect temporel pur.* l'idée sera ici de travailler avec la série chronologique issue du comptage des taches solaires. En particulier, on cherchera à comparer le modèle récent proposé par [59] avec une approche modélisation par un processus de comptage de type Poisson inhomogène où l'intensité dépendra du cycle solaire. L'objectif visé ici sera la caractérisation des cycles solaires (période moyenne,...) ainsi que des minima et maxima du nombre des taches solaires.

- *Description spatiale.* Dans cet axe, l'idée sera de supposer qu'à tout instant (ou sur une courte période) les données spatiales (i.e. la position des taches) correspondent à des réalisations indépendantes d'une variable aléatoire sur la sphère (par exemple loi normale sur la sphère [100]). Cet axe est très prometteur d'un point de vue recherche car il existe relativement peu de travaux de statistique inférentielle pour ce genre de problèmes ([20]). En particulier, on s'attachera à caractériser spatialement la moyenne et la dispersion des taches solaires par une construction de boules de confiance induites par la distance géodésique et le comportement limite de l'estimateur de la moyenne intrinsèque. Une question typique à laquelle nous souhaitons répondre via un outil de test d'hypothèses est la suivante : y a-t-il une évolution spatiale de la moyenne intrinsèque à deux périodes différentes et en particulier à deux périodes correspondant aux maxima du nombre de taches.

  Développons un peu de dernier point qui constitue un travail en cours avec Nicolas Le Bihan. Nous portons notre intérêt sur la loi normale définie sur une variété riemanienne proposée par Pennec [100]. Cette loi de probabilité définie pour de nombreuses variétés a été très peu étudiée et nous souhaitons dans un premier temps l'étudier sur des variétés simples telles que le cercle, la sphère ou le groupe $SO(3)$. Plus précisément, sur le cercle notre premier objectif est de comparer les propriétés théoriques de cette variable avec les lois très classiques et très utilisées dans le domaine des données circulaires [85] - la loi de

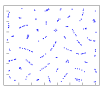
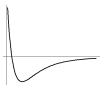
Von Mises (connue aussi sous le nom de “circular normal distribution”).

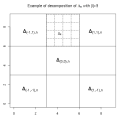
– *Approche spatiale pure.* A temps fixé, une autre approche consiste à modéliser les taches solaires par la réalisation d'un processus ponctuel sur la sphère. En particulier, fondé sur [22], on s'intéressera à caractériser l'inhomogénéité du processus ponctuel par une estimation non paramétrique de sa fonction intensité. La seconde question fondamentale sera alors de savoir s'il y a une interaction possible entre les taches solaires ou si un processus de Poisson spatial inhomogène est adéquat. D'un point de vue théorique, on s'intéressera à l'extension des propritétés asymptotiques d'estimateurs paramétriques et de techniques de validation de processus ponctuels de Gibbs, connues sur $\mathbb{R}^d$ (voir le Chapitre précédent), au cas de la sphère.

– *Approche spatio-temporelle.* L'objectif est d'agréger les connaissances des précédentes approches pour proposer et valider un modèle de processus ponctuel spatio-temporel sur la sphère. En particulier, on commencera par étudier un processus de Poisson inhomogène dont l'intensité conditionnelle dépend du temps et de l'espace de telle sorte que les marges en temps et en espace correspondent aux observations des précédentes approches.

Chapitre 4

Liste des articles en relation avec la Partie II

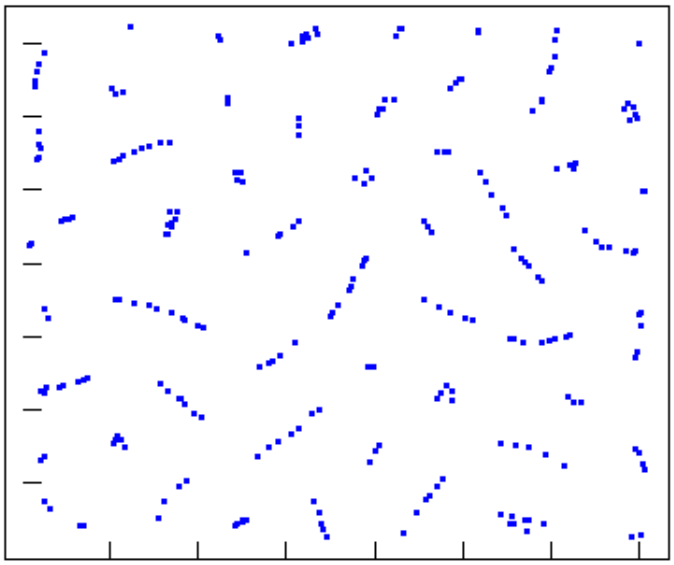
Articles parus	
	**[BCD08]** J.-M. Billiot, J.-F. Coeurjolly and R. Drouilhet. Maximum pseudolikelihood estimator for exponential family. *Electronic Journal of Statistics*, 2:234-264, 2008.
	**[HLTGDC09]** T. Hurtut, P.-E. Landes, J. Thollot, Y. Gousseau, R. Drouilhet and J.-F. Coeurjolly. Appearance-guided Synthesis of Element Arrangements by Example, In *7th International Symposium on Non-Photorealistic Animation and Rendering, NPAR 2009*, 2009.
	**[CD10]** J.-F. Coeurjolly and R. Drouilhet. Asymptotic properties of the maximum pseudo-likelihood estimator for stationary Gibbs point processes including the Lennard-Jones model. *Electronic Journal of Statistics*, 4:677-706 (2010).
	**[CDDL10]** J.-F. Coeurjolly, D. Dereudre, R. Drouilhet and F. Lavancier. Takacs-Fiksel method for stationary marked Gibbs point processes. *in revision for Scandinavian Journal of Statistics* (available at `http://arxiv.org/abs/1007.3351`), 2010.

Articles soumis	
	**[CL10]** J.-F. Coeurjolly and F. Lavancier. Residuals for stationary marked Gibbs point processes. *submitted for publication* (available at `http://arxiv.org/abs/1002.0857`), 2010.

Chapitre 5

Annex Part II : published and submitted papers

This chapter contains a copy of published or submitted papers related to Part II. These papers are chronologically ordered.



[BCD08]
J.-M. Billiot, J.-F. Coeurjolly and R. Drouilhet. Maximum pseudolikelihood estimator for exponential family. *Electronic Journal of Statistics*, 2:234-264, 2008.

Electronic Journal of Statistics
Vol. 2 (2008) 234–264
ISSN: 1935-7524
DOI: 10.1214/07-EJS160

Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes

Jean-Michel Billiot, Jean-François Coeurjolly and Rémy Drouilhet

SAGAG Team, Dept. of Statistics, LJK, Grenoble University, France
e-mail: Jean-Michel.Billiot@upmf-grenoble.fr;
Jean-Francois.Coeurjolly@upmf-grenoble.fr; Remy.Drouilhet@upmf-grenoble.fr

Abstract: This paper is devoted to the estimation of a vector $\boldsymbol{\theta}$ parametrizing an energy function of a Gibbs point process, via the maximum pseudolikelihood method. Strong consistency and asymptotic normality results of this estimator depending on a single realization are presented. In the framework of exponential family models, sufficient conditions are expressed in terms of the local energy function and are verified on a wide variety of examples.

AMS 2000 subject classifications: Primary 60G55; secondary 60J25.
Keywords and phrases: stationary marked Gibbs point processes, pseudolikelihood method, minimum contrast estimators.

Received December 2007.

1. Introduction

The class of Gibbs point processes is interesting because it allows us to introduce and study interactions between points through the modelling of an associated energy function. Historical aspects of the mathematical theory are covered briefly in Kallenberg (1983). When the energy function is parametrized, one among many other methods of estimation, is the maximization of the pseudolikelihood. Baddeley and Turner (2000) dealt with some practical aspects of such a parametric method and gave a survey on asymptotic results. They noticed that some classical examples (such as the area-interaction model, the Multi-Strauss model, the 2-type Strauss model) do not satisfy the assumptions of the existing asymptotic normality results. This paper aims at filling this gap.

Many proposals tried to estimate the energy function from the available point pattern data generated by some marked Gibbs point processes. If the energy belongs to a parametric family model, the most well-known methodology is the use of the likelihood function, see *e.g.* Møller and Waggepetersen (2003) and the references therein. The main drawback of this approach is that the likelihood function contains an unknown scaling factor whose value depends on the parameters and which is difficult to calculate. An alternative approach relies on the use of the pseudolikelihood. This idea originated from Besag (1974) in the study of lattice processes. Besag et al. (1982) further considered this method

for pairwise interaction point processes, while Jensen and Møller (1991) generalized it to the general class of marked Gibbs point processes. A general review of the problem of statistical inference on spatial point processes including the Takacs-Fiksel method (a parametric method based on a characteristic property of marked Gibbs point processes using Palm measure) and non-parametric methods can be found in the recent monograph of Møller and Waggepetersen (2003).

In order to underline our theoretical contributions, let us present the different papers discussing asymptotic properties of the maximum pseudolikelihood estimator. The first work was done by Jensen and Møller (1991). The authors obtained consistency for exponential family models of marked point processes. They mainly considered inhibition and hard-core models. They notably applied their results on the marked Strauss process. Jensen and Künsch (1994) and Mase (1995) focused on specific models with two parameters -the chemical potential and the inverse temperature- which can be viewed as particular exponential family models. Jensen and Künsch (1994) obtained an asymptotic normality result by first assuming the inhibition or hard-core property and then the finite range property. Mase (1995) established consistency for the class of superstable and lower regular potentials introduced by Ruelle (1970). Mase (2000) extended his work to the context of marked point processes and provided asymptotic normality by adding the assumption of finite range. Our goal is to deal with most classical models (see Baddeley and Turner (2000), Møller and Waggepetersen (2003) and Bertin et al. (1999b)) that could be interesting for practical purposes. They have been put into three categories according to their validity with respect to the previous existing works: 1) Overlap area point process 2) Multi-Strauss marked model, Strauss disc type process. 3) area-interaction, Geyer's triplet process, k-nearest-neighbour multi-Strauss marked model. Let us notice first that all the examples belong to the exponential family and satisfy the local stability and finite range properties. Due to the parametrization proposed by Jensen and Künsch (1994) and Mase (2000), they only consider the first category. By considering the exponential family, Jensen and Møller (1991) include a larger class of models. However, the required inhibition or hard-core type assumptions are only satisfied for examples 1 and 2. Examples 3 are only locally stable. What remains to be established is consistency for examples 3 and asymptotic normality for examples 2 and 3. In this paper, a general framework is proposed taking into consideration the previous remarks. Results are obtained using the general theory on minimum contrast estimators, *e.g.* Guyon (1995).

Section 2 introduces some background on marked Gibbs point processes. Our models are defined in Section 3. In the same spirit as Bertin et al. (1999a), providing existence results of stationary Gibbs states, assumptions on these models are expressed in terms of the local energy function. We also describe examples of interest of this work. Section 4 presents the pseudolikelihood method and our main results requiring two additional assumptions. The first one is an identifiability condition ensuring strong consistency. The second one is related to the definiteness of the asymptotic covariance matrix of the maximum pseudolikelihood estimator. These two assumptions allow us to derive a practical asymptotic

normality result. They are verified for all the considered examples in Section 5. It is the belief of the authors that these assumptions are not restrictive since they should be true for every well-parametrized model. Proofs have been postponed until Section 6.

2. Background on marked Gibbs point processes

For the sake of simplicity, the framework of this paper is restricted to two-dimensional marked Gibbs point processes. All the results must remain valid in the general d-dimensional ($d \geq 1$) case. Define $\mathcal{B}^2$ the Borel σ-algebra on $\mathbb{R}^2$, $\mathcal{B}_b^2$ the set of bounded Borel susbsets of $\mathbb{R}^2$ and λ^2 the Lebesgue measure on $\mathbb{R}^2$. Denote also by $\mathbb{M}$, $\mathcal{M}$ and λ^{m} the mark space and its corresponding σ-algebra and probability measure. Let $\mathbb{S} := \mathbb{R}^2 \times \mathbb{M}$, $\mathcal{B} := \mathcal{B}^2 \otimes \mathcal{M}$ and $\mu := \lambda^2 \otimes \lambda^{\mathrm{m}}$ denote respectively the state space and its corresponding σ-algebra and measure.

For shortness, let us denote $x^m = (x, m)$ for any $x \in \mathbb{R}^2$ and any mark $m \in \mathbb{M}$ and $|\Lambda| := \lambda^2(\Lambda)$ for any $\Lambda \in \mathcal{B}^2$. In addition, $|I|$ designates the number of elements of some countable set I, Λ^c is the complementary of some set Λ in $\mathbb{R}^2$ and $||\cdot||$ is the ℓ^2-norm. Let us define for all $i = (i_1, i_2) \in \mathbb{Z}^2$, $d > 0$ and $\rho \geq 0$ $\Delta_i(d) := \left\{z \in \mathbb{R}^2, d\left(i_j - \frac{1}{2}\right) \leq z_j \leq d\left(i_j + \frac{1}{2}\right), j = 1, 2\right\}$ and $\mathbb{B}(i, \rho) := \{k \in \mathbb{Z}^2 : |k - i| \leq \rho\}$ with $|i| := max(|i_1|, |i_2|)$.

Let $\widetilde{\Omega}$ denote the set of so-called configurations -of marked points- $\varphi := \{x_i^{m_i}\}_{i \in I}$ where I is a subset of $\mathbb{N}$ and $((x_i, m_i))_{i \in I}$ is a sequence of elements of $\mathbb{S}$. In particular, any element $\varphi \in \widetilde{\Omega}$ has the following representation $\varphi = \sum_{i \in I} \delta_{x_i^{m_i}}$ as an integer-valued measure on $\mathbb{S}$ such that for every $F \in \mathcal{B}_b^2$, $\varphi(F) \in \mathbb{N}$, where δ_{x^m} is the Dirac measure at some element $x^m \in \mathbb{S}$. The subset of $\widetilde{\Omega}$ with elements φ satisfying $|\varphi| := \varphi(\mathbb{S}) < +\infty$ is denoted by $\widetilde{\Omega}_f$. The space $\widetilde{\Omega}$ is equipped with the σ-algebra $\mathcal{F}$ generated by the family of sets $\left\{\varphi \in \widetilde{\Omega} : \varphi(F) = n\right\}$ with $n \in \mathbb{N}$ and $F \in \mathcal{B}_b^2$. For every $F \in \mathcal{B}^2$ and $\varphi \in \widetilde{\Omega}$ represented as $\varphi = \sum_{i \in I} \delta_{x_i^{m_i}}$, one introduces $\varphi_F := \sum_{i \in I, x_i^{m_i} \in F} \delta_{x_i^{m_i}}$ which can be viewed as the configuration of marked points of φ in F. Furthermore, for every $\Lambda \in \mathcal{B}_b^2$, φ_Λ conveniently denotes $\varphi_{\Lambda \times \mathbb{M}}$.

A marked point process is a $\widetilde{\Omega}$-valued random variable, denoted by Φ, with probability distribution P on $(\widetilde{\Omega}, \mathcal{F})$. The intensity measure N_P of P is defined as a measure on $\mathcal{B}^2$ such that for any $F \in \mathcal{B}_b^2$:

$$N_P(F) = \int_{\widetilde{\Omega}} \varphi(F) P(d\varphi) := \boldsymbol{E}(\Phi(F)).$$

In the stationary case, $N_P(F) = \nu_P \lambda^2(F)$ where ν_P is called the intensity of P. A marked Gibbs point process is usually defined using a family of local specifications with respect to a weight process (often a stationary marked Poisson process with distribution Q and intensity $\lambda_Q = 1$). Let Λ be a bounded region in $\mathbb{R}^2$. For such a process, given some configuration φ_{Λ^c} on Λ^c, the conditional

probability on Λ is of the form, for any $F \in \mathcal{F}$:

$$\Pi_\Lambda(\varphi, F) = \left\{ \frac{1}{Z_\Lambda(\varphi)} \int_{\widetilde{\Omega}_\Lambda} e^{-V(\psi|\varphi_{\Lambda^c})} 1_F(\psi \cup \varphi_{\Lambda^c}) Q_\Lambda(d\psi) \right\} 1_{R_\Lambda}(\varphi),$$

with the partition function

$$Z_\Lambda(\varphi) = \int_{\widetilde{\Omega}_\Lambda} e^{-V(\psi|\varphi_{\Lambda^c})} Q_\Lambda(d\psi)$$

and $R_\Lambda = \{\varphi \in \widetilde{\Omega} \,:\, 0 < Z_\Lambda(\varphi) < +\infty\}$ where

$$\int f(\psi) Q_\Lambda(\psi) := e^{-\mu(\Lambda \times \mathbb{M})} \sum_{n=0}^{+\infty} \frac{1}{n!} \int f(\underbrace{\{x_1^{m_1}, \ldots, x_n^{m_n}\}}_{\psi}) d\mu^{\otimes n}(x_1^{m_1}, \ldots, x_n^{m_n}).$$

Let us define the subset of all admissible configurations

$$\Omega := \left\{ \varphi \in \widetilde{\Omega} : \varphi \in \cap_{\Lambda \in \mathcal{B}_b^2} R_\Lambda \right\}$$

and denote by $\Omega_f := \widetilde{\Omega}_f \cap \Omega$. Whereas the finite energy function $V(\varphi)$ (for any $\varphi \in \Omega_f$) measures the cost of any configuration, the local energy $V(\psi|\varphi)$ (for any $\varphi, \psi \in \Omega_f$) represents the energy required to add the points of ψ in φ:

$$V(\psi|\varphi) = V(\psi \cup \varphi) - V(\varphi).$$

Let us notice that when ψ is a singleton $\{x^m\}$, we denote by a slight abuse $V(x^m|\varphi)$ instead of $V(\{x^m\}|\varphi)$. It is well-known that the collection of probability kernels $(\Pi_\Lambda)_{\Lambda \in \mathcal{B}_b^2}$ satisfies the set of compatibility and measurability conditions which define a local specification in the Preston's sense (Preston (1976)). The main condition is the consistency:

$$\Pi_\Lambda \Pi_{\Lambda'} = \Pi_\Lambda \quad \text{for} \quad \Lambda' \subset \Lambda.$$

Notice that some conditions are needed to ensure the existence of a probability measure P related to any local energy V and any weight process that satisfies the so-called Dobrushin-Lanford-Ruelle (D.L.R.) equations:

$$P(F|\mathcal{F}_{\Lambda^c})(\varphi) = \Pi_\Lambda(\varphi, F) \quad \text{for } P \text{ a.e. } \varphi \in \Omega \quad \text{for any } \Lambda \in \mathcal{B}_b^2 \text{ and } F \in \mathcal{F}.$$

For the general theory of Gibbs point processes, the reader may refer to Kallenberg (1983); Stoyan et al. (1995) and the references therein.

For some finite configuration φ (resp. some set G) and for all $x \in \mathbb{R}^2$, φ_x (resp. G_x) denotes the configuration φ (resp. the set G) translated of x. Finally, in this work a non-marked point process can be viewed as a particular case of marked point processes with $\mathbb{M} = \{0\}$.

3. Definitions and examples of marked Gibbs models

The framework of this paper is restricted to stationary marked Gibbs point processes based on an energy function invariant by translation, $V(\varphi;\boldsymbol{\theta})$, parametrized by some $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is some compact set of $\mathbb{R}^p$. The model is also assumed to belong to an exponential family, i.e.

$$V(\varphi;\boldsymbol{\theta}) = \boldsymbol{\theta}^T \boldsymbol{v}(\varphi), \tag{3.1}$$

where $\boldsymbol{v}(\varphi) = (v_1(\varphi),\dots,v_p(\varphi))$ is the vector of sufficient statistics. The local energy is then expressed as

$$V(x^m|\varphi;\boldsymbol{\theta}) = \boldsymbol{\theta}^T \boldsymbol{v}(x^m|\varphi), \tag{3.2}$$

where $\boldsymbol{v}(x^m|\varphi) = (v_1(x^m|\varphi),\dots,v_p(x^m|\varphi)) := \boldsymbol{v}(\varphi\cup\{x^m\}) - \boldsymbol{v}(\varphi)$.
Our models satisfy the general condition **[Mod]** described by the following statements:

[Mod:S] *Stability of the local energy*: there exists $K\geq 0$ such that for all $(m,\varphi)\in\mathbb{M}\times\Omega_f$

$$V(0^m|\varphi;\boldsymbol{\theta}) \geq -K.$$

[Mod:L] *Locality of the local energy*: there exists $D\geq 0$ such that for all $(m,\varphi)\in\mathbb{M}\times\Omega_f$

$$V(0^m|\varphi;\boldsymbol{\theta}) = V\left(0^m|\varphi_{B(0,D)};\boldsymbol{\theta}\right),$$

where $\mathcal{B}(x,r)$ denotes the ball centered at $x\in\mathbb{R}^2$ with radius $r>0$.

[Mod:I] *Integrability condition*: for $i=1,\dots,p$, there exist $\kappa_i^{(\sup)}\geq 0$, $k_i\in\mathbb{N}$ such that for all $(m,\varphi)\in\mathbb{M}\times\Omega_f$

$$v_i(0^m|\varphi) \leq \kappa_i^{(\sup)} |\varphi_{\mathcal{B}(0,D)}|^{k_i}.$$

Let us notice that, unlike **[Mod:I]**, the assumptions **[Mod:S]** and **[Mod:L]** cannot be directly expressed only in terms of the sufficient statistics. Nevertheless, **[Mod]** is satisfied as soon as for $i=1,\dots,p$, there exist $\kappa_i^{(\inf)},\kappa_i^{(\sup)}\geq 0$, $k_i\in\mathbb{N}$ such that one of both following assumptions is satisfied for all $(m,\varphi)\in\mathbb{M}\times\Omega_f$:

[Mod-1]

$$\theta_i\geq 0 \text{ and } -\kappa_i^{(\inf)} \leq v_i(0^m|\varphi) = v_i(0^m|\varphi_{\mathcal{B}(0,D)}) \leq \kappa_i^{(\sup)} |\varphi_{\mathcal{B}(0,D)}|^{k_i}.$$

[Mod-2]

$$-\kappa_i^{(\inf)} \leq v_i(0^m|\varphi) = v_i(0^m|\varphi_{\mathcal{B}(0,D)}) \leq \kappa_i^{(\sup)}.$$

Indeed, let I_1 and I_2 be the partition of $\{1,\dots,p\}$ such that for any $i\in I_1$ (resp. $i\in I_2$), v_i satisfies **[Mod-1]** (resp. **[Mod-2]**) then

$$\begin{aligned} V(0^m|\varphi;\boldsymbol{\theta}) &= \sum_{i\in I_1}\theta_i v_i(0^m|\varphi) + \sum_{i\in I_2}\theta_i v_i(0^m|\varphi) \\ &\geq -p\left(\max_{i\in I_1}\left(\theta_i\kappa_i^{(\inf)}\right) - \max_{i\in I_2}\left(|\theta_i|\times\max(\kappa_i^{(\inf)},\kappa_i^{(\sup)})\right)\right) := -K(\boldsymbol{\theta}) \end{aligned}$$

which ensures [**Mod:S**] with $K := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} K(\boldsymbol{\theta})$ ([**Mod:L**] and [**Mod:I**] are clearly satisfied).
Let us also point out that

- the well-known characteristics [**Mod:S**] and [**Mod:L**] associated with finite energies that are translation invariant ensure the existence of stationary measures (see Bertin et al. (1999a)).
- the local stability implies the Ruelle-bound correlation function (see Ruelle (1970)) leading to: $\boldsymbol{E}\left(\left|\Phi_{\mathcal{B}(0,D)}\right|^{\beta}\right) < +\infty$ for any $\beta > 0$.
- A key-ingredient of our proofs is the following one: for any $\alpha > 0$, any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and any $i = 1, \ldots, p$

$$\boldsymbol{E}\left(|v_i(0^M|\Phi)|^{\alpha} e^{-\boldsymbol{\theta}^T \boldsymbol{v}\left(0^M|\Phi\right)}\right) < +\infty. \tag{3.3}$$

  This condition is fulfilled under the assumptions [**Mod:S**] and [**Mod:I**].

Let us now present some examples. Except the one based on the k-nearest-neighbour graph, all examples are really classical and can be found *e.g.* in Baddeley and Turner (2000) and Møller and Waggepetersen (2003). For each example, we present the model through the sufficient statistics, the set of the parameters values (including $\boldsymbol{\Theta}$) for which the model is defined in the litterature and then we verify that [**Mod**] is satisfied. This proves in particular the existence of stationary Gibbs states in $\mathbb{R}^2$.

First of all, note that when $v_i(0^m|\varphi) := 1$, then v_i obviously satisfies [**Mod-2**]. Recall that for a non-marked point process $\mathbb{M} = \{0\}$.

Overlap area point process

- This non-marked process is defined for $p = 2$ and $R > 0$ by

$$v_1(\varphi) := |\varphi| \text{ and } v_2(\varphi) := \sum_{\{x,y\}\in\mathcal{P}_2(\varphi)} |\mathcal{B}(x, R/2) \cap \mathcal{B}(y, R/2)|,$$

  where $\mathcal{P}_k(\varphi)$ $(k \geq 1)$ is the set of all subsets of φ with k elements. Alternatively,

$$v_1(0|\varphi) := 1 \text{ and } v_2(0|\varphi) := \sum_{x\in\varphi} |\mathcal{B}(0, R/2) \cap \mathcal{B}(x, R/2)|.$$

  Let us notice that

$$\begin{aligned} &|\mathcal{B}(0, R/2) \cap \mathcal{B}(x, R/2)| \\ &\quad = \frac{1}{2}\left(R^2 Arcos\frac{||x||}{R} - ||x||\sqrt{R^2 - ||x||^2}\right)\mathbf{1}_{[0,R]}(||x||). \end{aligned}$$

- $\boldsymbol{\theta} \in \mathbb{R} \times \mathbb{R}^+$.
- v_2 satisfies [**Mod-1**] with $\kappa_2^{(\inf)} = 0$, $\kappa_2^{(\sup)} = \frac{\pi R^2}{4}$, $D = R$ and $k_2 = 1$.

Multi-Strauss marked point process

- Let $\mathbb{M} = \{1, \ldots, M\}$, λ^{m} the uniform probability measure on $\mathbb{M}$ and $p \geq 2$. Decompose $\varphi = \cup_{m=1}^{M} \varphi^m$ with $\varphi^m := \{x^m \in \Omega : x^m \in \varphi\}$ for any $m \in \mathbb{M}$. The finite energy of this process is defined by

$$\begin{aligned} V(\varphi; \boldsymbol{\theta}) &:= \sum_{m_1=1}^{M} \theta_1^{m_1,m_1} v_1^{m_1,m_1}(\varphi^{m_1}) \\ &\quad + \sum_{1 \leq m_1 \leq m_2 \leq M} \sum_{i=2}^{p^{m_1,m_2}} \theta_i^{m_1,m_2} v_i^{m_1,m_2}(\varphi^{m_1} \cup \varphi^{m_2}) \end{aligned}$$

with for all $m_1, m_2 \in \mathbb{M}$ and $i = 2, \ldots, p$

$$\begin{aligned} v_1^{m_1,m_1}(\varphi) &:= v_1^{m_1,m_1}(\varphi^{m_1}) := |\varphi^{m_1}| \\ v_i^{m_1,m_2}(\varphi) &:= v_i^{m_1,m_2}(\varphi^{m_1} \cup \varphi^{m_2}) \\ &= \sum_{\{x_1^{m_1}, x_2^{m_2}\} \in \mathcal{P}_2(\varphi)} \mathbf{1}_{[D_{i-1}^{m_1,m_2}, D_i^{m_1,m_2}[}(||x_1 - x_2||) \end{aligned} \quad (3.4)$$

where $0 \leq D_1^{m_1,m_2} < D_2^{m_1,m_2} < \ldots < D_{p^{m_1,m_2}}^{m_1,m_2} < +\infty$. In particular, the vector $\boldsymbol{\theta}$ could be ordered as follows:

$$\boldsymbol{\theta} = (\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^M) \text{ where } \boldsymbol{\theta}^m = (\boldsymbol{\theta}^{m,m}, \boldsymbol{\theta}^{m,m+1}, \ldots, \boldsymbol{\theta}^{m,M})$$

with

$$\boldsymbol{\theta}^{m_1,m_2} = \begin{cases} (\theta_1^{m_1,m_2}, \theta_2^{m_1,m_2}, \ldots, \theta_{p^{m_1,m_2}}^{m_1,m_2}) & \text{if } m_1 = m_2 \\ (\theta_2^{m_1,m_2}, \theta_3^{m_1,m_2}, \ldots, \theta_{p^{m_1,m_2}}^{m_1,m_2}) & \text{otherwise.} \end{cases}$$

where $p = M + \sum_{1 \leq m_1 \leq m_2 \leq M} (p^{m_1,m_2} - 1)$. One then derives the expression of the local energy

$$V(0^{m_1}|\varphi; \boldsymbol{\theta}) = \theta_1^{m_1,m_1} + \sum_{m_2=1}^{M} \sum_{i=2}^{p^{m_1,m_2}} \theta_i^{m_1,m_2} v_i^{m_1,m_2}(0^{m_1}|\varphi^{m_2})$$

where for convenience $\theta_i^{m_1,m_2}$ and $\theta_i^{m_2,m_1}$ denote the same parameter.
- For all $m_1, m_2 \in \mathbb{M}$ and $i = 2, \ldots, p^{m_1,m_2}$: $\theta_1^{m_1,m_1} \in \mathbb{R}$ and $\theta_i^{m_1,m_2} \in \begin{cases} \mathbb{R} & \text{when } D_1^{m_1,m_2} = \delta > 0 \\ \mathbb{R}^+ & \text{when } D_1^{m_1,m_2} = 0. \end{cases}$
- When $D_1^{m_1,m_2} = 0$ and $\theta_i^{m_1,m_2} \geq 0$, $v_i^{m_1,m_2}$ satisfies **[Mod-1]** with $\kappa_{i(m_1,m_2)}^{(\inf)} = 0$, $\kappa_{i(m_1,m_2)}^{(\sup)} = 1$, $D = \max_{m_1 \leq m_2} D_{p^{m_1,m_2}}^{m_1,m_2}$ and $k_{i(m_1,m_2)} = 1$, where $\theta_i^{m_1,m_2} = \theta_{i(m_1,m_2)}$ is the $i(m_1, m_2)-$th element of the vector $\boldsymbol{\theta}$. Under the hard-core assumption $D_1^{m_1,m_2} = \delta > 0$, $v_i^{m_1,m_2}$ satisfies **[Mod-2]** with $\kappa_{i(m_1,m_2)}^{(\inf)} = 0$, $\kappa_{i(m_1,m_2)}^{(\sup)} = \lceil \frac{D^2}{\delta^2} \rceil$ and $D = \max_{m_1 \leq m_2} D_{p^{m_1,m_2}}^{m_1,m_2}$.

$k-$nearest-neighbour multi-Strauss marked point process

- This marked point process is defined similarly as the multi-Strauss marked point process except that the complete graph $\mathcal{P}_2(\varphi)$ in (3.4) is replaced by the k-nearest-neighbour graph ($k \geq 1$).
- For all $m_1, m_2 \in \mathbb{M}$ and $i = 2, \ldots, p^{m_1,m_2}$: $\theta_i^{m_1,m_2} \in \mathbb{R}$.
- In Bertin et al. (1999b), it is proved that $v_i^{m_1,m_2}$ satisfies **[Mod-2]** with $\kappa_{i(m_1,m_2)}^{(\inf)} = \kappa_{i(m_1,m_2)}^{(\sup)} = 13k$ and $D = 2\max_{m_1 \leq m_2} D_{p^{m_1,m_2}}^{m_1,m_2}$.

Strauss type disc process

- Let $\mathbb{M} = [0, M_{\max}]$ with $0 < M_{\max} < +\infty$, λ^{m} the uniform probability measure on $\mathbb{M}$ and $p = 2$. This model is defined by

$$v_1(\varphi) = |\varphi| \text{ and } v_2(\varphi) = \sum_{\{x_1^{m_1}, x_2^{m_2}\} \in \mathcal{P}_2(\varphi)} \mathbf{1}_{[0, m_1+m_2]}(||x_2 - x_1||).$$

  Alternatively,

$$v_2(0^m|\varphi) = \sum_{x^{m'} \in \varphi} \mathbf{1}_{[0, m+m']}(||x||).$$

- $\boldsymbol{\theta} \in \mathbb{R} \times \mathbb{R}^+$.
- v_2 satisfies **[Mod-1]** with $\kappa_2^{(\inf)} = 0$, $\kappa_2^{(\sup)} = 1$, $D = M_{\max}$ and $k_2 = 1$.

Geyer's triplet interaction point process

- This non-marked point process is defined for $p = 3$ and $R > 0$ by

$$V(\varphi; \boldsymbol{\theta}) = \theta_1|\varphi| + \theta_2 v_2(\varphi) + \theta_3 v_3(\varphi)$$

  where

$$v_2(\varphi) = \sum_{\{x_1, x_2\} \in \mathcal{P}_2(\varphi)} \mathbf{1}_{[0,R]}(||x_1 - x_2||)$$

  and

$$v_3(\varphi) = \sum_{\xi \in \mathcal{P}_3(\varphi)} \prod_{\{x_1, x_2\} \in \mathcal{P}_2(\xi)} \mathbf{1}_{[0,R]}(||x_1 - x_2||).$$

  Note that $v_3(\varphi)$ represents the number of triangles of φ with edges of lengths lower than R. Alternatively

$$v_2(0|\varphi) = \sum_{x \in \varphi} \mathbf{1}_{[0,R]}(||x||)$$

  and

$$v_3(0|\varphi) = \sum_{\{x,y\} \in \mathcal{P}_2(\varphi)} \prod_{\{x_1, x_2\} \in \mathcal{P}_2(\{x,y,0\})} \mathbf{1}_{[0,R]}(||x_1 - x_2||).$$

- $\boldsymbol{\theta} \in \mathbb{R}^2 \times \mathbb{R}^+ \setminus \{0\}$.
- When $\theta_2 \geq 0$, v_2 and v_3 satisfy **[Mod-1]** with $D = R$, $\kappa_2^{(\inf)} = \kappa_3^{(\inf)} = 0$, $\kappa_2^{(\sup)} = \kappa_3^{(\sup)} = 1$, $k_2 = 1$ and $k_3 = 2$. When $\theta_2 < 0$, v_2 satisfies neither **[Mod-1]** nor **[Mod-2]**. However, Geyer (1999) proved that the local energy is stable and local (i.e. **[Mod:S]** and **[Mod:L]**) and **[Mod:I]** is satisfied with $D = R$, $\kappa_2^{(\sup)} = \kappa_3^{(\sup)} = 1$, $k_2 = 1$ and $k_3 = 2$.

Area interaction point process

- This model is the one-type marginal of the two-type Widom-Rowlinson model. Let $p = 2$ and $R > 0$

$$V(\varphi; \boldsymbol{\theta}) = \theta_1 |\varphi| + \theta_2 v_2(\varphi), \quad \text{with} \quad v_2(\varphi) := \left| \cup_{x \in \varphi} \mathcal{B}(x, R) \right|.$$

  Note that $v_2(\varphi)$ represents the area of the union of discs of radius R centered at the points. Alternatively,

$$v_2(0|\varphi) := \left| \cup_{x \in \varphi_{\mathcal{B}(0,2R)} \cup \{0\}} \mathcal{B}(x, R) \setminus \cup_{x \in \varphi_{\mathcal{B}(0,2R)}} \mathcal{B}(x, R) \right|.$$

- $\boldsymbol{\theta} \in \mathbb{R}^2$.
- v_2 satisfies **[Mod-2]** with $\kappa_2^{(\inf)} = 0$, $\kappa_2^{(\sup)} = \pi D^2$ and $D = 2R$.

Remark 1. *Jensen and Møller (1991) have already proved the consistency property in the inhibition case, i.e. **[Mod:S]** with $K = 0$. In particular, they did not consider the area-interaction model with negative parameters θ_1 and θ_2 (for instance). This gap has now been filled by extending the result in the case when **[Mod:S]** is satisfied. However, unlike these authors, we require the additional assumption **[Mod:I]**. In order to simplify our assumptions, we deliberately decided not to propose this particular case since this last assumption is not restrictive and is satisfied for all examples considered above.*

Remark 2. *Note that unlike the Multi-Strauss marked point process, neither inhibition nor hard-core assumption is required for the k−nearest-neighbour multi-Strauss marked point process since its local energy is naturally stable.*

Remark 3. *Concerning the Geyer's triplet process, the case $\theta_3 = 0$ is not considered since it is a particular case of a multi-Strauss marked point process.*

Remark 4. *Through these different examples, one may note that some parameters are assumed to be known: for example, the parameters $D_i^{m_1,m_2}$ for the multi-Strauss marked point process, the hard-core parameter δ, the parameter M_{max} for the Strauss type disc process, the parameter R for the Geyer's triplet process or the area interaction point process, ... Their estimations could be investigated by using ad hoc methods.*

4. MPLE: presentation and asymptotic results

4.1. Pseudolikelihood

As specified in the introduction, the idea of maximum pseudolikelihood is due to Besag (1974) who first introduced the concept for Markov random fields in order to avoid the normalizing constant. This work was then widely extended and Jensen and Møller (1991) (Theorem 2.2) obtained a general expression for marked Gibbs point processes. With our notation and up to a scalar factor, the pseudolikelihood defined for a configuration φ and a domain of observation Λ is denoted by $PL_\Lambda(\varphi;\boldsymbol{\theta})$ and given by

$$PL_\Lambda(\varphi;\boldsymbol{\theta}) = \exp\left(-\int_{\Lambda\times\mathbb{M}} e^{-V(x^m|\varphi;\boldsymbol{\theta})}\mu(dx^m)\right)\prod_{x^m\in\varphi_\Lambda} e^{-V(x^m|\varphi\setminus x^m;\boldsymbol{\theta})}. \quad (4.1)$$

It is more convenient to define (and work with) the log-pseudolikelihood, denoted by $LPL_\Lambda(\varphi;\boldsymbol{\theta})$.

$$LPL_\Lambda(\varphi;\boldsymbol{\theta}) = -\int_{\Lambda\times\mathbb{M}} e^{-V(x^m|\varphi;\boldsymbol{\theta})}\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda} V(x^m|\varphi\setminus x^m;\boldsymbol{\theta}). \quad (4.2)$$

Our data consist in the realization of a point process with energy function $V(\cdot;\boldsymbol{\theta}^\star)$ satisfying **[Mod]**. Thus, $\boldsymbol{\theta}^\star$ is the true parameter to be estimated and it is assumed that $\boldsymbol{\theta}^\star \in \mathring{\boldsymbol{\Theta}}$. The Gibbs measure will be denoted by $P_{\boldsymbol{\theta}^\star}$. Moreover the point process is assumed to be observed in a domain $\Lambda_n \oplus D^\vee = \cup_{x\in\Lambda_n}\mathcal{B}(x,D^\vee)$ for some $D^\vee \geq D$. For the asymptotic normality result, it is also assumed that $\Lambda_n \subset \mathbb{R}^2$ can de decomposed into $\cup_{i\in I_n}\Delta_i$ where $I_n = \mathbb{B}(0,n)$ and for $i=(i_1,i_2)\in\mathbb{Z}^2$, $\Delta_i = \Delta_i(\widetilde{D})$ for some $\widetilde{D}>0$ fixed from now on. As a consequence, as $n\to+\infty$, $\Lambda_n\to\mathbb{R}^2$ such that $|\Lambda_n|\to+\infty$ and $\dfrac{|\partial\Lambda_n|}{|\Lambda_n|}\to 0$.

Define for any configuration φ, $U_n(\varphi;\boldsymbol{\theta}) = -\frac{1}{|\Lambda_n|}LPL_{\Lambda_n}(\varphi;\boldsymbol{\theta})$. The maximum pseudolikelihood estimate (MPLE) denoted by $\widehat{\boldsymbol{\theta}}_n(\varphi)$ is then defined by

$$\widehat{\boldsymbol{\theta}}_n(\varphi) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\max}\ LPL_{\Lambda_n}(\varphi;\boldsymbol{\theta}) = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min}\, U_n(\varphi;\boldsymbol{\theta}).$$

We will also need the following basic notations:

- Gradient vector of U_n: $\boldsymbol{U}_n^{(1)}(\varphi;\boldsymbol{\theta}) := -|\Lambda_n|^{-1}\boldsymbol{LPL}_{\Lambda_n}^{(1)}(\varphi;\boldsymbol{\theta})$ where for any bounded Borel set Λ, $\left(\boldsymbol{LPL}_\Lambda^{(1)}(\varphi;\boldsymbol{\theta})\right)_j$ is defined for $j=1,\ldots,p$ by

$$\left(\boldsymbol{LPL}_\Lambda^{(1)}(\varphi;\boldsymbol{\theta})\right)_j = \int_{\Lambda\times\mathbb{M}} v_j(x^m|\varphi)e^{-V(x^m|\varphi;\boldsymbol{\theta})}\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda} v_j(x^m|\varphi\setminus x^m)$$

- Hessian matrix of U_n: $\underline{\boldsymbol{U}}_n^{(2)}(\varphi;\boldsymbol{\theta}) := -|\Lambda_n|^{-1}\underline{\boldsymbol{LPL}}_{\Lambda_n}^{(2)}(\varphi;\boldsymbol{\theta})$ where for any bounded Borel set Λ, $\left(\underline{\boldsymbol{LPL}}_{\Lambda}^{(2)}(\varphi;\boldsymbol{\theta})\right)_{j,k}$ is defined for $j,k=1,\dots,p$ by

$$\left(\underline{\boldsymbol{LPL}}_{\Lambda}^{(2)}(\varphi;\boldsymbol{\theta})\right)_{j,k} = \int_{\Lambda\times\mathbb{M}} v_j(x^m|\varphi)v_k(x^m|\varphi)e^{-V(x^m|\varphi;\boldsymbol{\theta})}\mu(dx^m)$$

Finally, note that from the decomposition of the observation domain Λ_n, one has

$$\boldsymbol{U}_n^{(1)}(\varphi;\boldsymbol{\theta}) = |\Lambda_n|^{-1}\sum_{i\in I_n}\boldsymbol{LPL}_{\Delta_i}^{(1)}(\varphi;\boldsymbol{\theta})$$

and

$$\underline{\boldsymbol{U}}_n^{(2)}(\varphi;\boldsymbol{\theta}) = |\Lambda_n|^{-1}\sum_{i\in I_n}\underline{\boldsymbol{LPL}}_{\Delta_i}^{(2)}(\varphi;\boldsymbol{\theta})\,.$$

4.2. Asymptotic results of the MPLE

This section provides consistency and asymptotic normality of the maximum pseudolikelihood estimator. Let us first consider the following assumption

[Ident] Identifiability condition: there exists $A_1,\dots,A_\ell$, $\ell\geq p$ events of Ω and $A_1^{\mathrm{m}},\dots,A_\ell^{\mathrm{m}}$ events of $\mathcal{M}$ such that:

- the ℓ events $B_j := A_j^{\mathrm{m}}\times A_j$ are disjoint and satisfy $\lambda^{\mathrm{m}}\otimes P_{\boldsymbol{\theta}^\star}(B_j)>0$
- for all $((m_1,\varphi_1),\dots,(m_\ell,\varphi_\ell))\in B_1\times\cdots\times B_\ell$ the (ℓ,p) matrix with entries $v_j(0^{m_i}|\varphi_i)$ is injective.

Theorem 1. *Under the assumptions **[Mod]** and **[Ident]**, for $P_{\boldsymbol{\theta}^\star}-$almost every φ, the maximum pseudolikelihood estimate $\widehat{\boldsymbol{\theta}}_n(\varphi)$ converges towards $\boldsymbol{\theta}^\star$ as n tends to infinity.*

For the next result consider

[SDP] For some $\overline{\Lambda} := \cup_{i\in\mathbb{B}\left(0,\left\lceil\frac{D}{\overline{D}}\right\rceil\right)}\Delta_i(\overline{D})$ with $\overline{D}>0$, there exists $A_0,\dots,A_\ell$, $\ell\geq p$ disjoint events of $\overline{\Omega} := \left\{\varphi\in\Omega : \varphi_{\Delta_i(\overline{D})}=\emptyset, 1\leq|i|\leq 2\left\lceil\frac{D}{\overline{D}}\right\rceil\right\}$ such that

- for $j=0,\dots,\ell$, $P_{\boldsymbol{\theta}^\star}(A_j)>0$.
- for all $(\varphi_0,\dots,\varphi_\ell)\in A_0\times\cdots\times A_\ell$ the (ℓ,p) matrix with entries $\left(\boldsymbol{LPL}_{\overline{\Lambda}}^{(1)}(\varphi_i;\boldsymbol{\theta})\right)_j - \left(\boldsymbol{LPL}_{\overline{\Lambda}}^{(1)}(\varphi_0;\boldsymbol{\theta})\right)_j$ is injective.

Theorem 2. *Under the assumptions **[Mod]** and **[Ident]**, we have, for any fixed $\widetilde{D}$, the following convergence in distribution as $n\to+\infty$*

$$|\Lambda_n|^{1/2}\,\underline{\boldsymbol{U}}_n^{(2)}(\Phi;\boldsymbol{\theta}^\star)\,\left(\widehat{\boldsymbol{\theta}}_n(\Phi)-\boldsymbol{\theta}^\star\right)\to\mathcal{N}\left(0,\underline{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^\star)\right), \tag{4.3}$$

where

$$\underline{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^\star) = \sum_{i\in\mathbb{B}(0,\lceil D\rceil)} \boldsymbol{E}\left(\boldsymbol{LPL}^{(1)}_{\Delta_0(1)}(\Phi;\boldsymbol{\theta}^\star)\,\boldsymbol{LPL}^{(1)}_{\Delta_i(1)}(\Phi;\boldsymbol{\theta}^\star)^T\right). \tag{4.4}$$

In addition under the assumption ***[SDP]***

$$|\Lambda_n|^{1/2}\,\widehat{\underline{\boldsymbol{\Sigma}}}_n(\Phi;D^\vee,\widetilde{D},\widehat{\boldsymbol{\theta}}_n(\Phi))^{-1/2}\,\underline{\boldsymbol{U}}^{(2)}_n(\Phi;\widehat{\boldsymbol{\theta}}_n(\Phi))\,\left(\widehat{\boldsymbol{\theta}}_n(\Phi)-\boldsymbol{\theta}^\star\right)\to\mathcal{N}\left(0,\underline{\boldsymbol{I}}_p\right), \tag{4.5}$$

where for some $\boldsymbol{\theta}$ *and any configuration* φ*, the matrix* $\widehat{\underline{\boldsymbol{\Sigma}}}_n(\varphi;D^\vee,\widetilde{D},\boldsymbol{\theta})$ *is defined by*

$$\widehat{\underline{\boldsymbol{\Sigma}}}_n(\varphi;D^\vee,\widetilde{D},\boldsymbol{\theta}) = |\Lambda_n|^{-1}\sum_{i\in I_n}\sum_{j\in\mathbb{B}\left(i,\left\lceil\frac{D^\vee}{D}\right\rceil\right)\cap I_n}\boldsymbol{LPL}^{(1)}_{\Delta_i}(\varphi;\boldsymbol{\theta})\,\boldsymbol{LPL}^{(1)}_{\Delta_j}(\varphi;\boldsymbol{\theta})^T. \tag{4.6}$$

Remark 5. *Let us underline that the scaling that yields to asymptotic normality corresponds to the usual parametric rate. Indeed, in the d-dimensional case one would obtain*

$$|\Lambda_n|^{1/2}=|\Lambda_n(\widetilde{D})|^{1/2}=|I_n|^{1/2}\times|\Delta_i(\widetilde{D})|^{1/2}=\widetilde{D}^{d/2}((2n+1)^d)^{1/2}\sim(2\widetilde{D})^{d/2}(n^d)^{1/2}.$$

Remark 6. *We would like to underline that* $\underline{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^\star)=\underline{\boldsymbol{\Sigma}}(\widetilde{D},\boldsymbol{\theta}^\star)=\underline{\boldsymbol{\Sigma}}(1,\boldsymbol{\theta}^\star)$ *where for all* $\widetilde{D}>0$

$$\underline{\boldsymbol{\Sigma}}(\widetilde{D},\boldsymbol{\theta}^\star)=\widetilde{D}^{-2}\sum_{i\in\mathbb{B}\left(0,\left\lceil\frac{D}{D}\right\rceil\right)}\boldsymbol{E}\left(\boldsymbol{LPL}^{(1)}_{\Delta_0}(\Phi;\boldsymbol{\theta}^\star)\,\boldsymbol{LPL}^{(1)}_{\Delta_i}(\Phi;\boldsymbol{\theta}^\star)^T\right). \tag{4.7}$$

5. Back to examples

This section is devoted to proving that all of our examples satisfy both assumptions **[Ident]** and **[SDP]**. For the assumption **[Ident]** $\underline{\boldsymbol{V}}$ denotes the matrix with entries $v_j(0^{m_i}|\varphi_i)$ where $(m_i,\varphi_i)\in B_i$ have to be defined according to the different examples. The assumption **[SDP]** may be rewritten for all $k=1,\ldots,\ell$ and for all $\varphi_k\in A_k$ and $\varphi_0\in A_0$:

$$\left(\forall\boldsymbol{y}\in\mathbb{R}^p,\boldsymbol{y}^T\left(\boldsymbol{LPL}^{(1)}_{\overline{\Lambda}}(\varphi_k;\boldsymbol{\theta}^\star)-\boldsymbol{LPL}^{(1)}_{\overline{\Lambda}}(\varphi_0;\boldsymbol{\theta}^\star)\right)=\boldsymbol{y}^T(\boldsymbol{L}(\varphi_k;\boldsymbol{\theta}^\star)-\boldsymbol{R}(\varphi_k))=0\right)\Rightarrow\boldsymbol{y}=0,$$

where for any configuration $\varphi\in\overline{\Omega}$ and $\varphi_0\in A_0$

$$\begin{aligned}\boldsymbol{L}(\varphi;\boldsymbol{\theta}^\star) &:= \int_{\overline{\Lambda}\times\mathbb{M}}\boldsymbol{v}(x^m|\varphi)\,e^{-\boldsymbol{\theta}^{\star T}\boldsymbol{v}(x^m|\varphi)}\mu(dx^m)\\ &\quad-\int_{\overline{\Lambda}\times\mathbb{M}}\boldsymbol{v}(x^m|\varphi_0)\,e^{-\boldsymbol{\theta}^{\star T}\boldsymbol{v}(x^m|\varphi_0)}\mu(dx^m)\\ \boldsymbol{R}(\varphi) &:= \sum_{x^m\in\varphi\cap\overline{\Lambda}}\boldsymbol{v}(x^m|\varphi\setminus x^m)-\sum_{x^m\in\varphi_0\cap\overline{\Lambda}}\boldsymbol{v}(x^m|\varphi_0\setminus x^m).\end{aligned}$$

Concerning this assumption, we choose $\overline{D} > D$ in all our examples.

5.1. Overlap area point process

Assumption ***[Ident]***

Consider

$$\begin{aligned} A_1 &:= \left\{\varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \emptyset\right\} \\ A_2 &:= \left\{\varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \{z\}, z \in \mathcal{B}((0, D/2), D/4)\right\}. \end{aligned}$$

We have for all $(\varphi_1, \varphi_2) \in A_1 \times A_2$

$$\underline{\boldsymbol{V}} = \begin{pmatrix} 1 & 0 \\ 1 & v_2(0|\varphi_2) \end{pmatrix}.$$

For every $\varphi_2 \in A_2$ such that $(\varphi_2)_{\mathcal{B}(0,D)} = \{z\}$ with for all $z \in \mathcal{B}((0, D/2), D/4)$, one remarks that $|\mathcal{B}(0, R/2) \cap \mathcal{B}(z, R/2)| := g_2(||z||) > 0$, then $\det(\underline{\boldsymbol{V}}) \neq 0$.

Assumption ***[SDP]***

Denote by A_0 any configuration set. Consider $A_n(\eta)$ for $n \geq 1$ and for some $0 < \eta < D$ the following configuration set

$$A_n(\eta) = \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \{z_1, \ldots, z_n\} \text{ with } z_1, \ldots, z_n \in \mathcal{B}(0, \eta)\right\}.$$

For any $\varphi_n \in A_n(\eta)$, we have

$$\left| \int_{\overline{\Lambda}} v_j(x|\varphi_n) e^{-\boldsymbol{\theta}^{\star T} \boldsymbol{v}(x|\varphi_n)} dx \right| \leq \begin{cases} |\overline{\Lambda}| e^K & \text{if j=1} \\ n\pi R^2 |\overline{\Lambda}| e^K & \text{if j=2.} \end{cases}$$

where K comes from the local stability property. Let us also remark that

$$\boldsymbol{y}^T \boldsymbol{R}(\varphi_n) = n y_1 + y_2 \times \sum_{x \in \varphi_n \cap \overline{\Lambda}} v_2(x|\varphi_n \setminus x) - \boldsymbol{y}^T \sum_{x \in \varphi_0 \cap \overline{\Lambda}} \boldsymbol{v}(x|\varphi_0 \setminus x)$$

with

$$0 < n(n-1) g_2(\eta) \leq \sum_{x \in \varphi_n \cap \overline{\Lambda}} v_2(x|\varphi_n \setminus x) \leq n(n-1) \frac{\pi R^2}{4}.$$

Therefore by combining these arguments, for every $\varepsilon > 0$, we have for n large enough

$$\begin{aligned} |y_2| g_2(\eta) &\leq \left| \frac{1}{n(n-1)} y_2 \sum_{x \in \varphi_n \cap \overline{\Lambda}} v_2(x|\varphi_n \setminus x) \right| \\ &= \left| \frac{1}{n(n-1)} \left(y_2 \sum_{x \in \varphi_n \cap \overline{\Lambda}} v_2(x|\varphi_n \setminus x) + \boldsymbol{y}^T (\boldsymbol{L}(\varphi_n; \boldsymbol{\theta}^\star) - \boldsymbol{R}(\varphi_n)) \right) \right| \leq \varepsilon. \end{aligned}$$

By choosing $\varepsilon = \frac{|y_2|g_2(\eta)}{2}$, this leads to $y_2 = 0$. Then, for every $\varepsilon' > 0$ we may obtain for n large enough

$$|y_1| = \left| y_1 + \frac{1}{n}(y_1, 0)^T (\boldsymbol{L}(\varphi_n; \boldsymbol{\theta}^\star) - \boldsymbol{R}(\varphi_n)) \right| \leq \varepsilon'.$$

By choosing $\varepsilon' = |y_1|/2$, this leads to $y_1 = 0$.

5.2. Multi-Strauss marked type models

Assumption ***[Ident]***

Define for any $m, m_1, m_2 \in \{1, \ldots, M\}$ with $m_2 \geq m_1$ and for any $i = 2, \ldots, p^{m_1,m_2}$

$$\begin{aligned} A_0 &:= \left\{\varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \emptyset\right\} \\ A_m^{\mathrm{m}} &:= \{m\} \\ A_i^{m_1,m_2} &:= \left\{\varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \{z^{m_1}\}, \text{ with } z \in \mathcal{B}(0, D_i^{m_1,m_2}) \setminus \mathcal{B}(0, D_{i-1}^{m_1,m_2})\right\} \end{aligned}$$

The following events $B_i^{m_1,m_2}$ are defined for any $i = 1, \ldots, p^{m_1,m_1}$ when $m_1 = m_2$ and any $i = 2, \ldots, p^{m_1,m_2}$ when $m_1 < m_2$ such that:

$$B_i^{m_1,m_2} = \begin{cases} A_{m_1}^{\mathrm{m}} \times A_0 & \text{if } m_1 = m_2 \text{ and } i = 1 \\ A_{m_1}^{\mathrm{m}} \times A_i^{m_1,m_2} & \text{otherwise} \end{cases}$$

One may order these $\ell = p$ events as $B_1, \ldots, B_\ell$ where $B_{k_i^{m_1,m_2}} := B_i^{m_1,m_2}$ with

$$k_i^{m_1,m_2} = (m_1 - 1 + \delta_{m_1,m_2}) + \left(i - 1 + \sum_{m_2'=m_1}^{m_2-1} (p^{m_1,m_2'} - 1) + \sum_{m_1'=1}^{m_1-1} \sum_{m_2=m_1'}^{M} (p^{m_1',m_2} - 1) \right).$$

The corresponding matrix is then $\underline{\boldsymbol{V}} = \begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & V_M \end{pmatrix}$ with

$$V_m = \begin{pmatrix} 1 & V_{m,m} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 1 & 0 & \cdots & 0 & V_{m,M} \end{pmatrix} \text{ and } V_{m_1,m_2} = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ 1 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

*Assumption **[SDP]***

Let us first introduce the following sets for any $\eta > 0$ and $d > 0$

$$A(\eta, d) = \left\{(z_1, z_2) \in \Delta_0(\overline{D})^2 : z_1 \in \mathcal{B}\left((0,0), \frac{\eta}{4}\right) \text{ and } z_2 \in \mathcal{B}\left((d,0), \frac{3\eta}{4}\right)\right\}$$

$$\begin{aligned} A_-(\eta, d) &= \left\{(z_1, z_2) \in \Delta_0(\overline{D})^2 : z_1 \in \mathcal{B}\left((0,0), \frac{\eta}{4}\right) \text{ and } z_2 \in \mathcal{B}\left((d - \frac{\eta}{2}, 0), \frac{\eta}{4}\right)\right\} \\ &\subset A(\eta, d) \end{aligned}$$

$$\begin{aligned} A_+(\eta, d) &= \left\{(z_1, z_2) \in \Delta_0(\overline{D})^2 : z_1 \in \mathcal{B}\left((0,0), \frac{\eta}{4}\right) \text{ and } z_2 \in \mathcal{B}\left((d + \frac{\eta}{2}, 0), \frac{\eta}{4}\right)\right\} \\ &\subset A(\eta, d). \end{aligned}$$

For any $i \in \{2, \ldots, p^{m_1,m_2}\}$, when η is small enough, the couple of points $(z_1, z_2) \in A(\eta, D_i^{m_1,m_2})$ (resp. $A_-(\eta, D_i^{m_1,m_2})$ and $A_+(\eta, D_i^{m_1,m_2})$) are such that $D_{i-1}^{m_1,m_2} < D_i^{m_1,m_2} - \eta < d(z_1, z_2) < D_i^{m_1,m_2} + \eta < D_{i+1}^{m_1,m_2}$ (resp. $D_{i-1}^{m_1,m_2} < D_i^{m_1,m_2} - \eta < d(z_1, z_2) < D_i^{m_1,m_2}$ and $D_i^{m_1,m_2} < d(z_1, z_2) < D_i^{m_1,m_2} + \eta < D_{i+1}^{m_1,m_2}$).

We now derive the following events

$$A_0 := \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \emptyset\right\}$$

$$A_i^{m_1,m_2}(\eta) = \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \{z_1^{m_1}, z_2^{m_2}\} \text{ with } (z_1, z_2) \in A(\eta, D_i^{m_1,m_2})\right\}$$

$$\begin{aligned} A_{i,-}^{m_1,m_2}(\eta) &= \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \{z_1^{m_1}, z_2^{m_2}\} \text{ with } (z_1, z_2) \in A_-(\eta, D_i^{m_1,m_2})\right\} \\ &\subset A_i^{m_1,m_2}(\eta) \end{aligned}$$

$$\begin{aligned} A_{i,+}^{m_1,m_2}(\eta) &= \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \{z_1^{m_1}, z_2^{m_2}\} \text{ with } (z_1, z_2) \in A_+(\eta, D_i^{m_1,m_2})\right\} \\ &\subset A_i^{m_1,m_2}(\eta). \end{aligned}$$

First of all note that for any $\varphi_0 \in A_0$,

$$\boldsymbol{y}^T \boldsymbol{L}\boldsymbol{P}\boldsymbol{L}_{\overline{\Lambda}}^{(1)}(\varphi_0; \boldsymbol{\theta}^\star) = \sum_{m=1}^{M} y_1^{m,m} e^{-\theta_1^{\star,m,m}} |\overline{\Lambda}| - 0$$

For any $\varphi_{i,-}^{m_1,m_2} \in A_{i,-}^{m_1,m_2}$ and any $\varphi_{i,+}^{m_1,m_2} \in A_{i,+}^{m_1,m_2}$ for $i = 2, \ldots, p^{m_1,m_2}$

$$\begin{aligned} \boldsymbol{y}^T \boldsymbol{R}(\varphi_{i,-}^{m_1,m_2}) &= y_1^{m_1,m_1} + y_1^{m_2,m_2} + 2y_i^{m_1,m_2} \\ \boldsymbol{y}^T \boldsymbol{R}(\varphi_{i,+}^{m_1,m_2}) &= y_1^{m_1,m_1} + y_1^{m_2,m_2} + 2y_i^{m_1,m_2}(1 - \delta_{i,p^{m_1,m_2}}) \end{aligned}$$

We leave the reader to check that for every $\varepsilon > 0$ there exists $\eta > 0$ small enough such that

$$\left|\boldsymbol{y}^T \boldsymbol{L}(\varphi_{i,+}^{m_1,m_2}; \boldsymbol{\theta}^\star) - \boldsymbol{y}^T \boldsymbol{L}(\varphi_{i,-}^{m_1,m_2}; \boldsymbol{\theta}^\star)\right| \leq \varepsilon.$$

Therefore for every $\varepsilon > 0$ we have for η small enough

$$\begin{aligned} 2|y_{p^{m_1,m_2}}^{m_1,m_2}| &= \Big| \boldsymbol{y}^T \left(\boldsymbol{L}(\varphi_{p^{m_1,m_2},+}^{m_1,m_2}; \boldsymbol{\theta}^\star) - \boldsymbol{R}(\varphi_{p^{m_1,m_2},+}^{m_1,m_2}) \right) \\ &\quad - \boldsymbol{y}^T \left(\boldsymbol{L}(\varphi_{p^{m_1,m_2},-}^{m_1,m_2}; \boldsymbol{\theta}^\star) - \boldsymbol{R}(\varphi_{p^{m_1,m_2},-}^{m_1,m_2}) \right) + 2y_{p^{m_1,m_2}}^{m_1,m_2} \Big| \leq \varepsilon. \end{aligned}$$

By choosing $\varepsilon = |y_{p^{m_1,m_2}}^{m_1,m_2}|$, this leads to $y_{p^{m_1,m_2}}^{m_1,m_2} = 0$. By iterating this argument, we obtain that for any $m_1, m_2 \in \{1, \ldots, M\}$, $y_2^{m_1,m_2} = \cdots = y_{p^{m_1,m_2}}^{m_1,m_2} = 0$. It remains to prove that $y_1^{1,1} = \cdots = y_1^{M,M} = 0$. For this, consider the following configuration set indexed by $n \geq 1$

$$A_n^{m_1} = \left\{ \varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \{z_1^{m_1}, \ldots, z_n^{m_1}\} \text{ with } z_1, \ldots, z_n \in \Delta_0(\overline{D}) \right\}.$$

For any $\varphi_n^{m_1} \in A_n^{m_1}$, we have

$$\begin{aligned} \boldsymbol{y}^T \boldsymbol{L}(\varphi_n^{m_1,m_1}; \boldsymbol{\theta}^\star) &= \sum_{m_1=1}^{M} y_1^{m_1,m_1} \int_{\overline{\Lambda} \times \mathbb{M}} e^{-V\left(x^m | \varphi_n^{m_1}; \boldsymbol{\theta}^\star\right)} \mu(dx^m) - |\overline{\Lambda}| e^{-\theta_1^{\star,m_1,m_1}} \\ \boldsymbol{y}^T \boldsymbol{R}(\varphi_n^{m_1,m_1}) &= n y_1^{m_1,m_1} \end{aligned}$$

Hence for every $\varepsilon > 0$ we have for n large enough by using the local stability property

$$\begin{aligned} |y_1^{m_1,m_1}| &= \left| \frac{1}{n} \boldsymbol{y}^T \left(\boldsymbol{L}(\varphi_n^{m_1,m_2}; \boldsymbol{\theta}^\star) + \boldsymbol{R}(\varphi_n^{m_1,m_2}) \right) + y_1^{m_1,m_1} \right| \\ &\leq \frac{2|\overline{\Lambda}|}{n} e^K \sum_{m_1=1}^{M} |y_1^{m_1,m_1}| \leq \varepsilon. \end{aligned}$$

By choosing $\varepsilon = |y_1^{m_1,m_1}|/2$, this leads to $y_1^{m_1,m_1} = 0$.

5.3. k−nearest-neighbour multi-Strauss marked point process

Assumption **[Ident]** (resp. **[SDP]**) is proven without any change in the proof of the multi-Strauss marked point process for every $k \geq 1$ (resp. $k \geq 2$). The proof of **[SDP]** for $k = 1$ is omitted.

5.4. Strauss disc process

*Assumption **[Ident]***

Consider

$$\begin{aligned} A_1 &:= \left\{ \varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \emptyset \right\} \\ A_2 &:= \left\{ \varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \{z^{m'}\},\ z \in \mathcal{B}(0, D/2),\ m' \in [D/2, D] \right\} \end{aligned}$$

and define $B_1 := \mathbb{M} \times A_1$ and $B_2 := \mathbb{M} \times A_2$. Then, for any $(m_1, \varphi_1, m_2, \varphi_2) \in B_1 \times B_2$

$$\underline{\boldsymbol{V}} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

which is injective since $\det(\underline{\boldsymbol{V}}) = 1$.

*Assumption **[SDP]***

Consider for $n \geq 1$

$$\begin{aligned} A_0 &:= \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \emptyset\right\} \\ A_n &:= \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \{z_1^{m_1}, \ldots, z_n^{m_n}\},\right. \\ &\qquad \left. z_1, \ldots, z_n \in \mathcal{B}(0, D/2),\ m_1, \ldots, m_n \in [D/2, D]\right\}. \end{aligned}$$

Note that for every $\varphi_0 \in A_0$ and any $\varphi_n \in A_n$

$$\boldsymbol{y}^T \boldsymbol{R}(\varphi_n) = n y_1 + n(n-1) y_2$$

Note also that from the local stability property $|\boldsymbol{y}^T \boldsymbol{L}(\varphi_n; \boldsymbol{\theta}^\star) \leq 2(|y_1| + |y_2|)|\overline{\Lambda}| e^K$. Then, for every $\varepsilon > 0$ we have for n large enough

$$|y_2| = \left| \frac{1}{n(n-1)} \boldsymbol{y}^T \left(\boldsymbol{L}(\varphi_n; \boldsymbol{\theta}^\star) - \boldsymbol{R}_n(\varphi_n)\right) + y_2 \right| \leq \varepsilon.$$

By choosing $\varepsilon = |y_2|/2$, this leads to $y_2 = 0$. Then, for every $\varepsilon' > 0$ we have for n large enough

$$|y_1| = \left| \frac{1}{n} (y_1, 0)^T \left(\boldsymbol{L}(\varphi_n; \boldsymbol{\theta}^\star) - \boldsymbol{R}_n(\varphi_n)\right) + y_1 \right| \leq \varepsilon'.$$

By choosing $\varepsilon' = |y_1|/2$, this leads to $y_1 = 0$.

5.5. Geyer's triplet point process

*Assumption **[Ident]***

By considering

$$A_1 := \left\{\varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \emptyset\right\},$$

$$A_2 := \left\{\varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \{z\},\ z \in \mathcal{B}((0, D/2), D/4), \varphi_{\mathcal{B}(0,2D)\setminus\mathcal{B}(0,D)} = \emptyset\right\}$$

and

$$A_3 := \left\{\varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \{z_1, z_2\}, z_1, z_2 \in \mathcal{B}((0, D/2), D/4), \varphi_{\mathcal{B}(0,2D)\setminus\mathcal{B}(0,D)} = \emptyset\right\},$$

we have for any $(\varphi_1, \varphi_2, \varphi_3) \in A_1 \times A_2 \times A_3$

$$\underline{\boldsymbol{V}} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix},$$

which is clearly injective since $\det \underline{\boldsymbol{V}} = 1$.

*Assumption **[SDP]***

Denote by A_0 any configuration set. Consider $A_n(\eta)$ for $n \geq 1$ and for some $0 < \eta < D$ the following configuration set

$$A_n(\eta) = \left\{ \varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \{z_1, \ldots, z_n\} \text{ with } z_1, \ldots, z_n \in \mathcal{B}(0, \eta) \right\}.$$

We leave the reader to chek that for $j = 1, \ldots, p$ and for any $\varphi_n \in A_n(\eta)$

$$\left| \int_{\Lambda} v_j(x|\varphi_n) e^{-\boldsymbol{\theta}^{\star T} \boldsymbol{v}(x|\varphi_n)} dx \right| \leq n^{j-1} |\overline{\Lambda}| e^K,$$

where K comes from the local stability property. Let us also remark that

$$\boldsymbol{y}^T \boldsymbol{R}(\varphi_n) = n y_1 + n(n-1) y_2 + \frac{n(n-1)(n-2)}{2} y_3 - \boldsymbol{y}^T \sum_{x \in \varphi_0 \cap \overline{\Lambda}} \boldsymbol{v}\left(x|\varphi_0 \setminus x\right).$$

Therefore by combining these two arguments, for every $\varepsilon > 0$, we have for n large enough

$$\frac{|y_3|}{2} = \left| \frac{y_3}{2} + \frac{\boldsymbol{y}^T (\boldsymbol{L}(\varphi_n; \boldsymbol{\theta}^\star) - \boldsymbol{R}(\varphi_n))}{n(n-1)(n-2)} \right| \leq \varepsilon.$$

By choosing $\varepsilon = \frac{|y_3|}{4}$, this leads to $y_3 = 0$. Then, for every ε', we may obtain for n large enough

$$|y_2| = \left| y_2 + \frac{(y_1, y_2, 0)^T (\boldsymbol{L}(\varphi_n; \boldsymbol{\theta}^\star) - \boldsymbol{R}(\varphi_n))}{n(n-1)} \right| \leq \varepsilon'.$$

By choosing $\varepsilon' = |y_2|/2$, this leads to $y_2 = 0$. And then for every ε'' for n large enough

$$|y_1| = \left| y_1 + \frac{(y_1, 0, 0)^T (\boldsymbol{L}(\varphi_n; \boldsymbol{\theta}^\star) - \boldsymbol{R}(\varphi_n))}{n} \right| \leq \varepsilon''.$$

By choosing $\varepsilon'' = \frac{|y_1|}{2}$, this finally leads to $y_1 = 0$.

5.6. Area-interaction model

Assumption ***[Ident]***

By considering

$$A_1 := \left\{\varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \emptyset\right\}$$

and for some small $\eta > 0$

$$A_2 := A_2(\eta) = \left\{\varphi \in \Omega : \varphi_{\mathcal{B}(0,D)} = \{z\}, z \in \mathcal{B}(0,\eta)\right\}.$$

we obtain for any $(\varphi_1, \varphi_2) \in A_1 \times A_2$

$$\underline{\boldsymbol{V}} = \begin{pmatrix} 1 & \pi R^2 \\ 1 & v_2(0|\varphi_2) \end{pmatrix}.$$

Since for any $\eta < R$, $0 < v_2(0|\varphi_2) < \pi R^2$, $\det(\underline{\boldsymbol{V}}) \neq 0$.

Assumption ***[SDP]***

Consider $A_k(\eta)$ for $k = 0, 1, 2$ and for some $0 < \eta < D$ the following configuration set

$$A_k(\eta) = \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0(\overline{D})} = \{z_1, \ldots, z_{k+1}\} \text{ with } z_1, \ldots, z_{k+1} \in \mathcal{B}(0,\eta)\right\}$$

For any $\varphi_k \in A_k(\eta)$ for $k = 0, 1, 2$,

$$\begin{aligned} R(\varphi_1) &= 2y_1 + y_2 g_1(\varphi_1)) - (y_1 + y_2 \pi R^2) \\ R(\varphi_2) &= 3y_1 + y_2 g_2(\varphi_2) - (y_1 + y_2 \pi R^2) \end{aligned}$$

For every $\varepsilon > 0$, there exists $\eta > 0$ such that for any $\varphi_k \in A_k(\eta)$ $(k = 1, 2)$, $|\boldsymbol{y}^T \boldsymbol{L}(\varphi_k; \boldsymbol{\theta}^\star)| \leq \varepsilon$ and such that $|y_2 g_k(\varphi_k)| \leq \varepsilon$. Then,

$$|y_1| = \left|y_1 - \boldsymbol{y}^T (L(\varphi_1; \boldsymbol{\theta}^\star) - L_2(\varphi_2; \boldsymbol{\theta}^\star)) - (R(\varphi_1) - R(\varphi_2))\right| \leq 4\varepsilon.$$

By choosing $\varepsilon = |y_1|/8$, this leads to $y_1 = 0$. Now,

$$|y_2|\pi R^2 = \left|y_2 \pi R^2 - (0, y_2)^T (L(\varphi_1; \boldsymbol{\theta}^\star) - R(\varphi_1))\right| \leq 2\varepsilon.$$

And by choosing $\varepsilon = \pi R^2 |y_2|/4$, this finally leads to $y_2 = 0$.

6. Annex: Proofs

6.1. Tools

Let us start by presenting a particular case of the Campbell Theorem combined with the Glötzl Theorem that is widely used in our future proofs.

Corollary 3. *Assume that the (marked) point process Φ with probability measure P is stationary. Let Λ be a bounded Borel set, let $\varphi \in \Omega$ and let g be a function satisfying $g(x^m, \varphi_x) = g(0^m, \varphi)$ for all $x^m \in \mathbb{S}$. Define M a random variable with its distribution λ^{m} and $f(m, \varphi) = g(0^m, \varphi)e^{-V(0^m|\varphi)}$ and assume that $f \in L^1(\lambda^{\mathrm{m}} \otimes P)$. Then,*

$$\boldsymbol{E}\Big(\sum_{x^m \in \Phi_\Lambda} g(x^m, \Phi \setminus x^m)\Big) = |\Lambda|\, \boldsymbol{E}\Big(\, g\left(0^M, \Phi\right) e^{-V\left(0^M|\Phi\right)} \, \Big) \tag{6.1}$$

Proof.

$$\begin{aligned}
\boldsymbol{E}\Big(\sum_{x^m \in \Phi_\Lambda} g(x^m, \Phi \setminus x^m)\Big) &= \int_\Omega \sum_{x^m \in \varphi} g(x^m, \varphi \setminus x^m)\mathbb{1}_\Lambda(x) P(d\varphi) \\
&= \int_{\mathbb{S}} \int_\Omega g(x^m, \varphi)\mathbb{1}_\Lambda(x) P^!_{x^m}(d\varphi) N_P(dx^m) \\
&= \int_{\mathbb{S}} \int_\Omega g(x^m, \varphi)\mathbb{1}_\Lambda(x) \nu_P(x^m) P^!_{x^m}(d\varphi)\mu(dx^m) \\
&= \int_{\mathbb{S}} \int_\Omega g(x^m, \varphi)\mathbb{1}_\Lambda(x) e^{-V(x^m|\varphi)} P(d\varphi)\mu(dx^m) \\
&= \int_{\Lambda\times\mathbb{M}} \int_\Omega g(x^m, \varphi) e^{-V(x^m|\varphi)} P(d\varphi)\mu(dx^m) \\
&= |\Lambda| \int_{\mathbb{M}} \int_\Omega g(x^m, \varphi) e^{-V(x^m|\varphi)} P(d\varphi)\lambda^{\mathrm{m}}(dm) \\
&= |\Lambda| \boldsymbol{E}\Big(g(0^M, \Phi) e^{-V\left(0^M|\Phi\right)} \Big)
\end{aligned}$$

where $\nu_P(\cdot)$ is the Radon-Nikodym derivative of N_P with respect to μ. ■

Let us now present a version of an ergodic theorem obtained by Nguyen and Zessin (1979) and widely used in this paper. Let $\widetilde{D} > 0$ and denote by Δ_0 the following fixed domain

Theorem 4 (Nguyen and Zessin (1979)). *Let $\{H_G, G \in \mathcal{B}_b^2\}$ be a family of random variables, which is covariant, that for all $x \in \mathbb{R}^2$,*

$$H_{G_x}(\varphi_x) = H_G(\varphi), \quad \textit{for a.e. } \varphi$$

and additive, that is for every disjoint $G_1, G_2 \in \mathcal{B}_b^2$,

$$H_{G_1 \cup G_2} = H_{G_1} + H_{G_2}, \quad a.s.$$

Let $\mathcal{I}$ be the sub$-\sigma-$algebra of $\mathcal{F}$ consisting of translation invariant (with probability 1) sets. Assume there exists a nonnegative and integrable random variable Y such that $|H_G| \le Y$ a.s. for every convex $G \subset \Delta_0$. Then,

$$\lim_{n\to+\infty} \frac{1}{|G_n|} H_{G_n} = \frac{1}{|\Delta_0|} E(H_{\Delta_0}|\mathcal{I}), \quad a.s.$$

for each regular sequence $G_n \to \mathbb{R}^2$.

6.2. Proof of Theorem 1

Due to the decomposition of stationary measures as a mixture of ergodic measures (see Preston (1976)), one only needs to prove Theorem 1 by assuming that $P_{\boldsymbol{\theta}^\star}$ is ergodic. From now on, $P_{\boldsymbol{\theta}^\star}$ is assumed to be ergodic. The tool used to obtain the almost sure convergence is a convergence theorem for minimum contrast estimators established by Guyon (1995).
We proceed in three stages.
Step 1. Convergence of $U_n(\Phi;\boldsymbol{\theta})$.
Decompose $U_n(\varphi;\boldsymbol{\theta}) = \frac{1}{|\Lambda_n|}\left(H_{1,\Lambda_n}(\varphi) + H_{2,\Lambda_n}(\varphi)\right)$ with

$$H_{1,\Lambda_n}(\varphi) = \int_{\Lambda_n\times\mathbb{M}} e^{-V(x^m|\varphi;\boldsymbol{\theta})}\mu(dx^m)$$

and

$$H_{2,\Lambda_n}(\varphi) = \sum_{x^m\in\Phi_{\Lambda_n}} V\left(x^m|\varphi\setminus x^m;\boldsymbol{\theta}\right).$$

Under the assumption **[Mod]**, one can apply Theorem 4 (Nguyen and Zessin (1979)) to the process H_{1,Λ_n}. And from Corollary 3, we obtain $P_{\boldsymbol{\theta}^\star}-$almost surely as $n\to+\infty$

$$\frac{1}{|\Lambda_n|}H_{1,\Lambda_n} \to \boldsymbol{E}\Big(\int_{\mathbb{M}} e^{-V(0^m|\Phi;\boldsymbol{\theta})}\lambda^{\mathbb{m}}(dm)\Big) = \boldsymbol{E}\Big(e^{-V\left(0^M|\Phi;\boldsymbol{\theta}\right)}\Big). \tag{6.2}$$

Now, let $G\subset\Delta_0$, we clearly have

$$|H_{2,G}(\varphi)| \le \sum_{x^m\in\Phi_G} |V\left(x^m|\varphi\setminus x^m;\boldsymbol{\theta}\right)| \le \sum_{x^m\in\varphi_{\Delta_0}} |V\left(x^m|\varphi\setminus x^m;\boldsymbol{\theta}\right)|$$

Under the assumption **[Mod]** and from Corollary 3, we have

$$\boldsymbol{E}\left(\sum_{x^m\in\Phi_{\Delta_0}} |V\left(x^m|\Phi\setminus x^m;\boldsymbol{\theta}\right)|\right) = |\Delta_0|\boldsymbol{E}\left(|V\left(0^M|\Phi;\boldsymbol{\theta}\right)|e^{-V\left(0^M|\Phi;\boldsymbol{\theta}^\star\right)}\right) < +\infty$$

This means that for all $G\subset\Delta_0$, there exists a random variable $Y\in L^1(P_{\boldsymbol{\theta}^\star})$ such that $|H_{2,G}|\le Y$. Thus, under the ssumption **[Mod]** and from Theorem 4 (Nguyen and Zessin (1979)) and from Corollary 3, we have $P_{\boldsymbol{\theta}^\star}-$almost surely

$$\begin{aligned}\frac{1}{|\Lambda_n|}H_{2,\Lambda_n} &\to \frac{1}{|\Delta_0|}\boldsymbol{E}\Big(\sum_{x^m\in\Phi_{\Delta_0}} V\left(x^m|\Phi\setminus x^m;\boldsymbol{\theta}\right)\Big)\\ &= \boldsymbol{E}\left(V\left(0^M|\Phi;\boldsymbol{\theta}\right)e^{-V\left(0^M|\Phi;\boldsymbol{\theta}^\star\right)}\right).\end{aligned} \tag{6.3}$$

We have the result by combining (6.2) and (6.3): $P_{\boldsymbol{\theta}^\star}-$almost surely

$$U_n(\cdot;\boldsymbol{\theta}) \to U(\boldsymbol{\theta}) = \boldsymbol{E}\Big(e^{-V\left(0^M|\Phi;\boldsymbol{\theta}\right)} + V\left(0^M|\Phi;\boldsymbol{\theta}\right)e^{-V\left(0^M|\Phi;\boldsymbol{\theta}^\star\right)}\Big) \tag{6.4}$$

Step 2. $U_n(\Phi;\cdot)$ a contrast function

Recall that $U_n(\cdot;\boldsymbol{\theta})$ is a contrast function if there exists a function $K(\cdot,\boldsymbol{\theta}^\star)$ (i.e. nonnegative function equal to zero if and only if $\boldsymbol{\theta}=\boldsymbol{\theta}^\star$) such that $P_{\boldsymbol{\theta}^\star}-$almost surely $U_n(\varphi;\boldsymbol{\theta})-U_n(\varphi;\boldsymbol{\theta})\to K(\boldsymbol{\theta},\boldsymbol{\theta}^\star)$. From Step 1, we have as $n\to+\infty$

$$\begin{aligned} K(\boldsymbol{\theta},\boldsymbol{\theta}^\star) &= \boldsymbol{E}\Big(e^{-V\left(0^M|\Phi;\boldsymbol{\theta}^\star\right)}\Big(e^{V\left(0^M|\Phi;\boldsymbol{\theta}\right)-V\left(0^M|\Phi;\boldsymbol{\theta}^\star\right)} \\ &\quad -\Big(1+V\left(0^M|\Phi;\boldsymbol{\theta}\right)-V\left(0^M|\Phi;\boldsymbol{\theta}^\star\right)\Big)\Big)\Big) \\ &= \boldsymbol{E}\Big(e^{-\boldsymbol{\theta}^{\star T}\boldsymbol{v}\left(0^M|\varphi\right)}\Big(e^{(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\boldsymbol{v}\left(0^M|\Phi\right)}-(1+(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\boldsymbol{v}\left(0^M|\Phi\right))\Big)\Big). \end{aligned} \tag{6.5}$$

Let $\boldsymbol{y}\in\mathbb{R}^{p+1}\setminus\{0\}$, and assume $\boldsymbol{y}^T\boldsymbol{v}\left(0^m|\varphi\right)=0$ for $\lambda^{\mathrm{m}}\otimes P_{\boldsymbol{\theta}^\star}-$a.e. (m,φ). By assuming [**Ident**], it follows that for $i=1,\ldots,\ell$ $(\ell\geq p)$ $\boldsymbol{y}^T\boldsymbol{v}\left(0^{m_i}|\varphi_i\right)=0$ for all $(m_i,\varphi_i)\in B_i$. From the injectivity of the matrix with entries $v_j(0^{m_i}|\varphi_i)$ for all $(m_1,\varphi_1,\ldots,m_\ell,\varphi_\ell)\in B_1\times\ldots\times B_\ell$ it comes that $\boldsymbol{y}=0$ which leads to a contradiction. Therefore, for $\boldsymbol{\theta}\neq\boldsymbol{\theta}^\star$, the assertion $(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\boldsymbol{v}\left(0^m|\varphi\right)\neq 0$ holds for $\lambda^{\mathrm{m}}\otimes P_{\boldsymbol{\theta}^\star}-$a.e. (m,φ).

By noticing that the function $t\mapsto e^t-(1+t)$ is nonnegative and is equal to zero if and only if $t=0$, one concludes that the random variable $e^{-\boldsymbol{\theta}^{\star T}\boldsymbol{v}\left(0^M|\Phi\right)}\times\Big(e^{(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\boldsymbol{v}\left(0^M|\Phi\right)}-(1+(\boldsymbol{\theta}-\boldsymbol{\theta}^\star)^T\boldsymbol{v}\left(0^M|\Phi\right))\Big)$ is almost surely positive for $\boldsymbol{\theta}\neq\boldsymbol{\theta}^\star$ and equals to zero when $\boldsymbol{\theta}=\boldsymbol{\theta}^\star$.

Before ending this step, note that for any φ, $U_n(\varphi;\cdot)$ and $K(\cdot,\boldsymbol{\theta}^\star)$ are clearly continuous functions.

Step 3. Modulus of continuity.

The modulus of continuity of the contrast process defined for all $\varphi\in\Omega$ and all $\eta>0$ by

$$W_n(\varphi,\eta)=\sup\left\{\left|U_n(\varphi;\boldsymbol{\theta})-U_n(\varphi;\boldsymbol{\theta}')\right|:\boldsymbol{\theta},\boldsymbol{\theta}'\in\boldsymbol{\Theta},||\boldsymbol{\theta}-\boldsymbol{\theta}'||\leq\eta\right\}$$

is such that there exists a sequence $(\varepsilon_k)_{k\geq 1}$, with $\varepsilon_k\to 0$ as $k\to+\infty$ such that for all $k\geq 1$

$$P\left(\limsup_{n\to+\infty}\left(W_n\left(\Phi,\frac{1}{k}\right)\geq\varepsilon_k\right)\right)=0. \tag{6.6}$$

Let us start to write $W_n\left(\varphi,\frac{1}{k}\right)\leq W_{1,n}\left(\varphi,\frac{1}{k}\right)+W_{2,n}\left(\varphi,\frac{1}{k}\right)$ with

$$\begin{aligned} W_{1,n}\left(\varphi,\frac{1}{k}\right) &= \sup\left\{\left|\frac{1}{|\Lambda_n|}\int_{\Lambda_n\times\mathbb{M}}\left(e^{-V(x^m|\varphi;\boldsymbol{\theta})}-e^{-V\left(x^m|\varphi;\boldsymbol{\theta}'\right)}\right)\mu(dx^m)\right|\right. \\ &\quad \left.:\boldsymbol{\theta},\boldsymbol{\theta}'\in\boldsymbol{\Theta},||\boldsymbol{\theta}-\boldsymbol{\theta}'||\leq\frac{1}{k}\right\} \end{aligned}$$

and

$$W_{2,n}\left(\varphi,\frac{1}{k}\right) = \sup\left\{\Big|\sum_{x^m\in\varphi_{\Lambda_n}} V\left(x^m|\varphi\setminus x^m;\boldsymbol{\theta}\right)-V\left(x^m|\varphi\setminus x^m;\boldsymbol{\theta}'\right)\Big| : \boldsymbol{\theta},\boldsymbol{\theta}'\in\boldsymbol{\Theta}, ||\boldsymbol{\theta}-\boldsymbol{\theta}'||\leq\frac{1}{k}\right\}.$$

From the assumption **[Mod]** it comes

$$\begin{aligned} W_{1,n}\left(\varphi,\frac{1}{k}\right) &\leq \sup\left\{\frac{1}{|\Lambda_n|}\int_{\Lambda_n\times\mathbb{M}}\left(|(\boldsymbol{\theta}-\boldsymbol{\theta}')^T\boldsymbol{v}\left(x^m|\varphi\right)|e^K\right)\mu(dx^m) : \boldsymbol{\theta},\boldsymbol{\theta}'\in\boldsymbol{\Theta}, ||\boldsymbol{\theta}-\boldsymbol{\theta}'||\leq\frac{1}{k}\right\} \\ &\leq \frac{e^K}{k}\frac{1}{|\Lambda_n|}\int_{\Lambda_n\times\mathbb{M}}||\boldsymbol{v}\left(x^m|\Phi\right)||\mu(dx^m). \end{aligned}$$

Under the assumption **[Mod]**, one can apply Theorem 4 (Nguyen and Zessin (1979)) to obtain as $n\to+\infty$

$$\frac{1}{|\Lambda_n|}\int_{\Lambda_n\times\mathbb{M}}||\boldsymbol{v}\left(x^m|\Phi\right)||\mu(dx^m)\to\boldsymbol{E}\left(||\boldsymbol{v}\left(0^M|\Phi\right)||\right),\ \text{for } P_{\boldsymbol{\theta}^\star}-a.e.\,\varphi.$$

So there exists $n_0^{(1)}(k)$ such that for all $n\geq n_0^{(1)}(k)$ we have

$$W_{1,n}\left(\varphi,\frac{1}{k}\right)\leq\frac{2e^K}{k}\,\boldsymbol{E}\left(||\boldsymbol{v}\left(0^M|\Phi\right)||\right),$$

for $P_{\boldsymbol{\theta}^\star}$-a.e. φ. By using the same arguments, one can prove that there exists $n_0^{(2)}(k)$ such that for all $n\geq n_0^{(2)}(k)$ we have

$$W_{2,n}\left(\varphi,\frac{1}{k}\right)\leq\frac{2}{k}\,\boldsymbol{E}\left(||\boldsymbol{v}\left(0^M|\Phi\right)||e^{-V\left(0^M|\Phi;\boldsymbol{\theta}^\star\right)}\right)\leq\frac{2e^K}{k}\,\boldsymbol{E}\left(||\boldsymbol{v}\left(0^M|\Phi\right)||\right),$$

for $P_{\boldsymbol{\theta}^\star}-$a.e. φ. And so for all $n\geq n_0(k)=\max(n_0^{(1)}(k),n_0^{(2)}(k))$, we have $P_{\boldsymbol{\theta}^\star}-$a.s.

$$W_n\left(\varphi,\frac{1}{k}\right)\leq W_{1,n}\left(\varphi,\frac{1}{k}\right)+W_{2,n}\left(\varphi,\frac{1}{k}\right)<\frac{\delta}{k},\ \text{for } P_{\boldsymbol{\theta}^\star}-a.e.\,\varphi.$$

with $\delta=4e^K\boldsymbol{E}\left(||\boldsymbol{v}\left(0^M|\Phi\right)||\right)$. Since,

$$\begin{aligned} \limsup_{n\to+\infty}\left\{W_n\left(\varphi,\frac{1}{k}\right)\geq\frac{\delta}{k}\right\} &= \bigcap_{m\in\mathbb{N}}\bigcup_{n\geq m}\left\{W_n\left(\varphi,\frac{1}{k}\right)\geq\frac{\delta}{k}\right\} \\ &\subset \bigcup_{n\geq n_0(k)}\left\{W_n\left(\varphi,\frac{1}{k}\right)\geq\frac{\delta}{k}\right\},\ \text{for } P_{\boldsymbol{\theta}^\star}-a.e.\,\varphi. \end{aligned}$$

the expected result (6.6) is proved.

Conclusion step. The Steps 1, 2 and 3 ensure the fact that we can apply a consistency result on minimum contrast estimators, see Guyon (1995).

6.3. Proof of Theorem 2

Step 1. Asymptotic normality of $\boldsymbol{U}_n^{(1)}(\Phi;\boldsymbol{\theta}^\star)$

The aim is to prove that for any fixed $\widetilde{D}$, the following convergence in distribution as $n \to +\infty$

$$|\Lambda_n|^{1/2}\, \boldsymbol{U}_n^{(1)}(\Phi;\boldsymbol{\theta}^\star) \to \mathcal{N}\left(0, \underline{\boldsymbol{\Sigma}}(\widetilde{D},\boldsymbol{\theta}^\star)\right) \tag{6.7}$$

where the matrix $\underline{\boldsymbol{\Sigma}}(\widetilde{D},\boldsymbol{\theta}^\star)$ is defined by (4.4).

The idea is to apply to $\boldsymbol{U}_n^{(1)}(\Phi;\boldsymbol{\theta}^\star)$ a central limit theorem obtained by Jensen and Künsch (1994), Theorem 2.1. The following conditions have to be fulfilled to apply this result. For all $j = 1,\dots,p+1$

(*i*) For all $i \in \mathbb{Z}^2$, $\boldsymbol{E}\left(\left(\boldsymbol{LPL}_{\Delta_i}^{(1)}(\Phi;\boldsymbol{\theta}^\star)\right)_j | \Phi_{\Delta_i^c}\right) = 0$.

(*ii*) For all $i \in \mathbb{Z}^2$, $\boldsymbol{E}\left(\left|\left(\boldsymbol{LPL}_{\Delta_i}^{(1)}(\Phi;\boldsymbol{\theta}^\star)\right)_j\right|^3\right) < +\infty$.

(*iii*) The matrix $\mathbb{V}\mathrm{ar}\left(|\Lambda_n|^{1/2}\boldsymbol{U}_n^{(1)}(\Phi;\boldsymbol{\theta}^\star)\right)$ converges to the matrix $\underline{\boldsymbol{\Sigma}}(\widetilde{D},\boldsymbol{\theta}^\star)$.

Condition (*i*): From the stationarity of the process, it is sufficient to prove that

$$\boldsymbol{E}\left(\left(\boldsymbol{LPL}_{\Delta_0}^{(1)}(\Phi;\boldsymbol{\theta}^\star)\right)_j | \Phi_{\Delta_0^c}\right) = 0.$$

Recall that for any configuration φ

$$\begin{aligned}\left(\boldsymbol{LPL}_{\Delta_0}^{(1)}(\varphi;\boldsymbol{\theta}^\star)\right)_j &= -\int_{\Delta_0\times\mathbb{M}} v_j(x^m|\varphi) e^{-V(x^m|\varphi;\boldsymbol{\theta}^\star)}\mu(dx^m) \\ &\quad + \int_{\Delta_0\times\mathbb{M}} v_j(x^m|\varphi\setminus x^m)\varphi(dx^m).\end{aligned} \tag{6.8}$$

Denote respectively by $G_1(\varphi)$ and $G_2(\varphi)$ the first and the second right-hand term of (6.8) and by $E_i = \boldsymbol{E}\left(G_i(\Phi)|\Phi_{\Delta_0^c} = \varphi_{\Delta_0^c}\right)$. From the definition of Gibbs point processes,

$$E_2 = \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})}\int_{\Omega_{\Delta_0}} Q(d\varphi_{\Delta_0})\int_{\mathbb{S}} \varphi_{\Delta_0}(dx^m)\mathbf{1}_{\Delta_0}(x) v_j(x^m|\varphi\setminus x^m) e^{-V\left(\varphi_{\Delta_0}|\varphi_{\Delta_0^c};\boldsymbol{\theta}^\star\right)}.$$

Denote by $\varphi' = (\varphi_{\Delta_0}, \varphi'_{\Delta_0^c})$. Since Q is a Poisson process we can write

$$\begin{aligned}E_2 &= \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})}\int_{\Omega} Q(d\varphi')\int_{\mathbb{S}} \varphi'(dx^m)\mathbf{1}_{\Delta_0}(x) v_j(x^m|\varphi\setminus x^m) e^{-V\left(\varphi_{\Delta_0}|\varphi_{\Delta_0^c};\boldsymbol{\theta}^\star\right)} \\ &= \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})}\int_{\Omega} Q(d\varphi') \\ &\quad \times \int_{\mathbb{S}} \varphi'(dx^m)\mathbf{1}_{\Delta_0}(x) v_j(x^m|\varphi'_{\Delta_0}\cup\varphi_{\Delta_0^c}\setminus x^m) e^{-V\left(\varphi'_{\Delta_0}|\varphi_{\Delta_0^c};\boldsymbol{\theta}^\star\right)}\end{aligned}$$

Now, from Campbell Theorem (applied to the Poisson measure Q)

$$\begin{aligned} E_2 &= \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})} \int_{\Delta_0\times\mathbb{M}} \mu(dx^m) \\ &\quad \times \int_\Omega Q^!_{x^m}(d\varphi') v_j(x^m|\varphi'_{\Delta_0} \cup \varphi_{\Delta_0^c}) e^{-V\left(\varphi'_{\Delta_0}\cup x^m|\varphi_{\Delta_0^c};\boldsymbol{\theta}^\star\right)}. \end{aligned}$$

Since from Slivnyak-Mecke Theorem, $Q = Q^!_x$, one can obtain

$$\begin{aligned} E_2 &= \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})} \int_\Omega Q(d\varphi') \int_{\Delta_0\times\mathbb{M}} \mu(dx^m)\, v_j(x^m|\varphi'_{\Delta_0} \cup \varphi_{\Delta_0^c}) e^{-V\left(\varphi'_{\Delta_0}\cup x^m|\varphi_{\Delta_0^c};\boldsymbol{\theta}^\star\right)} \\ &= \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})} \int_\Omega Q(d\varphi_{\Delta_0}) \int_{\Delta_0\times\mathbb{M}} \mu(dx^m) v_j(x^m|\varphi) e^{-V(x^m|\varphi;\boldsymbol{\theta}^\star)} e^{-V\left(\varphi_{\Delta_0}|\varphi_{\Delta_0^c};\boldsymbol{\theta}^\star\right)} \\ &= -E_1 \end{aligned}$$

Condition (*ii*): For any bounded domain Δ and any configuration φ, one may write for $j = 1, \ldots, p+1$

$$\begin{aligned} \left|\left(\boldsymbol{LPL}^{(1)}_\Delta(\varphi;\boldsymbol{\theta}^\star)\right)_j\right|^3 &\le 4\left|\int_{\Delta\times\mathbb{M}} v_j(x^m|\varphi) e^{-V(x^m|\varphi;\boldsymbol{\theta}^\star)} \mu(dx^m)\right|^3 \\ &\quad + 4\left|\sum_{x^m\in\varphi_\Delta} v_j(x^m|\varphi\setminus x^m)\right|^3 \end{aligned}$$

From [**Mod**], both right-hand terms are integrable with respect to $P_{\boldsymbol{\theta}^\star}$, which implies that for any domain Δ and in particular for Δ_i,

$$\boldsymbol{E}\left(\left|\left(\boldsymbol{LPL}^{(1)}_{\Delta_i}(\Phi;\boldsymbol{\theta}^\star)\right)_j\right|^3\right) < +\infty.$$

Condition (*iii*): let us start by noting that the vector $\boldsymbol{LPL}^{(1)}_{\Delta_i}(\varphi;\boldsymbol{\theta}^\star)$ depends only on φ_{Δ_j} for $j \in \mathbb{B}\left(i, \left\lceil \frac{D}{\widetilde{D}} \right\rceil\right)$. Let

$$E_{i,j} := \boldsymbol{E}\left(\boldsymbol{LPL}^{(1)}_{\Delta_i}(\Phi;\boldsymbol{\theta}^\star)\, \boldsymbol{LPL}^{(1)}_{\Delta_j}(\Phi;\boldsymbol{\theta}^\star)^T\right) = E_{0,j-i},$$

by using the stationarity of the process. From definitions, we can obtain

$$\begin{aligned} &\mathbb{V}\mathrm{ar}\left(|\Lambda_n|^{1/2}\boldsymbol{U}^{(1)}_n(\Phi;\boldsymbol{\theta}^\star)\right) \\ &= |\Lambda_n|^{-1}\mathbb{V}\mathrm{ar}\left(\sum_{i\in I_n} \boldsymbol{LPL}^{(1)}_{\Delta_i}(\Phi;\boldsymbol{\theta}^\star)\right) \\ &= |\Lambda_n|^{-1} \sum_{i,j\in I_n} E_{i,j} \\ &= |\Lambda_n|^{-1} \sum_{i\in I_n} \left(\sum_{j\in I_n\cap\mathbb{B}\left(i,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)} E_{i,j} + \sum_{j\in I_n\cap\mathbb{B}\left(i,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)^c} E_{i,j} \right). \end{aligned}$$

By using condition (i), one may assert that for any $j \in I_n \cap \mathbb{B}\left(i, \left\lceil \frac{D}{\widetilde{D}} \right\rceil\right)^c$

$$\begin{aligned} E &:= \boldsymbol{E}\left(\boldsymbol{LPL}^{(1)}_{\Delta_i}(\Phi;\boldsymbol{\theta}^\star)\,\boldsymbol{LPL}^{(1)}_{\Delta_j}(\Phi;\boldsymbol{\theta}^\star)^T\right) \\ &= \boldsymbol{E}\left(\boldsymbol{E}\left(\boldsymbol{LPL}^{(1)}_{\Delta_i}(\Phi;\boldsymbol{\theta}^\star)\,\boldsymbol{LPL}^{(1)}_{\Delta_j}(\Phi;\boldsymbol{\theta}^\star)^T|\Phi_{\Delta_j}\right)\right) \\ &= \boldsymbol{E}\left(\boldsymbol{E}\left(\boldsymbol{LPL}^{(1)}_{\Delta_i}(\Phi;\boldsymbol{\theta}^\star)\,|\Phi_{\Delta_j}\right)\boldsymbol{LPL}^{(1)}_{\Delta_j}(\Phi;\boldsymbol{\theta}^\star)^T\right) \\ &= 0 \end{aligned}$$

Denote by $\widetilde{I}_n$ the following set

$$\widetilde{I}_n = I_n \cap \left(\cup_{i\in\partial I_n}\mathbb{B}\left(i, \left\lceil \frac{D}{\widetilde{D}} \right\rceil\right)\right).$$

We now obtain

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left(|\Lambda_n|^{1/2}\boldsymbol{U}^{(1)}_n(\Phi;\boldsymbol{\theta}^\star)\right) &= |\Lambda_n|^{-1}\sum_{i\in I_n}\sum_{j\in I_n\cap\mathbb{B}\left(i,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)} E_{i,j} \\ &= |\Lambda_n|^{-1}\Bigg(\sum_{i\in I_n\setminus\widetilde{I}_n}\sum_{j\in I_n\cap\mathbb{B}\left(i,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)} E_{i,j} \\ &\qquad + \sum_{i\in\widetilde{I}_n}\sum_{j\in I_n\cap\mathbb{B}\left(i,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)} E_{i,j}\Bigg) \end{aligned}$$

Using the stationarity and the definition of the domain Λ_n, one obtains

$$\begin{aligned} |\Lambda_n|^{-1}\sum_{i\in I_n\setminus\widetilde{I}_n}\sum_{j\in I_n\cap\mathbb{B}\left(i,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)} E_{i,j} &= |\Lambda_n|^{-1}|I_n\setminus\widetilde{I}_n|\sum_{j\in\mathbb{B}\left(0,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)} E_{0,j} \\ &\to \underline{\boldsymbol{\Sigma}}(\widetilde{D},\boldsymbol{\theta}^\star), \end{aligned}$$

as $n \to +\infty$ and

$$|\Lambda_n|^{-1}\left|\sum_{i\in\widetilde{I}_n}\sum_{j\in I_n\cap\mathbb{B}\left(i,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)} E_{i,j}\right| \le |\Lambda_n|^{-1}|\widetilde{I}_n|\sum_{j\in\mathbb{B}\left(0,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)}|E_{0,j}| \to 0$$

as $n \to +\infty$. Hence as $n \to +\infty$

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(|\Lambda_n|^{1/2}\boldsymbol{U}_n^{(1)}(\Phi;\boldsymbol{\theta}^\star)\right) &= |\Lambda_n|^{-1}\sum_{i\in I_n}\sum_{j\in I_n\cap\mathbb{B}\left(i,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)} E_{i,j}\\ &\sim \underbrace{|I_n||\Lambda_n|^{-1}}_{\widetilde{D}^{-2}}\sum_{k\in\mathbb{B}\left(0,\left\lceil\frac{D}{\widetilde{D}}\right\rceil\right)} E_{0,k} = \underline{\boldsymbol{\Sigma}}(\widetilde{D},\boldsymbol{\theta}^\star) \quad (6.9)\end{aligned}$$

Step 2. Domination of $\underline{\boldsymbol{U}}_n^{(2)}(\Phi;\boldsymbol{\theta})$ *in a neighborhood of* $\boldsymbol{\theta}^\star$ *and convergence of* $\underline{\boldsymbol{U}}_n^{(2)}(\Phi;\boldsymbol{\theta}^\star)$

Let $j,k=1,\ldots,p$, recall that $\left(\underline{\boldsymbol{U}}_n^{(2)}(\varphi;\boldsymbol{\theta})\right)_{j,k}$ is defined for any configuration φ by

$$\left(\underline{\boldsymbol{U}}_n^{(2)}(\varphi;\boldsymbol{\theta})\right)_{j,k} = \frac{1}{|\Lambda_n|}\int_{\Lambda_n\times\mathbb{M}} v_j(x^m|\varphi)v_k(x^m|\varphi)e^{-V(x^m|\varphi;\boldsymbol{\theta})}\mu(dx^m).$$

Using the local stability property it comes

$$\left(\underline{\boldsymbol{U}}_n^{(2)}(\varphi;\boldsymbol{\theta})\right)_{j,k} \le \frac{e^K}{|\Lambda_n|}\int_{\Lambda_n\times\mathbb{M}} v_j(x^m|\varphi)v_k(x^m|\varphi)\mu(dx^m).$$

From **[Mod]**, one can apply Theorem 4 (Nguyen and Zessin (1979)). It follows that there exists n_0 such that for all $n \ge n_0$

$$\left|\left(\underline{\boldsymbol{U}}_n^{(2)}(\varphi;\boldsymbol{\theta})\right)_{j,k}\right| \le 2\times e^K\boldsymbol{E}\left(\left|v_j(0^M|\Phi)v_k(0^M|\Phi)\right|\right) \text{ for } P_{\boldsymbol{\theta}^\star}-a.e.\ \varphi.$$

Note that from Theorem 4 (Nguyen and Zessin (1979)), for all $\boldsymbol{\theta}$ (and in particular $\boldsymbol{\theta}=\boldsymbol{\theta}^\star$), $\underline{\boldsymbol{U}}_n^{(2)}(\cdot;\boldsymbol{\theta})$ converges almost surely as $n\to+\infty$ towards $\underline{\boldsymbol{U}}^{(2)}(\boldsymbol{\theta})$ which is the (p,p) matrix with elements

$$\left(\underline{\boldsymbol{U}}^{(2)}(\boldsymbol{\theta})\right)_{j,k} = \boldsymbol{E}\left(v_j(0^m|\Phi)v_k(0^m|\Phi)e^{-V(0^m|\varphi;\boldsymbol{\theta})}\right),\ \text{for } j,k=1,\ldots,p.$$

Let us underline that $\underline{\boldsymbol{U}}^{(2)}(\boldsymbol{\theta})$ is a symmetric definite positive matrix. Indeed, it is a positive matrix since for all $\boldsymbol{y}\in\mathbb{R}^{p+1}$

$$\begin{aligned}\boldsymbol{y}^T\underline{\boldsymbol{U}}^{(2)}(\boldsymbol{\theta})\boldsymbol{y} &= \sum_{j,k} y_j\boldsymbol{E}\left(v_j(0^M|\Phi)v_k(0^M|\Phi)e^{-V\left(0^M|\Phi;\boldsymbol{\theta}\right)}\right)y_k\\ &= \boldsymbol{E}\left(\left(\boldsymbol{y}^T\boldsymbol{v}\left(0^M|\Phi\right)\right)^2 e^{-V\left(0^M|\Phi;\boldsymbol{\theta}\right)}\right)\ge 0.\end{aligned}$$

And it is definite since, for all $\boldsymbol{y}\in\mathbb{R}^{p+1}\setminus\{0\}$ from **[Ident]**, $\boldsymbol{y}^T\boldsymbol{v}\left(0^m|\varphi\right)=0$ for $\lambda^{\mathrm{m}}\otimes P_{\boldsymbol{\theta}^\star}$−a.e. (m,φ) implies $\boldsymbol{y}=0$.

Conclusion Step Under the assumptions **[Mod]** and **[Ident]**, and using Steps 1 and 2, one can apply a classical result concerning asymptotic normality for minimum contrast estimators, see Guyon (1995) in order to obtain (4.3).

Now, the result (4.5) is proved in three substeps:

(i) We first prove that the matrix $\underline{\underline{\boldsymbol{\Sigma}}}(\boldsymbol{\theta}^\star) = \underline{\underline{\boldsymbol{\Sigma}}}(\overline{D}, \boldsymbol{\theta}^\star)$ is a symmetric definite positive matrix. From Equation (6.9), it is sufficient to prove that the matrix $\mathbb{V}\mathrm{ar}(|\Lambda_n(\overline{D})|^{-1/2}\boldsymbol{LPL}^{(1)}_{\Lambda_n(\overline{D})}(\Phi;\boldsymbol{\theta}^\star))$ is definite positive for n large enough. Let $\boldsymbol{y} \in \mathbb{R}^p \setminus \{0\}$, the aim is to prove that

$$V := \boldsymbol{y}^T \mathbb{V}\mathrm{ar}\left(|\Lambda_n(\overline{D})|^{-1/2}\boldsymbol{LPL}^{(1)}_{\Lambda_n(\overline{D})}(\Phi;\boldsymbol{\theta}^\star)\right)\boldsymbol{y} > 0.$$

Let $\overline{\Lambda} = \cup_{i\in\mathbb{B}\left(0,\left\lceil\frac{D}{\overline{D}}\right\rceil\right)}\Delta_i(\overline{D})$, using the same argument of Jensen and Künsch (1994) (Equation (3.2)), one can write

$$V \geq |\Lambda_n(\overline{D})|^{-1}\,\boldsymbol{E}\left(\mathbb{V}\mathrm{ar}\left(\boldsymbol{y}^T\boldsymbol{LPL}^{(1)}_{\Lambda_n(\overline{D})}(\Phi;\boldsymbol{\theta}^\star)\,|\Phi_{\Delta_k(\overline{D})}, k \notin (2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2\right)\right).$$

Note that for $i \neq j \in I_n$,

$$Cov\left(\boldsymbol{y}^T\boldsymbol{LPL}^{(1)}_{\Delta_i(\overline{D})}(\varphi;\boldsymbol{\theta}^\star), \boldsymbol{y}^T\boldsymbol{LPL}^{(1)}_{\Delta_j(\overline{D})}(\varphi;\boldsymbol{\theta}^\star)\,|\Phi_{\Delta_k(\overline{D})}, k \notin (2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2\right) = 0$$

due to the independence of $\boldsymbol{LPL}^{(1)}_{\Delta_i(\overline{D})}(\varphi;\boldsymbol{\theta}^\star)$ and $\boldsymbol{LPL}^{(1)}_{\Delta_j(\overline{D})}(\varphi;\boldsymbol{\theta}^\star)$ conditionally on $\Phi_{\Delta_k(\overline{D})}, k \notin (2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2$ when $i,j \in I_n \cap (2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2$ and $\boldsymbol{LPL}^{(1)}_{\Delta_i(\overline{D})}(\varphi;\boldsymbol{\theta}^\star)$ or $\boldsymbol{LPL}^{(1)}_{\Delta_j(\overline{D})}(\varphi;\boldsymbol{\theta}^\star)$ is constant when either i or $j \notin I_n \cap (2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2$. As a direct consequence,

$$\begin{aligned}
V &\geq |\Lambda_n(\overline{D})|^{-1}\boldsymbol{E}\left(\mathbb{V}\mathrm{ar}\left(\boldsymbol{y}^T\sum_{i\in I_n}\boldsymbol{LPL}^{(1)}_{\Delta_i(\overline{D})}(\Phi;\boldsymbol{\theta}^\star)\,\big|\Phi_{\Delta_k(\overline{D})}, k \notin (2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2\right)\right)\\
&= |\Lambda_n(\overline{D})|^{-1}\sum_{i\in I_n}\boldsymbol{E}\left(\mathbb{V}\mathrm{ar}\left(\boldsymbol{y}^T\boldsymbol{LPL}^{(1)}_{\Delta_i(\overline{D})}(\Phi;\boldsymbol{\theta}^\star)\,\big|\Phi_{\Delta_k(\overline{D})}, k \notin (2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2\right)\right)\\
&= |\Lambda_n(\overline{D})|^{-1}\sum_{\ell\in I_n\cap(2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2\setminus\widetilde{I_n}}\\
&\quad\boldsymbol{E}\left(\mathbb{V}\mathrm{ar}\left(\boldsymbol{y}^T\sum_{i\in I_n\cap\mathbb{B}\left(\ell,\left\lceil\frac{D}{\overline{D}}\right\rceil\right)}\boldsymbol{LPL}^{(1)}_{\Delta_i(\overline{D})}(\Phi;\boldsymbol{\theta}^\star)\,\big|\Phi_{\Delta_k(\overline{D})}, k \notin (2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2\right)\right)\\
&\quad+|\Lambda_n(\overline{D})|^{-1}\sum_{\ell\in(2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2\cap\widetilde{I_n}}\\
&\quad\boldsymbol{E}\left(\mathbb{V}\mathrm{ar}\left(\boldsymbol{y}^T\sum_{i\in I_n\cap\mathbb{B}\left(\ell,\left\lceil\frac{D}{\overline{D}}\right\rceil\right)}\boldsymbol{LPL}^{(1)}_{\Delta_i(\overline{D})}(\Phi;\boldsymbol{\theta}^\star)\,\big|\Phi_{\Delta_k(\overline{D})}, k \notin (2\left\lceil\frac{D}{\overline{D}}\right\rceil+1)\mathbb{Z}^2\right)\right)
\end{aligned}$$

Following the proof of Step 1, condition (*iii*) one may prove that the second right-hand term tends to 0 as $n \to +\infty$. Therefore by using the stationarity, we have for n large enough

$$\begin{aligned} V &\geq \frac{1}{2}|\Lambda_n(\overline{D})|^{-1}\left|I_n \cap (2\left\lceil \frac{D}{\overline{D}}\right\rceil + 1)\mathbb{Z}^2\right| \\ &\qquad \times \boldsymbol{E}\left(\mathbb{V}\mathrm{ar}\left(\boldsymbol{y}^T \boldsymbol{LPL}^{(1)}_{\Lambda}(\Phi;\boldsymbol{\theta}^\star)\,\middle|\Phi_{\Delta_k(\overline{D})}, 1 \leq |k| \leq 2\left\lceil \frac{D}{\overline{D}}\right\rceil\right)\right) \\ &= \frac{\overline{D}^{-2}}{2}\frac{|I_n \cap (2\left\lceil \frac{D}{\overline{D}}\right\rceil + 1)\mathbb{Z}^2|}{|I_n|} \\ &\qquad \times \boldsymbol{E}\left(\mathbb{V}\mathrm{ar}\left(\boldsymbol{y}^T \boldsymbol{LPL}^{(1)}_{\Lambda}(\Phi;\boldsymbol{\theta}^\star)\,\middle|\Phi_{\Delta_k(\overline{D})}, 1 \leq |k| \leq 2\left\lceil \frac{D}{\overline{D}}\right\rceil\right)\right) \\ &\geq \frac{\overline{D}^{-2}}{2}\left(\frac{3}{4\left\lceil \frac{D}{\overline{D}}\right\rceil + 1}\right)^2 \\ &\qquad \times \boldsymbol{E}\left(\mathbb{V}\mathrm{ar}\left(\boldsymbol{y}^T \boldsymbol{LPL}^{(1)}_{\Lambda}(\Phi;\boldsymbol{\theta}^\star)\,\middle|\Phi_{\Delta_k(\overline{D})}, 1 \leq |k| \leq 2\left\lceil \frac{D}{\overline{D}}\right\rceil\right)\right) \end{aligned}$$

Assume there exists some positive constant c such that $P_{\boldsymbol{\theta}^\star}-$a.s. $\boldsymbol{y}^T \boldsymbol{LPL}^{(1)}_{\Lambda}(\Phi;\boldsymbol{\theta}^\star) = c$ when the variables $\Phi_{\Delta_k(\overline{D})}, 1 \leq |k| \leq 2\left\lceil \frac{D}{\overline{D}}\right\rceil$ are (for example) fixed to $\emptyset$. By assuming **[SDP]** it follows that for any $\varphi_i \in A_i$ for $i = 0, \ldots, \ell$ (with $\ell \geq p$), $\boldsymbol{y}^T\left(\boldsymbol{LPL}^{(1)}_{\Lambda}(\varphi_i;\boldsymbol{\theta}^\star) - \boldsymbol{LPL}^{(1)}_{\Lambda}(\varphi_0;\boldsymbol{\theta}^\star)\right) = 0$. Since for all $(\varphi_0, \ldots, \varphi_\ell) \in A_0 \times \ldots \times A_\ell$, the matrix with entries $\left(\boldsymbol{LPL}^{(1)}_{\Lambda}(\varphi_i;\boldsymbol{\theta}^\star)\right)_j - \left(\boldsymbol{LPL}^{(1)}_{\Lambda}(\varphi_0;\boldsymbol{\theta}^\star)\right)_j$ is assumed to be injective, this leads to $\boldsymbol{y} = 0$ and hence to some contradiction. Therefore, when the variables $\Phi_{\Delta_k(\overline{D})}$, $1 \leq |k| \leq 2\left\lceil \frac{D}{\overline{D}}\right\rceil$ are fixed to $\emptyset$, the variable $\boldsymbol{y}^T \boldsymbol{LPL}^{(1)}_{\Lambda}(\Phi;\boldsymbol{\theta}^\star)$ is almost surely not a constant. Hence, $\underline{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^\star)$ is a symmetric definite positive matrix.

(*ii*) *Convergence of* $\widehat{\underline{\boldsymbol{\Sigma}}}_n(\varphi; D^\vee, \widetilde{D}, \boldsymbol{\theta})$.

Let us recall that for any $\varphi \in \Omega$, $D^\vee \geq D$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ we define

$$\widehat{\underline{\boldsymbol{\Sigma}}}_n(\varphi; D^\vee, \widetilde{D}, \boldsymbol{\theta}) = \frac{\widetilde{D}^{-2}}{|I_n|}\sum_{i \in I_n}\sum_{j \in I_n \cap \mathbb{B}\left(i, \left\lceil \frac{D^\vee}{D}\right\rceil\right)} \boldsymbol{LPL}^{(1)}_{\Delta_i}(\varphi;\boldsymbol{\theta})^T \boldsymbol{LPL}^{(1)}_{\Delta_j}(\varphi;\boldsymbol{\theta})$$

We also define

$$X_i(\varphi) := X_i(\varphi)^{k,\ell} = \sum_{j \in I_n \cap \mathbb{B}\left(i, \left\lceil \frac{D^\vee}{D}\right\rceil\right)} \left(\boldsymbol{LPL}^{(1)}_{\Delta_i}(\varphi;\boldsymbol{\theta})\right)_k \left(\boldsymbol{LPL}^{(1)}_{\Delta_j}(\varphi;\boldsymbol{\theta})\right)_\ell,$$

$Y_i(\varphi) := X_i(\varphi) - \boldsymbol{E}(X_i(\Phi))$ and $\overline{Y}_n(\varphi) = |I_n|^{-1}\sum_{i \in I_n} Y_i(\varphi)$. Since one may notice that $\boldsymbol{E}(X_i(\Phi)) = \widetilde{D}^2\left(\underline{\boldsymbol{\Sigma}}(\widetilde{D}, \boldsymbol{\theta})\right)_{k,\ell}$, we have

$$\overline{Y}_n(\varphi) = \widetilde{D}^2\left(\widehat{\underline{\boldsymbol{\Sigma}}}_n(\varphi; D^\vee, \widetilde{D}, \boldsymbol{\theta}) - \underline{\boldsymbol{\Sigma}}(\widetilde{D}, \boldsymbol{\theta})\right)_{k,\ell}.$$

Thus, the aim is to prove that as $n \to +\infty$, $\overline{Y}_n(\varphi) \to 0$ for $P_{\boldsymbol{\theta}^\star}-$a.e. φ. Since the process $\{Y_i, i \in \mathbb{Z}^2\}$ is stationary, it is sufficient to prove, see *e.g.* Guyon (1995)

- (a) $\boldsymbol{E}\left(Y_0(\Phi)^2\right) < +\infty$
- (b) $\boldsymbol{E}\left(|I_n|\overline{Y}_n(\Phi)^2\right) < +\infty$.

(a) We leave the reader to verify that **[Mod]** ensures this integrability condition.

(b) Note that $Y_i(\varphi)$ depends only on φ_{Δ_j} for $j \in \mathbb{B}\left(i, \left\lceil \frac{D^\vee}{\widetilde{D}} \right\rceil + \left\lceil \frac{D}{\widetilde{D}} \right\rceil\right)$. Hence, by choosing $j \in I_n \cap \mathbb{B}\left(i, \alpha_{D^\vee,D}\right)^c$ with $\alpha_{D^\vee,D} = \alpha_{D^\vee,D}(\widetilde{D}) := 2\left\lceil \frac{D^\vee}{\widetilde{D}} \right\rceil + \left\lceil \frac{D}{\widetilde{D}} \right\rceil$, the variables Y_i and Y_j are independent. Then, we obtain

$$\begin{aligned}
\boldsymbol{E}\left(|I_n|\overline{Y}_n(\Phi)^2\right) &= \frac{1}{|I_n|}\sum_{i,j\in I_n} \boldsymbol{E}\left(Y_i(\Phi)Y_j(\Phi)\right) \\
&= \frac{1}{|I_n|}\sum_{i\in I_n}\Bigg(\sum_{j\in I_n\cap\mathbb{B}\left(i,\alpha_{D^\vee,D}\right)} \boldsymbol{E}\left(Y_i(\Phi)Y_j(\Phi)\right) \\
&\qquad + \sum_{j\in I_n\cap\mathbb{B}\left(i,\alpha_{D^\vee,D}\right)^c} \boldsymbol{E}\left(Y_i(\Phi)Y_j(\Phi)\right)\Bigg) \\
&= \frac{1}{|I_n|}\sum_{i\in I_n}\sum_{j\in I_n\cap\mathbb{B}\left(i,\alpha_{D^\vee,D}\right)} \boldsymbol{E}\left(Y_i(\Phi)Y_j(\Phi)\right) \\
&\sim \sum_{k\in\mathbb{B}\left(0,\alpha_{D^\vee,D}\right)} \boldsymbol{E}\left(Y_0(\Phi)Y_k(\Phi)\right) \le \left(2\alpha_{D^\vee,D}+1\right)\boldsymbol{E}\left(Y_0(\Phi)^2\right).
\end{aligned}$$

Therefore, for all $D^\vee \ge D$ and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we have for $P_{\boldsymbol{\theta}^\star}-$a.e. φ as $n \to +\infty$

$$\underline{\widehat{\boldsymbol{\Sigma}}}_n(\varphi; D^\vee, \widetilde{D}, \boldsymbol{\theta}) \to \underline{\boldsymbol{\Sigma}}(\widetilde{D}, \boldsymbol{\theta}) = \underline{\boldsymbol{\Sigma}}(\boldsymbol{\theta}). \tag{6.10}$$

(iii) Since for any φ, the functions $\underline{\boldsymbol{U}}_n^{(2)}(\varphi;\cdot)$ and $\underline{\widehat{\boldsymbol{\Sigma}}}_n(\varphi; D^\vee, \widetilde{D}, \cdot)$ are continuous, it follows from Step 2 and (6.10) that one obtains for $P_{\boldsymbol{\theta}^\star}-$a.e. φ, as $n \to +\infty$

$$\underline{\boldsymbol{U}}_n^{(2)}(\varphi; \widehat{\boldsymbol{\theta}}) \to \underline{\boldsymbol{U}}^{(2)}(\boldsymbol{\theta}^\star) \quad \text{and} \quad \underline{\widehat{\boldsymbol{\Sigma}}}_n(\varphi; D^\vee, \widetilde{D}, \widehat{\boldsymbol{\theta}}) \to \underline{\boldsymbol{\Sigma}}(\boldsymbol{\theta}^\star).$$

Finally, note that the previous convergence also implies that for n large enough $\underline{\widehat{\boldsymbol{\Sigma}}}_n(\Phi; D^\vee, \widetilde{D}, \widehat{\boldsymbol{\theta}})$ is almost surely a symmetric definite positive matrix.

Acknowledgements

We are grateful to the referees and the associate editor for their helpful comments. We also thank Laurence Pierret for the time spent on improving the writing of this paper in English.

References

Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns. *Aust. N.Z. J. Statist.*, 42(3):283–322. MR1794056

Bertin, E., Billiot, J.-M., and Drouilhet, R. (1999a). Existence of "Nearest-Neighbour" Gibbs Point Models. *Adv. Appl. Prob.*, 31:895–909. MR1747447

Bertin, E., Billiot, J.-M., and Drouilhet, R. (1999b). k-Nearest-Neighbour Gibbs Point Processes. *Markov Processes and Related Fields*, 5(2):219–234. MR1762174

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice system. *J. R. Statist. Soc. Ser. B*, 26:192–236. MR0373208

Besag, J., Milne, R., and Zachary, S. (1982). Point process limits of lattice processes. 19:210–216. MR0644433

Geyer, C. (1999). Likelihood inference for spatial point processes. In O.E. Barndorff-Nielsen, W. K. . M. v. L., editor, *Stochastic Geometry: likelihood and Computation*, pages 79–140, Bocca Raton, Florida. MR1673118

Guyon, X. (1995). *Random Fields on a Network: Modeling, Statistics, and Applications*. Springer, New York. MR1344683

Jensen, J. and Künsch, H. (1994). On asymptotic normality of pseudo likelihood estimates of pairwise interaction processes. *Ann. Inst. Statist. Math.*, 46:475–486. MR1309718

Jensen, J. and Møller, J. (1991). Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.*, 1:445–461. MR1111528

Kallenberg, O. (1983). *Random Measures*. Academie-Verlag and Academic Press, Berlin, London. MR0818219

Mase, S. (1995). Consistency of maximum pseudo-likelihood estimator of continuous state space gibbsian process. *Ann. Appl. Probab.*, 5:603–612. MR1359821

Mase, S. (2000). Marked gibbs processes and asymptotic normality of maximum pseudo-likelihhod estimators. *Math. Nachr.*, 209:151–169. MR1734363

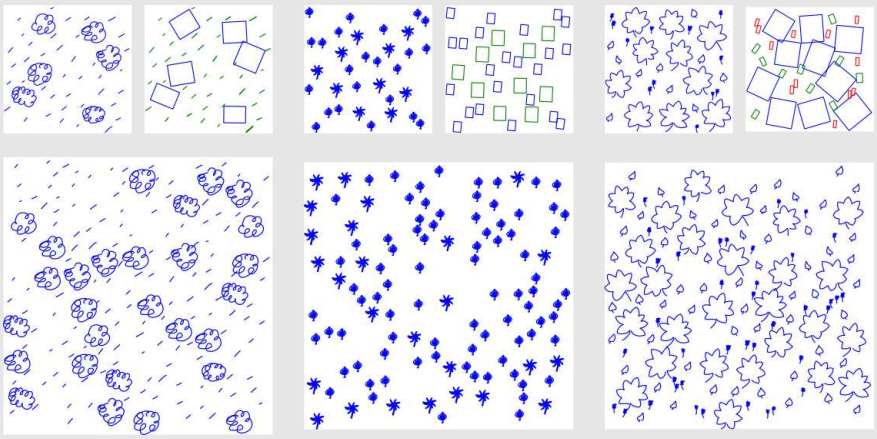
Møller, J. and Waggepetersen, R. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

Nguyen, X. and Zessin, H. (1979). Ergodic theorems for Spatial Process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 48:133–158. MR0534841

Preston, C. (1976). *Random Fields*, volume 534. Springer-Verlag, Berlin, Heidelberg, New York. MR0448630

Ruelle, D. (1970). Superstable interactions in classical statistical mechanics. *Commun. Math. Phys.*, 18:127–159. MR0266565

Stoyan, D., Kendall, W., and Mecke, J. (1995). *Stochastic Geometry and its Applications*. Wiley, Chichester, second edition.



[HLTGDC09]
T. Hurtut, P.-E. Landes, J. Thollot, Y. Gousseau, R. Drouilhet and J.-F. Coeurjolly. Appearance-guided Synthesis of Element Arrangements by Example, In *7th International Symposium on Non-Photorealistic Animation and Rendering, NPAR 2009*, 2009.

Appearance-guided Synthesis of Element Arrangements by Example

T. Hurtut, P.-E. Landes, J. Thollot
INRIA* - LJK† CNRS

Y. Gousseau
Telecom ParisTech - LTCI‡ CNRS

R. Drouillhet, J.-F. Coeurjolly
LJK† CNRS - SAGAG

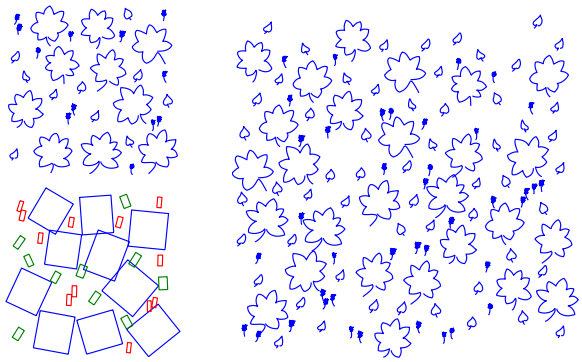


Figure 1: *Given a reference arrangement composed of vector elements* (top left)*, our analysis scheme divides the raw element set into appearance categories* (bottom left)*. Spatial interactions based on appearance can be learned by statistical modeling and exploited to yield visually similar arrangements* (right)*.*

Abstract

We present a technique for the analysis and re-synthesis of 2D arrangements of stroke-based vector elements. The capture of an artist's style by the sole posterior analysis of his/her achieved drawing poses a formidable challenge. Such by-example techniques could become one of the most intuitive tools for users to alleviate creation process efforts. Here, we propose to tackle this issue from a statistical point of view and take specific care of accounting for information usually overlooked in previous research, namely the elements' very appearance. Composed of curve-like strokes, we describe elements by a concise set of perceptually relevant features. After detecting appearance dominant traits, we can generate new arrangements that respect the captured appearance-related spatial statistics using multitype point processes. Our method faithfully reproduces visually similar arrangements and relies on neither heuristics nor post-processes to ensure statistical correctness.

Keywords: Vector texture synthesis, by-example synthesis, NPR.

1 Introduction

Automated stroke-based rendering systems are common in non photo-realistic rendering (NPR). Successful systems used to generate NPR depictions of 3D scenes or photographs are mainly based on heuristics or hard-coded rendering rules and it is up to the artist to take advantage of them to convey his/her own style. Fewer techniques, on the other hand, attempt to automatically learn it instead. In such approaches, the artist provides the system with an example, typically an eventually partially-finished drawing, which has to be analyzed in a way to grasp part of the user's style. The information extracted by this analysis is then used to automatically synthesize new examples visually similar to the original. Such approaches constitute very intuitive tools for artists to handle cumbersome and repetitive tasks, such as creating filling patterns. The main challenge of these techniques is to identify from a limited input what can be assimilated to style and capture it in a way that allows further synthesis. Moreover, to ensure a satisfactory variety of styles, priors that could restrict the scope of supported examples need to be avoided as much as possible.

We focus here on the synthesis of stroke-based elements arrangements. By arrangements, we mean distributions over the 2D plane of visual primitives that do not obey any placement rules or geometric constraints. In such cases, statistics over distances between elements are of primary importance and greatly characterize the input distribution. Our primary goal is then to faithfully reproduce these statistics in order to generate new resembling arrangements.

We claim that, more than sole spatial considerations, the distributed elements' appearance has to be investigated. More specifically, we believe that reproducing pair-wise occurrences of specific visual cues is mandatory to confer output arrangements the same "feel" as the given example. We therefore propose to model and take into account the elements' appearance in the synthesis. To achieve this, we concentrate on stroke-based elements defined as a set of of path-following strokes. Each stroke is a vector curve allowing us to take advantage of studies in line perception [Julesz 1986] to yield an effective element description.

Note that we assume that already-built elements are provided by the user, not individual strokes. They can either be directly drawn by an artist as a whole (this is the case of the examples shown in the paper), but could also be the result of a stroke clustering pre-process similar to the one proposed by [Barla et al. 2006b].

1.1 Related Work

Our technique exploits the vector elements' appearance to guide the synthesis of new arrangements. Related issues arise in various research fields in Computer Graphics, from raster texture synthesis to NPR stroke-based rendering systems. To provide users with intuitive manipulation handles, we favor example-based approaches over procedural techniques. Since texture synthesis is a rich Computer Science field, our review will focus on example-based methods only, before exploring line appearance encoding.

1.1.1 Pixel-based Texture Synthesis

Raster texture synthesis is very inspiring as it focused on example-based approaches early on, see for instance [Heeger and Bergen 1995]. In that case, elements are mere pixels and many successful techniques consider their colors as the realizations of a hidden Markov Random Field (MRF). Their objective is to simulate further sampling to generate new visually-close textures. Most techniques non-parametrically sample their input and use pixel neighborhood matching as an efficient way to implicitly capture its local behavior [Efros and Leung 1999; Wei and Levoy 2000; Ashikhmin 2001]. However, both the appearance and relative placements of such elements are quite limited. Though extra features can be embedded for improved matching [Wu and Yu 2004; Lefebvre and Hoppe 2006], pixels can only be assigned colors and are to follow the lattice structure imposed by the raster grid.

In our case, we aim at producing new arrangements of richer ele-

*ARTIS project-team
†Grenoble Universities, Laboratoire Jean Kuntzmann, CNRS UMR 5224
‡CNRS UMR 5141

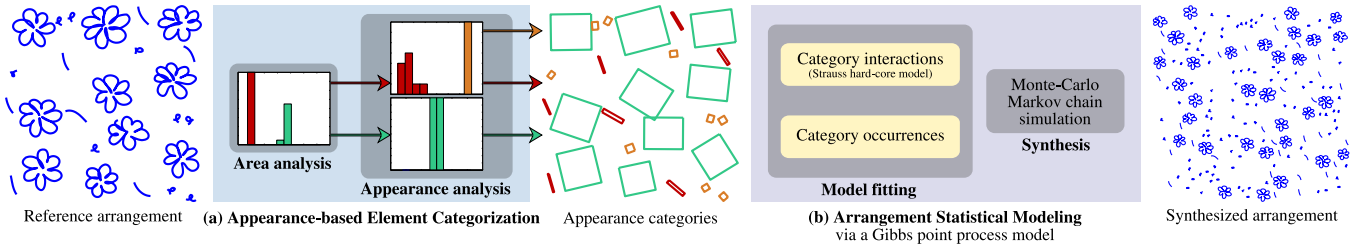


Figure 2: *Overview of our method. First, our* ***appearance-based element categorization*** (a) *examines the reference elements' shape and divides input elements into appearance categories. We perform this analysis by two successive grouping steps: according to the elements' area, and then to their visual appearance. Once the categories are established, we carry on with the* ***arrangement statistical modeling*** *of the user's input* (b)*. Its goal is to capture spatial interactions within, and between categories. For that aim, we infer the parameters of a Gibbs point process model from the categorized example. We can then generate new arrangements by establishing new realizations from the fitted model using Monte-Carlo Markov chain sampling.*

ments distributed over the 2D plane without any prior placement.

1.1.2 Patch-based Texture Synthesis

Motivated by the need to capture visual structures lost by the independent process of pixels, texture synthesis involving wider elements, namely pixel patches, were proposed. Few of them however take care of explicitly capturing and handling their relative spatial arrangements.

An interesting example is the case of the texture particle representation [Dischler et al. 2002]. The input bitmap texture is decomposed into a set of small blob-like elements, coined *particles*. Their relative placement is captured by the distances between neighboring particles' axis-aligned bounding boxes. Neighborhood relationships are determined by successive morphological dilation operations until contact between elements is established. Re-synthesis is achieved via a seeding procedure that uses non-parametric sampling for additions of new patches.

Earlier work by [Guo et al. 2001] further completes the analysis process as they acquire their elements, introduced as *textons*, by visual learning. They infer the parameters of a texture model defining the input image as the composed realizations of two stochastic *texton processes*. Elements' appearance, density and spatial arrangements are embedded in this unified model whose configuration likelihood is described by a Gibbs distribution. The parameters maximizing it are estimated by gradient ascent, the overall arrangement evolving according to a Markov chain process. Though powerful, their method requires the evaluation of many parameters and the output texton set gets visually relevant only after hundreds of iterations. We still aim at using similar statistical tools since they provide us with an elegant way of enforcing appearance-based statistics over the output. Not only transposing those techniques to vector elements, we also propose faster solutions.

Other techniques, dedicated to near-regular textures, strive to explicitly identify the underlying lattice structure in the input in order to obtain meaningful building elements [Liu et al. 2004a; Liu et al. 2004b]. Regularity between peaks of auto correlation is investigated and *tiles* – minimal set of patches whose periodic repetition defines the texture – are extracted accordingly. Their insight is that the number of possibilities of tiling the 2D Euclidean plane is limited to the finite number of wallpaper groups. However, such approaches are difficult to generalize to non-regular arrangements of vector elements such as those we want to re-synthesize.

1.1.3 Extension to Vector Primitives

Many generative NPR systems use *strokes* as their basic rendering building blocks. Such inputs (stipples, curves or brush strokes) can be handled in vector form. Compared to pixel patches, this representation grants a more subtle description of the elements' content. This enables us to find new approaches extending example-based raster synthesis procedures.

First attempts consist of parametric approaches. [Jodoin et al. 2002] first deal with the synthesis of hatching patterns by modeling 1D hatch sequences with an explicit MRF to reproduce local pair-wise distances between successive elements. The statistical modeling is elegant but difficult to extend to automatic 2D drawing analysis. Similarly, [Barla et al. 2006a] propose a method to synthesize 2D arrangements of both points and lines and enforce specific statistics on element in a corrective step.

As in texture synthesis, efficient non-parametric sampling techniques were devised, like in Barla and co-authors' subsequent work [Barla et al. 2006b]. Their main contribution is to yield an intermediate input representation by building elements out of strokes using proximity and continuation constraints. For re-synthesis, they first generate, for a given density, a 1D or 2D set of seed points. Input elements are then pasted to those locations by local neighborhood matching. The employed neighborhood system is the Delaunay triangulation over the elements' barycenters and additional perceptual measures determine the matching. Though used during element building, appearance attributes do not contribute during the synthesis step and supported distributions are uniform due to the Llyod relaxation performed on the seed points. [Ijiri et al. 2008] propose a similar, more synthesis-oriented method. New arrangements are created incrementally and rule-based heuristics ensure the correctness of the ongoing triangulation. Again, elements' visual attributes do not influence the distribution itself and most of Barla's perceptual matching considerations are gone for the sake of interactivity.

Our inputs are similar to Barla's or Ijiri's with subtle differences though. We directly have already-built elements at our disposal contrary to the former, while our elements are not explicitly labeled contrary to the latter. Our approach is also different from their work as we formalize arrangement analysis and re-synthesis as a statistical learning problem.

1.1.4 Line Appearance Encoding

One of our contributions is to use dominant element appearance traits as soft constraints influencing the synthesized distribution it-

self. All previously mentioned techniques only account for spatial considerations to determine their output arrangements. Even methods whose inputs allow relevant appearance analysis overlook this valuable information.

Element's appearance encoding is thus of primary importance. NPR research in style transfer accounts for that concern, a proper representation of the strokes' visual attributes largely contributing to the transfer success. Freeman's work on line drawing stylization uses an implicit representation of line appearance by using training data sets of lines and finding nearest neighbors in the target style set [Freeman et al. 2003]. The user's line drawing style is captured by the WYSIWYG NPR rendering system by encoding over-sketch as offsets relative to the line base path [Kalnins et al. 2002]. Style is then encoded as an explicit MRF which allows further 1D synthesis. [Hertzmann et al. 2002] extend their analogy framework to polyline stylization by example and match neighborhoods of the curve's points by comparing point positions and tangent magnitudes. Finally, [Brunn et al. 2007] capture line style as the details functions yielded by a wavelet-like decomposition of the lines.

In this paper, we dispose of compound elements composed of several path-following strokes. This representation of our input allows us to propose more elaborated measurements inspired by line perception studies and use those as relevant features for appearance categorization.

1.2 Contributions

Proposing new approaches for both arrangement analysis and synthesis, the contributions of our method are two-fold:

1. We propose a new algorithm to categorize the elements of a given arrangement using perceptually motivated measures.
2. Based on these measures, we use a multitype point process model to perform synthesis. We chose a model adapted to the capture and restitution of appearance statistics evaluated between, and within element categories.

The main advantages of our method are that it does not require any assumption concerning the input arrangement's distribution and that it performs accurate handling of the elements' appearance. We provide a detailed overview of our arrangement synthesis method's work-flow in Figure 2.

2 Appearance-based Element Categorization

The first step of our method aims at categorizing the example's elements according to their appearance. If some elements exhibit a similar appearance thorough the input, we want to recognize them as belonging to a same category. Elements considered as unique in the example will be grouped in an outlier category. Note that this step corresponds to an automated solution for the manual labeling of [Ijiri et al. 2008]. Our result could, therefore, be used as an input for their algorithm.

The reasons behind our appearance-driven element categorization are the following. According to the Gestalt law of similarity grouping, the Human Visual System tends to mentally perform perceptual categorization and build groups from isolated elements. Once those ensembles are established, strong visual interactions can arise. Not only elements can be perceived as interacting with the other members of its group, but interactions can also occur at the group level. This phenomenon is illustrated on Figure 1 where elements are visually split into three main appearance-based categories, namely the elongated, the cross-like, and the smaller strokes. Since all those elements do not overlap and are mixed quasi-regularly, inter-category interactions are considered repulsive here. On the other hand, intra-category interactions are different and could be described as follows. If considered only with respect to the other members of their own category, the elongated and small strokes seem regularly placed. For the cross-like elements, however, this placement rule does not apply. This distinction between inter- and intra-category visual interactions is mandatory to devise a good capture of the arrangement's visual attributes. We propose a method that can account for it.

2.1 Stroke-based Element Description

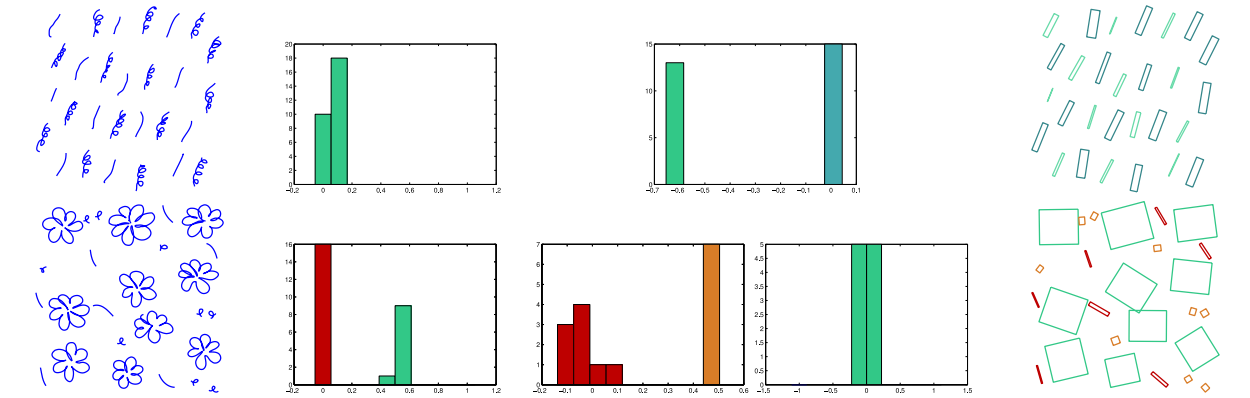
[Julesz 1986] studied human perceptual discrimination of textures composed of stroke-based elements, which he called *textons*. In his theory, discriminative features include the element's principal orientation, as well as its number of crossings and extremities. We use these features as our elements' appearance descriptors. Besides, Julesz's textons all shared the same size. To account for that, we add two features to each element's description: its area and elongation. In practice, elements' orientation, elongation and area are estimated on their fitted bounding box. Crossings and extremities are measured directly on the strokes constituting the element.

Next section tackles the issue of grouping together elements that meaningfully share similar characteristics in this feature space. This brings us to directly compare features that capture drastically different visual characteristics. This question is common to all clustering problems in heterogeneous spaces. Before comparison, features are normalized on $[0, 1]$. Care must be taken that the $[0, 1]$ interval still covers enough visual variation for each characteristic. Orientation is normalized by 2π. Elongation is defined as the ratio of the element's major axis over its minor axis and is normalized by 3. Elements whose elongation before normalization exceeds 3 are tagged as *very elongated* and their associated normalized elongation value is limited to 1. Similarly, elements whose area is larger than 5% of the reference arrangement's area are considered as *large* and attached a normalized surface value of 1. Since we have vector elements at our disposal, we can accurately estimate the curvature of their constitutive strokes. We embed this valuable shape information into our description by counting the number of points of strong curvature along the elements' curves. This feature intuitively corresponds to the number of perceived extremities and is normalized by 10. Lastly, we account for the number of crossings within each element and, as for the extremities, normalize it by 10.

In summary, this gives the following feature set:

Element features	Normalization constant
[Ijiri et al. 2008] Area	5% wrt reference arrangement surface
Principal orientation	2π
Elongation	3
Number of extremities	10
Number of crossings	10

Our description is highly discriminative and focuses on the lines' geometrical shape. Yet notice that the proposed line representation is by no means final, and incorporating other features could be possible. One needs to carefully choose those as the more features are added, the more observations in the input must be provided in order to devise meaningful statistics over a more highly-dimensional feature space. Correlation between descriptive components should also be as low as possible to reduce redundancy. For instance, embedding elements' colors in our descriptor set would thus require special care, such as palette extraction, to avoid the classical "curse of dimensionality" issue.



a) Reference arrangement b) Area histogram c) Appearance histogram d) Appearance categories

Figure 3: Categorization process *Given a reference arrangement* (a), *the modes of the area histogram are first detected* (b). *In this example, one mode is found for the top row, while two are identified for the bottom row corresponding to the small and large elements. For each resulting area category, modes of the appearance histogram are then computed* (c). *Here, two modes are detected for the first row, the most discriminative feature being the number of crossings. The bottom row displays the appearance histogram of small elements where two MM-modes appear (blue and green modes). Only one MM-mode has been detected on the appearance histogram (red mode) for the larger element set. The resulting categories are finally shown with corresponding colors* (d).

2.2 Detection of Meaningful Feature Modes

Our goal here is to categorize elements sharing common visual characteristics once represented in the previously introduced feature space. Our approach is based on two important perception considerations. First, visual perception argues that size is the first information to be perceived for visual recognition tasks. Consequently, we need to bring together elements whose area is roughly the same. Second, as Julesz observed in his studies, it often happens that, depending on the observed elements, not all the descriptive features participate to the perceptual categorization process. Not only useless for categorization, the remaining non-discriminative features also add noise in our sparse feature space which suffers from the usually low number of elements provided by the user. Identifying noise-inducing features is then crucial for ensuring a robust appearance-driven analysis.

Our categorization scheme thus falls into two stages. First, we categorize elements according to their area. Second, for each of the obtained groups considered individually, we perform another categorization step according to the elements' dominant appearance, computed via dimensionality reduction on the remaining four appearance features. Figure 3 illustrates this two-step scheme.

Element grouping according to element area and appearance is established by detecting relevant modes of those two features' density that we approximate by histograms evaluated over the reference arrangement. Because of the lack of prior concerning the number of expected categories, we rely on the *a contrario* method proposed in [Desolneux et al. 2003] as our mode-seeking procedure. We recall this method in the following section to make the paper self-contained.

***A contrario* Methods** *A contrario* approaches have been successfully applied to many Computer Vision problems among which the analysis of histogram modes. The main insight is to rely on a general perception law called the Helmholtz principle which states that an event is perceptually meaningful if it is unexpected. More formally, if the expectation of its occurrences is low under a random assumption.

In the case of histogram analysis, the random assumption is that the descriptor values are i.i.d. uniformly in the L histogram bins $\{b_1, \ldots, b_L\}$. Let us consider an interval noted $S_{i,j} = \{b_i, \ldots, b_j\}$ with $i \leq j$. The prior probability $p_{i,j}$ that an element has its feature descriptor in $S_{i,j}$ is then $p_{i,j} = (b_j - b_i + 1)/(b_L - b_1 + 1)$. Following [Desolneux et al. 2003], we define $S_{i,j}$ as an *ε-meaningful interval* if

$$N\mathcal{B}(p_{i,j}, N_E, k_{i,j}) < \varepsilon \tag{1}$$

where $N = L(L+1)/2$ is the number of possible connected sets of bins; N_E is the number of input elements; $k_{i,j}$ denotes the number of elements in $S_{i,j}$, and $\mathcal{B}$ is the tail of the binomial distribution:

$$\mathcal{B}(p, n, k) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

The quantity $N\mathcal{B}(p_{i,j}, N_E, k_{i,j})$ can be interpreted as the expectation of the bins from $S_{i,j}$ to occur by pure chance. If this estimate is very low, such bins constitute a meaningful interval. The ε parameter has been shown to cause a logarithmic ε-dependency on meaningfulness, making such approaches robust with respect to their unique parameter [Desolneux et al. 2000]. When set to 1, this leads to the following intuitive interpretation: bins appearing at least once in a random situation are considered as meaningful events.

In order to separate modes inside meaningful intervals, we can similarly define *ε-meaningful gaps* within the distribution histogram as the intervals containing fewer points than the expected average. We say that $S_{i,j}$ is an ε-meaningful gap if

$$N\mathcal{B}(1 - p_{i,j}, N_E, N_E - k_{i,j}) < \varepsilon$$

A *meaningful mode* is defined as a meaningful interval that does not include of meaningful gaps. Lastly, in order to forbid the case of non-disjoint meaningful modes, a meaningful mode is said to be *maximal* if it does not contain, and is not contained in another mode showing greater meaningfulness. Maximal meaningful modes are mentioned as MM-modes in the rest of the paper.

Categorization Algorithm Using *a contrario* histogram mode detection, we then obtain the following categorization algorithm.

First, we compute the MM-modes of the element area histogram estimated over the complete arrangement. This provides us with a preliminary set of categories. Any connected interval of bins which does not belong to an MM-mode is considered as an outlier area category.

Second, for each of these categories (including the possible outlier categories), we perform dimensionality reduction on the four remaining appearance features by Robust Principal Component Analysis [Hubert et al. 2002]. We then identify the MM-modes of these features after their projection onto their first principal component. Each found MM-mode defines an appearance category. Similarly to step one, for each appearance histogram, any connected interval of bins which does not belong to an MM-mode is considered as an outlier appearance category.

In the context of example-based methods, we consider arrangements that have a rather low number of elements, typically below a hundred, and thus a low number of distinct appearance categories. This restricts the precision of the histograms we can analyze. In all our experiments, the distribution of features is estimated on 10 bins, but the discretization scheme can be made more accurate as more input elements are provided by the artist. Likewise, if an area category contains less than 10 elements, we do not split it any further.

3 Statistical Arrangement Modeling

This section presents the statistical process that models the spatial arrangement of categorized elements. Once the parameters of this model are learned on the reference arrangement, the synthesis step consists in running realizations of this model at the desired scale, shape, or density needed by the user.

3.1 Multitype Point Process Model

With the input's appearance categories at hand, we now investigate the elements' relative positions from the perspective of their visual aspect. For that aim, we propose to capture their spatial arrangement via a multitype point process, a statistical model dedicated to the analysis of interactions between a finite set of typed categories. By considering pair-wise element distances as interactions between our established categories, we implicitly grasp the underlying correlation between the elements' appearance and their spatial organization. This model accounts for the interactions gathered over the whole input and supports a wide range of distributions, from stochastic to near-regular.

In our specific case, we assimilate the point data resulting from a realization of this model to the set of the N_E input elements $\mathbf{x} = \{x_1, \dots, x_{N_E}\}$. Given an element $x_i \in \mathbf{x}$, we associate its corresponding appearance category label m_i to it, m_i being taken from the N_C possible categories labels stored in the label set $\mathcal{M}$. It should be noted that, since $N_C < N_E$, the labels m_i, m_j, may refer to the very same appearance category even though they are related to two distinct elements x_i and x_j.

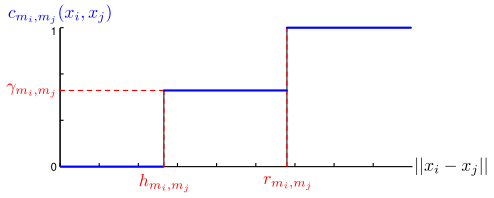
A way to construct a point process model is to write down its probability density function (PDF) with respect to a Completely Random Situation. Such point processes are called *Gibbs point processes* and offer many advantages. Manipulating their PDF to make them account for intricate interactions is easy and further simulation is ensured by well-known Monte-Carlo Markov Chain algorithms. Since we focus here on pair-wise interactions between element categories, we can define our model's PDF, noted $f(\mathbf{x})$, as follows (see [Ripley 1981] for further details):

$$f(\mathbf{x}) \propto \left[\prod_{x_i} d_{m_i}(x_i)\right] \left[\prod_{x_i \neq x_j} c_{m_i,m_j}(x_i, x_j)\right] \qquad (2)$$

where $d_m(\cdot)$ is the occurrence probability function of elements from the m category and $c_{m,m'}(\cdot,\cdot)$ is the interaction probability function between the m and m' categories.

A good rule of thumb for statistical modeling is to exploit models whose number of parameters does not exceed the number of observed data. Here, we thus use the simple *Strauss hard-core interaction* which directly relates interaction probability between appearance categories to the Euclidean distance between their elements:

$$c_{m_i,m_j}(x_i, x_j) = \begin{cases} 0 & \text{if } ||x_i - x_j|| < h_{m_i,m_j} \\ \gamma_{m_i,m_j} & \text{if } h_{m_i,m_j} \leq ||x_i - x_j|| < r_{m_i,m_j} \\ 1 & \text{otherwise} \end{cases}$$



The explicit definition of the interaction probability function of a category pair $m, m' \in \mathcal{M}$ then requires the estimation of three constant parameters noted $h_{m,m'}$, $r_{m,m'}$, and $\gamma_{m,m'}$. The first two are distance thresholds, called *hard-core distance* and *trend threshold* respectively. The last one is a scalar in range $[0, 1]$ defining the *interaction strength* and its tuning enables us to model a variety of arrangements from regular to random. Since our proposed interaction functions are symmetric, we just need to evaluate $3N_c(N_c + 1)/2$ interaction parameters to completely define our statistical arrangement model, with N_C being the total number of appearance categories.

3.2 Estimation of the Model Parameters

The multitype Strauss hard-core model is a generic descriptive model that can reproduce various spatial arrangements. This diversity is embedded in the parameters that need to be estimated from the input arrangement by likelihood maximization.

Given the limited set of provided elements, we need to make an important simplifying assumption to ensure a tractable statistical fitting. We suppose the reference arrangement is *stationary* which intuitively comes down to presuming that the artist draws homogeneously over the reference surface. Our re-synthesis still allows the creation of inhomogeneous element distributions. This simplification allows us to treat the categories' occurrence probability functions $d_m(\cdot)$ as constants during the estimation of the parameters. We denote this set of constants Δ.

Moreover, the statistical approach we adopt to estimate the parameters of the multitype point process is hazardous for extremely small categories. In practice, we assume that the user did not draw groups of similar elements containing less than three elements.

3.2.1 Hard-core Distances $h_{m,m'}$ Between Category Pairs

Given a pair of appearance categories $m, m' \in \mathcal{M}$, the hard-core distance obtained by likelihood maximization estimation $h^*_{m,m'}$

corresponds to the minimum distance between pairs of elements picked from the specified categories:

$$h^*_{m,m'} = \min_{\substack{x_i,x_j \\ m_i=m \\ m_j=m'}} ||x_i - x_j||$$

3.2.2 Trend Distances $r_{m,m'}$ Between Category Pairs

Maximizing the trend distances' likelihood estimator is more involving and intuitively corresponds to finding the circular window radius from which the reference arrangement is seen to be the most regular.

To compute that radius value, we use Ripley's $\mathcal{L}$ function which quantifies the deviation of the arrangement, when investigated at a specified scale, relative to a Completely Random Situation [Ripley 1981]. Here follows its formulation :

$$\mathcal{L}_{m,m'}(r) = \sqrt{\frac{\mathcal{K}_{m,m'}(r)}{\pi}} - r$$

$\mathcal{K}_{m,m'}(r)$ is the expected number of elements from the m category lying at a distance r of a randomly picked element of the m' category. As such, it gives an estimate of the element density evaluated at a given scale of a category with respect to another and is normalized in a way that a purely random distribution yields a constant value $\mathcal{L}_{m,m'}(r) = 0$ for all r. Distributions which exhibit more regularity present a negative $\mathcal{L}_{m,m'}$ profile. We thus look for r^*, the first value for which $\mathcal{L}_{m,m'}$ reaches a local minimum. This attests that regularity occurs with maximal amplitude at that scale.

3.2.3 Interaction Strengths $\gamma_{m,m'}$ and Category Occurrence Probabilities d_m

The estimation of the remaining models parameters involves the maximization of the PDF of our model evaluated over the reference arrangement $f(\mathbf{x})$. Finding the optimal parameter sets $\Gamma^* = (\gamma^*_{m,m'})$ and $\Delta^* = (d^*_m)$ comes down to find the best "explanation" by our statistical model of the observed input. However, as Equation (2) suggests, $f(\mathbf{x})$ is defined up to a normalization constant whose explicit evaluation is intractable. To circumvent this problem, we instead maximize the following log pseudo-likelihood involving ratios of f:

$$\sum_{x_i} \log\left(\frac{f(\mathbf{x})}{f(\mathbf{x}/\mathbf{x_i})}\right) - \frac{1}{N_c}\sum_{m=1}^{N_c}\int_{W_\mathcal{R}} \frac{f(\mathbf{x} \cup u_m)}{f(\mathbf{x})} du \qquad (3)$$

where $W_\mathcal{R}$ corresponds to the input drawing window and u_m to an element from the m^{th} appearance category. The involved PDF ratios can be understood as such: given the fixed element distribution $\mathbf{x}$, they quantify the conditional probability of observing an element at a specified location u. The first term of Equation (3) favors locations where observed elements actually lie, while the second term penalizes all the other locations within the drawing window W. The integral is usually estimated using a grid on $W_\mathcal{R}$ where locations u are the centers of each grid cell weighed by its surface. In our experiments, we use a regular grid.

This formula was first proposed by [Besag 1977] and later extended by [Jensen and Møller 1991]. Its suitability to a wide range of Gibbs point processes has been recently proved by [Billiot et al. 2008]. It admits a unique extremum in the (Γ, Δ) parameter space which we find using a Newton-Raphson approach.

3.3 Synthesis by Markov chain Monte-Carlo

As stated in Section 3.1, one noticeable strength of Gibbs point process models is their easy simulation using Markov chain Monte Carlo methods. This interesting property provides us with a convenient means to generate new arrangements that apparently obey the same stochastic process as the provided input. Since we cannot directly sample from $f(\mathbf{x})$, we construct a Markov chain whose set of vertices coincides with the set of elements from the reference arrangement $\mathbf{x}$ and whose equilibrium distribution is to converge to our fitted model's PDF $f(\mathbf{x})$.

We can now compute new realizations of our statistical model over a synthesis window $W_\mathcal{S}$, namely new element arrangements, by using a variant of the Metropolis-Hastings algorithm adapted to point processes [Geyer and Møller 1994].

- randomly initialize output arrangement $\mathbf{x}_0 = \mathbf{x}$ s.t. $f(\mathbf{x}) > 0$
- for time-steps t from 1 to T do:
  - alter current arrangement $\mathbf{x}_t$ using one of the two following equiprobable perturbations:
  Element birth:
    - add an element u at random location in $W_\mathcal{S}$
    - assign random category label m_u to u
    - candidate arrangement $\mathbf{x}' = \mathbf{x}_t \cup \{u\}$
    - compute acceptance rate $R_b = \frac{f(\mathbf{x}')}{f(\mathbf{x}_t)} \frac{n_t}{A_t}$
  Element death:
    - pick random element u from $\mathbf{x}_t$
    - candidate arrangement $\mathbf{x}' = \mathbf{x}_t / \{u\}$
    - compute acceptance rate $R_d = \frac{f(\mathbf{x}_t)}{f(\mathbf{x}')} \frac{A_t}{n_t}$
  - if $R_{b/d} > 1$, then accept perturbation $(\mathbf{x}_{t+1} \leftarrow \mathbf{x}')$
  else accept perturbation $(\mathbf{x}_{t+1} \leftarrow \mathbf{x}')$ with a probability $R_{b/d}$
  otherwise keep current arrangement unchanged $(\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t)$

Figure 4: *Arrangement synthesis by Metropolis-Hastings sampling*

In the pseudo-code provided Figure 4, we denote $\mathbf{x}_t$ the state of our Markov chain at time-step t. For a specified number T of iterations ($T = 10^5$ in our experiments), we slightly perturb $\mathbf{x}_t$ by introducing or removing one element and obtain a new candidate state for the chain $\mathbf{x}'$. These elementary perturbation events, respectively coined the *birth* or *death*, are effectively taken into account if they satisfy an acceptance rate criterion. Acceptance rates for births and deaths, called R_b and R_d, depend of the ratio of the model's PDF evaluated over $\mathbf{x}_t$ and $\mathbf{x}'$, as well as the current arrangement's area A_t and element number n_t.

The simulation output is a spatial distribution of category labelled elements. We finalize our synthesized arrangement by directly pasting onto each output element's location a reference element randomly picked from the correct appearance category.

4 Results and Discussion

We now present some results and put our technique into perspective with previous methods before discussing its current shortcomings.

4.1 Experimental Results

Examples of categorization and synthesis are shown Figure 5. Those examples attest that the Strauss hard-core process can reproduce various kinds of element distributions, from fairly regular to completely random (e.g., Figures 5-a and 5-e respectively). Thanks to the global multitype optimization procedure, distances between elements are adjusted according to the interactions occurring within

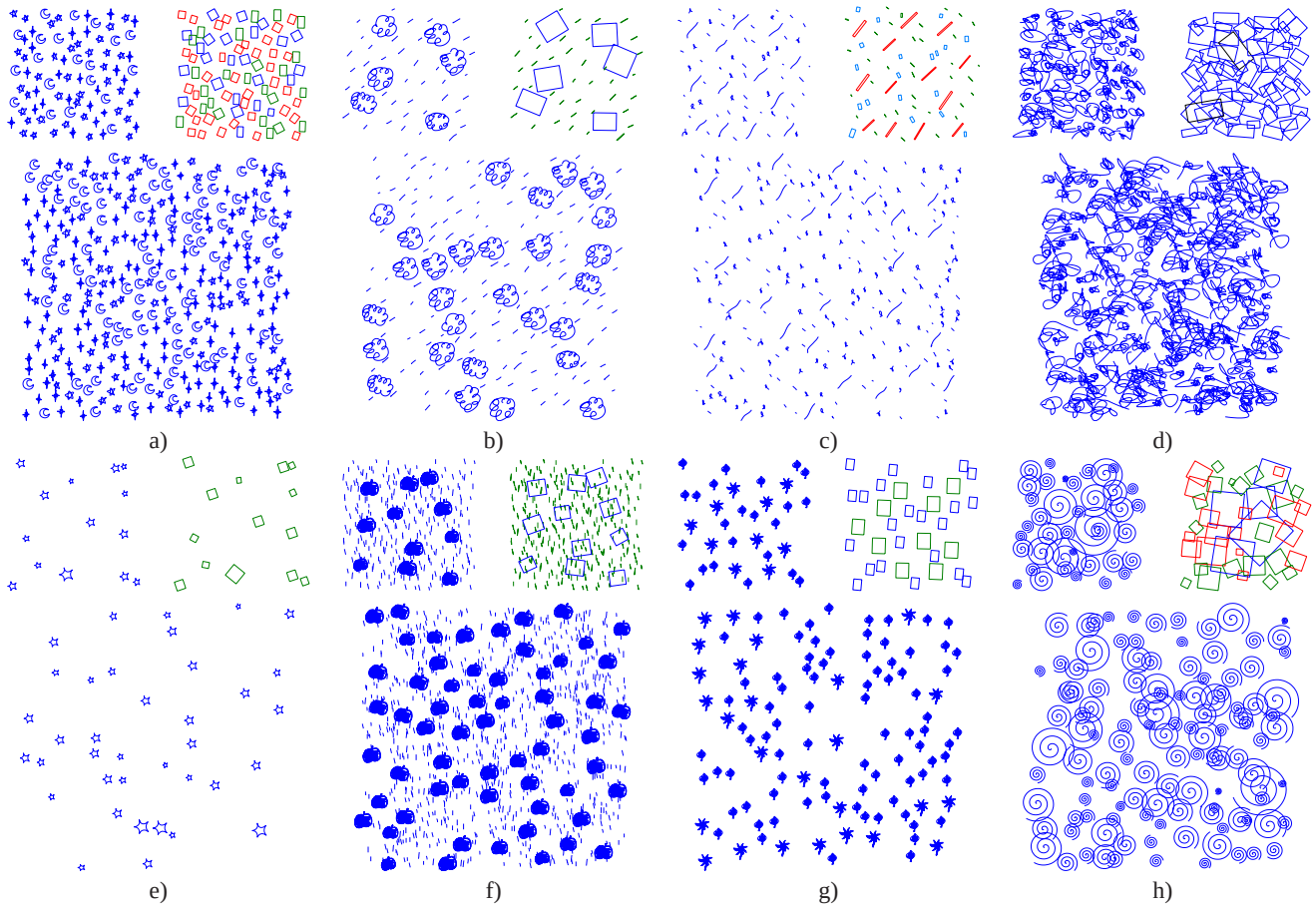


Figure 5: *Synthesis results attesting the variety of spatial distributions that our model can handle. The multitype Strauss hard-core model captures distributions ranging from fairly regular to random,* (a) *and* (e) *respectively. This diversity can be observed inside each category of elements. For instance, apples in* (f) *are regularly distributed, whereas the background is randomly arranged. Similarly, interactions between categories can also vary from repulsive to random,* (g) *and* (h) *respectively.*

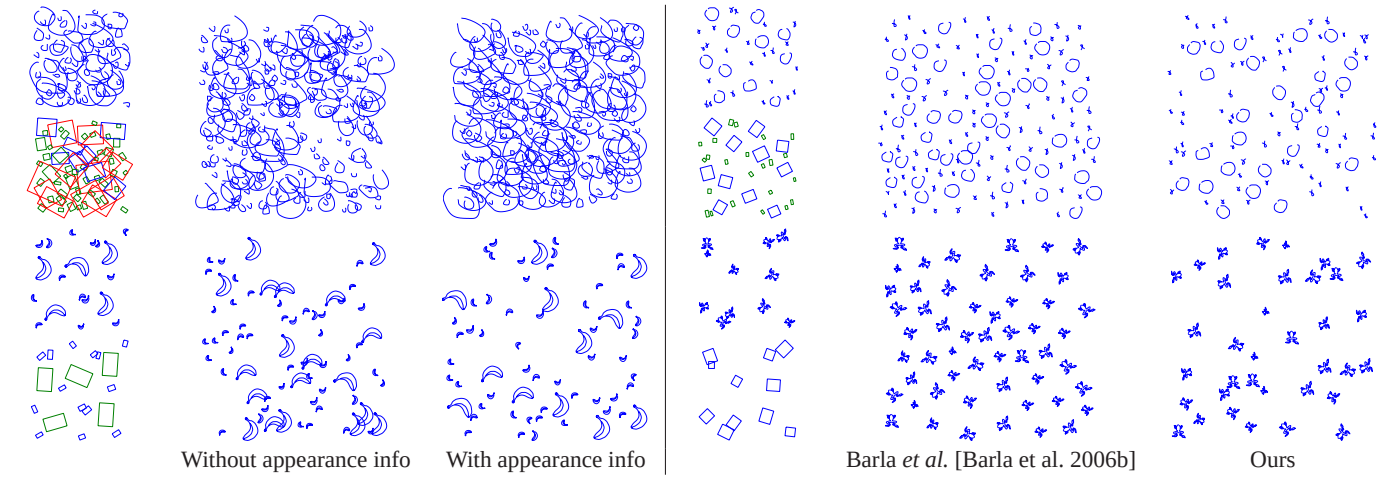


Figure 6: Comparisons (Left) *These examples motivate our appearance-driven categorization. All arrangements are generated by a Strauss interaction model, the ones in the center via a monotype point process overlooking interactions between elements of different appearance. Note the apparition of holes or overlaps which are absent from the input.* **(Right)** *The upper reference is the combination of a more and less regularly distributed sets of elements (the circles and crosses respectively). The approach in [Barla et al. 2006b] enforces regularity over the output while our approach captures the element interactions and preserves them. The lower example consists of a non-uniform arrangement which cannot be represented by the distribution obtained after Lloyd relaxation. A similar phenomenon would arise if using the seeding procedure by [Ijiri et al. 2008], as close seeds are merged below some distance threshold and additional seeds are created in empty regions.*

and between categories. This important property is especially visible on Figure 5-b where the large elements push back the smaller ones beyond the hard-core distance estimated between the two involved categories. Techniques relying on regular element distributions cannot respect such placement constraints, and in this case, element overlap would then occur.

Once our model's parameters have been estimated over the input, new arrangements can easily been synthesized onto various shapes, with possibly different element densities. A gray-level brush-like tool can then be used to intuitively draw arrangements of user-specified densities as illustrated on Figure 7.

Figure 7: *Example of user-drawn brush strokes whose intensity grants effective control over the synthesized densities of elements.*

In terms of performances, both appearance categorization and statistical fitting are interactive, the bottleneck of the method being the Metropolis-Hastings sampling procedure used for re-synthesis. All examples provided in the paper take about 5 seconds to be categorized and generated on a standard PC. Improvements can be investigated to reduce this computational load. More sophisticated perturbations in the MCMC procedure, such as translation and rotation of elements, could be investigated [Green 1995]. We could also use a spatially discretized grid during sampling, since our application does not meet accuracy requirements as high as the classical uses of such statistical tools.

4.2 Comparisons With Related Work

Improved Handling of Appearance The results on the left side of Figure 6 demonstrate the importance of our appearance analysis step. Arrangements generated by a *monotype* Strauss hard-core model, which considers all elements as visually equivalent, are not fully satisfactory. Even though the overall spatial distribution of the elements' locations is captured, undue holes or overlaps occur and compromise the resemblance of the results with the provided reference.

[Barla et al. 2006b] do account for visual similarity to some extend, by notably defining a perceptual distance used to compare element neighborhoods defined over the Delaunay graph and pick the input element to stitch to a given output location. Their measure, however, only relies on the elements' bounding boxes, whereas our approach integrates more perceptual features enabling the distinction of elements of comparable bounding box. The handling of the elements' appearance by [Ijiri et al. 2008] serves the very same purpose of guiding neighborhood matching. They do not provide any automatic analysis method though, as they require the user to manually label the different elements. Finally, our technique directly correlates statistics between appearance and spatial locations of the elements over the whole input and not just local neighborhoods.

Supported Element Distributions As already presented Figure 5, our method can faithfully reproduce a wide spectrum of element distributions. We notably capture *non-uniform* distributions, which is an improvement over the approaches proposed by Barla *et al.* and Ijiri *et al.*. Re-synthesis results displayed on the right side of Figure 6 attest that fact. Those consist of sets of irregularly distributed elements whose spatial organization cannot be captured by the point distribution resulting from a Llyod relaxation used in [Barla et al. 2006b]. As such, Barla's output arrangements always seem to follow an underlying hexagonal lattice structure. The procedural technique suggested by Ijiri *et al.* also leads to similar results as they devise several growing rules – such as the seed merging and empty space-filling rules – that force the output Delaunay triangulation to be unskewed.

The special case of *strongly regular* distributions is different, as it is not properly supported by our current Strauss model. While this shortcoming also exists in [Barla et al. 2006b], specific care has been provided by [Ijiri et al. 2008] to handle such arrangements, but to the price of user intervention, such as the manual correction of the extracted Delaunay triangulation. In our case, this issue is related to the fact we only consider pair-wise interactions. We believe that considering statistics of interactions involving more than two elements could help us lift that limitation.

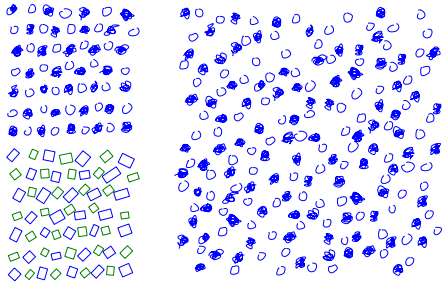


Figure 8: *This example illustrates the main limitation of the Strauss model. Since it only accounts to second order interactions, it cannot reproduce well strongly regular arrangements such as this reference whose elements are pasted onto on a rectangular lattice. More sophisticated models that uses higher order interactions could be investigated to push back this limitation.*

Automation vs. Versatility By locally modifying its element density parameters, our model inherently proposes some intuitive handles for the user to design the synthesized arrangements. For instance, it is straightforward to make our output distributions follow a specified path, typically drawn by the user via a intensity-varying brush tool. This allows the same kind of expressiveness as the *spray* and *boundary tools* in [Ijiri et al. 2008]. However, our local element density control is novel.

Further control on the elements' orientation, such as randomization or flow-guided harmonization, could also be added. Those would take the form of post-processing steps, however, since incorporating too many parameters would endanger the tractability of our statistical fitting. Considering other appearance features, like elements' size or color, could also lead to interesting results and is currently left to future work. Indeed, the focus of the present article is to increase the amount of automatically extractable information from a provided example and capture the interplay between the appearance and the spatial organization of a set of observed elements.

4.3 Limitations and Future Work

A group of perceptually similar elements can sometimes be over-categorized. This effect has yet no consequence on the synthesis step since it leads our model fitting to infer interactions between similar objects stored in different appearance groups. These interactions are reproduced in the synthesized arrangement, but remain unnoticed –as in the reference arrangement– because of the perceptual similarity of the elements (Figure 9-left). Actually, over-categorization on that specific example involves very small ele-

ments and a closer inspection on those tiny shapes does indicate differences between them. Such dissimilarities are hardly visible without explicit zooming though.

Another limitation of our approach is the spatial representation of the elements by their respective centroid. When elements are strongly elongated, this representation is not adapted and is misleading for the model. This is visible in Figure 9-right showing a hatching arrangement. The centroid distribution is well reproduced. Yet, interactions between centroids is not representative of interactions between the actual elements. This drawback could be circumvented by using the Hausdorff distance between bounding boxes instead of the Euclidean distance as the parameter for the interaction functions. Besides, the synthesis of such hatching patterns implies to answer another ill-posed problem. Would the artist expect the system to cover the output window with strokes picked from the example, or by directly adapt their length ?

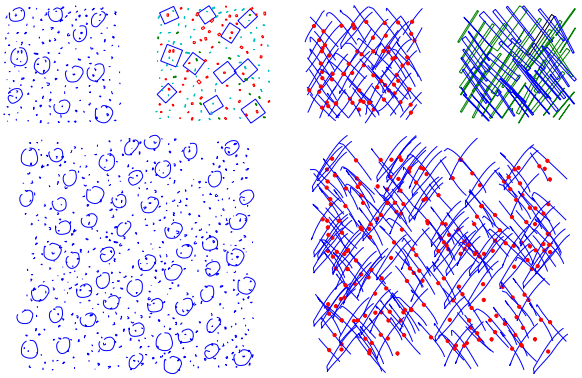


Figure 9: Limitations (Left) *Visually similar elements may sometimes end up in more categories than necessary, such as the small elements displayed here which fall into three distinct categories instead of one* (upper right). *We call that phenomenon* ***over-categorization****. It has close to no impact on the visual quality of our synthesis results, as the system then strives to reproduce unnoticeable interactions between categories containing resembling elements.* **(Right)** *Since elongated elements are not well represented by their sole centroids (red dots), our interaction model based on point-wise distances does not accurately account for the actual interactions between elements. It should noted that the distribution of the centroids is yet preserved.*

5 Conclusions

We have presented an example-based method to synthesize arrangements of vector elements that combines the appearance-guided categorization of the elements and the statistical modelling of the spatial interactions occurring within and between appearance categories. The categorization step is based on several perceptual principles and it could be profitably exploited in other methods such as the procedural approach of [Ijiri et al. 2008]. To the best of our knowledge, our statistical modelling for 2D element arrangements is novel. We believe that multitype point processes –and marked point processes in general– constitute interesting and flexible theoretical tools that could be further investigated.

Acknowledgements: we thank Xavier Descombes and Florence Forbes for their inspiring discussions.

References

ASHIKHMIN, M. 2001. Synthesizing natural textures. *I3D '01*, 217–226.

BARLA, P., BRESLAV, S., MARKOSIAN, L., AND THOLLOT, J. 2006. Interactive hatching and stippling by example. Tech. rep., INRIA.

BARLA, P., BRESLAV, S., THOLLOT, J., SILLION, F., AND MARKOSIAN, L. 2006. Stroke Pattern Analysis and Synthesis. *EUROGRAPHICS '06 25*, 3, 663–671.

BESAG, J. 1977. Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute 47*, 2, 77–92.

BILLIOT, J.-M., COEURJOLLY, J.-F., AND DROUILHET, R. 2008. Maximum pseudolikelihood estimator for exponential family models of marked gibbs point processes. *Electronic Journal of Statistics 2*, 234–264.

BRUNN, M., SOUSA, M. C., AND SAMAVATI, F. F. 2007. Capturing and re-using artistic styles with reverse subdivision-based multiresolution methods. *International Journal of Image and Graphics '07 7*, 4, 593–615.

DESOLNEUX, A., MOISAN, L., AND MOREL, J.-M. 2000. Meaningful alignments. *International Journal of Computer Vision 40*, 7–23.

DESOLNEUX, A., MOISAN, L., AND MORE, J. 2003. A grouping principle and four applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 4, 508–513.

DISCHLER, J.-M., MARITAUD, K., LÉVY, B., AND GHAZANFARPOUR, D. 2002. Texture particles. *EUROGRAPHICS '02 21*, 401–410.

EFROS, A. A., AND LEUNG, T. K. 1999. Texture synthesis by non-parametric sampling. *ICCV '99*, 1033–1038.

FREEMAN, W. T., TENENBAUM, J. B., AND PASZTOR, E. 2003. Learning style translation for the lines of a drawing. *ACM Transactions on Graphics 22*, 1, 33–46.

GEYER, C., AND MØLLER, J. 1994. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, 359–373.

GREEN, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 4, 711–732.

GUO, G.-E., ZHU, S.-C., AND WU, Y. 2001. Visual learning by integrating descriptive and generative methods. *ICCV '01 1*, 370–377.

HEEGER, D. J., AND BERGEN, J. R. 1995. Pyramid-based texture analysis/synthesis. *SIGGRAPH '95*, 229–238.

HERTZMANN, A., OLIVER, N., CURLESS, B., AND SEITZ, S. M. 2002. Curve analogies. *Eurographics workshop on Rendering '02*, 233–246.

HUBERT, M., ROUSSEEUW, P., AND VERBOVEN, S. 2002. A fast method for robust principal components with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems 60*, 1-2, 101–111.

IJIRI, T., MECH, R., IGARASHI, T., AND MILLER, G. 2008. An Example-based Procedural System for Element Arrangement. *EUROGRAPHICS '08 27*, 2, 429–436.

JENSEN, J., AND MØLLER, J. 1991. Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability 1*, 445–461.

JODOIN, P.-M., EPSTEIN, E., GRANGER-PICHÉ, M., AND OSTROMOUKHOV, V. 2002. Hatching by example: a statistical approach. *NPAR '02*, 29–36.

JULESZ, B. 1986. Texton gradients: The texton theory revisited. *Biological Cybernetics 54*, 4, 245–251.

KALNINS, R. D., MARKOSIAN, L., MEIER, B. J., KOWALSKI, M. A., LEE, J. C., DAVIDSON, P. L., WEBB, M., HUGHES, J. F., AND FINKELSTEIN, A. 2002. Wysiwyg npr: Drawing strokes directly on 3d models. *SIGGRAPH '02*, 755–762.

LEFEBVRE, S., AND HOPPE, H. 2006. Appearance-space texture synthesis. *SIGGRAPH '06 25*, 541–548.

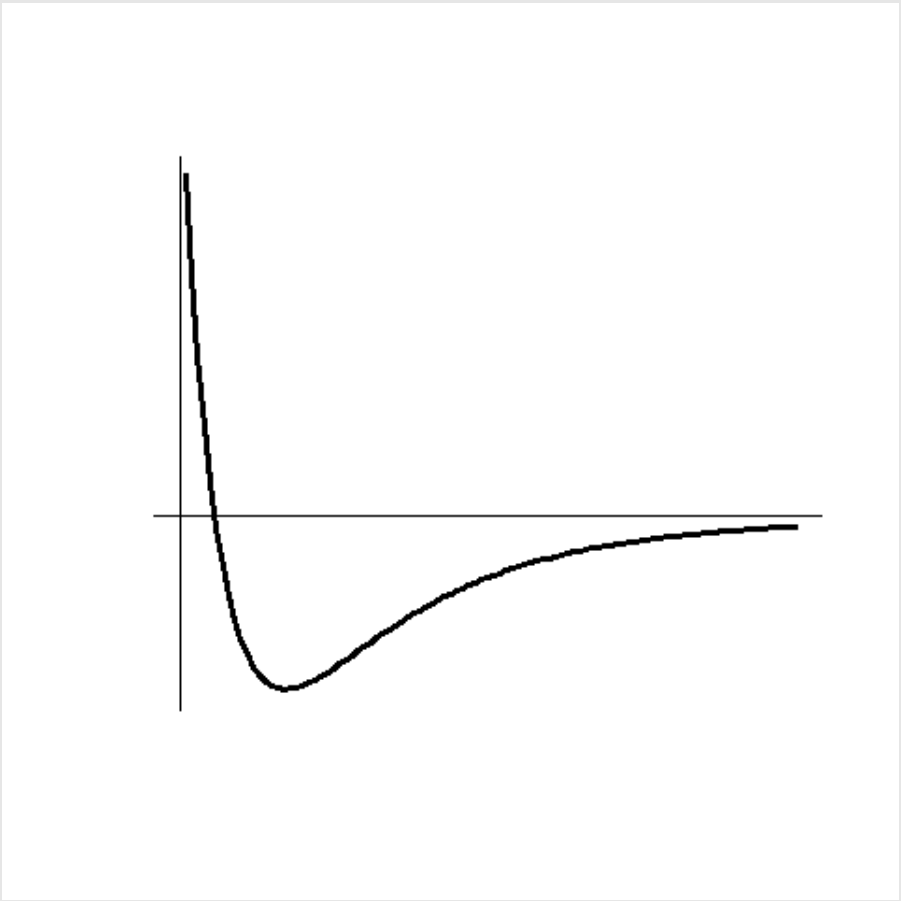
LIU, Y., COLLINS, R. T., AND TSIN, Y. 2004. A computational model for periodic pattern perception based on frieze and wallpaper groups. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26*, 354–371.

LIU, Y., LIN, W.-C., AND HAYS, J. 2004. Near-regular texture analysis and manipulation. *SIGGRAPH '04 23*, 368–376.

RIPLEY, B. D. 1981. *Spatial statistics*. Wiley New York.

WEI, L.-Y., AND LEVOY, M. 2000. Fast texture synthesis using tree-structured vector quantization. *SIGGRAPH '00*, 479–488.

WU, Q., AND YU, Y. 2004. Feature matching and deformation for texture synthesis. *SIGGRAPH '04 23*, 364–367.



[CD10]
J.-F. Coeurjolly and R. Drouilhet. Asymptotic properties of the maximum pseudo-likelihood estimator for stationary Gibbs point processes including the Lennard-Jones model. *Electronic Journal of Statistics*, 4:677-706 (2010).

Electronic Journal of Statistics
Vol. 4 (2010) 677–706
ISSN: 1935-7524
DOI: 10.1214/09-EJS494

Asymptotic properties of the maximum pseudo-likelihood estimator for stationary Gibbs point processes including the Lennard-Jones model

Jean-François Coeurjolly

Gipsa-lab, Grenoble University, France
LJK, Grenoble University, France
e-mail: Jean-Francois.Coeurjolly@upmf-grenoble.fr

Rémy Drouilhet

LJK, Grenoble University, France
e-mail: Remy.Drouilhet@upmf-grenoble.fr

Abstract: This paper presents asymptotic properties of the maximum pseudo-likelihood estimator of a vector parameterizing a stationary Gibbs point process. Sufficient conditions, expressed in terms of the local energy function defining a Gibbs point process, to establish strong consistency and asymptotic normality results of this estimator depending on a single realization, are presented. These results are general enough to no longer require the local stability and the linearity in terms of the parameters of the local energy function. We consider characteristic examples of such models, the Lennard-Jones and the finite range Lennard-Jones models. We show that the different assumptions ensuring the consistency are satisfied for both models whereas the assumptions ensuring the asymptotic normality are fulfilled only for the finite range Lennard-Jones model.

AMS 2000 subject classifications: Primary 60G55; secondary 60J25.
Keywords and phrases: Stationary Gibbs point processes, maximum pseudo-likelihood estimator, Lennard-Jones model.

Received September 2009.

This paper studies a method to estimate the parameters governing the distribution of a stationary marked Gibbs point process.

1. Introduction

These last years, much attention has been paid to spatial point pattern data, and especially to models and methodologies to fit them, see Møller (2008) for a recent overview of this topic and Daley and Vere-Jones (1988), Stoyan et al. (1987) Møller and Waagepetersen (2003) or Illian et al. (2008) for more general information. For spatial point pattern data, the reference model is the Poisson point process modelling a random configuration of points with no interaction

between points. In particular, this leads to the independence of any two random sub-configurations lying in two non-overlapping domains. A way to introduce dependence is to consider the class of Gibbs models. In a bounded domain, a Gibbs point process is defined by its probability measure whose density with respect to a Poisson point process measure is proportional to $e^{-V(\varphi)}$ where $V(\varphi)$ corresponds to the energy function (i.e. a cost function expressed in terms of interactions) of the configuration of points φ.

In the framework of parametric Gibbs models, when interested in asymptotic properties of estimators, it is essential to extend the definition of Gibbs models to $\mathbb{R}^d$. The probability measure of a Gibbs point process in $\mathbb{R}^d$ has to be defined by specifying its conditional density (indirectly expressed in terms of the energy function $V(\varphi)$), see *e.g.* Preston (1976) or Section 2 for more details.

The class of Gibbs point processes is extremely rich. The energy function can penalize points, pairs or triplets of points (see *e.g.* Baddeley and Turner (2000)). More sophisticated models can also be obtained by considering interactions based on the Delaunay or the $k-$nearest neighbor graphs (Bertin et al. (1999c,b)), Voronoï tessellations (Dereudre and Lavancier (2009)) or random sets (Kendall et al. (1999), Dereudre (2009)).

Following the definition of a parametric Gibbs point process, the natural question of efficiently estimating the parameters arises. Many proposals have tried to estimate the energy function from an available point pattern data. The most well-known method is the use of the likelihood function, see *e.g.* Møller and Waagepetersen (2003) and the references therein. The main drawback of this approach is that the likelihood function contains an unknown scaling factor whose value depends on the parameters. This parametric normalizing constant is difficult to calculate from a practical point of view. From a theoretical one, it also makes asymptotic results more complicated to obtain. An alternative approach relies on the use of the pseudo-likelihood function. The idea originated from Besag (1974) in the study of lattice processes. Besag et al. (1982) further considered this method for pairwise interaction point processes, and Jensen and Møller (1991) extended the definition of the pseudo-likelihood function to the general class of marked Gibbs point processes. The construction of the pseudo-likelihood function is based on the conditional densities which spare the computation of the scaling factor.

Our paper deals with asymptotic properties of the maximum pseudo-likelihood estimator. In order to underline our theoretical improvements, let us discuss the two main different papers discussing this topic:

- In Billiot et al. (2008), we obtain consistency and asymptotic normality for exponential family models of Gibbs point processes, that is, on models with energy functions that are linear in terms of the parameters. Moreover, we concentrate on models such that the local energy function is local and stable. The locality of the local energy expresses that the energy to insert a point x into φ, that is, $V(x|\varphi) = V(\varphi \cup x) - V(\varphi)$, depends only on the points of φ falling into some ball with a fixed radius whereas the stability of the local energy (property referred as the local stability) asserts

that $V(x|\varphi)$ is bounded from below by a finite negative constant. The paper Billiot et al. (2008) extends several papers (Jensen and Møller (1991), Jensen and Künsch (1994)) and includes a large class of examples of practical interest: area-interaction point process, Multi-Strauss marked point process based on the complete graph or the k-nearest-neighbors graph, or the Geyer's triplet point process to name a few.

- Another work has been undertaken by Mase. The consistency for non necessarily stable local energy functions (actually for superstable and lower regular ones introduced by Ruelle (1970)) is obtained in Mase (1995) for specific models with only two parameters -the chemical potential and the inverse temperature- which can be viewed as particular exponential family models. Mase (2000) extended his work to the context of marked point processes and provided asymptotic normality by adding the assumption of finite range.

Based on this literature, the main goal of this paper is to derive asymptotic properties similar to the ones presented before (consistency and asymptotic normality) but in a more general framework. We provide asymptotic results for general Gibbs point processes with *non (necessarily) linear* and *non (necessarily) stable* local energy functions. The characteristic example we have in mind is the *Lennard-Jones model*. This model, from statistical physics, is a stationary pairwise interaction Gibbs point process where the local energy to insert a point x into a configuration φ is parameterized as follows: for $\theta = (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3$ with $\theta_2, \theta_3 > 0$

$$V^{LJ}(x|\varphi;\theta) := \theta_1 + 4\theta_2 \sum_{y\in\varphi} \left(\left(\frac{\theta_3}{\|y-x\|} \right)^{12} - \left(\frac{\theta_3}{\|y-x\|} \right)^{6} \right).$$

Let us notice that Mase (1995) could only propose the estimation of θ_1 and θ_2 with known θ_3. The Lennard-Jones model is of great interest from several points of view. From a physical point of view, this model arises when theoretically modelling a pair of neutral atoms or molecules subject to two distinct forces in the limit of large separation and small separation: an attractive force at long ranges (van der Waals force, or dispersion force) and a repulsive force at short ranges (the result of overlapping electron orbitals, referred to as a Pauli repulsion from the Pauli exclusion principle). In this literature, the parameters θ_2 and θ_3 are often referred to as the depth potential and the (finite) distance at which the interparticle potential is zero. From a probabilistic point of view, this model constitutes the main example of superstable, regular and lower regular energies studied in Ruelle (1970) where the author proves the existence of ergodic measures for such models. Finally, from a statistical point of view, this model has been considered by several authors, see *e.g.* Ogata and Tanemura (1981), Goulard et al. (1996) for fitting spatial point patterns arising in forestry. In particular, let us note that, in Goulard et al. (1996), the model is fitted by using the maximum pseudo-likelihood method. As the authors do not endeavour to justify the theoretical performances of the procedure, the result proposed in Section 3.4 of this paper fills this gap.

The rest of the paper is organized as follows. Section 2 introduces some background and notation on Gibbs point processes (general definitions, examples). The maximum pseudo-likelihood method and asymptotic results of the derived estimator are proposed in Section 3. For general Gibbs point processes, sufficient conditions, expressed in terms of the local energy function to establish strong consistency and asymptotic normality results of this estimator are presented. While no general condition on the model is assumed to obtain the consistency, the characteristic finite range of the local energy function is required to establish the asymptotic normality. For the sake of simplicity, Section 3 (and the resulting proofs) would concentrate on non-marked Gibbs point processes. However, as we have shown in our paper Billiot et al. (2008), no real mathematical difficulty occurs with the introduction of marks. At end the end of Section 3, we apply the results to Lennard-Jones models. Proofs have been postponed until Appendices A (for the general results) and B (for the verifications of the different assumptions for Lennard-Jones models).

2. Background and notation

For the sake of simplicity, we consider Gibbs point processes in dimension $d = 2$.

2.1. General notation, configuration space

Subregions of $\mathbb{R}^2$ will typically be denoted by Λ or Δ and will always be assumed to be Borel with positive Lebesgue measure. We write $\Lambda \Subset \mathbb{R}^2$ if Λ is bounded. Λ^c denotes the complementary set of Λ inside $\mathbb{R}^2$. The notation $|.|$ will be used without ambiguity for different kind of objects. For a countable set $\mathcal{J}$, $|\mathcal{J}|$ represents the number of elements belonging to $\mathcal{J}$; For $\Lambda \Subset \mathbb{R}^2$, $|\Lambda|$ is the volume of Λ; For a vector $x \in \mathbb{R}^2$, $|x|$ corresponds to its uniform norm while $\|x\|$ is simply its euclidean norm. For all $x \in \mathbb{R}^2, \rho > 0$ and $i \in \mathbb{Z}^2$, let $\mathcal{B}(x, \rho) := \{y \in \mathbb{R}^2,\ |y - x| < \rho\}$ and $\mathbb{B}(i, \rho) := \mathcal{B}(i, \rho) \cap \mathbb{Z}^2$.

A configuration is a subset φ of $\mathbb{R}^2$ which is locally finite in that $\varphi_\Lambda := \varphi \cap \Lambda$ has finite cardinality $N_\Lambda(\varphi) := |\varphi_\Lambda|$ for all $\Lambda \Subset \mathbb{R}^2$. The space Ω of all configurations is equipped with the σ-algebra $\mathcal{F}$ that is generated by the counting variables $N_\Lambda(\varphi)$ with $\Lambda \Subset \mathbb{R}^2$. Finally, let $T = (\tau_x)_{x \in \mathbb{R}^2}$ be the shift group, where $\tau_x : \Omega \to \Omega$ is the translation by the vector $-x \in \mathbb{R}^2$.

2.2. Gibbs point processes

Our results will be expressed for general stationary Gibbs point processes. Since we are interested in asymptotic properties, we have to consider these point processes acting on the infinite volume $\mathbb{R}^2$. Let us briefly recall their definition.

A point process Φ is a Ω-valued random variable, with probability distribution P on $(\Omega, \mathcal{F})$. The most prominent point process is the (homogeneous) Poisson process with intensity $z > 0$. Recall that its probability measure π^z is the unique

probability measure on $(\Omega, \mathcal{F})$ such that the following holds for $\Lambda \Subset \mathbb{R}^2$: (i) N_Λ is Poisson distributed with parameter $z|\Lambda|$, and (ii) conditionally to $N_\Lambda = n$, the n points in Λ are independent with uniform distribution on Λ, for each interger $n \geq 1$. For $\Lambda \Subset \mathbb{R}^2$, let us denote by π_Λ^z the marginal probability measure in Λ of the Poisson process with intensity z.

Let $\theta \in \mathbb{R}^p$ (for some $p \geq 1$). For any $\Lambda \Subset \mathbb{R}^2$, let us consider the parametric function $V_\Lambda(.;\theta)$ from Ω into $\mathbb{R} \cup \{+\infty\}$. From a physical point of view, $V_\Lambda(\varphi;\theta)$ is the energy of φ_Λ in Λ given the outside configuration φ_{Λ^c}.

In this article, we focus on stationary point processes on $\mathbb{R}^2$, i.e. with T-invariant probability measure. For any $\Lambda \Subset \mathbb{R}^2$, we therefore consider $V_\Lambda(.;\theta))$ to be T-invariant, i.e. $V_\Lambda(\tau_x\varphi;\theta) = V_\Lambda(\varphi;\theta)$ for any $x \in \mathbb{R}^2$. Furthermore, we assume that the family of energies is hereditary, which means that for any $\Lambda \Subset \mathbb{R}^2$, $\varphi \in \Omega$, and $x \in \Lambda$: $V_\Lambda(\varphi;\theta)) = +\infty \Rightarrow V_\Lambda(\varphi \cup \{x\};\theta)) = +\infty$.

In such a context, a Gibbs measure is usually defined as follows (see Preston (1976)).

Definition 1. *A probability measure P_θ on Ω is a Gibbs measure for the family of energies $(V_\Lambda(.;\theta))_{\Lambda \Subset \mathbb{R}^2}$ if for every $\Lambda \Subset \mathbb{R}^2$, for P_θ-almost every outside configuration φ_{Λ^c}, the law of P_θ given φ_{Λ^c} admits the following density with respect to π_Λ^z:*

$$f_\Lambda(\varphi_\Lambda|\varphi_{\Lambda^c};\theta) = \frac{1}{Z_\Lambda(\varphi_{\Lambda^c};\theta)} e^{-V_\Lambda(\varphi;\theta)},$$

where $Z_\Lambda(\varphi_{\Lambda^c};\theta) := \int_{\Omega_\Lambda} e^{-V_\Lambda(\varphi_\Lambda \cup \varphi_{\Lambda^c};\theta)} \pi_\Lambda^z(d\varphi_\Lambda)$ is called the partition function.

Without loss of generality, the intensity of the Poisson process, z is fixed to 1 and we simply write π and π_Λ in place of π^1 and π_Λ^1. In the previous definition, we implicitly assume the consistency of the family $(f_\Lambda(.|.;\theta))_{\Lambda \Subset \mathbb{R}^2}$: for any $\Delta \subset \Lambda \Subset \mathbb{R}^2$

$$f_\Delta(\varphi_\Delta|\varphi_{\Delta^c};\theta) = \frac{f_\Lambda(\varphi_\Delta \cup \varphi_{\Lambda\setminus\Delta}|\varphi_{\Lambda^c};\theta)}{f_\Lambda(\varphi_{\Lambda\setminus\Delta}|\varphi_{\Lambda^c};\theta)} = \frac{f_\Lambda(\varphi_\Delta \cup \varphi_{\Lambda\setminus\Delta}|\varphi_{\Lambda^c};\theta)}{\int_{\Omega_\Delta} f_\Lambda(\psi_\Delta \cup \varphi_{\Lambda\setminus\Delta}|\varphi_{\Lambda^c};\theta)\pi_\Delta(d\psi_\Delta)}.$$

A sufficient condition to directly fulfill this basic ingredient is to assume the compatibility of the family $(V_\Lambda(.))_{\Lambda \Subset \mathbb{R}^2}$: for every $\Delta \subset \Lambda \Subset \mathbb{R}^2$, the function $\varphi \to V_\Lambda(\varphi;\theta) - V_\Delta(\varphi;\theta)$ from Ω into $\mathbb{R} \cup \{+\infty\}$ is measurable and only depends on φ_{Λ^c}.

The existence of a Gibbs measure on Ω which satisfies these conditional specifications is a difficult issue. We refer the interested reader to Ruelle (1969); Preston (1976); Bertin et al. (1999a); Dereudre (2005); Dereudre et al. (2010) for the technical and mathematical development of the existence problem. The minimal assumption of our paper is then:

> **[Mod-E]**: Our data consist in the realization of a point process Φ with Gibbs measure $P_{\theta^\star}$, where $\theta^\star \in \mathring{\boldsymbol{\Theta}}$, $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^p$ and, for any $\theta \in \boldsymbol{\Theta}$, there exists a stationary Gibbs measure P_θ for the family $(V_\Lambda(.;\theta))_{\Lambda \Subset \mathbb{R}^2}$.

In the rest of this paper, the reader has mainly to keep in mind the concept of local energy defined as the energy required to insert a point x into the configuration φ and expressed for any $\Lambda \ni x$ by

$$V\left(x|\varphi;\theta\right) := V_\Lambda(\varphi \cup \{x\}) - V_\Lambda(\varphi).$$

From the compatibility of the family of energies, the local energy does not depend on Λ.

Our asymptotic normality result will require the following locality property assumption.

[**Mod-L**]: There exists $D \geq 0$ such that for all $\varphi \in \Omega$

$$V\left(0|\varphi;\theta\right) = V\left(0|\varphi_{\mathcal{B}(0,D)};\theta\right).$$

2.3. Example: Lennard-Jones models

Let us present the main example studied in this paper. We call LJ-type model the stationary pairwise interaction point process defined for some $D \in]0,+\infty]$ by

$$V_\Lambda^{LJ}\left(\varphi;\theta\right) := \theta_1|\varphi_\Lambda| + H_\Lambda^{LJ}\left(\varphi;\theta\right) \text{ with } H_\Lambda^{LJ}\left(\varphi;\theta\right) := \sum_{\substack{x_1\in\varphi_\Lambda \\ x_2\in\varphi_{\Lambda^c}}} g^{LJ}(||x_1-x_2||;\theta)$$

and

$$g^{LJ}(r;\theta) := 4\theta_2\left(\left(\frac{\theta_3}{r}\right)^{12} - \left(\frac{\theta_3}{r}\right)^{6}\right)\mathbf{1}_{[0,D]}(r).$$

As a direct consequence, the local energy function is expressed as

$$V^{LJ}\left(x|\varphi;\theta\right) := \theta_1 + H^{LJ}\left(x|\varphi;\theta\right) \text{ with } H^{LJ}\left(x|\varphi;\theta\right) := \sum_{y\in\varphi} g^{LJ}(||x-y||;\theta).$$

where $\theta = (\theta_1,\theta_2,\theta_3) \in \mathbb{R}\times(\mathbb{R}^+)^2$. The cases $D=+\infty$ and $D<+\infty$ respectively correpond to the Lennard-Jones model (briefly presented in the introduction) and the Lennard-Jones model with finite range.

Ruelle (1970) has proved the existence of an ergodic measure for superstable, regular and lower regular potentials. The Lennard-Jones model (including the finite range one) is known to be the characteristic example of such a family of models for which Ruelle managed to prove the existence of ergodic measures for any $\theta \in \mathbb{R}\times(\mathbb{R}^+)^2$. In order to ensure [**Mod-E**], it is required to assume that $\theta_2^\star, \theta_3^\star > 0$. Finally, [**Mod-L**] is satisfied for the LJ-type model with $D<+\infty$ since the parameter D corresponds for pairwise interaction point processes to the range of the Gibbs point process.

3. Asymptotic results of the maximum pseudo-likelihood estimator

3.1. Maximum pseudo-likelihood method

The idea of maximum pseudo-likelihood is due to Besag (1974) who first introduced the concept for Markov random fields in order to avoid the normalizing constant. This work was then widely extended and Jensen and Møller (1991) (Theorem 2.2) obtained a general expression for Gibbs point processes. Using our notation and up to a scalar factor, the pseudo-likelihood defined for a configuration φ and a domain of observation Λ is denoted by $PL_\Lambda(\varphi;\theta)$ and given by

$$PL_\Lambda(\varphi;\theta) = \exp\left(-\int_\Lambda e^{-V(x|\varphi;\theta)}dx\right)\prod_{x\in\varphi_\Lambda} e^{-V(x|\varphi\setminus x;\theta)}. \tag{1}$$

It is more convenient to define and work with the log-pseudo-likelihood, denoted by $LPL_\Lambda(\varphi;\theta)$

$$LPL_\Lambda(\varphi;\theta) = -\int_\Lambda e^{-V(x|\varphi;\theta)}dx - \sum_{x\in\varphi_\Lambda} V(x|\varphi\setminus x;\theta). \tag{2}$$

The point process is assumed to be observed in a domain $\Lambda_n \oplus \widetilde{D} = \cup_{x\in\Lambda_n}\mathcal{B}(x,\widetilde{D})$ for some $\widetilde{D} < +\infty$. For the asymptotic normality result, it is also assumed that $\widetilde{D} \geq D$ and that $\Lambda_n \subset \mathbb{R}^2$ can be decomposed into $\cup_{i\in I_n}\Delta_i$ where $I_n = \mathbb{B}(0,n)$ and for $i\in\mathbb{Z}^2$, $\Delta_i = \Delta_i(\widetilde{D})$ is the square centered at i with side-length $\widetilde{D}$. As a consequence, as $n\to+\infty$, $\Lambda_n\to\mathbb{R}^2$ such that $|\Lambda_n|\to+\infty$ and $\frac{|\partial\Lambda_n|}{|\Lambda_n|}\to 0$.

Define for any configuration φ, $U_n(\varphi;\theta) = -\frac{1}{|\Lambda_n|}LPL_{\Lambda_n}(\varphi;\theta)$. The maximum pseudo-likelihood estimate (MPLE), denoted by $\widehat{\theta}_n(\varphi)$, is then defined by

$$\widehat{\theta}_n(\varphi) = \arg\max_{\theta\in\boldsymbol{\Theta}} LPL_{\Lambda_n}(\varphi;\theta) = \arg\min_{\theta\in\boldsymbol{\Theta}} U_n(\varphi;\theta).$$

The following basic notation are introduced: for $j,k=1,\ldots,p$ and $\Lambda \Subset \mathbb{R}^2$

- Gradient vector of U_n: $\mathbf{U}_n^{(1)}(\varphi;\theta) := -|\Lambda_n|^{-1}\mathbf{LPL}_{\Lambda_n}^{(1)}(\varphi;\theta)$ where

$$\left(\mathbf{LPL}_\Lambda^{(1)}(\varphi;\theta)\right)_j = \int_\Lambda \frac{\partial V}{\partial\theta_j}(x|\varphi;\theta)\, e^{-V(x|\varphi;\theta)}dx - \sum_{x\in\varphi_\Lambda}\frac{\partial V}{\partial\theta_j}(x|\varphi\setminus x;\theta).$$

- Hessian matrix of U_n: $\underline{\mathbf{U}}_n^{(2)}(\varphi;\theta) := -|\Lambda_n|^{-1}\underline{\mathbf{LPL}}_{\Lambda_n}^{(2)}(\varphi;\theta)$

$$\left(\underline{\mathbf{LPL}}_\Lambda^{(2)}(\varphi;\theta)\right)_{j,k} =$$

$$\int_\Lambda\left(\frac{\partial^2 V}{\partial\theta_j\partial\theta_k}(x|\varphi;\theta) - \frac{\partial V}{\partial\theta_j}(x|\varphi;\theta)\frac{\partial V}{\partial\theta_k}(x|\varphi;\theta)\right)e^{-V(x|\varphi;\theta)}dx$$
$$+\sum_{x\in\varphi_\Lambda}\frac{\partial V}{\partial\theta_j}(x|\varphi\setminus x;\theta)\frac{\partial V}{\partial\theta_k}(x|\varphi\setminus x;\theta).$$

Finally, note that from the decomposition of the observation domain Λ_n, one has

$$\mathbf{U}_n^{(1)}(\varphi;\theta) = -|\Lambda_n|^{-1} \sum_{i\in I_n} \mathbf{LPL}_{\Delta_i}^{(1)}(\varphi;\theta)$$

and

$$\underline{\mathbf{U}}_n^{(2)}(\varphi;\theta) = -|\Lambda_n|^{-1} \sum_{i\in I_n} \underline{\mathbf{LPL}}_{\Delta_i}^{(2)}(\varphi;\theta)\,.$$

3.2. Consistency of the MPLE

The assumption **[C]** gathers the following four assumptions:

[C1] For all $\theta \in \boldsymbol{\Theta}$,

$$\mathbf{E}\left(e^{-V(0|\Phi;\theta)}\right) < +\infty \quad \text{and} \quad \mathbf{E}\left(|V(0|\Phi;\theta)|\, e^{-V(0|\Phi;\theta^\star)}\right) < +\infty.$$

[C2] Identifiability condition : there exists $A_1,\ldots,A_\ell$, $\ell \geq p$ events in Ω such that:

- the ℓ events A_i are disjoint and satisfy $P_{\theta^\star}(B_i) > 0$
- for all $(\varphi_1,\ldots,\varphi_\ell) \in A_1 \times \cdots \times A_\ell$

$$\left\{ \begin{array}{l} D(0|\varphi_i;\theta) = 0 \\ i = 1\ldots,\ell \end{array} \right. \quad \Rightarrow \quad \theta = \theta^\star$$

where $D(0|\varphi_i;\theta) := V(0|\varphi_i;\theta) - V(0|\varphi_i;\theta^\star)$

[C3] The function $U_n(\varphi;\cdot)$ is continuous for $P_{\theta^\star}-$a.e. φ.

[C4] For all $\varphi \in \Omega$, $V(0|\varphi;\theta)$ is continuously differentiable in θ and for all $j = 1,\ldots,p$

$$\mathbf{E}\left(\max_{\theta\in\boldsymbol{\Theta}}\left(\left|\frac{\partial V}{\partial\theta_j}(0|\Phi;\theta)\right| e^{-V(0|\Phi;\theta)}\right)^2\right) \quad < \quad +\infty.$$

Theorem 1. *Under the assumptions **[Mod-E]** and **[C]**, for $P_{\theta^\star}-$almost every φ, the maximum pseudo-likelihood estimate $\widehat{\theta}_n(\varphi)$ converges towards $\theta^\star$ as n tends to infinity.*

3.3. Asymptotic normality of the MPLE

For establishing the asymptotic normality of the MPLE we need to assume the four additional following assumptions:

[N1] For all $\varphi \in \Omega$, $V(0|\varphi;\theta)$ is differentiable in $\theta = \theta^\star$. For all $k = 1,\ldots,3$ and for all $\lambda_1,\ldots,\lambda_k$, k positive integers such that $\sum_{i=1}^k \lambda_i = 3$ and for $\Delta \Subset \mathbb{R}^2$

$$\mathbf{E}\left(\int_{\Delta^k}\prod_{i=1}^k\left|\frac{\partial V}{\partial\theta_j}(0^M|\Phi;\theta^\star)\right|^{\lambda_i} e^{-V(\{x_1,\ldots,x_k\}|\Phi;\theta^\star)}dx_1\ldots dx_k\right) < +\infty.$$

[N2] There exists a neighbourhood $\mathcal{V}(\theta^\star)$ of $\theta^\star$ such that for all $\varphi \in \Omega$, $V\left(0|\varphi;\theta\right)$ is twice continuously differentiable in $\theta \in \mathcal{V}$ and, for all $j,k=1,\ldots,p$ and $\theta \in \mathcal{V}(\theta^\star)$,

$$\mathbf{E}\left(\left|\frac{\partial^2 V}{\partial\theta_j\partial\theta_k}\left(0|\Phi;\theta\right)\right| e^{-V\left(0|\Phi;\theta'\right)}\right) < +\infty, \text{ for } \theta'=\theta,\theta^\star$$

and

$$\mathbf{E}\left(\left(\left|\frac{\partial V}{\partial\theta_j}\left(0|\Phi;\theta\right)\right| e^{-V\left(0|\Phi;\theta\right)}\right)^2\right) < +\infty.$$

[N3] There exists $A_1,\ldots,A_\ell$, $\ell \geq p$ events in Ω such that:

- the ℓ events A_i are disjoint and satisfy $P_{\theta^\star}(A_i)>0$
- for all $(\varphi_1,\ldots,\varphi_\ell) \in A_1\times\cdots\times A_\ell$ the (ℓ,p) matrix with entries $\frac{\partial V}{\partial\theta_j}\left(0|\varphi_i;\theta^\star\right)$ is injective.

[N4] There exists $A_0,\ldots,A_\ell$, $\ell \geq p$ disjoint sub-events of $\overline{\Omega} := \{\varphi\in\Omega : \varphi_{\Delta_i}=\emptyset, 1\leq|i|\leq 2\}$ such that

- for $j=0,\ldots,\ell$, $P_{\theta^\star}(A_j)>0$.
- for all $(\varphi_0,\ldots,\varphi_\ell) \in A_0\times\cdots\times A_\ell$ the (ℓ,p) matrix with entries $\left(\mathbf{LPL}^{(1)}_{\overline{\Lambda}}\left(\varphi_i;\theta^\star\right)\right)_j - \left(\mathbf{LPL}^{(1)}_{\overline{\Lambda}}\left(\varphi_0;\theta^\star\right)\right)_j$ is injective, with $\overline{\Lambda}:=\cup_{i\in\mathbb{B}(0,1)}$.

The assumptions **[N3]** and **[N4]** will ensure (see Section A for more details) that the matrices $\underline{\mathbf{U}}^{(2)}(\theta^\star)$ and $\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star)$ respectively defined by

$$\left(\underline{\mathbf{U}}^{(2)}(\theta^\star)\right)_{j,k} := \mathbf{E}\left(\frac{\partial V}{\partial\theta_j}\left(0|\Phi;\theta^\star\right)\frac{\partial V}{\partial\theta_k}\left(0|\Phi;\theta^\star\right)e^{-V\left(0|\Phi;\theta^\star\right)}\right) \tag{3}$$

and

$$\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star) = \widetilde{D}^{-2}\sum_{i\in\mathbb{B}(0,1)}\mathbf{E}\left(\mathbf{LPL}^{(1)}_{\Delta_0}\left(\Phi;\theta^\star\right)\mathbf{LPL}^{(1)}_{\Delta_i}\left(\Phi;\theta^\star\right)^T\right), \tag{4}$$

are definite positive.

Observe that, when the energy function is linear, the expressions of the assumptions **[N1]** and **[N2]** are clearly simpler (see Billiot et al. (2008)) and that **[C2]** and **[N3]** are similar.

Theorem 2. *Under the assumptions **[Mod]**, **[C]**, **[N1]**, **[N2]** and **[N3]**, we have the following convergence in distribution as $n\to+\infty$*

$$|\Lambda_n|^{1/2}\,\underline{\mathbf{U}}^{(2)}(\theta^\star)\,\left(\widehat{\theta}_n(\Phi)-\theta^\star\right) \to \mathcal{N}\left(0,\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star)\right), \tag{5}$$

*where $\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star)$ is defined by (4). In addition under the assumption **[N4]***

$$|\Lambda_n|^{1/2}\,\widehat{\underline{\mathbf{\Sigma}}}_n(\Phi;\widehat{\theta}_n(\Phi))^{-1/2}\,\underline{\mathbf{U}}^{(2)}_n(\Phi;\widehat{\theta}_n(\Phi))\,\left(\widehat{\theta}_n(\Phi)-\theta^\star\right) \to \mathcal{N}\left(0,\underline{\mathbf{I}}_p\right), \tag{6}$$

where for some θ and any configuration φ, the matrix $\widehat{\underline{\mathbf{\Sigma}}}_n(\varphi;\theta)$ is defined by

$$\widehat{\underline{\mathbf{\Sigma}}}_n(\varphi;\theta) = |\Lambda_n|^{-1}\sum_{i\in I_n}\sum_{j\in\mathbb{B}(i,1)\cap I_n}\mathbf{LPL}^{(1)}_{\Delta_i}\left(\varphi;\theta\right)\mathbf{LPL}^{(1)}_{\Delta_j}\left(\varphi;\theta\right)^T. \tag{7}$$

In the following the assumption **[N]** will stand for the assumptions **[N1]**, **[N2]**, **[N3]** and **[N4]**.

3.4. Applications to Lennard-Jones models

The following proposition holds for the LJ-type model presented in Section 2.3.

Proposition 3.
(*i*) *Theorem 1 holds for the LJ-type model (with* $D \in]0, +\infty]$*), that is for the Lennard-Jones and the finite-range Lennard-Jones model.*
(*ii*) *Theorem 2 holds only for the finite-range Lennard-Jones model.*

The proof of Proposition 3 consists in verifying Assumptions **[C]** for the LJ-type model and **[N]** only for the finite range Lennard-Jones model. Two types of assumptions are distinguished:

- Integrabilility type assumptions, i.e. Assumptions **[C1]**, **[C4]**, **[N1]** and **[N2]**.
- Identifiability type assumptions, i.e. Assumptions **[C2]**, **[N3]** and **[N4]**.

Note that **[C3]** is obvious since $g^{LJ}(r, \cdot)$ is continuous. The proofs are somewhat technical and are postponed until Section B. For the integrability type assumptions, the following Lemma constitutes the main ingredient.

Lemma 4. *Let* Φ *be a stationary pairwise interaction Gibbs point process assumed to be superstable, regular and lower regular. For* $i = 1, 2$*, define* $H_i(x|\varphi) = \sum_{y \in \varphi} g_i(||x-y||)$ *with* g_i *a continuous function. Assume that there exists* $\varepsilon > 0$ *such that there exists a positive and decreasing function* $g(\cdot)$ *such that* $g_\varepsilon(r) := g_2(r) - \varepsilon|g_1(r)| \geq -g(r)$ *for all* $r > 0$ *and* $\int_0^{+\infty} rg(r)dr < +\infty$*. Then for all* $k \geq 0$*,*

$$\mathbf{E}\left(|H_1(0|\Phi)|^k e^{-H_2(0|\Phi)}\right) < +\infty.$$

Proof. For all finite configuration φ

$$\begin{aligned} |H_1(0|\varphi)|^k e^{-H_2(0|\varphi)} &= |H_1(0|\varphi)|^k \; e^{-\varepsilon|H_1(0|\varphi)|} \, e^{-(H_2(0|\varphi)-\varepsilon H_1(0|\varphi))} \\ &\leq c(\varepsilon,k) e^{-(H_2(0|\varphi)-\varepsilon H_1(0|\varphi))}, \quad \text{with } c(\varepsilon,k) = \left(\frac{k}{\varepsilon e}\right)^k \\ &\leq c(\varepsilon,k) e^{-H_\varepsilon(0|\varphi)}, \end{aligned}$$

where

$$H_\varepsilon(0|\varphi) := \sum_{x \in \varphi} g_\varepsilon(||x||).$$

Now, the assumptions ensure that g_ε is lower regular in the Ruelle sense. We may now apply the same argument as in Lemma 3 of Mase (1995) to prove the integrability of the random variable $e^{-H_\varepsilon(0|\Phi)}$. □

Appendix A: Proofs of Theorems 1 and 2

Let us start by presenting a particular case of the Campbell Theorem combined with the Glötz Theorem that is widely used in our future proofs.

Corollary 5. *Assume that the point process Φ with probability measure P is stationary. Let $\Lambda \Subset \mathbb{R}^2$, $\varphi \in \Omega$ and let g be a function satisfying $g(x,\varphi) = g(0,\tau_x\varphi)$ for all $x \in \mathbb{R}^2$. Define $f(\varphi) = g(0,\varphi)e^{-V(0|\varphi)}$ and assume that $f \in L^1(P)$. Then,*

$$\mathbf{E}\left(\sum_{x\in\Phi_\Lambda} g(x,\Phi\setminus x)\right) = \mathbf{E}\left(\int_\Lambda g(x,\Phi)e^{-V(x|\Phi)}dx\right) = |\Lambda|\,\mathbf{E}\Big(\,g\,(0,\Phi)\,e^{-V(0|\Phi)}\,\Big) \tag{8}$$

Proof. see Corollary 3 of Billiot et al. (2008) □

Let us now present a version of an ergodic theorem obtained by Nguyen and Zessin (1979) and widely used in this paper. Let Δ_0 be a fixed bounded domain

Theorem 6 (Nguyen and Zessin (1979))**.** *Let $\{H_G, G \in \mathcal{B}_b\}$ be a family of random variables, which is covariant, that is for all $x \in \mathbb{R}^2$,*

$$H_{\tau_x G}(\tau_x\varphi) = H_G(\varphi), \quad \text{for a.e. } \varphi$$

and additive, that is for every disjoint $G_1, G_2 \in \mathcal{B}_b$,

$$H_{G_1\cup G_2} = H_{G_1} + H_{G_2}, \quad a.s.$$

Let $\mathcal{I}$ be the sub-σ-algebra of $\mathcal{F}$ consisting of translation invariant (with probability 1) sets. Assume there exists a nonnegative and integrable random variable Y such that $|H_G| \le Y$ a.s. for every convex $G \subset \Delta_0$. Then,

$$\lim_{n\to+\infty} \frac{1}{|G_n|} H_{G_n} = \frac{1}{|\Delta_0|} E(H_{\Delta_0}|\mathcal{I}), \quad a.s.$$

for each regular sequence $G_n \to \mathbb{R}^2$.

A.1. Proof of Theorem 1

Due to the decomposition of stationary measures as a mixture of ergodic measures (see Preston (1976)), one only needs to prove Theorem 1 by assuming that $P_{\theta^\star}$ is ergodic. From now on, $P_{\theta^\star}$ is assumed to be ergodic. The tool used to obtain the almost sure convergence is a convergence theorem for minimum contrast estimators established by Guyon (1992).

We proceed in three stages.

Step 1. Convergence of $U_n(\Phi;\theta)$.

Decompose $U_n(\varphi;\theta) = \frac{1}{|\Lambda_n|}\left(H_{1,\Lambda_n}(\varphi) + H_{2,\Lambda_n}(\varphi)\right)$ with

$$H_{1,\Lambda_n}(\varphi) = \int_{\Lambda_n} e^{-V(x|\varphi;\theta)}dx \quad \text{and} \quad H_{2,\Lambda_n}(\varphi) = \sum_{x\in\Phi_{\Lambda_n}} V\left(x|\varphi\setminus x;\theta\right).$$

Under the assumption **[C1]**, one can apply Theorem 6 (Nguyen and Zessin (1979)) to the process H_{1,Λ_n}. And from Corollary 5, we obtain $P_{\theta^\star}-$almost surely as $n \to +\infty$

$$\frac{1}{|\Lambda_n|} H_{1,\Lambda_n}(\Phi) \to \mathbf{E}\Big(e^{-V(0|\Phi;\theta)}\Big). \tag{9}$$

Now, let $G \subset \Delta_0$, we clearly have

$$|H_{2,G}(\varphi)| \leq \sum_{x\in\varphi_G} |V\left(x|\varphi\setminus x;\theta\right)| \leq \sum_{x\in\varphi_{\Delta_0}} |V\left(x|\varphi\setminus x;\theta\right)|.$$

Under the assumption **[Mod]** and from Corollary 5, we have

$$\mathbf{E}\left(\sum_{x\in\Phi_{\Delta_0}} |V\left(x|\Phi\setminus x;\theta\right)|\right) = |\Delta_0|\mathbf{E}\left(|V\left(0|\Phi;\theta\right)|e^{-V(0|\Phi;\theta^\star)}\right) < +\infty$$

This means that for all $G \subset \Delta_0$, there exists a random variable $Y \in L^1(P_{\theta^\star})$ such that $|H_{2,G}(\Phi)| \leq Y$. Thus, under the assumption **[C1]** and from Theorem 6 (Nguyen and Zessin (1979)) and from Corollary 5, we have $P_{\theta^\star}-$almost surely

$$\frac{1}{|\Lambda_n|} H_{2,\Lambda_n}(\Phi) \to \frac{1}{|\Delta_0|}\mathbf{E}\Big(\sum_{x\in\Phi_{\Delta_0}} V\left(x|\Phi\setminus x;\theta\right)\Big) = \mathbf{E}\left(V\left(0|\Phi;\theta\right)e^{-V(0|\Phi;\theta^\star)}\right). \tag{10}$$

We have the result by combining (9) and (10): $P_{\theta^\star}-$almost surely

$$U_n(\Phi;\theta) \to U(\theta) = \mathbf{E}\Big(e^{-V(0|\Phi;\theta)} + V\left(0|\Phi;\theta\right)e^{-V(0|\Phi;\theta^\star)}\Big) \tag{11}$$

Step 2. $U_n(\cdot;\theta)$ is a contrast function

Recall that $U_n(\cdot;\theta)$ is a contrast function if there exists a function $K(\cdot,\theta^\star)$ (i.e. nonnegative function equal to zero if and only if $\theta=\theta^\star$) such that $P_{\theta^\star}-$almost surely $U_n(\Phi;\theta) - U_n(\Phi;\theta^\star) \to K(\theta,\theta^\star)$. From Step 1, we have

$$K(\theta,\theta^\star) = \mathbf{E}\Big(e^{-V(0|\Phi;\theta^\star)}\Big(e^{V(0|\Phi;\theta)-V(0|\Phi;\theta^\star)} - \Big(1+V\left(0|\Phi;\theta\right)-V\left(0|\Phi;\theta^\star\right)\Big)\Big)\Big). \tag{12}$$

Since the function $t \mapsto e^t - (1+t)$ is nonnegative and is equal to zero if and only if $t=0$, $K(\theta,\theta^\star) \geq 0$ and

$$\begin{aligned} K(\theta,\theta^\star) = 0 \quad &\Leftrightarrow \quad e^{V(0|\varphi;\theta)-V(0|\varphi;\theta^\star)} - \Big(1+V\left(0|\varphi;\theta\right)-V\left(0|\varphi;\theta^\star\right)\Big) = 0 \\ &\Leftrightarrow \quad D\left(0|\varphi;\theta\right) := V\left(0|\varphi;\theta\right) - V\left(0|\varphi;\theta^\star\right) = 0 \end{aligned}$$

for $P_{\theta^\star}$ – a.e. φ. Let us consider the ℓ events A_j $(j=1,\ldots,\ell)$ defined in Assumption **[C2]**. The previous equation is at least true for $\varphi_j \in A_j$, which leads under Assumption **[C2]** to $\theta=\theta^\star$. Therefore, $K(\theta,\theta^\star)=0 \Rightarrow \theta=\theta^\star$. The converse is trivial.

Before ending this step, note that the assumption **[C3]** asserts that for any φ, $U_n(\varphi;\cdot)$ and $K(\cdot,\theta^\star)$ are continuous functions.

Step 3. Modulus of continuity.

The modulus of continuity of the contrast process defined for all $\varphi \in \Omega$ and all $\eta > 0$ by

$$W_n(\varphi, \eta) = \sup\left\{\left|U_n(\varphi;\theta) - U_n(\varphi;\theta')\right| : \theta, \theta' \in \boldsymbol{\Theta}, ||\theta - \theta'|| \leq \eta\right\}$$

is such that there exists a sequence $(\varepsilon_k)_{k\geq 1}$, with $\varepsilon_k \to 0$ as $k \to +\infty$ such that for all $k \geq 1$

$$P\left(\limsup_{n\to+\infty}\left(W_n\left(\Phi, \frac{1}{k}\right) \geq \varepsilon_k\right)\right) = 0. \tag{13}$$

Let us start to write $W_n\left(\varphi, \frac{1}{k}\right) \leq W_{1,n}\left(\varphi, \frac{1}{k}\right) + W_{2,n}\left(\varphi, \frac{1}{k}\right)$ with

$$\begin{aligned} W_{1,n}\left(\varphi, \frac{1}{k}\right) &:= \sup\left\{W'_{1,\Lambda_n}(\varphi;\theta,\theta') : \theta, \theta' \in \boldsymbol{\Theta}, ||\theta - \theta'|| \leq \frac{1}{k}\right\} \\ W_{2,n}\left(\varphi, \frac{1}{k}\right) &:= \sup\left\{W'_{2,\Lambda_n}(\varphi;\theta,\theta') : \theta, \theta' \in \boldsymbol{\Theta}, ||\theta - \theta'|| \leq \frac{1}{k}\right\}. \end{aligned}$$

and

$$\begin{aligned} W'_{1,\Lambda_n}(\varphi;\theta,\theta') &:= \frac{1}{|\Lambda_n|}\int_{\Lambda_n}\left|e^{-V(x|\varphi;\theta)} - e^{-V\left(x|\varphi;\theta'\right)}\right|dx \\ W'_{2,\Lambda_n}(\varphi;\theta,\theta') &:= \frac{1}{|\Lambda_n|}\sum_{x\in\varphi_{\Lambda_n}}\left|V\left(x|\varphi\setminus x;\theta\right) - V\left(x|\varphi\setminus x;\theta'\right)\right|. \end{aligned}$$

Let $k \geq 1$ and let $\theta, \theta' \in \boldsymbol{\Theta}$ such that $||\theta - \theta'|| \leq \frac{1}{k}$, then under the assumption **[C1]** and from Theorem 6 and Corollary 5, we have $P_{\theta^\star}-$almost surely as $n \to +\infty$

$$\begin{aligned} W'_{1,\Lambda_n}(\Phi;\theta,\theta') &\longrightarrow \mathbf{E}\left(\left|e^{-V(0|\Phi;\theta)} - e^{-V\left(0|\Phi;\theta'\right)}\right|\right) \\ W'_{2,\Lambda_n}(\Phi;\theta,\theta') &\longrightarrow \mathbf{E}\left(\left|V\left(0|\Phi;\theta\right) - V\left(0|\Phi;\theta'\right)\right| e^{-V(0|\Phi;\theta^\star)}\right) \end{aligned}$$

Under Assumption **[C4]**, one may apply the mean value theorem in $\mathbb{R}^p$ as follows: there exist $\xi^{(1)}, \ldots, \xi^{(p)} \in \prod_{j=1}^p\left[\min(\theta_j, \theta'_j), \max(\theta_j, \theta'_j)\right]$ such that for all $\varphi \in \Omega$

$$e^{-V(0|\varphi;\theta)} - e^{-V\left(0|\varphi;\theta'\right)} = \sum_{j=1}^p\left(\theta_j - \theta'_j\right)\frac{\partial V}{\partial\theta_j}\left(0|\varphi;\xi^{(j)}\right)e^{-V\left(0|\varphi;\xi^{(j)}\right)}.$$

This leads, under Assumption **[C4]**, to the following inequality

$$\begin{aligned} &\mathbf{E}\left(\left|e^{-V(0|\Phi;\theta)} - e^{-V\left(0|\Phi;\theta'\right)}\right|\right)^2 \\ &\quad\leq \mathbf{E}\left(\left|e^{-V(0|\Phi;\theta)} - e^{-V\left(0|\Phi;\theta'\right)}\right|^2\right) \end{aligned}$$

$$\leq \mathbf{E}\left(||\theta-\theta'||^2\sum_{j=1}^{p}\left|\frac{\partial V}{\partial\theta_j}\left(0|\Phi;\xi^{(j)}\right)e^{-V\left(0|\Phi;\xi^{(j)}\right)}\right|^2.\right)$$
$$\leq \left(\frac{1}{k}\right)^2\gamma_1^2,$$

with $\gamma_1 := \mathbf{E}\left(\sum_{j=1}^{p}\max_{\theta\in\mathbf{\Theta}}\left|\frac{\partial V}{\partial\theta_j}(0|\Phi;\theta)\,e^{-V(0|\Phi;\theta)}\right|^2\right) < +\infty$. In such a way, one may also prove that

$$\mathbf{E}\left(\left|V(0|\Phi;\theta)-V(0|\Phi;\theta')\right|e^{-V(0|\Phi;\theta^\star)}\right)^2 \leq \left(\frac{1}{k}\right)^2\gamma_2^2,$$

with $\gamma_2 := \mathbf{E}\left(\sum_{j=1}^{p}\max_{\theta\in\mathbf{\Theta}}\left|\frac{\partial V}{\partial\theta_j}(0|\Phi;\theta)\,e^{-V(0|\Phi;\theta^\star)}\right|^2\right)$. Hence, for all $k\geq 1$ and for all $\theta,\theta'\in\mathbf{\Theta}$ such that $||\theta-\theta'||\leq\frac{1}{k}$ there exists $n_0(k)\geq 1$ such that for all $n\geq n_0(k)$, we have

$$W'_{1,\Lambda_n}(\varphi;\theta,\theta')\leq\frac{2}{k}\gamma_1 \qquad \text{and} \qquad W'_{2,\Lambda_n}(\varphi;\theta,\theta')\leq\frac{2}{k}\gamma_2,\ \text{for } P_{\theta^\star}-\text{a.e. } \varphi.$$

Since γ_1 and γ_2 are independent of θ and θ', we have for all $n\geq n_0(k)$

$$W_n\left(\varphi,\frac{1}{k}\right)\leq W_{1,n}\left(\varphi,\frac{1}{k}\right)+W_{2,n}\left(\varphi,\frac{1}{k}\right)\leq\frac{2}{k}(\gamma_1+\gamma_2):=\frac{c}{k},\ \text{for } P_{\theta^\star}-\text{a.e. } \varphi.$$

Finally, since

$$\begin{aligned}\limsup_{n\to+\infty}\left\{W_n\left(\varphi,\frac{1}{k}\right)\geq\frac{c}{k}\right\} &= \bigcap_{m\in\mathbb{N}}\bigcup_{n\geq m}\left\{W_n\left(\varphi,\frac{1}{k}\right)\geq\frac{c}{k}\right\}\\ &\subset \bigcup_{n\geq n_0(k)}\left\{W_n\left(\varphi,\frac{1}{k}\right)\geq\frac{c}{k}\right\}\end{aligned}$$

for $P_{\theta^\star}-$a.e. φ, the expected result (13) is proved.

Conclusion step. The Steps 1, 2 and 3 ensure the fact that we can apply Property 3.6 of Guyon (1992) which asserts the almost sure convergence for minimum contrast estimators.

A.2. Proof of Theorem 2

Step 1. Asymptotic normality of $\mathbf{U}_n^{(1)}(\Phi;\theta^\star)$

The aim is to prove the following convergence in distribution as $n\to+\infty$

$$|\Lambda_n|^{1/2}\,\mathbf{U}_n^{(1)}(\Phi;\theta^\star)\to\mathcal{N}\left(0,\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star)\right) \tag{14}$$

where the matrix $\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star)$ is defined by (4).

The idea is to apply to $\mathbf{U}_n^{(1)}(\Phi;\theta^\star)$ a central limit theorem obtained by Jensen and Künsch (1994), Theorem 2.1. The following conditions have to be fulfilled to apply this result. For all $j=1,\ldots,p$

(*i*) For all $i \in \mathbb{Z}^2$, $\mathbf{E}\left(\left(\mathbf{LPL}^{(1)}_{\Delta_i}(\Phi;\theta^\star)\right)_j|\Phi_{\Delta_i^c}\right) = 0$.

(*ii*) For all $i \in \mathbb{Z}^2$, $\mathbf{E}\left(\left|\left(\mathbf{LPL}^{(1)}_{\Delta_i}(\Phi;\theta^\star)\right)_j\right|^3\right) < +\infty$.

(*iii*) The matrix $\mathbb{V}\mathrm{ar}\left(|\Lambda_n|^{1/2}\mathbf{U}^{(1)}_n(\Phi;\theta^\star)\right)$ converges to the matrix $\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star)$.

Condition (*i*) : From the stationarity of the process, it is sufficient to prove that

$$\mathbf{E}\left(\left(\mathbf{LPL}^{(1)}_{\Delta_0}(\Phi;\theta^\star)\right)_j|\Phi_{\Delta_0^c}\right) = 0.$$

Recall that for any configuration φ

$$\begin{aligned}\left(\mathbf{LPL}^{(1)}_{\Delta_0}(\varphi;\theta^\star)\right)_j &= -\int_{\Delta_0}\frac{\partial V}{\partial\theta_j}(x|\varphi;\theta^\star)\,e^{-V(x|\varphi;\theta^\star)}dx \\ &\quad + \int_{\Delta_0}\frac{\partial V}{\partial\theta_j}(x|\varphi\setminus x;\theta^\star)\,\varphi(dx). \qquad (15)\end{aligned}$$

Denote respectively by $G_1(\varphi)$ and $G_2(\varphi)$ the first and the second right-hand term of (15) and by $E_i = \mathbf{E}\left(G_i(\Phi)|\Phi_{\Delta_0^c} = \varphi_{\Delta_0^c}\right)$. Let us define for any φ, the measure $\mu_\varphi := \sum_{x\in\varphi}\delta_x$. From the definition of Gibbs point processes,

$$E_2 = \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})}\int_{\Omega_{\Delta_0}}\pi_{\Delta_0}(d\varphi_{\Delta_0})\int_{\mathbb{R}^2}\mu_{\varphi_{\Delta_0}}(dx)\mathbf{1}_{\Delta_0}(x)\frac{\partial V}{\partial\theta_j}(x|\varphi\setminus x;\theta^\star)\,e^{-V_{\Delta_0}(\varphi;\theta^\star)}.$$

Since π is a Poisson process,

$$\int_{\Omega_{\Delta_0}}\pi_{\Delta_0}(d\varphi_{\Delta_0})f(\varphi) = \int_{\Omega_{\Delta_0}}\pi_{\Delta_0}(d\varphi_{\Delta_0})\int_{\Omega_{\Delta_0^c}}\pi_{\Delta_0^c}(d\varphi'_{\Delta_0^c})f(\varphi)$$

and therefore, by introducing $\psi := \varphi_{\Delta_0}\cup\varphi'_{\Delta_0^c}$

$$\begin{aligned}E_2 &= \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})}\int_\Omega\pi(d\psi)\int_{\mathbb{R}^2}\mu_\psi(dx)\mathbf{1}_{\Delta_0}(x)\frac{\partial V}{\partial\theta_j}\left(x|\psi_{\Delta_0}\cup\varphi_{\Delta_0^c}\setminus x;\theta^\star\right)\times \\ &\quad e^{-V_{\Delta_0}\left(\psi_{\Delta_0}\cup\varphi_{\Delta_0^c};\theta^\star\right)}.\end{aligned}$$

Now, from Campbell Theorem (applied to the Poisson measure π)

$$E_2 = \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})}\int_{\Delta_0}dx\int_\Omega\pi^!_x(d\psi)\frac{\partial V}{\partial\theta_j}\left(x|\psi_{\Delta_0}\cup\varphi_{\Delta_0^c};\theta^\star\right)e^{-V_{\Delta_0}\left(x\cup\psi_{\Delta_0}\cup\varphi_{\Delta_0^c};\theta^\star\right)},$$

where $\pi^!_x$ stands for the reduced Palm distribution of the Poisson point process. Since from Slivnyak-Mecke Theorem (see *e.g.* Møller and Waagepetersen (2003)), $\pi = \pi^!_x$, one can obtain

$$E_2 = \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})}\int_\Omega\pi(d\psi)\int_{\Delta_0}dx\,\frac{\partial V}{\partial\theta_j}\left(x|\psi_{\Delta_0}\cup\varphi_{\Delta_0^c};\theta^\star\right)e^{-V_{\Delta_0}\left(x\cup\psi_{\Delta_0}\cup\varphi_{\Delta_0^c};\theta^\star\right)}$$

$$= \frac{1}{Z_{\Delta_0}(\varphi_{\Delta_0^c})}\int_{\Omega_{\Delta_0}}\pi_{\Delta_0}(d\varphi_{\Delta_0})\int_{\Delta_0}dx\frac{\partial V}{\partial\theta_j}(x|\varphi;\theta^\star)\,e^{-V(x|\varphi;\theta^\star)}e^{-V_{\Delta_0}(\varphi;\theta^\star)}$$
$$= -E_1$$

Condition (ii) : For any bounded domain Δ one may write for $j=1,\ldots,p$

$$\left|\left(\mathbf{LPL}_\Delta^{(1)}(\Phi;\theta^\star)\right)_j\right|^3 \leq 4\left|\int_\Delta\frac{\partial V}{\partial\theta_j}(x|\Phi;\theta^\star)\,e^{-V(x|\Phi;\theta^\star)}dx\right|^3 + 4\left|\sum_{x\in\varphi_\Delta}\frac{\partial V}{\partial\theta_j}(x|\Phi\setminus x;\theta^\star)\right|^3.$$

The assumption **[N1]** ensures the integrability of the first right-hand term. For the second one, note that

$$\begin{aligned}T_2 &:= \left|\sum_{x\in\Phi_\Delta}\frac{\partial V}{\partial\theta_j}(x|\varphi\setminus x;\theta^\star)\right|^3\\ &\leq \sum_{\substack{x_1,x_2,x_3\in\varphi_\Delta\\ x_1\neq x_1,x_2\neq x_3,x_2\neq x_3}}\left|\frac{\partial V}{\partial\theta_j}(x_1|\varphi\setminus x_1;\theta^\star)\right|\left|\frac{\partial V}{\partial\theta_j}(x_2|\varphi\setminus x_2;\theta^\star)\right|\left|\frac{\partial V}{\partial\theta_j}(x_3|\varphi\setminus x_3;\theta^\star)\right|\\ &\quad+3\sum_{x_1,x_2\in\varphi_\Delta,x_1\neq x_2}\left|\frac{\partial V}{\partial\theta_j}(x_1|\varphi\setminus x_1;\theta^\star)\right|^2\left|\frac{\partial V}{\partial\theta_j}(x_2|\varphi\setminus x_2;\theta^\star)\right|\\ &\quad+\sum_{x_1\in\varphi_\Delta}\left|\frac{\partial V}{\partial\theta_j}(x_2|\varphi\setminus x_1;\theta^\star)\right|^3.\end{aligned}$$

The result is obtained by using the assumption **[N1]** and iterated versions of Corollary 5.

Condition (iii): let us start by noting that from the assumption **[Mod-L]**, the vector $\mathbf{LPL}_{\Delta_i}^{(1)}(\varphi;\theta^\star)$ depends only on φ_{Δ_j} for $j\in\mathbb{B}(i,1)$. Let $E_{i,j}:=\mathbf{E}\left(\mathbf{LPL}_{\Delta_i}^{(1)}(\Phi;\theta^\star)\,\mathbf{LPL}_{\Delta_j}^{(1)}(\Phi;\theta^\star)^T\right)$. Based on our definitions, we have

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left(|\Lambda_n|^{1/2}\mathbf{U}_n^{(1)}(\Phi;\theta^\star)\right) &= |\Lambda_n|^{-1}\mathbb{V}\mathrm{ar}\left(\sum_{i\in I_n}\mathbf{LPL}_{\Delta_i}^{(1)}(\Phi;\theta^\star)\right)\\ &= |\Lambda_n|^{-1}\sum_{i,j\in I_n}E_{i,j}\\ &= |\Lambda_n|^{-1}\sum_{i\in I_n}\left(\sum_{j\in I_n\cap\mathbb{B}(i,1)}E_{i,j}+\sum_{j\in I_n\cap\mathbb{B}(i,1)^c}E_{i,j}\right).\end{aligned}$$

Let $j\in I_n\cap\mathbb{B}(i,1)^c$, since $\mathbf{LPL}_{\Delta_i}^{(1)}(\varphi;\theta^\star)$ is a measurable function of $\varphi_{\Delta_i^c}$, we have by using condition (i):

$$\begin{aligned}
&\mathbf{E}\left(\mathbf{LPL}_{\Delta_i}^{(1)}\left(\Phi;\theta^\star\right)\mathbf{LPL}_{\Delta_j}^{(1)}\left(\Phi;\theta^\star\right)^T\right)\\
&\quad= \mathbf{E}\left(\mathbf{E}\left(\mathbf{LPL}_{\Delta_i}^{(1)}\left(\Phi;\theta^\star\right)\mathbf{LPL}_{\Delta_j}^{(1)}\left(\Phi;\theta^\star\right)^T|\Phi_{\Delta_i^c}\right)\right)\\
&\quad= \mathbf{E}\left(\mathbf{E}\left(\mathbf{LPL}_{\Delta_i}^{(1)}\left(\Phi;\theta^\star\right)|\Phi_{\Delta_i^c}\right)\mathbf{LPL}_{\Delta_j}^{(1)}\left(\Phi;\theta^\star\right)^T\right)\\
&\quad= 0
\end{aligned}$$

Denote by $\widetilde{I}_n$ the following set

$$\widetilde{I}_n = I_n \cap \left(\cup_{i\in\partial I_n}\mathbb{B}\left(i,1\right)\right).$$

We now obtain

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(|\Lambda_n|^{1/2}\mathbf{U}_n^{(1)}(\Phi;\theta^\star)\right) &= |\Lambda_n|^{-1}\sum_{i\in I_n}\sum_{j\in I_n\cap\mathbb{B}(i,1)}E_{i,j}\\
&= |\Lambda_n|^{-1}\left(\sum_{i\in I_n\setminus\widetilde{I}_n}\sum_{j\in I_n\cap\mathbb{B}(i,1)}E_{i,j}+\sum_{i\in\widetilde{I}_n}\sum_{j\in I_n\cap\mathbb{B}(i,1)}E_{i,j}\right)
\end{aligned}$$

Using the stationarity and the definition of the domain Λ_n, one obtains

$$|\Lambda_n|^{-1}\sum_{i\in I_n\setminus\widetilde{I}_n}\sum_{j\in I_n\cap\mathbb{B}(i,1)}E_{i,j} = |\Lambda_n|^{-1}|I_n\setminus\widetilde{I}_n|\sum_{j\in\mathbb{B}(0,1)}E_{0,j}\to\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star),$$

as $n\to+\infty$, and

$$|\Lambda_n|^{-1}\left|\sum_{i\in\widetilde{I}_n}\sum_{j\in I_n\cap\mathbb{B}\left(i,\lceil\frac{D}{\widetilde{D}}\rceil\right)}E_{i,j}\right| \le |\Lambda_n|^{-1}|\widetilde{I}_n|\sum_{j\in\mathbb{B}(0,1)}|E_{0,j}|\to 0 \quad \text{as } n\to+\infty.$$

Hence as $n\to+\infty$

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(|\Lambda_n|^{1/2}\mathbf{U}_n^{(1)}(\Phi;\theta^\star)\right) &= |\Lambda_n|^{-1}\sum_{i\in I_n}\sum_{j\in I_n\cap\mathbb{B}(i,1)}E_{i,j}\\
&\overset{n\to+\infty}{\longrightarrow} \underbrace{|I_n||\Lambda_n|^{-1}}_{\widetilde{D}^{-2}}\sum_{k\in\mathbb{B}(0,1)}E_{0,k}=\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star). \quad (16)
\end{aligned}$$

Step 2. Domination of $\underline{\mathbf{U}}_n^{(2)}(\Phi;\theta)$ in a neighborhood of $\theta^\star$ and convergence of $\underline{\mathbf{U}}_n^{(2)}(\Phi;\theta^\star)$ Let $j,k=1,\ldots,p$, recall that $\left(\underline{\mathbf{U}}_n^{(2)}(\varphi;\theta)\right)_{j,k}$ is defined in a neighborhood $\mathcal{V}(\theta^\star)$ of $\theta^\star$ for any configuration φ by

$$\begin{aligned}
\left(\underline{\mathbf{U}}_n^{(2)}(\varphi;\theta)\right)_{j,k} &= -\frac{1}{|\Lambda_n|}\int_{\Lambda_n}\frac{\partial^2 V}{\partial\theta_j\partial\theta_k}\left(x|\varphi;\theta\right)\exp\left(-V\left(x|\varphi;\theta\right)\right)dx\\
&\quad+\frac{1}{|\Lambda_n|}\int_{\Lambda_n}\frac{\partial V}{\partial\theta_j}\left(x|\varphi;\theta\right)\frac{\partial V}{\partial\theta_k}\left(x|\varphi;\theta\right)\exp\left(-V\left(x|\varphi;\theta\right)\right)dx
\end{aligned}$$

$$+\frac{1}{|\Lambda_n|}\sum_{x\in\varphi_{\Lambda_n}}\frac{\partial^2 V}{\partial\theta_j\partial\theta_k}(x|\varphi\setminus x;\theta)\,. \tag{17}$$

Under the assumption **[N1]** and **[N2]**, from Theorem 6 (Nguyen and Zessin (1979)) and from Corollary 5, there exists $n_0\in\mathbb{N}$ such that for all $n\geq n_0$

$$\begin{aligned}
&\left|\left(\underline{\mathbf{U}}_n^{(2)}(\varphi;\theta)\right)_{j,k}\right| \\
&\quad\leq\ 2\mathbf{E}\left(\left(\left|\frac{\partial^2 V}{\partial\theta_j\partial\theta_k}(0|\Phi;\theta)\right|+\left|\frac{\partial V}{\partial\theta_j}(0|\Phi;\theta)\,\frac{\partial V}{\partial\theta_k}(0|\Phi;\theta)\right|\right)e^{-V(0|\Phi;\theta)}\right) \\
&\qquad+2\times\mathbf{E}\left(\left|\frac{\partial^2 V}{\partial\theta_j\partial\theta_k}(0|\Phi;\theta)\right|e^{-V(0|\Phi;\theta^\star)}\right)
\end{aligned}$$

Note that from Theorem 6 (Nguyen and Zessin (1979)), $\underline{\mathbf{U}}_n^{(2)}(\cdot;\theta^\star)$ converges almost surely as $n\to+\infty$ towards $\underline{\mathbf{U}}^{(2)}(\theta^\star)$ defined by (3). Note that $\underline{\mathbf{U}}^{(2)}(\theta^\star)$ is a symmetric positive matrix since for all $\mathbf{y}\in\mathbb{R}^p$

$$\mathbf{y}^T\underline{\mathbf{U}}^{(2)}(\theta^\star)\mathbf{y}=\mathbf{E}\left(\left(\mathbf{y}^T\mathbf{V}^{(1)}(0|\Phi;\theta^\star)\right)^2 e^{-V(0|\Phi;\theta^\star)}\right)\geq 0,$$

where for $j=1,\ldots,p$, $\varphi\in\Omega$ and for $\theta\in\mathcal{V}(\theta^\star)$ $\left(\mathbf{V}^{(1)}(x|\varphi;\theta^\star)\right)_j:=\frac{\partial V}{\partial\theta_j}(x|\varphi;\theta)$ and it is a definite matrix under the assumption **[N3]**.
Conclusion Step Under the assumptions **[Mod]** and **[Ident]**, and using Steps 1 and 2, one can apply a classical result concerning asymptotic normality for minimum contrast estimators *e.g.* Proposition 3.7 of Guyon (1992) in order to obtain (5).

It remains to prove (6). This may de done in two different steps. The first one consists in verifying the positive definiteness of the matrix $\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star)$. The proof is strictly similar to the one of Billiot et al. (2008) (p. 261) except that the assumption [**SDP**] is now simply replaced by the more general one assumption **[N4]**. Now, the convergence in probability of $\underline{\widehat{\mathbf{\Sigma}}}_n(\Phi;\widehat{\theta}_n(\Phi))$ towards $\underline{\mathbf{\Sigma}}(\widetilde{D},\theta^\star)$ is obtained by applying Proposition 9 of Coeurjolly and Lavancier (2010).

Appendix B: Verifications of Assumptions [C] and [N] for LJ-type models

Before verifying the different assumptions, let us denote by

$$\theta_i^{\inf}:=\inf_{\theta\in\mathbf{\Theta}}\theta_i,\quad \theta_i^{\sup}:=\sup_{\theta\in\mathbf{\Theta}}\theta_i,$$

$$\theta^{\inf}:=\min(\theta_2^{\inf},\theta_3^{\inf})\quad\text{and}\quad\theta^{\sup}:=\max(\theta_2^{\sup},\theta_3^{\sup}).$$

Since $\mathbf{\Theta}$ is a compact set of $\mathbb{R}\times(]0,+\infty[)^2$, then $\theta^{\inf}>0$ and $\theta^{\sup}<+\infty$.

B.1. Assumptions [C]

B.1.1. Assumption [C1]

The first part is a direct application of Lemma 4. For the second part, one has to prove that for all $\theta \in \boldsymbol{\Theta}$

$$\mathbf{E}\left(|H^{LJ}(0|\Phi;\theta)\,|e^{-H^{LJ}(0|\Phi;\theta^\star)}\right) < +\infty$$

Let $g_\varepsilon(r) = g^{LJ}(r;\theta^\star) - \varepsilon|g^{LJ}(r;\theta)|$. We have

$$g_\varepsilon(r) := \begin{cases} 4\theta_2^\star\left(\frac{(\theta_3^\star)^{12}-\varepsilon\frac{\theta_2}{\theta_2^\star}\theta_3^{12}}{r^{12}} - \frac{(\theta_3^\star)^{6}-\varepsilon\frac{\theta_2}{\theta_2^\star}\theta_3^{6}}{r^{6}}\right) & \text{if } r \le \theta_3 \\ 4\theta_2^\star\left(\frac{(\theta_3^\star)^{12}+\varepsilon\frac{\theta_2}{\theta_2^\star}\theta_3^{12}}{r^{12}} - \frac{(\theta_3^\star)^{6}+\varepsilon\frac{\theta_2}{\theta_2^\star}\theta_3^{6}}{r^{6}}\right) & \text{if } r \ge \theta_3 \end{cases}$$

which satisfies the assumptions of Lemma 4 as soon as $\varepsilon < \left(\frac{\theta_3^\star}{\theta_3}\right)^{12}\frac{\theta_2^\star}{\theta_2}$, that is, as soon as $\varepsilon < \left(\frac{\theta^{\inf}}{\theta^{\sup}}\right)^{13}$.

B.1.2. Assumption [C2]

Let us denote for $n \ge 1$, $C_n = \mathcal{B}(0,n) \setminus \mathcal{B}(0,n-1)$ and define for $m, n \ge 1$ the following configuration sets

$$\begin{aligned} U_{m,n} &= \{\varphi \in \Omega : |\varphi_{C_n}| \le m|C_n|\} \\ U_m &= \cap_{n\ge 1} U_{m,n}. \end{aligned}$$

In order to prove **[C2]**, we need the following Lemma.

Lemma 7. *Let $R \in \mathbb{R}^+$, $\theta \in \boldsymbol{\Theta}$ and $\varphi \in U_m$, let us denote by*

$$Z(\varphi, R;\theta) := \sum_{x \in \varphi_{\mathcal{B}(0,R)^c}} g^{LJ}(||x||;\theta),$$

then for all $\delta > 0$ there exists R_0 such that for all $R \ge R_0$, $|Z(\varphi, R;\theta)| \le \delta$.

Proof.

$$\begin{aligned} Z(\varphi, R;\theta) = \left|\sum_{x\in\varphi_{\mathcal{B}(0,R)^c}} g^{LJ}(||x||;\theta)\right| &\le \sum_{n\ge\lceil R\rceil}\sum_{x\in\varphi_{C_n}} \left|g^{LJ}(||x||;\theta)\right| \\ &\le \sum_{n\ge\lceil R\rceil} |\varphi_{C_n}| \times \sup_{x\in C_n}\left|g^{LJ}(||x||;\theta^\star)\right|. \end{aligned}$$

There exists a constant $k = k(R)$ such that for all $n \ge \lceil R\rceil$,

$$\sup_{x\in C_n}\left|g^{LJ}(||x||;\theta^\star)\right| \le kn^{-6}.$$

Therefore,

$$\left| \sum_{x \in \varphi_{\mathcal{B}(0,R)^c}} g^{LJ}(||x||;\theta) \right| \leq km \sum_{n \geq \lceil R \rceil} |C_n| \times n^{-6} = \mathcal{O}\left(\sum_{n \geq \lceil R \rceil} n^{-5} \right),$$

which leads to the result since the previous series is convergent. □

Let $\theta \in \boldsymbol{\Theta} \setminus \theta^\star$ and consider the following configuration sets defined for $k \geq 1$ and for η small enough by

$$A_0 = \{\varphi \in \Omega : |\varphi \cap \mathcal{B}(0,D)| = 0\} \quad (18)$$

$$A_k(\eta) = \left\{\varphi \in \Omega : |\varphi \cap \mathcal{B}(0,D)| = |\varphi \cap \mathcal{B}((0,Dk^{-1/12}),\eta)| = 1\right\}, \quad (19)$$

where D is any positive real for the Lennard-Jones model and corresponds to the range of the function $g^{LJ}(\cdot)$ for the finite range Lennard-Jones model. There exists $m \geq 1$ such that for all $\eta > 0$ and for $k = 2, 4$

$$P_{\theta^\star}(A_0 \cap U_m) > 0 \quad \text{and} \quad P_{\theta^\star}(A_k(\eta) \cap U_m) > 0.$$

Now, let $\varphi_0 \in A_0 \cap U_m, \varphi_2 \in A_2(\eta) \cap U_m$ and $\varphi_4 \in A_4(\eta) \cap U_m$. First,

$$D(0|\varphi_0;\theta) = \theta_1 - \theta_1^\star + Z(\varphi_0, D; \theta) - Z(\varphi_0, D; \theta^\star) = 0.$$

For the Lennard-Jones model, according to Lemma 7 one has, for D large enough,

$$|Z(\varphi_0, D; \theta) - Z(\varphi_0, D; \theta^\star)| \leq \frac{1}{2}|\theta_1 - \theta_1^\star|.$$

Hence for η small enough, and for both models

$$\begin{aligned} 0 &= |D(0|\varphi_0;\theta)| \\ &\geq |\theta_1 - \theta_1^\star| - |Z(\varphi_0, D; \theta) - Z(\varphi_0, D; \theta^\star)| \\ &\geq \frac{1}{2}|\theta_1 - \theta_1^\star|, \end{aligned}$$

which leads to $\theta_1 = \theta_1^\star$. Moreover,

$$\begin{aligned} D(0|\varphi_2;\theta) &= 4\theta_2 \left(2\left(\frac{\theta_3}{D}\right)^{12} - \sqrt{2}\left(\frac{\theta_3}{D}\right)^{6}\right) - 4\theta_2^\star \left(2\left(\frac{\theta_3^\star}{D}\right)^{12} - \sqrt{2}\left(\frac{\theta_3^\star}{D}\right)^{6}\right) \\ &\quad + f_2(\varphi_2) + Z(\varphi_2, D; \theta) - Z(\varphi_2, D; \theta^\star) \\ D(0|\varphi_4;\theta) &= 4\theta_2 \left(4\left(\frac{\theta_3}{D}\right)^{12} - 2\left(\frac{\theta_3}{D}\right)^{6}\right) - 4\theta_2^\star \left(4\left(\frac{\theta_3^\star}{D}\right)^{12} - 2\left(\frac{\theta_3^\star}{D}\right)^{6}\right) \\ &\quad + f_4(\varphi_4) + Z(\varphi_4, D; \theta) - Z(\varphi_4, D; \theta^\star), \end{aligned}$$

where for any $\varphi_k \in A_k(\eta)$ $(k = 2, 4)$, there exists a positive function $\widetilde{f}_k(\eta)$ converging towards zero as $\eta \to 0$ such that $|f_k(\varphi_k)|$ is bounded by $\widetilde{f}_k(\eta)$. Now,

we have

$$\begin{aligned} 2D(0|\varphi_2;\theta) - D(0|\varphi_4;\theta) &= \frac{4(2-2\sqrt{2})}{D^6}\left(\theta_2\theta_3^6 - \theta_2^\star\theta_3^{\star 6}\right) + 2f(\varphi_2) - f_4(\varphi_4) \\ &\quad + Z'(\varphi_2,\varphi_4,D;\theta,\theta^\star) \\ &= 0 \end{aligned}$$

with

$$Z'(\varphi_2,\varphi_4,D;\theta,\theta^\star) := 2\left(Z(\varphi_2,D;\theta) - Z(\varphi_2,D;\theta^\star)\right) - \left(Z(\varphi_4,D;\theta) - Z(\varphi_4,D;\theta^\star)\right).$$

For η small enough, we have, for any $\varphi_k \in A_k(\eta)$ $(k=2,4)$,

$$|2f(\varphi_2) - f_4(\varphi_4)| \le 2\widetilde{f}_2(\eta) + \widetilde{f}_4(\eta) \le \frac{1}{4}\left|\frac{4(2-2\sqrt{2})}{D^6}\right| |\theta_2\theta_3^6 - \theta_2^\star\theta_3^{\star 6}|.$$

For the finite range Lennard-Jones model, $Z'(\varphi_2,\varphi_4,D;\theta,\theta^\star) = 0$. For the Lennard-Jones model, according to Lemma 7, one has for D large enough

$$|Z'(\varphi_2,\varphi_4,D;\theta,\theta^\star)| \le \frac{1}{4}\left|\frac{4(2-2\sqrt{2})}{D^6}\right| |\theta_2\theta_3^6 - \theta_2^\star\theta_3^{\star 6}|.$$

Hence for η small enough, and for both models

$$\begin{aligned} 0 &= \left|\frac{4(2-2\sqrt{2})}{D^6}\left(\theta_2\theta_3^6 - \theta_2^\star\theta_3^{\star 6}\right) + 2f(\varphi_2) - f_4(\varphi_4) + Z'(\varphi_2,\varphi_4,D;\theta,\theta^\star)\right| \\ &\ge \left|\frac{4(2-2\sqrt{2})}{D^6}\right| |\theta_2\theta_3^6 - \theta_2^\star\theta_3^{\star 6}| - |2f(\varphi_2) - f_4(\varphi_4)| - |Z'(\varphi_2,\varphi_4,D;\theta,\theta^\star)| \\ &\ge \frac{1}{2}\left|\frac{4(2-2\sqrt{2})}{D^6}\right| |\theta_2\theta_3^6 - \theta_2^\star\theta_3^{\star 6}| \end{aligned}$$

leading to $\theta_2\theta_3^6 = \theta_2^\star\theta_3^{\star 6}$. By considering the combination $\sqrt{2}D(0|\varphi_2;\theta) - D(0|\varphi_4;\theta)$ and using similar arguments as previously, one obtains: $\theta_2\theta_3^{12} = \theta_2^\star\theta_3^{\star 12}$. By computing the ratio of the two last equations, one obtains $\theta_3 = \theta_3^\star$ and then $\theta_2 = \theta_2^\star$.

B.1.3. Assumption [C4]

For all $\varphi \in \Omega$ and for any $\theta \in \boldsymbol{\Theta}$, $V^{LJ}(0|\varphi;\theta)$ is clearly differentiable in θ. First, note that **[C4]** is trivial for $j=1$. For $j=2,3$, let us define:

$$X_j(\varphi;\theta) := \left|\frac{\partial V^{LJ}}{\partial\theta_j}(0|\varphi;\theta)\right| e^{-V^{LJ}(0|\varphi;\theta)}.$$

Our aim will be to prove that for $j=2,3$ and for all $k>0$

$$\mathbf{E}\left(\max_{\theta\in\boldsymbol{\Theta}} X_j(\Phi;\theta)^k\right) < +\infty. \tag{20}$$

In particular, the Assumption **[C4]** corresponds to (20) with $k = 2$. Let us notice that for all $\varphi \in \Omega$ and for all $\theta \in \boldsymbol{\Theta}$

$$V^{LJ}(0|\varphi;\theta) \geq V^{\inf}(0|\varphi) := \theta^{\inf} + \sum_{x\in\varphi} g^{\inf}(||x||),$$

with for some $r > 0$, $g^{\inf}(r) := 4\theta^{\inf}\left(\frac{(\theta^{\inf})^{12}}{r^{12}} - \frac{(\theta^{\sup})^6}{r^6}\right)$. Let us also underline that for $j = 2, 3$

$$\frac{\partial g^{LJ}}{\partial \theta_j}(r;\theta) \geq \widetilde{g}_j^{\inf}(r) \quad \text{with } \widetilde{g}_j^{\inf}(r) := \begin{cases} 4\left(\frac{(\theta^{\inf})^{12}}{r^{12}} - \frac{(\theta^{\sup})^6}{r^6}\right) & \text{if } j = 2, \\ 4m\left(\frac{12(\theta^{\inf})^{11}}{r^{12}} - \frac{6(\theta^{\sup})^5}{r^6}\right) & \text{if } j = 3. \end{cases}$$

Therefore, by defining $\widetilde{V}_j^{\inf}(0|\varphi) := \sum_{x\in\varphi} \widetilde{g}_j^{\inf}(||x||)$, the result (20) will be ensured by proving

$$\mathbf{E}\left(\widetilde{V}_j^{\inf}(0|\Phi)e^{-V^{\inf}(0|\Phi)}\right) < +\infty.$$

According to Lemma 4, in order to prove this, let us denote by $g_{j,\varepsilon}(\cdot)$ the function defined for $j = 2, 3$, for some $\varepsilon > 0$ and for $r > 0$ by $g_{j,\varepsilon}(r) = \widetilde{g}_j^{\inf}(r) - \varepsilon\left|g^{\inf}(r)\right|$. On the one hand, one has

$$g_{2,\varepsilon}(r) = \begin{cases} 4\left(\frac{(\theta^{\inf})^{13} - \varepsilon(\theta^{\inf})^{12}}{r^{12}} - \frac{\theta^{\inf}(\theta^{\sup})^6 - \varepsilon(\theta^{\sup})^6}{r^6}\right) & \text{if } r \leq \frac{(\theta^{\inf})^2}{\theta^{\sup}}, \\ 4\left(\frac{(\theta^{\inf})^{13} + \varepsilon(\theta^{\inf})^{12}}{r^{12}} - \frac{\theta^{\inf}(\theta^{\sup})^6 + \varepsilon(\theta^{\sup})^6}{r^6}\right) & \text{if } r \geq \frac{(\theta^{\inf})^2}{\theta^{\sup}}, \end{cases}$$

which satisfies the assumptions of Lemma 4 as soon as $\varepsilon < \theta^{\inf}$. On the other hand

$$g_{3,\varepsilon}(r) = \begin{cases} 4\theta^{\inf}\left(\frac{(\theta^{\inf})^{12} - 12\varepsilon(\theta^{\inf})^{11}}{r^{12}} - \frac{(\theta^{\sup})^6 - 6\varepsilon(\theta^{\sup})^5}{r^6}\right) & \text{if } r \leq \left(2\frac{(\theta^{\inf})^{11}}{(\theta^{\sup})^5}\right)^{1/6} \\ 4\theta^{\inf}\left(\frac{(\theta^{\inf})^{12} + 12\varepsilon(\theta^{\inf})^{11}}{r^{12}} - \frac{(\theta^{\sup})^6 + 6\varepsilon(\theta^{\sup})^5}{r^6}\right) & \text{if } r \geq \left(2\frac{(\theta^{\inf})^{11}}{(\theta^{\sup})^5}\right)^{1/6}, \end{cases}$$

which satisfies the assumptions of Lemma 4 as soon as $\varepsilon < \theta^{\inf}/12$, which ends the proof.

B.2. Assumptions [N]

B.2.1. Assumption [N1]

Let us present two auxiliary lemmas.

Lemma 8. *Let φ be the realization of a stationary pairwise interaction point process with local energy function defined by*

$$V(x|\varphi;\theta) = \theta_1 + H(x|\varphi;\theta) \qquad \text{with} \qquad H(x|\varphi;\theta) = \sum_{y\in\varphi} g(||y-x||;\theta).$$

Let $K < +\infty$ and let $x_1,\ldots,x_K \in \mathbb{R}^2 \setminus \varphi$, $x_i \neq x_j$ for $i,j = 1,\ldots,K$ (where $K < +\infty$), then

$$\begin{aligned} H(\{x_1,\ldots,x_K\}|\varphi;\theta) &= \sum_{k=1}^{K} H(x_k|\varphi;\theta) + H(\{x_1,\ldots,x_K\};\theta) \\ V(\{x_1,\ldots,x_K\}|\varphi;\theta) &= \sum_{k=1}^{K} V(x_k|\varphi;\theta) + H(\{x_1,\ldots,x_K\};\theta) \end{aligned}$$

This result comes from the definition of the local energy.

Lemma 9. *Using the same notation and under the same assumptions of Lemma 8, assume that there exists g_{min} such that for all $r > 0$ and any $\theta \in \boldsymbol{\Theta}$, $g(r;\theta) \geq g_{min}$, then*

$$e^{-V(\{x_1,\ldots,x_K\}|\varphi;\theta)} \leq c_K \prod_{k=1}^{K} e^{-V(x_k|\varphi;\theta)} \qquad \text{with } c_K = e^{-\frac{K(K-1)}{2} g_{min}}$$

Proof. The proof is immediate since

$$H(\{x_1,\ldots,x_K\};\theta) = \sum_{i<j} g(||x_i - x_j||;\theta) \geq \frac{K(K-1)}{2} g_{min}.$$

□

Let $k = 1,\ldots,3$ and let $\lambda_1,\ldots,\lambda_k$, k positive integers such that $\sum_{i=1}^{k} \lambda_i = 3$ and define the random variable

$$A(\Phi) := \int_{\Delta^k} \prod_{i=1}^{k} \left| \frac{\partial V}{\partial \theta_j}(x_i|\Phi;\theta^\star) \right|^{\lambda_i} e^{-V(\{x_1,\ldots,x_k\}|\Phi;\theta^\star)} dx_i.$$

From Lemma 9, we have

$$\begin{aligned} \mathbf{E}(A(\Phi)) &\leq \mathbf{E}\left(c_k \int_{\Delta^k} \prod_{i=1}^{k} \left| \frac{\partial V}{\partial \theta_j}(x_i|\Phi;\theta^\star) \right|^{\lambda_i} e^{-V(x_i|\Phi;\theta^\star)} dx_i \right) \\ &= c_k \int_{\Delta^k} \mathbf{E}\left(\prod_{i=1}^{k} \left| \frac{\partial V}{\partial \theta_j}(x_i|\Phi;\theta^\star) \right|^{\lambda_i} e^{-V(x_i|\Phi;\theta^\star)} \right) dx_1 \ldots dx_k \\ &\leq c_k \int_{\Delta^k} \prod_{i=1}^{k} \mathbf{E}\left(\left| \frac{\partial V}{\partial \theta_j}(x_i|\Phi;\theta^\star) \right|^{k} e^{-\frac{k}{\lambda_i} V(x_i|\Phi;\theta^\star)} \right)^{1/k} dx_1 \ldots dx_k \end{aligned}$$

$$= c_k \prod_{i=1}^{k} \int_{\Delta} \mathbf{E}\left(\left| \frac{\partial V}{\partial \theta_j}(x_i|\Phi;\theta^\star) \right|^k e^{-\frac{k}{\lambda_i} V(x_i|\Phi;\theta^\star)} \right)^{1/k} dx_i$$

$$= c_k |\Delta|^k \prod_{i=1}^{k} \mathbf{E}\left(\left| \frac{\partial V}{\partial \theta_j}(0|\Phi;\theta^\star) \right|^k e^{-\frac{k}{\lambda_i} V(0|\Phi;\theta^\star)} \right)^{1/k}$$

by using Hölder's inequality and the stationarity of the process. The result is then a simple consequence of (20) and Lemma 4.

*B.2.2. Assumption [**N2**]*

For all $\varphi \in \Omega$, it is clear that for all $\theta \in \boldsymbol{\Theta}$, $V(0|\varphi;\theta)$ is twice continuously differentiable in θ. According to Lemma 4 and the fact that **[N1]** is satisfied, it is sufficient to prove that for all $j,k = 1,2,3$

$$\mathbf{E}\left(\left| \frac{\partial^2 V^{LJ}}{\partial \theta_j \partial \theta_k}(0|\Phi;\theta) \right| e^{-V^{LJ}(0|\Phi;\theta)} \right) < +\infty.$$

This is obvious when either j or k equals 1 and when $j = k = 2$ (since $\frac{\partial^2 g^{LJ}}{(\partial\theta_2)^2}(r;\cdot) = 0$). Now, for the other cases, define for $\theta \in \boldsymbol{\Theta}$ $g_{j,k,\varepsilon}(r) := g^{LJ}(r;\theta) - \varepsilon \left| \frac{\partial^2 g^{LJ}}{\partial\theta_j \partial\theta_k}(r;\theta) \right|$. We have

$$g_{2,3,\varepsilon}(r) = g_{3,2,\varepsilon}(r) = \begin{cases} 4\left(\frac{\theta_2\theta_3^{12} - 12\varepsilon\theta_3^{11}}{r^{12}} - \frac{\theta_3^6 - 6\varepsilon\theta_3^5}{r^6} \right) & \text{if } r \le 2^{1/6} \\ 4\left(\frac{\theta_2\theta_3^{12} + 12\varepsilon\theta_3^{11}}{r^{12}} - \frac{\theta_3^6 + 6\varepsilon\theta_3^5}{r^6} \right) & \text{otherwise} \end{cases}$$

which satisfies the assumptions of Lemma 4 as soon as $\varepsilon < \frac{\theta_2\theta_3}{12}$, that is, as soon as $\varepsilon < \frac{(\theta^{\inf})^2}{12}$. Finally,

$$g_{3,3,\varepsilon}(r) = \begin{cases} 4\left(\frac{\theta_2\theta_3^{12} - 132\varepsilon\theta_3^{10}}{r^{12}} - \frac{\theta_2\theta_3^6 - 30\varepsilon\theta_3^4}{r^6} \right) & \text{if } r \le \left(\frac{132}{30}\right)^{1/6}\theta_3 \\ 4\left(\frac{\theta_2\theta_3^{12} + 132\varepsilon\theta_3^{10}}{r^{12}} - \frac{\theta_2\theta_3^6 + 30\varepsilon\theta_3^4}{r^6} \right) & \text{otherwise} \end{cases}$$

which satisfies the assumptions of Lemma 4 as soon as $\varepsilon < \frac{\theta_2\theta_3^2}{132}$, that is, as soon as $\varepsilon < \frac{(\theta^{\inf})^3}{132}$.

*B.2.3. Assumption [**N3**]*

Let $\mathbf{y} = (y_1,y_2,y_3) \in \mathbb{R}^3$ and $g(\mathbf{y},\varphi) := \mathbf{y}^T \mathbf{V}_{LJ}^{(1)}(0|\varphi;\theta^\star)$. Let $\varphi_0 \in A_0$ and $\varphi_k(\eta) \in A_k(\eta)$ $(k = 2,4)$ where A_0 and $A_k(\eta)$ are defined by (18) and (19). Assume $g(\mathbf{y},\varphi_k) = 0$ for $k = 0,2,4$. Since, $g(\mathbf{y},\varphi_0) = y_1$, we have $y_1 = 0$. Now,

$$g(\mathbf{y},\varphi_2) = 4y_2\left(2\left(\frac{\theta_3^\star}{D}\right)^{12} - \sqrt{2}\left(\frac{\theta_3^\star}{D}\right)^6 \right) + 4y_3\theta_2^\star\left(2\frac{12\theta_3^{\star 11}}{D^{12}} - \sqrt{2}\frac{6\theta_3^{\star 5}}{D^6} \right)$$

$$
\begin{aligned}
& + f_2(\mathbf{y},\varphi_2) \\
g(\mathbf{y},\varphi_4) &= 4y_2\left(4\left(\frac{\theta_3^\star}{D}\right)^{12} - 2\left(\frac{\theta_3^\star}{D}\right)^{6}\right) + 4y_3\theta_2^\star\left(4\frac{12\theta_3^{\star 11}}{D^{12}} - 2\frac{6\theta_3^{\star 5}}{D^6}\right) \\
& + f_4(\mathbf{y},\varphi_4),
\end{aligned}
$$

where for any $\varphi_k \in A_k(\eta)$ $(k=2,4)$, there exists a positive function $\widetilde{f}_k(\mathbf{y},\eta)$ converging towards zero as $\eta \to 0$ such that $|f_k(\mathbf{y},\varphi_k)|$ is bounded by $\widetilde{f}_k(\mathbf{y},\eta)$. Now, we have

$$
2g(\mathbf{y},\varphi_2) - g(\mathbf{y},\varphi_4) = 4(2-2\sqrt{2})\frac{\theta_3^{\star 5}}{D^6}\left(\theta_3^\star y_2 + 6\theta_2^\star y_3\right) + 2f_2(\mathbf{y},\varphi_2) - f_4(\mathbf{y},\varphi_4) = 0.
$$

For η small enough, we have, for any $\varphi_k \in A_k(\eta)$ $(k=2,4)$,

$$
\begin{aligned}
|2f(\mathbf{y},\varphi_2) - f_4(\mathbf{y},\varphi_4)| &\leq 2|\widetilde{f}_2(\mathbf{y},\eta)| + |\widetilde{f}_4(\mathbf{y},\eta)| \\
&\leq \frac{1}{2}\left|4(2-2\sqrt{2})\frac{\theta_3^{\star 5}}{D^6}\left(\theta_3^\star y_2 + 6\theta_2^\star y_3\right)\right|.
\end{aligned}
$$

Hence for η small enough,

$$
0 = |2g(\mathbf{y},\varphi_2) - g(\mathbf{y},\varphi_4)| \geq \frac{1}{2}\left|4(2-2\sqrt{2})\frac{\theta_3^{\star 5}}{D^6}\left(\theta_3^\star y_2 + 6\theta_2^\star y_3\right)\right|,
$$

leading to the equation $\theta_3^\star y_2 + 6\theta_2^\star y_3 = 0$. By considering the linear combination $\sqrt{2}g(\mathbf{y},\varphi_2) - g(\mathbf{y},\varphi_4)$, we may obtain the equation $\theta_3^\star y_2 + 12\theta_2^\star y_3 = 0$ with similar arguments. Both equations lead to $y_2 = y_3 = 0$.

B.2.4. Assumption [N4]

The assumption **[N4]** may be rewritten for all $k = 1,\ldots,\ell$ and for all $\varphi_k \in A_k$ and $\varphi_0 \in A_0$: $\forall \mathbf{y} \in \mathbb{R}^3$

$$
\left(\mathbf{y}^T\left(\mathbf{LPL}_\Lambda^{(1)}(\varphi_k;\theta^\star) - \mathbf{LPL}_\Lambda^{(1)}(\varphi_0;\theta^\star)\right) = \mathbf{y}^T(\mathbf{L}(\varphi_k;\theta^\star) - \mathbf{R}(\varphi_k;\theta^\star)) = 0\right) \Longrightarrow \mathbf{y} = 0.
$$

where for any configuration $\varphi \in \overline{\Omega}$ and $\varphi_0 \in A_0$

$$
\begin{aligned}
\mathbf{L}(\varphi;\theta^\star) &:= \int_\Lambda \mathbf{V}_{LJ}^{(1)}(x|\varphi;\theta^\star)\, e^{-V^{LJ}(x|\varphi;\theta^\star)}dx - \int_\Lambda \mathbf{V}_{LJ}^{(1)}(x|\varphi_0;\theta^\star)\, e^{-V^{LJ}(x|\varphi_0;\theta^\star)}dx \\
\mathbf{R}(\varphi;\theta^\star) &:= \sum_{x\in\varphi\cap\overline{\Lambda}} \mathbf{V}_{LJ}^{(1)}(x|\varphi\setminus x;\theta^\star) - \sum_{x\in\varphi_0\cap\overline{\Lambda}} \mathbf{V}_{LJ}^{(1)}(x|\varphi_0\setminus x;\theta^\star).
\end{aligned}
$$

Concerning this assumption, we choose $\varphi_0 \in A_0 = \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0} = \emptyset\right\}$. Let $\mathbf{y} \in \mathbb{R}^3$ then

$$
\int_\Lambda \mathbf{y}^T \mathbf{V}_{LJ}^{(1)}(x|\varphi_0;\theta^\star)\, e^{-V^{LJ}(x|\varphi_0;\theta^\star)}dx = y_1 e^{-\theta_1^\star}\left|\overline{\Lambda}\right|
$$

and

$$\sum_{x\in\varphi_0\cap\overline{\Lambda}} \mathbf{y}^T \mathbf{V}_{LJ}^{(1)}(x|\varphi_0\setminus x;\theta^\star)=0.$$

Consider the following configuration set, defined for $\eta,\varepsilon>0$, by

$$A_2(\eta,\varepsilon)=\left\{\varphi\in\overline{\Omega}:\varphi_{\Delta_0}=\{z_1,z_2\} \text{ where } z_1\in\mathcal{B}(0,\eta), z_2\in\mathcal{B}((0,2\eta+\varepsilon),\eta)\right\}.$$

Note that for $z_1\in\mathcal{B}(0,\eta), z_2\in\mathcal{B}((0,2\eta+\varepsilon),\eta)$, $\varepsilon\leq||z_2-z_1||\leq\varepsilon+4\eta$. Let $\varphi_2\in A_2(\eta,\varepsilon)$ and $x\in\overline{\Lambda}$, then one may prove that for $j=2,3$

$$\begin{aligned} V^{LJ}(x|\varphi_2;\theta^\star) &= \theta_1^\star+2g^{LJ}(||x||;\theta^\star)+f(x,\eta,\varepsilon)\\ \frac{\partial V^{LJ}}{\partial\theta_j}(x|\varphi_2;\theta^\star) &= 2\frac{\partial g^{LJ}}{\partial\theta_j}(||x||;\theta^\star)+f_j(x,\eta,\varepsilon)\end{aligned}$$

where $f(x,\eta,\varepsilon)$ and $f_j(x,\eta,\varepsilon)$ are such that

$$\lim_{(\eta,\varepsilon)\to(0,0)} f(x,\eta,\varepsilon)=\lim_{(\eta,\varepsilon)\to(0,0)} f_j(x,\eta,\varepsilon)=0.$$

On the one hand, one may prove that there exists a function $f_L(\mathbf{y},\eta,\varepsilon)$ such that
$\lim_{(\eta,\varepsilon)\to(0,0)} f_L(\mathbf{y},\eta,\varepsilon)=0$ and such that

$$\mathbf{y}^T\mathbf{L}(\varphi_2;\theta^\star)=\mathbf{y}^T\mathbf{I}-y_1e^{-\theta_1^\star}|\overline{\Lambda}|+f_L(\mathbf{y},\eta,\varepsilon)$$

where

$$\mathbf{I}:=\int_{\overline{\Lambda}}\mathbf{h}(||x||;\theta^\star)e^{-\theta_1^\star-2g^{LJ}(||x||;\theta^\star)}dx$$

and

$$\mathbf{h}(r;\theta^\star):=\left(1,2\frac{\partial g^{LJ}}{\partial\theta_2}(r;\theta^\star),2\frac{\partial g^{LJ}}{\partial\theta_3}(r;\theta^\star)\right)^T.$$

On the other hand, there exists a function $f_R(\mathbf{y},\eta,\varepsilon)$ such that $\lim_{\eta\to0} f_R(\mathbf{y},\eta,\varepsilon)=0$

$$\mathbf{y}^T\mathbf{R}(\varphi_2;\theta^\star)=2y_1+2y_24\left(\left(\frac{\theta_3^\star}{\varepsilon}\right)^{12}-\left(\frac{\theta_3^\star}{\varepsilon}\right)^6\right)+2y_34\theta_2^\star\left(\frac{12\theta_3^{\star 11}}{\varepsilon^{12}}-\frac{6\theta_3^{\star 5}}{\varepsilon^6}\right)+f_R(\mathbf{y},\eta,\varepsilon).$$

Since

$$\begin{aligned}\varepsilon^{12}\mathbf{y}^T\left(\mathbf{L}(\varphi_2;\theta^\star)-\mathbf{R}(\varphi_2;\theta^\star)\right) &= \varepsilon^{12}\left(\mathbf{y}^T\mathbf{I}-y_1e^{-\theta_1^\star}|\overline{\Lambda}|+f_L(\mathbf{y},\eta,\varepsilon)-f_R(\mathbf{y},\eta,\varepsilon)\right)\\ &\quad -\varepsilon^6\left(2y_24\theta_3^{\star 6}+2y_34\theta_2^\star6\theta_3^\star5\right)+2y_24\theta_3^{\star 12}+2y_34\theta_2^\star12\theta_3^{\star 11}.\end{aligned}$$

For η and ε chosen small enough, one may prove that

$$0=\left|\varepsilon^{12}\mathbf{y}^T\left(\mathbf{L}(\varphi_2;\theta^\star)-\mathbf{R}(\varphi_2;\theta^\star)\right)\right|\geq\frac{1}{2}\left|2y_24\theta_3^{\star 12}+2y_34\theta_2^\star12\theta_3^{\star 11}\right|$$

leading to

$$2y_2 4\theta_3^{\star 12} + 2y_3 4\theta_2^\star 12\theta_3^{\star 11} = 0 \Leftrightarrow \theta_3^\star y_2 + 12\theta_2^\star y_3 = 0. \tag{21}$$

This means that

$$\mathbf{y}^T\mathbf{R}(\varphi_2;\theta^\star) = 2y_1 - \frac{1}{\varepsilon^6}\left(2y_2 4\theta_3^{\star 6} + 2y_3 4\theta_2^\star 6\theta_3^{\star 5}\right) + f_R(\mathbf{y},\eta,\varepsilon).$$

With similar arguments, we obtain that

$$2y_2 4\theta_3^{\star 6} + 2y_3 4\theta_2^\star 6\theta_3^{\star 5} = 0 \Leftrightarrow \theta_3^\star y_2 + 6\theta_2^\star y_3 = 0. \tag{22}$$

Equations (21) and (22) lead to $y_2 = y_3 = 0$. Now consider the following configuration set defined for some $k \geq 1$ and $\eta > 0$

$$A_k(\eta) = \left\{\varphi \in \overline{\Omega} : \varphi_{\Delta_0} = |\varphi \cap \mathcal{B}(0,\eta)| = k\right\}$$

and let $\varphi_k \in A_k(\eta)$. Then, one may prove that there exists a function $\widetilde{f}_L(\mathbf{y},\eta)$ such that $\lim_{\eta\to 0}\widetilde{f}_L(\mathbf{y},\eta) = 0$ and such that

$$\begin{aligned}\mathbf{y}^T\left(\mathbf{L}(\varphi_k;\theta^\star) - \mathbf{R}(\varphi_k;\theta^\star)\right) &= y_1\int_{\overline{\Lambda}} e^{-\theta_1^\star}\left(e^{-kg^{LJ}(||x||;\theta^\star)} - 1\right)dx - ky_1 \\ &\quad + \widetilde{f}_L(\mathbf{y},\eta) \\ &= 0\end{aligned}$$

Let us denote by $\Lambda_1 := \mathcal{B}(0,\min(\theta_3^\star, D))$ and $\Lambda_2 := \mathcal{B}(0,D)\setminus\Lambda_1$ Now let us consider two cases.

Case 1: $\theta_3^\star \leq D$. First note that for all $x \in \overline{\Lambda}$, $g^{LJ}(||x||;\theta^\star) \geq 0$. Then, for k large enough and for η small enough, we have

$$\left|\frac{1}{k}\int_{\Lambda_1} e^{-\theta_1^\star}\left(e^{-kg^{LJ}(||x||;\theta^\star)} - 1\right)dx\right| \leq \frac{|\Lambda_1|}{k}e^{-\theta_1^\star} \leq \frac{1}{4} \quad \text{and} \quad \left|\frac{1}{k}\widetilde{f}_L(\mathbf{y},\eta)\right| \leq \frac{|y_1|}{4}.$$

Hence for k large enough and for η small enough, we may obtain

$$\begin{aligned} 0 &= \frac{1}{k}\left|\mathbf{y}^T\left(\mathbf{L}(\varphi_k;\theta^\star) - \mathbf{R}(\varphi_k;\theta^\star)\right)\right| \\ &\geq |y_1| - \left|y_1\frac{1}{k}\int_{\Lambda_1} e^{-\theta_1^\star}\left(e^{-kg^{LJ}(||x||;\theta^\star)} - 1\right)dx + \frac{1}{k}\widetilde{f}_L(\mathbf{y},\eta)\right| \\ &\geq |y_1| - \frac{|y_1|}{4} - \frac{|y_1|}{4} = \frac{|y_1|}{2},\end{aligned}$$

which leads to $y_1 = 0$.

Case 2: $\theta_3^\star \geq D$**.** First note that for all $x \in \Lambda_2$,

$$g^{LJ}(||x||;\theta^\star) \leq g_m := g^{LJ}(D;\theta^\star) = 4\theta_2^\star\left(\left(\frac{\theta_3^\star}{D}\right)^{12} - \left(\frac{\theta_3^\star}{D}\right)^6\right) < 0.$$

On the one hand, for k large enough and for η small enough, we may have

$$\left|\frac{1}{k}y_1\int_{\Lambda_1} e^{-\theta_1^\star}\left(e^{-kg^{LJ}(||x||;\theta^\star)}-1\right)dx+\frac{1}{k}\widetilde{f}_L(\mathbf{y},\eta)-y_1\right|\leq\frac{|y_1|}{2}+|y_1|\leq\frac{3}{2}|y_1|.$$

On the other hand, we have for k large enough

$$\begin{aligned}\frac{1}{k}\left|y_1\int_{\Lambda_2} e^{-\theta_1^\star}\left(e^{-kg^{LJ}(||x||;\theta^\star)}-1\right)dx\right| &= \frac{|y_1|}{k}\int_{\Lambda_2} e^{-\theta_1^\star}\left(e^{-kg^{LJ}(||x||;\theta^\star)}-1\right)dx\\ &\geq \frac{|y_1|}{k}e^{-\theta_1^\star}|\Lambda_2|\left(e^{-kg_m}-1\right)\\ &= |y_1|e^{-\theta_1^\star}\frac{e^{k|g_m|}-1}{k}\\ &\geq 2|y_1|.\end{aligned}$$

Therefore for k large enough and for η small enough, we have

$$0=\frac{1}{k}\left|\mathbf{y}^T\left(\mathbf{L}(\varphi_k;\theta^\star)-\mathbf{R}(\varphi_k;\theta^\star)\right)\right|\geq 2|y_1|-\frac{3}{2}|y_1|=\frac{|y_1|}{2},$$

which leads to $y_1=0$.

Acknowledgements

We are grateful to the referee and the associate editor for their comments which helped us in improving a previous version.

References

Baddeley, A. and Turner, R.. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42: 283–322, 2000. MR1794056

Bertin, E., Billiot, J.-M., and Drouilhet, R.. Existence of "Nearest-Neighbour" Gibbs Point Models. *Ann. Appl. Probab.*, 31:895–909, 1999a. MR1747447

Bertin, E., Billiot, J.-M., and Drouilhet, R.. k-Nearest-Neighbour Gibbs Point Processes. *Markov Processes and Related Fields*, 5 (2):219–234, 1999b. MR1762174

Bertin, E., Billiot, J.-M., and Drouilhet, R.. Spatial Delaunay Gibbs Point Processes. *Stochastic Models*, 15(2):181–199, 1999c. MR1683602

Besag, J.. Spatial interaction and the statistical analysis of lattice system. *J. R. Statist. Soc. Ser. B*, 26:192–236, 1974. MR0373208

Besag, J., Milne, R., and Zachary, S.. Point process limits of lattice processes. *Ann. Appl. Prob.*, 19:210–216, 1982. MR0644433

Billiot, J.-M., Coeurjolly, J.-F., and Drouilhet, R.. Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes. *Electronic Journal of Statistics*, 2:234–264, 2008. MR2399195

Coeurjolly, J.-F. and Lavancier, F.. Residuals for stationary marked Gibbs point processes. *submitted for publication*, 2010. http://hal.archives-ouvertes.fr/hal-00453102/fr/.

Daley, D. and Vere-Jones, D.. *An introduction to the Theory of Point Processes*. Springer Verlag, New York, 1988. MR0950166

Dereudre, D.. Gibbs Delaunay tessellations with geometric hardcore condition. *J. Stat. Phys.*, 121(3-4):511–515, 2005.

Dereudre, D.. The existence of quermass-interaction processes for nonlocally stable interaction and nonbounded convex grains. *Adv. in Appl. Probab*, 41(3):664–681, 2009. MR2571312

Dereudre, D., Drouilhet, R., and Georgii, H.O.. Existence of Gibbsian point processes with geometry-dependent interactions. *submitted*, 2010. http://arxiv.org/abs/1003.2875.

Dereudre, D. and Lavancier, F.. Campbell equilibrium equation and pseudo-likelihood estimation for non-hereditary Gibbs point processes. *Bernoulli*, 15(4):1368–1396, 2009. MR2597597

Goulard, M., Särkkä, A., and Grabarnik, P.. Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scandinavian Journal of Statistics*, 23(3): 365–379, 1996.

Guyon, X.. *Champs aléatoires sur un réseau*. Masson, Paris, 1992.

Illian, J., Penttinen, A., and Stoyan, H.. *Statistical analysis and modelling of spatial point patterns*. Wiley-Interscience, 2008. MR2384630

Jensen, J.L. and Künsch, H.R.. On asymptotic normality of pseudo likelihood estimates of pairwise interaction processes. *Ann. Inst. Statist. Math.*, 46:475–486, 1994. MR1309718

Jensen, J.L. and Møller, J.. Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.*, 1:445–461, 1991. MR1111528

Kendall, W.S., Van Lieshout, M.N.M., and Baddeley, A.J.. Quermass-interaction processes conditions for stability. *Advances on applied probability*, 31:315–342, 1999. MR1724554

Mase, S.. Consistency of maximum pseudo-likelihood estimator of continuous state space gibbsian process. *Ann. Appl. Probab.*, 5:603–612, 1995. MR1359821

Mase, S.. Marked gibbs processes and asymptotic normality of maximum pseudo-likelihhod estimators. *Math. Nachr.*, 209:151–169, 2000. MR1734363

Møller, J.. Parametric methods for spatial point processes. Technical Report Research Report R-2008-04, Department of Mathematical Sciences, Aalborg University, 2008.

Møller, J. and Waagepetersen, R.. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, 2003. MR2004226

Nguyen, X.X. and Zessin, H.. Ergodic theorems for Spatial Process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 48: 133–158, 1979. MR0534841

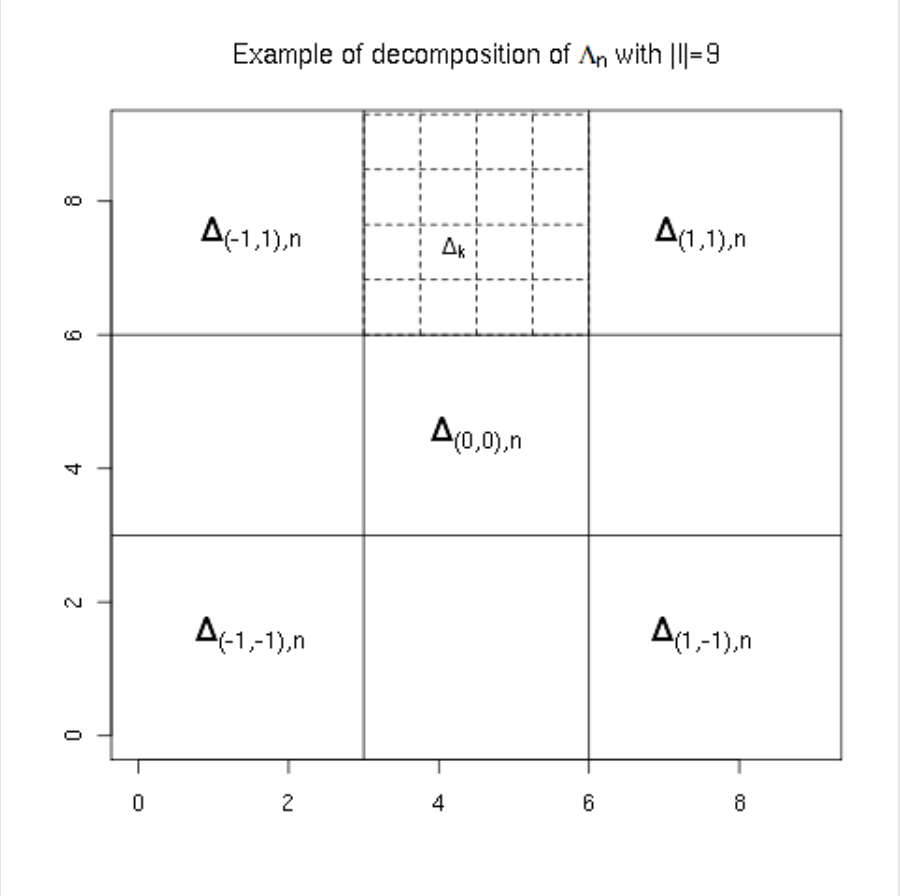
Ogata, Y. and Tanemura, M.. Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Annals of the Institute of Statistical Mathematics*, 33(1):315–338, 1981.

Preston, C.J.. *Random fields*. Springer Verlag, 1976. MR0448630

Ruelle, D.. *Statistical Mechanics*. Benjamin, New York-Amsterdam, 1969. MR0289084

Ruelle, D.. Superstable interactions in classical statistical mechanics. *Commun. Math. Phys.*, 18:127–159, 1970. MR0266565

Stoyan, D., Kendall, W.S., Mecke, J., and Ruschendorf, L.. *Stochastic geometry and its applications*. John Wiley and Sons, Chichester, 1987. MR0895588



[CL10]
J.-F. Coeurjolly and F. Lavancier. Residuals for stationary marked Gibbs point processes. *submitted for publication* (available at `http://arxiv.org/abs/1002.0857`), 2010.

Residuals and goodness-of-fit tests for stationary marked Gibbs point processes

Jean-François Coeurjolly[1,2] and Frédéric Lavancier[3]
[1] Gipsa-lab, CNRS, Grenoble, France
[2] Laboratory Jean Kuntzmann, Grenoble, France.
[3] Laboratoire Jean Leray, Nantes, France.

Abstract

The inspection of residuals is a fundamental step to investigate the quality of adjustment of a parametric model to data. For spatial point processes, the concept of residuals has been recently proposed by Baddeley et al. Baddeley et al. (2005) as an empirical counterpart of the *Campbell equilibrium* equation for marked Gibbs point processes. The present paper focuses on stationary marked Gibbs point processes and deals with asymptotic properties of residuals for such processes. In particular, the consistency and the asymptotic normality are obtained for a wide class of residuals including the classical ones (raw residuals, inverse residuals, Pearson residuals). Based on these asymptotic results, we define goodness-of-fit tests with Type-I error theoretically controlled. One of these tests constitutes an extension of the quadrat counting test widely used to test the null hypothesis of a homogeneous Poisson point process.

AMS 2000 subject classifications: Primary 62M30, 60G55; secondary 60K35, 62F03, 62F05, 62F12

Keywords: stationary marked Gibbs point processes, residuals, goodness-of-fit test, quadrat counting test, maximum pseudolikelihood estimator, Campbell Theorem, Georgii-Nguyen-Zessin formula, central limit theorem for spatial random fields

1 Introduction

Recent works on statistical methods for spatial point pattern makes parametric inference feasible for a wide range of models, see Møller (2008) for an overview of this topic and more generally the books of Daley and Vere-Jones (1988), Stoyan et al. (1987) Møller and Waagepetersen (2003) or Illian et al. (2008) for a survey on spatial point processes. The question is then to know whether the model is well-fitted to data or not. For classical parametric models, this is usually done via the inspection of residuals. They play a central role in parametric inference, see Atkinson (1985) for instance. This notion is quite complex for spatial point processes and has been recently proposed by Baddeley et al. Baddeley et al. (2005), following ideas from a previous work of Stoyan and Grabarnik Stoyan and Grabarnik (1991).

The definition of residuals for spatial point processes is a natural generalization of the well-known residuals for point processes in one-dimensional time, used in survival analysis (see Fleming and Harrington (1991) or Andersen et al. (1993) for an overview). For example, a simple measure of the adequacy of a one-dimensional point process model consists in computing the difference between the number of events in an interval $[0, t]$ and the conditional intensity (or hazard rate of the lifetime distri-

bution) parametrically estimated and integrated from 0 to t. The extension in higher dimension requires further developments due to the lack of natural ordering. It may be done for point processes admitting a conditional density with respect to the Poisson process. These point processes correspond to the Gibbs measures. The equilibrium in one dimension between the number of events and the integrated hazard rate may be replaced in higher dimension by the *Campbell equilibrium* equation or *Georgii-Nguyen-Zessin* formula (see Georgii (1976), Nguyen and Zessin (1979a) and Section 2.3), which is the basis for defining the class of $h-$residuals where h represents a test function. In particular, Baddeley et al. Baddeley et al. (2005) consider different choices of h leading to the so-called raw residuals, inverse residuals and Pearson residuals, and show that they share similarities with the residuals obtained for generalized linear models.

Thanks to various diagnostic plots developped in the seminal paper Baddeley et al. (2005) and implementation within the `R` package `spatstat` Baddeley and Turner (2005), residuals appear to be a very convenient tool in practice. Some properties of the residuals process are exhibited in Baddeley et al. (2005) and Baddeley et al. (2008), including a conditional independence property and variance formulae in particular cases. In these two papers, the authors conjecture that a strong law of large numbers and a central limit theorem should hold for the residuals as the sampling window expands.

Our paper addresses this question for $d-$dimensional stationary marked Gibbs point processes. We obtain the strong consistency and the asymptotic normality in several contexts for a large class of test functions h. The $h-$residuals crucially depend on an estimate of the parameter vector. We consider the natural framework where the estimate is computed with the same data over which the $h-$residuals are assessed. The assumptions are very general and we show that they are fulfilled for several classical models, including the area interaction point process, the multi-Strauss marked point process, the Strauss type disc process, the Geyer's triplet point process, etc. The assumptions on the estimator are quite natural and we show that they are fulfilled in particular by the maximum pseudolikelihood estimator (in short MPLE) (see Baddeley and Turner (2000) for instance), for which asymptotic properties are now well-known (see Jensen and Møller (1991), Jensen and Künsch (1994), Billiot et al. (2008), Dereudre and Lavancier (2009) and Coeurjolly and Drouilhet (2009)).

Moreover, based on these asymptotic results, we propose statistical goodness-of-fit tests for which the Type-I error is asymptotically controlled. To the best of our knowledge, this is the first attempt in this direction. Such tests exist for rejecting the assumption of a homogeneous or inhomogeneous Poisson point process, but for general marked Gibbs point processes, the existing validation methods are either graphical (for example by using the QQ-plot proposed by Baddeley et al. (2005)) or rely on Monte-Carlo based simulations. We present two tests based on the computation of the residuals on different subdomains of the observation window. They extend in a very natural way the quadrat counting test for homogeneous Poisson distributions (see Diggle (2003) for instance). Besides, we present a test which combines several different $h-$residuals (associated to different functions h), computed on the entire observation window. The next step will be to implement these testing procedures to assess their power, compare them and reveal their limits. A thorough study will require extensive simulations and should deserve a separate paper.

The rest of the paper is organized as follows. Section 2 gathers the main notation used in this paper and briefly displays the general background. The definition of marked Gibbs point processes is given. They depend exclusively on the choice of an energy function or equivalently, of a local energy function. All the assumptions are based on this function. The *Georgii-Nguyen-Zessin* formula is recalled, leading to the definition of the $h-$innovations and $h-$residuals. Some examples are presented including the classical residuals considered by Baddeley et al. (2005) and new ones connected to the well-known empty space function (or spherical contact distribution), denoted in the literature by F, see Møller and Waagepetersen (2003) for instance.

Section 3 deals with asymptotic properties and presents our main results. A parametric stationary d-dimensional marked Gibbs point process is observed in a domain, denoted by Λ_n, assumed to

increase up to $\mathbb{R}^d$. Sufficient conditions expressed in terms of the test function and the local energy function are given in order to derive the strong consistency result. We also propose an asymptotic control in probability of the departure of the h-residuals process from the h-innovations through the departure of the estimate from the true parameter vector (see Proposition 3). This allows us to deduce asymptotic normality results. Two different frameworks are considered: for the first one, the initial domain is splitted into a fixed finite number of subdomains (with volumes aimed at converging to $+\infty$) and we consider the vector composed of the $h-$residuals computed on each subdomain. For the second framework, we consider the vector composed of the h_j-residuals (for $j = 1, \dots, s$) computed on the same domain Λ_n, where $h_1, \dots, h_s$ are different test functions.

The asymptotic normality results depend on unknown asymptotic covariance matrices. The important question of estimating these matrices is addressed in Section 4. We give a general condition under which these matrices are definite-positive and propose a consistent estimate.

Section 5 exploits the asymptotic results obtained before. Some goodness-of-fit tests are proposed, based on normalized residuals computed in the two previous frameworks. They are shown to converge to some χ^2 distribution. Framework 1 leads to a generalization of the quadrat counting test for homogeneous Poisson distributions. Framework 2 yields a test which combines the information coming from several residuals, as for instance residuals coming from the estimation of the empty space function at several points.

The different assumptions made in the previous sections to obtain asymptotic results are discussed in Section 6. When considering classical test functions, exponential family models and the MPLE, the regularity and integrability type assumptions are shown to be satisfied for a wide class of examples. The testing procedures require moreover an identifiability condition to provide a proper normalization. Proposition 16 shows that this condition is easy to check for the first proposed test and appears to be not restrictive in this case. For the other tests, checking this condition depends more specifically on the model and the test function. We show, in Proposition 18, how this condition can be verified on two examples of models with several choices of test functions.

In Section 7, the very special situation where the energy function is not hereditary is considered. The GNZ formula is not valid any more in this setting but, provided a slight modification, it has been recently extended in Dereudre and Lavancier (2009). This leads to a natural generalization of the residuals to the non-hereditary setting.

Proofs of our main results are postponed to Section 8. The main material is composed of an ergodic theorem obtained by Nguyen and Zessin (1979b) and a new multivariate central limit theorem for spatial processes. Our setting actually involves some non stationary conditional centered random fields. A general central limit theorem adapted to this context has been obtained in Comets and Janzura (1998) for self-normalized sums (see also Jensen and Künsch (1994) in the stationary case and without self-normalization). But, contrary to these papers where the observation domain is assumed to be of the form $[-n, n]^d$, we consider domains that may increase continuously up to $\mathbb{R}^d$. This particularity, which seems more relevant, requires an extension of the results in Comets and Janzura (1998) and Jensen and Künsch (1994) to triangular arrays. This new central limit theorem is presented in Appendix A.

2 Background on marked Gibbs point processes and definition of residuals

2.1 General notation, configuration space

We denote by $\mathcal{B}(\mathbb{R}^d)$ the space of bounded Borel sets in $\mathbb{R}^d$. For any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, Λ^c denotes the complementary set of Λ inside $\mathbb{R}^d$. The norm $|.|$ will be used without ambiguity for different kind of objects. For a vector $\mathbf{x}$, $|\mathbf{x}|$ represents the uniform norm of $\mathbf{x}$; For a countable set $\mathcal{J}$, $|\mathcal{J}|$ represents the number of elements belonging to $\mathcal{J}$; For a set $\Delta \in \mathcal{B}(\mathbb{R}^d)$, $|\Delta|$ is the volume of Δ.

Let $\underline{\mathbf{M}}$ be a matrix, we denote by $\|\underline{\mathbf{M}}\|$ the Frobenius norm of $\underline{\mathbf{M}}$ defined by $\|\underline{\mathbf{M}}\|^2 = Tr(\underline{\mathbf{M}}^T\underline{\mathbf{M}})$, where Tr is the trace operator. For a vector $\mathbf{x}$, $\|\mathbf{x}\|$ is simply its euclidean norm.

For all $\mathbf{x} \in \mathbb{R}^d$ and $\rho > 0$, $\mathcal{B}(\mathbf{x}, \rho) := \{\mathbf{y},\ |\mathbf{y} - \mathbf{x}| < \rho\}$. Let us also consider the short notation, for $i \in \mathbb{Z}^d$, $\mathbb{B}_i(\rho) = \mathcal{B}(i, \rho) \cap \mathbb{Z}^d$.

The space $\mathbb{R}^d$ is endowed with the Borel σ-algebra and the Lebesgue measure λ. Let $\mathbb{M}$ be a measurable space, which aims at being the mark space, endowed with the σ-algebra $\mathcal{M}$ and the probability measure λ^{m}. The state space of the point processes will be $\mathbb{S} := \mathbb{R}^d \times \mathbb{M}$ measured by $\mu := \lambda \otimes \lambda^{\mathrm{m}}$. We shall denote for short $x^m = (x, m)$ an element of $\mathbb{S}$.

The space of point configurations will be denoted by $\Omega = \Omega(\mathbb{S})$. This is the set of simple integer-valued measures on $\mathbb{S}$. It is endowed with the σ-algebra $\mathcal{F}$ generated by the sets $\{\varphi \in \Omega,\ \varphi(\Lambda \times A) = n\}$ for all $n \in \mathbb{N}$, for all $A \in \mathcal{M}$ and for all $\Lambda \in \mathcal{B}(\mathbb{R}^d)$. For any $x^m \in \mathbb{S}$ and $\varphi \in \Omega$, we denote $x^m \in \varphi$ if $\varphi(x^m) > 0$. For any $\varphi \in \Omega$ and any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, we denote $\varphi_\Lambda := \varphi_{\Lambda\times\mathbb{M}}$ the projection of φ onto $\Lambda \times \mathbb{M}$, which is just the mesure $\sum_{x^m \in \varphi \cap (\Lambda\times\mathbb{M})} \delta_{x^m}$, where δ_x is the Dirac measure at x. We will use without ambiguity some set notation for elements in Ω, e.g. $\varphi \cup \{x^m\} = \varphi \cup x^m := \varphi + \delta_{x^m}$ and for $x^m \in \varphi$, $\varphi \setminus \{x^m\} = \varphi \setminus x^m := \varphi - \delta_{x^m}$. For any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, the number of elements of φ_Λ is denoted by $|\varphi_\Lambda| := \varphi(\Lambda \times \mathbb{M})$.

2.2 Marked Gibbs point processes

The framework of this paper is restricted to stationary marked Gibbs point processes. Since we are interested in asymptotic properties, we consider these point processes on the infinite volume $\mathbb{R}^d$. Let us briefly recall their definition.

A marked point process Φ is a Ω-valued random variable, with probability distribution P on $(\Omega, \mathcal{F})$. The most prominent marked point process is the marked Poisson process π^ν with intensity measure ν on $\mathbb{R}^d$ (and mark density λ^{m}). The homogeneous marked Poisson process arises when $\nu = z\lambda$, with $z > 0$.

Let $\theta \in \mathbf{\Theta}$, where $\mathbf{\Theta}$ is some compact set of $\mathbb{R}^p$ (for some $p \geq 1$). For any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, let us consider the parametric function $V_\Lambda(.;\theta)$ from Ω into $\mathbb{R} \cup \{+\infty\}$. For fixed θ, $(V_\Lambda(.;\theta))_{\Lambda\in\mathcal{B}(\mathbb{R}^d)}$ constitutes a compatible family of energies if, for every $\Lambda \subset \Lambda'$ in $\mathcal{B}(\mathbb{R}^d)$, there exists a measurable function $\psi_{\Lambda,\Lambda'}$ from Ω into $\mathbb{R} \cup \{+\infty\}$ such that

$$\forall \varphi \in \Omega \quad V_{\Lambda'}(\varphi;\theta) = V_\Lambda(\varphi;\theta) + \psi_{\Lambda,\Lambda'}(\varphi_{\Lambda^c};\theta). \tag{1}$$

From a physical point of view, $V_\Lambda(\varphi_\Lambda;\theta)$ is the energy of φ_Λ in Λ given the outside configuration φ_{Λ^c}. The following definition is the classical way to define Gibbs measures through their conditional specifications (see Preston (1976)).

Definition 1. *A probability measure P_θ on Ω is a marked Gibbs measure for the compatible family of energies $(V_\Lambda(.;\theta))_{\Lambda\in\mathcal{B}(\mathbb{R}^d)}$ and the intensity ν if for every $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, for P_θ-almost every outside configuration φ_{Λ^c}, the law of P_θ given φ_{Λ^c} admits the following conditional density with respect to π^ν:*

$$f_\Lambda(\varphi_\Lambda|\varphi_{\Lambda^c};\theta) = \frac{1}{Z_\Lambda(\varphi_{\Lambda^c};\theta)} e^{-V_\Lambda(\varphi;\theta)},$$

where $Z_\Lambda(\varphi_{\Lambda^c};\theta)$ is a normalization called the partition function.

The existence of a Gibbs measure on Ω which satisfies these conditional specifications is a difficult issue. We do not want to open this discussion here and we will assume that the Gibbs measures we consider exist. We refer the interested reader to Ruelle (1969); Preston (1976); Bertin et al. (1999); Dereudre (2005); Dereudre et al. (2010), see also Section 6 for several examples.

In this article, we focus on stationary marked point processes on $\mathbb{S}$, i.e. on point processes admitting a conditional density with respect to the homogeneous marked Poisson process π. Moreover, without loss of generality, the intensity of the Poisson process, z, is fixed to 1. We assume in

a first step that the family of energies is hereditary, which means that for any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$, for any $\varphi \in \Omega$, and for all $x^m \in \Lambda \times \mathbb{M}$,

$$V_\Lambda(\varphi;\theta)) = +\infty \Rightarrow V_\Lambda(\varphi \cup \{x^m\};\theta)) = +\infty, \tag{2}$$

or equivalently, for all $x^m \in \varphi_\Lambda$, $f_\Lambda(\varphi_\Lambda|\varphi_{\Lambda^c};\theta) > 0 \Rightarrow f_\Lambda(\varphi_\Lambda \setminus \{x^m\}|\varphi_{\Lambda^c};\theta) > 0$. The non-hereditary case will be considered in Section 7. The main assumption is then the following.

> **[Mod-E]**: For any $\theta \in \mathbf{\Theta}$, the compatible family of energies $(V_\Lambda(.;\theta))_{\Lambda\in\mathcal{B}(\mathbb{R}^d)}$ is hereditary, invariant by translation, and such that an associated Gibbs measure P_θ exists and is stationary. Our data consist in the realization of a point process with Gibbs measure $P_{\theta^\star}$. The vector $\theta^\star$ is thus the true parameter to be estimated, assumed to be in $\mathring{\mathbf{\Theta}}$.

The local energy to insert a marked point x^m into the configuration φ is defined for any Λ containing x^m by

$$V\left(x^m|\varphi;\theta\right) := V_\Lambda(\varphi \cup \{x^m\}) - V_\Lambda(\varphi).$$

From the compatibility of the family of energies, i.e. (1), this definition does not depend on Λ. We restrict our study to finite-range interaction point processes, which is the main limitation of this paper.

> **[Mod-L]**: There exists $D \geq 0$ such that for all $(m,\varphi) \in \mathbb{M} \times \Omega$
>
> $$V\left(0^m|\varphi;\theta\right) = V\left(0^m|\varphi_{\mathcal{B}(0,D)};\theta\right).$$

2.3 Definitions of residuals for spatial point processes

The basic ingredient for the definition of residuals is the so-called GNZ formula stated below.

Theorem 1 (Georgii-Nguyen-Zessin Formula)**.** *Under **[Mod-E]**, for any function $h(\cdot,\cdot;\theta) : \mathbb{S} \times \Omega \to \mathbb{R}$ (eventually depending on some parameter θ) such that the following quantities are finite, then*

$$\mathbf{E}\left(\int_{\mathbb{R}^d\times\mathbb{M}} h\left(x^m,\Phi;\theta\right) e^{-V\left(x^m|\Phi;\theta^\star\right)}\mu(dx^m)\right) = \mathbf{E}\left(\sum_{x^m\in\Phi} h\left(x^m,\Phi\setminus x^m;\theta\right)\right), \tag{3}$$

where $\mathbf{E}$ denotes the expectation with respect to $P_{\theta^\star}$.

For stationary marked Gibbs point processes, (3) reduces to

$$\mathbf{E}\left(h\left(0^M,\Phi;\theta\right) e^{-V\left(0^M|\Phi;\theta^\star\right)}\right) = \mathbf{E}\left(h\left(0^M,\Phi\setminus 0^M;\theta\right)\right) \tag{4}$$

where M denotes a random variable with probability distribution λ^{m}. The following definition is based on empirical versions of both terms appearing in (4).

Definition 2. *For any bounded domain Λ, let us define the $h-$innovations (denoted by I_Λ) and the $h-$residuals (denoted by R_Λ and depending on an estimate $\widehat{\theta}$ of $\theta^\star$) by*

$$\begin{aligned}
I_\Lambda\left(\varphi;h,\theta^\star\right) &:= \int_{\Lambda\times\mathbb{M}} h\left(x^m,\varphi;\theta^\star\right) e^{-V\left(x^m|\varphi;\theta^\star\right)}\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda} h\left(x^m,\varphi\setminus x^m;\theta^\star\right)\\
R_\Lambda\left(\varphi;h,\widehat{\theta}\right) &:= \int_{\Lambda\times\mathbb{M}} h\left(x^m,\varphi;\widehat{\theta}\right) e^{-V\left(x^m|\varphi;\widehat{\theta}\right)}\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda} h\left(x^m,\varphi\setminus x^m;\widehat{\theta}\right).
\end{aligned}$$

From a practical point of view, the last notion is the most interesting since it provides a computable measure. The main examples considered by Baddeley et al. in Baddeley et al. (2005) (in the context of stationary point processes) are obtained by setting $h\left(x^m, \varphi; \theta\right) = 1$ for the raw residuals, $h\left(x^m, \varphi; \theta\right) = e^{V(x^m|\varphi;\theta)}$ for the inverse residuals and $h\left(x^m, \varphi; \theta\right) = e^{V(x^m|\varphi;\theta)/2}$ for the Pearson residuals. In particular, one may note that the raw residuals constitutes a difference of two estimates of the intensity of the point process (up to a normalisation by $|\Lambda|$): the first one is a parametric one and depends on the model while the second one is a nonparametric one (since it is equal to $|\varphi_\Lambda|$). Another more evolved example is to consider the function defined for $r > 0$ by

$$h_r(x^m, \varphi; \theta) := \mathbf{1}_{[0,r]}(d(x^m, \varphi))\, e^{V(x^m|\varphi;\theta)}$$

where $d(x^m, \varphi) = \inf_{y^m \in \varphi} \|y - x\|$. Considering this function leads to

$$R_\Lambda(\varphi; h_r, \widehat{\theta}) = \int_{\Lambda \times \mathbb{M}} \mathbf{1}_{[0,r]}(d(x^m, \varphi))\mu(dx^m) - \sum_{x^m \in \varphi_\Lambda} h_r(x^m, \varphi \setminus x^m; \widehat{\theta}). \tag{5}$$

Then for a large window $R(\varphi; h_h, \widehat{\theta})/|\Lambda|$ leads to a difference of two estimates of the well-known empty space function F at distance r. Recall that for a marked stationary point process (see Møller and Waagepetersen (2003) for instance)

$$F(r) := P\left(d(0^M, \Phi) \leq r\right).$$

The first term in the right hand side of (5) corresponds to the natural nonparametric estimator of $F(r)$ while the second one is a parametric estimator of $F(r)$.

3 Asymptotic properties

From now on, we assume that the point process satisfies **[Mod-E]** and **[Mod-L]**, that is **[Mod]**. The realization of $\Phi \sim P_{\theta^\star}$ is assumed to be observed in a domain $\Lambda_n \oplus D^+$, with $D^+ \geq D$, aimed at growing up to $\mathbb{R}^d$ as $n \to +\infty$. According to the locality assumption **[Mod-L]**, we are thus ensured that the $h-$innovations and $h-$residuals can be computed.

The aim of this section is to present several asymptotic properties for I_{Λ_n} and R_{Λ_n}. We prove their consistency and we propose two asymptotic normality results within different frameworks:

- Framework 1: for a fixed test function h, Λ_n is a cube, divided into a fixed finite number of sub-cubes (which will increase with Λ_n). The purpose is then to obtain the asymptotic normality for the vector composed of the $h-$residuals computed in each sub-cube.
- Framework 2: we consider $h_1, \ldots, h_s$ s different test functions and the aim is to obtain the asymptotic normality of the vector composed of the h_j-residuals computed on Λ_n.

In both frameworks, an estimate of $\theta^\star$ is involved. We assume that it is computed from the full domain Λ_n with the same data used to evaluate the $h-$residuals, which is a natural setting in practice. Moreover, contrary to the previous works dealing with asymptotic properties on Gibbs point processes (*e.g.* Jensen and Künsch (1994), Comets and Janzura (1998) or Billiot et al. (2008)), where Λ_n is assumed to be of the discrete form $[-n, n]^d$, we consider general domains that may grow continuously up to $\mathbb{R}^d$.

The asymptotic results obtained in this section are the basis to derive goodness-of-fit tests, as presented in Section 5.

3.1 Consistency of the residuals process

We obtain consistency results for $I_{\tilde{\Lambda}_n}\left(\Phi; h, \theta^\star\right)$ and $R_{\tilde{\Lambda}_n}\left(\Phi; h, \widehat{\theta}_n(\Phi)\right)$, where for all $n \geq 1$, $\tilde{\Lambda}_n \subset \Lambda_n$, $(\tilde{\Lambda}_n)_{n\geq 1}$ and $(\Lambda_n)_{n\geq 1}$ are regular sequences whose size increases to ∞.

The assumption **[C]** gathers the two following assumptions:

[**C1**]

$$\mathbf{E}\left(\left|h\left(0^M,\Phi;\theta^\star\right)\right| e^{-V\left(0^M|\Phi;\theta^\star\right)}\right) < +\infty.$$

[**C2**] For all $(m,\varphi)\in\mathbb{M}\times\Omega$, the functions $h\left(0^m,\varphi;\theta\right)$ and $f\left(0^m,\varphi;\theta\right) := h\left(0^m,\varphi;\theta\right) e^{-V\left(0^m|\varphi;\theta\right)}$ are continuously differentiable with respect to θ in a neighborhood $\mathcal{V}(\theta^\star)$ of $\theta^\star$ and

$$\mathbf{E}\left(\left\|\mathbf{f}^{(1)}\left(0^M,\Phi;\theta^\star\right)\right\|\right) < +\infty \quad \text{and} \quad \mathbf{E}\left(\left\|\mathbf{h}^{(1)}\left(0^M,\Phi;\theta^\star\right)\right\| e^{-V\left(0^M|\Phi;\theta^\star\right)}\right) < +\infty,$$

where $\mathbf{f}^{(1)}$ denotes the gradient vector of f with respect to θ.

Concerning the residuals process, we also need to assume

[**E1**] The estimator $\widehat{\theta}_n(\varphi)$ of $\theta^\star$, computed from the full observation domain Λ_n, converges for $P_{\theta^\star}-$a.e. φ towards $\theta^\star$, as $n\to+\infty$.

Proposition 2. *Assuming **[Mod]**, we have as $n\to+\infty$*

*(a) Under **[C1]**: for $P_{\theta^\star}-$a.e. φ, $|\tilde{\Lambda}_n|^{-1} I_{\tilde{\Lambda}_n}\left(\varphi;h,\theta^\star\right)$ converges towards 0.*

*(b) Under **[C]** and **[E1]**: for $P_{\theta^\star}-$a.e. φ, $|\tilde{\Lambda}_n|^{-1} R_{\tilde{\Lambda}_n}\left(\varphi;h,\widehat{\theta}_n(\varphi)\right)$ converges towards 0.*

Remark 1. *Assumption **[Mod-L]**, while useful to allow the computation of the residuals in practice, is actually useless to prove their consistency.*

3.2 Asymptotic control in probability of the residuals process

We provide in this section a control for the departure of the residuals from the innovations and $(\widehat{\theta}_n-\theta^\star)$. This is a crucial result to investigate the asymptotic normality of the residuals. We need the folllowing assumptions.

[**N1**] For all $(m,\varphi)\in\mathbb{M}\times\Omega$, the functions $h\left(0^m,\varphi;\theta\right)$ and $f\left(0^m,\varphi;\theta\right)$ (defined in [**C1**]) are twice continuously differentiable with respect to θ in a neighborhood $\mathcal{V}(\theta^\star)$ of $\theta^\star$ and

$$\mathbf{E}\left(\left\|\underline{\mathbf{f}}^{(2)}\left(0^M,\Phi;\theta^\star\right)\right\|\right) < +\infty \quad \text{and} \quad \mathbf{E}\left(\left\|\underline{\mathbf{h}}^{(2)}\left(0^M,\Phi;\theta^\star\right)\right\| e^{-V\left(0^M|\Phi;\theta^\star\right)}\right) < +\infty,$$

where $\underline{\mathbf{g}}^{(2)}\left(0^m,\varphi;\theta^\star\right) = \left(\frac{\partial^2}{\partial\theta_j\partial\theta_k} g\left(0^m,\varphi;\theta^\star\right)\right)_{1\le j,k\le p}$ for $g=f,h$.

[**E2**] There exists a random vector $\mathbf{T}$ such that the following convergence holds as $n\to+\infty$

$$|\Lambda_n|^{1/2}\left(\widehat{\theta}_n(\Phi)-\theta^\star\right) \xrightarrow{d} \mathbf{T}.$$

Proposition 3. *Under assumptions **[C]**, **[N1]** and **[E1-2]**, assuming that $|\tilde{\Lambda}_n| = \mathcal{O}(|\Lambda_n|)$, then as $n\to+\infty$,*

$$R_{\tilde{\Lambda}_n}\left(\Phi;h,\widehat{\theta}_n(\Phi)\right) = I_{\tilde{\Lambda}_n}\left(\Phi;h,\theta^\star\right) - |\tilde{\Lambda}_n|\left(\widehat{\theta}_n(\Phi)-\theta^\star\right)^T \mathcal{E}\left(h;\theta^\star\right) + o_P(|\tilde{\Lambda}_n|^{1/2}), \tag{6}$$

where $\mathcal{E}\left(h;\theta^\star\right)$ is the vector defined by

$$\mathcal{E}\left(h;\theta^\star\right) := \mathbf{E}\left(h\left(0^M,\Phi;\theta^\star\right)\mathbf{V}^{(1)}\left(0^M|\Phi;\theta^\star\right) e^{-V\left(0^M|\Phi;\theta^\star\right)}\right). \tag{7}$$

The notation $X_n(\Phi) = o_P(w_n)$ means that $w_n^{-1}X_n(\Phi)$ converges in probability towards 0 as n tends to infinity.

Remark 2. *Note that for exponential family models, $\mathbf{V}^{(1)}(x^m|\varphi;\theta^\star)$ corresponds to the vector of sufficient statistics (see Section 6 for more details).*

3.3 Assumptions required for the asymptotic normality results

Apart from the assumptions [**Mod**], [**C**] and [**N1**] on the model, we will need to assume [**N2-4**] below. All these assumptions are fulfilled by many models as proved in Section 6.

[**N2**] For any bounded domain Λ, for any $\theta \in \mathcal{V}(\theta^\star)$,

$$\mathbf{E}\left(\left|I_\Lambda\left(\Phi;h,\theta^\star\right)\right|^3\right) < +\infty.$$

[**N3**] For any sequence of bounded domains Γ_n such that $\Gamma_n \to 0$ when $n \to \infty$, for any $\theta \in \mathcal{V}(\theta^\star)$,

$$\mathbf{E}\left(I_{\Gamma_n}\left(\Phi;h,\theta\right)^2\right) \longrightarrow 0.$$

[**N4**] For any $\varphi \in \Omega$ and any bounded domain Λ, $I_\Lambda\left(\varphi;\theta\right)$ depends only on $\varphi_{\Lambda\oplus D}$.

Concerning the properties required for the estimator $\widehat{\theta}_n$, we need its consistency through [**E1**] and to refine [**E2**] into [**E2(bis)**] below. Note that the maximum pseudolikelihood estimator satisfies these assumptions for many models (see section 6.2).

[**E2(bis)**] The estimate admits the following expansion

$$\widehat{\theta}_n(\Phi) - \theta^\star = \frac{1}{|\Lambda_n|}\mathbf{U}_{\Lambda_n}\left(\Phi;\theta^\star\right) + o_P(|\Lambda_n|^{-1/2}),$$

where, for any $\theta \in \mathcal{V}(\theta^\star)$,

(i) for any $\varphi \in \Omega$ and for two disjoint bounded domains Λ_1, Λ_2,

$$\mathbf{U}_{\Lambda_1\cup\Lambda_2}\left(\varphi;\theta\right) = \mathbf{U}_{\Lambda_1}\left(\varphi;\theta\right) + \mathbf{U}_{\Lambda_2}\left(\varphi;\theta\right),$$

(ii) for all $j = 1,\dots,p$ and any bounded domain Λ

$$\mathbf{E}\left(\left|\left(\mathbf{U}_\Lambda\left(\Phi;\theta\right)\right)_j\right|^3\right) < +\infty,$$

(iii) for all $j = 1,\dots,p$ and for any bounded domain Λ

$$\mathbf{E}\left(\left(\mathbf{U}_\Lambda\left(\Phi;\theta\right)\right)_j \middle| \Phi_{\Lambda^c}\right) = 0,$$

(iv) for all $j = 1,\dots,p$ and for any sequence of bounded domains Γ_n,

$$\mathbf{E}\left(\left(\mathbf{U}_{\Gamma_n}\left(\Phi;\theta\right)\right)_j^2\right) \longrightarrow 0 \quad \text{as} \quad \Gamma_n \to 0,$$

(v) for any $\varphi \in \Omega$ and any bounded domain Λ, $\mathbf{U}_\Lambda\left(\varphi;\theta\right)$ depends only on $\varphi_{\Lambda\oplus D}$.

Remark 3. *Assumption **[E2(bis)]** implies **[E2]**. Indeed, under this assumption one may apply Theorem 2.1 of Jensen and Künsch (1994) and assert: there exists a matrix $\underline{\mathbf{\Sigma}}$ such that $|\Lambda_n|^{-1/2}\mathbf{U}_{\Lambda_n}\left(\Phi;\theta^\star\right) \xrightarrow{d} \mathcal{N}(0,\underline{\mathbf{\Sigma}})$, as $n \to +\infty$.*

3.4 Asymptotic normality of the $h-$residuals computed on subdomains of Λ_n

In this framework, we give ourself a test function h and we compute the $h-$residuals on disjoint subdomains of Λ_n. In this context, we assume that the domain Λ_n is a cube and is divided into a fixed number of subdomains as follows

$$\Lambda_n := \bigcup_{j\in\mathcal{J}} \Lambda_{j,n}$$

where $\mathcal{J}$ is a finite set and all the $\Lambda_{j,n}$ are disjoint cubes with the same volume $|\Lambda_{0,n}|$ increasing up to $+\infty$. Let us denote by $\mathbf{R}_{\mathcal{J},n}\left(\varphi;h,\widehat{\theta}_n\right)$ the vector of the residuals computed on each subdomain, i.e. $\mathbf{R}_{\mathcal{J},n}\left(\varphi;h,\widehat{\theta}_n\right) = \left(R_{\Lambda_{j,n}}\left(\varphi;h,\widehat{\theta}_n\right)\right)_{j\in\mathcal{J}}$.

According to Proposition 3 and in view of **[E2(bis)]**, we introduce the following notation

$$R_{\infty,\Lambda}(\varphi;h,\theta) := I_\Lambda(\varphi;h,\theta) - \mathbf{U}_\Lambda\left(\varphi;\theta\right)^T \mathcal{E}\left(h;\theta\right) \tag{8}$$

for any $\varphi\in\Omega$, for any bounded domain Λ and for any $\theta\in\mathbf{\Theta}$.

Proposition 4. *Assume that*

- *The parametric model satisfies **[Mod]**.*
- *The energy function and the test function h satisfy **[C]** and **[N1-4]**.*
- *The energy function and the estimate $\widehat{\theta}_n$ satisfy **[E1]** and **[E2(bis)]**.*

Then, the following convergence in distribution holds, as $n\to+\infty$

$$|\Lambda_{0,n}|^{-1/2}\,\mathbf{R}_{\mathcal{J},n}\left(\Phi;h,\widehat{\theta}_n\right) \xrightarrow{d} \mathcal{N}\left(0,\underline{\mathbf{\Sigma}}_1(\theta^\star)\right), \tag{9}$$

where $\underline{\mathbf{\Sigma}}_1(\theta^\star) = \lambda_{Inn}\,\underline{\mathbf{I}}_{|\mathcal{J}|} + |\mathcal{J}|^{-1}(\lambda_{Res}-\lambda_{Inn})\,\underline{\mathbf{J}}$ with $\underline{\mathbf{J}} = \mathbf{e}\mathbf{e}^T$ and $\mathbf{e} = (1,\ldots,1)^T$. The constants λ_{Inn} and λ_{Res} are respectively defined by

$$\lambda_{Inn} = D^{-d}\sum_{|k|\le 1}\mathbf{E}\left(I_{\Delta_0(D)}\left(\Phi;h,\theta^\star\right) I_{\Delta_k(D)}\left(\Phi;h,\theta^\star\right)\right), \tag{10}$$

$$\lambda_{Res} = D^{-d}\sum_{|k|\le 1}\mathbf{E}\left(R_{\infty,\Delta_0(D)}(\Phi;h,\theta^\star)R_{\infty,\Delta_k(D)}(\Phi;h,\theta^\star)\right), \tag{11}$$

where, for all $k\in\mathbb{Z}^d$, $\Delta_k(D)$ is the cube centered at kD with side-length D.

From this asymptotic normality result, we can deduce the convergence for the norm of the centered residuals. This is the basis for a generalization of the quadrat counting test discussed in Section 5. We denote by $\overline{\mathbf{R}}_{\mathcal{J},n}(\varphi;h)$ the mean residuals over all subdomains, that is $\overline{\mathbf{R}}_{\mathcal{J},n}(\varphi;h) = |\mathcal{J}|^{-1}\sum_{j\in\mathcal{J}} R_{\Lambda_{j,n}}\left(\varphi;h,\widehat{\theta}_n\right)$.

Corollary 5. *Under the assumptions of Proposition 4,*

$$|\Lambda_{0,n}|^{-1}\|\mathbf{R}_{\mathcal{J},n}(\Phi;h) - \overline{\mathbf{R}}_{\mathcal{J},n}(\Phi;h)\|^2 \xrightarrow{d} \lambda_{Inn}\;\chi^2_{|\mathcal{J}|-1}. \tag{12}$$

Proof. An easy computation shows that λ_{Inn} and λ_{Res} are the two eigenvalues of $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ with respective order $|\mathcal{J}|-1$ and 1. Let $\underline{\mathbf{P}}_{Inn}$ be the matrix of orthonormalized eigenvectors associated to λ_{Inn}. This matrix of size $(|\mathcal{J}|,|\mathcal{J}|-1)$ satisfies by definition $\underline{\mathbf{P}}^T_{Inn}\underline{\mathbf{P}}_{Inn} = \underline{\mathbf{I}}_{|\mathcal{J}|-1}$ and, from (9), $|\Lambda_{0,n}|^{-1}\;\|\underline{\mathbf{P}}^T_{Inn}\mathbf{R}_{\mathcal{J},n}(\varphi;h)\|^2 \xrightarrow{d} \lambda_{Inn}\;\chi^2_{|\mathcal{J}|-1}$. Moreover, it is easy to check that $\underline{\mathbf{P}}_{Inn}\underline{\mathbf{P}}^T_{Inn} = \underline{\mathbf{I}}_{|\mathcal{J}|} - |\mathcal{J}|^{-1}\underline{\mathbf{J}}_{|\mathcal{J}|}$ which leads to $\|\underline{\mathbf{P}}^T_{Inn}\mathbf{R}_{\mathcal{J},n}(\varphi;h)\|^2 = \|\mathbf{R}_{\mathcal{J},n}(\varphi;h) - \overline{\mathbf{R}}_{\mathcal{J},n}(\varphi;h)\|^2$. ■

Remark 4. *The asymptotic covariance matrix $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ and λ_{Inn} involve only the covariance structure of the innovations (or the residuals) in a finite box around 0. This comes from the locality assumption **[Mod-L]**, also involved in **[N4]** and **[E2(bis)]**. A challenging task in practice is to estimate λ_{Inn} and λ_{Res} (and so $\underline{\mathbf{\Sigma}}_1(\theta^\star)$), this issue is investigated in Section 4.*

3.5 Asymptotic normality of the $(h_j)_{j=1,\dots,s}-$residuals computed on Λ_n

In this framework, we consider s different test functions and we compute all h_j-residuals on the same domain Λ_n, which is assumed to be a cube growing up to $\mathbb{R}^d$ when $n \to +\infty$.

We present an asymptotic normality result for the random vector $\left(R_{\Lambda_n}\left(\Phi; h_j, \widehat{\theta}_n\right)\right)_{j=1,\dots,s}$.

Proposition 6. *Assume that*

- *The parametric model satisfies **[Mod]**.*
- *The energy function and the test functions h_j (for $j = 1, \dots, s$) satisfy **[C]** and **[N1-4]**.*
- *The energy function and the estimate $\widehat{\theta}_n$ satisfy **[E1]** and **[E2(bis)]**.*

Then, the following convergence in distribution holds, as $n \to +\infty$

$$|\Lambda_n|^{-1/2}\left(R_{\Lambda_n}\left(\Phi; h_j, \widehat{\theta}_n\right)\right)_{j=1,\dots,s} \xrightarrow{d} \mathcal{N}\left(0, \underline{\mathbf{\Sigma}}_2(\theta^\star)\right), \tag{13}$$

where $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ is the (s,s) matrix given by

$$\underline{\mathbf{\Sigma}}_2(\theta^\star) = D^{-d}\sum_{|k|\le 1}\mathbf{E}\left(\mathbf{R}_{\infty,\Delta_0(D)}(\Phi; \mathbf{h}, \theta^\star)\,\mathbf{R}_{\infty,\Delta_k(D)}(\Phi; \mathbf{h}, \theta^\star)^T\right), \tag{14}$$

where $\mathbf{R}_{\infty,\Lambda}(\varphi, \mathbf{h}, \theta^\star) := \left(R_{\infty,\Lambda}(\varphi; h_j, \theta^\star)\right)_{j=1,\dots,s}$, see (8), and where, for all $k \in \mathbb{Z}^d$, $\Delta_k(D)$ is the cube centered at kD with side-length D.

4 Estimation and positivity of the asymptotic covariance matrices

4.1 Statement of the problem

The aim of this section is to provide a condition under which, on the one hand the matrices $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ and $\underline{\mathbf{\Sigma}}_2(\theta^\star)$, defined in Propositions 4 and 6, are positive-definite, and on the other hand λ_{Inn}, involved in Corollary 5, is positive. Then we define estimators of $\underline{\mathbf{\Sigma}}_1^{-1/2}(\theta^\star)$, λ_{Inn}^{-1} and $\underline{\mathbf{\Sigma}}_2^{-1/2}(\theta^\star)$. As a consequence, we will be in position to normalize and estimate the quantities arising in (9), (12) and (13) so that they converge to a free law.

Before this, let us focus on the particular form of the matrix $\underline{\mathbf{\Sigma}}_1(\theta^\star)$. This $(|\mathcal{J}|, |\mathcal{J}|)$ matrix has two eigenvalues λ_{Inn} and λ_{Res} (respectively defined by (10) and (11)), whose multiplicity is $|\mathcal{J}|-1$ for λ_{Inn} and 1 for λ_{Res}. By using the Gram-Schmidt process for orthonormalizing the eigenvectors of $\underline{\mathbf{\Sigma}}_1(\theta^\star)$, one obtains the explicit form for the squared inverse of this matrix, provided λ_{Inn} and λ_{Res} do not vanish:

$$\underline{\mathbf{\Sigma}}_1^{-1/2}(\theta^\star) = \frac{1}{\sqrt{\lambda_{Inn}}}\,\underline{\mathbf{I}}_{|\mathcal{J}|} + \frac{1}{|\mathcal{J}|}\left(\frac{1}{\sqrt{\lambda_{Res}}} - \frac{1}{\sqrt{\lambda_{Inn}}}\right)\underline{\mathbf{J}},$$

where $\underline{\mathbf{J}} = \mathbf{e}\mathbf{e}^T$ and $\mathbf{e} = (1, \dots, 1)^T$. Therefore, estimating $\underline{\mathbf{\Sigma}}_1^{-1/2}(\theta^\star)$ can be reduced to the estimation of these two eigenvalues λ_{Inn} and λ_{Res}.

Consequently, the estimation of λ_{Inn} and the covariance matrices $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ and $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ is achieved by estimating (10), (11) and (14), which can be viewed as a particular case of estimating the matrix (actually a constant for the two first expressions)

$$\underline{\mathbf{M}}(\theta^\star) = D^{-d} \sum_{|k|\leq 1} \mathbf{E}\left(\mathbf{Y}_{\Delta_0(D)}(\Phi;\theta^\star)\,\mathbf{Y}_{\Delta_k(D)}(\Phi;\theta^\star)^T\right),$$

where, according to the assumptions involved in Propositions 4 and 6, for any bounded domain Λ, $\mathbf{Y}_\Lambda(\Phi;\theta)$ is a random vector of dimension q ($q=1$ or s) depending on θ, such that for any bounded domains $\Lambda, \Lambda_1, \Lambda_2$ (Λ_1, Λ_2 disjoint), for any $\theta \in \mathcal{V}(\theta^\star)$, for any $j = 1, \ldots, q$ and any $\varphi \in \Omega$

(i) $\mathbf{Y}_{\Lambda_1\cup\Lambda_2}(\varphi;\theta) = \mathbf{Y}_{\Lambda_1}(\varphi;\theta) + \mathbf{Y}_{\Lambda_2}(\varphi;\theta)$,

(ii) $\mathbf{E}\left((\mathbf{Y}_\Lambda(\Phi;\theta))_j^2\right) < +\infty$,

(iii) $\mathbf{E}\left((\mathbf{Y}_\Lambda(\Phi;\theta))_j \,\middle|\, \Phi_{\Lambda^c}\right) = 0$,

(iv) for any sequence of bounded domains Γ_n, $\mathbf{E}\left((\mathbf{Y}_{\Gamma_n}(\Phi;\theta))_j^2\right) \longrightarrow 0 \quad \text{as} \quad \Gamma_n \to 0$,

(v) $\mathbf{Y}_\Lambda(\varphi;\theta)$ depends only on $\varphi_{\Lambda\oplus D}$.

4.2 Positive definiteness of $\underline{\mathbf{M}}(\theta^\star)$

Let us consider the following assumption.

[PD] For some $\overline{\Lambda} := \cup_{|i|\leq\lceil\frac{D}{\overline{\delta}}\rceil}\Delta_i(\overline{\delta})$ with $\overline{\delta} > 0$, there exists $B \in \mathcal{F}$ and $A_0, \ldots, A_\ell$, ($\ell \geq 1$) disjoint events of $\overline{\Omega}_B := \left\{\varphi \in \Omega : \varphi_{\Delta_i(\overline{\delta})} \in B, 1 \leq |i| \leq 2\left\lceil\frac{D}{\overline{\delta}}\right\rceil\right\}$ such that

- for $j = 0, \ldots, \ell$, $P_{\theta^\star}(A_j) > 0$.
- for all $(\varphi_0, \ldots, \varphi_\ell) \in A_0 \times \cdots \times A_\ell$ the (ℓ, q) matrix with entries $\left(\mathbf{Y}_{\overline{\Lambda}}(\varphi_i;\theta^\star)\right)_j - \left(\mathbf{Y}_{\overline{\Lambda}}(\varphi_0;\theta^\star)\right)_j$ is injective, which means:

$$\left(\forall \mathbf{y} \in \mathbb{R}^q, \mathbf{y}^T\left(\mathbf{Y}_{\overline{\Lambda}}(\varphi_i;\theta^\star) - \mathbf{Y}_{\overline{\Lambda}}(\varphi_0;\theta^\star)\right) = 0\right) \Longrightarrow \mathbf{y} = 0.$$

Proposition 7. *From the definition of $\mathbf{Y}_\Lambda(\Phi;\theta)$ and under **[PD]**, the matrix $\underline{\mathbf{M}}(\theta^\star)$ is positive-definite.*

Remark 5. *The assumption **[PD]** is associated to some characteristics of the point process Φ. The parameter $\overline{\delta}$ is independent of the parameters involved in the different estimators (e.g. $D^\vee$ or δ arising in the next section). Given a model, the event B and $\overline{\delta}$ are chosen in order to let the different configurations sets $A_0, A_1, \ldots, A_\ell$ as simple as possible. For most models, a convenient choice is $B = \emptyset$ and $\overline{\delta} \geq D$ (see the examples treated in Appendix B for instance).*

4.3 Estimation of $\underline{\mathbf{M}}(\theta^\star)$

The dependence of $\underline{\mathbf{M}}(\theta^\star)$ on D may be lightened thanks to the following lemma, whose proof is relegated to section 8.5.

Lemma 8. *The matrix $\underline{\mathbf{M}}(\theta^\star)$ can be rewritten for any $\delta > 0$ and any $D^\vee \geq D$ as*

$$\underline{\mathbf{M}}(\theta^\star) = \delta^{-d} \sum_{|k|\leq\lceil\frac{D^\vee}{\delta}\rceil} \mathbf{E}\left(\mathbf{Y}_{\Delta_0(\delta)}(\Phi;\theta^\star)\,\mathbf{Y}_{\Delta_k(\delta)}(\Phi;\theta^\star)^T\right),$$

where $\Delta_k(\delta)$ is the cube with side-length δ centered at $k\delta$.

From this result, to achieve an estimation of $\underline{\mathbf{M}}(\theta^\star)$, it is required to estimate the involved expectation and $\theta^\star$ (by $\widehat{\theta}_n$). This is enough for the estimation of λ_{Inn} for which $\mathbf{Y}_\Lambda(\varphi;\theta) = I_\Lambda(\varphi;\theta)$. But when $\mathbf{Y}_\Lambda(\varphi;\theta) = R_{\infty,\Lambda}(\varphi;h,\theta)$ or $\mathbf{Y}_\Lambda(\varphi;\theta) = \mathbf{R}_{\infty,\Lambda}(\varphi;\mathbf{h},\theta)$, which appears in $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ and $\underline{\mathbf{\Sigma}}_2(\theta^\star)$, it can be noticed that $\mathbf{Y}_\Lambda$ still depends on an expectation with respect to $P_{\theta^\star}$, through the vector $\mathcal{E}(h,\theta^\star)$ defined by (7). Moreover, the vector $\mathbf{U}_\Lambda$ in **[E2(bis)]** may also depend on such a term (this is the case for example when considering the maximum pseudolikelihood estimate as shown in Section 6.2). This means that $\mathbf{Y}_\Lambda(\varphi;\theta)$ cannot be estimated only by $\mathbf{Y}_\Lambda(\varphi;\widehat{\theta}_n)$, but by $\widehat{\mathbf{Y}}_{n,\Lambda}(\varphi;\widehat{\theta}_n)$, where $\widehat{\mathbf{Y}}_n$ is an estimator of $\mathbf{Y}$. We assume in the sequel that $\widehat{\mathbf{Y}}_n$ satisfies the same properties $(i)-(v)$ as $\mathbf{Y}$ and is a good estimator of $\mathbf{Y}$ (see Proposition 9). The explicit form of $\widehat{\mathbf{Y}}_n$ depends strongly on the estimate $\widehat{\theta}_n$ (e.g. through $\mathbf{U}_\Lambda$ in **[E2(bis)]**). When $\widehat{\theta}_n$ is the maximum pseudolikelihood estimator, we provide explicit formulas for $\widehat{\mathbf{Y}}_n$ in Section 6.3.

Let us now specify an estimator of $\underline{\mathbf{M}}(\theta^\star)$. Assume that the point process is observed in the domain $\Lambda_{n_0} \oplus D^+$ where $D^+ \geq D$ and Λ_{n_0} is a cube. For any δ such that $|\Lambda_{n_0}|\delta^{-d} \in \mathbb{N}$, we may consider the decomposition $\Lambda_{n_0} = \cup_{k\in\mathcal{K}_{n_0}} \Delta_k(\delta)$, where the $\Delta_k(\delta)$'s are disjoint cubes with side-length δ and centered, without loss of generality, at $k\delta$. For any such δ, according to Lemma 8, a natural estimator of $\underline{\mathbf{M}}(\theta^\star)$ is, for any $D^\vee \geq D$,

$$\widehat{\underline{\mathbf{M}}}_{n_0}(\varphi;\widehat{\theta}_{n_0}(\varphi),\delta,D^\vee) = |\Lambda_{n_0}|^{-1} \sum_{k\in\mathcal{K}_{n_0}} \sum_{j\in\mathbb{B}_k\left(\lceil \frac{D^\vee}{\delta} \rceil\right)\cap\mathcal{K}_{n_0}} \widehat{\mathbf{Y}}_{n_0,\Delta_j(\delta)}\left(\varphi;\widehat{\theta}_{n_0}(\varphi)\right) \widehat{\mathbf{Y}}_{n_0,\Delta_k(\delta)}\left(\varphi;\widehat{\theta}_{n_0}(\varphi)\right)^T. \tag{15}$$

Remark 6. *As suggested by Lemma 8, the parameter δ in (15) may be chosen arbitrarily. Yet, while $\underline{\mathbf{M}}(\theta^\star)$ is actually independent of δ, its estimate $\widehat{\underline{\mathbf{M}}}_{n_0}$ may depend on it due to edge effects.*

The following proposition provides a framework to study the asymptotic properties of (15) and shows the consistency of $\widehat{\underline{\mathbf{M}}}_{n_0}$ when the domain Λ_n increases up to ∞ as $n \to \infty$. Its proof is relegated to section 8.7.

Proposition 9. *Under **[Mod]**, **[E1]**, assume that for any θ in a neighborhood $\mathcal{V}(\theta^\star)$ of $\theta^\star$, for any bounded domain Λ, for any $\varphi \in \Omega$ and for $j = 1,\dots,p$, $\left(\widehat{\mathbf{Y}}_{n,\Lambda}(\varphi;\cdot)\right)_j$ is a continuous function. Assume moreover that*

$$\sup_{k\in\mathcal{K}_n} \left|\widehat{\mathbf{Y}}_{n,\Delta_k(\delta_n)}(\Phi;\theta) - \mathbf{Y}_{\Delta_k(\delta_n)}(\Phi;\theta)\right| \xrightarrow{\mathbb{P}} 0, \tag{16}$$

where, for any $\delta > 0$ as above, $(\delta_n)_{n\in\mathbb{N}}$ is a sequence satisfying $|\Lambda_n|\delta_n^{-d} \in \mathbb{N}$, $\delta_{n_0} = \delta$ and $\delta_n \to \delta$ as $n \to \infty$. Then, for any $D^\vee \geq D$,

$$\widehat{\underline{\mathbf{M}}}_n\left(\Phi;\widehat{\theta}_n(\Phi),\delta_n,D^\vee\right) \xrightarrow{\mathbb{P}} \underline{\mathbf{M}}(\theta^\star).$$

Remark 7. *The choice of the sequence $(\delta_n)_{n\in\mathbb{N}}$ is always possible (see the proof). Since we allow the domain Λ_n to grow continuously up to $\mathbb{R}^d$, its decomposition in sub-cubes with side-length δ is not always possible. The sequence $(\delta_n)_{n\in\mathbb{N}}$ is thus mandatory to make a decomposition of the domain available when n increases. We chose it by respecting as most as possible the initial choice of the practicioner.*

5 Goodness-of-fit tests for stationary marked Gibbs point processes

We present in this section three goodness-of-fit tests, based on the residuals computed according to the different frameworks considered in Section 3. We assume that the point process is observed in the domain $\Lambda_{n_0} \oplus D^+$ where $D^+ \geq D$ and Λ_{n_0} is a cube.

5.1 Quadrat-type test with $|J|-1$ degrees of freedom

According to the setting detailed in Section 3.4, we divide the domain Λ_{n_0} into a fixed number of subdomains, namely $\Lambda_{n_0} := \bigcup_{j\in\mathcal{J}} \Lambda_{j,n_0}$ where $\mathcal{J}$ is a finite set and all the Λ_{j,n_0} are disjoint cubes with the same volume $|\Lambda_{0,n_0}|$. Moreover, in each sub-domain, we consider the decomposition $\Lambda_{j,n_0} = \cup_{k\in\mathcal{K}_{j,n_0}} \Delta_k(\delta)$, for any δ such that $|\Lambda_{0,n_0}|\delta^{-d} \in \mathbb{N}$, where the $\Delta_k(\delta)$'s are disjoint cubes with side-length δ.

Following (15), we consider, for any $\delta > 0$ as above and any $D^\vee \geq D$, the estimator

$$\widehat{\lambda}_{n_0,Inn} = |\Lambda_{n_0}|^{-1} \sum_{i\in\mathcal{K}_{n_0}} \sum_{j\in\mathbb{B}_i\left(\lceil \frac{D^\vee}{\delta} \rceil\right)\cap\mathcal{K}_{n_0}} I_{\Delta_i(\delta)}\left(\varphi;\widehat{\theta}_{n_0}(\varphi)\right) I_{\Delta_j(\delta)}\left(\varphi;\widehat{\theta}_{n_0}(\varphi)\right), \tag{17}$$

where $\mathcal{K}_{n_0} = \cup_{j\in\mathcal{J}}\mathcal{K}_{j,n_0}$. Note that $I_{\Delta_i(\delta)}\left(\varphi;\widehat{\theta}_{n_0}(\varphi)\right) = R_{\Delta_i(\delta)}\left(\varphi;\widehat{\theta}_{n_0}(\varphi)\right)$ but we preserve this redundant notation in the sequel.

The following corollary is an immediate consequence of Corollary 5 and Proposition 9.

Corollary 10. *Under the assumptions of Proposition 4 and if [PD] holds for* $\mathbf{Y}_{\overline{\Lambda}}(\Phi;\theta^\star) = I_{\overline{\Lambda}}(\Phi;\theta^\star)$*, then, for any* $\delta > 0$*, one can construct a sequence* $(\delta_n)_{n\in\mathbb{N}}$ *satisfying* $|\Lambda_{0,n}|\delta_n^{-d} \in \mathbb{N}$*,* $\delta_{n_0} = \delta$ *and* $\delta_n \to \delta$*, such that as* $n \to +\infty$

$$T_{1,n} := |\Lambda_{0,n}|^{-1}\, \widehat{\lambda}_{n,Inn}^{-1} \times \|\mathbf{R}_{\mathcal{J},n}(\Phi;h) - \overline{\mathbf{R}}_{\mathcal{J},n}(\Phi;h)\|^2 \xrightarrow{d} \chi^2(|\mathcal{J}|-1). \tag{18}$$

This result leads to a goodness-of-fit test for $H_0 : \Phi \sim P_{\theta^\star}$ versus $H_1 : \Phi \nsim P_{\theta^\star}$. Let us briefly summarize the different steps to implement the test for a given asymptotic level $\alpha \in (0,1)$.

- **Step 1** Consider a parametric model of a stationary marked Gibbs point process with finite range D observed on the domain $\Lambda_{n_0} \oplus D^+$ with $D^+ \geq D$.
- **Step 2** Choose an estimation method satisfying the assumptions **[E1]**, **[E2(bis)]** (for example the MPLE) and compute the estimate $\widehat{\theta}_{n_0}$ on Λ_{n_0}.
- **Step 3**
  a) Consider a test function h (satisfying **[C1-2]**, **[N1-3]** and **[PD]**), divide Λ_{n_0} into $|\mathcal{J}|$ cubes and compute the $h-$residuals on each different cube.
  b) Estimate λ_{Inn} by (17).
  c) Compute the test statistic T_{1,n_0} involved in (18).
- **Step 4** Reject the model if $T_{1,n_0}(\varphi) > \chi^2_{1-\alpha}(|\mathcal{J}|-1)$.

Let us note that in the particular case of a homogeneous Poisson point process with intensity z and when considering the raw residuals ($h = 1$), this test is exactly the Poisson dispersion test applied to the $|\mathcal{J}|$ quadrat counts, also called quadrat counting test, see Diggle (2003) for instance. Indeed, in this case, $\mathbf{R}_{\mathcal{J},n}(\varphi;h) - \overline{\mathbf{R}}_{\mathcal{J},n}(\varphi;h)$ is the vector of quadrat counts and $\lambda_{Inn} = z$. Considering $|\Lambda_{0,n}|\widehat{\lambda}_{n_0,Inn}$ as an estimation of the intensity on $\Lambda_{0,n}$, the statistic $T_{1,n}$ reduces to the ratio of the sum of squares of the quadrat counts over their estimated mean.

Remark 8. *The condition [PD] in Corollary 10 has to be verified with* $\mathbf{Y}_{\overline{\Lambda}}(\Phi;\theta^\star) = I_{\overline{\Lambda}}(\Phi;\theta^\star)$ *which is not so difficult (see Proposition 16 for a general result). Indeed, contrarily to Corollary 11 and 12, this condition does not depend on the form of the estimator* $\widehat{\theta}_n$*. Moreover, as emphasized in Section 3.4, the assumptions of Proposition 4 are satisfied for many models (this will be explored in details for exponential models in Section 6.1). This means that the proposed goodness-of-fit test based on (18) may be used for many models and many choices of function* h*.*

Remark 9. *The weakeness of this testing procedure (and the next ones) could be the estimation (17) of λ_{Inn} (and in general the estimator (15)). The choice of the parameters δ and $D^{\vee}$ in (17) is crucial. For instance, for fixed n, in the extreme cases $\delta \to 0$ or $D^{\vee} \to \infty$, we get $\widehat{\lambda}_{n,Inn} \approx 0$. A careful simulation study should help for these choices. Another improvement could be to estimate λ_{Inn} via Monte-Carlo methods.*

5.2 Quadrat-type test with $|J|$ degrees of freedom

Under the same setting as above, assume moreover that **[PD]** holds for $\mathbf{Y}_{\overline{\Lambda}}(\varphi;\theta^\star) = R_{\infty,\overline{\Lambda}}(\varphi;h,\theta^\star)$. Let us define the normalized residuals

$$\widetilde{\mathbf{R}}_{1,n_0}(\varphi;h) := \widehat{\lambda}_{n_0,Inn}^{-1/2}\mathbf{R}_{\mathcal{J},n_0}(\varphi;h) + \left(\widehat{\lambda}_{n_0,Res}^{-1/2} - \widehat{\lambda}_{n_0,Inn}^{-1/2}\right)\overline{\mathbf{R}}_{\mathcal{J},n_0}(\varphi;h),$$

where $\widehat{\lambda}_{n_0,Inn}$ is defined in (17) and $\widehat{\lambda}_{n_0,Res}$ is an estimate of λ_{Res} following (15). When considering the MPLE, explicit formulas for $\widehat{\lambda}_{n_0,Res}$ are given in Section 6.3. It is easy to check that $\widetilde{\mathbf{R}}_{1,n_0}(\varphi;h) = \widehat{\underline{\mathbf{\Sigma}}}_{1n_0}^{-1/2}\,\mathbf{R}_{\mathcal{J},n_0}(\varphi;h)$. Therefore the following corollary is deduced from Propositions 4 and 9.

Corollary 11. *Under the assumptions of Propositions 4 and 9, assuming that **[PD]** holds for $\mathbf{Y}_{\overline{\Lambda}}(\Phi;\theta^\star) = I_{\overline{\Lambda}}(\Phi;\theta^\star)$ and $\mathbf{Y}_{\overline{\Lambda}}(\Phi;\theta^\star) = R_{\infty,\overline{\Lambda}}(\Phi;h,\theta^\star)$, then, for any $\delta > 0$, one can construct a sequence $(\delta_n)_{n\in\mathbb{N}}$ which satisfies $|\Lambda_{0,n}|\delta_n^{-d} \in \mathbb{N}$, $\delta_{n_0} = \delta$ and $\delta_n \to \delta$ as $n \to \infty$, such that as $n \to +\infty$,*

$$\widetilde{T}_{1,n}(\Phi) := |\Lambda_{0,n}|^{1/2}\|\widetilde{\mathbf{R}}_{1,n}(\Phi;h)\|^2 \xrightarrow{d} \chi^2(|\mathcal{J}|) \tag{19}$$

A goodness-of-fit test with asymptotic size $\alpha \in (0,1)$ is deduced similarly as in the previous section. The steps to follow in practice are the same except that in **Step 3** b), one has to estimate both λ_{Inn} and λ_{Res}, and in **Step 4** we reject the model if $\widetilde{T}_{1,n_0}(\varphi) > \chi^2_{1-\alpha}(|\mathcal{J}|)$.

Remark 10. *Let us emphasize that, with respect to Corollary 10, Corollary 11 involves an additional more complex assumption: **[PD]** has to be satisfied for $\mathbf{Y}_{\overline{\Lambda}}(\Phi;\theta^\star) = R_{\infty,\overline{\Lambda}}(\Phi;h,\theta^\star)$. This kind of assumption deeply depends on the nature of the estimate $\widehat{\theta}$. This problem is investigated in Proposition 18 for particular examples. Furthermore, we show in Proposition 17 that $\lambda_{Res} = 0$ occurs for many models and many choices of h including the Poisson model when $h = 1$. These two remarks underline the fact that the test relying on $\widetilde{T}_{1,n}$ is more restrictive than the previous one with $T_{1,n}$.*

5.3 Empty space function type test

Let us consider the setting of section 3.5, where s different residuals are computed on the same full domain Λ_{n_0}. We consider the decomposition $\Lambda_{n_0} = \cup_{k\in\mathcal{K}_{n_0}}\Delta_k(\delta)$, for any δ such that $|\Lambda_{n_0}|\delta^{-d} \in \mathbb{N}$, where the $\Delta_k(\delta)$'s are disjoint cubes with side-length δ.

Under the notation of Proposition 6, assuming **[PD]** holds for $\mathbf{Y}_{\overline{\Lambda}}(\varphi;\theta^\star) = \mathbf{R}_{\infty,\overline{\Lambda}}(\varphi,\mathbf{h},\theta^\star)$, let us define

$$\widetilde{\mathbf{R}}_{2,n_0}(\varphi;\mathbf{h},\widehat{\theta}) := \widehat{\underline{\mathbf{\Sigma}}}_{2n_0}^{-1/2}\,\left(R_{\Lambda_{n_0}}(\varphi;h_j,\widehat{\theta})\right)_{j=1,\dots,s}$$

where $\widehat{\underline{\mathbf{\Sigma}}}_{2n_0}^{-1/2} := \widehat{\underline{\mathbf{\Sigma}}}_{2n_0}^{-1/2}(\varphi,\widehat{\theta};\delta,D^{\vee})$ is an estimation of $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ as in (15). See explicit formulas in Section 6.3 when considering the MPLE.

From Propositions 6 and 9, we get the following corollary.

Corollary 12. *Assuming **[PD]** with $\mathbf{Y}_{\overline{\Lambda}}(\varphi;\theta^\star) = \mathbf{R}_{\infty,\overline{\Lambda}}(\varphi,\mathbf{h},\theta^\star)$, under the assumptions of Propositions 6 and 9, then, for any $\delta > 0$ as above, one can construct a sequence $(\delta_n)_{n\in\mathbb{N}}$ which satisfies $|\Lambda_n|\delta_n^{-d} \in \mathbb{N}$, $\delta_{n_0} = \delta$ and $\delta_n \to \delta$ as $n \to \infty$, such that, as $n \to +\infty$,*

$$\widetilde{T}_{2,n}(\Phi) := |\Lambda_n|^{1/2}\|\widetilde{\mathbf{R}}_{2,n}(\Phi;\mathbf{h},\widehat{\theta})\|^2 \xrightarrow{d} \chi^2(s). \tag{20}$$

A goodness-of-fit test for $H_0 : \Phi \sim P_{\theta^\star}$ versus $H_1 : \Phi \nsim P_{\theta^\star}$, with asymptotic size $\alpha \in (0,1)$ is deduced as before. From a practical point of view, the steps detailed in 5.1 are modified into:

- **Step 3(framework 2)**
  a) Consider s different test functions (satisfying **[C1-2]**, **[N1-3]** and **[PD]**), and compute the s different h_j-residuals on the same initial domain Λ_{n_0}.
  b) Estimate the matrix $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ by (15) and compute $\widehat{\underline{\mathbf{\Sigma}}_{2n_0}}^{-1/2}$ with any numerical routine (e.g. a choleski decomposition or a singular value decomposition).
  c) Compute the test statistic $\widetilde{T}_{2,n_0}(\varphi)$ defined by (20).
- **Step 4** Fix $\alpha \in (0,1)$ and reject the model if $\widetilde{T}_{2,n_0}(\varphi) > \chi^2_{1-\alpha}(s)$.

6 Application to exponential models and the MPLE

Through Sections 3, 4 and 5 three sets of assumptions have been considered. The first one deals with integrability and regularity of the model and the test function(s) and gathers **[Mod]**, **[C]** and **[N1-4]**. The second one is about the estimator $\widehat{\theta}_n$ and involves **[E1]** and **[E2(bis)]**. Finally, the third one, assumption **[PD]** is very specific and deals with the positive definiteness of covariance matrices. We prove in this section that these assumptions are in general fulfilled for exponential family models and the MPLE.

6.1 Assumptions [Mod], [C] and [N1-4] for exponential family models

The energy function of exponential family models is given for any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$ by $V_\Lambda(\varphi;\theta) = \theta^T \mathbf{v}_\Lambda(\varphi)$, where $\mathbf{v}_\Lambda(\varphi)$ is the vector of sufficient statistics given by $\mathbf{v}_\Lambda(\varphi) = (v_{1,\Lambda}(\varphi), \ldots, v_{p,\Lambda}(\varphi))^T$. The local energy is then expressed as $V\left(x^m|\varphi;\theta\right) = \theta^T \mathbf{v}\left(x^m|\varphi\right)$, where $\mathbf{v}\left(x^m|\varphi\right) = (v_1(x^m|\varphi), \ldots, v_p(x^m|\varphi)) := \mathbf{v}_\Lambda(\varphi \cup \{x^m\}) - \mathbf{v}_\Lambda(\varphi)$. Let us consider the following assumption:

[Exp] For $i = 1, \cdots, p$, there exist $\kappa_i^{(\inf)}, \kappa_i^{(\sup)} \geq 0$, $k_i \in \mathbb{N}$ such that one of both following assumptions is satisfied for all $(m,\varphi) \in \mathbb{M} \times \Omega$:

$$\theta_i \geq 0 \text{ and } -\kappa_i^{(\inf)} \leq v_i(0^m|\varphi) = v_i(0^m|\varphi_{\mathcal{B}(0,D)}) \leq \kappa_i^{(\sup)} |\varphi_{\mathcal{B}(0,D)}|^{k_i}.$$

or

$$-\kappa_i^{(\inf)} \leq v_i(0^m|\varphi) = v_i(0^m|\varphi_{\mathcal{B}(0,D)}) \leq \kappa_i^{(\sup)}.$$

The assumption **[Exp]** has already been considered by Billiot et al. (2008). It is fulfilled for a large class of examples including the overlap area point process, the multi-Strauss marked point process, the $k-$nearest-neighbor multi-Strauss marked point process, the Strauss type disc process, the Geyer's triplet point process, the area interaction point process,....

Proposition 13. *Under **[Exp]**, the assumptions **[Mod]**, **[C]** and **[N1-4]** are satisfied for the raw residuals, inverse residuals, Pearson residuals or residuals based on the empty space function.*

Proof. The assumption **[Exp]** implies that the local energy function is local and stable, which, from results of Bertin et al. (1999), implies that **[Mod]** is fulfilled. A direct consequence of **[Exp]** is that for every $\alpha > 0$, for all $\theta \in \mathbf{\Theta}$ and for all $i = 1, \ldots, p$

$$\mathbf{E}\left(|v_i(0^M|\Phi)|^\alpha e^{-\theta^T \mathbf{v}\left(0^M|\Phi\right)}\right) < +\infty, \tag{21}$$

which ensures the integrability assumptions **[C]** and **[N1-2]** for the residuals considered in the proposition. The locality assumption **[N4]** is contained in **[Exp]**. Finally, an application of the dominated convergence theorem, with the help of (21), shows **[N3]**. ■

Remark 11. *Our setting is not restricted to locally stable exponential family models. As an example, following ideas of Coeurjolly and Drouilhet (2009), one may prove that* ***[C]*** *and* ***[N1-4]*** *are fulfilled for Lennard-Jones type models.*

6.2 Assumptions [E1] and [E2(bis)] for the MPLE

Among the different parametric estimation methods available for spatial point patterns, the maximum pseudolikelihood is of particular interest. Indeed, unlike the maximum likelihood estimation method, it does not require the computation of the partition function, it is quite easy to implement and asymptotic results are now well-known (see Jensen and Møller (1991), Jensen and Künsch (1994), Billiot et al. (2008), Dereudre and Lavancier (2009) and Coeurjolly and Drouilhet (2009)). The MPLE is obtained by maximizing the log-pseudolikelihood contrast, given for exponential models by

$$LPL_{\Lambda_n}(\varphi;\theta) = -\int_{\Lambda_n\times\mathbb{M}} e^{-\theta^T \mathbf{v}(x^m|\varphi)}\mu(dx^m) \;-\; \theta^T \sum_{x\in\varphi}\mathbf{v}(x^m|\varphi\setminus x^m). \tag{22}$$

Proposition 14. *Under assumption* ***[Exp]*** *(and an additional indentifiability condition),* ***[E1]*** *and* ***[E2(bis)]*** *are fulfilled for the MPLE. The vector* $\mathbf{U}_\Lambda(\varphi;\theta^\star)$ *in* ***[E2(bis)]*** *is then expressed as follows*

$$\mathbf{U}_\Lambda(\varphi;\theta^\star) := \underline{\mathbf{H}}(\theta^\star)^{-1}\mathbf{LPL}^{(1)}_\Lambda(\varphi;\theta^\star), \tag{23}$$

where $\mathbf{LPL}^{(1)}_\Lambda(\varphi;\theta^\star)$ *is the gradient vector of the log-pseudolikelihood given by*

$$\mathbf{LPL}^{(1)}_\Lambda(\varphi;\theta^\star) := \int_{\Lambda\times\mathbb{M}} \mathbf{v}(x^m|\varphi;\theta^\star)e^{-\theta^{\star T}\mathbf{v}(x^m|\varphi)}\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda}\mathbf{v}(x^m|\varphi\setminus x^m;\theta^\star) \tag{24}$$

and where $\underline{\mathbf{H}}(\theta^\star)$ *is the symmetric matrix*

$$\underline{\mathbf{H}}(\theta^\star) := \mathbf{E}\left(\mathbf{v}(0^M|\Phi;\theta^\star)\mathbf{v}(0^M|\Phi;\theta^\star)^T\, e^{-V\left(0^M|\Phi;\theta^\star\right)}\right). \tag{25}$$

Proof. **[E1]** is proved by Billiot et al. (2008) (under **[Exp]** and the identifiability condition **[Ident]** arising p.244 in Billiot et al. (2008)). Let $\mathbf{Z}_n(\varphi;\theta^\star) := -|\Lambda_n|^{-1}\mathbf{LPL}_{\Lambda_n}(\varphi;\theta^\star)$. If $\widehat{\theta}_n(\varphi) = \widehat{\theta}_n^{MPLE}(\varphi)$ denotes the maximum pseudolikelihood estimate, one derives

$$\mathbf{Z}^{(1)}_n(\varphi;\widehat{\theta}_n) - \mathbf{Z}^{(1)}_n(\varphi;\theta^\star) = 0 - \mathbf{Z}^{(1)}_n(\varphi;\theta^\star) = \underline{\mathbf{H}}_n(\varphi;\theta^\star,\widehat{\theta}_n)(\widehat{\theta}_n(\varphi)-\theta^\star)$$

with $\underline{\mathbf{H}}_n(\varphi;\theta^\star,\widehat{\theta}_n) = \int_0^t \underline{\mathbf{Z}}^{(2)}_n\left(\varphi;\theta^\star + t(\widehat{\theta}_n(\varphi)-\theta^\star)\right)dt$. Under assumptions **[Exp]** and **[Ident]**, then, for n large enough, $\underline{\mathbf{H}}_n$ is invertible and converges almost surely towards the matrix $\underline{\mathbf{H}}(\theta^\star)$ given by (25). Moreover, following the proof of Theorem 2 of Billiot et al. (2008) (see condition (iii) p.257-258), we derive $Var(\mathbf{Z}^{(1)}_n(\Phi;\theta^\star)) = \mathcal{O}(|\Lambda_n|^{-1})$. So

$$\begin{aligned}|\Lambda_n|^{1/2}\Big((\widehat{\theta}_n(\Phi)-\theta^\star) + \underline{\mathbf{H}}^{-1}(\theta^\star)\mathbf{Z}^{(1)}_n(\Phi;\theta^\star)\Big) &= -|\Lambda_n|^{1/2}\left(\underline{\mathbf{H}}_n^{-1}(\Phi;\widehat{\theta}_n,\theta^\star) - \underline{\mathbf{H}}^{-1}(\theta^\star)\right)\mathbf{Z}^{(1)}_n(\Phi;\theta^\star)\\ &\to\; 0,\end{aligned}$$

in probability as $n\to+\infty$. This implies (23). Finally, $\mathbf{U}_\Lambda(\varphi;\theta^\star)$ fulfills properties $(i)-(v)$ in **[E2(bis)]** for the same reasons as in the proof of Proposition 13 and, for (iii), from the proof of Theorem 2 (step 1, p. 257) in Billiot et al. (2008). ■

Remark 12. *In the same spirit as Remark 11, let us underline that the MPLE also satisfies* ***[E1]*** *and* ***[E2(bis)]*** *for some non locally stable and non exponential family models, including Lennard-Jones type models (provided a locality assumption).*

6.3 Estimation of asymptotic covariance matrices when considering the MPLE

We still focus on exponential family models. As in Section 4.3, we assume that the point process is observed in the domain $\Lambda_{n_0} \oplus D^+$ where $D^+ \geq D$ and Λ_{n_0} is a cube. Moreover, we consider the decomposition $\Lambda_{n_0} = \cup_{k \in \mathcal{K}_{n_0}} \Delta_k(\delta)$, for any δ such that $|\Lambda_{n_0}|\delta^{-d} \in \mathbb{N}$, where the $\Delta_k(\delta)$'s are disjoint cubes with side-length δ and centered, without loss of generality, at $k\delta$.

From (8) and (23), we have under the assumptions **[Exp]** and when considering the MPLE

$$R_{\infty,\Lambda}(\varphi; h, \theta^\star) := I_\Lambda(\varphi; h, \theta^\star) - \mathbf{LPL}^{(1)}(\varphi; \theta^\star)^T \mathbf{W}(h, \theta^\star) \tag{26}$$

where $\mathbf{W}(h,\theta) := \underline{\mathbf{H}}(\theta)^{-1}\mathcal{E}(h,\theta)$. A natural estimator of $\mathbf{W}(h,\theta^\star)$ is given by $\widehat{\mathbf{W}}_{n_0}(\varphi; h, \widehat{\theta}_{n_0}) := \underline{\widehat{\mathbf{H}}}_{n_0}(\varphi; \widehat{\theta}_{n_0})^{-1}\widehat{\mathcal{E}}_{n_0}(\varphi; h, \widehat{\theta}_{n_0})$ where

$$\begin{aligned}
\underline{\widehat{\mathbf{H}}}_{n_0}(\varphi; \widehat{\theta}_{n_0}) &= |\Lambda_{n_0}|^{-1} \int_{\Lambda_{n_0}\times\mathbb{M}} \mathbf{v}(x^m|\varphi)\mathbf{v}(x^m|\varphi)^T e^{-\widehat{\theta}_{n_0}^T \mathbf{v}(x^m|\varphi)} \mu(dx^m), \\
\widehat{\mathcal{E}}_{n_0}(\varphi; h, \widehat{\theta}_{n_0}) &= |\Lambda_{n_0}|^{-1} \int_{\Lambda_{n_0}\times\mathbb{M}} h(x^m, \varphi; \widehat{\theta}_{n_0})\mathbf{v}(x^m|\varphi) e^{-\widehat{\theta}_{n_0}^T \mathbf{v}(x^m|\varphi)} \mu(dx^m).
\end{aligned}$$

In this spirit, let $\widehat{R}_{n_0,\infty,\Lambda}(\varphi; h, \widehat{\theta}_{n_0}) := I_\Lambda(\varphi; h, \widehat{\theta}_{n_0}) - \mathbf{LPL}_\Lambda^{(1)}(\varphi; \widehat{\theta}_{n_0})^T \widehat{\mathbf{W}}_{n_0}(\varphi; h, \widehat{\theta}_{n_0})$ and $\widehat{\mathbf{R}}_{n_0,\infty,\Lambda}(\varphi; \mathbf{h}, \widehat{\theta}_{n_0}) := \left(\widehat{R}_{n_0,\infty,\Lambda}(\varphi; h_j, \widehat{\theta}_{n_0})\right)_{j=1,\dots,s}$. Based on these notation, we obtain the following estimations for $\lambda_{Inn}, \lambda_{Res}$ and $\underline{\mathbf{\Sigma}}_2(\theta^\star)$

$$\begin{aligned}
\widehat{\lambda}_{n_0,Inn}(\varphi, \widehat{\theta}_{n_0}(\varphi), \delta, D^\vee) &= |\Lambda_{n_0}|^{-1} \sum_{i\in\mathcal{K}_{n_0}} \sum_{j\in\mathbb{B}_i\left(\lceil \frac{D^\vee}{\delta} \rceil\right)\cap\mathcal{K}_{n_0}} I_{\Delta_i(\delta)}\left(\varphi; \widehat{\theta}_{n_0}(\varphi)\right) I_{\Delta_j(\delta)}\left(\varphi; \widehat{\theta}_{n_0}(\varphi)\right), \\
\widehat{\lambda}_{n_0,Res}(\varphi, \widehat{\theta}_{n_0}(\varphi), \delta, D^\vee) &= |\Lambda_{n_0}|^{-1} \sum_{i\in\mathcal{K}_{n_0}} \sum_{j\in\mathbb{B}_i\left(\lceil \frac{D^\vee}{\delta} \rceil\right)\cap\mathcal{K}_{n_0}} \widehat{R}_{\infty,\Delta_i(\delta)}(\varphi; h, \widehat{\theta}_{n_0}) \widehat{R}_{\infty,\Delta_j(\delta)}(\Phi; h, \widehat{\theta}_{n_0}), \\
\widehat{\underline{\mathbf{\Sigma}}}_{2n_0}(\varphi, \widehat{\theta}_{n_0}(\varphi), \delta, D^\vee) &= |\Lambda_{n_0}|^{-1} \sum_{i\in\mathcal{K}_{n_0}} \sum_{j\in\mathbb{B}_i\left(\lceil \frac{D^\vee}{\delta} \rceil\right)\cap\mathcal{K}_{n_0}} \widehat{\mathbf{R}}_{\infty,\Delta_i(\delta)}(\varphi; \mathbf{h}, \widehat{\theta}_{n_0}) \widehat{\mathbf{R}}_{\infty,\Delta_j(\delta)}(\varphi; \mathbf{h}, \widehat{\theta}_{n_0})^T.
\end{aligned}$$

Corollary 15. *Under the notation and assumptions of Propositions 4 and 6, and under **[Exp]**, then, for any $\delta > 0$ as above, one can consider a sequence δ_n which satisfies $\delta_{n_0} = \delta$ and $\delta_n \to \delta$, such that for any $D^\vee \geq D$, the estimators $\widehat{\lambda}_{n,Inn}(\Phi, \widehat{\theta}_n(\Phi), \delta_n, D^\vee)$, $\widehat{\lambda}_{n,Res}(\Phi, \widehat{\theta}_n(\Phi), \delta_n, D^\vee)$ and $\widehat{\underline{\mathbf{\Sigma}}}_{2n}(\Phi, \widehat{\theta}_n(\Phi), \delta_n, D^\vee)$ converge in probability (as $n \to +\infty$) towards respectively λ_{Inn}, λ_{Res} and $\underline{\mathbf{\Sigma}}_2(\theta^\star)$.*

Proof. We apply Proposition 9, where for any $\theta \in \mathcal{V}(\theta^\star)$, we set

- for λ_{Inn}: $\widehat{\mathbf{Y}}_\Lambda(\varphi;\theta) = \mathbf{Y}_\Lambda(\varphi;\theta) = I_\Lambda(\varphi; h, \theta)$.
- for λ_{Res}: $\mathbf{Y}_\Lambda(\varphi;\theta) = R_{\infty,\Lambda}(\varphi; h, \theta)$ and $\widehat{\mathbf{Y}}_{n,\Lambda}(\varphi;\theta) = \widehat{R}_{n,\infty,\Lambda}(\varphi; h, \theta)$.
- for $\underline{\mathbf{\Sigma}}_2(\theta^\star)$: $\mathbf{Y}_\Lambda(\varphi;\theta) = \mathbf{R}_{\infty,\Lambda}(\varphi; \mathbf{h}, \theta)$ and $\widehat{\mathbf{Y}}_{n,\Lambda}(\varphi;\theta) = \widehat{\mathbf{R}}_{n,\infty,\Lambda}(\varphi; \mathbf{h}, \theta)$.

The result is obvious for λ_{Inn}. For λ_{Res} (the proof is similar for $\underline{\mathbf{\Sigma}}_2(\theta^\star)$), it remains to prove that for any $\theta \in \mathbf{\Theta}$, $\sup_{k\in\mathcal{K}_n} \left|\widehat{R}_{n,\infty,\Delta_k(\delta_n)}(\Phi; h, \theta) - R_{n,\infty,\Delta_k(\delta_n)}(\Phi; h, \theta)\right| \to 0$ in probability as $n \to +\infty$. For any $k \in \mathcal{K}_n$, we derive

$$\widehat{R}_{n,\infty,\Delta_k(\delta_n)}(\Phi; h, \theta) - R_{\infty,\Delta_k(\delta_n)}(\Phi; h, \theta) = \mathbf{LPL}^{(1)}_{\Delta_k(\delta_n)}(\varphi; \theta)^T \left(\mathbf{W}(h,\theta) - \widehat{\mathbf{W}}_n(\varphi; h, \theta)\right).$$

Assumption **[E2(bis)]** (implied by **[Exp]**, see Proposition 14) ensures that $|\mathbf{LPL}^{(1)}_{\Delta_k(\delta_n)\setminus\Delta_k(\delta)}(\Phi;\theta)|$ converges to 0 in quadratic mean. In particular, the convergence of $\mathbf{LPL}^{(1)}_{\Delta_k(\delta_n)}(\Phi;\theta)$ towards $\mathbf{LPL}^{(1)}_{\Delta_k(\delta)}(\Phi;\theta)$ holds in probability. Moreover under the assumptions **[N1]** and **[E2(bis)]**, the ergodic theorem of Nguyen and Zessin (1979b) may be applied to prove that $\widehat{\mathbf{W}}_n(\Phi;h,\theta)$ converges almost surely towards $\mathbf{W}(h,\theta)$, as $n\to+\infty$. Slutsky's theorem ends the proof. ■

Remark 13. *If the model is not an exponential model, Corollary 15 still holds by replacing the vector of the sufficient statistics, $\mathbf{v}(x^m|\varphi)$, by the gradient vector of the local energy function, $\mathbf{V}^{(1)}(x^m|\varphi)$ in the different definitions.*

6.4 Positive definiteness of covariance matrices when considering the MPLE

Let us now focus on the positive-definitess of the above quantities. According to Proposition 7 the key assumption to check is **[PD]**.

As adressed in Remark 8, we begin by giving a general result ensuring that $\lambda_{Inn}>0$.

Proposition 16. *Under the assumption **[Exp]**, then $\lambda_{Inn}>0$ for the raw residuals, the Pearson residuals and the inverse residuals.*

Proof. In **[PD]**, we fix $\overline{\delta}=D$ and $B=\emptyset$. Let us write $\overline{\Omega}:=\overline{\Omega}_\emptyset$. Consider the following events for some $n\geq 1$

$$A_0=\left\{\varphi\in\overline{\Omega}:\varphi(\Delta_0(\overline{\delta}))=0\right\}\quad\text{and}\quad A_n=\left\{\varphi\in\overline{\Omega}:\varphi(\Delta_0(\overline{\delta}))=n\right\},$$

and let $\varphi_0\in A_0$ and $\varphi_n\in A_n$. Recall that the local stability property (ensured by **[Exp]**) asserts that there exists $K\geq 0$ such that $V(x^m|\varphi;\theta^\star)\geq -K$ for any $x^m\in\mathbb{S}$ and any $\varphi\in\Omega$. Now, let us consider the three type of residuals.

Raw residuals ($h=1$). From the local stability property

$$|I_{\overline{\Lambda}}(\varphi_n;h,\theta^\star)-I_{\overline{\Lambda}}(\varphi_0;h,\theta^\star)|\geq n-\left|\int_{\overline{\Lambda}\times\mathbb{M}}e^{-V(x^m|\varphi_n;\theta^\star)}-e^{-V(x^m|\varphi_0;\theta^\star)}\mu(dx^m)\right|\geq n-2|\overline{\Lambda}|e^K>0,$$

for n large enough. And so assuming that the left-hand-side is zero leads to a contradiction, which proves **[PD]**.

Inverse residuals ($h=e^V$). Again, from the local stability property

$$|I_{\overline{\Lambda}}(\varphi_n;h,\theta^\star)-I_{\overline{\Lambda}}(\varphi_0;h,\theta^\star)|=\left|\sum_{x^m\in\varphi_n{}_{\overline{\Lambda}}}e^{V(x^m|\varphi_n\setminus x^m;\theta^\star)}\right|\geq ne^{-K}>0,$$

which proves **[PD]** similarly to the previous case.

Pearson residuals ($h=e^{V/2}$). From the same argument

$$\begin{aligned}|I_{\overline{\Lambda}}(\varphi_n;h,\theta^\star)-I_{\overline{\Lambda}}(\varphi_0;h,\theta^\star)| &\geq \left|\sum_{x^m\in\varphi_n{}_{\overline{\Lambda}}}e^{V(x^m|\varphi_n\setminus x^m;\theta^\star)/2}\right|-\\ &\quad\left|\int_{\overline{\Lambda}\times\mathbb{M}}e^{-V(x^m|\varphi_n;\theta^\star)/2}-e^{-V(x^m|\varphi_0;\theta^\star)/2}\mu(dx^m)\right|\\ &\geq ne^{-K/2}-2|\overline{\Lambda}|e^{K/2}>0,\end{aligned}$$

for n large enough, which ends the proof. ■

Proposition 16 asserts that **[PD]** is fullfilled for $\mathbf{Y}_{\overline{\Lambda}}(\Phi;\theta^\star)=I_{\overline{\Lambda}}(\Phi;\theta^\star)$. Therefore, the combination of Propositions 13, 14 and 16 and Corollary 15 ensures all the conditions of Corollary 10

hold. So a goodness-of-fit test based on (18) is valid for exponential family models satisfying [**Exp**] and for the raw residuals, the Pearson residuals and the inverse ones.

Now, let us focus on tests based on Corollary 11 and 12. The following result is important from a practical point of view. It asserts that λ_{Res} (and so $\underline{\mathbf{\Sigma}}_1(\theta^\star)$), and $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ may fail to be positive-definite for an inappropriate choice of test function.

Proposition 17. *Let us consider an exponential family model, let $\widehat{\theta} := \widehat{\theta}^{MPLE}$ and let us choose a test function of the form $h(x^m, \varphi; \theta) = \omega^T \mathbf{v}(x^m|\varphi)$ for some $\omega \in \mathbb{R}^p \setminus 0$, then $\lambda_{Res} = 0$ and the matrices $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ and $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ in Propositions 4 and 6 are only semidefinite-positive matrices.*

Proof. The result is proved by noticing that

$$\begin{aligned}
\underline{\mathbf{H}}(\theta^\star)\,\omega &= \mathbf{E}\left(\mathbf{v}(0^M|\Phi)\mathbf{v}(0^M|\Phi)^T e^{-V\left(0^M|\Phi;\theta^\star\right)}\right)\,\omega \\
&= \mathbf{E}\left(\mathbf{v}(0^M|\Phi)\left(\omega^T\mathbf{v}(0^M|\Phi)\right)^T e^{-V\left(0^M|\Phi;\theta^\star\right)}\right) \\
&= \mathbf{E}\left(h(0^M,\Phi;\theta^\star)\mathbf{v}(0^M|\Phi)e^{-V\left(0^M|\Phi;\theta^\star\right)}\right) = \mathcal{E}(\omega^T\mathbf{v},\theta^\star).
\end{aligned}$$

Therefore, $\mathbf{W}\left(\omega^T\mathbf{v},\theta^\star\right) = \underline{\mathbf{H}}(\theta^\star)^{-1}\mathcal{E}(\omega^T\mathbf{v},\theta^\star) = \omega$, which means that for any $\varphi \in \Omega$ and any bounded domain Λ

$$R_{\infty,\Lambda}(\varphi;\omega^T\mathbf{v},\theta^\star) = I_\Lambda(\varphi;\omega^T\mathbf{v},\theta^\star) - \mathbf{LPL}^{(1)}_\Lambda(\varphi;\theta^\star)^T\omega = 0.$$

This means that if, for the framework 1, the test function is of the form $h = \omega^T\mathbf{v}$ then $\lambda_{Res} = 0$ and if one of the test functions, for the framework 2, is of the form $h = \omega^T\mathbf{v}$, then $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ is necessary singular. ■

Remark 14. *As for Corollary 15, the result of Proposition 17 still holds in general by replacing the vector $\mathbf{v}(x^m|\varphi)$ by the gradient vector of the local energy function $\mathbf{V}^{(1)}(x^m|\varphi)$.*

As a consequence of Proposition 17, the two goodness-of-fit tests based on $T'_{1,n}$ and $T_{2,n}$ in Section 5.2 and 5.3 are not available (for the MPLE) if the test function h is a linear combination of the sufficient statistics $\mathbf{v}(x^m|\varphi)$. Since for most classical models, the value 1 can be obtained from a linear combination of $\mathbf{v}(x^m|\varphi)$, the raw residuals ($h = 1$) are not an appropriate choice for these two tests. This is the case for the two following examples: the area-interaction point process and the 2-type marked Strauss point process, which are presented in details in Appendix B. The following result proves that for a different choice of $h-$residuals, $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ and $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ are positive-definite.

Proposition 18. *For the 2-type marked Strauss point process and the area-interaction point process, when considering the MPLE as an estimator of $\theta^\star$, then*

- *the matrix $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ obtained in Framework 1 from the inverse residuals $h = e^V$,*
- *the matrix $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ obtained in Framework 2 from the empty space residuals, which are constructed for $0 < r_1 < \ldots < r_s < +\infty$ from the family of test functions*

$$h_j(x^m,\varphi;\theta) = \mathbf{1}_{[0,r_j]}(d(x^m,\varphi))e^{V(x^m|\varphi;\theta)}, \quad j = 1,\ldots,s,$$

are positive-definite.

The proof of this result is postponed in Appendix B. The combination of Propositions 13, 14, 18 and Corollary 15 ensures all the conditions of Corollary 11 and 12 hold. So a goodness-of-fit test based on (19) (resp. (20)) is valid for the 2-type marked Strauss point process and the area-interaction point process and for the inverse residuals (resp. the family of test functions based on the empty space function).

Following the Proof of Proposition 18, it is the belief of the authors that such a result holds for other models and other choices of test functions. However, another model and/or test functions will lead to a specific proof. Therefore, this result cannot be as general as the one presented in Proposition 16.

7 The non-hereditary case

Up to here, we have assumed through **[Mod-E]** that the family of energies is hereditary. We consider in this section the non-hereditary case. This particular situation can only occur in presence of a hardcore interaction. From a general point of view, we say that a family of energies involves a hardcore interaction if some point configurations have an infinite energy. Many classical models of Gibbs measures include a hardcore part, as the hard ball model.

A family of energies involving a hardcore part is hereditary if (2) holds. This is a common assumption done for Gibbs energies and it appears to be fulfilled in most classical models, including the hard ball model. However, one may encounter some non-hereditary models, in the sense that (2) does not hold. Intuitively, in this case, when one removes a point from an allowed point configuration, it is possible to obtain a forbidden point configuration. This occurs for instance for Gibbs Delaunay-Voronoï tessellations or forced-clustering processes (see Dereudre and Lavancier (2009) and Dereudre and Lavancier (2010)).

In the non-hereditary case, the GNZ formula (3), which is the basis to define the residuals, becomes false (see Remark 2 in Dereudre and Lavancier (2009)). It is extended to non-hereditary interactions in Dereudre and Lavancier (2009). This generalization requires to introduce the notion of removable points.

Definition 3. *Let $\varphi \in \Omega$ and $x \in \varphi$, then x is removable from φ if there exists $\Lambda \in \mathcal{B}(\mathbb{R}^d)$ such that $x \in \Lambda$ and $V_\Lambda(\varphi - x; \theta) < \infty$. The set of removable points in φ is denoted by $\mathcal{R}(\varphi)$.*

Notice that in the hereditary case, $\mathcal{R}(\varphi) = \varphi$.

The GNZ formula is then generalized to the non-hereditary case as follows. Assuming for any $\theta \in \mathbf{\Theta}$ that a Gibbs measure exists for the family of energies $(V_\Lambda(.;\theta))_{\Lambda \in \mathcal{B}(\mathbb{R}^d)}$, then, for any function $h(\cdot,\cdot,\theta) : \mathbb{S} \times \Omega \to \mathbb{R}^2$ such that the following quantities are finite,

$$\mathbf{E}\left(\int_{\mathbb{R}^d \times \mathbb{M}} h\left(x^m, \Phi; \theta\right) e^{-V(x^m|\Phi;\theta^\star)} \mu(dx^m)\right) = \mathbf{E}\left(\sum_{x^m \in \mathcal{R}(\Phi)} h\left(x^m, \Phi \setminus x^m; \theta\right)\right). \tag{27}$$

We can therefore define the $h-$residuals for (possibly) non-hereditary interactions. For any bounded domain Λ, if $\widehat{\theta}$ is an estimate of $\theta^\star$, the $h-$residuals are

$$R_\Lambda\left(\varphi; h, \widehat{\theta}\right) = \int_{\Lambda \times \mathbb{M}} h\left(x^m, \varphi; \widehat{\theta}\right) e^{-V\left(x^m|\varphi;\widehat{\theta}\right)} \mu(dx^m) - \sum_{x^m \in \mathcal{R}(\varphi_\Lambda)} h\left(x^m, \varphi \setminus x^m; \widehat{\theta}\right). \tag{28}$$

If the set of removable points $\mathcal{R}(\varphi)$ does not depend on θ, it is straightforward to extend all the asymptotic results obtained for the residuals in the preceding sections to (28).

If the set of removable points depends on θ, this is false. Even in the hereditary case, if θ is a hardcore parameter (as the hardcore distance in the hard ball model) then $\widehat{\theta}$ behaves as an estimator of the support of the distribution P_θ. In this case assumption **[E2]** has typically few chances to hold and the asymptotic law of the residuals is unknown. In Dereudre and Lavancier (2010) Figure 15, some simulations of raw-residuals for Gibbs Voronoï tessellations are presented, involving an estimated hardcore parameter in a non-hereditary setting : they show that the distribution of the residuals does not seem to be gaussian in this case.

8 Proofs

Since any stationary Gibbs measure can be represented as a mixture of ergodic measures (see Preston (1976)), it is sufficient to prove the different convergences involved in this paper for ergodic measures. We therefore assume from now on that $P_{\theta^\star}$ is ergodic.

8.1 Proof of Proposition 2

(a) Under **[C1]**, the ergodic theorem of Nguyen and Zessin (1979b) holds for both terms appearing in the definition of $I_{\tilde{\Lambda}_n}(\varphi; h, \theta^\star)$. Then, as $n \to +\infty$, one has $P_{\theta^\star}-$a.s.

$$|\tilde{\Lambda}_n|^{-1} I_{\tilde{\Lambda}_n}(\Phi; h, \theta^\star) \to \mathbf{E}\left(h\left(0^M, \Phi; \theta^\star\right) e^{-V\left(0^M|\Phi;\theta^\star\right)}\right) - \mathbf{E}\left(h\left(0^M, \Phi \setminus 0^M; \theta^\star\right)\right),$$

which equals to 0 from the GNZ formula (4).

(b) The aim is to prove that the difference $|\tilde{\Lambda}_n|^{-1} R_{\tilde{\Lambda}_n}\left(\varphi; h, \widehat{\theta}_n(\varphi)\right) - |\tilde{\Lambda}_n|^{-1} I_{\tilde{\Lambda}_n}(\varphi; h, \theta^\star)$ converges towards 0 for $P_{\theta^\star}-$a.e. φ. Let us write

$$R_{\tilde{\Lambda}_n}\left(\varphi; h, \widehat{\theta}_n(\varphi)\right) - I_{\tilde{\Lambda}_n}(\Phi; h, \theta^\star) := T_1(\varphi) - T_2(\varphi)$$

with

$$T_1(\varphi) := \int_{\tilde{\Lambda}_n \times \mathbb{M}} \left(f\left(x^m, \varphi; \widehat{\theta}_n(\varphi)\right) - f\left(x^m, \varphi; \theta^\star\right)\right) \mu(dx^m) \tag{29}$$

$$T_2(\varphi) := \sum_{x^m \in \varphi_{\tilde{\Lambda}_n}} h\left(x^m, \varphi \setminus x^m; \widehat{\theta}_n(\varphi)\right) - h\left(x^m, \varphi \setminus x^m; \theta^\star\right). \tag{30}$$

Under the Assumptions **[C2]** and **[E1]**, from the ergodic theorem and the GNZ formula, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$\begin{aligned} |\tilde{\Lambda}_n|^{-1} T_1(\varphi) &\leq \frac{2}{|\tilde{\Lambda}_n|} \int_{\tilde{\Lambda}_n \times \mathbb{M}} \left(\widehat{\theta}_n(\varphi) - \theta^\star\right)^T \mathbf{f}^{(1)}\left(x^m, \varphi; \theta^\star\right) \mu(dx^m) \\ &\leq 2\|\widehat{\theta}_n(\varphi) - \theta^\star\| \times \frac{1}{|\tilde{\Lambda}_n|} \int_{\tilde{\Lambda}_n \times \mathbb{M}} \|\mathbf{f}^{(1)}\left(x^m, \varphi; \theta^\star\right)\| \mu(dx^m) \\ &\leq 4\|\widehat{\theta}_n(\varphi) - \theta^\star\| \times \mathbf{E}\left(\|\mathbf{f}^{(1)}\left(0^M, \Phi; \theta^\star\right)\|\right), \end{aligned} \tag{31}$$

and

$$\begin{aligned} |\tilde{\Lambda}_n|^{-1} T_2(\varphi) &\leq \frac{2}{|\tilde{\Lambda}_n|} \sum_{x^m \in \varphi_{\tilde{\Lambda}_n}} \left(\widehat{\theta}_n(\varphi) - \theta^\star\right)^T \mathbf{h}^{(1)}\left(x^m, \varphi \setminus x^m; \theta^\star\right) \\ &\leq 4\|\widehat{\theta}_n(\varphi) - \theta^\star\| \times \mathbf{E}\left(\|\mathbf{h}^{(1)}\left(0^M, \Phi; \theta^\star\right)\| e^{-V\left(0^M|\Phi;\theta^\star\right)}\right). \end{aligned} \tag{32}$$

Equations (31) and (32) lead to

$$|\tilde{\Lambda}_n|^{-1} R_{\tilde{\Lambda}_n}\left(\varphi; h, \widehat{\theta}_n(\varphi)\right) - |\tilde{\Lambda}_n|^{-1} I_{\tilde{\Lambda}_n}(\varphi; h, \theta^\star) \leq c\|\widehat{\theta}_n(\varphi) - \theta^\star\|,$$

for n large enough, with $c = 4 \times \mathbf{E}\left(\|\mathbf{f}^{(1)}\left(0^M, \Phi; \theta^\star\right)\| + \|\mathbf{h}^{(1)}\left(0^M, \Phi; \theta^\star\right)\| e^{-V\left(0^M|\Phi;\theta^\star\right)}\right)$.

8.2 Proof of Proposition 3

Recall that

$$R_{\tilde{\Lambda}_n}\left(\varphi; h, \widehat{\theta}_n(\varphi)\right) - I_{\tilde{\Lambda}_n}(\varphi; h, \theta^\star) = T_1(\varphi) - T_2(\varphi)$$

where $T_1(\varphi)$ and $T_2(\varphi)$ are defined by (29) and (30). Let us write

$$\begin{aligned} T_1(\varphi) &= \int_{\tilde{\Lambda}_n \times \mathbb{M}} \left(\widehat{\theta}_n(\varphi) - \theta^\star\right)^T \mathbf{f}^{(1)}\left(x^m, \varphi; \theta^\star\right) \mu(dx^m) + T_1'(\varphi) \\ T_2(\varphi) &= \sum_{x^m \in \varphi_{\tilde{\Lambda}_n}} \left(\widehat{\theta}_n(\varphi) - \theta^\star\right)^T \mathbf{h}^{(1)}\left(x^m, \varphi \setminus x^m; \theta^\star\right) + T_2'(\varphi), \end{aligned}$$

with

$$\begin{aligned} T_1'(\varphi) &:= \int_{\tilde{\Lambda}_n \times \mathbb{M}} A_1\left(x^m, \varphi; \widehat{\theta}_n(\varphi)\right) \mu(dx^m) \\ T_2'(\varphi) &= \sum_{x^m \in \varphi_{\tilde{\Lambda}_n}} A_2\left(x^m, \varphi \setminus x^m; \widehat{\theta}_n(\varphi)\right) \end{aligned}$$

and

$$\begin{aligned} A_1\left(x^m, \varphi; \widehat{\theta}_n(\varphi)\right) &:= f\left(x^m, \varphi; \widehat{\theta}_n\right) - f\left(x^m, \varphi; \theta^\star\right) - \left(\widehat{\theta}_n(\varphi) - \theta^\star\right)^T \mathbf{f}^{(1)}\left(x^m, \varphi; \theta^\star\right) \\ A_2\left(x^m, \varphi; \widehat{\theta}_n(\varphi)\right) &:= h\left(x^m, \varphi; \widehat{\theta}_n\right) - h\left(x^m, \varphi; \theta^\star\right) - \left(\widehat{\theta}_n(\varphi) - \theta^\star\right)^T \mathbf{h}^{(1)}\left(x^m, \varphi; \theta^\star\right). \end{aligned}$$

From the mean value theorem, there exist for $j = 1, \ldots, p$,
$\xi_{1,j}, \xi_{2,j} \in [\min(\widehat{\theta}_1, \theta_1^\star), \max(\widehat{\theta}_1, \theta_1^\star)] \times \ldots \times [\min(\widehat{\theta}_p, \theta_p^\star), \max(\widehat{\theta}_p, \theta_p^\star)]$ such that

$$A_1\left(x^m, \varphi; \widehat{\theta}_n(\varphi)\right) = \sum_{j=1}^{p} (\widehat{\theta}_j - \theta_j^\star)\left(f_j^{(1)}\left(x^m, \varphi; \xi_{1,j}\right) - f_j^{(1)}\left(x^m, \varphi; \theta^\star\right)\right) \tag{33}$$

$$A_2\left(x^m, \varphi \setminus x^m; \widehat{\theta}_n(\varphi)\right) = \sum_{j=1}^{p} (\widehat{\theta}_j - \theta_j^\star)\left(h_j^{(1)}\left(x^m, \varphi \setminus x^m; \xi_{2,j}\right) - h_j^{(1)}\left(x^m, \varphi \setminus x^m; \theta^\star\right)\right). \tag{34}$$

Let $j \in \{1, \ldots, p\}$, again from the mean value theorem, there exist for $\ell = 1, 2$ and for $k = 1, \ldots, p$,
$\eta_{\ell,j,k} \in [\min(\xi_{\ell,j,1}, \theta_1^\star), \max(\xi_{\ell,j,1}, \theta_1^\star)] \times \ldots \times [\min(\xi_{\ell,j,p}, \theta_p^\star), \max(\xi_{\ell,j,p}, \theta_p^\star)]$ such that

$$f_j^{(1)}\left(x^m, \varphi; \xi_{1,j}\right) - f_j^{(1)}\left(x^m, \varphi; \theta^\star\right) = \sum_{k=1}^{p} \left(\xi_{1,j,k} - \theta_k^\star\right) f_{jk}^{(2)}\left(x^m, \varphi; \eta_{1,j,k}\right) \tag{35}$$

$$h_j^{(1)}\left(x^m, \varphi \setminus x^m; \xi_{2,j}\right) - h_j^{(1)}\left(x^m, \varphi \setminus x^m; \theta^\star\right) = \sum_{k=1}^{p} \left(\xi_{2,j,k} - \theta_k^\star\right) h_{jk}^{(2)}\left(x^m, \varphi \setminus x^m; \eta_{2,j,k}\right) \tag{36}$$

By combining (33), (34), (35) and (36) and under **[N1]**, we can deduce the existence of $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, one has for $P_{\theta^\star}-$a.e. φ

$$\begin{aligned} |\tilde{\Lambda}_n|^{-1}|T_1'(\varphi)| &\leq \frac{2}{|\tilde{\Lambda}_n|} \int_{\tilde{\Lambda}_n \times \mathbb{M}} \sum_{j,k} \left|(\widehat{\theta}_j - \theta_j^\star)(\widehat{\theta}_k - \theta_k^\star) f_{jk}^{(2)}\left(x^m, \varphi; \theta^\star\right)\right| \mu(dx^m) \\ &\leq 2\|\widehat{\theta}_n(\varphi) - \theta^\star\|^2 \times \frac{1}{|\tilde{\Lambda}_n|} \int_{\tilde{\Lambda}_n \times \mathbb{M}} \|\underline{\mathbf{f}}^{(2)}\left(x^m, \varphi; \theta^\star\right)\| \mu(dx^m) \\ &\leq 4\|\widehat{\theta}_n(\varphi) - \theta^\star\|^2 \times \mathbf{E}\left(\|\underline{\mathbf{f}}^{(2)}\left(0^M, \Phi; \theta^\star\right)\|\right) \end{aligned}$$

and

$$\begin{aligned} |\tilde{\Lambda}_n|^{-1}|T_2'(\varphi)| &\leq \frac{2}{|\tilde{\Lambda}_n|} \sum_{x^m \in \varphi_{\tilde{\Lambda}_n}} \sum_{j,k} \left|(\widehat{\theta}_j - \theta_j^\star)(\widehat{\theta}_k - \theta_k^\star) h_{jk}^{(2)}\left(x^m, \varphi \setminus x^m; \theta^\star\right)\right| \\ &\leq 2\|\widehat{\theta}_n(\varphi) - \theta^\star\|^2 \times \frac{1}{|\tilde{\Lambda}_n|} \sum_{x^m \in \varphi_{\tilde{\Lambda}_n}} \|\underline{\mathbf{h}}^{(2)}\left(x^m, \varphi \setminus x^m; \theta^\star\right)\| \\ &\leq 4\|\widehat{\theta}_n(\varphi) - \theta^\star\|^2 \times \mathbf{E}\left(\|\underline{\mathbf{h}}^{(2)}\left(0^M, \Phi; \theta^\star\right)\| e^{-V\left(0^M | \Phi; \theta^\star\right)}\right) \end{aligned}$$

Since

$$|\tilde{\Lambda}_n|^{1/2}\|\widehat{\theta}_n(\varphi) - \theta^\star\|^2 = \left(\frac{|\tilde{\Lambda}_n|}{|\Lambda_n|}\right)^{1/2} \| \, |\Lambda_n|^{1/2}(\widehat{\theta}_n(\varphi) - \theta^\star)\| \times \|\widehat{\theta}_n(\varphi) - \theta^\star\|$$

then, under the assumptions of Proposition 3, one has, from Slustsky's theorem, the following convergence in probability as $n \to +\infty$

$$|\tilde{\Lambda}_n|^{1/2} \|\widehat{\theta}_n(\Phi) - \theta^\star\|^2 \xrightarrow{P} 0.$$

By combining all these results, one obtains the following convergence in probability, as $n \to +\infty$

$$|\tilde{\Lambda}_n|^{-1/2} \left(T_1(\Phi) - T_2(\Phi) - |\tilde{\Lambda}_n| \left(\widehat{\theta}_n(\Phi) - \theta^\star\right)^T \mathbf{X}_{\tilde{\Lambda}_n}(\Phi) \right) = |\tilde{\Lambda}_n|^{-1/2} \left(T_1'(\Phi) - T_2'(\Phi) \right) \xrightarrow{P} 0.$$

where $\mathbf{X}_{\tilde{\Lambda}_n}(\Phi)$ is the random vector defined for all $j = 1, \ldots, p$ by

$$\left(\mathbf{X}_{\tilde{\Lambda}_n}(\Phi)\right)_j := \frac{1}{|\tilde{\Lambda}_n|} \int_{\tilde{\Lambda}_n \times \mathbb{M}} f_j^{(1)}\left(x^m, \Phi; \theta^\star\right) \mu(dx^m) - \frac{1}{|\tilde{\Lambda}_n|} \sum_{x^m \in \Phi_{\tilde{\Lambda}_n}} h_j^{(1)}\left(x^m, \Phi \setminus x^m; \theta^\star\right).$$

By using the ergodic theorem and the GNZ formula, one has $P_{\theta^\star}-$a.s. as $n \to +\infty$

$$\left(\mathbf{X}_{\tilde{\Lambda}_n}(\Phi)\right)_j \to \mathbf{E}\left(f_j^{(1)}\left(0^M, \Phi; \theta^\star\right) - h_j^{(1)}\left(0^M, \Phi; \theta^\star\right) e^{-V\left(0^M | \Phi; \theta^\star\right)} \right).$$

Finally, let us notice that for all $(m, \varphi) \in \mathbb{M} \times \Omega$ and for all $j = 1, \ldots, p$

$$\begin{aligned} f_j^{(1)}\left(0^m, \varphi; \theta^\star\right) &= \frac{\partial}{\partial \theta_j} \left(h\left(0^m, \varphi; \theta\right) e^{-V\left(0^m | \varphi; \theta\right)} \right)\Big|_{\theta = \theta^\star} \\ &= h_j^{(1)}\left(0^m, \varphi; \theta^\star\right) e^{-V\left(0^m | \varphi; \theta^\star\right)} - h\left(0^m, \varphi; \theta\right) V_j^{(1)}\left(0^m | \varphi; \theta^\star\right) e^{-V\left(0^m | \varphi; \theta^\star\right)}. \end{aligned}$$

Therefore $\left(\mathbf{X}_{\tilde{\Lambda}_n}(\Phi)\right)_j \to -\mathcal{E}_j(h, \theta^\star)$ $P_{\theta^\star}-$a.s. as $n \to +\infty$. This finally leads to the following convergence in probability, as $n \to +\infty$

$$|\tilde{\Lambda}_n|^{-1/2} \left(T_1(\Phi) - T_2(\Phi) + |\tilde{\Lambda}_n| \left(\widehat{\theta}_n(\Phi) - \theta^\star\right)^T \mathcal{E}(h; \theta^\star)) \right) \xrightarrow{P} 0.$$

8.3 Proof of Proposition 4

Let us first state a result widely used in the following.

Lemma 19. *For any bounded domain Λ and for any test function h*

$$\mathbf{E}\left(I_\Lambda(\Phi; h, \theta^\star) | \Phi_{\Lambda^c} \right) = 0. \tag{37}$$

The proof of Lemma 19 is omitted since it corresponds to the proof of Theorem 2 (Step 1, p. 257) of Billiot et al. (2008) by subsituting $v_j(x^m|\varphi)$ by the test function $h(x^m, \varphi; \theta^\star)$.

For all $n \in \mathbb{N}$, the domain Λ_n is assumed to be a cube divided as $\Lambda_n = \bigcup_{j \in \mathcal{J}} \Lambda_{j,n}$ where for all $j \in \mathcal{J}$, the $\Lambda_{j,n}$'s are disjoint cubes. So $|\Lambda_n| = |\mathcal{J}||\Lambda_{j,n}| = |\mathcal{J}||\Lambda_{0,n}|$. Moreover, for all $j \in \mathcal{J}$, we can decompose each $\Lambda_{j,n}$ in the following way :

$$\Lambda_{j,n} := \bigcup_{k \in \mathcal{K}_{j,n}} \Delta_k(D_n) \tag{38}$$

where the $\Delta_k(D_n)$'s are disjoint cubes with side-length D_n and $\mathcal{K}_{j,n} \subset \mathbb{Z}^d$. The side-length D_n is chosen greater than D and as close as possible to D, leading to

$$D_n = \frac{|\Lambda_n|^{1/d}}{|\mathcal{J}|^{1/d} \left\lfloor \frac{|\Lambda_n|^{1/d}}{|\mathcal{J}|^{1/d} D} \right\rfloor}.$$

This choice implies $D_n \to D$ when $n \to \infty$ and guarantees $D \leq D_n \leq 2D$ as soon as $|\Lambda_n| \geq |\mathcal{J}|D^d$. The cubes $\Lambda_{j,n}$'s are therefore divided into $|\mathcal{K}_{j,n}| = |\Lambda_{0,n}|D_n^{-d}$ cubes whose volumes are closed to D^d. Denoting $\mathcal{K}_n = \bigcup_{j\in\mathcal{J}} \mathcal{K}_{j,n}$, we have $|\mathcal{K}_n| = |\Lambda_n|D_n^{-d} = |\mathcal{J}||\mathcal{K}_{j,n}|$ and finally

$$\Lambda_n = \bigcup_{j\in\mathcal{J}} \bigcup_{k\in\mathcal{K}_{j,n}} \Delta_k(D_n) = \bigcup_{k\in\mathcal{K}_n} \Delta_k(D_n). \tag{39}$$

From Proposition 3 and under Assumption **[E2(bis)]**, one has for any $j \in \mathcal{J}$

$$|\Lambda_{j,n}|^{-1/2} R_{\Lambda_{j,n}}\left(\Phi; h, \widehat{\theta}_n(\Phi)\right) = |\Lambda_{j,n}|^{-1/2} R_{\infty,\Lambda_{j,n}}\left(\Phi; h, \theta^\star\right) + o_P(1),$$

where $R_{\infty,\Lambda_{j,n}}\left(\Phi; h, \theta^\star\right)$ is defined in (8).

Therefore the proof of Proposition 4 reduces to the proof of the asymptotic normality of the vector $\left(|\Lambda_{j,n}|^{-1/2} R_{\infty,\Lambda_{j,n}}\left(\Phi; h, \theta^\star\right)\right)_{j\in\mathcal{J}}$. Now

$$\begin{aligned}
|\Lambda_{j,n}|^{-1/2} R_{\infty,\Lambda_{j,n}}\left(\Phi; h, \theta^\star\right) &= |\Lambda_{0,n}|^{-1/2}\left(I_{\Lambda_{j,n}}\left(\Phi; h, \theta^\star\right) - \frac{|\Lambda_{0,n}|}{|\Lambda_n|}\mathbf{U}_{\Lambda_n}\left(\Phi; \theta^\star\right)^T \mathcal{E}(h; \theta^\star)\right) \\
&= \frac{|\Lambda_{0,n}|^{1/2}}{|\Lambda_n|}\left(|\mathcal{J}| \times I_{\Lambda_{j,n}}\left(\Phi; h, \theta^\star\right) - \mathbf{U}_{\Lambda_n}\left(\Phi; \theta^\star\right)^T \mathcal{E}(h; \theta^\star)\right) \\
&= \frac{1}{D_n^{d/2}|\mathcal{J}|^{1/2}} \frac{1}{|\mathcal{K}_n|^{1/2}} \sum_{k\in\mathcal{K}_n} W_{j,n,\Delta_k(D_n)}\left(\Phi; \theta^\star\right),
\end{aligned} \tag{40}$$

where for any $\varphi \in \Omega$

$$W_{j,n,\Delta_k(D_n)}\left(\varphi; \theta^\star\right) = \begin{cases} W^{(1)}_{\Delta_k(D_n)}\left(\varphi; \theta^\star\right) := |\mathcal{J}| \times I_{\Delta_k(D_n)}\left(\varphi; h, \theta^\star\right) \\ \qquad\qquad -\mathbf{U}_{\Delta_k(D_n)}\left(\varphi; \theta^\star\right)^T \mathcal{E}(h; \theta^\star) \qquad \text{if } k \in \mathcal{K}_{j,n}, \\ W^{(2)}_{\Delta_k(D_n)}\left(\varphi; \theta^\star\right) := -\mathbf{U}_{\Delta_k(D_n)}\left(\varphi; \theta^\star\right)^T \mathcal{E}(h; \theta^\star) \qquad \text{if } k \in \mathcal{K}_n \setminus \mathcal{K}_{j,n}. \end{cases} \tag{41}$$

Therefore, to prove a central limit theorem for the vector $\left(|\Lambda_{j,n}|^{-1/2} R_{\infty,\Lambda_{j,n}}\left(\Phi; h, \theta^\star\right)\right)_{j\in\mathcal{J}}$, it suffices to apply Theorem 21 (see Appendix A), where in its statement we choose $\mathbf{Z}_{n,k} = (W_{j,n,\Delta_k(D_n)}(\Phi; \theta^\star))_{j\in\mathcal{J}}$, $X_{n,i} = \Phi_{\Delta_i(D_n)}$ and $p = |\mathcal{J}|$. For this, we first have to specify the asymptotic variance matrix $\underline{\underline{\mathbf{\Sigma}}}$, then to check the assumptions of Theorem 21.

First step: computation of the asymptotic variance.

Let us fix a cartesian coordinate system such that 0 is the center of Λ_n. We assume, without loss of generality, that $|\mathcal{J}|$ is odd. Moreover, we can always choose an odd number $|\mathcal{K}_{j,n}|$ of cubes $\Delta_k(D_n)$ in (55). Consequently, $\Lambda_{0,n}$ may be centered at 0 and each $\Delta_k(D_n)$ is centered at kD_n, $k \in \mathbb{Z}^d$. Note that if $|\mathcal{J}|$ was even, each $\Delta_k(D_n)$ would be centered at $kD_n/2$. So, in this system, $\mathcal{K}_n$ is a subset of $\mathbb{Z}^d$, independent of D_n, with $|\mathcal{K}_n| = |\mathcal{J}| \left\lfloor \frac{|\Lambda_n|^{1/d}}{|\mathcal{J}|^{1/d} D} \right\rfloor^d$ elements.

Set, for all $k, k' \in \mathbb{Z}^d$,

$$\begin{cases} E^{(1)}_{k,k'}(D_n) := \mathbf{E}\left(W^{(1)}_{\Delta_k(D_n)}\left(\Phi; \theta^\star\right) W^{(1)}_{\Delta_{k'}(D_n)}\left(\Phi; \theta^\star\right)\right) \\ E^{(12)}_{k,k'}(D_n) := \mathbf{E}\left(W^{(1)}_{\Delta_k(D_n)}\left(\Phi; \theta^\star\right) W^{(2)}_{\Delta_{k'}(D_n)}\left(\Phi; \theta^\star\right)\right) \\ E^{(2)}_{k,k'}(D_n) := \mathbf{E}\left(W^{(2)}_{\Delta_k(D_n)}\left(\Phi; \theta^\star\right) W^{(2)}_{\Delta_{k'}(D_n)}\left(\Phi; \theta^\star\right)\right) \end{cases}$$

Note that from the stationarity of the point process, we have $E^{(l)}_{k,k'}(D_n) = E^{(l)}_{0,k-k'}(D_n)$, for $l = 1, 12, 2$. Moreover, under Assumptions **[N4]** and **[E2(bis)]**, for any $k \in \mathcal{K}_n$ and for any configuration φ, since $D_n \geq D$, $W^{(i)}_{\Delta_k(D_n)}\left(\varphi; \theta^\star\right)$, $i = 1, 2$, depends only on $\varphi_{\Delta_l(D_n)}$ for $|l - k| \leq 1$

that is $l \in \mathbb{B}_k(1)$. As a consequence, if $k' \in \mathbb{B}_k^c(1)$, $W^{(i)}_{\Delta_{k'}(D_n)}(\Phi;\theta^\star)$ is a measurable function of $\Phi_{\Delta_k^c(D_n)}$. This leads, for $i,j=1,2$, to

$$\begin{aligned}\mathbf{E}\left(W^{(i)}_{\Delta_k(D_n)}(\Phi;\theta^\star)\,W^{(j)}_{\Delta_{k'}(D_n)}(\Phi;\theta^\star)\right) &= \mathbf{E}\left(\mathbf{E}\left(W^{(i)}_{\Delta_k(D_n)}(\Phi;\theta^\star)\,W^{(j)}_{\Delta_{k'}(D_n)}(\Phi;\theta^\star)\,|\Phi_{\Delta_k^c(D_n)}\right)\right)\\ &= \mathbf{E}\left(W^{(i)}_{\Delta_{k'}(D_n)}(\Phi;\theta^\star)\,\mathbf{E}\left(W^{(j)}_{\Delta_k(D_n)}(\Phi;\theta^\star)\,|\Phi_{\Delta_k^c(D_n)}\right)\right).\end{aligned} \tag{42}$$

From Lemma 19 and under **[E2(bis)]** then for any $k\in\mathbb{Z}^d$ and for $i=1,2$,

$$\mathbf{E}\left(W^{(i)}_{\Delta_k(D_n)}(\Phi;\theta^\star)\,|\Phi_{\Delta_k^c(D_n)}\right)=0. \tag{43}$$

From (42) and (43), we deduce that, for $l=1,12,2$,

$$k'\in\mathbb{B}_k^c(1)\Longrightarrow E^{(l)}_{k,k'}(D_n)=0. \tag{44}$$

We are now in position to compute the covariance. For any i and j in $\mathcal{J}$, from (40),

$$\begin{aligned}&cov\left(|\Lambda_{i,n}|^{-1/2}R_{\infty,\Lambda_{i,n}}(\Phi;h,\theta^\star),|\Lambda_{j,n}|^{-1/2}R_{\infty,\Lambda_{j,n}}(\Phi;h,\theta^\star)\right)\\ &\qquad=\frac{1}{D_n^d|\mathcal{J}|}\mathbf{E}\left(\frac{1}{|\mathcal{K}_n|}\sum_{k\in\mathcal{K}_n}\sum_{k'\in\mathcal{K}_n}W_{i,n,\Delta_k(D_n)}(\Phi;\theta^\star)\,W_{j,n,\Delta_{k'}(D_n)}(\Phi;\theta^\star)\right).\end{aligned} \tag{45}$$

Let us first consider the case $i=j$. We may write

$$\begin{aligned}&\mathbf{E}\left(\frac{1}{|\mathcal{K}_n|}\sum_{k\in\mathcal{K}_n}\sum_{k'\in\mathcal{K}_n}W_{i,n,\Delta_k(D_n)}(\Phi;\theta^\star)^2\right)\\ &\qquad=\frac{1}{|\mathcal{K}_n|}\Bigg(\underbrace{\sum_{k,k'\in\mathcal{K}_{i,n}}E^{(1)}_{k,k'}(D_n)}_{:=S_1}+2\underbrace{\sum_{k\in\mathcal{K}_{i,n},k'\in\mathcal{K}_n\setminus\mathcal{K}_{i,n}}E^{(12)}_{k,k'}(D_n)}_{:=S_2}+\underbrace{\sum_{k,k'\in\mathcal{K}_n\setminus\mathcal{K}_{i,n}}E^{(2)}_{k,k'}(D_n)}_{:=S_3}\Bigg).\end{aligned}$$

The following lemma will be useful to drop the dependence on D_n in each term S_1, S_2, S_3 above.

Lemma 20. *For any $i,j=1,2$, denoting $\overline{\Delta}_0(\tau)=\cup_{k\in\mathbb{B}_0(1)}\Delta_k(\tau)$ (for some $\tau>0$), we have*

$$W^{(i)}_{\Delta_0(D_n)}(\Phi;\theta^\star)\,W^{(j)}_{\overline{\Delta}_0(D_n)}(\Phi;\theta^\star)\xrightarrow{L_1}W^{(i)}_{\Delta_0(D)}(\Phi;\theta^\star)\,W^{(j)}_{\overline{\Delta}_0(D)}(\Phi;\theta^\star).$$

Proof. For any $i=1,2$, $W^{(i)}_{\Delta_0(D_n)}$ is a linear combination of $I_{\Delta_0(D_n)}$ and $\mathbf{U}_{\Delta_0(D_n)}$, which converge respectively in L^2 to $I_{\Delta_0(D)}$ and $\mathbf{U}_{\Delta_0(D)}$ by **[N3]** and **[E2(bis)]**, since $D_n\to D$. Thus $W^{(i)}_{\Delta_0(D_n)}$ converges in L^2 to $W^{(i)}_{\Delta_0(D)}$ as $n\to\infty$. Similarly, for any $j=1,2$, $W^{(j)}_{\overline{\Delta}_0(D_n)}$ tends in L^2 to $W^{(j)}_{\overline{\Delta}_0(D)}$. The convergence stated in Lemma 20 then follows. ■

Let us focus on the asymptotic of each term S_1, S_2, S_3.

Term S_1: from (44),

$$S_1=\sum_{k\in\mathcal{K}_{i,n}}\Bigg(\sum_{k'\in\mathbb{B}_k(1)\cap\mathcal{K}_{i,n}}E^{(1)}_{k,k'}(D_n)+\underbrace{\sum_{k'\in\mathbb{B}_k^c(1)\cap\mathcal{K}_{i,n}}E^{(1)}_{k,k'}(D_n)}_{=0}\Bigg)=\sum_{k\in\mathcal{K}_{i,n}}\sum_{k'\in\mathbb{B}_k(1)\cap\mathcal{K}_{i,n}}E^{(1)}_{k,k'}(D_n).$$

Let $\widetilde{\mathcal{K}}_{i,n} := \mathcal{K}_{i,n} \cap \left(\cup_{j\in\partial\mathcal{K}_{i,n}} \mathbb{B}_j(1)\right)$ and note that $\frac{|\widetilde{\mathcal{K}}_{i,n}|}{|\mathcal{K}_{i,n}|} \to 0$ as $n \to +\infty$. Then,

$$S_1 = \sum_{k\in\mathcal{K}_{i,n}\setminus\widetilde{\mathcal{K}}_{i,n}} \sum_{k'\in\mathbb{B}_k(1)\cap\mathcal{K}_{i,n}} E^{(1)}_{k,k'}(D_n) + \underbrace{\sum_{k\in\widetilde{\mathcal{K}}_{i,n}} \sum_{k'\in\mathbb{B}_k(1)\cap\mathcal{K}_{i,n}} E^{(1)}_{k,k'}(D_n)}_{:=A_1}.$$

Since,

$$\frac{1}{|\mathcal{K}_n|} \times |A_1| \leq \frac{|\widetilde{\mathcal{K}}_{i,n}|}{|\mathcal{K}_n|} \sum_{k\in\mathbb{B}_0(1)} |E^{(1)}_{0,k}(D_n)| \overset{n\to+\infty}{\longrightarrow} 0,$$

(because $D \leq D_n \leq 2D$ and $\frac{|\widetilde{\mathcal{K}}_{i,n}|}{|\mathcal{K}_{i,n}|} \to 0$), we obtain, as $n \to +\infty$,

$$\frac{1}{|\mathcal{K}_n|} S_1 \quad \sim \quad \frac{|\mathcal{K}_{i,n} \setminus \widetilde{\mathcal{K}}_{i,n}|}{|\mathcal{K}_n|} \sum_{k\in\mathbb{B}_0(1)} E^{(1)}_{0,k}(D_n) \sim \frac{|\mathcal{K}_{i,n}|}{|\mathcal{K}_n|} \sum_{k\in\mathbb{B}_0(1)} E^{(1)}_{0,k}(D_n).$$

From Lemma 20,

$$\sum_{k\in\mathbb{B}_0(1)} E^{(1)}_{0,k}(D_n) = \mathbf{E}\left(W^{(1)}_{\Delta_0(D_n)}(\Phi;\theta^\star)\, W^{(1)}_{\overline{\Delta}_0(D_n)}(\Phi;\theta^\star)\right) \longrightarrow \sum_{k\in\mathbb{B}_0(1)} E^{(1)}_{0,k}(D).$$

Therefore,

$$\frac{1}{|\mathcal{K}_n|} S_1 \sim \frac{1}{|\mathcal{J}|} \sum_{k\in\mathbb{B}_0(1)} E^{(1)}_{0,k}(D).$$

Term S_2: with similar arguments as above, we obtain

$$\begin{aligned} S_2 \quad = \quad & \underbrace{\sum_{k\in\mathcal{K}_{i,n}\setminus\widetilde{\mathcal{K}}_{i,n}} \sum_{k'\in\mathcal{K}_n\setminus\mathcal{K}_{i,n}} E^{(12)}_{k,k'}(D_n)}_{=0} \\ & + \sum_{k\in\widetilde{\mathcal{K}}_{i,n}} \left(\sum_{k'\in\mathbb{B}_k(1)\cap(\mathcal{K}_n\setminus\mathcal{K}_{i,n})} E^{(12)}_{k,k'}(D_n) + \underbrace{\sum_{k'\in\mathbb{B}^c_k(1)\cap(\mathcal{K}_n\setminus\mathcal{K}_{i,n})} E^{(12)}_{k,k'}(D_n)}_{=0} \right) \end{aligned}$$

Therefore, since $\frac{|\widetilde{\mathcal{K}}_{i,n}|}{|\mathcal{K}_n|} \to 0$ and $D \leq D_n \leq 2D$,

$$\frac{1}{|\mathcal{K}_n|} S_2 \leq \frac{|\widetilde{\mathcal{K}}_{i,n}|}{|\mathcal{K}_n|} \sum_{k\in\mathbb{B}_0(1)} |E^{(12)}_{0,k}(D_n)| \overset{n\to+\infty}{\longrightarrow} 0.$$

Term S_3:

$$S_3 = \sum_{k\in\mathcal{K}_n\setminus\mathcal{K}_{i,n}} \sum_{k'\in\mathbb{B}_k(1)\cap(\mathcal{K}_n\setminus\mathcal{K}_{i,n})} E^{(2)}_{k,k'}(D_n) + \underbrace{\sum_{k\in\mathcal{K}_n\setminus\mathcal{K}_{i,n}} \sum_{k'\in\mathbb{B}^c_k(1)\cap(\mathcal{K}_n\setminus\mathcal{K}_{i,n})} E^{(2)}_{k,k'}(D_n)}_{=0}.$$

Let $\widetilde{\mathcal{K}}_n = (\mathcal{K}_n \setminus \mathcal{K}_{i,n}) \cap \left(\cup_{j\in\partial(\mathcal{K}_n\setminus\mathcal{K}_{i,n})} \mathbb{B}_j(1)\right)$ and note that $\frac{|\widetilde{\mathcal{K}}_n|}{|\mathcal{K}_n|} \to 0$, as $n \to +\infty$. Then,

$$S_3 = \sum_{k\in\mathcal{K}_n\setminus\widetilde{\mathcal{K}}_n} \sum_{k'\in\mathbb{B}_k(1)\cap(\mathcal{K}_n\setminus\mathcal{K}_{i,n})} E^{(2)}_{k,k'}(D_n) + \underbrace{\sum_{k\in\widetilde{\mathcal{K}}_n} \sum_{k'\in\mathbb{B}_k(1)\cap(\mathcal{K}_n\setminus\mathcal{K}_{i,n})} E^{(2)}_{k,k'}(D_n)}_{:=A_3}.$$

Since,

$$\frac{1}{|\mathcal{K}_n|}|A_3| \leq \frac{|\widetilde{\mathcal{K}}_n|}{|\mathcal{K}_n|} \sum_{k\in\mathbb{B}_0(1)} |E^{(2)}_{0,k}(D_n)| \overset{n\to+\infty}{\longrightarrow} 0,$$

we obtain, from Lemma 20,

$$\frac{1}{|\mathcal{K}_n|}S_3 \sim \frac{|\mathcal{K}_n \setminus \widetilde{\mathcal{K}}_n|}{|\mathcal{K}_n|} \sum_{k\in\mathbb{B}_0(1)} E^{(2)}_{0,k}(D_n) \sim \frac{|\mathcal{J}|-1}{|\mathcal{J}|} \sum_{k\in\mathbb{B}_0(1)} E^{(2)}_{0,k}(D).$$

Combining the three terms S_1, S_2 and S_3, we have, as $n \to +\infty$

$$\mathbf{E}\left(\frac{1}{|\mathcal{K}_n|}\sum_{k\in\mathcal{K}_n}\sum_{k'\in\mathcal{K}_n} W_{i,n,\Delta_k(D_n)}\left(\Phi;\theta^\star\right)^2\right) \sim \sum_{k\in\mathbb{B}_0(1)}\left(\frac{1}{|\mathcal{J}|}E^{(1)}_{0,k}(D) + \frac{|\mathcal{J}|-1}{|\mathcal{J}|}E^{(2)}_{0,k}(D)\right). \tag{46}$$

When $i \neq j$, there are three main cases in (45), according to $k,k' \in \mathcal{K}_{i,n}$, $k,k' \in \mathcal{K}_{j,n}$, or $k,k' \in \mathcal{K}_n \setminus (\mathcal{K}_{i,n} \cup \mathcal{K}_{j,n})$. As for the case $i=j$ treated before, the other situations involve non-zero correlations on edges sets like $\widetilde{\mathcal{K}}_{i,n}$, which are negligible with respect to $|\mathcal{K}_n|$. The covariance is therefore equivalent, up to $D_n^d|\mathcal{J}|$, to

$$\frac{1}{|\mathcal{K}_n|}\sum_{k,k'\in\mathcal{K}_{i,n}} E^{(12)}_{k,k'}(D_n) + \frac{1}{|\mathcal{K}_n|}\sum_{k,k'\in\mathcal{K}_{j,n}} E^{(12)}_{k,k'}(D_n) + \frac{1}{|\mathcal{K}_n|}\sum_{k,k'\in\mathcal{K}_n\setminus(\mathcal{K}_{i,n}\cup\mathcal{K}_{j,n})} E^{(2)}_{k,k'}(D_n).$$

The simplification occurs as for the case $i=j$ and, since $|\mathcal{K}_{i,n}| = |\mathcal{K}_{j,n}|$, we obtain the asymptotic equivalent for the covariance (45)

$$\frac{1}{D^d|\mathcal{J}|}\sum_{k\in\mathbb{B}_0(1)}\left(\frac{2}{|\mathcal{J}|}E^{(12)}_{0,k}(D) + \frac{|\mathcal{J}|-2}{|\mathcal{J}|}E^{(2)}_{0,k}(D)\right). \tag{47}$$

Finally, from (46) and (47), we deduce that $\underline{\mathbf{\Sigma}}_1(\theta^\star)$, defined in Proposition 4, corresponds to the asymptotic variance of $\left(|\Lambda_{j,n}|^{-1/2}R_{\infty,\Lambda_{j,n}}\left(\Phi;h,\theta^\star\right)\right)_{j\in\mathcal{J}}$.

Second step: application of Theorem 21.

We apply Theorem 21 with $\mathbf{Z}_{n,k} = (W_{j,n,\Delta_k(D_n)})_{j\in\mathcal{J}}$, $X_{n,i} = \Phi_{\Delta_i(D_n)}$, $p = |\mathcal{J}|$ and $\underline{\mathbf{\Sigma}} = \underline{\mathbf{\Sigma}}_1(\theta^\star)$, which is a symmetric positive-semidefinite matrix as the limit of a covariance matrix (from the first step of the proof).

The assumption (54) holds from **[N4]**, **[E2(bis)]** and because $D_n \geq D$. Assumptions (i), (ii) and (iii) are direct consequences of **[E2(bis)]**, **[N2]** and Lemma 19. It remains to prove (iv). Assuming $\underline{\mathbf{\Sigma}} = (\Sigma_{ij})$ for $1 \leq i,j \leq p$, from the definition of the Frobenius norm, we have

$$\begin{aligned}\mathbf{E}\left\|\,|\mathcal{K}_n|^{-1}\sum_{k\in\mathcal{K}_n}\sum_{k'\in\mathbb{B}_k(1)\cap\mathcal{K}_n}\mathbf{Z}_{n,k}\mathbf{Z}_{n,k'}{}^T - \underline{\mathbf{\Sigma}}\right\| \\ \leq \sum_{i=1}^p\sum_{j=1}^p \mathbf{E}\left|\,|\mathcal{K}_n|^{-1}\sum_{k\in\mathcal{K}_n}\sum_{k'\in\mathbb{B}_k(1)\cap\mathcal{K}_n} W_{i,n,\Delta_k(D_n)}W_{j,n,\Delta_{k'}(D_n)} - \Sigma_{ij}\right|. \end{aligned}\tag{48}$$

Let us first assume that $i \neq j$ are fixed and denote $Y_{n,k}(D_n) = W_{i,n,\Delta_k(D_n)}$, $S^k_n(D_n) = \sum_{k'\in\mathbb{B}_k(1)\cap\mathcal{K}_n} W_{j,n,\Delta_{k'}(D_n)}$. We have

$$\begin{aligned}\mathbf{E}\left|\,|\mathcal{K}_n|^{-1}\sum_{k\in\mathcal{K}_n}\sum_{k'\in\mathbb{B}_k(1)\cap\mathcal{K}_n} W_{i,n,\Delta_k(D_n)}W_{j,n,\Delta_{k'}(D_n)} - \Sigma_{ij}\right| &= |\mathcal{K}_n|^{-1}\mathbf{E}\left|\sum_{k\in\mathcal{K}_n} Y_{n,k}(D_n)S^k_n(D_n) - \Sigma_{ij}\right| \\ &\leq E_1 + E_2 + E_3 + E_4,\end{aligned}$$

where

$$E_1 = \frac{|\mathcal{K}_{i,n}|}{|\mathcal{K}_n|}\mathbf{E}\left||\mathcal{K}_{i,n}|^{-1}\sum_{k\in\mathcal{K}_{i,n}}(Y_{n,k}(D_n)S_n^k(D_n)-\mathbf{E}(Y_{n,k}(D_n)S_n^k(D_n)))\right|,$$

$$E_2 = \frac{|\mathcal{K}_{j,n}|}{|\mathcal{K}_n|}\mathbf{E}\left||\mathcal{K}_{j,n}|^{-1}\sum_{k\in\mathcal{K}_{j,n}}(Y_{n,k}(D_n)S_n^k(D_n)-\mathbf{E}(Y_{n,k}(D_n)S_n^k(D_n)))\right|,$$

$$E_3 = \frac{|\mathcal{K}_n\setminus(\mathcal{K}_{i,n}\cup\mathcal{K}_{j,n})|}{|\mathcal{K}_n|}\times \mathbf{E}\left||\mathcal{K}_n\setminus(\mathcal{K}_{i,n}\cup\mathcal{K}_{j,n})|^{-1}\sum_{k\in\mathcal{K}_n\setminus(\mathcal{K}_{i,n}\cup\mathcal{K}_{j,n})}(Y_{n,k}(D_n)S_n^k(D_n)-\mathbf{E}(Y_{n,k}(D_n)S_n^k(D_n)))\right|,$$

$$E_4 = \left|\frac{|\mathcal{K}_{i,n}|}{|\mathcal{K}_n|}\sum_{k\in\mathcal{K}_{i,n}}\mathbf{E}(Y_{n,k}(D_n)S_n^k(D_n)) + \frac{|\mathcal{K}_{i,n}|}{|\mathcal{K}_n|}\sum_{k\in\mathcal{K}_{j,n}}\mathbf{E}(Y_{n,k}(D_n)S_n^k(D_n)) + \frac{|\mathcal{K}_n\setminus(\mathcal{K}_{i,n}\cup\mathcal{K}_{j,n})|}{|\mathcal{K}_n|}\sum_{k\in\mathcal{K}_n\setminus(\mathcal{K}_{i,n}\cup\mathcal{K}_{j,n})}\mathbf{E}(Y_{n,k}(D_n)S_n^k(D_n))-\Sigma_{ij}\right|.$$

The first three terms E_1, E_2 and E_3 can be handled similarly. Let us focus on E_1:

$$\begin{aligned}\frac{|\mathcal{K}_n|}{|\mathcal{K}_{i,n}|}E_1 \le{}& |\mathcal{K}_{i,n}|^{-1}\sum_{k\in\mathcal{K}_{i,n}}\mathbf{E}\left|Y_{n,k}(D_n)S_n^k(D_n)-Y_{n,k}(D)S_n^k(D)\right| \\ &+|\mathcal{K}_{i,n}|^{-1}\mathbf{E}\left|\sum_{k\in\mathcal{K}_{i,n}}(Y_{n,k}(D)S_n^k(D)-\mathbf{E}(Y_{n,k}(D)S_n^k(D)))\right| \\ &+|\mathcal{K}_{i,n}|^{-1}\sum_{k\in\mathcal{K}_{i,n}}\left|\mathbf{E}\left(Y_{n,k}(D)S_n^k(D)\right)-\mathbf{E}\left(Y_{n,k}(D_n)S_n^k(D_n)\right)\right|. \qquad (49)\end{aligned}$$

Up to the edge effects which are negligible with respect to $|\mathcal{K}_{i,n}|$, $\left(Y_{n,k}(D)S_n^k(D)\right)_k$ is stationary when $k\in\mathcal{K}_{i,n}$, since in this case, from (41), $W_{i,n,\Delta_k(D)} = W^{(1)}_{\Delta_k(D)}$ does not depend on n. Therefore the second term in (49) tends to 0 by the mean ergodic theorem. For a fixed n, we have also by stationarity (up to the edge effects)

$$\begin{aligned}\mathbf{E}\left|Y_{n,k}(D_n)S_n^k(D_n)-Y_{n,k}(D)S_n^k(D)\right| &= \mathbf{E}\left|Y_{n,0}(D_n)S_n^0(D_n)-Y_{n,0}(D)S_n^0(D)\right| \\ &= \mathbf{E}\left|W^{(1)}_{\Delta_0(D_n)}W^{(2)}_{\overline{\Delta}_0(D_n)}-W^{(1)}_{\Delta_0(D)}W^{(2)}_{\overline{\Delta}_0(D)}\right|,\end{aligned}$$

where $\overline{\Delta}_0(D_n) = \cup_{k'\in\mathbb{B}_0(1)}\Delta_{k'}(D_n)$. From Lemma 20, this term tends to 0, therefore the first term in (49) asymptotically vanishes. The same argument shows that the third term in (49) also tends to 0 as $n\to\infty$. As a consequence, E_1 tends to 0.

The same decomposition as in (49) may be done for E_2 and E_3, which leads by similar arguments to $E_2\to 0$ and $E_3\to 0$. The last term E_4 involves the difference between Σ_{ij} and its empirical counterpart. The same calculations as in the first step of the proof shows that $E_4\to 0$.

Therefore, we have proved that the terms in the double-sum (48) corresponding to $i\neq j$ asymptotically vanish. The same result can be proved similarly when $i=j$. Thus assumption (iv) in Theorem 21 holds and the convergence in law is deduced.

8.4 Proof of Proposition 6

We can decompose Λ_n in the following way :

$$\Lambda_n := \bigcup_{k \in \mathcal{K}_n} \Delta_k(D_n)$$

where the Δ_k's are disjoint cubes with side-length D_n and $\mathcal{K}_n \subset \mathbb{Z}^d$ satisfies $|\mathcal{K}_n| = |\Lambda_n| D_n^{-d}$. Similarly as in the proof of Proposition 4, we choose

$$D_n = \frac{|\Lambda_n|^{1/d}}{\left\lfloor \frac{|\Lambda_n|^{1/d}}{D} \right\rfloor},$$

which implies $D_n \to D$ when $n \to \infty$ and guarantees $D \leq D_n \leq 2D$ as soon as $|\Lambda_n| \geq D^d$.

From Proposition 3 and under Assumption **[E2(bis)]**, for all $i = 1, \ldots, s$,

$$\begin{aligned} |\Lambda_n|^{-1/2} R_{\Lambda_n}\left(\Phi; h_i, \widehat{\theta}_n(\Phi)\right) &= |\Lambda_n|^{-1/2} R_{\infty,\Lambda_n}\left(\Phi; h_i, \theta^\star\right) + o_P(1) \\ &= |\Lambda_n|^{-1/2}\left(I_{\Lambda_n}\left(\Phi; h_i, \theta^\star\right) - \mathbf{U}_{\Lambda_n}\left(\Phi; \theta^\star\right)^T \mathcal{E}(h_i; \theta^\star)\right) + o_P(1) \\ &= \frac{1}{D_n^{d/2}} \frac{1}{|\mathcal{K}_n|^{1/2}} \sum_{k \in \mathcal{K}_n} W_{\Delta_k(D_n)}\left(\Phi; h_i, \theta^\star\right) + o_P(1), \end{aligned}$$

where for any $\varphi \in \Omega$

$$W_{\Delta_k(D_n)}\left(\varphi; h_i, \theta^\star\right) := I_{\Delta_k(D_n)}\left(\varphi; h_i, \theta^\star\right) + \mathbf{U}_{\Delta_k(D_n)}\left(\varphi; \theta^\star\right)^T \mathcal{E}(h_i; \theta^\star).$$

We apply Theorem 21 in the simpler case when $f_{n,k} = f$ for all $n \in \mathbb{N}$ and all $k \in \mathcal{K}_n$. If $D_n = D$ for all n, this framework would reduce to a stationary setting similar to Theorem 2.1 in Jensen and Künsch (1994). But as Λ_n is allowed to increase continuously up to $\mathbb{R}^d$, $D_n \equiv D$ is impossible. We will therefore apply Theorem 21 in Appendix A with $\mathbf{Z}_{n,k} = \left(W_{\Delta_k(D_n)}\left(\Phi; h_j, \theta^\star\right)\right)_{j=1\ldots s}$, $X_{n,i} = \Phi_{\Delta_i(D_n)}$ and $p = s$.

Let us first compute the covariance matrix of $\left(|\Lambda_n|^{-1/2} R_{\infty,\Lambda_n}\left(\Phi; h_i, \theta^\star\right)\right)_{i=1,\ldots,s}$. By the same calculations as for the term S_1 in the proof of Proposition 4, we obtain

$$\begin{aligned} &cov\left(|\Lambda_n|^{-1/2} R_{\infty,\Lambda_n}\left(\Phi;\ (h_i, \theta^\star), |\Lambda_n|^{-1/2} R_{\infty,\Lambda_n}\left(\Phi; h_j, \theta^\star\right)\right)\right. \\ &\qquad = \frac{1}{D_n^d} \mathbf{E}\left(\frac{1}{|\mathcal{K}_n|} \sum_{k \in \mathcal{K}_n} \sum_{k' \in \mathcal{K}_n} W_{\Delta_k(D_n)}\left(\Phi; h_i, \theta^\star\right) W_{\Delta_{k'}(D_n)}\left(\Phi; h_j, \theta^\star\right)\right) \\ &\qquad \sim \frac{1}{D^d} \sum_{k \in \mathbb{B}_0(1)} \mathbf{E}\left(W_{\Delta_0(D)}\left(\Phi; h_i, \theta^\star\right) W_{\Delta_k(D)}\left(\Phi; h_j, \theta^\star\right)\right). \end{aligned} \tag{50}$$

The asymptotic covariance matrix is thus $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ defined in Proposition 6. We can now apply Theorem 21 in the appendix with $\underline{\mathbf{\Sigma}} = \underline{\mathbf{\Sigma}}_2(\theta^\star)$. The assumption (54) hods because $D_n \geq D$ and from **[N4]** and **[E2(bis)]**. The assumptions (i), (ii) and (iii) follow from **[E2(bis)]**, **[N2]** and Lemma 19. Assumption (iv) may be checked easily as in the second step of the proof of Proposition 4, by using (50).

8.5 Proof of Lemma 8

For simplicity, let $\mathbf{Y}_\Lambda := \mathbf{Y}_\Lambda(\Phi; \theta)$. Let us denote $\overline{\Delta}(\delta, D^\vee) := \cup_{|j| \leq \lceil D^\vee/\delta \rceil} \Delta_j(\delta)$. From the additivity property of $\mathbf{Y}$, proving Lemma 8 reduces to prove that for any $\delta > 0$ and any $D^\vee \geq D$, $D^d A(\delta, D^\vee) = \delta^d A(D, D)$ where

$$A(\delta, D^\vee) := \mathbf{E}\left(\mathbf{Y}_{\Delta_0(\delta)} \mathbf{Y}_{\overline{\Delta}(\delta, D^\vee)}{}^T\right).$$

Since $D^\vee \geq D$, we can write $\overline{\Delta}(\delta, D^\vee) = \overline{\Delta}(\delta, D) \cup \Delta'$, where $\Delta' \subset \left(\overline{\Delta}(\delta, D)\right)^c$. From the locality assumption, $\mathbf{Y}_{\Delta'}$ is only a function of $\Phi_{\Delta_0^c(\delta)}$. So

$$\mathbf{E}\left(\mathbf{Y}_{\Delta_0(\delta)}\mathbf{Y}_{\Delta'}{}^T\right) = \mathbf{E}\left(\mathbf{E}\left(\mathbf{Y}_{\Delta_0(\delta)}\mathbf{Y}_{\Delta'}{}^T|\Phi_{\Delta_0^c(\delta)}\right)\right) = \mathbf{E}\left(\mathbf{E}\left(\mathbf{Y}_{\Delta_0(\delta)}|\Phi_{\Delta_0^c(\delta)}\right)\mathbf{Y}_{\Delta'}{}^T\right) = 0, \quad (51)$$

which yields $A(\delta, D^\vee) = A(\delta, D)$. By denoting $A(\delta) := A(\delta, D^\vee) = A(\delta, D)$ and $\overline{\Delta}(\delta) := \overline{\Delta}(\delta, D)$, we must prove $D^d A(\delta) = \delta^d A(D)$.

Let us first assume $\delta = kD$ with $k \in \mathbb{N}$. We may write $\overline{\Delta}(kD) = (\Delta_0(kD) \oplus D) \cup \Delta'$ and may assert that $\mathbf{Y}_{\Delta'}$ depends only on a function of $\Phi_{\Delta_0^c(kD)}$. By a similar argument as in (51), we obtain $A(\delta) = \mathbf{E}(\mathbf{Y}_{\Delta_0(kD)}\mathbf{Y}_{\Delta_0(kD)\oplus D}{}^T)$. From the disjoint decomposition $\Delta_0(kD) = \cup_{j\in\mathcal{K}}\Delta_j(D)$ where $|\mathcal{K}| = k^d$, we have, by the same decorrelation argument as above and by stationarity,

$$\begin{aligned} A(\delta) &= \sum_{j\in\mathcal{K}} \mathbf{E}(\mathbf{Y}_{\Delta_j(D)}\mathbf{Y}_{\Delta_0(kD)\oplus D}{}^T) = \sum_{j\in\mathcal{K}} \mathbf{E}(\mathbf{Y}_{\Delta_j(D)}\mathbf{Y}_{\Delta_j(D)\oplus D}{}^T) \\ &= k^d \mathbf{E}(\mathbf{Y}_{\Delta_0(D)}\mathbf{Y}_{\Delta_0(D)\oplus D}{}^T) = \frac{\delta^d}{D^d} A(D). \end{aligned}$$

Let us now assume $D = k\delta$ with $k \in \mathbb{N}$. First notice that in this case $\overline{\Delta}(D) = \overline{\Delta}(\delta) \oplus \frac{D}{2}(1-1/k)$. The following decomposition holds: $\Delta_0(D) = \cup_{j\in\mathcal{K}}\Delta_j(\delta)$ where $|\mathcal{K}| = k^d$. For any $j \in \mathcal{K}$, $|j| \leq \frac{D}{2}(1-1/k)$, so $\overline{\Delta}(D)$ contains any translation of the set $\overline{\Delta}(\delta)$ with respect to j. Let us denote this translated set by $\tau_j\overline{\Delta}(\delta)$. From the same decorrelation argument as above and by stationarity, we have

$$A(D) = \sum_{j\in\mathcal{K}} \mathbf{E}(\mathbf{Y}_{\Delta_j(\delta)}\mathbf{Y}_{\overline{\Delta}(D)}{}^T) = \sum_{j\in\mathcal{K}} \mathbf{E}(\mathbf{Y}_{\Delta_j(\delta)}\mathbf{Y}_{\tau_j\overline{\Delta}(\delta)}{}^T) = \sum_{j\in\mathcal{K}} \mathbf{E}(\mathbf{Y}_{\Delta_0(\delta)}\mathbf{Y}_{\overline{\Delta}(\delta)}{}^T) = \frac{D^d}{\delta^d} A(\delta). \quad (52)$$

Let us now consider the case $D/\delta = k'/k$, where $(k, k') \in \mathbb{N}^2$. Let $\delta' = \delta/k$, then $D = k'\delta'$ and according to (52), $\delta'^d A(D) = D^d A(\delta')$. In the same way as we have proved $D^d A(\delta) = \delta^d A(D)$ when $\delta = kD$, this is not difficult to show that for any $\delta = k\delta'$ with $\delta' \leq D$, $\delta'^d A(\delta) = \delta^d A(\delta')$. As a consequence when $D/\delta = k'/k$, we obtain

$$A(D) = \frac{D^d}{\delta'^d} A(\delta') = \frac{D^d}{\delta^d} A(\delta). \quad (53)$$

In the general case, one may find a sequence of rational numbers $(q_n)_{n\in\mathbb{N}}$ which converges to D/δ. Let $\delta_n = q_n D$, we have from (53), $A(D) = \frac{D^d}{\delta_n^d} A(\delta_n)$. Since we have assumed $\mathbf{E}(\mathbf{Y}_{\Gamma_n}^2) \to 0$ when $\Gamma_n \to 0$, the additivity of $\mathbf{Y}$ and $\delta_n \to \delta$ yield

$$A(\delta_n) = \mathbf{E}\left(\mathbf{Y}_{\Delta_0(\delta_n)}\mathbf{Y}_{\overline{\Delta}(\delta_n)}{}^T\right) \to \mathbf{E}\left(\mathbf{Y}_{\Delta_0(\delta)}\mathbf{Y}_{\overline{\Delta}(\delta)}{}^T\right) = A(\delta)$$

as n goes to infinity. Therefore, the identity (53) holds for any $\delta > 0$, which concludes the proof.

8.6 Proof of Proposition 7

The proof follows arguments presented by Jensen and Künsch in Jensen and Künsch (1994). Let $C_n(\overline{\delta}) = [-n\overline{\delta} - \overline{\delta}/2, n\overline{\delta} + \overline{\delta}/2]^d$, so $C_n(\overline{\delta}) = \cup_{k\in\mathcal{K}_n}\Delta_k(\overline{\delta})$, where $\mathcal{K}_n = [-n, n]^d \cap \mathbb{Z}^d$ and $\Delta_k(\overline{\delta})$ is the cube centered at $k\overline{\delta}$ with side-length $\overline{\delta}$. We have

$$\begin{aligned} Var\left(|C_n(\overline{\delta})|^{-1/2}\mathbf{Y}_{C_n(\overline{\delta})}(\Phi;\theta^\star)\right) &= |C_n(\overline{\delta})|^{-1} \sum_{i,j\in\mathcal{K}_n} \mathbf{E}\left(\mathbf{Y}_{\Delta_i(\overline{\delta})}(\Phi;\theta^\star)\,\mathbf{Y}_{\Delta_j(\overline{\delta})}(\Phi;\theta^\star)^T\right) \\ &= |C_n(\overline{\delta})|^{-1} \sum_{i\in\mathcal{K}_n} \sum_{j\in\mathbb{B}_i\left(\lceil\frac{D}{\delta}\rceil\right)\cap\mathcal{K}_n} \mathbf{E}\left(\mathbf{Y}_{\Delta_i(\overline{\delta})}(\Phi;\theta^\star)\,\mathbf{Y}_{\Delta_j(\overline{\delta})}(\Phi;\theta^\star)^T\right). \end{aligned}$$

Since $|C_n(\overline{\delta})| = \overline{\delta}^d |\mathcal{K}_n|$, from the ergodic theorem,

$$Var\left(|C_n(\overline{\delta})|^{-1/2}\mathbf{Y}_{C_n(\overline{\delta})}(\Phi;\theta^\star)\right) \longrightarrow \overline{\delta}^{-d} \sum_{|k|\leq\left\lceil\frac{D}{\overline{\delta}}\right\rceil} \mathbf{E}\left(\mathbf{Y}_{\Delta_0(\overline{\delta})}(\Phi;\theta^\star)\,\mathbf{Y}_{\Delta_k(\overline{\delta})}(\Phi;\theta^\star)^T\right)$$

which is $\underline{\mathbf{M}}(\theta^\star)$ by Lemma 8.

Therefore, to prove that $\underline{\mathbf{M}}(\theta^\star)$ is positive-definite, it is sufficient to prove that the covariance matrix $Var\left(|C_n(\overline{\delta})|^{-1/2}\mathbf{Y}_{C_n(\overline{\delta})}(\Phi;\theta^\star)\right)$ is positive-definite for n large enough. Let $\mathbf{x}\in\mathbb{R}^q\setminus\{0\}$, we must show that

$$V := \mathbf{x}^T Var\left(|C_n(\overline{\delta})|^{-1/2}\mathbf{Y}_{C_n(\overline{\delta})}(\Phi;\theta^\star)\right)\mathbf{x} > 0.$$

Since, for two random variables X, X' with finite variance

$$Var(X) = \mathbf{E}(Var(X|X')) \;+\; Var(\mathbf{E}(X|X')) \geq \mathbf{E}(Var(X|X')),$$

we have, by denoting $L := \left(2\left\lceil\frac{D}{\overline{\delta}}\right\rceil+1\right)\mathbb{Z}^d$,

$$\begin{aligned} V &\geq |C_n(\overline{\delta})|^{-1}\,\mathbf{E}\left(Var\left(\mathbf{x}^T\mathbf{Y}_{C_n(\overline{\delta})}(\Phi;\theta^\star)|\;\Phi_{\Delta_k(\overline{\delta})}, k\notin L\right)\right)\\ &= |C_n(\overline{\delta})|^{-1}\mathbf{x}^T\mathbf{E}\Bigg(Var\Bigg(\sum_{\ell\in L\cap\mathcal{K}_n}\underbrace{\sum_{i\in\mathbb{B}_\ell\left(\left\lceil\frac{D}{\overline{\delta}}\right\rceil\right)\cap\mathcal{K}_n}\mathbf{Y}_{\Delta_i(\overline{\delta})}(\Phi;\theta^\star)}_{:=\mathbf{S}_{\ell,n}(\Phi)}\;|\;\Phi_{\Delta_k(\overline{\delta})}, k\notin L\Bigg)\Bigg)\mathbf{x}\end{aligned}$$

Note that from the locality property, $\mathbf{S}_{\ell,n}(\Phi)$ depends only on $\Phi_{\Delta_j(\overline{\delta})}$ for $j\in\mathbb{B}_\ell\left(2\left\lceil\frac{D}{\overline{\delta}}\right\rceil\right)$. Therefore, conditionally on $\Phi_{\Delta_k(\overline{\delta})}, k\notin L$, the variables $\mathbf{S}_{\ell,n}(\Phi)$ and $\mathbf{S}_{\ell',n}(\Phi)$ (for $\ell\neq\ell'$) are independent. Now, let $\overline{\Delta}(\overline{\delta}) := \cup_{|i|\leq\left\lceil\frac{D}{\overline{\delta}}\right\rceil}\Delta_i(\overline{\delta})$, from the stationarity we have for n large enough

$$\begin{aligned} V &\geq |C_n(\overline{\delta})|^{-1}\mathbf{x}^T\sum_{\ell\in L\cap\mathcal{K}_n}\mathbf{E}\Bigg(Var\Bigg(\mathbf{S}_{\ell,n}(\Phi)|\;\Phi_{\Delta_k(\overline{\delta})}, k\notin L\Bigg)\Bigg)\mathbf{x}\\ &\geq \frac{\overline{\delta}^{-d}}{2}\frac{|L\cap\mathcal{K}_n|}{|\mathcal{K}_n|}\times\mathbf{E}\left(Var\left(\mathbf{x}^T\mathbf{Y}_{\overline{\Delta}(\overline{\delta})}(\Phi;\theta^\star)\,\middle|\Phi_{\Delta_k(\overline{\delta})}, 1\leq|k|\leq 2\left\lceil\frac{D}{\overline{\delta}}\right\rceil\right)\right)\\ &\geq \kappa(\overline{\delta},D,d)\times\mathbf{E}\left(Var\left(\mathbf{x}^T\mathbf{Y}_{\overline{\Delta}(\overline{\delta})}(\Phi;\theta^\star)\,\middle|\Phi_{\Delta_k(\overline{\delta})}, 1\leq|k|\leq 2\left\lceil\frac{D}{\overline{\delta}}\right\rceil\right)\right),\end{aligned}$$

where $\kappa(\overline{\delta},D,d)$ is a positive constant. Assume there exists some positive constant c such that $P_{\theta^\star}-$a.s. $\mathbf{x}^T\mathbf{Y}_{\overline{\Delta}(\overline{\delta})}(\Phi;\theta^\star) = c$ when the variables $\Phi_{\Delta_k(\overline{\delta})}, 1\leq|k|\leq 2\left\lceil\frac{D}{\overline{\delta}}\right\rceil$ are fixed to belong to B, where $B\in\mathcal{F}$ is involved in **[PD]**. It follows that for any $\varphi_i\in A_i$ for $i=0,\ldots,\ell$ (with $\ell\geq 1$), where the A_i's come from **[PD]**, $\mathbf{x}^T\left(\mathbf{Y}_{\overline{\Delta}(\overline{\delta})}(\varphi_i;\theta^\star)-\mathbf{Y}_{\overline{\Delta}(\overline{\delta})}(\varphi_0;\theta^\star)\right)=0$. Since for all $(\varphi_0,\ldots,\varphi_\ell)\in A_0\times\ldots\times A_\ell$, the matrix with entries $\left(\mathbf{Y}_{\overline{\Delta}(\overline{\delta})}(\varphi_i;\theta^\star)\right)_j - \left(\mathbf{Y}_{\overline{\Delta}(\overline{\delta})}(\varphi_0;\theta^\star)\right)_j$ is assumed to be injective, this leads to $\mathbf{x}=0$ and hence to some contradiction. Therefore, when the variables $\Phi_{\Delta_k(\overline{\delta})}$, $1\leq|k|\leq 2\left\lceil\frac{D}{\overline{\delta}}\right\rceil$ are for example assumed to belong to B, the variable $\mathbf{x}^T\mathbf{Y}_{\overline{\Delta}(\overline{\delta})}(\Phi;\theta^\star)$ is almost surely not a constant and so $V>0$, which proves that $\underline{\mathbf{M}}(\theta^\star)$ is a symmetric positive-definite matrix.

8.7 Proof of Proposition 9

Since for any $\varphi \in \Omega$, $\widehat{\underline{\mathbf{M}}}_n(\varphi;\cdot,\delta,D^\vee)$ is continuous in a neighborhood $\mathcal{V}(\theta^\star)$ of $\theta^\star$ and according to [**E1**], it is sufficient to prove that for any $\theta \in \mathcal{V}(\theta^\star)$, $\widehat{\underline{\mathbf{M}}}_n(\Phi;\theta,\delta_n,D^\vee)$ converges in probability towards $\underline{\mathbf{M}}(\theta)$.

We choose the sequence δ_n as follows :

$$\delta_n = \frac{|\Lambda_n|^{1/d}}{\left\lfloor \frac{|\Lambda_n|^{1/d}}{\delta} \right\rfloor},$$

which guarantees $\delta_{n_0} = \delta$, since $|\Lambda_{n_0}|\delta^{-d} \in \mathbb{N}$, $\delta \le \delta_n \le 2\delta$ for n sufficiently large, and $\delta_n \to \delta$ as $n \to \infty$. This choice allows us to consider, for all $n \in \mathbb{N}$, the decomposition $\Lambda_n = \cup_{k\in\mathcal{K}_n}\Delta_k(\delta_n)$, where the $\Delta_k(\delta_n)$'s are disjoint cubes with side-length δ_n and centered at $k\delta_n$. Moreover, since $\delta_n \ge \delta$ and $\delta_n \to \delta$, we have $\left\lceil \frac{D^\vee}{\delta_n} \right\rceil = \left\lceil \frac{D^\vee}{\delta} \right\rceil$ when n is large enough, which is assumed in the sequel of the proof.

Let $\widetilde{\mathcal{K}}_n := \mathcal{K}_n \cap \left(\cup_{j\in\partial\mathcal{K}_n}\mathbb{B}_j\left(\left\lceil \frac{D^\vee}{\delta}\right\rceil\right)\right)$. Since $|\Lambda_n| = \delta_n^d|\mathcal{K}_n|$, we have

$$\left|\delta_n^d\widehat{\underline{\mathbf{M}}}_n(\Phi;\theta,\delta_n,D^\vee) - \delta^d\underline{\mathbf{M}}(\theta)\right| \le X_1 + X_2 + X_3 + X_4,$$

where by setting $\overline{\Delta}_k(\tau) = \cup_{j\in\mathbb{B}_k(\lceil\frac{D^\vee}{\delta}\rceil)}\Delta_j(\tau)$ (for some $\tau > 0$),

$$\begin{aligned}
X_1 &= \left||\mathcal{K}_n|^{-1}\sum_{k\in\widetilde{\mathcal{K}}_n}\sum_{j\in\mathbb{B}_k(\lceil\frac{D^\vee}{\delta}\rceil)\cap\mathcal{K}_n}\widehat{\mathbf{Y}}_{n,\Delta_k(\delta_n)}(\Phi;\theta)\,\widehat{\mathbf{Y}}_{n,\Delta_j(\delta_n)}(\Phi;\theta)^T\right| \\
X_2 &= \left||\mathcal{K}_n|^{-1}\sum_{k\in\mathcal{K}_n\setminus\widetilde{\mathcal{K}}_n}\left(\widehat{\mathbf{Y}}_{n,\Delta_k(\delta_n)}(\Phi;\theta)\,\widehat{\mathbf{Y}}_{n,\overline{\Delta}_k(\delta_n)}(\Phi;\theta)^T - \mathbf{Y}_{\Delta_k(\delta_n)}(\Phi;\theta)\,\mathbf{Y}_{\overline{\Delta}_k(\delta_n)}(\Phi;\theta)^T\right)\right| \\
X_3 &= \left||\mathcal{K}_n|^{-1}\sum_{k\in\mathcal{K}_n\setminus\widetilde{\mathcal{K}}_n}\left(\mathbf{Y}_{n,\Delta_k(\delta_n)}(\Phi;\theta)\,\mathbf{Y}_{n,\overline{\Delta}_k(\delta_n)}(\Phi;\theta)^T - \mathbf{Y}_{\Delta_k(\delta)}(\Phi;\theta)\,\mathbf{Y}_{\overline{\Delta}_k(\delta)}(\Phi;\theta)^T\right)\right| \\
X_4 &= \left||\mathcal{K}_n|^{-1}\sum_{k\in\mathcal{K}_n\setminus\widetilde{\mathcal{K}}_n}\mathbf{Y}_{\Delta_k(\delta)}(\Phi;\theta)\,\mathbf{Y}_{\overline{\Delta}_k(\delta)}(\Phi;\theta)^T - \mathbf{E}\left(\mathbf{Y}_{\Delta_0(\delta)}(\Phi;\theta)\,\mathbf{Y}_{\overline{\Delta}_0(\delta)}(\Phi;\theta)^T\right)\right|.
\end{aligned}$$

We have from the additivity and the stationartiy of $\widehat{\mathbf{Y}}_n$,

$$\begin{aligned}
\mathbf{E}|X_1| &\le |\mathcal{K}_n|^{-1}\sum_{k\in\widetilde{\mathcal{K}}_n}\sum_{j\in\mathbb{B}_k(\lceil\frac{D^\vee}{\delta}\rceil)\cap\mathcal{K}_n}\mathbf{E}\left|\widehat{\mathbf{Y}}_{n,\Delta_k(\delta_n)}(\varphi;\theta)\,\widehat{\mathbf{Y}}_{n,\Delta_j(\delta_n)}(\varphi;\theta)^T\right| \\
&\le \frac{|\widetilde{\mathcal{K}}_n|}{|\mathcal{K}_n|}\mathbf{E}\left|\widehat{\mathbf{Y}}_{n,\Delta_0(\delta_n)}(\Phi;\theta)\widehat{\mathbf{Y}}_{n,\overline{\Delta}_0(\delta_n)}(\Phi;\theta)^T\right|,
\end{aligned}$$

which tends to 0 as $n \to +\infty$, because $\frac{|\widetilde{\mathcal{K}}_n|}{|\mathcal{K}_n|} \to 0$ and $\delta \le \delta_n \le 2\delta$. Therefore, X_1 converges in probability to 0. The second term converges also to 0 in probability from the additivity of $\mathbf{Y}$ and $\widehat{\mathbf{Y}}$ and from (16). The expectation of the third term converges to 0 by following the proof of Lemma 20. Finally, from the stationarity of $\mathbf{Y}$ and since $|\mathcal{K}_n| \sim |\mathcal{K}_n \setminus \widetilde{\mathcal{K}}_n|$, the mean ergodic theorem applies to $\mathbf{E}|X_4|$, which, in particular, shows that $X_4 \to 0$ in probability. This proves that

$$\delta_n^d\widehat{\underline{\mathbf{M}}}_n(\Phi;\theta,\delta_n,D^\vee) \longrightarrow \delta^d\underline{\mathbf{M}}(\theta),$$

in probability, as $n \to \infty$. Since δ_n is a deterministic sequence converging to δ, the conclusion of Proposition 9 follows.

A Central Limit Theorem

The following result is a central limit theorem for conditionnally centered random fields. It generalizes Theorem 2.1 in Jensen and Künsch (1994) to a non-stationary and non-ergodic setting. A general result has been proved by Comets and Janzura (1998) for self normalized sums, provided a fourth moment condition. Our result is in the same spirit but it is proved for triangular array and without self-normalization, which is well-adapted to the residuals process framework. This allows in particular to avoid the fourth moment assumption.

Theorem 21. *Let $X_{n,i}$, $n \in \mathbb{N}$, $i \in \mathbb{Z}^d$, be a triangular array field in a measurable space S. For $n \in \mathbb{N}$, let $\mathcal{K}_n \subset \mathbb{Z}^d$ and for $k \in \mathcal{K}_n$, assume*

$$\mathbf{Z}_{n,k} = f_{n,k}\left(X_{n,k+i},\ i \in \mathcal{I}_0\right), \tag{54}$$

where $\mathcal{I}_0 = \{i \in \mathbb{Z}^d,\ |i| \leq 1\}$ and $f_{n,k} : S^{\mathcal{I}_0} \to \mathbb{R}^p$. Let $\mathbf{S}_n = \sum_{k\in\mathcal{K}_n} \mathbf{Z}_{n,k}$. If

(i) $c_3 := \sup_{n\in\mathbb{N}} \sup_{k\in\mathcal{K}_n} \mathbf{E}|\mathbf{Z}_{n,k}|^3 < \infty$,

(ii) $\forall n \in \mathbb{N}$, $\forall k \in \mathcal{K}_n$, $\mathbf{E}(\mathbf{Z}_{n,k}|X_{n,j},\ j \neq k) = 0$,

(iii) $|\mathcal{K}_n| \to +\infty$ *as* $n \to \infty$,

(iv) There exists a symmetric matrix $\underline{\mathbf{\Sigma}} \geq 0$ such that

$$\mathbf{E}\left\| |\mathcal{K}_n|^{-1} \sum_{k\in\mathcal{K}_n} \sum_{j\in\mathbb{B}_k(1)\cap\mathcal{K}_n} \mathbf{Z}_{n,k}\mathbf{Z}_{n,j}{}^T - \underline{\mathbf{\Sigma}} \right\| \to 0,$$

then $|\mathcal{K}_n|^{-1/2}\mathbf{S}_n \xrightarrow{d} \mathcal{N}(0, \underline{\mathbf{\Sigma}})$ *as* $n \to \infty$.

Proof.

Let us first assume that $\underline{\mathbf{\Sigma}}$ is a positive-definite matrix (i.e. $\underline{\mathbf{\Sigma}} > 0$). According to the Stein's method (see also Bolthausen (1982)), it suffices to prove that, for all $\mathbf{e} \in \mathbb{R}^p$ such that $\|\mathbf{e}\| = 1$ and for all $\lambda \in \mathbb{R}$,

$$\mathbf{E}\left(\left(i\lambda - \mathbf{e}^T|\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}\mathbf{S}_n\right)\exp\left(i\lambda\mathbf{e}^T|\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}\mathbf{S}_n\right)\right) \to 0.$$

Denoting $\mathbf{u} = \lambda\mathbf{e}$, this is equivalent to prove that for all $\mathbf{u} \in \mathbb{R}^p$,

$$\mathbf{E}\left(\underbrace{\left(i\mathbf{u} - |\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}\mathbf{S}_n\right)\exp(i\mathbf{u}^T|\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}\mathbf{S}_n)}_{:=\mathbf{A}}\right) \to 0.$$

We decompose the term $\mathbf{A}$ in the same spirit as in Bolthausen (1982), Jensen and Künsch (1994) and Comets and Janzura (1998). Let us denote by $\underline{\mathbf{I}}_p$ the identity matrix of size p and $\mathbf{S}_n^k = \sum_{j\in\mathbb{B}_k(1)\cap\mathcal{K}_n} \mathbf{Z}_{n,j}$. Noting that $\mathbf{u}^T\underline{\mathbf{\Sigma}}^{-1/2}\mathbf{S}_n^k = \mathbf{S}_n^{kT}\underline{\mathbf{\Sigma}}^{-1/2^T}\mathbf{u}$, the decomposition is $\mathbf{E}(\mathbf{A}) = \mathbf{E}(\mathbf{A}_1 - \mathbf{A}_2 - \mathbf{A}_3)$ where

$$\mathbf{A}_1 = i\exp(i\mathbf{u}^T|\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}\mathbf{S}_n)\left[\underline{\mathbf{I}}_p - |\mathcal{K}_n|^{-1}\underline{\mathbf{\Sigma}}^{-1/2}\sum_{k\in\mathcal{K}_n}\mathbf{Z}_{n,k}\mathbf{S}_n^{kT}\right]\mathbf{u},$$

$$\mathbf{A}_2 = \exp(i\mathbf{u}^T|\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}\mathbf{S}_n)|\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2} \times \sum_{k\in\mathcal{K}_n}\mathbf{Z}_{n,k}\left(1 - \exp(-i\mathbf{u}^T|\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}\mathbf{S}_n^k) - i\mathbf{u}^T|\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}\mathbf{S}_n^k\right),$$

$$\mathbf{A}_3 = |\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}\sum_{k\in\mathcal{K}_n}\mathbf{Z}_{n,k}\exp\left[i\mathbf{u}^T|\mathcal{K}_n|^{-1/2}\underline{\mathbf{\Sigma}}^{-1/2}(\mathbf{S}_n - \mathbf{S}_n^k)\right].$$

The two last terms $\mathbf{A}_2$ and $\mathbf{A}_3$ can be handled as in Jensen and Künsch (1994): $\mathbf{E}(\mathbf{A}_3) = 0$ by (ii) and $|\mathbf{E}(\mathbf{A}_2)| \to 0$ from the same inequalities therein and the sub-multiplicative property of the Frobenius norm. These inequalities rely on two facts: $\forall x \in \mathbb{R}, |1 - e^{-ix} - ix| \leq x^2/2$ and for all n and for all $(k_1, k_2, k_3) \in \mathcal{K}_n$, $\mathbf{E}(|\mathbf{Z}_{n,k_1}||\mathbf{Z}_{n,k_2}||\mathbf{Z}_{n,k_3}|) \leq \left(\mathbf{E}(|\mathbf{Z}_{n,k_1}|^3)\mathbf{E}(|\mathbf{Z}_{n,k_2}|^3)\mathbf{E}(|\mathbf{Z}_{n,k_3}|^3)\right)^{1/3}$ which is less than c_3 by (i).

For $\mathbf{A}_1$, we cannot use a mean ergodic theorem as in Jensen and Künsch (1994), but Assumption (iv) is sufficient. Indeed,

$$\begin{aligned}\|\mathbf{E}(\mathbf{A}_1)\| &\leq \|\mathbf{u}\|\mathbf{E}\left\|\underline{\mathbf{I}}_p - |\mathcal{K}_n|^{-1}\underline{\mathbf{\Sigma}}^{-1/2}\sum_{k\in\mathcal{K}_n}\sum_{j\in\mathbb{B}_k(1)\cap\mathcal{K}_n}\mathbf{Z}_{n,k}\mathbf{Z}_{n,j}{}^T\underline{\mathbf{\Sigma}}^{-1/2^T}\right\| \\ &\leq \|\mathbf{u}\|\left\|\underline{\mathbf{\Sigma}}^{-1/2}\right\|^2\mathbf{E}\left\||\mathcal{K}_n|^{-1}\sum_{k\in\mathcal{K}_n}\sum_{j\in\mathbb{B}_k(1)\cap\mathcal{K}_n}\mathbf{Z}_{n,k}\mathbf{Z}_{n,j}{}^T - \underline{\mathbf{\Sigma}}\right\|\end{aligned}$$

which tends to 0 by (iv).

Now, if $\underline{\mathbf{\Sigma}}$ is not a positive-definite matrix, one can find an orthonormal basis $(\mathbf{f}_1, \ldots, \mathbf{f}_p)$ of $\mathbb{R}^p$, where the $\mathbf{f}_i$'s are eigenvectors of $\underline{\mathbf{\Sigma}}$. We agree that, if $r < p$ denotes the rank of $\underline{\mathbf{\Sigma}}$, then $(\mathbf{f}_1, \ldots, \mathbf{f}_r)$ is a basis of the image of $\underline{\mathbf{\Sigma}}$, while $(\mathbf{f}_{r+1}, \ldots, \mathbf{f}_p)$ is a basis of its kernel.

Let us denote by $\underline{\mathbf{V}}_{Im}$ the matrix whose columns are $(\mathbf{f}_1, \ldots, \mathbf{f}_r)$ and $\underline{\mathbf{V}}_{Ker}$ the matrix whose columns are $(\mathbf{f}_{r+1}, \ldots, \mathbf{f}_p)$. Similarly, for any $\mathbf{u} \in \mathbb{R}^p$, let us denote by u_i its i-th coordinate in the basis $(\mathbf{f}_1, \ldots, \mathbf{f}_p)$ and $\mathbf{u}_{Im} = (u_1, \ldots, u_r)$, $\mathbf{u}_{Ker} = (u_{r+1}, \ldots, u_p)$. Hence $\mathbf{u} = \underline{\mathbf{V}}_{Im}\mathbf{u}_{Im} + \underline{\mathbf{V}}_{Ker}\mathbf{u}_{Ker}$.

The convergence in law of $|\mathcal{K}_n|^{-1/2}\mathbf{S}_n$ to a Gaussian vector reduces to the convergence of $\mathbf{u}^T|\mathcal{K}_n|^{-1/2}\mathbf{S}_n$ for all $\mathbf{u} \in \mathbb{R}^p$. We have

$$\mathbf{u}^T|\mathcal{K}_n|^{-1/2}\mathbf{S}_n = \mathbf{u}_{Im}{}^T\underline{\mathbf{V}}_{Im}{}^T|\mathcal{K}_n|^{-1/2}\mathbf{S}_n + \mathbf{u}_{Ker}{}^T\underline{\mathbf{V}}_{Ker}{}^T|\mathcal{K}_n|^{-1/2}\mathbf{S}_n. \tag{55}$$

From (iv) and since $\underline{\mathbf{V}}_{Ker}{}^T\underline{\mathbf{\Sigma}}\,\underline{\mathbf{V}}_{Ker} = 0$, we deduce that

$$\mathbf{E}\left\||\mathcal{K}_n|^{-1}\sum_{k\in\mathcal{K}_n}\sum_{j\in\mathbb{B}_k(1)\cap\mathcal{K}_n}\underline{\mathbf{V}}_{Ker}{}^T\mathbf{Z}_{n,k}\mathbf{Z}_{n,j}{}^T\underline{\mathbf{V}}_{Ker}\right\| \longrightarrow 0,$$

which means that $\underline{\mathbf{V}}_{Ker}{}^T|\mathcal{K}_n|^{-1/2}\mathbf{S}_n$ tends to 0 in quadratic mean.

On the other hand, the assumptions of Theorem 21 imply that $(i)-(iv)$ remain true when one replaces $\mathbf{Z}_{n,k}$ by $\underline{\mathbf{V}}_{Im}{}^T\mathbf{Z}_{n,k}$ and $\underline{\mathbf{\Sigma}}$ by $\underline{\mathbf{V}}_{Im}{}^T\underline{\mathbf{\Sigma}}\,\underline{\mathbf{V}}_{Im}$. Since $\underline{\mathbf{V}}_{Im}{}^T\underline{\mathbf{\Sigma}}\,\underline{\mathbf{V}}_{Im}$ is positive-definite, the convergence in law of $\underline{\mathbf{V}}_{Im}{}^T|\mathcal{K}_n|^{-1/2}\mathbf{S}_n$ holds for the same reasons as in the first part of the proof.

Therefore, we have proved that for all $\mathbf{u} \in \mathbb{R}^p$, $\mathbf{u}^T|\mathcal{K}_n|^{-1/2}\mathbf{S}_n \xrightarrow{d} \mathcal{N}\left(0, \mathbf{u}_{Im}{}^T\underline{\mathbf{V}}_{Im}{}^T\underline{\mathbf{\Sigma}}\,\underline{\mathbf{V}}_{Im}\mathbf{u}_{Im}\right)$. It is easy to check that $\mathbf{u}_{Im}{}^T\underline{\mathbf{V}}_{Im}{}^T\underline{\mathbf{\Sigma}}\,\underline{\mathbf{V}}_{Im}\mathbf{u}_{Im} = \mathbf{u}^T\underline{\mathbf{\Sigma}}\mathbf{u}$, which concludes the proof. ■

B Assumption [PD] on two examples

In this section, we focus on the two following models, belonging to the exponential family :

1. **Two-type marked Strauss point process** : $\mathbb{M} = \{1, 2\}$ and for $\theta = \left(\theta_1^1, \theta_1^2, \theta_2^{1,1}, \theta_2^{1,2}, \theta_2^{2,2}\right)^T$, for any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$,

$$V_\Lambda(\varphi;\theta) = \theta_1^1 \underbrace{|\varphi_\Lambda^1|}_{:=v_{\Lambda,1}^1(\varphi)} + \theta_1^2 \underbrace{|\varphi_\Lambda^2|}_{:=v_{\Lambda,1}^2(\varphi)} + \sum_{\substack{m_1,m_2=1\\ m_1\leq m_2}}^{2}\theta_2^{m_1,m_2}\underbrace{\sum_{\substack{\{x_1^{m_1},x_2^{m_2}\}\in\mathcal{P}_2(\varphi)\\ \{x_1^{m_1},x_2^{m_2}\}\cap\Lambda\neq\emptyset}}\mathbf{1}_{[0,D^{m_1,m_2}]}(\|x_2 - x_1\|)}_{:=v_{\Lambda,2}^{m_1,m_2}(\varphi)}.$$

Alternatively,

$$V\left(x^m|\varphi;\theta\right) = \theta_1^m + \sum_{m'=1}^{2} \theta_2^{m,m'} \sum_{y^{m'}\in\varphi} \mathbf{1}_{[D_0^{m_1,m_2},D^{m_1,m_2}]}(\|y-x\|).$$

This process is well-defined when $\theta_2^{m_1,m_2} \geq 0$ and $D_0^{m_1,m_2} = 0$ (inhibition assumption), or when $\theta_2^{m_1,m_2} \in \mathbb{R}^2$ and $D_0^{m_1,m_2} = \delta > 0$ (hard-core assumption), see Proposition 13 for instance. The range of the local energy function equals $D = \max\left(D^{1,1}, D^{1,2}, D^{2,2}\right)$.

2. **Area interaction point process** : $\mathbb{M} = \{0\}$ and for $R > 0$, $\theta = (\theta_1, \theta_2)$ and any $\Lambda \in \mathcal{B}(\mathbb{R}^d)$,

$$V_\Lambda\left(\varphi;\theta\right) = \theta_1|\varphi_\Lambda| + \theta_2 v_2(\varphi_\Lambda), \quad \text{with} \quad v_2(\varphi_\Lambda) := \left|\cup_{x\in\varphi_\Lambda} B(x,R)\right|.$$

Alternatively,

$$v_1(0|\varphi) := 1, \qquad v_2(0|\varphi) := \left|\cup_{x\in(\varphi_{\mathcal{B}(0,2R)}\cup\{0\})}\mathcal{B}(x,R) \setminus \cup_{x\in\varphi_{\mathcal{B}(0,2R)}}\mathcal{B}(x,R)\right|.$$

This model is well-defined for $\theta \in \mathbb{R}^2$ (see Proposition 13 for instance) and the range of the local energy equals $D = 2R$.

Both these models satisfy **[C]** and **[N1-4]**. The aim of the sequel is to prove Proposition 18, which claims that $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ and $\underline{\mathbf{\Sigma}}_2(\theta^\star)$, involved respectively in Proposition 4 and 6, are positive-definite for these models, when considering the maximum pseudolikelihood estimate for $\widehat{\theta}_n$ and the two following frameworks

- Framework 1 (for $\underline{\mathbf{\Sigma}}_1(\theta^\star)$): we consider the inverse residuals ($h = e^V$). Let us recall that Proposition 17 asserts that **[PD]** fails for the raw residuals ($h = 1$) for both the area-interaction and 2-type marked Strauss models.
- Framework 2 (for $\underline{\mathbf{\Sigma}}_2(\theta^\star)$): we consider the family of test functions given for $j = 1, \ldots, s$ and $0 < r_1 < \ldots < r_s < +\infty$ by

$$h_j(x^m, \varphi; \theta) = \mathbf{1}_{[0,r_j]}(d(x^m,\varphi))e^{V(x^m|\varphi;\theta)},$$

related to parametric and nonparametric estimations of the empty space function at distance r_j.

When considering the MPLE, $R_{\infty,\Lambda}(\varphi; h, \theta^\star)$ is given by (26) with $\mathbf{LPL}^{(1)}$, $\underline{\mathbf{H}}$ and $\mathcal{E}$ respectively given by (24), (25) and (7).

B.1 2-type marked Strauss point process

We only deal with the inhibition case, that is $\mathbf{\Theta} = \mathbb{R}^2 \times \mathbb{R}_+^3$ and $D_0^{m_1,m_2} = 0$. The following proofs could easily be extended to the hard-core case and to the multi-Strauss marked point process (see *e.g.* Billiot et al. (2008)). For any vector $\mathbf{z}$ of length 5, we sometimes reparameterize it similarly as the parameter vector, that is $\mathbf{z} = (z_1^1, z_1^2, z_2^{1,1}, z_2^{1,2}, z_2^{2,2})^T$.

B.1.1 Proof that $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ is positive-definite for the two-type Strauss model

From Proposition 7, proving that $\underline{\mathbf{\Sigma}}_1(\theta^\star)$ is positive-definite in Framework 1 reduces to check Assumption **[PD]** with $h = e^V$ and

(i) $\mathbf{Y}_\Lambda(\varphi;\theta^\star) = I_\Lambda(\varphi; e^V, \theta^\star)$,

(ii) $\mathbf{Y}_\Lambda(\varphi;\theta^\star) = R_{\infty,\Lambda}(\varphi; e^V, \theta^\star)$.

(i) is ensured by Proposition 16.

(ii) We fix $\overline{\delta} = D$ and $B = \emptyset$ in **[PD]**. Let $\overline{\Omega} := \overline{\Omega}_{\emptyset}$. Without loss of generality, one may assume that $\theta_2^{\star 1,1} > 0$. Let us define for $n \geq 1$

$$A_{n,-}(\eta) = \Big\{\varphi \in \overline{\Omega} : \varphi(\Delta_0(\overline{\delta}) \times \{1\}) = 2n, \varphi(\Delta_0(\overline{\delta}) \times \{2\}) = 0, \\ \varphi\left(\mathcal{B}\left((0,0), \frac{\eta}{4}\right)\right) = n, \varphi\left(\mathcal{B}\left((D^{1,1} - \frac{\eta}{2}, 0), \frac{\eta}{4}\right)\right) = n\Big\},$$

$$A_{n,+}(\eta) = \Big\{\varphi \in \overline{\Omega} : \varphi(\Delta_0(\overline{\delta}) \times \{1\}) = 2n, \varphi(\Delta_0(\overline{\delta}) \times \{2\}) = 0, \\ \varphi\left(\mathcal{B}\left((0,0), \frac{\eta}{4}\right)\right) = n, \varphi\left(\mathcal{B}\left((D^{1,1} + \frac{\eta}{2}, 0), \frac{\eta}{4}\right)\right) = n\Big\}.$$

Let $\varphi_{n,-} \in A_{n,-}$ and $\varphi_{n,+} \in A_{n,+}$. Then for η small enough

$$I_{\overline{\Lambda}}(\varphi_{n,\bullet}; e^V, \theta^\star) = |\overline{\Lambda}| - \begin{cases} 2ne^{\theta_1^{\star 1} + (2n-1)\theta_2^{\star 1,1}} & \text{if } \bullet = -, \\ 2ne^{\theta_1^{\star 1} + (n-1)\theta_2^{\star 1,1}} & \text{if } \bullet = +. \end{cases}$$

$$\left(\mathbf{LPL}_{\overline{\Lambda}}^{(1)}(\varphi_{n,\bullet}; \theta^\star)\right)_1^{m'} = \int_{\overline{\Lambda}\times\mathbb{M}} v_1^{m'}(x^m|\varphi_{n,\bullet}) e^{-V(x^m|\varphi_{n,\bullet};\theta^\star)} \mu(dx^m) - \begin{cases} 2n & \text{if } m' = 1, \\ 0 & \text{if } m' = 2. \end{cases}$$

$$\left(\mathbf{LPL}_{\overline{\Lambda}}^{(1)}(\varphi_{n,-}; \theta^\star)\right)_2^{m_1',m_2'} = \int_{\overline{\Lambda}\times\mathbb{M}} v_2^{m_1,m_2}(x^m|\varphi_{n,-}) e^{-V(x^m|\varphi_{n,-};\theta^\star)} \mu(dx^m) \\ - \begin{cases} 2n(2n-1) & \text{if } m_1 = m_2 = 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$\left(\mathbf{LPL}_{\overline{\Lambda}}^{(1)}(\varphi_{n,+}; \theta^\star)\right)_2^{m_1,m_2} = \int_{\overline{\Lambda}\times\mathbb{M}} v_2^{m_1,m_2}(x^m|\varphi_{n,+}) e^{-V(x^m|\varphi_{n,+};\theta^\star)} \mu(dx^m) \\ - \begin{cases} n(n-1) & \text{if } m_1 = m_2 = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Now,

$$\begin{aligned} &\Delta R_{\infty,\overline{\Lambda}}(\varphi_{n,-}, \varphi_{n,+}) := R_{\infty,\overline{\Lambda}}(\varphi_{n,-}; e^V, \theta^\star) - R_{\infty,\overline{\Lambda}}(\varphi_{n,+}; e^V, \theta^\star) \\ &= 2n\left(e^{\theta_1^{\star 1} + (n-1)\theta_2^{\star 1,1}} - e^{\theta_1^{\star 1} + (2n-1)\theta_2^{\star 1,1}}\right) + \left(\mathbf{W}(e^V, \theta^\star)\right)_2^{1,1} (2n(2n-1) - n(n-1)) \\ &\quad + f(\varphi_{n,-}, \varphi_{n,+}, \mathbf{W}, \eta) \\ &= \underbrace{2ne^{\theta_1^{\star 1} + (n-1)\theta_2^{\star 1,1}}(1 - e^{n\theta_2^{\star 1,1}}) + n(3n-1)\left(\mathbf{W}(e^V, \theta^\star)\right)_2^{1,1}}_{:=x_n} + f(\varphi_{n,-}, \varphi_{n,+}, \mathbf{W}, \eta). \end{aligned}$$

Fix $\varepsilon > 0$, there exists $n_0 \geq 1$ such that for all $n \geq n_0$, $x_n < -\varepsilon$. Now by a continuity argument, there exists $\eta_0 = \eta_0(n_0)$ such that for all $\eta \leq \eta_0(n_0)$, $|f(\varphi_{n_0,-}, \varphi_{n_0,+}, \mathbf{W}, \eta)| \leq \varepsilon/2$. Therefore by assuming that $\Delta R_{\infty,\overline{\Lambda}}(\varphi_{n_0,-}, \varphi_{n_0,+}) = 0$, we obtain for $\eta \leq \eta_0$

$$0 = |\Delta R_{\infty,\overline{\Lambda}}(\varphi_{n_0,-}, \varphi_{n_0,+})| \geq |x_{n_0}| - |f(\varphi_{n_0,-}, \varphi_{n_0,+}, \mathbf{W}, \eta)| \geq \varepsilon/2 > 0$$

which leads to a contradiction and proves **[PD]**.

B.1.2 Proof that $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ is positive-definite for the two-type Strauss model

From Proposition 7, proving that $\underline{\mathbf{\Sigma}}_2(\theta^\star)$ is positive-definite in Framework 2 reduces to check Assumption **[PD]** with

$$\mathbf{Y}_\Lambda(\varphi; \theta^\star) = \mathbf{R}_{\infty,\Lambda}(\varphi; \mathbf{h}; \theta^\star),$$

where, for all $j = 1, \ldots, s$, $\left(\mathbf{R}_{\infty,\Lambda}(\varphi; \mathbf{h}; \theta^\star)\right)_j = R_{\infty,\Lambda}(\varphi; h_j, \theta^\star)$, h_j is the test function given by $h_j(x^m, \varphi; \theta) = \mathbf{1}_{[0,r_j]}(d(x^m, \varphi))e^{V(x^m|\varphi;\theta)}$. We fix as before $\overline{\delta} = D$ and $B = \emptyset$ in **[PD]**.

Let $0 < r_1 < \ldots < r_s < +\infty$. Let us also assume that $r_i \neq D$ for $i = 1, \ldots, s$ and define

$$A^{1,1}_{i,-}(\eta) = \Big\{\varphi \in \overline{\Omega} : \varphi(\Delta_0(\overline{D})) = 2, \varphi\left(\mathcal{B}\left((0,0), \frac{\eta}{4}\right) \times \{1\}\right) = 1, \varphi\left(\mathcal{B}\left((r_i - \frac{\eta}{2}, 0), \frac{\eta}{4}\right) \times \{1\}\right) = 1\Big\},$$
$$A^{1,1}_{i,+}(\eta) = \Big\{\varphi \in \overline{\Omega} : \varphi(\Delta_0(\overline{D})) = 2, \varphi\left(\mathcal{B}\left((0,0), \frac{\eta}{4}\right) \times \{1\}\right) = 1, \varphi\left(\mathcal{B}\left((r_i + \frac{\eta}{2}, 0), \frac{\eta}{4}\right) \times \{1\}\right) = 1\Big\}.$$

Let $\varphi_{i,\bullet} \in A^{1,1}_{i,\bullet}(\eta)$ for $\bullet = -, +$ and i$= 1, \ldots, s$. Let κ_i the constant given by

$$\kappa_i = \begin{cases} 2e^{\theta^{\star 1}_1 + \theta^{\star 1,1}_2} & \text{if } r_i < D, \\ 2e^{\theta^{\star 1}_1} & \text{otherwise.} \end{cases}$$

Then for $i, j = 1, \ldots, s$ and for η small enough

$$I_{\overline{\Lambda}}(\varphi_{i,-}; h_j, \theta^\star) = \int_{\overline{\Lambda}\times\mathbb{M}} h_j(x^m, \varphi_{i,-}) e^{-V(x^m|\varphi_{i,-};\theta^\star)} \mu(dx^m) - \begin{cases} \kappa_i & \text{if } i \leq j, \\ 0 & \text{otherwise.} \end{cases}$$
$$I_{\overline{\Lambda}}(\varphi_{i,+}; h_j, \theta^\star) = \int_{\overline{\Lambda}\times\mathbb{M}} h_j(x^m, \varphi_{i,+}) e^{-V(x^m|\varphi_{i,+};\theta^\star)} \mu(dx^m) - \begin{cases} \kappa_i & \text{if } i < j, \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand

$$\left(\mathbf{LPL}^{(1)}_{\overline{\Lambda}}(\varphi_{i,\bullet}; \theta^\star)\right)^{m'}_1 = \int_{\overline{\Lambda}\times\mathbb{M}} v_1^{m'}(x^m|\varphi_{i,\bullet} e^{-V(x^m|\varphi_{i,\bullet};\theta^\star)} \mu(dx^m) - \begin{cases} 2 & \text{if } m' = 1, \\ 0 & \text{if } m' = 2. \end{cases}$$
$$\left(\mathbf{LPL}^{(1)}_{\overline{\Lambda}}(\varphi_{i,\bullet}; \theta^\star)\right)^{m_1,m_2}_2 = \int_{\overline{\Lambda}\times\mathbb{M}} v_2^{m_1,m_2}(x^m|\varphi_{i,\bullet} e^{-V(x^m|\varphi_{i,\bullet};\theta^\star)} \mu(dx^m) - \begin{cases} 2 & \text{if } m_1 = m_2 = 1 \\ & \text{and } r_i < D \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbf{x} \in \mathbb{R}^s \setminus \{0\}$, then from previous computations

$$\mathbf{x}^T \left(\mathbf{R}_{\infty,\overline{\Lambda}}(\varphi_{i,+}; \mathbf{h}, \theta^\star) - \mathbf{R}_{\infty,\overline{\Lambda}}(\varphi_{i,-}; \mathbf{h}, \theta^\star)\right) = 2\kappa_i x_i + f(\mathbf{x}, \varphi_{i,+}, \varphi_{i,-}, \mathbf{h}). \tag{56}$$

By using a continuity argument, one may prove that for every $\varepsilon > 0$ there exists $\eta > 0$ such that $|f((\mathbf{x}, \varphi_{i,+}, \varphi_{i,-}, \mathbf{h})| \leq \varepsilon$. Therefore, assuming that the l.h.s. of (57) equals 0 leads to $x_i = 0$ for $i = 1, \ldots, s$.

B.2 Area-interaction point process

We fix for simplicity $d = 2$, though the proofs may be extended easily to higher dimensions.

B.2.1 Proof that $\underline{\Sigma}_1(\theta^\star)$ is positive-definite for the area-interaction model

From Proposition 7, the proof reduces to check Assumption **[PD]** with $h = e^V$, $\overline{\delta} = D$, $B = \emptyset$ and

(i) $\mathbf{Y}_\Lambda(\varphi; \theta^\star) = I_\Lambda(\varphi; e^V, \theta^\star)$,

(ii) $\mathbf{Y}_\Lambda(\varphi; \theta^\star) = R_{\infty,\Lambda}(\varphi; e^V, \theta^\star)$.

Again (i) is ensured by Proposition 16 since this model satisfies **[Exp]**.

(ii) Let us consider for some $\eta, \omega > 0$ the two following events:

$$\begin{aligned} A_1(\eta, \omega) &:= \left\{\varphi \in \overline{\Omega} : \varphi(\Delta_0(\overline{\delta})) = 2, \varphi(\mathcal{B}((0,0), \eta)) = 1, \varphi(\mathcal{B}((0,\omega), \eta)) = 1\right\} \\ A_2(\eta, \omega) &:= \left\{\varphi \in \overline{\Omega} : \varphi(\Delta_0(\overline{\delta})) = 3, \varphi(\mathcal{B}((0,0), \eta)) = 1, \varphi(\mathcal{B}((0,\omega), \eta)) = 2\right\} \end{aligned}$$

Fix η, ω, let $\varphi_j \in A_j(\eta,\omega)$ and denote by $\widetilde{e^V}(\varphi) := \sum_{x\in\varphi_{\overline{\Lambda}}} e^{V(x|\varphi\setminus x;\theta^\star)}$

$$I_{\overline{\Lambda}}(\varphi_j; e^V, \theta^\star) \quad = \quad |\overline{\Lambda}| - \widetilde{e^V}(\varphi_j).$$

When $\eta \to 0$,

$$\widetilde{e^V}(\varphi_1) \to 2e^{\theta_1^\star+\theta_2^\star g(\omega)} \quad \text{and} \quad \widetilde{e^V}(\varphi_2) \to 2e^{\theta_1^\star} + e^{\theta_1^\star+\theta_2^\star g(\omega)}$$

where $g(\omega) := |\mathcal{B}((0,0),R)\cup\mathcal{B}((0,\omega),R)|-|\mathcal{B}((0,0),R)|$. Moreover, by denoting $\widetilde{v}_2(\varphi) = \sum_{x\in\varphi} v_2(x|\varphi\setminus x)$

$$\begin{aligned}
\left(\mathbf{LPL}_{\overline{\Lambda}}^{(1)}(\varphi_j;\theta^\star)\right)_1 &= \int_{\overline{\Lambda}} e^{-V(x|\varphi_j;\theta^\star)}dx - \left\{\begin{array}{ll} 2 & \text{if } j=1 \\ 3 & \text{if } j=2 \end{array}\right. \\
\left(\mathbf{LPL}_{\overline{\Lambda}}^{(1)}(\varphi_j;\theta^\star)\right)_2 &= \int_{\overline{\Lambda}} v_2(x|\varphi_j)e^{-V(x|\varphi_j;\theta^\star)}dx - \widetilde{v}_2(\varphi_j).
\end{aligned}$$

Again, when $\eta \to 0$, one may note that for $k = 1, 2$

$$\int_{\overline{\Lambda}} v_k(x|\varphi_1)e^{-V(x|\varphi_1;\theta^\star)}dx - \int_{\overline{\Lambda}} v_k(x|\varphi_2)e^{-V(x|\varphi_2;\theta^\star)}dx \to 0$$

and

$$\widetilde{v}_2(\varphi_1) \to 2g(\omega) \quad \text{and} \quad \widetilde{v}_2(\varphi_2) \to g(\omega)$$

These computations lead to

$$\begin{aligned}
R_{\infty,\overline{\Lambda}}(\varphi_1; e^V, \theta^\star) - R_{\infty,\overline{\Lambda}}(\varphi_2; e^V, \theta^\star) &= 2e^{\theta_1^\star} - e^{\theta_1^\star+\theta_2^\star g(\omega)} - \left(\mathbf{W}(e^V,\theta^\star)\right)_1 + g(\omega)\left(\mathbf{W}(e^V,\theta^\star)\right)_2 \\
&\quad + f(\varphi_1,\varphi_2,\mathbf{W}),
\end{aligned}$$

where the function f is such that for all $\varepsilon > 0$, there exists η small enough such that $|f(\varphi_1,\varphi_2,\mathbf{W})| \leq \varepsilon$. Let $\varphi_j \in A_j(\eta,0)$, then, since $g(0) = 0$, assuming that the l.h.s. of the previous equation equals 0 leads to $\left(\mathbf{W}(e^V,\theta^\star)\right)_1 = e^{\theta_1^\star}$. Now, let $\omega > 0$ and again assume that $R_{\infty,\overline{\Lambda}}(\varphi_1; e^V, \theta^\star) = R_{\infty,\overline{\Lambda}}(\varphi_2; e^V, \theta^\star)$, we therefore obtain (by the continuity argument)

$$\left(\mathbf{W}(e^V,\theta^\star)\right)_2 = \frac{e^{\theta_1^\star+\theta_2^\star g(\omega)} - e^{\theta_1^\star}}{g(\omega)}.$$

But $\left(\mathbf{W}(e^V,\theta^\star)\right)_2$ is a constant and so cannot depend on ω. Therefore one of the assumptions made before is untrue, which proves **[PD]**.

B.2.2 Proof that $\underline{\Sigma}_2(\theta^\star)$ is positive-definite for the area-interaction model

From Proposition 7, it suffices to check Assumption **[PD]** with

$$\mathbf{Y}_\Lambda(\varphi;\theta^\star) = \mathbf{R}_{\infty,\Lambda}(\varphi;\mathbf{h};\theta^\star),$$

where, for all $j = 1,\ldots,s$, $\left(\mathbf{R}_{\infty,\Lambda}(\varphi;\mathbf{h};\theta^\star)\right)_j = R_{\infty,\Lambda}(\varphi;h_j,\theta^\star)$, h_j is the test function given by $h_j(x^m,\varphi;\theta) = \mathbf{1}_{[0,r_j]}(d(x^m,\varphi))e^{V(x^m|\varphi;\theta)}$, and where, again, we choose $\overline{\delta} = D$ and $B = \emptyset$.

The proof is quite similar to the one proposed for the 2-type marked Strauss point process (see B.2.2). Let $0 < r_1 < \ldots < r_s < +\infty$. Let us also assume that $r_i \neq D$ for $i = 1,\ldots,s$

$$\begin{aligned}
A_{i,-}(\eta) &= \left\{\varphi\in\overline{\Omega} : \varphi(\Delta_0(\overline{D})) = 2, \varphi\left(\mathcal{B}\left((0,0),\frac{\eta}{4}\right)\right) = 1, \varphi\left(\mathcal{B}\left((r_i-\frac{\eta}{2},0),\frac{\eta}{4}\right)\right) = 1\right\}, \\
A_{i,+}(\eta) &= \left\{\varphi\in\overline{\Omega} : \varphi(\Delta_0(\overline{D})) = 2, \varphi\left(\mathcal{B}\left((0,0),\frac{\eta}{4}\right)\right) = 1, \varphi\left(\mathcal{B}\left((r_i+\frac{\eta}{2},0),\frac{\eta}{4}\right)\right) = 1\right\}.
\end{aligned}$$

Let $i,j \in \{1,\ldots,s\}$ and $k \in \{1,2\}$, let $\varphi_{i,-} \in A_{i,-}$ and $\varphi_{i,+} \in A_{i,+}$, then

$$\begin{aligned}
I_{\overline{\Lambda}}(\varphi_{i,-};h_j,\theta^\star) &= \int_{\overline{\Lambda}} h_j(x,\varphi_{i,-};\theta^\star)e^{-V(x|\varphi_{i,-};\theta^\star)}dx - \begin{cases} \widetilde{e^V}(\varphi_{i,-}) & \text{if } i \le j \\ 0 & \text{otherwise.} \end{cases} \\
I_{\overline{\Lambda}}(\varphi_{i,+};h_j,\theta^\star) &= \int_{\overline{\Lambda}} h_j(x,\varphi_{i,+};\theta^\star)e^{-V(x|\varphi_{i,+};\theta^\star)}dx - \begin{cases} \widetilde{e^V}(\varphi_{i,+}) & \text{if } i < j \\ 0 & \text{otherwise.} \end{cases} \\
\left(\mathbf{LPL}^{(1)}_{\overline{\Lambda}}(\varphi_{i;\bullet};\theta^\star)\right)_k &= \int_{\overline{\Lambda}} v_k(x|\varphi_{i,\bullet})e^{-V(x|\varphi_{i,\bullet};\theta^\star)}dx - \sum_{x\in\varphi_{i,\bullet}} v_k(x|\varphi_{i,\bullet}\setminus x),
\end{aligned}$$

for $\bullet = -,+$. It is expected that for small η, $\left(\mathbf{LPL}^{(1)}_{\overline{\Lambda}}(\varphi_{i,-};\theta^\star)\right)_k \simeq \left(\mathbf{LPL}^{(1)}_{\overline{\Lambda}}(\varphi_{i,+};\theta^\star)\right)_k$ and $\widetilde{e^V}(\varphi_{i,-}) \simeq \widetilde{e^V}(\varphi_{i,+}) \simeq \kappa_i := 2e^{\theta_1^\star+\theta_2^\star|\mathcal{B}(0,R)\cup\mathcal{B}(r_i,R)|}$. Let $\mathbf{x} \in \mathbb{R}^s \setminus \{0\}$, then from previous computations

$$\mathbf{x}^T\left(\mathbf{R}_{\infty,\overline{\Lambda}}(\varphi_{i,+};\mathbf{h},\theta^\star) - \mathbf{R}_{\infty,\overline{\Lambda}}(\varphi_{i,-};\mathbf{h},\theta^\star)\right) = 2\kappa_i x_i + f(\mathbf{x},\varphi_{i,+},\varphi_{i,-},\mathbf{h}) \tag{57}$$

where for every $\varepsilon > 0$ there exists $\eta > 0$ such that $|f((\mathbf{x},\varphi_{i,+},\varphi_{i,-},\mathbf{h})| \le \varepsilon$. Therefore, assuming that the l.h.s. of (57) equals 0 leads to $x_i = 0$ for $i = 1,\ldots,s$.

References

P.K. Andersen, Ø Borgan, R.D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer-Verlag, New York, 1993.

A. Atkinson. *Plots, Transformations and Regression*. Oxford Statistical Sciences. Oxford University Press/ Clarendon, 1985.

A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42:283–322, 2000.

A. Baddeley and R. Turner. Modelling spatial point patterns in R. *Journal of Statistical Software*, 12(6):1–42, 2005.

A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes. *Journal of the Royal Statistical Society (series B)*, 67:1–35, 2005.

A. Baddeley, J. Møller, and A.G. Pakes. Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics*, 60(3):627–649, 2008.

E. Bertin, J.-M. Billiot, and R. Drouilhet. Existence of "Nearest-Neighbour" Gibbs Point Models. *Ann. Appl. Probab.*, 31:895–909, 1999.

J.-M. Billiot, J.-F. Coeurjolly, and R. Drouilhet. Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes. *Electronic Journal of Statistics*, 2: 234–264, 2008.

E. Bolthausen. On the central limit theorem for stationary mixing random fields. *The Annals of Probability*, 10(4):1047–1050, 1982.

J.-F. Coeurjolly and R. Drouilhet. Maximum pseudo-likelihood estimator for general marked Gibbs point processes and applications to the lennard-jones model. *submitted*, submitted (http://hal.archives-ouvertes.fr/hal-00442750/en/), 2009.

F. Comets and M. Janzura. A central limit theorem for conditionally centred random fields with an application to Markov fields. *Journal of Applied Probability*, 35(3):608–621, 1998.

D. Daley and D. Vere-Jones. *An introduction to the Theory of Point Processes*. Springer Verlag, New York, 1988.

D. Dereudre. Gibbs Delaunay tessellations with geometric hardcore condition. *J. Stat. Phys.*, 121 (3-4):511–515, 2005.

D. Dereudre and F. Lavancier. Campbell equilibrium equation and pseudo-likelihood estimation for non-hereditary Gibbs point processes. *Bernoulli*, 15(4):1368–1396, 2009.

D. Dereudre and F. Lavancier. Practical simulation and estimation for Gibbs Delaunay-voronoï tessellations with geometric hardcore interaction. *submitted (http://hal.archives-ouvertes.fr/hal-00402029/en/)*, 2010.

D. Dereudre, R. Drouilhet, and H.O. Georgii. Existence of Gibbsian point processes with geometry-dependent interactions. *submitted*, 2010.

P.J. Diggle. *Statistical analysis of spatial point patterns*. Academic Press, London, 2003.

T.R. Fleming and D.P. Harrington. *Counting processes and survival analysis*. Wiley, New York, 1991.

H.O. Georgii. Canonical and grand canonical Gibbs states for continuum systems. *Communications in Mathematical Physics*, 48(1):31–51, 1976.

J. Illian, A. Penttinen H., and Stoyan. *Statistical analysis and modelling of spatial point patterns*. Wiley-Interscience, 2008.

J.L. Jensen and H.R. Künsch. On asymptotic normality of pseudo likelihood estimates of pairwise interaction processes. *Ann. Inst. Statist. Math.*, 46:475–486, 1994.

J.L. Jensen and J. Møller. Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.*, 1:445–461, 1991.

J. Møller. Parametric methods for spatial point processes. Technical Report Research Report R-2008-04, Department of Mathematical Sciences, Aalborg University, 2008.

J. Møller and R. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, 2003.

X. Nguyen and H. Zessin. Integral and differential characterizations Gibbs processes. *Mathematische Nachrichten*, 88(1):105–115, 1979a.

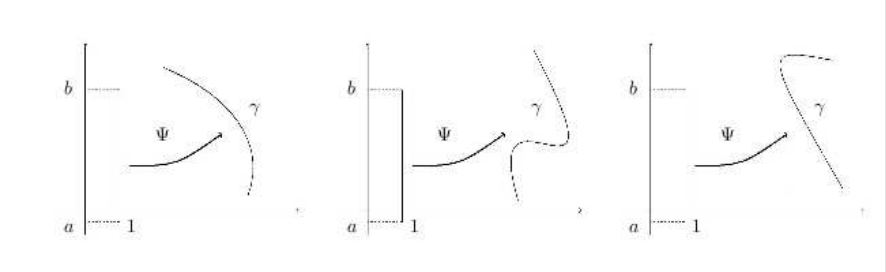
X.X. Nguyen and H. Zessin. Ergodic theorems for Spatial Process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 48:133–158, 1979b.

C.J. Preston. *Random fields*. Springer Verlag, 1976.

D. Ruelle. *Statistical Mechanics*. Benjamin, New York-Amsterdam, 1969.

D. Stoyan and P. Grabarnik. Second-order characteristics for stochastic structures connected with Gibbs point processes. *Mathematische Nachrichten*, 151:95–100, 1991.

D. Stoyan, W.S. Kendall, J. Mecke, and L. Ruschendorf. *Stochastic geometry and its applications*. John Wiley and Sons, Chichester, 1987.



[CDDL10]
J.-F. Coeurjolly, D. Dereudre, R. Drouilhet and F. Lavancier. Takacs-Fiksel method for stationary marked Gibbs point processes. *in revision for Scandinavian Journal of Statistics* (available at `http://arxiv.org/abs/1007.3351`), 2010.

Takacs Fiksel method for stationary marked Gibbs point processes

J.-F. Coeurjolly[1,2], D. Dereudre[3], R. Drouilhet[2] and F. Lavancier[4]

[1] GIPSA-lab, Grenoble University, France,
[2] Laboratoire Jean Kuntzmann, Grenoble University, France
[3] LAMAV, Valenciennes University, France,
[4] Laboratoire de Mathématiques Jean Leray, Nantes University, France.

July 21, 2010

Abstract

This paper studies a method to estimate the parameters governing the distribution of a stationary marked Gibbs point process. This method, known as the Takacs-Fiksel method, is based on the estimation of the left and right hand sides of the Georgii-Nguyen-Zessin formula and leads to a family of estimators due to the possible choices of test functions. We propose several examples illustrating the interest and flexibility of this procedure. We also provide sufficient conditions based on the model and the test functions to derive asymptotic properties (consistency and asymptotic normality) of the resulting estimator. The different assumptions are discussed for exponential family models and for a large class of test functions.

Keywords: stationary marked Gibbs point processes, parametric estimation, Takacs-Fiksel method, asymptotic properties, ergodic theorem, central limit theorem.

1 Introduction

Spatial point pattern data arise in a wide range of applications and various statistical methods have been developed to study these kinds of data (see [32], [33] or [25] for a review). In particular, a spatial point process is often modelled as the realization of a Gibbs distribution, defined through an interaction function, also called Hamiltonian. Gibbs models are extensions of the well-known Poisson process since they constitute a way to introduce dependence between points. Inference for parametric models in this setting has known a large development during the last decade. The most popular method to estimate the parameters is certainly the maximum likelihood estimator (MLE). It involves an intractable normalizing constant, but recent developments in computational statistics, in particular perfect simulations, make inference feasible for many Gibbs models (see [4]). Although the MLE suffers from a lack of theoretical justifications (only some results for sparse patterns are proposed in [29]), some comparison studies, as in [17], have shown that it outperforms the other estimation methods. Nevertheless, the computation of the MLE remains very time-consuming and even extremely difficult to perform for some models. It is thus necessary to have alternative quick estimators at one's disposal, at least to propose relevant initial values for the MLE computation. The maximum pseudo-likelihood estimator (MPLE for short) constitutes one of them. Proposed by Besag in [7] and popularized by J.L. Jensen and J. Moller in [26] and A. Baddeley and R. Turner in [1] for spatial point processes, this method has the advantage of being theoretically well understood (see [9], [15]) and is much faster to compute than the MLE. Another estimation procedure is the Takacs-Fiksel estimator, which arised from [18], [39], [19]. It can be viewed, in some sense, as a

generalization of the MPLE. As a matter of fact, the Takacs-Fiksel method is not very popular, nor really used in practice. The main reason is certainly its relative poor performances, in terms of mean square error, observed for some particular cases as in [17]. However, we think that this procedure deserves some consideration for several reasons that we expose below.

The Takacs-Fiksel procedure is based on the Georgii-Nguyen-Zessin formula (GNZ formula for short). Empirical counterparts of the left and the right hand sides of this equation are considered, and the induced estimator is such that the difference of these two terms is close to zero. Since the GNZ formula is valid for any test functions, the Takacs-Fiksel procedure does not only lead to one particular estimator but to a family of estimators, depending on the choice of the test functions. This flexibility is the main advantage of the procedure. We present several examples in the present paper. Let us summarize them in order to underline the interest of this method (see Section 3.2 for more details). First, this procedure can allow us to achieve estimations that likelihood-type methods cannot. As an example, we focus on the quermass model, which gathers the area interaction point process as a particular case (see [27], [13]). This model is sometimes used for geometric random objects. From a data set, one typically does not observe the point pattern but only some geometric sets arising from these points. The non-observability of the points makes the likelihood-type inference unfeasible. As we will show, this issue may be solved thanks to the Takacs-Fiksel procedure, provided that the test functions are chosen properly.

Another motivation is the possibility to choose some test functions depending on the Hamiltonian in order to construct quicker estimators which do not require the computation of an integral for each value of the parameter. This improvement appears crucial for rigid models, as those involved in stochastic geometry (see [16]), for which the MLE is prohibitively time-consuming and the MPLE still remains difficult to implement. Moreover, for some models, it is even possible to obtain explicit estimators which do not require any simulation nor optimization. This is illustrated for the Strauss model.

Therefore, it appears important to us to understand the theoretical properties of this procedure. This problem is the main objective of the present paper. We prove the consistency and the asymptotic normality of the induced estimator in a very large setting. In particular, we obtain a central limit theorem for the Takacs-Fiksel estimator with the classical rate of convergence, i.e. square root of the volume of the observation domain. This asymptotic result leads to the following comment: as a quick consistent estimator, the Takacs-Fiksel estimator appears to be a very good starting point for refined algorithms. Among them, let us mention the first step Newton Raphson algorithm (as used in [24]) which allows an accurate approximation of the MLE, starting from a consistent estimator, in only one step. Although the theoretical justifications are missing in the Gibbs framework, it is well-known that this procedure leads to an efficient estimator in the classical iid case (see [28]). Another possibility could be to exploit the local asymptotic normality of the model in order to construct an adaptive estimator from the Takacs-Fiksel estimator. The LAN property has only been proved for restrictive models in [29], but one can hope that it remains true for most Gibbs models. This procedure also leads to an efficient estimator. All these possibilities are interesting prospectives for future investigations.

Some asymptotic properties of the Takacs-Fiksel procedure have already been investigated in two previous studies: by L. Heinrich in [23] and by J.-M. Billiot in [8]. These papers have different frameworks and are based on different tools but they both involve regularity and integrability type assumptions on the Hamiltonian and a theoretical condition which ensures that the contrast function (associated to the Takacs-Fiksel procedure) has a unique minimum. In [23], the consistency and the asymptotic normality are obtained for a quite large class of test functions. These results are, however, proved under the Dobrushin condition (see Theorems 2 and 3 in [23]) which implies the uniqueness of the underlying Gibbs measure and some mixing properties. This condition imposes a dramatic reduction of the space of possible values for the parameters of the model. In [8], the author focuses only on pairwise interaction point processes (which excludes the Quermass model for example). The author mainly obtained the consistency for a specific class of test functions. In the case of a multi-Strauss pairwise interaction point process, the author also proved that the identifiability condition holds for the class of test functions he considered.

In contrast, our asymptotic results are proved in a very large setting, i.e. for a large class of stationary marked Gibbs models and test functions. The main restriction, involved only for the asymptotic

normality, is the finite-range of the Hamiltonian. There are no limitations on the space of parameters and, in particular, the possible presence of phase transition does not affect the asymptotic behavior of the estimator. Moreover, the test functions may depend on the parameters. This extension seems important to us because, as emphasized in Section 3.2.2, such test functions can lead to quick and/or explicit estimators. All the general hypotheses assumed for the asymptotic results are discussed. For this, we focus on exponential family models, that is, on models whose interaction function is linear in the parameters. We show that our integrability and regularity assumptions are not restrictive since they are valid for a large class of models such as the Multi-Strauss marked point process, the Strauss-disc type point process, the Geyer's triplet point process, the Quermass model and for all test functions used as a motivation of this work. In the exponential family models setting, we also discuss the classical identifiability condition which is required for the Takacs-Fiksel procedure. To the best of our knowledge, this is the first attempt to discuss it : which choices of test functions (and how many) lead to a unique minimum of the contrast function? We propose some general criteria and provide some examples. It seems commonly admitted that to achieve the identification of the Takacs-Fiksel procedure, one should at least choose as many test functions as the number of parameters. As a consequence of our study, it appears that one should generally strictly choose more test functions than the number of parameters to achieve identification.

The rest of the paper is organized as follows. Section 2 introduces some notation and a short background on marked Gibbs point processes. The Takacs-Fiksel method is presented in Section 3. It is based on the GNZ formula which is recalled. Several examples of test functions are given. They aim at illustrating our interest to consider the Takacs-Fiksel procedure. Asymptotic results of the induced estimator are proposed in Section 4. Our results are obtained from a single realization observed in a domain whose volume is supposed to increase to infinity. Some integrability and regularity assumptions made for the Hamiltion and for the test functions are discussed in this section while in Section 5 the identifiability condition is specifically dealt with. In Section 6, the very special situation where the energy function is not hereditary is considered. The GNZ formula is no longer valid in this setting, but it has been recently extended in [15] thanks to a slight modification. This leads to a natural generalization of the Takacs-Fiksel procedure. Finally, Section 7 contains the proofs of the asymptotic results.

2 Background and notation

2.1 General notation, configuration space

Subregions of $\mathbb{R}^d$ will typically be denoted by Λ or Δ and will always be assumed to be Borel with positive Lebesgue measure. We write $\Lambda \Subset \mathbb{R}^d$ if Λ is bounded. Λ^c denotes the complementary set of Λ inside $\mathbb{R}^d$. The notation $|.|$ will be used without ambiguity for different kind of objects. For a countable set $\mathcal{J}$, $|\mathcal{J}|$ represents the number of elements belonging to $\mathcal{J}$; For $\Lambda \Subset \mathbb{R}^d$, $|\Lambda|$ is the volume of Λ; For $x \in \mathbb{R}^d$, $|x|$ corresponds to its uniform norm while $\|x\|$ is simply its euclidean norm. For all $x \in \mathbb{R}^d, \rho > 0$ let $\mathcal{B}(x,\rho) := \{y \in \mathbb{R}^d,\ |y-x| < \rho\}$. For a matrix $\underline{\mathbf{M}}$, let $\|\underline{\mathbf{M}}\|$ be the Frobenius norm of $\underline{\mathbf{M}}$ defined by $\|\underline{\mathbf{M}}\|^2 = Tr(\underline{\mathbf{M}}^T\underline{\mathbf{M}})$, where Tr is the trace operator.

The space $\mathbb{R}^d$ is endowed with the Borel σ-algebra and the Lebesgue measure λ. Let $\mathbb{M}$ be a measurable space, which aims at being the mark space, endowed with the σ-algebra $\mathcal{M}$ and the probability measure λ^{m}. The state space of the point processes will be $\mathbb{S} := \mathbb{R}^d \times \mathbb{M}$ measured by $\mu := \lambda \otimes \lambda^{\mathrm{m}}$. We shall denote for short $x^m = (x,m)$ an element of $\mathbb{S}$.

A configuration is a subset φ of $\mathbb{S}$ which is locally finite in that $\varphi_\Lambda := \varphi \cap (\Lambda \times \mathbb{M})$ has finite cardinality $N_\Lambda(\varphi) := |\varphi_\Lambda|$ for all $\Lambda \Subset \mathbb{R}^d$. The space $\Omega = \Omega(\mathbb{S})$ of all configurations is equipped with the σ-algebra $\mathcal{F}$ that is generated by the counting variables $N_\Lambda(\varphi)$ with $\Lambda \Subset \mathbb{R}^d$. Finally, let $T = (\tau_x)_{x\in\mathbb{R}^d}$ be the shift group, where $\tau_x : \Omega \to \Omega$ is the translation by the vector $-x \in \mathbb{R}^d$. For the sake of simplicity, we set $\varphi \cup x^m := \varphi \cup \{x^m\}$ and $\varphi \setminus x^m := \varphi \setminus \{x^m\}$.

2.2 Marked Gibbs point processes

Our results will be expressed for general stationary Gibbs point processes. Since we are interested in asymptotic properties, we have to consider these point processes acting on the infinite volume $\mathbb{R}^d$. Let us briefly recall their definition.

A marked point process Φ is a Ω-valued random variable, with probability distribution P on $(\Omega, \mathcal{F})$. The most prominent marked point process is the marked Poisson process π^z with intensity measure $z\lambda \otimes \lambda^{\mathrm{m}}$ on \$, with $z > 0$ (see for example [30] for definition and properties). For $\Lambda \Subset \mathbb{R}^d$, let us denote by π^z_Λ the marginal probability measure in Λ of the Poisson process with intensity z. Without loss of generality, the intensity z is fixed to 1 and we simply write π and π_Λ in place of π^1 and π^1_Λ.

Let $\theta \in \mathbb{R}^p$ (for some $p \geq 1$). For any $\Lambda \Subset \mathbb{R}^d$, let us consider the parametric function $V_\Lambda(.;\theta)$ from Ω into $\mathbb{R} \cup \{+\infty\}$. From a physical point of view, $V_\Lambda(\varphi;\theta)$ is the energy of φ_Λ in Λ given the outside configuration φ_{Λ^c}. In this paper, we focus on stationary point processes on $\mathbb{R}^d$, i.e. with stationary (i.e. T-invariant) probability measure. For any $\Lambda \Subset \mathbb{R}^d$, we therefore consider $V_\Lambda(.;\theta))$ to be T-invariant, i.e. $V_\Lambda(\tau_x\varphi;\theta) = V_\Lambda(\varphi;\theta)$ for any $x \in \mathbb{R}^d$. Furthermore, $(V_\Lambda(.;\theta))_{\Lambda \Subset \mathbb{R}^d}$ is a compatible family of energies, i.e for every $\Lambda \subset \Lambda' \Subset \mathbb{R}^d$, there exists a measurable function $\psi_{\Lambda,\Lambda'}$ from Ω into $\mathbb{R} \cup \{+\infty\}$ such that

$$\forall \varphi \in \Omega \quad V_{\Lambda'}(\varphi;\theta) = V_\Lambda(\varphi;\theta) + \psi_{\Lambda,\Lambda'}(\varphi_{\Lambda^c};\theta). \tag{1}$$

A stationary marked Gibbs point process is characterized by a stationary marked Gibbs measure usually defined as follows (see [35]).

Definition 1 *A probability measure P_θ on Ω is a stationary marked Gibbs measure for the compatible family of T-invariant energies $(V_\Lambda(.;\theta))_{\Lambda \Subset \mathbb{R}^d}$ if for every $\Lambda \Subset \mathbb{R}^d$, for P_θ-almost every outside configuration φ_{Λ^c}, the law of P_θ given φ_{Λ^c} admits the following conditional density with respect to π_Λ:*

$$f_\Lambda(\varphi_\Lambda|\varphi_{\Lambda^c};\theta) = \frac{1}{Z_\Lambda(\varphi_{\Lambda^c};\theta)} e^{-V_\Lambda(\varphi;\theta)},$$

where $Z_\Lambda(\varphi_{\Lambda^c};\theta)$ is a normalization called the partition function.

The existence of a Gibbs measure on Ω which satisfies these conditional specifications is a difficult issue. We refer the interested reader to [36, 35, 5, 12, 14] for the technical and mathematical development of the existence problem.

In a first step, we assume that the family of energies is hereditary (the non-hereditary case will be considered in Section 6), which means that for any $\Lambda \Subset \mathbb{R}^d$, for any $\varphi \in \Omega$, and for all $x^m \in \Lambda \times \mathbb{M}$,

$$V_\Lambda(\varphi;\theta) = +\infty \Rightarrow V_\Lambda(\varphi \cup x^m;\theta) = +\infty, \tag{2}$$

or equivalently, for all $x^m \in \varphi_\Lambda$, $f_\Lambda(\varphi_\Lambda|\varphi_{\Lambda^c};\theta) > 0 \Rightarrow f_\Lambda(\varphi_\Lambda \setminus x^m|\varphi_{\Lambda^c};\theta) > 0$.

The minimal assumption of our paper is then:

> [**Mod**]: Our data consist in the realization of a marked point process Φ with stationary marked Gibbs measure $P_{\theta^\star}$, where $\theta^\star \in \mathring{\mathbf{\Theta}}$, $\mathbf{\Theta}$ is a compact subset of $\mathbb{R}^p$ and, for any $\theta \in \mathbf{\Theta}$, there exists a stationary marked Gibbs measure P_θ for the compatible T-invariant hereditary family $(V_\Lambda(.;\theta))_{\Lambda \Subset \mathbb{R}^d}$.

Let us note that [**Mod**] ensures the existence of at least one stationary Gibbs measure. When this Gibbs measure is not unique, we say that the phase transition occurs. In this situation the set of Gibbs measures is a Choquet simplex and any Gibbs measure is a mixing of extremal ergodic Gibbs measures. If the Gibbs measure is unique, it is necessary ergodic (see [21] for more details about these properties).

In the rest of this paper, the reader has mainly to keep in mind the concept of local energy defined as the energy required to insert a point x^m into the configuration φ and expressed for any $\Lambda \ni x$ by

$$V\left(x^m|\varphi;\theta\right) := V_\Lambda(\varphi \cup x^m;\theta) - V_\Lambda(\varphi;\theta). \tag{3}$$

From the compatibility of the family of energies, i.e. (1), this definition does not depend on Λ.

3 The Takacs-Fiksel estimation procedure

3.1 Presentation

The basic ingredient for the definition of the Takacs-Fiksel method is the so-called GNZ formula (see [20]) stated below.

Theorem 1 (Georgii-Nguyen-Zessin Formula) *Under **[Mod]**, for any function $h(\cdot,\cdot;\theta) : \mathbb{S}\times\Omega \to \mathbb{R}$ such that the following quantities are finite, then*

$$\mathbf{E}\left(\int_{\mathbb{R}^d\times\mathbb{M}} h\left(x^m,\Phi;\theta\right) e^{-V(x^m|\Phi;\theta^\star)}\mu(dx^m)\right) = \mathbf{E}\left(\sum_{x^m\in\Phi} h\left(x^m,\Phi\setminus x^m;\theta\right)\right), \tag{4}$$

where $\mathbf{E}$ denotes the expectation with respect to $P_{\theta^\star}$.

For stationary marked Gibbs point processes, (4) reduces to

$$\mathbf{E}\left(h\left(0^M,\Phi;\theta\right) e^{-V\left(0^M|\Phi;\theta^\star\right)}\right) = \mathbf{E}\left(h\left(0^M,\Phi\setminus 0^M;\theta\right)\right) \tag{5}$$

where M denotes a random variable with probability distribution λ^{m}.

Let $h(\cdot,\cdot;\theta) : \mathbb{S}\times\Omega\to\mathbb{R}$ and let us define for any $\varphi\in\Omega$, $\theta\in\boldsymbol{\Theta}$ and $\Lambda\Subset\mathbb{R}^d$

$$C_\Lambda(\varphi;h,\theta) := \int_{\Lambda\times\mathbb{M}} h(x^m,\varphi;\theta)e^{-V(x^m|\varphi;\theta)}\mu(dx^m) - \sum_{x^m\in\varphi_\Lambda} h(x^m,\varphi\setminus x^m;\theta) \tag{6}$$

Assume that we observe the realization of a marked point process Φ satisfying **[Mod]** in a domain Λ_n (aimed at converging towards $\mathbb{R}^d$). For appropriate choices of the functional h and a sequence of domain Λ_n, then it is possible to apply an ergodic theorem of [34] to prove that the first and second terms of $|\Lambda_n|^{-1}C_{\Lambda_n}(\Phi;h,\theta)$ respectively converge to the left and right terms of (4). Here is the basic argument to define the Takacs-Fiksel method: let us give K functions $h_k(\cdot,\cdot;\theta) : \mathbb{S}\times\Omega\to\mathbb{R}$ (for $k=1,\ldots,K$), then the Takacs-Fiksel estimator is simply defined by

$$\widehat{\theta}_n(\varphi) := \widehat{\theta}_n^{TF}(\varphi) = \arg\min_{\theta\in\boldsymbol{\Theta}} \sum_{k=1}^K C_{\Lambda_n}(\varphi;h_k,\theta)^2. \tag{7}$$

3.2 Some examples

In this section, some examples of models and test functions h, involved in (6), are provided. The choices made in previous studies are presented in 3.2.1. The two examples presented in 3.2.2 and 3.2.3 show the relevance of the Takacs-Fiksel procedure to provide quick estimates. The last example, in 3.2.4, shows that appropriate choices of test functions can solve identification problems when points are not observed.

3.2.1 Classical examples

Let us first quote the particular case when the Takacs-Fiksel estimator reduces to the maximum pseudo-likelihood estimator. The MPLE is obtained by maximizing the log-pseudo-likelihood contrast function, given by

$$LPL_{\Lambda_n}(\varphi;\theta) = -\int_{\Lambda_n\times\mathbb{M}} e^{-V(x^m|\varphi;\theta)}\mu(dx^m) \;-\; \sum_{x\in\varphi_{\Lambda_n}} V(x^m|\varphi\setminus x^m;\theta). \tag{8}$$

Therefore, with the choice $h_k(x^m,\varphi;\theta) = \frac{\partial}{\partial\theta_k}V(x^m|\varphi;\theta)$, $k=1,\ldots,p$, the estimator $\widehat{\theta}_n$, defined by (7), solves the system $C_{\Lambda_n}(\varphi;h_k,\theta)=0$, $k=1,\ldots,p$, which means that $\widehat{\theta}_n$ is the root of the gradient vector of LPL_{Λ_n}, i.e. $\widehat{\theta}_n$ is the MPLE.

The Takacs-Fiksel estimator is implemented in [17] and is compared to other estimators. In this study, the two parameters of a pairwise interaction function are estimated from different test functions $h_{r_1},\cdots,h_{r_K}$, where for $r>0$,

$$h_r(x,\varphi;\theta) = \left|\varphi_{\mathcal{B}(x,r)}\right| = \sum_{y\in\varphi}\mathbf{1}_{[0,r]}(|y-x|).$$

The integral term involved in (6) is approximated by discretization and the induced estimation (7) is then assessed. Note that with this choice of test functions, the sum term in (6), when normalized by $|\Lambda|^{-1}$, is

an estimation of $\rho^2\mathcal{K}(r)$, where $\mathcal{K}(\cdot)$ is the reduced second order function and ρ denotes the intensity of the stationary point process Φ, i.e. for all $B \in \mathcal{B}(\mathbb{R}^d)$, $\mathbf{E}(\Phi(B)) = \rho|B|$.

The latter choice requires the computation of the integral in (6) for all θ. A more convenient choice could be the one first proposed by Fiksel:

$$h_r(x, \varphi; \theta) = \left|\varphi_{\mathcal{B}(x,r)}\right| e^{V(x|\varphi;\theta)} = e^{V(x|\varphi;\theta)} \sum_{y\in\varphi} \mathbf{1}_{[0,r]}(|y - x|). \tag{9}$$

In the stationary case, this leads to the following approximation thanks to the ergodic theorem

$$\frac{1}{|\Lambda|}\int_{\Lambda\times\mathbb{M}} h_r(x, \varphi; \theta)e^{-V(x|\varphi;\theta)}\mu(dx^m) \approx \mathbf{E}\left|\Phi_{\mathcal{B}(0,r)}\right| = \rho\pi r^2. \tag{10}$$

The integral term in (6) is thus easily approximated by $|\varphi_\Lambda|\pi r^2$ for all θ, while the sum can be explicitly computed.

3.2.2 Some choices leading to quick estimations

The main advantage of the Takacs-Fiksel procedure is to provide quick consistent estimators, that might supply initial values for a more evolved procedure. A simple way to achieve this goal is to generalize (9) and consider test functions of the form

$$h(x^m, \varphi; \theta) = \tilde{h}(x^m, \varphi)e^{V(x^m|\varphi;\theta)},$$

where $\tilde{h}(x^m, \varphi)$ does not depend on θ. So, the integral term in (6) has to be computed only once and not for all θ, while the sum term in (6) does not require any approximation. Hence, the optimisation problem (7) may be resolved very quickly.

In some particular examples, explicit formulas may even be obtained for the integral term, as in (10). In the same spirit, an explicit estimator for the Strauss interaction is provided below.

3.2.3 An example of explicit estimator

The (non-marked) Strauss process with range of interaction $R > 0$ is given for any $\Lambda \Subset \mathbb{R}^d$ by

$$V_\Lambda(\varphi; \theta) = \theta_1|\varphi_\Lambda| + \theta_2 \sum_{\substack{\{x,y\}\in\mathcal{P}_2(\varphi)\\ \{x,y\}\cap\Lambda\neq\emptyset}} \mathbf{1}_{[0,R]}(|x - y|),$$

where $\theta_1 \in \mathbb{R}$ and $\theta_2 > 0$ are the two parameters of the model. Alternatively,

$$V(x|\varphi; \theta) = \theta_1 + \theta_2 \sum_{y\in\varphi} \mathbf{1}_{[0,R]}(|y - x|).$$

Let us consider the following family of test functions, for $k \in \mathbb{N}$,

$$h_k(x, \varphi; \theta) = \begin{cases} e^{k\theta_2} & \text{if } \left|\varphi_{\mathcal{B}(x,R)}\right| = k,\\ 0 & \text{otherwise.}\end{cases} \tag{11}$$

This choice gives in (6)

$$C_\Lambda(\varphi; h_k, \theta) = e^{-\theta_1}V_{k,\Lambda}(\varphi) - e^{k\theta_2}N_{k,\Lambda}(\varphi),$$

where $N_{k,\Lambda}(\varphi)$ denotes the number of points $x \in \varphi_\Lambda$ such that $|\mathcal{B}(x, R)\cap\varphi| = k$ and $V_{k,\Lambda}(\varphi)$ denotes the volume of the set $\{y \in \Lambda, |\mathcal{B}(y, R)\cap\varphi| = k\}$.

Several explicit estimators may be obtained following (7) from (at least) two test functions as above. Let us quote the simplest one, corresponding to the choice h_0 and h_1 in (7). This leads to the contrast function $C_\Lambda(\varphi; h_0, \theta)^2 + C_\Lambda(\varphi; h_1, \theta)^2$ which vanishes at the unique point $(\widehat{\theta}_{1,n}(\varphi), \widehat{\theta}_{2,n}(\varphi))$ with

$$\widehat{\theta}_{1,n}(\varphi) = \ln\left(\frac{V_{0,\Lambda}(\varphi)}{N_{0,\Lambda}(\varphi)}\right), \quad \widehat{\theta}_{2,n}(\varphi) = \ln\left(\frac{V_{1,\Lambda}(\varphi)}{N_{1,\Lambda}(\varphi)}\right) - \ln\left(\frac{V_{0,\Lambda}(\varphi)}{N_{0,\Lambda}(\varphi)}\right).$$

This estimator of (θ_1, θ_2) is completely explicit, provided the quantities $N_{k,\Lambda}(\varphi)$ and $V_{k,\Lambda}(\varphi)$ are available. They can be easily approximated by computational geometry tools.

3.2.4 A solution for unobservability issues

The Quermass model introduced in [27] is a marked point process which aims at modelling random sets in $\mathbb{R}^2$. This is a generalization of the well-known Boolean model to interacting random balls. Let us denote by x^R a marked point where x and $R > 0$ (i.e. the mark) respectively represent the center and the radius of the associated ball $\mathcal{B}(x, R)$. For a finite configuration φ, i.e. with a finite support instead of $\mathbb{R}^2$, the Quermass energy is defined for $(\theta_1, \theta_2, \theta_3, \theta_4) \in \mathbb{R}^4$ by:

$$V(\varphi; \theta) = \theta_1 |\varphi| + \theta_2 \, \mathcal{P}(\Gamma) + \theta_3 \, \mathcal{A}(\Gamma) + \theta_4 \, \mathcal{E}(\Gamma) \quad \text{where } \Gamma = \bigcup_{(x,R)\in\varphi} \mathcal{B}(x, R)$$

and $\mathcal{P}(\Gamma)$, $\mathcal{A}(\Gamma)$ and $\mathcal{E}(\Gamma)$ denote respectively the perimeter, the area and the Euler-Poincaré characteristic (i.e. number of components minus number of holes) of the set Γ. To extend this definition to the infinite support $\mathbb{R}^2$, it is convenient to suppose that the radius of the balls are almost surely uniformly bounded (i.e. $\lambda^{\mathrm{m}}([0, R_0]) = 1$ for some $R_0 > 0$). In this case the family of energies (V_Λ) is defined by

$$V_\Lambda(\varphi; \theta) = V\left(\varphi_{\Lambda \oplus B(0,2R_0)}; \theta\right) - V\left(\varphi_{\Lambda \oplus B(0,2R_0)\setminus\Lambda}; \theta\right).$$

This definition may be extended to unbounded radius, though a restriction to the so-called tempered configurations is needed to ensure the existence of the associated Gibbs measure. We refer to [13] for more details.

When $\theta_2 = \theta_3 = \theta_4 = 0$, this model reduces to the Boolean model (see [38] for a survey). The area process (see [3]) is also a particular case, taking $\theta_2 = \theta_4 = 0$.

In practice, one only observes the random set Γ, so the marked points x^R in φ are unknown. A challenging task is then to estimate the parameters $(\theta_1, \theta_2, \theta_3, \theta_4)$ in the presence of this unobservability issue. In particular, a direct application of the maximum likelihood or pseudo-likelihood method is impossible to estimate all the parameters, and especially θ_1 which requires the observation of the number of points in φ. For the other parameters, which are related to the observable functionals $\mathcal{P}$, $\mathcal{A}$ and $\mathcal{E}$, the MLE has been investigated in [31].

Let us show that the Takacs-Fiksel procedure may be used to estimate θ_1 in spite of this unobservability issue. Indeed, it is possible to choose some test function h such that both the integral and the sum in (6) are computable. The integral term actually always involves observable quantities, since the point x^R in the integral is added to the configuration φ. The unobservability issue may occur for the sum term. Let us consider the following example of test function:

$$h_{per}(x^R, \varphi; \theta) = \mathcal{P}\left(\mathcal{C}(x, R) \cap \Gamma^c\right), \tag{12}$$

where $\mathcal{C}(x, R)$ is the sphere $\{y, |y - x| = R\}$. For any finite configuration φ, we then have

$$\sum_{x^R \in \varphi} h_{per}(x^R, \varphi \setminus x^R; \theta) = \mathcal{P}(\Gamma),$$

so that this sum is computable even if each term $h_{per}(x^R, \varphi \setminus x^R; \theta)$ is not. If the configuration φ is infinite then for any bounded set Λ, $\sum_{x^R \in \varphi_\Lambda} h_{per}(x^R, \varphi \setminus x^R; \theta)$ is equal to the perimeter of Γ restricted to Λ plus a boundary term which is asymptotically negligible with respect to the volume of Λ.

Consequently, assuming $(\theta_2, \theta_3, \theta_4)$ known, θ_1 may be estimated thanks to (7) with the above choice as test function.

The joint estimation of all four parameters might be achieved by a combination of the MLE and the Takacs-Fiksel method, or thanks to additional test functions sharing the same property as above, i.e. such that the sum in (6) is observable.

In the particular case of the area process, $\theta_2 = \theta_4 = 0$ and R is constant (i.e. $\lambda^{\mathrm{m}} = \delta_R$), it suffices to find one more test function to ensure an identifiable estimation (see Example 2 in Section 5 for more details). A possible additional test function is

$$h_{iso}(x^R, \varphi; \theta) = \begin{cases} 1 & \text{if} \quad \mathcal{P}\left(\mathcal{C}(x, R)\right) = 2\pi R \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

In this case $\sum_{x^R \in \varphi} h_{iso}(x^R, \varphi \setminus x^R; \theta)$ corresponds to the number of isolated balls in Γ.

4 Asymptotic results for the Takacs-Fiksel estimator

We present in this section asymptotic results for the Takacs-Fiksel estimator for a point process satisfying **[Mod]** and assumed to be observed in a domain Λ_n, where $(\Lambda_n)_{n\geq 1}$ is a sequence of increasing cubes whose size goes to $+\infty$ as n goes to $+\infty$.

First, for a function g depending on θ, we denote by $\mathbf{g}^{(1)}(\theta)$ (resp. $\underline{\mathbf{g}}^{(2)}(\theta)$) the gradient vector of length p (resp. the Hessian matrix of size (p,p)) evaluated at θ. Let us rewrite the Takacs-Fiksel estimator as

$$\widehat{\theta}_n(\varphi) = \arg\min_{\theta\in\mathbf{\Theta}} U_{\Lambda_n}(\varphi;\mathbf{h},\theta),$$

with $U_{\Lambda_n}(\varphi;\mathbf{h},\theta) = |\Lambda_n|^{-2}\ \sum_{k=1}^K C_{\Lambda_n}(\varphi;h_k,\theta)^2$, where $\mathbf{h}=(h_1,\ldots,h_K)$ and C_{Λ_n} is given by (6).

4.1 Consistency

The consistency is obtained under the following assumptions, denoted by **[C]**: for any Gibbs measure $P_{\theta^\star}$, for all $\theta\in\mathbf{\Theta}$, $k=1,\ldots,K$ and $\varphi\in\Omega$

[C1] $\mathbf{E}\left(\left|h_k\left(0^M,\Phi;\theta\right)\right| e^{-V\left(0^M|\Phi;\theta'\right)}\right) < +\infty,$ for $\theta'=\theta,\theta^\star$.

[C2] $U_{\Lambda_n}(\varphi;\mathbf{h},\cdot)$ is a continuous function for $P_{\theta^\star}-$a.e. φ.

[C3]

$$\sum_{k=1}^K \mathbf{E}\left(h_k(0^M,\Phi;\theta)\left(e^{-V(0^M|\Phi;\theta)} - e^{-V(0^M|\Phi;\theta^\star)}\right)\right)^2 = 0 \quad \Longrightarrow \theta=\theta^\star. \tag{14}$$

[C4] h_k and f_k, defined by $f_k(x^m,\varphi;\theta) := h_k(x^m,\varphi;\theta)e^{-V(x^m|\varphi;\theta)}$, are continuously differentiable and

$$\mathbf{E}\left(\max_{\theta\in\mathbf{\Theta}}|f_k(0^M,\Phi;\theta)|\right) < +\infty \quad \text{and} \quad \mathbf{E}\left(\max_{\theta\in\mathbf{\Theta}}|h_k(0^M,\Phi;\theta)|e^{-V(0^M|\Phi;\theta^\star)}\right) < +\infty,$$

$$\mathbf{E}\left(\max_{\theta\in\mathbf{\Theta}}\|\mathbf{f_k}^{(1)}(0^M,\Phi;\theta)\|\right) < +\infty \quad \text{and} \quad \mathbf{E}\left(\max_{\theta\in\mathbf{\Theta}}\|\mathbf{h_k}^{(1)}(0^M,\Phi;\theta)\|e^{-V(0^M|\Phi;\theta^\star)}\right) < +\infty.$$

Proposition 2 *Assuming **[Mod]** and **[C]** then, as $n\to+\infty$, the Takacs-Fiksel estimator $\widehat{\theta}_n(\varphi)$ converges towards $\theta^\star$ for $P_{\theta^\star}-$a.e. φ.*

Assumptions **[C1]**, **[C2]** and **[C4]** are related to the regularity and the integrability of the different test functions and the local energy function. Some general criteria may be proposed to verify these assumptions, see Section 4.3 for a discussion. Assumption **[C3]** corresponds to an identifiability condition and requires much more attention. It is well-known that such an assumption is fulfilled when $\mathbf{h}=\mathbf{V}^{(1)}$ (leading to the MPLE) under mild assumptions (see Assumption **[Ident]** proposed by [9]). The question to know if this remains true for more general test functions is difficult (actually it is untrue in several cases). This will be discussed specifically in Section 5.

4.2 Asymptotic normality

We need the following assumptions denoted by **[N]**: For any Gibbs measure $P_{\theta^\star}$, $k=1,\ldots,K$, $\Lambda\Subset\mathbb{R}^d$, $\varphi\in\Omega$ and θ in a neighborhood $\mathcal{V}(\theta^\star)$ of $\theta^\star$:

[N1] $\mathbf{E}\left(|C_\Lambda(\Phi;h_k,\theta^\star)|^3\right) < +\infty.$

[N2] For any sequence of bounded domains Γ_n such that $\Gamma_n\to 0$ as $n\to+\infty$, $\mathbf{E}\left(C_{\Gamma_n}(\Phi;h_k,\theta^\star)^2\right)\to 0$.

[N3] $C_\Lambda(\varphi;h_k,\theta^\star)$ depends only on $\varphi_{\Lambda\oplus B(0,D)}$ for some $D\geq 0$ (which is uniform in Λ, φ, $\theta^\star$).

[N4] h_k and f_k (defined in **[C4]**) are twice continuously differentiable in θ and

$$\mathbf{E}\left(\|\underline{\mathbf{h}}^{(2)}(0^M,\Phi;\theta)\|e^{-V\left(0^M|\Phi;\theta^\star\right)}\right)<+\infty \quad \text{and} \quad \mathbf{E}\left(\|\underline{\mathbf{f}}^{(2)}(0^M,\Phi;\theta)\|\right)<+\infty.$$

Let us remark that Assumption **[N3]** leads to consider in general that V has a finite range, which means that there exists $D\geq 0$ such that for all $(m,\varphi)\in\mathbb{M}\times\Omega$ and all $\theta\in\mathbf{\Theta}$

$$V\left(0^m|\varphi;\theta\right)=V\left(0^m|\varphi_{\mathcal{B}(0,D)};\theta\right).$$

The same kind of finite range property is also expected for (h_k).

Proposition 3 *Under Assumptions **[Mod]**, **[C]** and **[N]**, for any ergodic Gibbs measure $P_{\theta^\star}$ the following convergence in distribution holds as $n\to+\infty$*

$$|\Lambda_n|^{1/2}\underline{\mathcal{E}}(\mathbf{h},\theta^\star)\underline{\mathcal{E}}(\mathbf{h},\theta^\star)^T\left(\widehat{\theta}_n(\Phi)-\theta^\star\right)\xrightarrow{d}\mathcal{N}\left(0,\underline{\mathcal{E}}(\mathbf{h},\theta^\star)\underline{\mathbf{\Sigma}}(\mathbf{h},\theta^\star)\underline{\mathcal{E}}(\mathbf{h},\theta^\star)^T\right), \tag{15}$$

where $\underline{\mathcal{E}}(\mathbf{h},\theta^\star)$ is the (p,K) matrix defined for $i=1,\ldots,p$ and $k=1,\ldots,K$ by

$$\left(\underline{\mathcal{E}}(\mathbf{h},\theta^\star)\right)_{ik}=\mathbf{E}\left(h_k(0^M,\Phi;\theta^\star)\left(\mathbf{V}^{(1)}(0^M|\Phi;\theta^\star)\right)_i e^{-V(0^M|\Phi;\theta^\star)}\right)$$

and where $\underline{\mathbf{\Sigma}}(\mathbf{h},\theta^\star)$ is the (K,K) matrix defined by

$$\underline{\mathbf{\Sigma}}(\mathbf{h},\theta^\star)=D^{-d}\sum_{|\ell|\leq 1}\mathbf{E}\left(\widetilde{\mathbf{C}}_{\Delta_0(D)}(\Phi;\mathbf{h},\theta^\star)\widetilde{\mathbf{C}}_{\Delta_\ell(D)}(\Phi;\mathbf{h},\theta^\star)^T\right), \tag{16}$$

where, for all $k\in\mathbb{Z}^d$, $\Delta_k(D)$ is the cube centered at kD with side-length D and where, for some bounded domain Λ, $\widetilde{\mathbf{C}}_\Lambda(\Phi;\mathbf{h},\theta^\star):=\left(C_\Lambda(\Phi;h_k,\theta^\star)\right)_{k=1,\ldots,K}$.

Remark 1 *While Assumptions **[N1-3]** will ensure that a central limit theorem holds for $\mathbf{U}^{(1)}_{\Lambda_n}(\Phi;\mathbf{h},\theta^\star)$, Assumption **[N4]** is required to prove that $\underline{\mathbf{U}}^{(2)}_{\Lambda_n}(\Phi;\mathbf{h},\theta)$ is uniformly bounded in a neighborhood of $\theta^\star$. These two statements allow us to apply a general central limit theorem for minimum contrast estimators (e.g. Theorem 3.4.5 of [22]).*

Remark 2 *In Proposition 3, if the Gibbs measure $P_{\theta^\star}$ is stationary (and then not necessarily ergodic), it is a mixing of ergodic measures and the left hand term in (15) converges in distribution to a mixing of normal distributions. We may also propose an asymptotic result valid in the presence of a phase transition. Indeed, if the matrix $\underline{\mathcal{E}}(\mathbf{h},\theta^\star)\underline{\mathbf{\Sigma}}(\mathbf{h},\theta^\star)\underline{\mathcal{E}}(\mathbf{h},\theta^\star)^T$ is positive definite (for any extremal ergodic measure of the Choquet simplex of stationary Gibbs measures), then we may define a consistent empirical version $S(\Phi)$ of $\left[\underline{\mathcal{E}}(\mathbf{h},\theta^\star)\underline{\mathbf{\Sigma}}(\mathbf{h},\theta^\star)\underline{\mathcal{E}}(\mathbf{h},\theta^\star)^T\right]^{-1/2}\underline{\mathcal{E}}(\mathbf{h},\theta^\star)\underline{\mathcal{E}}(\mathbf{h},\theta^\star)^T$ (see Section 4 of [11] for more details) to obtain:*

$$|\Lambda_n|^{1/2}S(\Phi)\left(\widehat{\theta}_n(\Phi)-\theta^\star\right)\xrightarrow{d}\mathcal{N}\left(0,\underline{\mathbf{I}}\right).$$

Remark 3 *Following Section 3.2.1, let us underline that (15) is coherent with the asymptotic normality of the MPLE established in [10], i.e. with the case $K=p$ and $\mathbf{h}=\mathbf{V}^{(1)}$. Indeed, with similar assumptions to the ones presented in the present paper, Theorem 2 Equation (4.4) in [10] states that*

$$|\Lambda_n|^{1/2}\underline{\mathbf{A}}(\theta^\star)\left(\widehat{\theta}_n(\Phi)-\theta^\star\right)\right)\xrightarrow{d}\mathcal{N}\left(0,\underline{\mathbf{\Sigma}}(\mathbf{V}^{(1)},\theta^\star)\right), \tag{17}$$

where $\underline{\mathbf{A}}(\theta^\star)$ is the symmetric (p,p) matrix given for $i,k=1,\ldots,p$ by

$$\left(\underline{\mathbf{A}}(\theta^\star)\right)_{ik}=\mathbf{E}\left(\left(\mathbf{V}^{(1)}(0^M|\Phi;\theta^\star)\right)_k\left(\mathbf{V}^{(1)}(0^M|\Phi;\theta^\star)\right)_i e^{-V(0^M|\Phi;\theta^\star)}\right).$$

Since $\mathbf{h}=\mathbf{V}^{(1)}$, then $\underline{\mathcal{E}}(\mathbf{V}^{(1)},\theta^\star):=\underline{\mathbf{A}}(\theta^\star)$ and therefore (15) reduces to

$$|\Lambda_n|^{1/2}\underline{\mathbf{A}}(\theta^\star)^2\left(\widehat{\theta}_n(\Phi)-\theta^\star\right)\right)\xrightarrow{d}\mathcal{N}\left(0,\underline{\mathbf{A}}(\theta^\star)\underline{\mathbf{\Sigma}}(\mathbf{V}^{(1)},\theta^\star)\underline{\mathbf{A}}(\theta^\star)\right)$$

which is exactly (17) by assuming that $\underline{\mathbf{A}}(\theta^\star)$ is invertible.

Remark 4 *The question how to choose the test functions in order to minimize the norm of the asymptotic covariance matrix is difficult to answer, still open and is a perspective for future work.*

4.3 Discussion

The present paragraph is devoted to the discussion of Assumptions [**Mod**], [**C**] (except [**C3**]) and [**N**]. In the previous sections, we have expressed the different assumptions in a very general way. Our aim, here, is to make these assumptions concrete for a wide range of models and a wide range of test functions, in order to illustrate that our setting is not restrictive. In particular, we will focus on exponential family models having a local energy of the form:

$$V(x^m|\varphi;\theta) := \theta^T \mathbf{V}(x^m|\varphi) = \theta_1 V_1(x^m|\varphi) + \ldots + \theta_p V_p(x^m|\varphi), \tag{18}$$

with $\mathbf{V} = (V_1, \ldots, V_p)$ a vector function from $\mathbb{S} \times \Omega(\mathbb{S})$ to $\mathbb{R}^p$.

Let us consider the following assumptions: for all $(m, \varphi) \in \mathbb{M} \times \Omega$.

[**Exp**] For $i = 1, \cdots, p$, there exist $\kappa_i^{(\inf)}, \kappa_i^{(\sup)} \geq 0$, $k_i \in \mathbb{N}, D > 0$ such that one of the two following assumptions is satisfied :

$$\theta_i \geq 0 \text{ and } -\kappa_i^{(\inf)} \leq V_i(0^m|\varphi) = V_i(0^m|\varphi_{\mathcal{B}(0,D)}) \leq \kappa_i^{(\sup)} |\varphi_{\mathcal{B}(0,D)}|^{k_i}.$$

or

$$-\kappa_i^{(\inf)} \leq V_i(0^m|\varphi) = V_i(0^m|\varphi_{\mathcal{B}(0,D)}) \leq \kappa_i^{(\sup)}.$$

$\widetilde{[\mathbf{Exp}]}$ Assumption [**Exp**] with $k_i = 0$ or 1 for all i (when $\theta_i \geq 0$).

[**H**] There exist $\kappa > 0$, $k \in \mathbb{N}, D > 0$ such that $h(0^m, \varphi; \theta) = h(0^m, \varphi_{\mathcal{B}(0,D)}; \theta)$, such that $h(0^m, \varphi; \cdot)$ is twice continuously differentiable in θ and such that $|Y(\varphi, m)| \leq \kappa |\varphi_{\mathcal{B}(0,D)}|^k$, where

$$Y(\varphi, m) := \max\left(|h(0^m, \varphi; \theta)|, \|\mathbf{h}^{(1)}(0^m, \varphi; \theta)\|, \|\underline{\mathbf{h}}^{(2)}(0^m, \varphi; \theta)\|\right). \tag{19}$$

$\widetilde{[\mathbf{H}]}$ $h(0^m, \varphi; \theta) = \widetilde{h}(0^m, \varphi; \theta) e^{\theta^T \mathbf{V}(0^m|\varphi)}$ with $\widetilde{h}$ satisfying [**H**].

Let us underline that the different constants involved in these assumptions are assumed to be independent of m, φ, θ. Note also that if the test function h is independent of θ, $Y(\varphi, m)$ obviously reduces to $|h(0^m, \varphi)|$.

These assumptions are common and very simple to check. Assumption [**Exp**] has already been investigated in [9]. It includes a wide variety of models such as the overlap area point process, the multi-Strauss marked point process, the $k-$nearest-neighbor multi-Strauss marked point process, the Strauss type disc process, the Geyer's triplet point process, the area process, some special cases of Quermass process (for instance when $\lambda^{\mathbb{m}}$ has a compact support not containing 0 and $\theta_4 = 0$), etc. Among these models, the only one that does not satisfy $\widetilde{[\mathbf{Exp}]}$ is the Geyer's triplet point process (see [9] p.242).

On the other hand, the test functions $h(x^m|\varphi) = 1$, $h(x^m|\varphi) = |\varphi_{\mathcal{B}(x,r)}|$, $h(x^m|\varphi) = \mathbf{V}_k^{(1)}(x^m|\varphi;\theta) = V_k(x^m|\varphi)$ satisfy [**H**] (for the last one, it is implied by [**Exp**]). Note that the functional described by (11) for the Strauss model depending on θ, the functionals h_{per} and h_{iso} in (12), (13) also satisfy [**H**]. In a similar way, test functions like $e^{\theta^T \mathbf{V}(x^m|\varphi)}$, $e^{\theta^T \mathbf{V}(x^m|\varphi)/2}$, $|\varphi_{\mathcal{B}(x,r)}| e^{\theta^T \mathbf{V}(x^m|\varphi)}$, $\mathbf{1}_{[0,r]}(d(x^m, \varphi)) e^{\theta^T \mathbf{V}(x^m|\varphi)}$ satisfy $\widetilde{[\mathbf{H}]}$.

We show that most of all the assumptions required in Propositions 2 and 3 are not too restrictive.

Proposition 4 (*i*) *Under Assumption **[Exp]**, Assumption **[Mod]** is fulfilled.*
(*ii*) *For a test function satisfying **[H]** (resp. $\widetilde{[\mathbf{H}]}$) and a model satisfying **[Exp]** (resp. $\widetilde{[\mathbf{Exp}]}$), then Assumptions **[C]** (excepted **[C3]**) and **[N]** are fulfilled.*

Proof. (*i*) We refer the reader to [9] (in particular Section 3) where it is proved that [**Exp**] ensures the stability and the finite range of the local energy, which ensures the existence of ergodic measures (see *e.g.* [5] for details).

(*ii*) [**C2**] and [**N3**] are quite obvious to check. Now, recall that the stability of the local energy means that for all $(m, \varphi) \in \mathbb{M} \times \Omega$, $\theta^T \mathbf{V}(0^m|\varphi) \geq -\rho$ for $\rho < +\infty$, independent of m, φ, θ. Let us also underline

that the local stability property ensures that for every $\Lambda \Subset \mathbb{R}^d$, every $c \in \mathbb{R}$, $\mathbf{E}(e^{c|\Phi_\Lambda|}) < +\infty$ (see e.g. Proposition 11 of [6]), which obviously implies that $\mathbf{E}(|\Phi_\Lambda|^\alpha) < +\infty$ and $\mathbf{E}(|\Phi_\Lambda|^\alpha e^{c|\Phi_\Lambda|}) < +\infty$ for every $\alpha > 0$. Now, under **[H]** and **[Exp]** (or $\widetilde{\mathbf{[H]}}$ and $\widetilde{\mathbf{[Exp]}}$), the expectations in **[C1]**, **[C4]**, **[N1]** and **[N4]** are clearly finite. Let us focus on **[C1]** for example (the justification for the other assumptions is similar). We have for any $\theta, \theta' \in \mathbf{\Theta}$

$$\mathbf{E}\left(|h(0^M,\Phi;\theta)|e^{-\theta'^T\mathbf{v}(0^M|\Phi)}\right)$$
$$\leq \begin{cases} e^\rho \mathbf{E}\left(|h(0^M,\Phi;\theta)|\right) \leq c \times e^\rho \mathbf{E}\left(|\Phi_{\mathcal{B}(0,D)}|^\alpha\right) & \text{under } \mathbf{[H]} \text{ and } \mathbf{[Exp]} \\ e^\rho \mathbf{E}\left(|\widetilde{h}(0^M,\Phi;\theta)|e^{\theta^T\mathbf{v}(0^M|\Phi)}\right) \leq c \times e^\rho \mathbf{E}\left(|\Phi_{\mathcal{B}(0,D)}|^\alpha e^{c|\Phi_{\mathcal{B}(0,D)}|}\right) & \text{under } \widetilde{\mathbf{[H]}} \text{ and } \widetilde{\mathbf{[Exp]}}, \end{cases}$$

for some constants α and c. Note that, if we had not assumed $\widetilde{\mathbf{[Exp]}}$ for test functions of the form $\widetilde{\mathbf{[H]}}$, then one would have had expectations of the form $\mathbf{E}\left(e^{|\Phi_\Lambda|^k}\right)$ for some $k > 1$ which is not necessary finite under the local stability property. Asumption **[N2]** is proved similarly and by using the dominated convergence theorem. ■

Remark 5 *By following ideas in [10], it is possible to fulfill the integrability type assumptions for more complicated models such as the Lennard-Jones model (which is not locally stable and nonlinear in terms of the parameters). The using of Ruelle's estimates [37] plays a crucial role in this case of superstable interaction. For the sake of conciseness and simplicity, we do not investigate this in the present paper.*

5 Identifiability : Assumption [C3]

The Assumption **[C3]** is related to the identifiability of the estimation procedure. It is more complicated to verify than the other assumptions and an investigation to obtain a criterion or a characterization seems necessary. We address this question in this section.

In the following, we consider that the interaction has an exponential form as in (18). Then **[C3]** is equivalent to: $\theta = \theta^\star$ is the unique solution of the nonlinear system of equations in θ defined by

$$\mathbf{E}\left(h_k(0^M,\Phi;\theta)\left(e^{-\theta^T\mathbf{V}(0^M|\Phi)} - e^{-\theta^{\star T}\mathbf{V}(0^M|\Phi)}\right)\right) = 0, \qquad 1 \leq k \leq K. \tag{20}$$

If h_k and $\mathbf{V}$ are sufficiently regular, each equation in (20) gives a $(p-1)$-dimensional manifold of solutions in $\mathbf{\Theta}$ containing $\theta^\star$. So it is clear that the choice $K \geq p$ is in general necessary to prove that the system (20) admits the unique solution $\theta^\star$.

In Section 5.1, we investigate the delicate case $K = p$ in detail. In opposition to the linear case where p hyperplanes in $\mathbb{R}^p$ have in general a unique common point, the intersection of p $(p-1)$-dimensional manifolds does not generally reduce to a single point. So, when $K = p$, there is no guarantee that (20) has a unique solution $\theta^\star$. This is illustrated by a simple example at the beginning of Section 5.1. In Proposition 5, we provide a criterion to ensure that the system in (20) admits the only one solution $\theta^\star$. Some examples, for which the criterion is available, are presented and the rigidity of the criterion, when $p \geq 3$, is also evoked. In the case where $p = 2$, we show that our criterion is not far from being necessary.

The case $K > p$ is studied in Section 5.2. The identification problem should be simpler since, in general, $p+1$ $(p-1)$-dimensional manifolds in $\mathbb{R}^p$ have no common point. We give a sufficient criterion to prove the identification but we think that it is far from being necessary.

Before presenting these two sections, let us give further notation. We denote by P_V the law of $\mathbf{V}(0^M,\Phi)$ in $\mathbb{R}^p$. We also define the function Ψ_θ, for each $\theta \in \mathbf{\Theta}$, by

$$\begin{array}{rccc} \Psi_\theta : & \mathbb{R}^p & \longrightarrow & \mathbb{R}^K \\ & \mathbf{v} & \longmapsto & \begin{pmatrix} \mathbf{E}\Big(h_1(0^M,\Phi;\theta)\Big|\mathbf{V}(0^M|\Phi) = \mathbf{v}\Big) \\ \vdots \\ \mathbf{E}\Big(h_K(0^M,\Phi;\theta)\Big|\mathbf{V}(0^M|\Phi) = \mathbf{v}\Big) \end{pmatrix}. \end{array} \tag{21}$$

We will see that this function plays a crucial role in the identification problem.

5.1 The case $K = p$

First of all, let us give a simple example to show that the identification problem is delicate in the situation where $K = p$. Let us consider that $K = p = 2$, $V_1 = 1$, and let us choose the simple test functions $h_1 = 1$ and $h_2 = e^{\theta^T \mathbf{V}(x^m|\phi)}$. Then $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2)$ with $\tilde{\theta}_1 = \theta_1^\star - \ln(\mathbf{E}(e^{-\theta_2^\star V_2(0^M|\Phi)})$ and $\tilde{\theta}_2 = 0$ is always a solution of the system in (20). Therefore if $\theta_2^\star \neq 0$ and if θ defined before is in $\mathbf{\Theta}$, then the system in (20) admits at least two solutions.

In the following, we first give a sufficient criterion to prove the identifiability and propose some examples. Next, we show the rigidity of our criterion which seems constraining when $p \geq 3$.

5.1.1 Criterion for identifiability

Assumption [**Det**] gathers the two following assumptions:

[**Det**($\neq$)] For every θ in $\mathbf{\Theta}$, $det(\mathbf{v}_1, \ldots, \mathbf{v}_p) det\big(\Psi_\theta(\mathbf{v}_1), \ldots, \Psi_\theta(\mathbf{v}_p)\big)$ is not $(P_V)^{\otimes p}$-a.s. identically null

[**Det**($\geq$)] For every θ in $\mathbf{\Theta}$, there exists $\epsilon = \pm 1$ such that for $(P_V)^{\otimes p}$-a.s. every $(\mathbf{v}_1, \ldots, \mathbf{v}_p)$ in $(\mathbb{R}^p)^p$

$$\epsilon \; det(\mathbf{v}_1, \ldots, \mathbf{v}_p) det\big(\Psi_\theta(\mathbf{v}_1), \ldots, \Psi_\theta(\mathbf{v}_p)\big) \geq 0.$$

When $\epsilon = 1$ (respectively $\epsilon = -1$), [**Det**($\geq$)] means that Ψ_θ preserves the sign (respectively the opposite sign) of the determinant.

The criterion is the following.

Proposition 5 *If $K = p$ then Assumption* [**Det**] *ensures that Assumption **[C3]** holds.*

Proof. Denoting by ζ the real function $x \mapsto \ln\left(\frac{e^x - 1}{x}\right)$ with the convention $\zeta(0) = 0$, the equations (20) become

$$(\theta^\star - \theta)^T \mathbf{X}_k(\theta, \theta^\star) = 0, \qquad 1 \leq k \leq p, \tag{22}$$

where the vector $\mathbf{X}_k(\theta, \theta^\star)$ is defined by

$$\mathbf{X}_k(\theta, \theta^\star) = \mathbf{E}\left(h_k(0^M, \Phi; \theta) e^{-\theta^{\star T} \mathbf{V}(0^M|\Phi)} e^{\zeta\left((\theta^\star - \theta)^T \mathbf{V}(0^M|\Phi)\right)} \mathbf{V}(0^M|\Phi) \right). \tag{23}$$

Therefore the system (20) admits the unique solution $\theta^\star$ if the family of vectors $(\mathbf{X}_k(\theta, \theta^\star))_{1 \leq k \leq p}$ is independent in $\mathbb{R}^p$. Let us give a formula of the determinant of these vectors which shows that it is not null.

Conditioning by the law of $\mathbf{V}(0^M|\Phi)$ and using the multi-linearity of the determinant, we obtain

$$\begin{aligned}
& det(\mathbf{X}_1(\theta, \theta^\star), \ldots, \mathbf{X}_p(\theta, \theta^\star)) \\
= & \int \cdots \int det\left(\mathbf{E}\left[h_1(0^M, \Phi; \theta) e^{-\theta^{\star T} \mathbf{v}_1} e^{\zeta\left((\theta^\star - \theta)^T \mathbf{v}_1\right)} \mathbf{v}_1 \middle| \mathbf{V}(0^M|\Phi) = \mathbf{v}_1 \right], \ldots, \right. \\
& \left. \mathbf{E}\left[h_p(0^M, \Phi; \theta) e^{-\theta^{\star T} \mathbf{v}_p} e^{\zeta\left((\theta^\star - \theta)^T \mathbf{v}_p\right)} \mathbf{v}_p \middle| \mathbf{V}(0^M|\Phi) = \mathbf{v}_p \right] \right) P_V(d\mathbf{v}_1) \cdots P_V(d\mathbf{v}_p) \\
= & \int \cdots \int e^{\sum_{k=1}^p -\theta^{\star T} \mathbf{v}_k + \zeta\left((\theta^\star - \theta)^T \mathbf{v}_k\right)} det\left(\mathbf{v}_1, \ldots, \mathbf{v}_p\right) \prod_{k=1}^p \mathbf{E}\left[h_k(0^M, \Phi; \theta) | \mathbf{v}_k \right] P_V(d\mathbf{v}_1) \cdots P_V(d\mathbf{v}_p) \\
= & \frac{1}{p!} \sum_{\sigma \in S_p} \int \cdots \int e^{\sum_{k=1}^p -\theta^{\star T} \mathbf{v}_{\sigma(k)} + \zeta\left((\theta^\star - \theta)^T \mathbf{v}_{\sigma(k)}\right)} det\left(\mathbf{v}_{\sigma(1)}, \ldots, \mathbf{v}_{\sigma(p)}\right) \\
& \prod_{k=1}^p \mathbf{E}\left[h_k(0^M, \Phi; \theta) | \mathbf{v}_{\sigma(k)} \right] P_V(d\mathbf{v}_1) \cdots P_V(d\mathbf{v}_p),
\end{aligned}$$

where S_p is the set of all permutations in $\{1,\dots,p\}$. Denoting by $\epsilon(\sigma)$ the signature of σ, we obtain

$$
\begin{aligned}
& det(\mathbf{X}_1(\theta,\theta^\star),\dots,\mathbf{X}_p(\theta,\theta^\star)) \qquad (24)\\
= \;& \frac{1}{p!}\int\cdots\int e^{\sum_{k=1}^p -\theta^{\star T}\mathbf{v}_k+\zeta\left((\theta^\star-\theta)^T\mathbf{v}_k\right)} det\left(\mathbf{v}_1,\dots,\mathbf{v}_p\right)\sum_{\sigma\in S_p}\epsilon(\sigma)\prod_{k=1}^p \mathbf{E}\left[h_k(0^M,\Phi;\theta)|\mathbf{v}_{\sigma(k)}\right]\\
& P_V(d\mathbf{v}_1)\cdots P_V(d\mathbf{v}_p)\\
= \;& \frac{1}{p!}\int\cdots\int e^{\sum_{k=1}^p -\theta^{\star T}\mathbf{v}_k+\zeta\left((\theta^\star-\theta)^T\mathbf{v}_k\right)} det(\mathbf{v}_1,\dots,\mathbf{v}_p) det\big(\Psi_\theta(\mathbf{v}_1),\dots,\Psi_\theta(\mathbf{v}_p)\big)P_V(d\mathbf{v}_1)\cdots P_V(d\mathbf{v}_p).
\end{aligned}
$$

From Assumption **[Det]** , this determinant is not null. The Proposition is proved. ■

Now let us give some examples for which the criterion is available.

Example 1 (linear case) *If the function Ψ_θ is linear and invertible then Assumption* **[Det(≥)]** *is clearly satisfied and* **[Det(≠)]** *holds as soon as the support of P_V is not included in a hyperplane. In particular, if $h_k = V_k$ for every $1 \le k \le p$ then Ψ_θ is equal to the identity function. This situation corresponds to the pseudo-likelihood procedure for which we regain the identifiability via our criterion.*

Example 2 (area process) *For the area process defined in Section 3.2.4 with $\lambda^{\mathrm{m}} = \delta_R$ (i.e. the radii of balls are constant), it is easy to check that the functions h_{per} and h_{iso} respectively defined by* (12) *and* (13) *give a function Ψ which satisfies Assumption* **[Det]** *. Indeed the support of P_V is the segment $\{1\}\times[0,\pi R^2]$ in $\mathbb{R}^2$ and for a vector $\mathbf{v}=(1,v_2)$ the image $\Psi(\mathbf{v})$ is $(\psi_1(v_2),0)$ if $V_2 \neq \pi R^2$ and $(2\pi R,1)$ if $v_2 = \pi R^2$. Therefore, it follows that* **[Det(≥)]** *is satisfied and noting that $0 < P_V((1,\pi R^2)) < 1$ we deduce that* **[Det(≠)]** *holds too.*

Example 3 (a general example with $p=2$) *Example 2 is included in a more general setting when $p=2$. Indeed let us suppose that function Ψ_θ has the form $\Psi_\theta(v_1,v_2) = \big(g_\theta(v_1,v_2), g_\theta(v_1,v_2)f_\theta(v_2/v_1)\big)$ where g_θ is a nonnegative scalar function and f_θ is a monotone scalar function. Then Ψ_θ satisfies* **[Det(≥)]** *and* **[Det(≠)]** *holds if $g_\theta(v_1,v_2)f_\theta(v_2/v_1)$ is not $P_V^{\otimes 2}$-a.s. constant when $g_\theta(v_1,v_2)$ is not null.*

Example 4 (functions of the type $h_k e^{\theta^T\mathbf{V}}$) *Let us suppose that the functions (h_k) ensure that Ψ_θ satisfies* **[Det(≥)]** *then for any nonnegative function g_θ from $\mathbb{R}^p$ to $\mathbb{R}$ the functions $(\tilde{h}_k) = \big(g_\theta(\mathbf{V})h_k\big)$ also provide a function $\tilde{\Psi}_\theta$ satisfying* **[Det(≥)]** *. This remark is related to Section 3.2.2 where it is suggested to choose functions $(\tilde{h}_k)$ of the form $(e^{\theta^T\mathbf{V}}h_k)$ to simplify the integral in (6).*
As an immediate consequence, the test functions $\big(V_k e^{\theta^T\mathbf{V}}\big)$, considered in [8] for the particular multi-Strauss point process, satisfy **[Det(≥)]** *.*

5.1.2 Rigidity of the criterion

In this section, we give some comments about the rigidity of the criterion. In Proposition 6 below, we show that a function Ψ_θ, satisfying **[Det(≥)]** , has a strong linear structure since, under very reasonable assumptions, the image of any hyperplane is included in a hyperplane. For example, in the classical setting where $V_1 = 1$, the function Ψ_θ is defined from the affine space $H = \{1\}\times\mathbb{R}^{p-1}$ and if Ψ is assumed to be continuous then the image of any $p-2$ dimensional affine space in H is included in a hyperplane. This property clearly shows that Ψ_θ is very rigid when $p \ge 3$.

However, when $p=2$, we show in Proposition 7 that our criterion is not far from being necessary. Indeed, we present a large class of examples which do not satisfy our criterion and for which the identifiability fails.

Proposition 6 *Let Ψ be a continuous function from $\mathcal{D}$ to $\mathbb{R}^p$ satisfying* **[Det(≥)]** *, where the domain $\mathcal{D}$ is a subset of $\mathbb{R}^p$ with the following property: for any $(x_i)_{1\le i\le p}\in\mathcal{D}^p$ such that $det(x_1,\dots,x_p)=0$, then for any neighborhood $\mathcal{V}$ of (x_i), there exist (x_i^+) and (x_i^-) in $\mathcal{V}\cap\mathcal{D}^p$ such that $det(x_1^+,\dots,x_p^+)>0$ and $det(x_1^-,\dots,x_p^-)<0$. Then for any hyperplane H in $\mathbb{R}^p$ the image $\Psi(H\cap\mathcal{D})$ is included in a hyperplane of $\mathbb{R}^p$.*

Proof. Let H be a hyperplane in $\mathbb{R}^p$. To prove that $\Psi(H \cap \mathcal{D})$ is included in a hyperplane, it is sufficient to prove that the dimension of the vectorial space generated by the vectors in $\Psi(H \cap \mathcal{D})$ is not equal to p. Let us suppose that it is equal to p, then there exists $(x_i)_{1\leq i\leq p}$ in $(H \cap \mathcal{D})^p$ such that $det(\Psi(x_1), \ldots, \Psi(x_p)) \neq 0$. Since $dim(H) = p - 1$, we have $det(x_1, \ldots, x_p) = 0$. By continuity of Ψ and by the local properties of $\mathcal{D}$ assumed in Proposition 6, we find $(x_i^+)_{1\leq i\leq p}$ and $(x_i^-)_{1\leq i\leq p}$ in $\mathcal{D}^p$ such that $det(x_1^+, \ldots, x_p^+)det(\Psi(x_1^+), \ldots, \Psi(x_p^+)) > 0$ and $det(x_1^-, \ldots, x_p^-)det(\Psi(x_1^-), \ldots, \Psi(x_p^-)) < 0$, which contradicts Assumption **[Det(≥)]** . ■

In the case where $\mathcal{D} = \mathbb{R}^p$, if we assume that $\Psi(\mathbb{R}^p)$ is not reduced to a hyperplane and that Ψ is differentiable at the origin, then we can show that Ψ satisfies **[Det(≥)]** if and only if $\Psi(x) = g(x)Ax$, where A is an invertible matrix and g a nonnegative scalar function. It means that Ψ is quasi linear and so the rigidity of Ψ is very strong.

Now let us focus on the case where $p = 2$ and let us show that, while our criterion seems very constraining, it is not far from being necessary in this case. We suppose that $p = 2$, $V_1 = 1$ and that the support of V_2 is included in an interval $[a, b]$. Let us remark that this case occurs for the area process with $[a, b] = [0, \pi R^2]$. First of all, it is easy to check visually, depending on the geometry of γ_θ defined by the curve $\Psi_\theta(\{1\} \times [a, b])$, whether Ψ_θ satisfies **[Det]** (see figure 1 for examples).

Moreover let us show that the criterion is not far from being necessary. Suppose that the functions (h_i) do not depend on θ and that $\Psi := \Psi_\theta$ satisfies for $\epsilon = \pm 1$ Assumption $[\widetilde{\mathbf{Det}}]$ decomposed into the three following assumptions:

$[\widetilde{\mathbf{Det}}(\neq)]$ $det(\mathbf{v}_1, \mathbf{v}_2)\, det(\Psi(\mathbf{v}_1), \Psi(\mathbf{v}_2))$ is not $P_V^{\otimes 2}$-a.s. identically null

$[\widetilde{\mathbf{Det}}(\geq)]$ there exists $\delta > 0$ such that for $P_V^{\otimes 2}$-a.s. every $(\mathbf{v}_1, \mathbf{v}_2)$ in $(\{1\} \times [a, a + \delta])^2$

$$\epsilon\, det(\mathbf{v}_1, \mathbf{v}_2)\, det(\Psi(\mathbf{v}_1), \Psi(\mathbf{v}_2)) \geq 0$$

$[\widetilde{\mathbf{Det}}(\leq)]$ there exists $\delta > 0$ such that for $P_V^{\otimes 2}$-a.s. every $(\mathbf{v}_1, \mathbf{v}_2)$ in $(\{1\} \times [b - \delta, b])^2$

$$\epsilon\, det(\mathbf{v}_1, \mathbf{v}_2)\, det(\Psi(\mathbf{v}_1), \Psi(\mathbf{v}_2)) \leq 0.$$

See Figure 1 for an example of such Ψ. Obviously, this situation is not exactly the opposite of Assumption **[Det]**, but it is strongly related to it. Then, we have the following proposition which proves that the identifiability fails for this large class of examples.

Proposition 7 *If the functions (h_i) are nonnegative, if Ψ satisfies $[\widetilde{\mathbf{Det}}]$ and if $det\Big(E_{P_V}(\Psi(\mathbf{v})\mathbf{v}^T)\Big) \neq 0$ then **[C3]** fails.*

Let us note that even if the assumption $det\Big(E_{P_V}(\Psi(\mathbf{v})\mathbf{v}^T)\Big) \neq 0$ seems unnatural, it is in general satisfied.

Proof. Let us show that (20) admits another solution than $\theta^\star$. We only give here the main lines of the proof.

We denote by $\mathcal{O}$ the set in $\mathbb{R}^2$ containing the vectors $\mathbf{u}$ which are orthogonal to at least one vector $\mathbf{v}$ in $\{1\} \times [a, b]$, i.e. $\mathcal{O} = \{\mathbf{u} \in \mathbb{R}^2, \exists \mathbf{v} \in \{1\} \times [a, b], \mathbf{u}^T\mathbf{v} = 0\}$. In fact, $\mathcal{O}$ is the union of a cone $\mathcal{O}^+$ in the upper half plane and a cone $\mathcal{O}^-$ in the lower half plane. For any $\delta > 0$, the expression of the determinant in (24) can be split in two parts

$$\begin{aligned} & det(\mathbf{X}_1(\theta, \theta^\star), \mathbf{X}_2(\theta, \theta^\star)) \qquad (25)\\ = \;& \frac{1}{2}\int\int_{[a,a+\delta]^2} e^{\sum_{k=1}^2 -\theta^{\star T}\mathbf{v}_k + \zeta\left((\theta^\star-\theta)^T\mathbf{v}_k\right)} det(\mathbf{v}_1, \mathbf{v}_2) det\big(\Psi(\mathbf{v}_1), \Psi(\mathbf{v}_2)\big) P_V(d\mathbf{v}_1) P_V(d\mathbf{v}_2)\\ & + \frac{1}{2}\int\int_{[a,b]^2\setminus[a,a+\delta]^2} e^{\sum_{k=1}^2 -\theta^{\star T}\mathbf{v}_k + \zeta\left((\theta^\star-\theta)^T\mathbf{v}_k\right)} det(\mathbf{v}_1, \mathbf{v}_2) det\big(\Psi(\mathbf{v}_1), \Psi(\mathbf{v}_2)\big) P_V(d\mathbf{v}_1) P_V(d\mathbf{v}_2). \end{aligned}$$

Let $\mathbf{u} \neq 0$ in $\mathcal{O}^+$ and $\theta = \theta^\star + \alpha\mathbf{u}$ with $\alpha > 0$. From $[\widetilde{\mathbf{Det}}(\neq)]$ and $[\widetilde{\mathbf{Det}}(\geq)]$ since ζ is increasing and $\zeta(x) \sim x$ as $x \to +\infty$, we deduce that the first integral in (25) dominates the second one when α

goes to infinity. Therefore $det(\mathbf{X}_1(\theta,\theta^\star),\mathbf{X}_2(\theta,\theta^\star))$ has the sign ϵ (defined before $[\widetilde{\mathbf{Det}}]$) when α is large enough. Similarly, if $\mathbf{u}$ is in $\mathcal{O}^-$ then from $[\widetilde{\mathbf{Det}(\neq)}]$ and $[\widetilde{\mathbf{Det}(\leq)}]$ $det(\mathbf{X}_1(\theta,\theta^\star),\mathbf{X}_2(\theta,\theta^\star))$ has the sign $-\epsilon$ when α is large enough and it implies that the sign of $det(\mathbf{X}_1(\theta,\theta^\star),\mathbf{X}_2(\theta,\theta^\star))$, with $\theta=\theta^\star+\mathbf{u}$, is different for $\mathbf{u}$ in $\mathcal{O}^+$ or in $\mathcal{O}^-$ as soon as $|\mathbf{u}|$ is large enough. We deduce that there exists a continuous curve $t\mapsto\mathbf{u}(t)$ which crosses $\mathcal{O}$ such that $det(\mathbf{X}_1(\theta(t),\theta^\star),\mathbf{X}_2(\theta(t),\theta^\star))=0$ for every $\theta(t)=\theta^\star+\mathbf{u}(t)$. Let us note that the assumption $det\Big(E_{P_V}(\Psi(\mathbf{v})\mathbf{v}^T)\Big)\neq 0$ ensures that $det(\mathbf{X}_1(\theta^\star,\theta^\star),\mathbf{X}_2(\theta^\star,\theta^\star))\neq 0$, and so $\mathbf{u}(t)$ is never null. Let us show that there exists t_0 such that $\theta(t_0)$ is a solution of the system in (20).

Since the functions (h_i) are nonnegative and from Definition (23), we obtain that for every t, $\mathbf{X}_1(\theta(t),\theta^\star)$ is collinear to a vector in $\{1\}\times[a,b]$. By continuity of the function $t\mapsto\mathbf{X}_1(\theta(t),\theta^\star)^T\mathbf{u}(t)$ and by the mean value theorem, there exists t_0 such that $\mathbf{u}(t_0)$ is orthogonal to $\mathbf{X}_1(\theta(t_0),\theta^\star)$. Since the determinant $det(\mathbf{X}_1(\theta(t_0),\theta^\star),\mathbf{X}_2(\theta(t_0),\theta^\star))=0$, $\mathbf{u}(t_0)$ is also orthogonal to $\mathbf{X}_2(\theta(t_0),\theta^\star)$ and it follows that the system in (22) provides at least two solutions $\theta^\star$ and $\theta(t_0)$. Identification assumption (20) or **[C3]** fail. ■

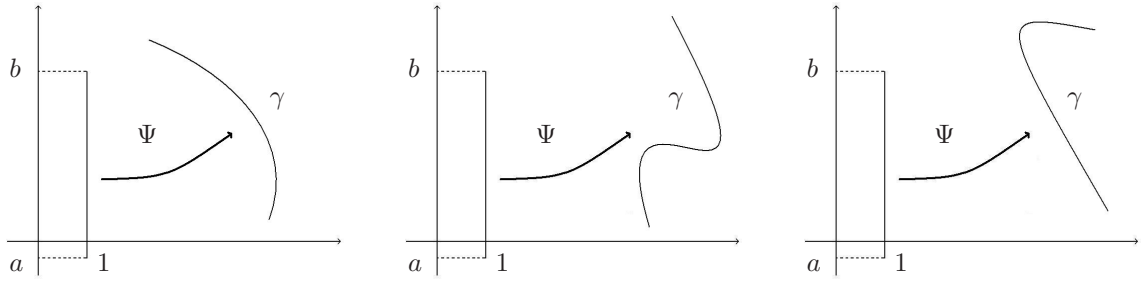


Figure 1: on the left (respectively in the middle), an example of function Ψ satisfying (respectively not satisfying) $[\mathbf{Det}(\geq)]$. On the right, an example of Ψ satisfying $[\widetilde{\mathbf{Det}}]$.

5.2 The case K>p

In the case where $K>p$, we noticed, in the introduction, that the identification problem should be simpler. Nevertheless, we did not find a satisfactory criterion to prove it. The following Proposition 8 gives a sufficient criterion which is probably far from being necessary. It is based on a slight modification of Assumption **[Det]** which does not seem to be the appropriate tool in this setting. However, in the case where $p=2$ and $K=3$, this condition reduces to a nice geometrical property which can be checked easily.

First, let us present the criterion. We denote by $\mathcal{A}$ the set of all subsets with p elements in $\{1,\dots,K\}$,

$$\mathcal{A}=\Big\{I\subset\{1,\dots,K\},\ \text{such that } \#(I)=p\Big\}.$$

We say that Assumption **[Det']** is satisfied if, for every θ in $\mathbf{\Theta}$, there exists a family of real coefficients $(c_I)_{I\in\mathcal{A}}$ such that the two following assumptions hold:

[Det'(≠)] $\displaystyle\sum_{I\in\mathcal{A}} c_I det(\mathbf{v}_1,\dots,\mathbf{v}_p)det\big(\Psi_\theta^I(\mathbf{v}_1),\dots,\Psi_\theta^I(\mathbf{v}_p)\big)$ is not $(P_V)^{\otimes p}$-a.s. identically null

[Det'(≥)] for $(P_V)^{\otimes p}$-a.s. every $(\mathbf{v}_1,\dots,\mathbf{v}_p)$ in $(\mathbb{R}^p)^p$

$$\sum_{I\in\mathcal{A}} c_I det(\mathbf{v}_1,\dots,\mathbf{v}_p)det\big(\Psi_\theta^I(\mathbf{v}_1),\dots,\Psi_\theta^I(\mathbf{v}_p)\big)\geq 0,$$

where $\Psi_\theta^I(\mathbf{v})$ denotes the p-dimensional vector extracted from $\Psi_\theta(\mathbf{v})$ for the coordinates given by I. In the particular case where $K = p$, Assumption **[Det']** becomes Assumption **[Det]** exactly.

Our criterion is the following

Proposition 8 *Assumption* **[Det']** *ensures that Assumption [C3] holds.*

Proof. As in the proof of Proposition 5, to show that (20) admits the unique solution $\theta^\star$, it is sufficient to prove that there exists I in $\mathcal{A}$ such that, for all $\theta \neq \theta^\star$, $det\big((\mathbf{X}_i(\theta,\theta^\star))_{i\in I}\big) \neq 0$. It is equivalent to: for every $\theta \neq \theta^\star$ in $\mathbf{\Theta}$ there exists a family of real coefficients $(c_I)_{I\in\mathcal{A}}$ such that

$$\sum_{I\in\mathcal{A}} c_I det\big((\mathbf{X}_i(\theta,\theta^\star))_{i\in I}\big) > 0. \tag{26}$$

With calculations as in (24) we obtain

$$\begin{aligned} & p! \sum_{I\in\mathcal{A}} c_I det\big((\mathbf{X}_i(\theta,\theta^\star))_{i\in I}\big) \\ = & \int e^{\sum_{k=1}^p -\theta^{\star T}\mathbf{v}_k + \zeta\left((\theta^\star-\theta)^T\mathbf{v}_k\right)} \left(\sum_{I\in\mathcal{A}} C_I det(\mathbf{v}_1,\ldots,\mathbf{v}_p) det\big(\Psi_\theta^I(\mathbf{v}_1),\ldots,\Psi_\theta^I(\mathbf{v}_p)\big)\right) P_V(d\mathbf{v}_1)\cdots P_V(d\mathbf{v}_p). \end{aligned}$$

Thanks to **[Det']** , this quantity is positive. ■

In the case where $p = 2$ and $K = 3$, **[Det'(≥)]** is satisfied if and only if there exist a, b, c in $\mathbb{R}^3$ such that for every $\mathbf{v}_1$ and $\mathbf{v}_2$ with $det(\mathbf{v}_1,\mathbf{v}_2) > 0$

$$a\ det\left(\Psi_\theta^{\{1,2\}}(\mathbf{v}_1),\Psi_\theta^{\{1,2\}}(\mathbf{v}_2)\right) + b\ det\left(\Psi_\theta^{\{1,3\}}(\mathbf{v}_1),\Psi_\theta^{\{1,3\}}(\mathbf{v}_2)\right) + c\ det\left(\Psi_\theta^{\{2,3\}}(\mathbf{v}_1),\Psi_\theta^{\{2,3\}}(\mathbf{v}_2)\right) \geq 0. \tag{27}$$

If we denote by $\wedge$ the vectorial product in $\mathbb{R}^3$, the inequality in (27) means that the following set

$$\Big\{\Psi_\theta(\mathbf{v}_1)\wedge\Psi_\theta(\mathbf{v}_2),\ \text{for all } \mathbf{v}_1,\mathbf{v}_2 \text{ such that } det(\mathbf{v}_1,\mathbf{v}_2) > 0\Big\} \tag{28}$$

is included in the half space in $\mathbb{R}^3$ with equation $cx - by + az \geq 0$. In the setting where $V_1 = 1$ and V_2 is included in an interval $[a, b]$, as for the area process, this condition becomes a geometrical characteristic of the curve $\gamma_\theta = \Psi_\theta(\{1\}\times[a,b])$ in $\mathbb{R}^3$, which is easy to check visually.

6 Extension in the presence of non-hereditary interaction

In several recent papers, Gibbs processes with non hereditary interactions are considered, in particular in the domain of stochastic geometry (see [12, 14]). The parametric estimation of such models has also been investigated. The first results in this direction have been given in [15] via a pseudo-likelihood procedure based on a generalization of the Georgii-Nguyen-Zessin formula (4). The same kind of generalization is possible for the Takacs-Fiksel procedure. We address this improvement in this section.

In the following, we do not assume that the energy $V_\Lambda(\varphi,\theta)$ satisfies the heredity assumption (2). The first consequences are that the local energy $V(x^m|\varphi;\theta)$ is not defined in general and that the Georgii-Nguyen-Zessin formula is not available. Let us begin by presenting the generalization of this formula, as stated in [15] Proposition 2, which is valid in the hereditary and non-hereditary settings.

We first need to recall the concept of removable points which has been introduced in [15] Definition 3.

Definition 2 *A point x^m in a configuration φ is called removable if there exists a bounded set Λ containing x such that $V_\Lambda(\varphi\backslash x^m,\theta) < +\infty$. We denote by $\mathcal{R}_\theta(\varphi)$ the set of removable points in φ.*

Let us remark that the removable set is only related to the support of the underlying Gibbs measure. The local energy $V(x^m|\varphi\backslash x^m;\theta)$ of any removable point $x^m \in \mathcal{R}_\theta(\varphi)$ can then be defined by the classical expression (3) where Λ comes from Definition 2. In the hereditary case, all the points of φ are removable and we regain the classical definition of the local energy.

The generalization of the Georgii-Nguyen-Zessin formula is the following equation

$$\mathbf{E}\left(\int_{\mathbb{R}^d\times\mathbb{M}} h\left(x^m,\Phi;\theta\right)e^{-V(x^m|\Phi;\theta^\star)}\mu(dx^m)\right)=\mathbf{E}\left(\sum_{x^m\in\mathcal{R}_{\theta^\star}(\Phi)} h\left(x^m,\Phi\setminus x^m;\theta\right)\right). \tag{29}$$

Let us notice that the only difference with the classical formula is that the sum is restricted to the removable points. Now, let us present the consequences of this formula on the Takacs-Fiksel procedure. We have to consider the two following cases:

- When the support of the Gibbs measure does not depend on θ: the set of removable points $\mathcal{R}_\theta(\varphi)$ does not depend on θ either. In this case, the Takacs-Fiksel estimor is defined by (7), and C_Λ is as in (6) to the exception of the sum which is restricted to the removable points:

$$C_\Lambda(\varphi;h,\theta):=\int_{\Lambda\times\mathbb{M}} h(x^m,\varphi;\theta)e^{-V(x^m|\varphi;\theta)}\mu(dx^m)-\sum_{x^m\in\mathcal{R}_{\theta^\star}(\varphi)\cap\Lambda} h(x^m,\varphi\setminus x^m;\theta). \tag{30}$$

  The sum is computable because by assumption, the set $\mathcal{R}_{\theta^\star}(\varphi)$ does not depend on $\theta^\star$. In this situation, with the same assumptions **[C]** and **[N]**, the consistency and the asymptotic normality of the estimator may be proved as in Section 7.

- When the support of the Gibbs measure depends on some parameters $\theta_{hc}=(\theta_1,\ldots,\theta_q)$, $q\leq p$ (called the hardcore parameters): the remaining parameters $\theta_{sm}=(\theta_{q+1},\ldots,\theta_p)$ are supposed to parametrize the classical (or smooth) interaction between points. The set of removable points $\mathcal{R}_\theta(\varphi)$ therefore depends on θ_{hc} only. The estimation issue is more complicated in this case. Indeed, Assumption **[C2]** requires some regularities of the interaction with respect to the parameter θ, such as continuity, which clearly fail to be true for the support parameter θ_{hc}. The Takacs-Fiksel procedure is therefore unable to estimate θ_{hc}. Note that this problem is not specific to the presence of non-hereditary interactions, but arises as soon as some hardcore parameters have to be estimated. In [15], the authors solve this problem in both the hereditary and non-hereditary setting, by introducing a two-step estimation procedure. We can follow the same strategy here. In a first step, the estimator $\hat{\theta}_{hc}$ of the hardcore parameter is defined in a natural way according to the observed support of the point process (see Section 4.2.1 in [15]). Then, in a second step, the Takacs-Fiksel estimator $\hat{\theta}_{sm}$ is defined by (7) with

$$\begin{aligned} C_\Lambda(\varphi;h,\theta_{sm}) \;:=\; & \int_{\Lambda\times\mathbb{M}} h(x^m,\varphi;\theta_{sm},\hat{\theta}_{hc})e^{-V(x^m|\varphi;\theta_{sm},\hat{\theta}_{hc})}\mu(dx^m) \\ & -\sum_{x^m\in\mathcal{R}_{\hat{\theta}_{hc}}(\varphi)\cap\Lambda} h(x^m,\varphi\setminus x^m;\theta_{sm},\hat{\theta}_{hc}). \end{aligned} \tag{31}$$

  Let us remark that the estimator $\hat{\theta}_{hc}$ is plugged in the computation of C_Λ. In particular, the removable points are determined with respect to $\hat{\theta}_{hc}$. As in [15], the regularity and integrability assumptions of type **[C]** for $\hat{\theta}_{sm}$ and conditions on the support of the Gibbs measure are required in order to obtain the consistency of $(\hat{\theta}_{hc},\hat{\theta}_{sm})$. The asymptotic normality is more difficult to obtain and no general results are available. In fact, it seems that there is no hope to expect asymptotic normality without managing the rate of convergence of $\hat{\theta}_{hc}$, which should strongly depend on the model.

7 Proofs of asymptotic results

7.1 Proof of Proposition 2

Let $P_{\theta^\star}$ be an ergodic Gibbs measure (if $P_{\theta^\star}$ is not ergodic, we finish the proof by a classical argument of mixing). The proof is split into two steps.

Step 1. U_{Λ_n} is a contrast function. Under **[C1]**, the ergodic theorem obtained in [34] and the GNZ formula given in (5) may be applied to prove that $P_{\theta^\star}$-a.s.

$$\begin{aligned}|\Lambda_n|^{-1}C_{\Lambda_n}(\Phi,h_k;\theta) &\to \mathbf{E}\left(h_k(0^M,\Phi;\theta)e^{-V(0^M|\Phi;\theta)}\right)-\mathbf{E}\left(h_k(0^M,\Phi\setminus 0^M)\right)\\ &= \mathbf{E}\left(h_k(0^M,\Phi;\theta)\left(e^{-V(0^M|\Phi;\theta)}-e^{-V(0^M|\Phi;\theta^\star)}\right)\right).\end{aligned}$$

Therefore, as $n\to+\infty$, one obtains $P_{\theta^\star}-$a.s.

$$U_{\Lambda_n}(\Phi;\mathbf{h},\theta)\to U(\theta):=\sum_{k=1}^K\mathbf{E}\left(h_k(0^M,\Phi;\theta)\left(e^{-V(0^M|\Phi;\theta)}-e^{-V(0^M|\Phi;\theta^\star)}\right)\right)^2.$$

Note that $U(\theta^\star)=0$. In addition with Assumptions **[C2]** and **[C3]**, this proves that U_{Λ_n} is a continuous contrast function vanishing only at $\theta^\star$.

Step 2. Modulus of continuity

The modulus of continuity of the contrast process is defined for all $\varphi\in\Omega$ and all $\eta>0$ by

$$W_n(\varphi,\eta)=\sup\left\{\left|U_{\Lambda_n}(\varphi;\mathbf{h},\theta)-U_{\Lambda_n}(\varphi;\mathbf{h},\theta')\right|:\theta,\theta'\in\mathbf{\Theta},||\theta-\theta'||\le\eta\right\}.$$

This step aims at proving that there exists a sequence $(\varepsilon_\ell)_{\ell\ge1}$, with $\varepsilon_\ell\to0$ as $\ell\to+\infty$ such that for all $\ell\ge1$

$$P\left(\limsup_{n\to+\infty}\left(W_n\left(\Phi,\frac{1}{\ell}\right)\ge\varepsilon_\ell\right)\right)=0. \tag{32}$$

Let $\theta,\theta'\in\mathbf{\Theta}$, then

$$\begin{aligned}|U_{\Lambda_n}(\varphi;\mathbf{h},\theta)-U_{\Lambda_n}(\varphi;\mathbf{h},\theta')| \le \sum_{k=1}^K\Big\{&|\Lambda_n|^{-1}|C_{\Lambda_n}(\varphi;h_k,\theta)-C_{\Lambda_n}(\varphi;h_k,\theta')|\times\\ &|\Lambda_n|^{-1}\big(|C_{\Lambda_n}(\varphi;h_k,\theta)|+|C_{\Lambda_n}(\varphi;h_k,\theta')|\big)\Big\}.\end{aligned} \tag{33}$$

Under Assumption **[C4]**, there exists $n_0\ge1$ such that for all $n\ge n_0$ we have for $P_{\theta^\star}-$a.e. φ

$$\begin{aligned}|\Lambda_n|^{-1}|C_{\Lambda_n}(\varphi;h_k,\theta)| &\le |\Lambda_n|^{-1}\int_{\Lambda_n\times\mathbb{M}}\max_{\theta\in\mathbf{\Theta}}|h_k(x^m,\varphi;\theta)|e^{-V(x^m|\varphi;\theta)}\mu(dx^m)\\ &\quad+|\Lambda_n|^{-1}\sum_{x^m\in\varphi_{\Lambda_n}}\max_{\theta\in\mathbf{\Theta}}|h_k(x^m,\varphi\setminus x^m;\theta)|\\ &\le \gamma_1,\end{aligned}$$

where

$$\gamma_1:=2\times\max_{k=1,\dots,K}\left(\mathbf{E}\left(\max_{\theta\in\mathbf{\Theta}}|f_k(0^M,\Phi;\theta)|\right)+\mathbf{E}\left(\max_{\theta\in\mathbf{\Theta}}|h_k(0^M,\Phi;\theta)|e^{-V(0^M|\Phi;\theta^\star)}\right)\right).$$

Therefore for all $n\ge n_0$

$$|U_{\Lambda_n}(\varphi;\mathbf{h},\theta)-U_{\Lambda_n}(\varphi;\mathbf{h},\theta')|\le\gamma_1\sum_{k=1}^K\left(A_{\Lambda_n}(\varphi;h_k,\theta,\theta')+B_{\Lambda_n}(\varphi;h_k,\theta,\theta')\right),$$

where

$$\begin{aligned}A_{\Lambda_n}(\varphi;h_k,\theta,\theta') &:= \int_{\Lambda_n\times\mathbb{M}}|f_k(x^m,\varphi;\theta)-f_k(x^m,\varphi;\theta')|\mu(dx^m)\\ B_{\Lambda_n}(\varphi;h_k,\theta,\theta') &= \sum_{x^m\in\varphi_{\Lambda_n}}|h_k(x^m,\varphi\setminus x^m;\theta)-h_k(x^m,\varphi\setminus x^m;\theta')|.\end{aligned}$$

Under Assumption **[C4]**, one may apply the mean value theorem in $\mathbb{R}^p$ as follows: there exist $\xi^{(1)},\ldots,\xi^{(p)} \in [\min(\theta_1,\theta_1'),\max(\theta_1,\theta_1')] \times \ldots \times \left[\min(\theta_p,\theta_p'),\max(\theta_p,\theta_p')\right]$ such that for all $\varphi \in \Omega$

$$\begin{aligned} A_{\Lambda_n}(\varphi;h_k,\theta,\theta') &= \int_{\Lambda_n\times\mathbb{M}} \sum_{j=1}^{p} (\theta_j-\theta_j')(\mathbf{f_k}^{(1)}(x^m,\varphi;\xi_j^{(1)}))_j \; \mu(dx^m) \\ &\leq p \times \|\theta-\theta'\| \int_{\Lambda_n\times\mathbb{M}} \max_{\theta\in\mathbf{\Theta}} \|\mathbf{f_k}^{(1)}(x^m,\varphi;\theta)\| \; \mu(dx^m) \end{aligned}$$

In a similar way, one may prove that for $P_{\theta^\star}-$a.e. φ

$$B_{\Lambda_n}(\varphi;h_k,\theta,\theta') \leq p \times \|\theta-\theta'\| \sum_{x^m\in\varphi_{\Lambda_n}} \max_{\theta\in\mathbf{\Theta}} \|\mathbf{h_k}^{(1)}(x^m,\varphi\setminus x^m;\theta)\|.$$

Under Assumption **[C4]**, there exists $n_1(k)\geq 1$ such that for all $n\geq n_1(k)$, we have for $P_\theta^\star-$a.e. φ

$$A_{\Lambda_n}(\varphi;h_k,\theta,\theta') + B_{\Lambda_n}(\varphi;h_k,\theta,\theta') \leq \gamma_2\|\theta-\theta'\|$$

where

$$\gamma_2 := 2p \times \max_{k=1,\ldots,K} \left(\mathbf{E}\left(\max_{\theta\in\mathbf{\Theta}} \|\mathbf{f_k}^{(1)}(0^M,\Phi;\theta)\| \right) + \mathbf{E}\left(\max_{\theta\in\mathbf{\Theta}} \|\mathbf{h_k}^{(1)}(0^M,\Phi;\theta)\| e^{-V(0^M|\Phi;\theta^\star)} \right) \right).$$

We finally obtain the following upper-bound for $P_{\theta^\star}-$a.e. φ, for all θ,θ' such that $\|\theta-\theta'\|\leq 1/\ell$ and for all $n\geq N=\max(n_0,\max_k n_1(k))$

$$|U_{\Lambda_n}(\varphi;\mathbf{h},\theta) - U_{\Lambda_n}(\varphi;\mathbf{h},\theta')| \leq \gamma \times \frac{1}{\ell},$$

with $\gamma = K\times\gamma_1\times\gamma_2$ and therefore $W_n(\varphi,1/\ell)\leq \gamma\times\frac{1}{\ell}$. Finally, since

$$\limsup_{n\to+\infty} \left\{ W_n\left(\varphi,\frac{1}{\ell}\right) \geq \frac{\gamma}{\ell} \right\} = \bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} \left\{ W_n\left(\varphi,\frac{1}{\ell}\right) \geq \frac{\gamma}{\ell} \right\} \subset \bigcup_{n\geq N} \left\{ W_n\left(\varphi,\frac{1}{\ell}\right) \geq \frac{\gamma}{\ell} \right\}$$

for $P_{\theta^\star}-$a.e. φ, the expected result (32) is proved.
Conclusion step. Steps 1 and 2 ensure the fact that we can apply Theorem 3.4.3 of [22] which asserts the almost sure convergence for minimum contrast estimators.

7.2 Proof of Proposition 3

The proof is based on a classical result concerning asymptotic normality for minimum contrast estimators *e.g.* Theorem 3.4.5 of [22]. We split it in two different steps.

Step 1. Asymptotic normality of $\mathbf{U}_{\Lambda_n}^{(1)}(\Phi;\mathbf{h},\theta^\star)$.
We start with the following Lemma.

Lemma 9 *Under the Assumptions **[N1]**, **[N2]** and **[N3]**, the following convergence holds in distribution, as $n\to+\infty$*

$$|\Lambda_n|^{-1/2}\; \widetilde{\mathbf{C}}_{\Lambda_n}(\Phi;\mathbf{h},\theta^\star) \xrightarrow{d} \mathcal{N}(0,\underline{\mathbf{\Sigma}}(\mathbf{h},\theta^\star)), \tag{34}$$

where $\underline{\mathbf{\Sigma}}(\mathbf{h},\theta^\star)$ and $\widetilde{\mathbf{C}}_{\Lambda_n}(\varphi;\mathbf{h},\theta^\star)$ are defined in Proposition 3.

Proof. The vector $\widetilde{\mathbf{C}}_{\Lambda_n}(\varphi;\mathbf{h},\theta^\star)$ (of length K) corresponds to the vector of the h_k-residuals for $k=1,\ldots,K$ computed on the same domain Λ_n with $\widehat{\theta}=\theta^\star$, see [2] for a definition and practical study of this concept of residuals. The asymptotic behavior of the residuals process has been investigated in [11]. In particular, with the notation of the present paper, the asymptotic normality of the vector $\widetilde{\mathbf{C}}_{\Lambda_n}(\Phi;\mathbf{h},\widehat{\theta})$ for general $\widehat{\theta}$ is obtained (see Proposition 4 in [11]). When $\widehat{\theta}=\theta^\star$, the assumptions and the asymptotic covariance matrix of Proposition 4 in [11] respectively reduce to **[N1-3]** and (16). ■

Now, according to the definition of $U_{\Lambda_n}(\varphi;\mathbf{h},\theta^\star)$, we have

$$\mathbf{U}^{(1)}_{\Lambda_n}(\varphi;\mathbf{h},\theta^\star) = 2|\Lambda_n|^{-2}\sum_{k=1}^{K}\mathbf{C}^{(1)}_{\Lambda_n}(\varphi;h_k,\theta^\star)C_{\Lambda_n}(\varphi;h_k,\theta^\star).$$

Under Assumption **[C4]**, one may apply an ergodic theorem (see [34]), in order to derive $P^\star_\theta-$a.s., as $n\to+\infty$

$$|\Lambda_n|^{-1}\mathbf{C}^{(1)}_{\Lambda_n}(\Phi;h_k,\theta^\star)\to\mathbf{E}\left(\mathbf{f_k}^{(1)}(0^M,\Phi;\theta^\star)+\mathbf{h_k}^{(1)}(0^M,\Phi;\theta^\star)e^{-V(0^M|\Phi;\theta^\star)}\right). \tag{35}$$

It is easily checked that this expectation reduces to the vector of length p defined by $\mathcal{E}(h_k,\theta^\star):=\mathbf{E}\left(h_k(0^M,\Phi;\theta^\star)\mathbf{V}^{(1)}(0^M|\Phi;\theta^\star)e^{-V(0^M|\Phi;\theta^\star)}\right)$. Let us denote by $\underline{\mathcal{E}}(\mathbf{h},\theta^\star)$ the (p,K) matrix $(\mathcal{E}(h_1,\theta^\star),\ldots,\mathcal{E}(h_K,\theta^\star))$, then we get the following decomposition

$$\begin{aligned}
|\Lambda_n|^{1/2}\mathbf{U}^{(1)}_{\Lambda_n}(\Phi;\mathbf{h},\theta^\star) &= 2|\Lambda_n|^{1/2}\times|\Lambda_n|^{-2}\sum_{k=1}^{K}\mathbf{C}^{(1)}_{\Lambda_n}(\Phi;h_k,\theta^\star)C_{\Lambda_n}(\Phi;h_k,\theta^\star)\\
&= 2|\Lambda_n|^{-1/2}\underline{\mathcal{E}}(\mathbf{h},\theta^\star)\widetilde{\mathbf{C}}_{\Lambda_n}(\Phi;\mathbf{h},\theta^\star)\\
&\quad +2\sum_{k=1}^{K}\left(|\Lambda_n|^{-1}\mathbf{C}^{(1)}_{\Lambda_n}(\Phi;h_k,\theta^\star)-\mathcal{E}(h_k,\theta^\star)\right)|\Lambda_n|^{-1/2}C_{\Lambda_n}(\Phi;h_k,\theta^\star).
\end{aligned}$$

According to (35) and Lemma 9, Slutsky's Theorem implies that for any $k\in\{1,\ldots,K\}$,

$$\left(|\Lambda_n|^{-1}\mathbf{C}^{(1)}_{\Lambda_n}(\Phi;h_k,\theta^\star)-\mathcal{E}(h_k,\theta^\star)\right)|\Lambda_n|^{-1/2}C_{\Lambda_n}(\Phi;h_k,\theta^\star)\xrightarrow{P}0,$$

as $n\to+\infty$, the zero here being a vector of length p. Using again Lemma 9, we finally reach the following convergence in distribution as $n\to+\infty$

$$|\Lambda_n|^{1/2}\mathbf{U}^{(1)}_{\Lambda_n}(\Phi;\mathbf{h},\theta^\star)\xrightarrow{d}\mathcal{N}\left(0,4\underline{\mathcal{E}}(\mathbf{h},\theta^\star)\,\underline{\mathbf{\Sigma}}(\mathbf{h},\theta^\star)\,\underline{\mathcal{E}}(\mathbf{h},\theta^\star)^T\right),$$

where $\underline{\mathbf{\Sigma}}(\mathbf{h},\theta^\star)$ is defined by (16).

Step 2. Convergence of $\underline{\mathbf{U}}^{(2)}_{\Lambda_n}(\Phi;\mathbf{h},\theta)$ for $\theta\in\mathcal{V}(\theta^\star)$

According to our definition and Assumption **[N4]**, the (p,p) matrix $\underline{\mathbf{U}}^{(2)}_{\Lambda_n}(\varphi;\mathbf{h},\theta)$ is defined for $i,j=1,\ldots,K$ by

$$\left(\underline{\mathbf{U}}^{(2)}_{\Lambda_n}(\varphi;\mathbf{h},\theta^\star)\right)_{ij}=2|\Lambda_n|^{-2}\sum_{k=1}^{K}\left\{\left(\underline{\mathbf{C}}^{(2)}_{\Lambda_n}(\varphi;h_k;\theta)\right)_{ij}C_{\Lambda_n}(\varphi;h_k,\theta)+\left(\mathbf{C}^{(1)}_{\Lambda_n}(\varphi;h_k,\theta)\right)_i\left(\mathbf{C}^{(1)}_{\Lambda_n}(\varphi;h_k,\theta)\right)_j\right\}.$$

Note also that $\mathbf{C}^{(1)}_{\Lambda_n}(\varphi;h_k,\theta)$ and $\underline{\mathbf{C}}^{(2)}_{\Lambda_n}(\varphi;h_k,\theta)$ are defined by

$$\begin{aligned}
\mathbf{C}^{(1)}_{\Lambda_n}(\varphi;h_k,\theta) &= \int_{\Lambda_n\times\mathbb{M}}\mathbf{f}^{(1)}_k(x^m,\varphi;\theta)\mu(dx^m)-\sum_{x^m\in\varphi_{\Lambda_n}}\mathbf{h}^{(1)}_k(x^m,\varphi\setminus x^m;\theta)\\
\underline{\mathbf{C}}^{(2)}_{\Lambda_n}(\varphi;h_k,\theta) &= \int_{\Lambda_n\times\mathbb{M}}\underline{\mathbf{f}}^{(2)}_k(x^m,\varphi;\theta)\mu(dx^m)-\sum_{x^m\in\varphi_{\Lambda_n}}\underline{\mathbf{h}}^{(2)}_k(x^m,\varphi\setminus x^m;\theta).
\end{aligned}$$

Under Assumption **[N4]**, then for all $i,j=1,\ldots,p$ and for any $k=1,\ldots,K$, each normalized term $|\Lambda_n|^{-1}C_{\Lambda_n}(\Phi;h_k,\theta)$, $|\Lambda_n|^{-1}\mathbf{C}^{(1)}_{\Lambda_n}(\Phi;h_k,\theta)$ and $|\Lambda_n|^{-1}\underline{\mathbf{C}}^{(2)}_{\Lambda_n}(\Phi;h_k,\theta)$ satisfies an ergodic theorem. Therefore, for any $\theta\in\mathcal{V}(\theta^\star)$, there exists a matrix $\underline{\mathbf{U}}^{(2)}(\mathbf{h},\theta)$ such that $P^\star_\theta-$a.s.

$$\underline{\mathbf{U}}^{(2)}_{\Lambda_n}(\Phi;\mathbf{h},\theta)\to\underline{\mathbf{U}}^{(2)}(\mathbf{h},\theta).$$

This justifies that, for n large enough, in a neighborhood of $\theta^\star$, $\left(\underline{\mathbf{U}}^{(2)}_{\Lambda_n}(\varphi;\mathbf{h},\theta)\right)_{ij}$ is uniformly bounded by $2\times\max_{\theta\in\mathcal{V}(\theta^\star)}\left|\left(\underline{\mathbf{U}}^{(2)}(\mathbf{h},\theta)\right)_{ij}\right|$ for $P^\star_\theta-$a.e. φ. When $\theta=\theta^\star$, recall, from (5), that $|\Lambda_n|^{-1}C_{\Lambda_n}(\Phi;h_k,\theta^\star)$

converges almost surely to zero and that (35) holds. Hence, $\underline{\mathbf{U}}^{(2)}(\theta^\star)$ reduces to $2\underline{\mathcal{E}}(\mathbf{h},\theta^\star)\underline{\mathcal{E}}(\mathbf{h},\theta^\star)^T$.

Conclusion Step. From Theorem 3.4.5 of [22], Steps 1 and 2 ensure that the normalized difference $|\Lambda_n|^{1/2}\left(\underline{\mathbf{U}}^{(2)}(\mathbf{h},\theta^\star)\left(\widehat{\theta}_n(\Phi)-\theta^\star\right)-\mathbf{U}^{(1)}_{\Lambda_n}(\Phi;\mathbf{h},\theta^\star)\right)$ converges in probability to 0, which is the expected result.

References

[1] A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42:283–322, 2000.

[2] A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes. *Journal of the Royal Statistical Society (series B)*, 67:1–35, 2005.

[3] A. J. Baddeley and M. N. M. Van Lieshout. Area-interaction point processes. *Ann. Inst. Statist. Math.*, 47(4):601–619, 1995.

[4] K.K. Berthelsen and J. Møller. Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, 30(3):549–564, 2003.

[5] E. Bertin, J.-M. Billiot, and R. Drouilhet. Existence of "Nearest-Neighbour" Gibbs Point Models. *Ann. Appl. Probab.*, 31:895–909, 1999.

[6] E. Bertin, J.-M. Billiot, and R. Drouilhet. R-local Delaunay inhibition model. *Journal of Statistical Physics*, 132(4):649–667, 2008.

[7] J. Besag. Spatial interaction and the statistical analysis of lattice system. *Journal of the Royal Statistical Society (series B)*, 26:192–236, 1974.

[8] J.-M. Billiot. Asymptotic properties of takacs-fiksel estimation method for gibbs point processes. *Statistics*, 30(1):69–89, 1997.

[9] J.-M. Billiot, J.-F. Coeurjolly, and R. Drouilhet. Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes. *Electronic Journal of Statistics*, 2:234–264, 2008.

[10] J.-F. Coeurjolly and R. Drouilhet. Asymptotic properties of the maximum pseudo-likelihood estimator for stationary gibbs point processes including the lennard-jones model. *submitted*, 2009. `http://hal.archives-ouvertes.fr/hal-00442750/`.

[11] J.-F. Coeurjolly and F. Lavancier. Residuals for stationary marked Gibbs point processes. *submitted*, 2010. `http://hal.archives-ouvertes.fr/hal-00453102/`.

[12] D. Dereudre. Gibbs Delaunay tessellations with geometric hardcore condition. *J. Stat. Phys.*, 121(3-4):511–515, 2005.

[13] D. Dereudre. The existence of quermass-interaction processes for nonlocally stable interaction and nonbounded convex grains. *Adv. Appli. Prob.*, 41(3):664–681, 2009.

[14] D. Dereudre, R. Drouilhet, and H.O. Georgii. Existence of Gibbsian point processes with geometry-dependent interactions. *submitted*, 2010.

[15] D. Dereudre and F. Lavancier. Campbell equilibrium equation and pseudo-likelihood estimation for non-hereditary Gibbs point processes. *Bernoulli*, 15(4):1368–1396, 2009.

[16] D. Dereudre and F. Lavancier. Practical simulation and estimation for Gibbs Delaunay-voronoï tessellations with geometric hardcore interaction. *to appear in Computational Stat. and Data Analysis*, 2010.

[17] P. J. Diggle, T. Fiksel, P. Grabarnik, Y. Ogata, D. Stoyan, and M. Tanemura. On parameter estimation for pairwise interaction point processes. *International Statistical Review*, 62(1):99–117, 1994.

[18] T. Fiksel. Estimation of parameterized pair potentials of marked and non-marked gibbsian point processes. *Elektron. Inform. Kybernet.*, 20:270–278, 1984.

[19] T. Fiksel. Estimation of interaction potentials of gibbsian point processes. *Math. Operationsf. Statist. Ser. Statist.*, 19:77–86, 1988.

[20] H.O. Georgii. Canonical and grand canonical Gibbs states for continuum systems. *Communications in Mathematical Physics*, 48(1):31–51, 1976.

[21] H.O. Georgii. *Gibbs measures and Phase Transitions*. de Gruyter, Berlin, 1988.

[22] X. Guyon. *Random fields on a network. Modeling, Statistics and Applications*. Springer Verlag, New York, 1995.

[23] L. Heinrich. Mixing properties of gibbsian point processes and asymptotic normality of takacs-fiksel estimates. *preprint*, 1992.

[24] F. Huang and Y. Ogata. Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics*, 8(3):510–530, 1999.

[25] J. Illian, A. Penttinen H., and Stoyan. *Statistical analysis and modelling of spatial point patterns*. Wiley-Interscience, 2008.

[26] J.L. Jensen and J. Møller. Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.*, 1:445–461, 1991.

[27] W. S. Kendall, M. N. M. Van Lieshout, and A. J. Baddeley. Quermass-interaction processes conditions for stability. *Adv. Appli. Prob.*, 31:315–342, 1999.

[28] E.L. Lehman. *Elements of large-sample theorey*. Springer, 1999.

[29] S. Mase. Uniform lan condition of planar gibbsian point processes and optimality of maximum likelihood estimators of soft-core potential function. *Probab. Theory and Related Fields*, 92:51–67, 1992.

[30] K Matthes, J. Kerstan, and J. Mecke. *Infinitely Divisible Point Process*. J. Wiley, 1978.

[31] J. Møller and K. Helisová. Power diagrams and interaction processes for union of discs. *Adv. Appli. Prob.*, 40(2):321–347, 2008.

[32] J. Møller and R. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, 2003.

[33] J. Møller and R. Waagepetersen. Modern statistics for spatial point processes. *Scan. J. Stat.*, 34:643–684, 2007.

[34] X.X. Nguyen and H. Zessin. Ergodic theorems for Spatial Process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 48:133–158, 1979.

[35] C.J. Preston. *Random fields*. Springer Verlag, 1976.

[36] D. Ruelle. *Statistical Mechanics*. Benjamin, New York-Amsterdam, 1969.

[37] D. Ruelle. Superstable interactions in classical statistical mechanics. *Comm. Math. Phys.*, 1970.

[38] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic geometry and its applications*. John Wiley and Sons, Chichester, 1995.

[39] R. Takacs. Estimator for the pair-potential of a gibbsian point process. *Math. Operationsf. Statist. Ser. Statist.*, 17:429–433, 1986.

Troisième partie

Autour de la statistique appliquée et de l'enseignement de la statistique

Chapitre 1

Thèmes et problèmes considérés

Sommaire

Ce chapitre résume mes contributions dans différents domaines dont le thème commun nous semble être la statistique appliquée. La Section 1.1 résume de nombreuses réflexions menées avec Rémy Drouilhet sur l'enseignement de la statistique. Les Sections 1.2 et 1.3 relatent des activités autour du co-encadrement de deux étudiants en thèse (J.F. Robineau et M. Nguile Makao). La Section 1.4 traite de mon activité de recherche dans une université de sciences sociales avec des collaborations avec des économistes et psychologues.

1.1 Autour de l'enseignement de la statistique

En collaboration avec Rémy Drouilhet, nous avons depuis 2001 (bien que ce travail ait été mené seul par R. Drouilhet depuis 1999) développé une approche expérimentale pour tenter de visualiser des concepts statistiques (telle que la notion de test d'hypothèses ou celle d'intervalle de confiance) et probabilistes (notion d'estimateur, notion de densité de probabilité, notion de modèle,...) ceci afin (principalement) d'éviter de trop lourdes techniques mathématiques souvent mises en avant dans notre domaine et malheureusement pas toujours adaptées au public économiste qui est le nôtre. Dans cette optique nous avons donc développé d'une part des logiciels visuels à caractère pédagogique et d'autre part un système de notations (ainsi que des supports de cours) adaptés à notre discours. A notre avis, ce travail constitue une contribution à la didactique de la statistique. Il a d'ailleurs fait l'objet d'une présentation à un congrés international à Alger [CD04] et de deux papiers aux 37-èmes journées de statistique (Pau, juin 2005) [CD05b, CD05a] : le premier est associé à la notion d'approche expérimentale des probabilités pour appréhender des notions telles que la densité de probabilité (voir Figure 1.1) et l'intervalle de confiance et le second l'utilisation de ces principes pour construire étape par étape un test d'hypothèses, comprendre et visualiser les risques d'erreurs de décision qui lui sont associées et appréhender la notion de p-valeur ou risque critique.

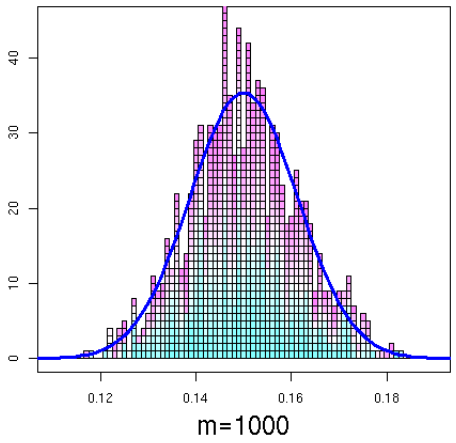


FIGURE 1.1 – Histogramme sous la forme d'un mur de briques de $m = 1000$ réalisations de la moyenne de n variables discrètes et son approximation gaussienne. Chaque cube de surface $1/m$ est associée à une réalisation dont la valeur peut être lue par l'abscisse du centre du cube. Plus m grandit et plus les cubes réduisent de surface et plus n augmente plus les cubes deviennent des points. Cette visualisation nous permet de plus facilement interpréter une densité de probabilité comme un "tas" de réalisations et une réalisation d'une variable aléatoire "comme" l'abscisse d'un point choisi au hasard sous la courbe.

Les réflexions menées avec Rémy Drouilhet et les étudiants nous ont conduits à penser que dans le cadre des tests d'hypothèses paramétriques le cadre gaussien est bien souvent trop mis en avant. Certes, il est agréable pour des étudiants ayant des connaissances en calculs de probabilités car ils peuvent établir la loi de la statistique de Student pour un test de moyenne par exemple mais pour un étudiant souhaitant simplement répondre à une question qu'il se pose quant à la valeur d'un paramètre, peu lui importe. Et d'ailleurs dans de nombreuses applications, le caractère gaussien est loin d'être vérifié.

Lorsque l'on considère la statistique de Student pour tester une moyenne (ou une différence de moyennes), si l'on possède un nombre conséquent de données peu importe de savoir si les variables sont gaussiennes ou pas car cette statistique est robuste dans le sens où $\mathcal{S}t(n-1) \simeq \mathcal{N}(0,1)$ lorsque n est grand. Si l'on souhaite tester une variance simple (ou un rapport de variances), la statistique du χ^2 $((n-1)\widehat{\sigma}^2/\sigma_0^2)$ peut s'éloigner de manière très importante d'une loi $\chi^2(n-1)$ lorsque les données ne sont plus gaussiennes. Et il est à noter qu'un logiciel tel que `R` ne propose aucune procédure pour tester une variance simple dans un cadre non gaussien. En collaboration avec R. Drouilhet, Pierre Lafaye de Micheaux et Jean-François Robineau, nous avons réalisé une modeste contribution à ces réflexions en développant le package `R asymptest` qui permet de tester et d'estimer des intervalles de confiance pour un paramètre θ tel que la moyenne, la variance, la différence ou le rapport de moyennes ou variances en utilisant le fait que (pour ces paramètres) $(\widehat{\theta}-\theta)/\widehat{\sigma}_{\widehat{\theta}} \xrightarrow{d} \mathcal{N}(0,1)$ où $\widehat{\sigma}^2_{\widehat{\theta}}$ est un estimateur de la variance de l'estimateur. Une publication sur la méthodologie générale, la description du package et son utilisation a été réalisée en commun [CDdMR09] (à noter qu'une version plus longue, contenant de nombreuses simulations et les preuves simples, est disponible sur arxiv `http://arxiv.org/abs/0902.0506`).

1.2 Problèmes de sélection de variables (parmi un grand nombre) dans un cadre de discrimination

Ce paragraphe vise à présenter brièvement le travail réalisé en collaboration avec Rémy Drouilhet et Jean-François Robineau [CDR07], à la suite de la thèse de ce dernier (thèse co-encadrée par J.-F. Coeurjolly, R. Drouilhet et C. Garbay, et soutenue en Décembre 2004). Cette thèse est intitulée "Méthodes de sélection de variables (parmi un grand nombre) dans un cadre de discrimination".

Le but consistait à sélectionner un nombre relativement restreint de variables parmi $X_1, \ldots, X_p$ ($p >> n$, qui peuvent être aussi bien continues, discrètes ou catégorielles) dans l'objectif de discriminer une variable Y (catégorielle). La stratégie adoptée a été réalisée en deux étapes :

1. Etape de quantification supervisée de manière à rendre la variable (ou un ensemble de variables) moins complexe et à la rendre la plus ressemblante à Y.
2. Utilisation des versions quantifiées pour construire des méthodes de type ascendante de sélection de variables.

L'originalité de la thèse réside dans la volonté d'inscrire ces deux étapes dans la théorie de l'information. En effet, pour réaliser ces deux étapes, nous avons utilisé des mesures entropiques (adaptées aux vecteurs discrets ou catégoriels) permettant de mesurer une balance entre la quantité d'information contenue dans un vecteur $\mathbf{X}$ pour expliquer Y et la complexité du couple $(\mathbf{X}, Y)$, afin d'arrêter l'algorithme de quantification et d'ajouter/supprimer des variables dans notre sélection. A ces fins, dans [CDR07], nous avons défini et étudié sous un angle théorique des divergences d'information (normalisées) qui s'écrivent sous la forme : pour deux vecteurs $\mathbf{U}, \mathbf{V}$ (discrets ou catégoriels)

$$\Delta_{\mathbf{U},\mathbf{V}} = C_{\mathbf{U},\mathbf{V}} - I_{\mathbf{U},\mathbf{V}} \quad \text{et} \quad \delta_{\mathbf{U},\mathbf{V}} = \frac{C_{\mathbf{U},\mathbf{V}} - I_{\mathbf{U},\mathbf{V}}}{C_{\mathbf{U},\mathbf{V}}},$$

où $I_{\mathbf{U},\mathbf{V}}$ désigne l'information mutuelle entre $\mathbf{U}$ et $\mathbf{V}$ et où $C_{\mathbf{U},\mathbf{V}}$ désigne un terme de complexité qui va assurer un certain nombre de propriétés de base souhaitées pour δ et Δ : divergences symétriques, positives, minimales lorsque $\mathbf{U}$ et $\mathbf{V}$ partagent la même information, maximales lorsque $\mathbf{U}$ et $\mathbf{V}$ sont indépendantes, invariantes par transformations continues strictement croissantes, normalisées (i.e. appartenant à $[0,1]$ pour δ). Nous avons envisagé une très large classe de termes de complexité. Parmi eux, citons $C_{\mathbf{U},\mathbf{V}} = H_{\mathbf{U},\mathbf{V}}, \max(H_{\mathbf{U}}, H_{\mathbf{V}}), (H_{\mathbf{U}} + H_{\mathbf{V}})/2, \sqrt{H_{\mathbf{U}} H_{\mathbf{V}}}$. Notons que les deux premiers choix (dans leurs versions non normalisées) ont été introduits par Li et Vitanyi dans [82]. Notre travail a consisté à définir un certain nombre de propriétés utiles dans un cadre de prédiction comme par exemple l'obtention d'une inégalité triangulaire.

1.3 Modélisation et prédiction de la pneumonie nosocomiale

Ce travail s'inscrit dans la thèse de Molière Nguile Makao, co-encadrée par J.-F. Timsit (CHU Grenoble), Benoît Liquet (INSERM Bordeaux) et moi-même. Cette thèse s'intitule "Pneumonie nosocomiale acquise sous ventilation mécanique : prédiction du diagnostic et influence sur le pronostic" et sera soutenue en Novembre 2010.

La pneumonie nosocomiale est l'infection nosocomiale la plus fréquente et la plus grave rencontrée en réanimation, tout particulièrement chez les patients intubés et en ventilation mécanique. Le risque de pneumonie nosocomiale dépend de nombreux facteurs :

- La durée de ventilation mécanique est le risque le plus important : le risque cumulé d'acquisition de pneumonie augmente donc avec la durée de ventilation mécanique. Cependant, le risque d'acquisition à un moment donné (ou risque instantané) de la maladie n'est pas constant. En effet, le risque est maximal dans la première semaine de ventilation mécanique puis décroît progressivement.

- Plusieurs facteurs de risque liés au terrain ont aussi été rapportés et devront être pris en compte (âge, sexe masculin, bronchopathie chronique). D'autres sont liés à la gravité initiale (scores de gravité, utilisation de procédures invasives, intensité des suppléances d'organes).
- L'antibiothérapie reçue par le patient a une importance fondamentale qui nécessite une analyse poussée. L'antibiothérapie reçue dans les premiers jours de séjour en réanimation est un facteur protecteur rapporté dans plusieurs études. Cependant l'effet protecteur disparaît après la première semaine de ventilation, et l'antibiothérapie devient alors un facteur de risque. Par ailleurs, l'antibiothérapie préalable à la pneumonie nosocomiale modifie considérablement les micro-organismes rencontrés. Les germes de la pneumonie seront résistants à l'antibiothérapie préalable reçue par le malade.
- D'autres événements ou traitements utilisés en cours de séjour modifient encore le risque, en particulier l'extubation accidentelle, l'absence de position demi-assise chez un patient en nutrition entérale, l'utilisation de traitement modifiant le pH gastrique.

Les signes cliniques et para-cliniques d'appel de la pneumonie sont relativement simples (fièvre, expectoration purulente hyperleucocytose et leucopénie, infiltrat radiologique pulmonaire). Ils ne sont cependant pas du tout spécifiques puisque environ la moitié des patients suspects de pneumonie nosocomiale ont, en fait, des maladies infectieuses et non infectieuses qui miment la pneumonie mais justifient un traitement très différent (oedème pulmonaire cardiogénique, embolie pulmonaire). Même des réanimateurs experts, devant la présence de ces signes, se trompent une fois sur trois sur le diagnostic de pneumonie. Enfin, l'adéquation et la précocité de l'antibiothérapie conditionnent le pronostic, en particulier chez les patients de gravité intermédiaire.

Un outil permettant d'estimer le risque de pneumonie nosocomiale dans l'avenir proche pourrait ainsi permettre de repérer les patients pouvant justifier de tests diagnostiques précoces et d'un traitement immédiatement adapté. Dans ce contexte, l'objectif du travail est de développer un modèle statistique de prédiction des pneumonies nosocomiales en réanimation, i.e. de développer un outil et des procédures de suivi qui permettront d'identifier avec précision et de façon personnalisée, les patients qui courent le plus grand risque d'acquisition des pneumonies nosocomiales dans le but d'empêcher sa contraction.

Les données, sur lesquelles s'appuie cette thèse, proviennent de la base de données OUTCOMEREA. Cette base contient des informations diverses sur le séjour des patients qui ont été hospitalisés dans 11 hôpitaux de Paris et de sa région, ainsi que de la région Rhône-Alpes depuis 1997. Cette base est actualisée en permanence. L'échantillon d'intérêt décrit les données de 2873 patients dont 434 ont contracté une pneumonie nosocomiale durant leur séjour en réanimation ainsi que plus de 200 covariables (fixes comme le sexe, l'âge et dépendantes du temps comme l'évolution du score de gravité ou la prise d'antibiotiques par exemple).

L'état de santé d'un patient est modélisé par un modèle multi-états défini par un processus stochastique $\{X(t); t \geq 0\}$ à valeurs sur $S = \{0, 1, 2, 3\}$ (voir Figure 1.2) représentant respectivement être en réanimation sans pneumonie nosocomiale acquise sous ventilation mécanique (PN), être en réanimation avec une pneumonie nosocomiale, décéder en réanimation et sortir du service de réanimation. L'avantage de l'utilisation de l'approche multi-états comme modèle de survie des patients en réanimation est que le problème de l'hypothèse des censures non informatives, causées par les événements compétitifs quand on ne s'intéresse qu'à un seul événement à la fois, ne se pose plus. Dans ce modèle, le temps t représente le temps passé dans le service de réanimation et T1 représente le temps d'entrée dans l'état 1. A l'entrée dans le service de réanimation, tous les patients sont sains (sans pneumonie nosocomiale, i.e. $X(0) = 0$).

Deux approches ont été considérées pour étudier ce modèle : un modèle de Markov non-homogène et un modèle semi-markovien (les intensités de transition ne dépendent que de la durée de séjour dans un état). Les covariables (qu'il a fallu sélectionner par des études annexes et des procédures statistiques) sont introduites via un modèle de Cox. Les intensités de transition ont alors été estimés soient paramétriquement soit non paramétriquement (en estimant au préalable les

paramètres associés au modèle de Cox via la vraisemblance partielle). Ces méthodes et approches sont comparées et discutées. Par la suite, nous avons développé un estimateur individualisé de la probabilité à un instant donné de déclarer une pneumonie dans les trois jours suivants (cette prospective à trois jours est liée à la durée d'incubation de la maladie) et évalué sa pertinence et son intérêt pour le praticien. Un article en commun sur ce sujet est en préparation, il est associé à une conférence présentée à la SFDS [MCLT09].

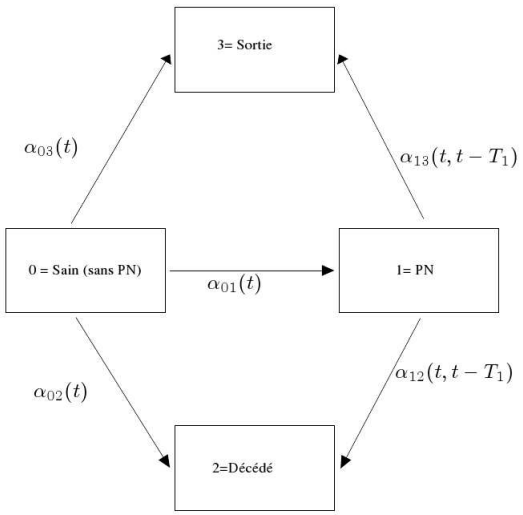


FIGURE 1.2 – Modèle multi-états utilisé pour la modélisation de la pneumonie nosocomiale

1.4 Statistique appliquée dans les domaines de l'économie et de la psychologie sociale

En collaboration avec Rémy Drouilhet, Sophie Ebermeyer et Mélanie Sevin (travaillant, à l'époque, chez Economie et Humanisme `http ://www.economie-humanisme.org/`), nous avons dépouillé des enquêtes de données de panel concernant le devenir de travailleurs bénéficiaires de contrats d'insertion. Ce travail a été concrétisé par la rédaction de trois rapports techniques pour la DRTEFP et la DR ANPE [KES+04a, KES+04b, KES+04c].

Dans le cadre d'un contrat entre le Laboratoire Interuniversitaire de Psychologie et le LJK (J.-F. Coeurjolly et R. Drouilhet), nous avons mis en place une collaboration avec Aurélie Derbier pour mettre en place et effectuer les traitements statistiques lors de sa thèse (soutenue en Juin 2010) intitulée "Etude épidémiologique des facteurs de risque et de protection du suicide. Mise en place et évaluation du programme Coping And Support Training au collège et au lycée". Une question méthodologique soulevée par l'application nous a amenés à proposer un nouvel outil.

Développons un peu ceci. L'étude menée par Aurélie Derbier consistait à faire passer tout une série de questionnaires à une population de collégiens et lycéens. Ces questionnaires, déjà utilisés au Canada et aux Etats-Unis, mesurent le bien être des élèves, leur caractère dépressif, leur degré de consommation d'alcool et de drogue, leur niveau de sociabilisation,... De nombreux scores individuels sont établis à partir de ces questionnaires et un niveau de risque suicidaire individuel est obtenu. Le travail d'ordre clinique (réalisé par Aurélie) a consisté à réaliser un certain nombre d'ateliers de groupes (discussion, réflexions individuelles, mises en situation,...) d'élèves jugés suicidaires et non suicidaires, à refaire passer l'ensemble des questionnaires au maximum

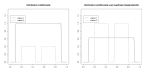
d'élèves et d'essayer de mesurer l'efficacité des ateliers en analysant l'évolution des scores. En particulier, il nous fallait mettre en place des tests statistiques pour comparer par sous-groupe (suicidaires, non suicidaires) l'évolution d'un score entre les deux séries de questionnaires. Nous nous sommes heurtés à plusieurs difficultés : premièrement les scores sont clairement des réalisations d'une variable discrète parfois bi-modale (très loin d'une gaussienne) et deuxièmement nous avions de nombreuses données manquantes (plus de la moitié) ; certains élèves ayant passé la première série de questionnaires et pas la seconde, et inversement.

Pour pouvoir répondre à notre problématique, nous ne pouvions pas utiliser les méthodes d'imputation des données manquantes utilisant la vraisemblance. Nous nous sommes orientés vers une stratégie non paramétrique et asymptotique. Nous avons considéré plusieurs stratégies pour imputer les données manquantes et tenter d'améliorer l'estimateur de la différence de moyenne obtenu en prenant en compte uniquement les données appariées : des méthodes consistant à proposer pour toute donnée manquante en session 2 (par exemple) la moyenne des variables de la session 2 observées ou la prédiction par un modèle de régression linéaire estimé avec les données appariées. A notre connaissance, ce problème n'avait jamais été regardé sous cet angle et nous avons obtenu des résultats simples sur la variance des nouveaux estimateurs ainsi que la loi asymptotique de la statistique centrée réduite. Nous avons tenté de comparer ces différentes méthodologies d'un point de vue théorique en fonction des différentes tailles d'échantillon (i.e. du nombre de données manquantes en session 1, du nombre de données appariées et du nombre de données manquantes en session 2) et du niveau de corrélation des scores en session 1 et 2. Il est difficile d'énoncer une conclusion générale. Néanmoins, deux commentaires très instructifs peuvent être établis : premièrement imputer par la moyenne n'améliore pas forcément l'estimateur initial et peut même le dégrader. Deuxièmement, imputer par une méthode de type régression linéaire est une stratégie toujours meilleure que celle de ne rien faire. Pour plus de détails sur la définition des estimateurs, le calcul des variances asymptotiques et les comparaisons théoriques et numériques (essentiellement en fonction du niveau de corrélation des variables avant et après) ainsi que le TCL obtenu, nous renvoyons le lecteur au travail joint à ce mémoire [CDD10] qui constitue une partie du travail en cours.

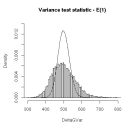
Chapitre 2

Liste des travaux en relation avec la Partie III

Articles parus

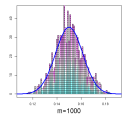


[CDR07]
J.-F. Coeurjolly, R. Drouilhet and J.-F. Robineau. Normalized information-based divergences. *Problems of Information Transmission*, 43(3):167-189, 2007.

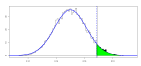


[CDLR09]
J.-F. Coeurjolly and R. Drouilhet and P. Lafaye de Micheaux and J.-F. Robineau. asympTest: an `R` package for performing parametric statistical tests and confidence intervals based on the central limit theorem. *R Journal*. 2:26-30 (Long version (19p.) at `http://arxiv.org/abs/0902.0506`), 2009.

Actes de congrés

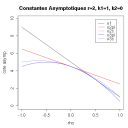


[CD05a]
J.-F. Coeurjolly and R. Drouilhet. Approche expérimentale des probabilités comme complément à une approche plus classique, In *37èmes journées de statistiques*, Pau, France, 2005



[CD05b]
J.-F. Coeurjolly and R. Drouilhet. Construction d'un test d'hypothèses par une approche visuelle et une approche expérimentale des probabilités, In *37èmes journées de statistiques*, Pau, France, 2005

Rapport technique

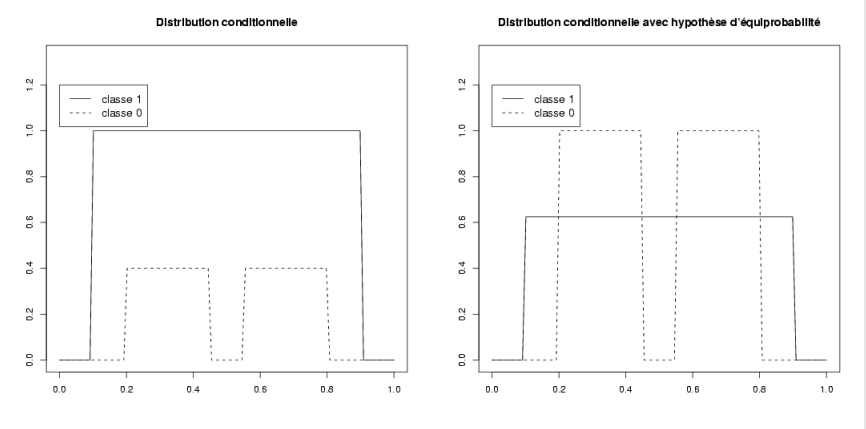


[CDD10]
J.-F. Coeurjolly, A. Derbier and R. Drouilhet. Comparaison de différentes stratégies pour estimer une différence de moyennes en présence de données appariées et indépendantes dans un cadre non nécessairement gaussien, *technical report*, 2010.

Chapitre 3

Annex Part III : published and submitted papers

This chapter contains a copy of published or submitted papers related to Part III. These papers are chronologically ordered.



[CDR07]
J.-F. Coeurjolly, R. Drouilhet and J.-F. Robineau. Normalized information-based divergences. *Problems of Information Transmission*, 43(3):167-189, 2007.

ISSN 0032-9460, Problems of Information Transmission, 2007, Vol. 43, No. 3, pp. 167–189. © Pleiades Publishing, Inc., 2007.
Original Russian Text © J.-F. Coeurjolly, R. Drouilhet, J.-F. Robineau, 2007, published in Problemy Peredachi Informatsii, 2007, Vol. 43, No. 3, pp. 3–27.

INFORMATION THEORY

Normalized Information-Based Divergences

J.-F. Coeurjolly, R. Drouilhet, and J.-F. Robineau

Université Pierre Mendès-France, Grenoble, France
Jean-Francois.Coeurjolly@upmf-grenoble.fr
Remy.Drouilhet@upmf-grenoble.fr

Received April 11, 2006; in final form, May 16, 2007

Abstract—This paper is devoted to the mathematical study of some divergences based on mutual information which are well suited to categorical random vectors. These divergences are generalizations of the "entropy distance" and "information distance." Their main characteristic is that they combine a complexity term and the mutual information. We then introduce the notion of (normalized) information-based divergence, propose several examples, and discuss their mathematical properties, in particular, in some prediction framework.

DOI: 10.1134/S0032946007030015

1. INTRODUCTION

Shannon's information theory, usually just called *information theory*, was introduced in 1948 [1]. The theory is aimed at providing means for measuring information. More precisely, the amount of information in an object may be measured by its *entropy* and may be interpreted as the length of the description of the object in some encoding way. In the Shannon approach, the objects to be encoded are assumed to be outcomes of a known source. Shannon's theory also provides the notion of *mutual information* (related to two objects), which plays a central role in many applications, from lossy compression to machine learning methods.

Several authors noted that it would be useful to modify the mutual information such that the resulting quantity becomes a metric in a strict sense. As a first example, [2, 3] introduced the *entropy distance* defined as the sum of conditional entropies. Other interesting measures are the *information distance* [4] and its normalized version, the *similarity metric*, introduced in [5] in the context of the Kolmogorov complexity theory. More precisely, the information distance is defined as the maximum of the conditional Kolmogorov complexities. The similarity metric is universal in the sense defined by the authors and is not computable since it is based on an uncomputable notion of the Kolmogorov complexity.

Recent papers have demonstrated that applications of suitable versions of the similarity metric are of use in areas as diverse as genomics, virology, languages, literature, music, handwritten digits, and astronomy [6]. To apply the metric to real data, the authors have to replace the use of the non-computable Kolmogorov complexity with an approximation obtained by using standard real-world compressors: GenCompress for genomics [7], the *Normalized Compression Distance* (NCD) for music clustering [8], and the *Normalized Google Distance* (NGD) for automatic meaning discovery [9] are examples of effective compressors. To include the information distance and similarity metric in a framework based on information theory concepts, we make use of the principle that *the expected Kolmogorov complexity equals the Shannon entropy*; an interested reader is referred to [10–12] for more details.

Consequently, both the entropy distance and information distance are expressed in terms of conditional entropies: the first one as their sum and the second as their maximum. In [13] there is given a proof of the triangle inequality for these distances and their respective normalized versions.

In the supervised learning framework, the use of some method of selecting covariables among a large number is required when it is assumed that the data size is too small with respect to the number of available covariables (in order to apply any existing discriminant analysis method). Such a problem has been widely treated (see, e.g., [14]). The approach undertaken in [15] is mainly based on three kinds of methodological tools. The first one is a supervised quantization method consisting in the simplification of too complex covariables (in particular, with a too large number of possible values). Indeed, our main belief is that in order to predict the class variable generally representing a small number of categories of data, each possibly predictive covariable must be not too complex. The second one is a more usual step-by-step selection method combining the simplified covariables together in order to detect a cluster of data of the same class. The last one is aimed at detecting redundancy among the set of covariables. These three tasks can be realized using the entropy or information distances (or their normalized versions). Let us emphasize some properties allowing one to understand the usefulness of these criteria in such a context. The entropy and information distances, D^E and D^I, can be rewritten as the difference between some term (respectively, the joint entropy and the maximum of the marginal entropies) and the mutual information. The first term may be interpreted as a complexity term. Moreover, both are independence measures with the particular property to be minimal (in fact equal to 0) when random vectors share exactly the same information. In [15] it was then proposed to extend the definition of the entropy and information distances by introducing the notion of information-based divergence $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$ between two categorical random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, defined as the difference of some complexity term $C_{\boldsymbol{X},\boldsymbol{Y}}$ and the mutual information $I_{\boldsymbol{X},\boldsymbol{Y}}$ and such that $C_{\boldsymbol{X},\boldsymbol{Y}}$ is an upper bound for $I_{\boldsymbol{X},\boldsymbol{Y}}$ reached when $\boldsymbol{X}$ and $\boldsymbol{Y}$ share exactly the same information. The notion of the normalized information-based divergence $\delta_{\boldsymbol{X},\boldsymbol{Y}}$ is directly derived by dividing the associated information-based divergence $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$ by the complexity term $C_{\boldsymbol{X},\boldsymbol{Y}}$. The normalized versions d^E and d^I of D^E and D^I are particular examples. Other examples are given in [15]. Among them, one is of particular interest since its complexity term C^S is the mean of marginal entropies. The associated (unnormalized) information-based divergence Δ^S is not so different from D^E since it corresponds to its half. Nevertheless, the expression for its complexity term C^S really differs from the complexity term C^E of D^E (i.e., the joint entropy). For practical purposes, we may argue that D^I, D^E, and Δ^S are not well suited in the prediction framework since a small value of these distances means that both the explained and explicative variables have a good knowledge of each other. This is due to the fact that both conditional entropies have at least the same weight.

In this paper, this drawback is weakened by introducing a natural extension $C^{S,\alpha}$ of the complexity term C^S defined as a weighted mean (by α and $1-\alpha$ for some $0 < \alpha \leq 1$) of the minimum and maximum of marginal entropies. This kind of complexity term leads to an expected I-divergence $\Delta^{S,\alpha}$, which is the weighted mean of the minimum and maximum of conditional entropies.

The paper is organized as follows. In Section 2 we recall the definitions and the main properties of the entropy and information distances (and their normalized versions). Similarly to [16], we extract the main characteristics to define some general concept of information divergence which could theoretically be applied in a more general setting (continuous, discrete, etc.). In Section 3 we concentrate on categorical data (and in particular discrete) random vectors, since this is usually the case in most applications that use entropy or information distance. We give the definition of the (normalized) information-based divergence and propose several examples. We study their mathematical properties in the general context and propose some sufficient conditions for these divergences to verify some triangle-type inequality. Finally, in Section 4 we exhibit some properties

of information-based divergences in the special prediction framework. In particular, we show that these divergences are useful to detect redundancy.

2. NORMALIZED ENTROPY DISTANCE AND NORMALIZED INFORMATION DISTANCE

Let us denote by Γ the set of categorical random vectors, that is, discrete-valued random vectors with finite entropy. In the following, $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ are three elements of Γ.

2.1. Notation

We denote by $H_{\boldsymbol{X}}$ (when it exists) the Shannon entropy of $\boldsymbol{X}$ given by

$$H_{\boldsymbol{X}} = -\sum_{\boldsymbol{x}\in\Omega_{\boldsymbol{X}}} p_{\boldsymbol{X}}(\boldsymbol{x})\log(p_{\boldsymbol{X}}(\boldsymbol{x})), \quad \text{with} \quad p_{\boldsymbol{X}}(\boldsymbol{x}) = \mathbf{P}(\boldsymbol{X}=\boldsymbol{x}).$$

In the same way, one can define the joint entropy of $\boldsymbol{X}$ and $\boldsymbol{Y}$, denoted by $H_{\boldsymbol{X},\boldsymbol{Y}}$, and the conditional entropy of $\boldsymbol{X}$ by $\boldsymbol{Y}$ (respectively, $\boldsymbol{Y}$ by $\boldsymbol{X}$), denoted by $H_{\boldsymbol{X}|\boldsymbol{Y}}$ (respectively, $H_{\boldsymbol{Y}|\boldsymbol{X}}$). Finally, we denote by $I_{\boldsymbol{X},\boldsymbol{Y}}$ the mutual information between the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$. When these different quantities exist, there are the following relations (see, e.g., [17]):

$$H_{\boldsymbol{X},\boldsymbol{Y}} = H_{\boldsymbol{X}} + H_{\boldsymbol{Y}|\boldsymbol{X}} = H_{\boldsymbol{Y}} + H_{\boldsymbol{X}|\boldsymbol{Y}}, \tag{1}$$

$$I_{\boldsymbol{X},\boldsymbol{Y}} = H_{\boldsymbol{X}} - H_{\boldsymbol{X}|\boldsymbol{Y}} = H_{\boldsymbol{Y}} - H_{\boldsymbol{Y}|\boldsymbol{X}} = H_{\boldsymbol{X}} + H_{\boldsymbol{Y}} - H_{\boldsymbol{X},\boldsymbol{Y}}. \tag{2}$$

2.2. Definitions and Some Basic Properties

We now present some measures allowing us to overcome some drawbacks of the mutual information. As a first generalization, several authors noted that it would be useful to modify the mutual information such that the resulting quantity becomes a metric in a strict sense. Two such measures exist and are well known in the literature. The first one, called the "entropy distance," is derived from the domain of information theory. The second one, called "information distance," originates in works around the Kolmogorov complexity. These measures are defined (when they exist) for two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ as follows:

- Entropy distance
$$D^E_{\boldsymbol{X},\boldsymbol{Y}} = H_{\boldsymbol{X}|\boldsymbol{Y}} + H_{\boldsymbol{Y}|\boldsymbol{X}}; \tag{3}$$
- Information distance
$$D^I_{\boldsymbol{X},\boldsymbol{Y}} = \max(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}}). \tag{4}$$

Both measures are indeed some modifications of the mutual information since from (1) and (2) we have

$$D^E_{\boldsymbol{X},\boldsymbol{Y}} = H_{\boldsymbol{X},\boldsymbol{Y}} - I_{\boldsymbol{X},\boldsymbol{Y}} \qquad \text{and} \qquad D^I_{\boldsymbol{X},\boldsymbol{Y}} = \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}) - I_{\boldsymbol{X},\boldsymbol{Y}}. \tag{5}$$

The quantities $H_{\boldsymbol{X},\boldsymbol{Y}}$ and $\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ are upper bounds for the mutual information $I_{\boldsymbol{X},\boldsymbol{Y}}$ and are reached when $\boldsymbol{X}$ and $\boldsymbol{Y}$ share exactly the same information. In other words, these two measures are nonnegative and vanish if and only if $H_{\boldsymbol{Y}|\boldsymbol{X}} = H_{\boldsymbol{X}|\boldsymbol{Y}} = 0$, expressing the fact that $\boldsymbol{X}$ (respectively, $\boldsymbol{Y}$) predicts $\boldsymbol{Y}$ (respectively, $\boldsymbol{X}$) with probability 1.

These measures satisfy

$$D^E_{\boldsymbol{X},\boldsymbol{Y}} \le H_{\boldsymbol{X},\boldsymbol{Y}} \qquad \text{and} \qquad D^I_{\boldsymbol{X},\boldsymbol{Y}} \le \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}), \tag{6}$$

where the equality holds if the vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent. In [18, 19] it was noted that in bioinformatics an unnormalized distance may not be a proper evolutionary distance measure. To overcome this problem within the algorithmic framework, they form two normalized versions of distances, D^E and D^I. Their Shannon versions were proposed and studied in [13].

Definition 1. When they exist, one defines the following two measures:

- Normalized entropy distance
$$d^E_{\boldsymbol{X},\boldsymbol{Y}} = \frac{H_{\boldsymbol{X}|\boldsymbol{Y}} + H_{\boldsymbol{Y}|\boldsymbol{X}}}{H_{\boldsymbol{X},\boldsymbol{Y}}};$$
- Normalized information distance
$$d^I_{\boldsymbol{X},\boldsymbol{Y}} = \frac{\max(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}})}{\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})}.$$

Since $H_{\boldsymbol{X},\boldsymbol{Y}} = 0 \Leftrightarrow H_{\boldsymbol{X}} = H_{\boldsymbol{Y}} = 0 \Leftrightarrow \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}) = 0$, we set by convention $d^E_{\boldsymbol{X},\boldsymbol{Y}} = 0$ (respectively, $d^I_{\boldsymbol{X},\boldsymbol{Y}} = 0$) when $H_{\boldsymbol{X}} = H_{\boldsymbol{Y}} = 0$.

We are encouraged to define the following equivalence class: the vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be equivalent if $\boldsymbol{X}$ (respectively, $\boldsymbol{Y}$) predicts $\boldsymbol{Y}$ (respectively, $\boldsymbol{X}$) with probability 1; we denote

$$\boldsymbol{X} \sim \boldsymbol{Y} \quad \Longleftrightarrow \quad H_{\boldsymbol{Y}|\boldsymbol{X}} = H_{\boldsymbol{X}|\boldsymbol{Y}} = 0 \quad \Longleftrightarrow \quad I_{\boldsymbol{X},\boldsymbol{Y}} = H_{\boldsymbol{X},\boldsymbol{Y}} = H_{\boldsymbol{X}} = H_{\boldsymbol{Y}}. \tag{7}$$

Due to the previous convention,

$$d^E_{\boldsymbol{X},\boldsymbol{Y}} = 0 \quad \Longleftrightarrow \quad d^I_{\boldsymbol{X},\boldsymbol{Y}} = 0 \quad \Longleftrightarrow \quad \boldsymbol{X} \sim \boldsymbol{Y}.$$

From (1) and (2), one can obtain the following expressions for these two measures, allowing some new interpretations.

Proposition 1. *We have the following expressions for $d^E_{\boldsymbol{X},\boldsymbol{Y}}$ and $d^I_{\boldsymbol{X},\boldsymbol{Y}}$:*

$$d^E_{\boldsymbol{X},\boldsymbol{Y}} = 1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{H_{\boldsymbol{X},\boldsymbol{Y}}}, \tag{8}$$

$$d^I_{\boldsymbol{X},\boldsymbol{Y}} = 1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})} \tag{9}$$

$$= \max\left(\frac{H_{\boldsymbol{X}|\boldsymbol{Y}}}{H_{\boldsymbol{X}}}, \frac{H_{\boldsymbol{Y}|\boldsymbol{X}}}{H_{\boldsymbol{Y}}}\right). \tag{10}$$

Proposition 2. *The measures d^E and d^I constitute two distances bounded by* 1.

To our knowledge, these results were proved in [13]. The proofs are very similar to those of [20], where algorithmic version of these distances were considered. The proof is therefore omitted, but in Section 3.3 we propose a result extending this one in the sense that we give conditions for measures that can be written as (8) and (9) to constitute a metric.

2.3. Concept of Information Divergence

From the previous study related to D^I, D^E, d^I, and d^E, we can exhibit some characteristics useful for an attempt to define the concept of information divergence Δ in a more general setting. Let us first consider a similarity measure $\mathcal{I}_{\boldsymbol{X},\boldsymbol{Y}}$ (not necessarily the mutual information) which is minimal (in fact equal to 0) when $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent and maximal (in fact equal to $\mathcal{I}_{\boldsymbol{X},\boldsymbol{X}} = \mathcal{I}_{\boldsymbol{Y},\boldsymbol{Y}}$) when the distributions of $\boldsymbol{X}$ given $\boldsymbol{Y} = \boldsymbol{y}$ and $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ are trivial. An information divergence $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$ could satisfy the following properties:

[P1] symmetry: $\Delta_{\boldsymbol{X},\boldsymbol{Y}} = \Delta_{\boldsymbol{Y},\boldsymbol{X}}$;
[P2] nonnegativeness: $\Delta_{\boldsymbol{X},\boldsymbol{Y}} \geq 0$;
[P3] $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$ is minimal (i.e., $\Delta_{\boldsymbol{X},\boldsymbol{Y}} = 0$) if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ share exactly the same information (i.e., $\mathcal{I}_{\boldsymbol{X},\boldsymbol{Y}}$ is maximal);
[P4] $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$ is maximal if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent (i.e., $\mathcal{I}_{\boldsymbol{X},\boldsymbol{Y}} = 0$).

Other supplementary properties could be that $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$:

[P5] is normalized: $\Delta_{\boldsymbol{X},\boldsymbol{Y}} \in [0,1]$ and $\Delta_{\boldsymbol{X},\boldsymbol{Y}} = 1$ when $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent;
[P6] satisfies the triangle inequality: $\Delta_{\boldsymbol{X},\boldsymbol{Y}} \leq \Delta_{\boldsymbol{X},\boldsymbol{Z}} + \Delta_{\boldsymbol{Z},\boldsymbol{Y}}$;
[P7] is invariant under continuous and strictly increasing transformations $\boldsymbol{\varphi}(\cdot)$ and $\boldsymbol{\psi}(\cdot)$ of the vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ whenever they are quantitative random vectors.

There exists a vast literature discussing criteria that satisfy the stated properties. We may cite [21] or a recent work [16], where it is proposed to detect the dependence between two possibly nonlinear processes through the Bhattacharya–Matusita–Hellinger measure of dependence, given by

$$S_\rho = \frac{1}{2} \iint \Big(\sqrt{f_1(\boldsymbol{x},\boldsymbol{y})} - \sqrt{f_2(\boldsymbol{x},\boldsymbol{y})} \Big)^2 d\boldsymbol{x}\, d\boldsymbol{y},$$

where f_1 (respectively, f_2) is the joint density (respectively, product of marginal densities) of $\boldsymbol{X}$ and $\boldsymbol{Y}$. This measure, which has another advantage to be applicable to both continuous or discrete variables, satisfies properties **[P1]**–**[P7]** (in fact, let us precise that **[P7]** is only valid if $\boldsymbol{\varphi}(\cdot) = \boldsymbol{\psi}(\cdot)$).

In some framework where the purpose is to predict some reference variable, one may find interesting to work with a divergence $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$ which combines the minimization of a nonnegative complexity term denoted by $\mathcal{C}_{\boldsymbol{X},\boldsymbol{Y}}$ and the maximization of a nonnegative information term $\mathcal{I}_{\boldsymbol{X},\boldsymbol{Y}}$. The quantity $\mathcal{C}_{\boldsymbol{X},\boldsymbol{Y}}$ is called a complexity term since it is assumed to be expressed as a function of $\mathcal{H}_{\boldsymbol{X}}$, $\mathcal{H}_{\boldsymbol{Y}}$, and $\mathcal{H}_{\boldsymbol{X},\boldsymbol{Y}}$ measuring in some way the complexity of the vectors $\boldsymbol{X}$, $\boldsymbol{Y}$, and $(\boldsymbol{X},\boldsymbol{Y})$, respectively. In other words, we may expect that an information divergence $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$ could also satisfy the following properties:

[P8] When $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same complexity (in the sense that $\mathcal{C}_{\boldsymbol{Y},\boldsymbol{X}_1} = \mathcal{C}_{\boldsymbol{Y},\boldsymbol{X}_2}$), we have $\Delta_{\boldsymbol{Y},\boldsymbol{X}_1} < \Delta_{\boldsymbol{Y},\boldsymbol{X}_2}$ whenever $\boldsymbol{X}_1$ has a better knowledge about $\boldsymbol{Y}$ than $\boldsymbol{X}_2$ (i.e., $\mathcal{I}_{\boldsymbol{Y},\boldsymbol{X}_1} > \mathcal{I}_{\boldsymbol{Y},\boldsymbol{X}_2}$);
[P9] When $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same knowledge about $\boldsymbol{Y}$ (i.e., $\mathcal{I}_{\boldsymbol{Y},\boldsymbol{X}_1} = \mathcal{I}_{\boldsymbol{Y},\boldsymbol{X}_2}$), we have $\Delta_{\boldsymbol{Y},\boldsymbol{X}_1} < \Delta_{\boldsymbol{Y},\boldsymbol{X}_2}$ whenever $\boldsymbol{X}_1$ is simpler than $\boldsymbol{X}_2$ in the sense that $\mathcal{C}_{\boldsymbol{Y},\boldsymbol{X}_1} < \mathcal{C}_{\boldsymbol{Y},\boldsymbol{X}_2}$. Moreover, in this particular situation
[P10] $\mathcal{C}_{\boldsymbol{Y},\boldsymbol{X}_1} \leq \mathcal{C}_{\boldsymbol{Y},\boldsymbol{X}_2}$ must be equivalent to $\mathcal{H}_{\boldsymbol{X}_1} \leq \mathcal{H}_{\boldsymbol{X}_2}$;
[P11] When $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ share almost exactly the same information (i.e., $\mathcal{I}_{\boldsymbol{X}_1,\boldsymbol{X}_2}$ is almost maximal and $\Delta_{\boldsymbol{X}_1,\boldsymbol{X}_2} \simeq 0$), the difference between the divergences $\Delta_{\boldsymbol{Y},\boldsymbol{X}_1}$ and $\Delta_{\boldsymbol{Y},\boldsymbol{X}_2}$ is almost zero (i.e., $\Delta_{\boldsymbol{Y},\boldsymbol{X}_1} \simeq \Delta_{\boldsymbol{Y},\boldsymbol{X}_2}$).

A class of candidates that satisfy **[P8]** and **[P9]** could be of the form

$$\Delta_{\boldsymbol{X},\boldsymbol{Y}} = \frac{\mathcal{C}_{\boldsymbol{X},\boldsymbol{Y}} - \mathcal{I}_{\boldsymbol{X},\boldsymbol{Y}}}{\mathcal{W}_{\boldsymbol{X},\boldsymbol{Y}}}, \tag{11}$$

where $\mathcal{W}_{\boldsymbol{X},\boldsymbol{Y}}$ is a positive term. When $\mathcal{W}_{\boldsymbol{X},\boldsymbol{Y}} = \mathcal{C}_{\boldsymbol{X},\boldsymbol{Y}}$, we obtain a normalized information divergence. Properties **[P2]** and **[P3]** and relation (11) imply that $\mathcal{C}_{\boldsymbol{X},\boldsymbol{Y}}$ is an upper bound for $\mathcal{I}_{\boldsymbol{X},\boldsymbol{Y}}$, reached when $\boldsymbol{X}$ and $\boldsymbol{Y}$ share exactly the same information.

In the rest of this paper we concentrate on divergences of the form (11) that are in addition well suited to categorical random variables (and in particular discrete random variables). In such a framework, we shall only describe some entropic-based criteria (i.e., $\mathcal{H}_{\boldsymbol{X}} = H_{\boldsymbol{X}}$), and so the information term will be set to the mutual information $I_{\boldsymbol{X},\boldsymbol{Y}}$.

3. INFORMATION-BASED DIVERGENCES AND THEIR NORMALIZED VERSIONS

3.1. Definition and Examples

Definition 2. Two criteria Δ and δ are called, respectively, an information-based divergence and a normalized information-based divergence (for short, I-divergence and NI-divergence) if they

can be written as

$$\Delta_{\boldsymbol{X},\boldsymbol{Y}} = C_{\boldsymbol{X},\boldsymbol{Y}} - I_{\boldsymbol{X},\boldsymbol{Y}}, \tag{12}$$

$$\delta_{\boldsymbol{X},\boldsymbol{Y}} = \frac{C_{\boldsymbol{X},\boldsymbol{Y}} - I_{\boldsymbol{X},\boldsymbol{Y}}}{C_{\boldsymbol{X},\boldsymbol{Y}}} = 1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{C_{\boldsymbol{X},\boldsymbol{Y}}}, \tag{13}$$

where the term $C_{\boldsymbol{X},\boldsymbol{Y}}$ constitutes a complexity term satisfying

(i) $C_{\boldsymbol{X},\boldsymbol{Y}} = C_{\boldsymbol{Y},\boldsymbol{X}}$;
(ii) $I_{\boldsymbol{X},\boldsymbol{Y}} \leq C_{\boldsymbol{X},\boldsymbol{Y}}$, and this bound is achieved if and only if the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are equivalent, i.e., if and only if $\boldsymbol{X} \sim \boldsymbol{Y}$.

We set by convention $\delta_{\boldsymbol{X},\boldsymbol{Y}} = 0$ when $C_{\boldsymbol{X},\boldsymbol{Y}} = I_{\boldsymbol{X},\boldsymbol{Y}} = 0$.

This definition implies automatically that an I-divergence $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$ (respectively, NI-divergence $\delta_{\boldsymbol{X},\boldsymbol{Y}}$) satisfies properties **[P1]**–**[P4]** (respectively, **[P1]**–**[P5]**). In the rest of the paper, the term $C_{\boldsymbol{X},\boldsymbol{Y}}$ is expressed as

$$C_{\boldsymbol{X},\boldsymbol{Y}} = f_C(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}}, I_{\boldsymbol{X},\boldsymbol{Y}}), \tag{14}$$

where $f_C(\cdot,\cdot,\cdot)$ is a nonnegative function. Under this expression for $C_{\boldsymbol{X},\boldsymbol{Y}}$, property **[P7]** is ensured since the conditional entropies and mutual information depend only on the joint probability distribution of the categorical random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$.

From now on, we propose a series of examples where we adopt the following convention: an I-divergence (respectively, NI-divergence) satisfying the triangle inequality is denoted by D (respectively, d) rather than Δ (respectively, δ). Moreover, each example will be particularized by some discriminating additional letter in the same manner as D^E and D^I (respectively, d^E and d^I), which clearly constitute I-divergences (respectively, NI-divergences).

In [15], we investigate two new entropic criteria naturally expressed by

$$\delta^D_{\boldsymbol{X},\boldsymbol{Y}} = \frac{1}{2}\left(\frac{H_{\boldsymbol{X}|\boldsymbol{Y}}}{H_{\boldsymbol{X}}} + \frac{H_{\boldsymbol{Y}|\boldsymbol{X}}}{H_{\boldsymbol{Y}}}\right) \qquad \text{and} \qquad \delta^S_{\boldsymbol{X},\boldsymbol{Y}} = \frac{H_{\boldsymbol{X}|\boldsymbol{Y}} + H_{\boldsymbol{Y}|\boldsymbol{X}}}{H_{\boldsymbol{X}} + H_{\boldsymbol{Y}}},$$

which can be rewritten as NI-divergences:

$$\delta^D_{\boldsymbol{X},\boldsymbol{Y}} = 1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{C^D_{\boldsymbol{X},\boldsymbol{Y}}} \quad \text{with} \quad C^D_{\boldsymbol{X},\boldsymbol{Y}} = \left(\frac{1}{2}\left(\frac{1}{H_{\boldsymbol{X}}} + \frac{1}{H_{\boldsymbol{Y}}}\right)\right)^{-1}, \tag{15}$$

$$\delta^S_{\boldsymbol{X},\boldsymbol{Y}} = 1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{C^S_{\boldsymbol{X},\boldsymbol{Y}}} \quad \text{with} \quad C^S_{\boldsymbol{X},\boldsymbol{Y}} = \frac{1}{2}(H_{\boldsymbol{X}} + H_{\boldsymbol{Y}}). \tag{16}$$

Their unnormalized versions are expressed as $\Delta^D_{\boldsymbol{X},\boldsymbol{Y}} = C^D_{\boldsymbol{X},\boldsymbol{Y}} - I_{\boldsymbol{X},\boldsymbol{Y}}$ and $D^S_{\boldsymbol{X},\boldsymbol{Y}} = C^S_{\boldsymbol{X},\boldsymbol{Y}} - I_{\boldsymbol{X},\boldsymbol{Y}}$.

In this paper, we are interested in a large family of I-divergences or NI-divergences with complexity terms of the form

$$C^\alpha_{\boldsymbol{X},\boldsymbol{Y}} = g^{-1}\big(\alpha g(m_{\boldsymbol{X},\boldsymbol{Y}}) + (1-\alpha) g(M_{\boldsymbol{X},\boldsymbol{Y}})\big), \tag{17}$$

with $m_{\boldsymbol{X},\boldsymbol{Y}} = \min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ and $M_{\boldsymbol{X},\boldsymbol{Y}} = \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ and where $0 \leq \alpha < 1$ and $g(\cdot)$ is any monotone function on $\mathbb{R}^+$. When this is not ambiguous, we set $m = m_{\boldsymbol{X},\boldsymbol{Y}}$ and $M = M_{\boldsymbol{X},\boldsymbol{Y}}$. To be convinced that I-divergences and NI-divergences with complexity terms of the form (17) satisfy Definition 2(ii), let us note that

$$I_{\boldsymbol{X},\boldsymbol{Y}} = g^{-1}\big(\alpha g(I_{\boldsymbol{X},\boldsymbol{Y}}) + (1-\alpha) g(I_{\boldsymbol{X},\boldsymbol{Y}})\big) \leq g^{-1}\big(\alpha g(m) + (1-\alpha) g(M)\big).$$

When $\alpha = 0$, the complexity term C^α corresponds to C^I. When $\alpha = 1$, the complexity term defined as $\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ and denoted by $C^{\min}_{\boldsymbol{X},\boldsymbol{Y}}$ does not satisfy Definition 2(ii) and therefore **[P3]**. The associated $\Delta^{\min}$ (respectively, $\delta^{\min}$) is not an I-divergence (respectively, NI-divergence).

Now we pay particular attention to the complexity terms $C^{D,\alpha}$, $C^{S,\alpha}$, $C^{R,\alpha}$, and $C^{P,\alpha}$ of the form (17) with, respectively, $g^D(\cdot) = 1/\cdot$, $g^S(\cdot) = \cdot$, $g^R(\cdot) = \sqrt{\cdot}$, and $g^P(\cdot) = \log(\cdot)$:

$$C^{D,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} = \left(\alpha \frac{1}{\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})} + (1-\alpha)\frac{1}{\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})}\right)^{-1}, \tag{18}$$

$$C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} = \alpha\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}) + (1-\alpha)\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}), \tag{19}$$

$$C^{R,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} = \left(\alpha\sqrt{\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})} + (1-\alpha)\sqrt{\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})}\,\right)^2, \tag{20}$$

$$C^{P,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} = \min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})^{\alpha}\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})^{1-\alpha}. \tag{21}$$

The previous measures Δ^S, δ^S, Δ^D, and δ^D are particular examples of such a family since the value of $\alpha = \frac{1}{2}$ leads to $C^{1/2}_{\boldsymbol{X},\boldsymbol{Y}} = g^{-1}\big(\frac{1}{2}g(H_{\boldsymbol{X}}) + \frac{1}{2}g(H_{\boldsymbol{Y}})\big)$. When $\alpha = \frac{1}{2}$, $\Delta^{\bullet,\alpha}$ and $\delta^{\bullet,\alpha}$ will simply be denoted by $\Delta^{\bullet}$ and $\delta^{\bullet}$, where $\bullet$ stands for S, R, P, and D.

Let us first comment on the particular expressions of the divergences $\Delta^{S,\alpha}$ and $\delta^{D,\alpha}$ associated to $C^{D,\alpha}$ and $C^{S,\alpha}$ given by

$$\begin{aligned}
\Delta^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} &= \alpha\min(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}}) + (1-\alpha)\max(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}}) \\
&= \alpha\Delta^{\min}_{\boldsymbol{X},\boldsymbol{Y}} + (1-\alpha)D^{I}_{\boldsymbol{X},\boldsymbol{Y}}, \\
\delta^{D,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} &= \alpha\min\begin{pmatrix} H_{\boldsymbol{X}|\boldsymbol{Y}} & H_{\boldsymbol{Y}|\boldsymbol{X}} \\ H_{\boldsymbol{X}} & H_{\boldsymbol{Y}} \end{pmatrix} + (1-\alpha)\max\begin{pmatrix} H_{\boldsymbol{X}|\boldsymbol{Y}} & H_{\boldsymbol{Y}|\boldsymbol{X}} \\ H_{\boldsymbol{X}} & H_{\boldsymbol{Y}} \end{pmatrix} \\
&= \alpha\delta^{\min}_{\boldsymbol{X},\boldsymbol{Y}} + (1-\alpha)d^{I}_{\boldsymbol{X},\boldsymbol{Y}}.
\end{aligned}$$

Clearly, the previous representation of $\Delta^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}}$ (respectively, $\delta^{D,\alpha}_{\boldsymbol{X},\boldsymbol{Y}}$) as a convex combination of $\Delta^{\min}_{\boldsymbol{X},\boldsymbol{Y}}$ and $D^{I}_{\boldsymbol{X},\boldsymbol{Y}}$ (respectively, $\delta^{\min}_{\boldsymbol{X},\boldsymbol{Y}}$ and $d^{I}_{\boldsymbol{X},\boldsymbol{Y}}$) introduces a degree of freedom that could be useful for practical purposes in the prediction framework, where $\boldsymbol{Y}$ could represent some class variable. According to the parameter α, one may choose between one or two prediction terms, $H_{\boldsymbol{X}|\boldsymbol{Y}}$ and $H_{\boldsymbol{Y}|\boldsymbol{X}}$ (respectively, $\begin{smallmatrix} H_{\boldsymbol{X}|\boldsymbol{Y}} \\ H_{\boldsymbol{X}} \end{smallmatrix}$ and $\begin{smallmatrix} H_{\boldsymbol{Y}|\boldsymbol{X}} \\ H_{\boldsymbol{Y}} \end{smallmatrix}$). This possibility to introduce a nonuniform mixing of the entropic contributions in the expression of the complexity terms seems to be not feasible by a direct adaptation of $C^{I}_{\boldsymbol{X},\boldsymbol{Y}}$.

Remark 1. By choosing $g(\cdot) = (\cdot)^{\gamma}$ for some $\gamma > 0$, the complexity term is given by $C^{\gamma,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} = \big\|(\alpha^{\frac{1}{\gamma}}m, (1-\alpha)^{\frac{1}{\gamma}}M)\big\|_{\gamma}$, where $\|\boldsymbol{x}\|_{\gamma} = \Big(\sum\limits_{i=1}^{2}|x_i|^{\gamma}\Big)^{1/\gamma}$ denotes the norm of some vector $\boldsymbol{x}$ of length 2. Note that for any $0 \le \alpha \le 1$, we have

$$(\alpha^{\wedge})^{\frac{1}{\gamma}}\|(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})\|_{\gamma} \le C^{\gamma,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \le (\alpha^{\vee})^{\frac{1}{\gamma}}\|(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})\|_{\gamma},$$

with $\alpha^{\wedge} = \min(\alpha, 1-\alpha)$ and $\alpha^{\vee} = \max(\alpha, 1-\alpha)$. When γ goes to infinity, $C^{\gamma,\alpha}_{\boldsymbol{X},\boldsymbol{Y}}$ converges towards $C^{I}_{\boldsymbol{X},\boldsymbol{Y}}$.

Remark 2. The complexity term C^{α} is invariant under linear transformations of g. In particular, g and $-g$ provide the same complexity term. Consequently, without loss of generality we could restrict g to be an increasing function.

Let us now propose a result to arrange these different examples considered in this paper. First, some preliminary result is given.

Lemma 1. *Let $C^{(1)}$ and $C^{(2)}$ be two complexity terms of the form* (17) *with functions g_1 and g_2. Assume that either the function $g_1 \circ g_2^{-1}$ is concave or $g_2 \circ g_1^{-1}$ is convex. Then $C^{(1)}_{\boldsymbol{X},\boldsymbol{Y}} \le C^{(2)}_{\boldsymbol{X},\boldsymbol{Y}}$.*

Proof. By rewriting $g_1 = (g_1 \circ g_2^{-1}) \circ g_2$ when $g_1 \circ g_2^{-1}$ is concave and $g_1^{-1} = g_2^{-1} \circ (g_2 \circ g_1^{-1})$ when $(g_2 \circ g_1^{-1})$ is convex, one gets

$$g_1^{-1}(\alpha g_1(m) + (1-\alpha) g_1(M)) \leq \begin{cases} g_2^{-1}(\alpha (g_2 \circ g_1^{-1}) \circ g_1(m) + (1-\alpha)(g_2 \circ g_1^{-1}) \circ g_1(M)) \\ g_1^{-1}(g_1 \circ g_2^{-1}(\alpha g_2(m) + (1-\alpha) g_2(M))) \end{cases}$$
$$\leq g_2^{-1}(\alpha g_2(m) + (1-\alpha) g_2(M)),$$

where $m = \min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ and $M = \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$. $\triangle$

Proposition 3. *For any I-divergences $\Delta^{(1)}$ or $\Delta^{(2)}$ or any NI-divergences $\delta^{(1)}$ or $\delta^{(2)}$ with respective complexity terms $C^{(1)}$ and $C^{(2)}$, we have the following equivalence*:

$$\Delta^{(1)}_{\boldsymbol{X},\boldsymbol{Y}} \leq \Delta^{(2)}_{\boldsymbol{X},\boldsymbol{Y}} \iff \delta^{(1)}_{\boldsymbol{X},\boldsymbol{Y}} \leq \delta^{(2)}_{\boldsymbol{X},\boldsymbol{Y}} \iff C^{(1)}_{\boldsymbol{X},\boldsymbol{Y}} \leq C^{(2)}_{\boldsymbol{X},\boldsymbol{Y}}. \tag{22}$$

Since for any $0 \leq \alpha \leq \alpha' \leq 1$ we have

$$C^{\alpha'}_{\boldsymbol{X},\boldsymbol{Y}} \leq C^{\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \leq C^{I}_{\boldsymbol{X},\boldsymbol{Y}}, \tag{23}$$

the associated I-divergences and NI-divergences are ordered according to equation (22). *Furthermore, a similar result holds for the main examples of this paper since*

$$C^{D,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \leq C^{P,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \leq C^{R,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \leq C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \leq C^{I}_{\boldsymbol{X},\boldsymbol{Y}} \leq C^{E}_{\boldsymbol{X},\boldsymbol{Y}}. \tag{24}$$

Proof. Equation (22) is direct. The left-hand side of (23) comes from

$$C^{\alpha}_{\boldsymbol{X},\boldsymbol{Y}} = g^{-1}(\alpha g(\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})) + (1-\alpha) g(\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})))$$
$$\leq g^{-1}(g(\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}))) = C^{I}_{\boldsymbol{X},\boldsymbol{Y}},$$

and the right-hand side is direct. Since $g^P \circ (g^D)^{-1}(\cdot) = -\log(\cdot)$, $g^R \circ (g^P)^{-1}(\cdot) = \exp\left(\frac{1}{2}\cdot\right)$, and $g^S \circ (g^R)^{-1}(\cdot) = (\cdot)^2$ are convex functions, (24) is a direct consequence of Lemma 1. $\triangle$

Remark 3. By assuming that either $g(\cdot)$ is a concave function or $g^{-1}(\cdot)$ is a convex function, we have the inequality

$$C^{\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \leq \alpha m + (1-\alpha) M = C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}},$$

which means that any Δ^{α} (respectively, δ^{α}) (satisfying the previous assumption) is upper bounded by $\Delta^{S,\alpha}$ (respectively, $\delta^{S,\alpha}$).

The following proposition gives a larger class of examples of I-divergences and NI-divergences.

Proposition 4. *Let $(\alpha^{(j)})_{j=1,\ldots,J}$ be some vector of probability weights for some $J \geq 1$.*

(i) *Let $\delta^{(1)}, \ldots, \delta^{(J)}$ be NI-divergences. Then the measure defined by*

$$\delta_{\boldsymbol{X},\boldsymbol{Y}} = \sum_{j=1}^{J} \alpha^{(j)} \delta^{(j)}_{\boldsymbol{X},\boldsymbol{Y}} \tag{25}$$

is an NI-divergence with complexity term given by

$$C_{\boldsymbol{X},\boldsymbol{Y}} = \left(\sum_{j=1}^{J} \frac{\alpha^{(j)}}{C^{(j)}_{\boldsymbol{X},\boldsymbol{Y}}} \right)^{-1}. \tag{26}$$

(ii) *Let $\Delta^{(1)}, \ldots, \Delta^{(j)}$ be I-divergences and $\delta^{(1)}, \ldots, \delta^{(j)}$ be NI-divergences with complexity terms $C^{(1)}_{\boldsymbol{X},\boldsymbol{Y}}, \ldots, C^{(J)}_{\boldsymbol{X},\boldsymbol{Y}}$. Then the measures defined by*

$$\Delta_{\boldsymbol{X},\boldsymbol{Y}} = C_{\boldsymbol{X},\boldsymbol{Y}} - I_{\boldsymbol{X},\boldsymbol{Y}} \qquad \text{and} \qquad \delta_{\boldsymbol{X},\boldsymbol{Y}} = 1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{C_{\boldsymbol{X},\boldsymbol{Y}}}, \quad \text{with} \quad C_{\boldsymbol{X},\boldsymbol{Y}} = \sum_{j=1}^{J} \alpha^{(j)} C^{(j)}_{\boldsymbol{X},\boldsymbol{Y}}, \tag{27}$$

are also, respectively, an I-divergence and NI-divergence.

The proof is immediate.

3.2. Around Property [**P3**]

The fact that an I-divergence Δ (respectively, NI-divergence δ) satisfies property [**P3**] may be expressed as follows: $\Delta_{\boldsymbol{X},\boldsymbol{Y}} = 0 \Leftrightarrow D^I_{\boldsymbol{X},\boldsymbol{Y}} = 0$ (respectively, $\delta_{\boldsymbol{X},\boldsymbol{Y}} = 0 \Leftrightarrow d^I_{\boldsymbol{X},\boldsymbol{Y}} = 0$). In fact, [**P3**] should be extended to a more useful assumption: $\Delta_{\boldsymbol{X},\boldsymbol{Y}}$ (or $\delta_{\boldsymbol{X},\boldsymbol{Y}}$) is near from the minimum 0 if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ share almost the same information. This may be translated by the following implications related to an I-divergence Δ (respectively, NI-divergence δ):

- For all $\gamma > 0$ there exists $\varepsilon > 0$ such that for all $(\boldsymbol{X}, \boldsymbol{Y}) \in \Upsilon$

$$\Delta_{\boldsymbol{X},\boldsymbol{Y}} \le \varepsilon \implies D^I_{\boldsymbol{X},\boldsymbol{Y}} \le \gamma \qquad (\text{respectively,} \quad \delta_{\boldsymbol{X},\boldsymbol{Y}} \le \varepsilon \implies d^I_{\boldsymbol{X},\boldsymbol{Y}} \le \gamma);$$

- For all $\varepsilon > 0$ there exists $\gamma > 0$ such that for all $(\boldsymbol{X}, \boldsymbol{Y}) \in \Upsilon$

$$D^I_{\boldsymbol{X},\boldsymbol{Y}} \le \gamma \implies \Delta_{\boldsymbol{X},\boldsymbol{Y}} \le \varepsilon \qquad (\text{respectively,} \quad d^I_{\boldsymbol{X},\boldsymbol{Y}} \le \gamma \implies \delta_{\boldsymbol{X},\boldsymbol{Y}} \le \varepsilon).$$

An I-divergence Δ (respectively, NI-divergence δ) inherits the previous property if it satisfies:

[P3′(Υ, k_1, k_2)] There exist some positive constants k_1 and k_2 ($k_1 \le k_2$) such that for all pairs $(\boldsymbol{X}, \boldsymbol{Y}) \in \Upsilon \subset \Gamma^2$ we have

$$k_1 D^I_{\boldsymbol{X},\boldsymbol{Y}} \le \Delta_{\boldsymbol{X},\boldsymbol{Y}} \le k_2 D^I_{\boldsymbol{X},\boldsymbol{Y}} \qquad (\text{respectively,} \quad k_1 d^I_{\boldsymbol{X},\boldsymbol{Y}} \le \delta_{\boldsymbol{X},\boldsymbol{Y}} \le k_2 d^I_{\boldsymbol{X},\boldsymbol{Y}}). \tag{28}$$

Among our examples, we assert that D^E and d^E both satisfy [**P3′($\Gamma^2, 1, 2$)**], that is,

$$D^I_{\boldsymbol{X},\boldsymbol{Y}} \le D^E_{\boldsymbol{X},\boldsymbol{Y}} \le 2 D^I_{\boldsymbol{X},\boldsymbol{Y}} \qquad (\text{respectively,} \quad d^I_{\boldsymbol{X},\boldsymbol{Y}} \le d^E_{\boldsymbol{X},\boldsymbol{Y}} \le 2 d^I_{\boldsymbol{X},\boldsymbol{Y}}).$$

Most of complexity terms considered in this paper are of the particular form (17) where the function $g(\cdot)$ is a monotone function on $\mathbb{R}^+$. From (23), we can point out that for such complexity terms (expressed in terms of Δ or δ), the constant k_2 is equal to 1. Moreover, we assert that if Δ satisfies [**P3′($\Upsilon, k_1, 1$)**], then the associated δ also satisfies [**P3′($\Upsilon, k_1, 1$)**] since

$$k_1 d^I_{\boldsymbol{X},\boldsymbol{Y}} = \frac{k_1 D^I_{\boldsymbol{X},\boldsymbol{Y}}}{C^I_{\boldsymbol{X},\boldsymbol{Y}}} \le \frac{\Delta_{\boldsymbol{X},\boldsymbol{Y}}}{C_{\boldsymbol{X},\boldsymbol{Y}}} = \delta_{\boldsymbol{X},\boldsymbol{Y}}.$$

Therefore, the results presented hereafter in the rest of this section will be expressed for I-divergences only.

Furthermore, we now consider only complexity terms of the form (17) defined through a function $g(\cdot)$ continuously differentiable on some set $\mathcal{D}^g \subset \mathbb{R}^+$. Let us first introduce the following two subsets of $\mathcal{D}^g$:

$$\mathcal{E}_1^g = \{\Theta \subset \mathcal{D}^g : \ 0 < \varkappa^g_{\inf,\Theta} < \varkappa^g_{\sup,\Theta} < +\infty\} \qquad \text{and} \qquad \mathcal{E}_2^{g,\alpha} = \left\{\Theta \subset \mathcal{E}_1^g : \ \frac{\alpha \varkappa^g_{\sup,\Theta}}{\varkappa^g_{\inf,\Theta}} < 1\right\},$$

with $\varkappa^g_{\inf,\Theta} = \inf\limits_{x\in\Theta} |g'(x)|$ and $\varkappa^g_{\sup,\Theta} = \sup\limits_{x\in\Theta} |g'(x)|$. Denote also $\alpha^\wedge = \min(\alpha, 1-\alpha)$.

In the following, two results ensuring that an I-divergence Δ^α of the form (17) satisfies property [**P3′(Υ, k_1, k_2)**] are proposed. The difference relies upon the framework: the constants k_1 and k_2 differ whenever the set Υ differs.

Proposition 5. *For any $\Theta \in \mathcal{E}_1^g$, the I-divergence Δ^α satisfies* [**P3′($\Upsilon_\Theta, \alpha^\wedge \frac{\varkappa^g_{\inf,\Theta}}{\varkappa^g_{\sup,\Theta}}, 1$)**] *with* $\Upsilon_\Theta = \{(\boldsymbol{X}, \boldsymbol{Y}) \in \Gamma^2 : \ H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}, I_{\boldsymbol{X},\boldsymbol{Y}} \in \Theta\}$.

Proof. Denote $x = \min(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}})$, $y = \max(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}})$, and $z = I_{\boldsymbol{X},\boldsymbol{Y}}$. There exist c_1, c_2, and c_3 such that

$$\begin{aligned} g^{-1}\big(\alpha g(x+z) + (1-\alpha)g(y+z)\big) - z & \\ &= (\alpha(g(x+z) - g(z)) + (1-\alpha)(g(y+z) - g(z)))(g^{-1})'(c_1) \\ &= \alpha|g'(c_2)||(g^{-1})'(c_1)|x + (1-\alpha)|g'(c_3)||(g^{-1})'(c_1)|y, \end{aligned}$$

with $c_1 \in \big[\min\big(g(z), \alpha g(x+z)+(1-\alpha)g(y+z)\big), \max\big(g(z), \alpha g(x+z)+(1-\alpha)g(y+z)\big)\big]$, $c_2 \in [z, x+z]$, and $c_3 \in [z, y+z]$. Then for all x, y, and z we obtain

$$g^{-1}\big(\alpha g(x+z) + (1-\alpha)g(y+z)\big) - z \geq \alpha^{\wedge \varkappa^{g}_{\inf,\Theta}}_{\varkappa^{g}_{\sup,\Theta}} \max(x, y),$$

which means that $\alpha^{\wedge \varkappa^{g}_{\inf,\Theta}}_{\varkappa^{g}_{\sup,\Theta}} D^I_{\boldsymbol{X},\boldsymbol{Y}} \leq \Delta^{\alpha}_{\boldsymbol{X},\boldsymbol{Y}}$. $\triangle$

Proposition 6. *For any $\Theta \in \mathcal{E}_2^g$, the I-divergence Δ^α satisfies* $\mathbf{[P3'(\Gamma^2_\Theta, 1-\alpha^{\wedge \varkappa^{g}_{\sup,\Theta}}_{\varkappa^{g}_{\inf,\Theta}}, 1)]}$ *with* $\Gamma_\Theta = \{\boldsymbol{Z} \in \Gamma : H_{\boldsymbol{Z}} \in \Theta\}$.

Proof. We have

$$\begin{aligned} D^I_{\boldsymbol{X},\boldsymbol{Y}} - \Delta^{\alpha}_{\boldsymbol{X},\boldsymbol{Y}} = C^I_{\boldsymbol{X},\boldsymbol{Y}} - C_{\boldsymbol{X},\boldsymbol{Y}} &= \alpha(g^{-1})'(c_1)(g(\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})) - g(\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}))) \\ &= \alpha|(g^{-1})'(c_1)||g'(c_2)||H_{\boldsymbol{X}} - H_{\boldsymbol{Y}}|, \end{aligned}$$

with $c_1 \in [g(\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})), g(\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}))]$ and $c_2 \in [\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}), \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})]$. Then we obtain

$$D^I_{\boldsymbol{X},\boldsymbol{Y}} - \Delta^{\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \leq \alpha^{\varkappa^{g}_{\sup,\Theta}}_{\varkappa^{g}_{\inf,\Theta}} D^I_{\boldsymbol{X},\boldsymbol{Y}},$$

which leads to the result. $\triangle$

For the sake of simplicity, we use the notation $\varkappa^{\bullet}_{\inf,\Theta}$ and $\varkappa^{\bullet}_{\sup,\Theta}$ instead of $\varkappa^{g^\bullet}_{\inf,\Theta}$ and $\varkappa^{g^\bullet}_{\sup,\Theta}$.

The following result is devoted to our different examples. We apply the previous two propositions and present a new result obtained by taking into account the specific form of each example.

Proposition 7. *The I-divergence $\Delta^{\bullet,\alpha}$ satisfies properties* $\mathbf{[P3'(\Upsilon_\Theta, k_1^{a,\bullet}, 1)]}$ *(from Proposition* 5*),* $\mathbf{[P3'(\Gamma^2_\Theta, k_1^{b,\bullet}, 1)]}$ *(from Proposition* 6*), and* $\mathbf{[P3'(\Gamma^2_\Theta, k_1^{c,\bullet}, 1)]}$*, where $\bullet$ stands for S, R, P, and D, and where*

$\bullet$	Θ	$\varkappa^{\bullet}_{\inf,\Theta}$	$\varkappa^{\bullet}_{\sup,\Theta}$	$k_1^{a,\bullet} = \alpha^{\wedge \varkappa_{\inf,\Theta}}_{\varkappa^{\bullet}_{\sup,\Theta}}$	$k_1^{b,\bullet} = 1 - \alpha^{\varkappa_{\sup,\Theta}}_{\varkappa^{\bullet}_{\inf,\Theta}}$	$k_1^{c,\bullet}$
S	$\mathbb{R}^+$	1	1	$\alpha^\wedge$	$1-\alpha$	
R	$[c_1, c_2]$	$\frac{1}{2\sqrt{c_2}}$	$\frac{1}{2\sqrt{c_1}}$	$\alpha^{\wedge \sqrt{\rho}}$	$1 - \alpha\sqrt{\rho}$ $(if\ \rho < \frac{1}{\alpha^2})$	$(1-\alpha)\Big(1 - \frac{\alpha}{(1+\frac{1}{\sqrt{\rho}})^2}\Big)$
R	$\mathbb{R}^+$					$(1-\alpha)^2$
P	$[c_1, c_2]$	$\frac{1}{c_2}$	$\frac{1}{c_1}$	$\alpha^{\wedge \rho}$	$1 - \alpha\rho$ $(if\ \rho < \frac{1}{\alpha})$	$\frac{\rho^{1-\alpha} - 1}{\rho - 1}$
D	$[c_1, c_2]$	$\frac{1}{c_2^2}$	$\frac{1}{c_1^2}$	$\alpha^{\wedge \rho^2}$	$1 - \alpha\rho^2$ $(if\ \rho < \frac{1}{\sqrt{\alpha}})$	$\frac{1}{1 + \frac{\alpha}{1-\alpha}\rho}$

with $0 < c_1 \leq c_2 < +\infty$ and $\rho = \frac{c_2}{c_1}$.

Proof. The computation of $k_1^{a,\bullet}$ and $k_1^{b,\bullet}$ follows from Propositions 5 and 6. Hence, let us only concentrate on $k_1^{c,\bullet}$ for the complexity terms $C^{R,\alpha}$, $C^{P,\alpha}$, and $C^{D,\alpha}$. We denote $m = \min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ and $M = \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$.

- Complexity term $C^{R,\alpha}$:

$$\begin{aligned} D^I_{\boldsymbol{X},\boldsymbol{Y}} - \Delta^{R,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} &= \alpha(1-\alpha)(\sqrt{M}-\sqrt{m})^2 + \alpha(M-m) \\ &= \alpha(1-\alpha)\frac{(M-m)^2}{(\sqrt{M}+\sqrt{m})^2} + \alpha(M-m) \\ &\le \alpha(1-\alpha)\frac{(D^I_{\boldsymbol{X},\boldsymbol{Y}})^2}{(\sqrt{M}+\sqrt{m})^2} + \alpha D^I_{\boldsymbol{X},\boldsymbol{Y}}. \end{aligned}$$

  Thus,

$$\Delta^{R,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \ge (1-\alpha) D^I_{\boldsymbol{X},\boldsymbol{Y}} \left(1 - \alpha \frac{D^I_{\boldsymbol{X},\boldsymbol{Y}}}{(\sqrt{M}+\sqrt{m})^2}\right).$$

  The result is obtained by noting that

$$\frac{D^I_{\boldsymbol{X},\boldsymbol{Y}}}{(\sqrt{m}+\sqrt{M})^2} \le \frac{M}{(\sqrt{m}+\sqrt{M})^2} = \frac{1}{\left(1+\sqrt{\frac{m}{M}}\right)^2} \le \frac{1}{\left(1+\sqrt{\frac{c_1}{c_2}}\right)^2} \le 1.$$

- Complexity term $C^{P,\alpha}$: by using the Taylor expansion with integral rest, one obtains

$$\begin{aligned} D^I_{\boldsymbol{X},\boldsymbol{Y}} - \Delta^{P,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} &= M^{1-\alpha}(M^\alpha - m^\alpha) \\ &= M^{1-\alpha}(M-m)\int_0^1 \frac{\alpha}{(m+t(M-m))^{1-\alpha}}\,dt \\ &\le (M-m)\int_0^1 \frac{\alpha}{\left(\frac{1}{\rho}+t\left(1-\frac{1}{\rho}\right)\right)^{1-\alpha}}\,dt \\ &\le D^I_{\boldsymbol{X},\boldsymbol{Y}} \frac{1}{1-\frac{1}{\rho}}\left[\left(\frac{1}{\rho}+t\left(1-\frac{1}{\rho}\right)\right)^{\alpha}\right]_0^1 = D^I_{\boldsymbol{X},\boldsymbol{Y}} \frac{1-\left(\frac{1}{\rho}\right)^{\alpha}}{1-\frac{1}{\rho}}. \end{aligned}$$

  Thus,

$$\Delta^{P,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \ge D^I_{\boldsymbol{X},\boldsymbol{Y}} \left(1 - \frac{1-\left(\frac{1}{\rho}\right)^{\alpha}}{1-\frac{1}{\rho}}\right) = D^I_{\boldsymbol{X},\boldsymbol{Y}} \frac{\rho^{1-\alpha}-1}{\rho-1}.$$

- Complexity term $C^{D,\alpha}$:

$$D^I_{\boldsymbol{X},\boldsymbol{Y}} - \Delta^{D,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} = M - \frac{mM}{\alpha M + (1-\alpha)m} = \frac{\alpha M}{\alpha M + (1-\alpha)m}(M-m) \le \frac{1}{1+\frac{1-\alpha}{\alpha}\frac{c_1}{c_2}} D^I_{\boldsymbol{X},\boldsymbol{Y}}. \quad \triangle$$

3.3. Around the Triangle Inequality

The question arises now whether an I-divergence or NI-divergence satisfies property **[P6]**, i.e., the triangle inequality. The following proposition establishes sufficient conditions for such measures to constitute a metric.

Lemma 2. *We have*

$$H_{\boldsymbol{X},\boldsymbol{Y}} \le H_{\boldsymbol{X},\boldsymbol{Z}} + H_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}}, \tag{29}$$

$$I_{\boldsymbol{X},\boldsymbol{Y}} \ge I_{\boldsymbol{X},\boldsymbol{Z}} + I_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}}. \tag{30}$$

Proof. From the general properties on entropy, one can obtain

$$H_{\boldsymbol{X},\boldsymbol{Y}} \le H_{\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}} = H_{\boldsymbol{X},\boldsymbol{Z}} + H_{\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{Z}} \le H_{\boldsymbol{X},\boldsymbol{Z}} + H_{\boldsymbol{Y}|\boldsymbol{Z}} = H_{\boldsymbol{X},\boldsymbol{Z}} + H_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}}. \tag{31}$$

Equation (30) directly follows from (2). △

Proposition 8. *Assume that the complexity term defining an I-divergence satisfies the following property*:

$$C_{\boldsymbol{X},\boldsymbol{Y}} \le C_{\boldsymbol{X},\boldsymbol{Z}} + C_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}}. \tag{32}$$

Then the associated I-divergence satisfies the triangle inequality, that is,

$$\Delta_{\boldsymbol{X},\boldsymbol{Y}} \le \Delta_{\boldsymbol{X},\boldsymbol{Z}} + \Delta_{\boldsymbol{Y},\boldsymbol{Z}}. \tag{33}$$

In addition, if C satisfies

$$C_{\boldsymbol{X},\boldsymbol{Z}} \ge \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Z}}), \tag{34}$$

then the associated NI-divergence also satisfies the triangle inequality, that is,

$$\delta_{\boldsymbol{X},\boldsymbol{Y}} \le \delta_{\boldsymbol{X},\boldsymbol{Z}} + \delta_{\boldsymbol{Y},\boldsymbol{Z}}. \tag{35}$$

Proof. Since the quantity

$$A = -(C_{\boldsymbol{X},\boldsymbol{Y}} - I_{\boldsymbol{X},\boldsymbol{Y}}) + (C_{\boldsymbol{X},\boldsymbol{Z}} - I_{\boldsymbol{X},\boldsymbol{Z}}) + (C_{\boldsymbol{Y},\boldsymbol{Z}} - I_{\boldsymbol{Y},\boldsymbol{Z}}),$$

is nonnegative from (30) and (32), equation (33) immediately follows. Moreover, we have the following equation:

$$\delta_{\boldsymbol{X},\boldsymbol{Y}} \le 1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{C_{\boldsymbol{X},\boldsymbol{Y}} + A}. \tag{36}$$

Now it is also easy to see from (34) that

$$A + C_{\boldsymbol{X},\boldsymbol{Y}} \ge C_{\boldsymbol{X},\boldsymbol{Z}} + C_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} \ge \max(C_{\boldsymbol{X},\boldsymbol{Z}}, C_{\boldsymbol{Y},\boldsymbol{Z}}).$$

From (36) it follows that

$$\delta_{\boldsymbol{X},\boldsymbol{Y}} \le \frac{C_{\boldsymbol{X},\boldsymbol{Z}} - I_{\boldsymbol{X},\boldsymbol{Z}} + C_{\boldsymbol{Y},\boldsymbol{Z}} - I_{\boldsymbol{Y},\boldsymbol{Z}}}{\max(C_{\boldsymbol{X},\boldsymbol{Z}}, C_{\boldsymbol{Y},\boldsymbol{Z}})} \le \frac{C_{\boldsymbol{X},\boldsymbol{Z}} - I_{\boldsymbol{X},\boldsymbol{Z}}}{C_{\boldsymbol{X},\boldsymbol{Z}}} + \frac{C_{\boldsymbol{Y},\boldsymbol{Z}} - I_{\boldsymbol{Y},\boldsymbol{Z}}}{C_{\boldsymbol{Y},\boldsymbol{Z}}} = \delta_{\boldsymbol{X},\boldsymbol{Z}} + \delta_{\boldsymbol{Y},\boldsymbol{Z}}. \quad \triangle$$

Remark 4. In Proposition 8 there is no implication between (32) and (34). Indeed, one may check that the NI-divergence δ^S (with $\alpha = 1/2$ for example) satisfies the first inequality but not the second one. Now consider an NI-divergence with complexity term $C_{\boldsymbol{X},\boldsymbol{Y}} = \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}) + H_{\boldsymbol{X}|\boldsymbol{Y}} H_{\boldsymbol{Y}|\boldsymbol{X}}$. By choosing $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ such that $H_{\boldsymbol{X}|\boldsymbol{Z}} = H_{\boldsymbol{Y}|\boldsymbol{Z}} = 0$ and $H_{\boldsymbol{X}|\boldsymbol{Y}} = H_{\boldsymbol{Y}|\boldsymbol{X}} = I_{\boldsymbol{X},\boldsymbol{Y}} = H_{\boldsymbol{Z}}/3 = H_{\boldsymbol{X},\boldsymbol{Y}}/3 > 2$, one sees that (34) is satisfied but (32) is not.

Remark 5. Let us consider an NI-divergence δ with the complexity term

$$C_{\boldsymbol{X},\boldsymbol{Y}} = C'_{\boldsymbol{X},\boldsymbol{Y}} + \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$$

such that $C'_{\boldsymbol{X},\boldsymbol{Y}} \ge 0$ (necessarily $C'_{\boldsymbol{X},\boldsymbol{Y}} = 0$ whenever $\boldsymbol{X} \sim \boldsymbol{Y}$). Then, Δ and δ satisfy the triangle inequality if C' also satisfies the triangle inequality. However, this is not a necessary condition. Indeed, the triangle inequality is not satisfied for the same example of the previous remark with $C'_{\boldsymbol{X},\boldsymbol{Y}} = H_{\boldsymbol{X}|\boldsymbol{Y}} H_{\boldsymbol{Y}|\boldsymbol{X}}$, for which $C'_{\boldsymbol{X},\boldsymbol{Z}} = C'_{\boldsymbol{Y},\boldsymbol{Z}} = 0$, whereas $C'_{\boldsymbol{X},\boldsymbol{Y}} > 0$.

Let us now propose some examples and consequences through the following corollary.

Corollary 1. (i) *The measures D^E and D^I satisfy condition* (32) *and so are metrics.*

(ii) *The measures d^E and d^I satisfy conditions* (32) *and* (34) *and so are metrics.*

(iii) *The measure $D^{S,\alpha}$ for $\alpha \le \frac{1}{2}$ satisfies condition* (32) *and so is a metric. Moreover, when $\alpha > \frac{1}{2}$, this measure does not satisfy* (32).

(iv) *Let $(\alpha^{(j)})_{j=1,\dots,J}$ be some vector of probability weights for some $J \ge 1$. Let $\Delta^{(1)}, \dots, \Delta^{(J)}$ be I-divergences (respectively, $\delta^{(1)}, \dots, \delta^{(J)}$ be NI-divergences) with complexity terms $C^{(1)}_{\boldsymbol{X},\boldsymbol{Y}}, \dots, C^{(J)}_{\boldsymbol{X},\boldsymbol{Y}}$ satisfying* (32) (*respectively,* (32) *and* (34)). *Then the measures defined by* (27) *satisfy the triangle inequality.*

Proof. (i), (ii) Equation (29) corresponds exactly to (32) for $C^E_{\boldsymbol{X},\boldsymbol{Y}} = H_{\boldsymbol{X},\boldsymbol{Y}}$. Since $H_{\boldsymbol{X},\boldsymbol{Z}} \ge \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Z}})$, we have proved that D^E and d^E are metrics. Concerning D^I and d^I, the complexity term corresponds to $C^I_{\boldsymbol{X},\boldsymbol{Y}} = \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$. Thus, it suffices to prove (32), which is quite obvious. Indeed,

$$\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Z}}) + \max(H_{\boldsymbol{Y}}, H_{\boldsymbol{Z}}) - H_{\boldsymbol{Z}} \ge \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}).$$

(iii) Let $m = \min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ and $M = \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$. We distinguish three cases.

- $H_{\boldsymbol{Z}} < m$:

$$C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + C^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} = (2\alpha - 1)H_{\boldsymbol{Z}} + (1-\alpha)(m+M).$$

If $\alpha > \frac{1}{2}$ and $H_{\boldsymbol{X}} = H_{\boldsymbol{Y}}$, the right-hand side of the previous equation equals $(1-2\alpha) \times (C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} - H_{\boldsymbol{Z}}) + C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} < C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}}$. Thus, (32) can never be satisfied for $\alpha > \frac{1}{2}$. Now, if $\alpha \le \frac{1}{2}$, we have

$$C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + C^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} > (1-\alpha)(m+M) \ge C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}}.$$

- $H_{\boldsymbol{Z}} > M$:

$$C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + C^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} = (2\alpha - 1)H_{\boldsymbol{Z}} + (1-\alpha)(m+M) \ge \alpha m + (1-\alpha)M = C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}}.$$

- $m \le H_{\boldsymbol{Z}} \le M$:

$$C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + C^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} = \alpha m + (1-\alpha)M = C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}}.$$

(iv) This is trivial. $\triangle$

We claim that the measures $\Delta^{R,\alpha}$, $\Delta^{P,\alpha}$, and $\Delta^{D,\alpha}$ (and so $\delta^{R,\alpha}$, $\delta^{P,\alpha}$, and $\delta^{D,\alpha}$) do not satisfy condition (32). Consider, for example, $\Delta^{D,\alpha}$. Let us choose $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ such that $H_{\boldsymbol{Z}} > \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ and $H_{\boldsymbol{Z}} = \frac{2-\alpha}{1-\alpha} H_{\boldsymbol{X}} = \frac{2-\alpha}{1-\alpha} H_{\boldsymbol{Y}}$. This leads to

$$C^{D,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + C^{D,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} = H_{\boldsymbol{Z}} \left(\frac{H_{\boldsymbol{X}}}{\alpha H_{\boldsymbol{X}} + (1-\alpha)H_{\boldsymbol{Z}}} + \frac{H_{\boldsymbol{Y}}}{\alpha H_{\boldsymbol{Y}} + (1-\alpha)H_{\boldsymbol{Z}}} - 1 \right) = 0 < C^{D,\alpha}_{\boldsymbol{X},\boldsymbol{Y}},$$

which contradicts (32).

Concerning these divergences and the measures $\Delta^{S,\alpha}$ (for $\alpha > \frac{1}{2}$) and $\delta^{S,\alpha}$, we do not know if they satisfy the triangle inequality. We can only say that our tool cannot be applied to prove it. We suggest to weaken property **[P6]** in the following way in order to obtain more results. Let an I-divergence or NI-divergence satisfy

[P6′(Υ, c)] There exists $c \ge 1$ such that for all $(\boldsymbol{X},\boldsymbol{Y}), (\boldsymbol{Y},\boldsymbol{Z}), (\boldsymbol{X},\boldsymbol{Z}) \in \Upsilon$

$$\Delta_{\boldsymbol{X},\boldsymbol{Y}} \le c(\Delta_{\boldsymbol{X},\boldsymbol{Z}} + \Delta_{\boldsymbol{Y},\boldsymbol{Z}}).$$

Property [**P6**] is then equivalent to [$\mathbf{P6'(\Gamma^2, 1)}$], and we already know that D^E, d^E, D^I, d^I, and $D^{S,\alpha}$ (for $\alpha \leq \frac{1}{2}$) satisfy [$\mathbf{P6'(\Gamma^2, 1)}$]. When $\Upsilon \subsetneq \Gamma^2$, property [**P6′**] is in a sense local, whereas it is global (as the classical triangle inequality) when $\Upsilon = \Gamma^2$.

Note that if an I-divergence (or NI-divergence) satisfies [$\mathbf{P3'(\Upsilon, k_1, k_2)}$], then [$\mathbf{P6'(\Upsilon, \frac{k_2}{k_1})}$] is satisfied since

$$\Delta_{\boldsymbol{X},\boldsymbol{Y}} \leq k_2 D^I_{\boldsymbol{X},\boldsymbol{Y}} \leq k_2 (D^I_{\boldsymbol{X},\boldsymbol{Z}} + D^I_{\boldsymbol{Y},\boldsymbol{Z}}) \leq \frac{k_2}{k_1} (\Delta_{\boldsymbol{X},\boldsymbol{Z}} + \Delta_{\boldsymbol{Y},\boldsymbol{Z}}).$$

We then inherit a lot of results from Proposition 7 related to our examples. In particular, $\Delta^{\bullet,\alpha}$ and $\delta^{\bullet,\alpha}$ (where $\bullet$ stands for S, R, P, and D) both satisfy [$\mathbf{P6'(\Upsilon_\Theta, \frac{1}{k_1^a})}$], [$\mathbf{P6'(\Gamma^2_\Theta, \frac{1}{k_1^b})}$], and [$\mathbf{P6'(\Gamma^2_\Theta, \frac{1}{k_1^c})}$].

In the rest of this section, we attempt to ensure the global property [$\mathbf{P6'(\Gamma^2, c)}$]. From Proposition 7 (with $\Theta = \mathbb{R}^+$) we see that the divergences $\Delta^{S,\alpha}$ (when $\alpha > \frac{1}{2}$) and $\delta^{S,\alpha}$ (respectively, $\Delta^{R,\alpha}$ and $\delta^{R,\alpha}$) satisfy [$\mathbf{P6'(\Gamma^2, \frac{1}{1-\alpha})}$] (respectively, [$\mathbf{P6'(\Gamma^2, \frac{1}{(1-\alpha)^2})}$]).

When $\alpha \leq \frac{1}{2}$, we could improve the previous result on $\Delta^{R,\alpha}$ by proving that it satisfies property [$\mathbf{P6'(\Gamma^2, \frac{1}{\alpha^2 + (1-\alpha)^2})}$], in the same spirit as in the proof leading to [**P3′**]. Indeed,

$$\begin{aligned} D^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} - \Delta^{R,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} &= \alpha(1-\alpha)(m + M - 2\sqrt{mM} \\ &\leq 2\alpha(1-\alpha)(D^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} + I_{\boldsymbol{X},\boldsymbol{Y}} - \sqrt{mM}) \\ &\leq 2\alpha(1-\alpha) D^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}}, \end{aligned}$$

which leads to $\Delta^{R,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \geq (\alpha^2 + (1-\alpha)^2) D^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}}$. Finally, let us note that

$$\Delta^{R,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \leq D^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \leq D^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + D^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} \leq \frac{1}{\alpha^2 + (1-\alpha)^2} (\Delta^{R,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + \Delta^{R,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}}).$$

Now we give a more general result, allowing us, in particular, to improve [$\mathbf{P6'(\Gamma^2, \frac{1}{1-\alpha})}$] for $\Delta^{S,\alpha}$ when $\alpha > \frac{1}{2}$.

Proposition 9. *Let us consider the following assumptions on a complexity term*: *there exists a constant* $c \geq 1$ *such that*

$$cC_{\boldsymbol{X},\boldsymbol{Z}} + cC_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} - (c-1)(I_{\boldsymbol{X},\boldsymbol{Z}} + I_{\boldsymbol{Y},\boldsymbol{Z}}) \geq C_{\boldsymbol{X},\boldsymbol{Y}}, \tag{37}$$

$$cC_{\boldsymbol{X},\boldsymbol{Z}} + cC_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} - (c-1)(I_{\boldsymbol{X},\boldsymbol{Z}} + I_{\boldsymbol{Y},\boldsymbol{Z}}) \geq \max(C_{\boldsymbol{X},\boldsymbol{Y}}, C_{\boldsymbol{X},\boldsymbol{Z}}, C_{\boldsymbol{Y},\boldsymbol{Z}}). \tag{38}$$

If an I*-divergence satisfies* (37) *or an* NI*-divergence satisfies* (38), *then they satisfy property* [$\mathbf{P6'(\Gamma^2, c)}$].

Proof. Let us introduce

$$A = -(C_{\boldsymbol{X},\boldsymbol{Y}} - I_{\boldsymbol{X},\boldsymbol{Y}}) + c(C_{\boldsymbol{X},\boldsymbol{Z}} - I_{\boldsymbol{X},\boldsymbol{Z}}) + c(C_{\boldsymbol{Y},\boldsymbol{Z}} - I_{\boldsymbol{Y},\boldsymbol{Z}}).$$

From (30) and (37), one may assert that

$$A \geq cC_{\boldsymbol{X},\boldsymbol{Z}} + cC_{\boldsymbol{Y},\boldsymbol{Z}} - C_{\boldsymbol{X},\boldsymbol{Y}} - H_{\boldsymbol{Z}} - (c-1)(I_{\boldsymbol{X},\boldsymbol{Z}} + I_{\boldsymbol{Y},\boldsymbol{Z}}) \geq 0,$$

which implies that the result is valid for Δ. Now from (38) one may write

$$A + C_{\boldsymbol{X},\boldsymbol{Y}} \geq \max(C_{\boldsymbol{X},\boldsymbol{Y}}, C_{\boldsymbol{X},\boldsymbol{Z}}, C_{\boldsymbol{Y},\boldsymbol{Z}}) \geq \max(C_{\boldsymbol{X},\boldsymbol{Z}}, C_{\boldsymbol{Y},\boldsymbol{Z}}),$$

which leads to

$$\delta_{\boldsymbol{X},\boldsymbol{Y}} \le \frac{c(C_{\boldsymbol{X},\boldsymbol{Z}} - I_{\boldsymbol{X},\boldsymbol{Z}}) + c(C_{\boldsymbol{Y},\boldsymbol{Z}} - I_{\boldsymbol{Y},\boldsymbol{Z}})}{\max(C_{\boldsymbol{X},\boldsymbol{Z}}, C_{\boldsymbol{Y},\boldsymbol{Z}})} \le c\delta_{\boldsymbol{X},\boldsymbol{Z}} + c\delta_{\boldsymbol{Y},\boldsymbol{Z}}. \quad \triangle$$

Corollary 2. *The measures* $\Delta^{S,\alpha}$ *for* $\alpha > \frac{1}{2}$ *satisfy* $[\mathbf{P6'(\Gamma^2, \frac{\alpha}{1-\alpha})}]$.

Proof. Let us concentrate on $\Delta^{S,\alpha}$ for $\alpha > \frac{1}{2}$. Let

$$A = cC^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + cC^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} - (c-1)(I_{\boldsymbol{X},\boldsymbol{Z}} + I_{\boldsymbol{Y},\boldsymbol{Z}}).$$

Without loss of generality, we assume that $H_{\boldsymbol{X}} \le H_{\boldsymbol{Y}}$. We distinguish three cases.

- $H_{\boldsymbol{Z}} \le H_{\boldsymbol{X}} \le H_{\boldsymbol{Y}}$: since $C^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} \ge I_{\boldsymbol{Y},\boldsymbol{Z}}$, we have

  $$A \ge c(1-\alpha)H_{\boldsymbol{X}} + (1-\alpha)H_{\boldsymbol{Y}} + (c\alpha + \alpha - 1)H_{\boldsymbol{Z}} - (c-1)I_{\boldsymbol{X},\boldsymbol{Z}}.$$

  Then

  $$A - C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \ge (c(1-\alpha) - \alpha)H_{\boldsymbol{X}} + (c\alpha + \alpha - 1)H_{\boldsymbol{Z}} - (c-1)I_{\boldsymbol{X},\boldsymbol{Z}} \ge (c-1)(H_{\boldsymbol{Z}} - I_{\boldsymbol{X},\boldsymbol{Z}}) \ge 0$$

  as soon as $c \ge \frac{\alpha}{1-\alpha}$.
- $H_{\boldsymbol{X}} \le H_{\boldsymbol{Y}} \le H_{\boldsymbol{Z}}$: we have

  $$A \ge \alpha H_{\boldsymbol{X}} + c\alpha H_{\boldsymbol{Y}} + ((1-\alpha) + c(1-\alpha) - 1)H_{\boldsymbol{Z}} - (c-1)I_{\boldsymbol{Y},\boldsymbol{Z}}.$$

  Then

  $$A - C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \ge (c\alpha - (1-\alpha))H_{\boldsymbol{Y}} + ((1-\alpha) + c(1-\alpha) - 1)H_{\boldsymbol{Z}} - (c-1)I_{\boldsymbol{Y},\boldsymbol{Z}} \ge (c-1)(H_{\boldsymbol{Y}} - I_{\boldsymbol{Y},\boldsymbol{Z}}) \ge 0$$

  as soon as $c \ge \frac{\alpha}{1-\alpha}$.
- $H_{\boldsymbol{X}} < H_{\boldsymbol{Z}} < H_{\boldsymbol{Y}}$: we have

  $$A \ge c\alpha H_{\boldsymbol{X}} + (1-\alpha)H_{\boldsymbol{Y}} + (c(1-\alpha)H_{\boldsymbol{Z}} + \alpha - 1) - (c-1)I_{\boldsymbol{X},\boldsymbol{Z}}.$$

  Then

  $$A - C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \ge (c-1)\alpha H_{\boldsymbol{X}} + (c-1)(1-\alpha)H_{\boldsymbol{Z}} - I_{\boldsymbol{X},\boldsymbol{Z}} \ge 0.$$

Hence, we obtain $A - C^{S,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} \ge 0$ for $c = \frac{\alpha}{1-\alpha}$. $\triangle$

Remark 6. The tool presented in Proposition 9 cannot be applied to the I-divergence $\Delta^{D,\alpha}$ and NI-divergence $\delta^{D,\alpha}$. Indeed, let us be given some $c \ge 1$ and consider the quantity

$$A = cC^{D,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + cC^{D,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} - (c-1)(I_{\boldsymbol{X},\boldsymbol{Z}} + I_{\boldsymbol{Y},\boldsymbol{Z}}).$$

In fact, one can always find $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ such that for all $c \ge 1$ the quantity A is negative. Indeed, let us choose $\boldsymbol{Z}$ independent of $\boldsymbol{X}$ and $\boldsymbol{Y}$ and such that $\alpha H_{\boldsymbol{Z}} + (1-\alpha)H_{\boldsymbol{X}} = 3cH_{\boldsymbol{X}}$ and $\alpha H_{\boldsymbol{Z}} + (1-\alpha)H_{\boldsymbol{Y}} = 3cH_{\boldsymbol{Y}}$. It is easy to see that $H_{\boldsymbol{Z}} \ge \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ and therefore $A = H_{\boldsymbol{Z}}\left(\frac{1}{3} + \frac{1}{3} - 1\right) < 0$. In the same manner, the tool is inapplicable to the I-divergence $\Delta^{P,\alpha}$ and NI-divergence $\delta^{P,\alpha}$. Indeed, let us take $\boldsymbol{Z}$ independent of $\boldsymbol{X}$ and $\boldsymbol{Y}$ and such that $H_{\boldsymbol{X}} = H_{\boldsymbol{Y}} = \left(\frac{1}{3c}\right)^{1/\alpha} H_{\boldsymbol{Z}}$; then $H_{\boldsymbol{Z}} \ge \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$ and

$$A = cC^{P,\alpha}_{\boldsymbol{X},\boldsymbol{Z}} + cC^{P,\alpha}_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} - (c-1)(I_{\boldsymbol{X},\boldsymbol{Z}} + I_{\boldsymbol{Y},\boldsymbol{Z}}) = -\frac{1}{3}H_{\boldsymbol{Z}} < 0.$$

The following result is an extension of Proposition 9, well suited to be applied to $\delta^{D,\alpha}$.

Proposition 10. *Let us assume that there exist two positive integers, I and Jsuch that an NI-divergence $\delta_{\boldsymbol{X},\boldsymbol{Y}}$ can be expressed as*

$$\delta_{\boldsymbol{X},\boldsymbol{Y}} = \sum_{i=1}^{I} \frac{S^{(i)}_{\boldsymbol{X},\boldsymbol{Y}}}{U^{(i)}_{\boldsymbol{X},\boldsymbol{Y}}} = \sum_{j=1}^{J} \alpha^{(j)} \left(1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{C^{(j)}_{\boldsymbol{X},\boldsymbol{Y}}}\right),$$

where $(\alpha^{(j)})_{j=1,\dots,J}$ is some vector of probability weights. Denote $S_{\boldsymbol{X},\boldsymbol{Y}} = \sum\limits_{i=1}^{I} S^{(i)}_{\boldsymbol{X},\boldsymbol{Y}}$ and $U_{\boldsymbol{X},\boldsymbol{Y}} = \max\limits_{i=1,\dots,I} U^{(i)}_{\boldsymbol{X},\boldsymbol{Y}}$. Then, if there exists some real number $c \geq 1$ such that for any $j = 1, \dots, J$ we have

(i) $A^{(j)} = I_{\boldsymbol{X},\boldsymbol{Y}} - C^{(j)}_{\boldsymbol{X},\boldsymbol{Y}} + c(S_{\boldsymbol{X},\boldsymbol{Z}} + S_{\boldsymbol{Z},\boldsymbol{Y}}) \geq 0$,
(ii) $A^{(j)} + C^{(j)}_{\boldsymbol{X},\boldsymbol{Y}} \geq \max(U_{\boldsymbol{X},\boldsymbol{Z}}, U_{\boldsymbol{Z},\boldsymbol{Y}})$,

then δ satisfies $[\mathbf{P6'(\Gamma^2, \boldsymbol{c})}]$.

Proof. Using assumptions (i) and (ii), one can prove that for all $j = 1, \dots, J$ we have

$$1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{C^{(j)}_{\boldsymbol{X},\boldsymbol{Y}}} \leq 1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{C^{(j)}_{\boldsymbol{X},\boldsymbol{Y}} + A^{(j)}} \leq c \frac{S_{\boldsymbol{X},\boldsymbol{Z}} + S_{\boldsymbol{Y},\boldsymbol{Z}}}{\max(U_{\boldsymbol{X},\boldsymbol{Z}}, U_{\boldsymbol{Z},\boldsymbol{Y}})} \leq c(\delta_{\boldsymbol{X},\boldsymbol{Z}} + \delta_{\boldsymbol{Y},\boldsymbol{Z}}).$$

It follows that

$$\delta_{\boldsymbol{X},\boldsymbol{Y}} = \sum_{j=1}^{J} \alpha^{(j)} \left(1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{C^{(j)}_{\boldsymbol{X},\boldsymbol{Y}}}\right) \leq \sum_{j=1}^{J} \alpha^{(j)} c(\delta_{\boldsymbol{X},\boldsymbol{Z}} + \delta_{\boldsymbol{Y},\boldsymbol{Z}}) = c(\delta_{\boldsymbol{X},\boldsymbol{Z}} + \delta_{\boldsymbol{Y},\boldsymbol{Z}}). \quad \triangle$$

Corollary 3. *The measure $\delta^{D,\alpha}$ satisfies* $[\mathbf{P6'(\Gamma^2, \frac{1}{\alpha^\wedge})}]$.

Proof. We have

$$\begin{aligned}
\delta^{D,\alpha}_{\boldsymbol{X},\boldsymbol{Y}} &= \alpha \min\left(\frac{H_{\boldsymbol{X}|\boldsymbol{Y}}}{H_{\boldsymbol{X}}}, \frac{H_{\boldsymbol{Y}|\boldsymbol{X}}}{H_{\boldsymbol{Y}}}\right) + (1-\alpha) \max\left(\frac{H_{\boldsymbol{X}|\boldsymbol{Y}}}{H_{\boldsymbol{X}}}, \frac{H_{\boldsymbol{Y}|\boldsymbol{X}}}{H_{\boldsymbol{Y}}}\right) \\
&= \alpha \frac{\min(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}})}{\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})} + (1-\alpha) \frac{\max(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}})}{\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})} \\
&= \alpha \left(1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})}\right) + (1-\alpha) \left(1 - \frac{I_{\boldsymbol{X},\boldsymbol{Y}}}{\max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})}\right).
\end{aligned}$$

Using the notation introduced in Proposition 10, we have $I = J = 2$, $S^{(1)}_{\boldsymbol{X},\boldsymbol{Y}} = \alpha \min(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}})$, $S^{(2)}_{\boldsymbol{X},\boldsymbol{Y}} = (1-\alpha)\max(H_{\boldsymbol{X}|\boldsymbol{Y}}, H_{\boldsymbol{Y}|\boldsymbol{X}})$, $U^{(1)}_{\boldsymbol{X},\boldsymbol{Y}} = \min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$, $U^{(2)}_{\boldsymbol{X},\boldsymbol{Y}} = \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$, $C^{(1)}_{\boldsymbol{X},\boldsymbol{Y}} = \min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$, and $C^{(2)}_{\boldsymbol{X},\boldsymbol{Y}} = \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}})$. Let us fix c to the value $\frac{1}{\alpha^\wedge}$. We have

$$\begin{aligned}
A^{(1)} = I_{\boldsymbol{X},\boldsymbol{Y}} - \min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}) + \frac{1}{\alpha^\wedge} \Big(&\alpha \min(H_{\boldsymbol{X}|\boldsymbol{Z}}, H_{\boldsymbol{Z}|\boldsymbol{X}}) + (1-\alpha)\max(H_{\boldsymbol{X}|\boldsymbol{Z}}, H_{\boldsymbol{Z}|\boldsymbol{X}}) \\
&+ \alpha \min(H_{\boldsymbol{Y}|\boldsymbol{Z}}, H_{\boldsymbol{Z}|\boldsymbol{Y}}) + (1-\alpha)\max(H_{\boldsymbol{Y}|\boldsymbol{Z}}, H_{\boldsymbol{Z}|\boldsymbol{Y}})\Big).
\end{aligned}$$

Clearly, from (29) we have

$$\begin{aligned}
A^{(1)} &\geq \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}) - H_{\boldsymbol{X},\boldsymbol{Y}} + 2H_{\boldsymbol{X},\boldsymbol{Z}} + 2H_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{X}} - H_{\boldsymbol{Y}} - 2H_{\boldsymbol{Z}} \\
&\geq H_{\boldsymbol{X},\boldsymbol{Z}} + H_{\boldsymbol{Y},\boldsymbol{Z}} - \min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}) - H_{\boldsymbol{Z}} \geq 0.
\end{aligned}$$

Furthermore,

$$\min(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}) + A^{(1)} \geq H_{\boldsymbol{X},\boldsymbol{Z}} + H_{\boldsymbol{Y},\boldsymbol{Z}} - H_{\boldsymbol{Z}} \geq \max(H_{\boldsymbol{X}}, H_{\boldsymbol{Y}}, H_{\boldsymbol{Z}}) = \max(U_{\boldsymbol{X},\boldsymbol{Z}}, U_{\boldsymbol{Y},\boldsymbol{Z}}).$$

It follows that $A^{(1)}$ satisfies conditions (i) and (ii) of Proposition 10 with $c = \frac{1}{\alpha^{\wedge}}$. The proof is strictly similar for $A^{(2)}$. $\triangle$

4. PREDICTION FRAMEWORK

Let us consider properties related to the prediction of some fixed random vector $\boldsymbol{Y}$.

4.1. Prediction Framework

Recall that our purpose is to find a random vector $\boldsymbol{X}$ that minimizes $\Delta_{\boldsymbol{Y},\boldsymbol{X}}$ (respectively, $\delta_{\boldsymbol{Y},\boldsymbol{X}}$), which combines a complexity term $C_{\boldsymbol{X},\boldsymbol{Y}}$ (to minimize) and an information term $I_{\boldsymbol{X},\boldsymbol{Y}}$ (to maximize). Let us assume that we have already got some $\boldsymbol{X}_1$ and its associated measure $\Delta_{\boldsymbol{Y},\boldsymbol{X}_1}$ (respectively, $\delta_{\boldsymbol{Y},\boldsymbol{X}_1}$). After evaluating $\Delta_{\boldsymbol{Y},\boldsymbol{X}_2}$ (respectively, $\delta_{\boldsymbol{Y},\boldsymbol{X}_2}$), we may be interested in describing conditions under which $\boldsymbol{X}_2$ is better (or worse) than $\boldsymbol{X}_1$.

Proposition 11. *The following cases may occur.*

Case 1. We choose $\boldsymbol{X}_2$ instead of $\boldsymbol{X}_1$ when

$$\begin{aligned} \Delta_{\boldsymbol{Y},\boldsymbol{X}_2} < \Delta_{\boldsymbol{Y},\boldsymbol{X}_1} \quad &\Longleftrightarrow \quad C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1} < I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}, \\ \delta_{\boldsymbol{Y},\boldsymbol{X}_2} < \delta_{\boldsymbol{Y},\boldsymbol{X}_1} \quad &\Longleftrightarrow \quad \frac{C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1}}{C_{\boldsymbol{Y},\boldsymbol{X}_1}} < \frac{I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}}{I_{\boldsymbol{Y},\boldsymbol{X}_1}}. \end{aligned} \tag{39}$$

Case 2. We keep $\boldsymbol{X}_1$ and reject $\boldsymbol{X}_2$ when

$$\begin{aligned} \Delta_{\boldsymbol{Y},\boldsymbol{X}_2} \geq \Delta_{\boldsymbol{Y},\boldsymbol{X}_1} \quad &\Longleftrightarrow \quad C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1} \geq I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}, \\ \delta_{\boldsymbol{Y},\boldsymbol{X}_2} \geq \delta_{\boldsymbol{Y},\boldsymbol{X}_1} \quad &\Longleftrightarrow \quad \frac{C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1}}{C_{\boldsymbol{Y},\boldsymbol{X}_1}} \geq \frac{I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}}{I_{\boldsymbol{Y},\boldsymbol{X}_1}}. \end{aligned}$$

This result implies automatically that properties **[P8]** and **[P9]** are satisfied. Let us comment more precisely on the previous proposition.

- Case 1 holds when
  1. $\boldsymbol{X}_2$ is simpler than $\boldsymbol{X}_1$ (i.e., $C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1} < 0$) and $\boldsymbol{X}_2$ is at least as informative as $\boldsymbol{X}_1$ (i.e., $I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1} \geq 0$).
  2. $\boldsymbol{X}_2$ and $\boldsymbol{X}_1$ have the same complexity (i.e., $C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1} = 0$) and $\boldsymbol{X}_2$ is more informative than $\boldsymbol{X}_1$ (i.e., $I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1} > 0$).
  3. $\boldsymbol{X}_2$ is simpler and less informative than $\boldsymbol{X}_1$ and such that the absolute (respectively, relative) excess of complexity is less than the absolute (respectively, relative) gain of information, i.e., $C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1} < I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1} < 0$ (respectively, $\frac{C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1}}{C_{\boldsymbol{Y},\boldsymbol{X}_1}} < \frac{I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}}{I_{\boldsymbol{Y},\boldsymbol{X}_1}} < 0$).
  4. $\boldsymbol{X}_2$ is more complex and more informative than $\boldsymbol{X}_1$ and such that the absolute (respectively, relative) excess of complexity is less than the absolute (respectively, relative) gain of information, i.e., $0 < C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1} < I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}$ (respectively, $0 < \frac{C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1}}{C_{\boldsymbol{Y},\boldsymbol{X}_1}} < \frac{I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}}{I_{\boldsymbol{Y},\boldsymbol{X}_1}}$).
- Case 2 holds when
  1. $\boldsymbol{X}_2$ is at least as complex as $\boldsymbol{X}_1$ (i.e., $C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1} \geq 0$) and $\boldsymbol{X}_2$ is at most as informative as $\boldsymbol{X}_1$ (i.e., $I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1} \leq 0$).

2. $\boldsymbol{X}_2$ is simpler and less informative than $\boldsymbol{X}_1$ and such that the absolute (respectively, relative) excess of complexity is greater than or equal to the absolute (respectively, relative) gain of information, i.e., $0 > C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1} \geq I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}$ (respectively, $0 > \frac{C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1}}{C_{\boldsymbol{Y},\boldsymbol{X}_1}} \geq \frac{I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}}{I_{\boldsymbol{Y},\boldsymbol{X}_1}}$).
3. $\boldsymbol{X}_2$ is more complex and more informative than $\boldsymbol{X}_1$ and such that the absolute (respectively, relative) excess of complexity is greater than or equal to the absolute (respectively, relative) gain of information, i.e., $C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1} \geq I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1} > 0$ (respectively, $\frac{C_{\boldsymbol{Y},\boldsymbol{X}_2} - C_{\boldsymbol{Y},\boldsymbol{X}_1}}{C_{\boldsymbol{Y},\boldsymbol{X}_1}} \geq \frac{I_{\boldsymbol{Y},\boldsymbol{X}_2} - I_{\boldsymbol{Y},\boldsymbol{X}_1}}{I_{\boldsymbol{Y},\boldsymbol{X}_1}} > 0$).

Proposition 12. *Any complexity term C^α of the form* (17) *with $\alpha \in]0,1[$ satisfies* **[P10]**.

Proof. Without loss of generality, the function $g(\cdot)$ that defines C^α is assumed to be increasing. It is clear that $H_{\boldsymbol{X}_2} \geq H_{\boldsymbol{X}_1} \Rightarrow C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_2} \geq C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_1}$ and respectively (by symmetry) $H_{\boldsymbol{X}_1} \geq H_{\boldsymbol{X}_2} \Rightarrow C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_1} \geq C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_2}$, which is equivalent to $C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_2} > C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_1} \Rightarrow H_{\boldsymbol{X}_2} > H_{\boldsymbol{X}_1}$. Thus, it remains to check that $C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_2} = C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_1} \Rightarrow H_{\boldsymbol{X}_2} = H_{\boldsymbol{X}_1}$:

$$\begin{aligned} C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_2} - C^\alpha_{\boldsymbol{Y},\boldsymbol{X}_1} = 0 \quad &\Longleftrightarrow \quad \alpha(g(m_1) - g(m_2)) + (1-\alpha)(g(M_1) - g(M_2)) = 0 \\ &\Longleftrightarrow \quad g(m_1) = g(m_2) \quad \text{and} \quad g(M_1) = g(M_2) \\ &\Longleftrightarrow \quad m_1 = m_2 \quad \text{and} \quad M_1 = M_2 \\ &\Longleftrightarrow \quad H_{\boldsymbol{X}_2} = H_{\boldsymbol{X}_1}. \end{aligned} \tag{40}$$

Equation (40) is obtained by observing that $g(m_1) - g(m_2)$ and $g(M_1) - g(M_2)$ are of the same sign. $\triangle$

Remark 7. The complexity terms C^E and C^I (which corresponds to the case of any C^α with $\alpha = 0$) do not satisfy property **[P10]** in the general case. Indeed, there is no implication for C^E, and one can only prove that $H_{\boldsymbol{X}_1} \geq H_{\boldsymbol{X}_2} \Rightarrow C^I_{\boldsymbol{Y},\boldsymbol{X}_1} \geq C^I_{\boldsymbol{Y},\boldsymbol{X}_2}$. However, we point out that when $I_{\boldsymbol{Y},\boldsymbol{X}_1} = I_{\boldsymbol{Y},\boldsymbol{X}_2}$, then both C^E and C^I satisfy **[P10]**.

More specifically, two frameworks may be of special interest.

- $\boldsymbol{X}_2$ is as informative as $\boldsymbol{X}_1$ (i.e., $I_{\boldsymbol{Y},\boldsymbol{X}_1} = I_{\boldsymbol{Y},\boldsymbol{X}_2}$): we expect to select the random variable with the smallest entropy. This is effectively what happens when **[P10]** is satisfied; from Proposition 12 and Remark 7 (in this framework), this is the case for $C^\bullet$ with $\bullet = I, S, R, P, D$ in the general case and for C^E in this framework, since $H_{\boldsymbol{Y},\boldsymbol{X}_2} - H_{\boldsymbol{Y},\boldsymbol{X}_1} = H_{\boldsymbol{X}_2} - H_{\boldsymbol{X}_1}$.
- $\boldsymbol{X}_1 = g(\boldsymbol{X}_2)$ with g some surjective (but not injective) mapping: $\boldsymbol{X}_2$ is more complex than $\boldsymbol{X}_1$, and $\boldsymbol{X}_2$ is at least as informative as $\boldsymbol{X}_1$. Consequently, this case is nontrivial since both the absolute (respectively, relative) excess of complexity and absolute (respectively, relative) gain of information are competing. Let us give two important examples of such a context.
  1. Quantization problem: given a quantized version $\boldsymbol{X}_1$ of some (continuous) random variable with its associated partition $\mathcal{A}_1$, the problem is to know whether some new quantized version $\boldsymbol{X}_2$ with an associated partition $\mathcal{A}_2$ finer than $\mathcal{A}_1$ should be preferred to predict $\boldsymbol{Y}$.
  2. Variable selection problem: assume that one wants to construct an ascending selection method. The vector $\boldsymbol{X}_1$ could represent some selected set of covariables and $\boldsymbol{X}_2 = (\boldsymbol{X}_1, \boldsymbol{X}'_2)$ a larger set of covariables. The aim is therefore to know if $\boldsymbol{X}'_2$ should be integrated to the selected set or not.

Some simple algorithms of quantization and selection methods are proposed in [15] using these results.

4.2. Around the Redundancy of Two Random Vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$

In the future use of an *I*-divergence or *NI*-divergence, one would expect that if two discrete-valued random vectors $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have the same (or almost the same) information with respect to an *I*-divergence or *NI*-divergence, then both have the same effect on the prediction of another vector $\boldsymbol{Y}$. This requirement, expressed by property **[P11]**, could be used for example in a variable selection problem in the context of discrimination to detect redundant variables.

In order to make property **[P11]** applicable for practical purpose, we may find interesting to have a bound on the difference $|\Delta_{\boldsymbol{Y},\boldsymbol{X}_1} - \Delta_{\boldsymbol{Y},\boldsymbol{X}_2}|$ (respectively, $|\delta_{\boldsymbol{Y},\boldsymbol{X}_1} - \delta_{\boldsymbol{Y},\boldsymbol{X}_2}|$) expressed in terms of $D^I_{\boldsymbol{X}_1,\boldsymbol{X}_2}$ (respectively, $d^I_{\boldsymbol{X}_1,\boldsymbol{X}_2}$). More precisely, the question may arise whether there exists a function $h(\cdot)$ satisfying $h(x) \to 0$ as $x \to 0$ and such that $|\Delta_{\boldsymbol{Y},\boldsymbol{X}_1} - \Delta_{\boldsymbol{Y},\boldsymbol{X}_2}| \le h(D^I_{\boldsymbol{X}_1,\boldsymbol{X}_2})$ (respectively, $|\delta_{\boldsymbol{Y},\boldsymbol{X}_1} - \delta_{\boldsymbol{Y},\boldsymbol{X}_2}| \le h(d^I_{\boldsymbol{X}_1,\boldsymbol{X}_2})$). Here, according to our examples, we only concentrate on linear functions $h(\cdot)$.

We then propose to translate property **[P11]** to an *I*-divergence Δ (respectively, *NI*-divergence δ) as follows:

[P11′($\boldsymbol{\Upsilon}, \boldsymbol{k}$)] There exists some positive constant k such that for all $(\boldsymbol{X}_1, \boldsymbol{X}_2) \in \Upsilon \subset \Gamma^2$ we have

$$|\Delta_{\boldsymbol{Y},\boldsymbol{X}_1} - \Delta_{\boldsymbol{Y},\boldsymbol{X}_2}| \le k D^I_{\boldsymbol{X}_1,\boldsymbol{X}_2}. \tag{41}$$

As a first answer, let us note that if the *I*-divergence (respectively, *NI*-divergence) satisfies the triangle inequality **[P6′($\boldsymbol{\Gamma^2}$, 1)]** or **[P3′($\boldsymbol{\Upsilon}, \boldsymbol{k_1}, \boldsymbol{k_2}$)]**, then it satisfies **[P11′($\boldsymbol{\Upsilon}, \boldsymbol{k_2}$)]** due to the following equivalent expression of the triangle inequality:

$$|D_{\boldsymbol{Y},\boldsymbol{X}_1} - D_{\boldsymbol{Y},\boldsymbol{X}_2}| \le D_{\boldsymbol{X}_1,\boldsymbol{X}_2} \qquad (\text{respectively,} \quad |d_{\boldsymbol{Y},\boldsymbol{X}_1} - d_{\boldsymbol{Y},\boldsymbol{X}_2}| \le d_{\boldsymbol{X}_1,\boldsymbol{X}_2}).$$

A priori, if an *I*-divergence or *NI*-divergence only satisfies **[P6′($\boldsymbol{\Gamma^2}, \boldsymbol{c}$)]** with some $c > 1$, then this property does no more seem to be true: indeed, for all $\boldsymbol{Y}$, $\boldsymbol{X}_1$, and $\boldsymbol{X}_2$, one may prove for an *I*-divergence, for instance, that

$$|\Delta_{\boldsymbol{Y},\boldsymbol{X}_1} - \Delta_{\boldsymbol{Y},\boldsymbol{X}_2}| \le c\Delta_{\boldsymbol{X}_1,\boldsymbol{X}_2} + (c-1)\min(\Delta_{\boldsymbol{Y},\boldsymbol{X}_1}, \Delta_{\boldsymbol{Y},\boldsymbol{X}_2}) \not\le c\Delta_{\boldsymbol{X}_1,\boldsymbol{X}_2}.$$

Actually, this apparently disappointing result only expresses that the "redundancy" property cannot (always) be derived from a triangle-type inequality.

The following proposition gives some sufficient conditions on a complexity term ensuring that the associated Δ and δ satisfy property **[P11′]**.

Proposition 13. (i) *Assume that there exists some positive constant $\varkappa_1$ such that the complexity term of an I-divergence for all $(\boldsymbol{X}_1, \boldsymbol{X}_2) \in \Upsilon$ satisfies*

$$|C_{\boldsymbol{Y},\boldsymbol{X}_1} - C_{\boldsymbol{Y},\boldsymbol{X}_2}| \le \varkappa_1 |H_{\boldsymbol{X}_1} - H_{\boldsymbol{X}_2}|. \tag{42}$$

Then Δ satisfies **[P11′($\boldsymbol{\Upsilon}, \boldsymbol{1 + \varkappa_1}$)]**.

(ii) *If in addition there exists some positive constant $\varkappa_2$ such that for all $(\boldsymbol{X}_1, \boldsymbol{X}_2) \in \Upsilon$*

$$\max(C_{\boldsymbol{Y},\boldsymbol{X}_1}, C_{\boldsymbol{Y},\boldsymbol{X}_2}) \ge \varkappa_2 C^I_{\boldsymbol{X}_1,\boldsymbol{X}_2}, \tag{43}$$

then the associated NI-divergence satisfies **[P11′($\boldsymbol{\Upsilon}, \frac{\boldsymbol{1 + \varkappa_1}}{\boldsymbol{\varkappa_2}}$)]**.

Proof. (i) Let us first write

$$|\Delta_{\boldsymbol{Y},\boldsymbol{X}_1} - \Delta_{\boldsymbol{Y},\boldsymbol{X}_2}| \le |I_{\boldsymbol{Y},\boldsymbol{X}_1} - I_{\boldsymbol{Y},\boldsymbol{X}_2}| + |C_{\boldsymbol{Y},\boldsymbol{X}_1} - C_{\boldsymbol{Y},\boldsymbol{X}_2}|. \tag{44}$$

Now note that

$$I_{\boldsymbol{Y},\boldsymbol{X}_1} \geq I_{\boldsymbol{Y},\boldsymbol{X}_2} + I_{\boldsymbol{X}_1,\boldsymbol{X}_2} - H_{\boldsymbol{X}_2},$$

from which one can deduce

$$|I_{\boldsymbol{Y},\boldsymbol{X}_1} - I_{\boldsymbol{Y},\boldsymbol{X}_2}| \leq \max(H_{\boldsymbol{X}_1}, H_{\boldsymbol{X}_2}) - I_{\boldsymbol{X}_1,\boldsymbol{X}_2} = \max(H_{\boldsymbol{X}_1|\boldsymbol{X}_2}, H_{\boldsymbol{X}_2|\boldsymbol{X}_1}) = D^I_{\boldsymbol{X}_1,\boldsymbol{X}_2}. \tag{45}$$

The result is then obtained by combining (42), (44), and (45).

(ii) We can obtain the following result:

$$\begin{aligned} |\delta_{\boldsymbol{Y},\boldsymbol{X}_1} - \delta_{\boldsymbol{Y},\boldsymbol{X}_2}| &\leq \frac{\min(C_{\boldsymbol{Y},\boldsymbol{X}_1}, C_{\boldsymbol{Y},\boldsymbol{X}_2})(|I_{\boldsymbol{Y},\boldsymbol{X}_1} - I_{\boldsymbol{Y},\boldsymbol{X}_2}| + |C_{\boldsymbol{Y},\boldsymbol{X}_1} - C_{\boldsymbol{Y},\boldsymbol{X}_2}|)}{C_{\boldsymbol{Y},\boldsymbol{X}_1} C_{\boldsymbol{Y},\boldsymbol{X}_2}} \\ &\leq \frac{|I_{\boldsymbol{Y},\boldsymbol{X}_1} - I_{\boldsymbol{Y},\boldsymbol{X}_2}| + |C_{\boldsymbol{Y},\boldsymbol{X}_1} - C_{\boldsymbol{Y},\boldsymbol{X}_2}|}{\max(C_{\boldsymbol{Y},\boldsymbol{X}_1}, C_{\boldsymbol{Y},\boldsymbol{X}_2})}. \end{aligned}$$

The result then comes from (42), (43), and (45). $\triangle$

Let us apply the previous result to our different examples.

Corollary 4. *Let $\boldsymbol{X}_1, \boldsymbol{X}_2 \in \Gamma_\Theta$ with $\Theta = [c_1, c_2]$; define γ_i, $i = 1, 2$, such that $c_i = \gamma_i H_{\boldsymbol{Y}}$. Then*

$$|\Delta^\bullet_{\boldsymbol{Y},\boldsymbol{X}_1} - \Delta^\bullet_{\boldsymbol{Y},\boldsymbol{X}_2}| \leq (1 + \varkappa^\bullet_{1,\Theta}) D^I_{\boldsymbol{X}_1,\boldsymbol{X}_2} \qquad \text{and} \qquad |\delta^\bullet_{\boldsymbol{Y},\boldsymbol{X}_1} - \delta^\bullet_{\boldsymbol{Y},\boldsymbol{X}_2}| \leq \begin{matrix} 1 + \varkappa^\bullet_{1,\Theta} \\ \varkappa^\bullet_{2,\Theta} \end{matrix} d^I_{\boldsymbol{X}_1,\boldsymbol{X}_2}, \tag{46}$$

where $\bullet$ stands for S, R, P, and D, and where the different constants are expressed by

$\bullet$	$\varkappa^\bullet_{1,\Theta}$	$\varkappa^\bullet_{2,\Theta}$
S	$\alpha^\vee$	$(1-\alpha) + \alpha\gamma_{1,2}$
R	$\alpha^{\vee 2} + \frac{\alpha(1-\alpha)}{\sqrt{\gamma_1}}$	$\left((1-\alpha) + \alpha\sqrt{\gamma_{1,2}}\right)^2$
P	$\max\left(\frac{1-\alpha}{\gamma_1^\alpha}, \frac{\alpha}{\gamma_1^{1-\alpha}}, \mathbf{1}_{]0,1]}(\gamma_1)\right)$	$\gamma_{1,2}^\alpha$
D	$\frac{\alpha^\vee}{(\alpha^\wedge)^2} \frac{1}{(1+\gamma_{1,2})^2}$	$\left(\frac{\alpha}{\gamma_{1,2}} + (1-\alpha)\right)^{-1}$

with $\gamma_{1,2} = \min\left(\gamma_1, \frac{1}{\gamma_2}\right)$.

Proof. For the sake of simplicity, let us denote $m_i = \min(H_{\boldsymbol{Y}}, H_{\boldsymbol{X}_i})$ and $M_i = \max(H_{\boldsymbol{Y}}, H_{\boldsymbol{X}_i})$ for $i = 1, 2$. The proofs mainly rely upon the two following tools:

- Since $m_1 - m_2$ and $M_1 - M_2$ are of the same sign, we have

$$|(m_1 - m_2)| + |M_1 - M_2| = |(m_1 - m_2) + (M_1 - M_2)| = |H_{\boldsymbol{X}_1} - H_{\boldsymbol{X}_2}|$$

  and then for any $f_1, f_2 \geq 0$

$$|f_1(m_1 - m_2) + f_2(M_1 - M_2)| \leq f^\vee |H_{\boldsymbol{X}_1} - H_{\boldsymbol{X}_2}|, \tag{47}$$

  with $f^\vee = \max(f_1, f_2)$.
- For $i = 1, 2$, we have

$$m_i \geq \left\{ \begin{matrix} \min(1, \gamma_1) M_i \\ \min\left(1, \frac{1}{\gamma_2}\right) M_i \end{matrix} \right\} \geq \min\left(1, \gamma_1, \frac{1}{\gamma_2}\right) M_i = \gamma_{1,2} M_i,$$

  since $\gamma_{1,2} < 1$ as a direct consequence of $\gamma_1 < \gamma_2$.

For different cases, we need in particular to check (43), which holds whenever there exists some positive constant $\varkappa$ such that $C^{\bullet,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_i} \geq \varkappa M_i$ for $i = 1, 2$, since $\max(C^{\bullet,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_1}, C^{\bullet,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_2}) \geq \varkappa \max(M_1, M_2) \geq \varkappa C^{I}_{\boldsymbol{X}_1,\boldsymbol{X}_2}$.

- Complexity term $C^{S,\alpha}$: we have

$$\begin{aligned} |C^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_1} - C^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_2}| &= |\alpha m_1 + (1-\alpha)M_1 - \alpha m_2 - (1-\alpha)M_2| \\ &= |\alpha(m_1 - m_2) + (1-\alpha)(M_1 - M_2)| \\ &\leq \alpha^{\vee} |H_{\boldsymbol{X}_1} - H_{\boldsymbol{X}_2}| \end{aligned}$$

  from (47), with $f_1 = \alpha$ and $f_2 = (1-\alpha)$. Moreover, for $i = 1, 2$,

$$C^{S,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_i} = \alpha m_i + (1-\alpha)M_i \geq ((1-\alpha) + \alpha\gamma_{1,2})M_i.$$

- Complexity term $C^{R,\alpha}$: we have

$$\left|C^{R,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_1} - C^{R,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_2}\right| = \left|\alpha^2(m_1 - m_2) + (1-\alpha)^2(M_1 - M_2) + 2\alpha(1-\alpha)\sqrt{H_{\boldsymbol{Y}}}\left(\sqrt{H_{\boldsymbol{X}_1}} - \sqrt{H_{\boldsymbol{X}_2}}\right)\right|.$$

  Furthermore, we may obtain

$$|\alpha^2(m_1 - m_2) + (1-\alpha)^2(M_1 - M_2)| \leq \alpha^{\vee 2} |H_{\boldsymbol{X}_1} - H_{\boldsymbol{X}_2}|$$

  and

$$\left|\sqrt{H_{\boldsymbol{Y}}}\left(\sqrt{H_{\boldsymbol{X}_1}} - \sqrt{H_{\boldsymbol{X}_2}}\right)\right| = \frac{\sqrt{H_{\boldsymbol{Y}}}}{2\sqrt{\min(H_{\boldsymbol{X}_1}, H_{\boldsymbol{X}_2})}} |H_{\boldsymbol{X}_1} - H_{\boldsymbol{X}_2}| \leq \frac{1}{2\sqrt{\gamma_1}} |H_{\boldsymbol{X}_1} - H_{\boldsymbol{X}_2}|.$$

  Hence,

$$\left|C^{R,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_1} - C^{R,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_2}\right| \leq \left(\alpha^{\vee 2} + \frac{\alpha(1-\alpha)}{\sqrt{\gamma_1}}\right) |H_{\boldsymbol{X}_1} - H_{\boldsymbol{X}_2}|$$

  from (47), with $f_1 = \alpha^2$ and $f_2 = (1-\alpha)^2$. Moreover, one can prove for $i = 1, 2$

$$C^{R,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_i} = \left(\alpha\sqrt{m_i} + (1-\alpha)\sqrt{M_i}\right)^2 \geq \left((1-\alpha) + \alpha\sqrt{\gamma_{1,2}}\right)^2 M_i.$$

- Complexity term $C^{P,\alpha}$: we have (by assuming that $H_{\boldsymbol{X}_2} > H_{\boldsymbol{X}_1}$)

$$\begin{aligned} \left|C^{P,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_1} - C^{P,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_2}\right| &= \left|m_1^{\alpha} M_1^{1-\alpha} - m_2^{\alpha} M_2^{1-\alpha}\right| \\ &= \begin{cases} H_{\boldsymbol{Y}}^{\alpha}(H_{\boldsymbol{X}_2}^{1-\alpha} - H_{\boldsymbol{X}_1}^{1-\alpha}) & \text{if } H_{\boldsymbol{Y}} \leq \min(H_{\boldsymbol{X}_1}, H_{\boldsymbol{X}_2}), \\ H_{\boldsymbol{Y}}^{1-\alpha}(H_{\boldsymbol{X}_2}^{\alpha} - H_{\boldsymbol{X}_1}^{\alpha}) & \text{if } H_{\boldsymbol{Y}} \geq \max(H_{\boldsymbol{X}_1}, H_{\boldsymbol{X}_2}), \\ H_{\boldsymbol{Y}}^{\alpha} H_{\boldsymbol{X}_2}^{1-\alpha} - H_{\boldsymbol{X}_1}^{\alpha} H_{\boldsymbol{Y}}^{1-\alpha} & \text{otherwise.} \end{cases} \end{aligned}$$

  Note that the third case cannot occur if $\gamma_1 \geq 1$. We have

$$\begin{aligned} \left|C^{P,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_1} - C^{P,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_2}\right| &\leq \begin{cases} \frac{1-\alpha}{\gamma_1^{\alpha}}(H_{\boldsymbol{X}_2} - H_{\boldsymbol{X}_1}) & \text{if } H_{\boldsymbol{Y}} \leq \min(H_{\boldsymbol{X}_1}, H_{\boldsymbol{X}_2}), \\ \frac{\alpha}{\gamma_1^{1-\alpha}}(H_{\boldsymbol{X}_2} - H_{\boldsymbol{X}_1}) & \text{if } H_{\boldsymbol{Y}} \geq \max(H_{\boldsymbol{X}_1}, H_{\boldsymbol{X}_2}), \\ H_{\boldsymbol{X}_2} - H_{\boldsymbol{X}_1} & \text{otherwise} \end{cases} \\ &\leq \max\left(\frac{1-\alpha}{\gamma_1^{\alpha}}, \frac{\alpha}{\gamma_1^{1-\alpha}}, \mathbf{1}_{]0,1]}(\gamma_1)\right) |H_{\boldsymbol{X}_2} - H_{\boldsymbol{X}_1}|. \end{aligned}$$

  Moreover, we may obtain for $i = 1, 2$

$$C^{P,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_i} = m_i^{\alpha} M_i^{1-\alpha} \geq \gamma_{1,2}^{\alpha} M_i.$$

- Complexity term $C^{D,\alpha}$: we have

$$\begin{aligned}\left|C^{D,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_1} - C^{D,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_2}\right| &= \frac{|\alpha M_1 M_2(m_1-m_2)+(1-\alpha)m_1m_2(M_1-M_2)|}{(\alpha M_1+(1-\alpha)m_1)(\alpha M_2+(1-\alpha)m_2)} \\ &\le \frac{\alpha^\vee}{(\alpha^\wedge)^2}\frac{M_1M_2}{(m_1+M_1)(m_2+M_2)}|H_{\boldsymbol{X}_2}-H_{\boldsymbol{X}_1}| \qquad (48)\\ &\le \frac{\alpha^\vee}{(\alpha^\wedge)^2}\frac{1}{(1+\gamma_{1,2})^2}|H_{\boldsymbol{X}_2}-H_{\boldsymbol{X}_1}|.\end{aligned}$$

Equation (48) is obtained using (47) with $f_1 = \alpha M_1 M_2$ and $f_2 = (1-\alpha)m_1m_2$. Finally, we also have for $i = 1, 2$

$$C^{D,\alpha}_{\boldsymbol{Y},\boldsymbol{X}_i} = \left(\frac{\alpha}{m_i} + \frac{1-\alpha}{M_i}\right)^{-1} \ge \left(\frac{\alpha}{\gamma_{1,2}} + (1-\alpha)\right)^{-1} M_i. \quad \triangle$$

Remark 8. Note that when $\alpha \le \frac{1}{2}$, the measure $\Delta^{S,\alpha}$ is a metric and so we derive (46) directly from **[P3′]**.

The authors would like to thank the referee whose suggestions and fruitful comments helped them to improve the paper.

REFERENCES

1. Shannon, C.E., A Mathematical Theory of Communication (continued), *Bell Syst. Tech. J.*, 1948, vol. 27, no. 4, pp. 623–656.
2. Crutchfield J.P., Information and Its Metric, *Nonlinear Structures in Physical Systems: Pattern Formation, Chaos, and Waves (Proc. 2nd Woodward Conf., San Jose State Univ., 1989)*, Lam, L. and Morris, H.C., Eds., New York: Springer, 1990, pp. 119–130.
3. Hillman, C., A Formal Theory of Information: I. Statics, *Preprint*, 1998. Available from `http://citeseer.ist.psu.edu/89520.html`.
4. Bennett, C.H., Gács, P., Li, M., Vitányi, P.M.B., and Zurek, W.H., Information Distance, *IEEE Trans. Inform. Theory*, 1998, vol. 44, no. 4, pp. 1407–1423.
5. Li, M., Chen, X., Li, X., Ma, B., and Vitányi, P.M.B., The Similarity Metric, *IEEE Trans. Inform. Theory*, 2004, vol. 50, no. 12, pp. 3250–3264.
6. Cilibrasi, R. and Vitányi, P.M.B., Clustering by Compression, *IEEE Trans. Inform. Theory*, 2005, vol. 51, no. 4, pp. 1523–1545.
7. Li, M., Badger, J.H., Chen, X., Kwong, S., Kearney, P., and Zhang, H., An Information-Based Sequence Distance and Its Application to Whole Mitochondrial Genome Phylogeny, *Bioinformatics*, 2001, vol. 17, no. 1, pp. 149–154.
8. Cilibrasi, R., Vitányi, P., and de Wolf, R., Algorithmic Clustering of Music, 2003, e-print `arxiv.org/cs.SD/0303025`.
9. Cilibrasi, R. and Vitányi, P., Automatic Meaning Discovery Using Google, 2005, e-print `arxiv.org/cs.CL/0412098`.
10. Grünwald, P. and Vitányi, P., Shannon Information and Kolmogorov Complexity, 2004, e-print `arxiv.org/cs.IT/0410002`.
11. Leung-Yan-Cheong, S.K. and Cover, T.M., Some Equivalences between Shannon Entropy and Kolmogorov Complexity, *IEEE Trans. on Inf. Theory*, 1978, vol. 24, no. 3, pp. 331–338.
12. Hammer, D., Romashchenko, A., Shen, A., and Vereshchagin, N., Inequalities for Shannon Entropy and Kolmogorov Complexity, *J. Comput. Syst. Sci.*, 2000, vol. 60, no. 2, pp. 442–464.

13. Kraskov, A., Stögbauer, H., Andrzejak, R.G., and Grassberger, P., Hierarchical Clustering Based on Mutual Information, 2003, e-print arxiv.org/q-bio/0311039.

14. Liu, H. and Motoda, H., *Feature Selection for Knowledge Discovery and Data Mining*, Boston: Kluwer, 1998.

15. Robineau, J.-F., Méthodes de sélection de variables (parmi un grand nombre) dans un cadre de discrimination, *PhD Thesis*, Grenoble, France: Univ. Joseph Fourier, 2004.

16. Granger, C.W., Maasoumi, E., and Racine, J., A Dependence Metric for Possibly Nonlinear Processes, *J. Time Ser. Anal.*, 2004, vol. 25, no. 5, pp. 649–669.

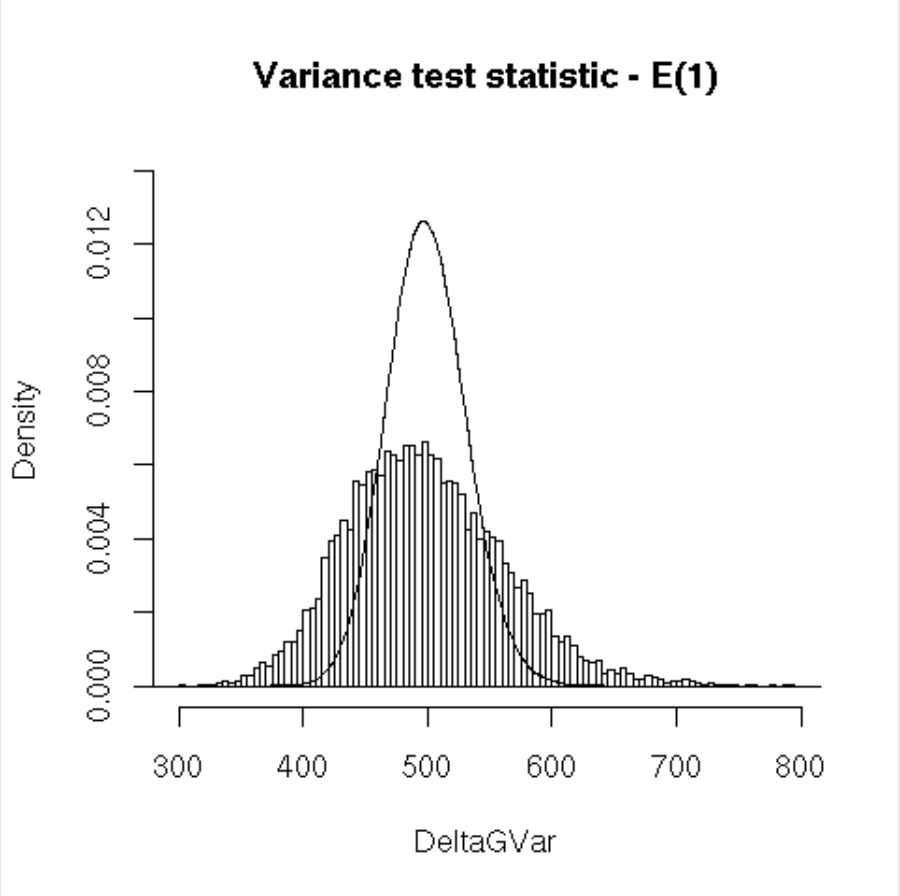
17. Cover, T.M. and Thomas, J.A., *Elements of Information Theory*, New York: Wiley, 1991.

18. Kaltchenko, A., Algorithms for Estimation of Information Distance with Application to Bioinformatics and Linguistics, in *Proc. 2004 Canadian Conf. on Electrical and Computer Engineering*, 2004, vol. 4, p. 2255. Available from arxiv.org/cs.CC/0404039.

19. Li, M. and Vitányi, P., *An Introduction to Kolmogorov Complexity and Its Applications*, New York: Springer, 1997, 2nd ed.

20. Li, M, Chen, X., Li, X., Ma, B., and Vitányi, P., The Similarity Metric, in *Proc. 14th Annual ACM–SIAM Symp. on Discrete Algorithms (SODA-03), New York*, 2003, pp. 863–872.

21. Ullah, A., Entropy, Divergence and Distance Measures with Econometric Applications, *J. Stat. Plan. Inference*, 1996, vol. 49, no. 1, pp. 137–162.



[CDLR09]
J.-F. Coeurjolly and R. Drouilhet and P. Lafaye de Micheaux and J.-F. Robineau. asympTest: an `R` package for performing parametric statistical tests and confidence intervals based on the central limit theorem. *R Journal*. 2:26-30 (Long version (19p.) at `http://arxiv.org/abs/0902.0506`), 2009.

asympTest: A Simple R Package for Classical Parametric Statistical Tests and Confidence Intervals in Large Samples

by J.-F. Coeurjolly, R. Drouilhet, P. Lafaye de Micheaux and J.-F. Robineau

Abstract: **asympTest** is an R package implementing large sample tests and confidence intervals. One and two sample mean and variance tests (differences and ratios) are considered. The test statistics are all expressed in the same form as the Student t-test, which facilitates their presentation in the classroom. This contribution also fills the gap of a robust (to non-normality) alternative to the chi-square single variance test for large samples, since no such procedure is implemented in standard statistical software.

Introduction

It is sometimes desirable to compare two variances rather than two averages. To cite a few examples (Dean and Illowsky (2009)): college administrators would like two college professors grading exams to have the same variation in their grading; in order for a lid to fit a container, the variation in the lid and the container should be the same; a supermarket might be interested in the variability of check-out times for two checkers.

Now usually, a first course on statistical inference presents mean tests in both Gaussian and asymptotical frameworks (Table 1), but variance tests are often presented only in the Gaussian case (Table 2).

Population law		Test statistic	Law
Gaussian	σ^2 known	$\frac{\overline{Y}_n - \mu_{ref}}{\sigma/\sqrt{n}}$	$\mathcal{N}(0,1)$
	σ^2 unknown	$\frac{\overline{Y}_n - \mu_{ref}}{S_n/\sqrt{n}}$	$t(n-1)$
Unknown ($n > 30$)	σ^2 known	$\frac{\overline{Y}_n - \mu_{ref}}{\sigma/\sqrt{n}}$	$\approx\mathcal{N}(0,1)$ asympt.
	σ^2 unknown	$\frac{\overline{Y}_n - \mu_{ref}}{S_n/\sqrt{n}}$	$\approx\mathcal{N}(0,1)$ asympt.

Table 1: Testing $H_0 : \mu = \mu_{ref}$ for both the Gaussian and large sample cases.

Test statistic	Law
$(n-1)S_n^2/\sigma_{ref}^2$	$\chi^2_{(n-1)}$
S_1^2/S_2^2	$F_{(n_1-1,n_2-1)}$

Table 2: Testing $H_0 : \sigma^2 = \sigma_{ref}^2$ or $H_0 : \sigma_1^2 = \sigma_2^2$ for the Gaussian case (σ^2, σ_1^2, σ_2^2 unknown; σ_{ref}^2 known).

An important point to be noticed is that students are usually told that mean tests are robust to non-normality for large samples as indicated by the asymptotic $\mathcal{N}(0,1)$ distribution in the last two cells of Table 1 (see e.g. Ozgur and Strasser (2004)). They could think that this also occurs for variance tests. Indeed, many practitioners use the classical chi-square single variance test or Fisher's two variances test, even if the Gaussian assumption fails. This could lead to heavy errors, **even for large samples**, as shown in Figure 1. Miller (1997, p. 264) describes this situation as "catastrophic".

To have a better idea of the type I error in the classical single variance test, let us test for example $H_0 : \sigma^2 = 1$ versus $H_1 : \sigma^2 < 1$, by simulating 10000 samples of size 1000 from an $\mathcal{E}(1)$ distribution (i.e. under H_0) and using $\alpha = 5\%$. We obtained a percentage of rejection of the null of 21.53%, thus showing a type I error far greater than α. The percentage for the asymptotic test (described later) is 9.05% which is not too far from α. For a $\mathcal{U}([0,5])$, the classical single variance test leads to a type I error far lesser than α (0.44%). Our test still behaves correctly with a type I error near α (5.39%). This is mainly due to the departure of the kurtosis of the distribution from 3 (for more theoretical details see *e.g.* Section 2.2 of Coeurjolly et al. (2009)).

Note that the problem of the robustness (to departures from normality) of tests for comparing two (or more) variances has been widely treated in the literature, see e.g. Box (1953), Conover et al. (1981), Tiku and Akkaya (2004), Pan (1999) and the references therein. These authors built specific test statistics. Note also that in the one sample (non Gaussian) case, to the best of our knowledge, no statistical tool is available to compare a population variance to a reference value.

Now, it is well-known, see e.g. Casella and Berger (2001, p. 492), that a common method for constructing a large sample test statistic may be based on an estimator that has an asymptotic normal distribution. Suppose we wish to test a hypothesis about a parameter θ, and $\hat{\theta}_n$ is some estimator of θ based on a sample of size n. If we can prove some form of the central limit theorem to show that, as $n \to +\infty$,

$$(\hat{\theta} - \theta)/\hat{\sigma}_{\hat{\theta}} \xrightarrow{d} \mathcal{N}(0,1) \qquad (1)$$

where $\hat{\sigma}_{\hat{\theta}}$ is the usual standard error, which is a convergent (in probability) estimate of $\sigma_{\hat{\theta}} = \sqrt{Var(\hat{\theta}_n)}$, then one has the basis for an approximate test.

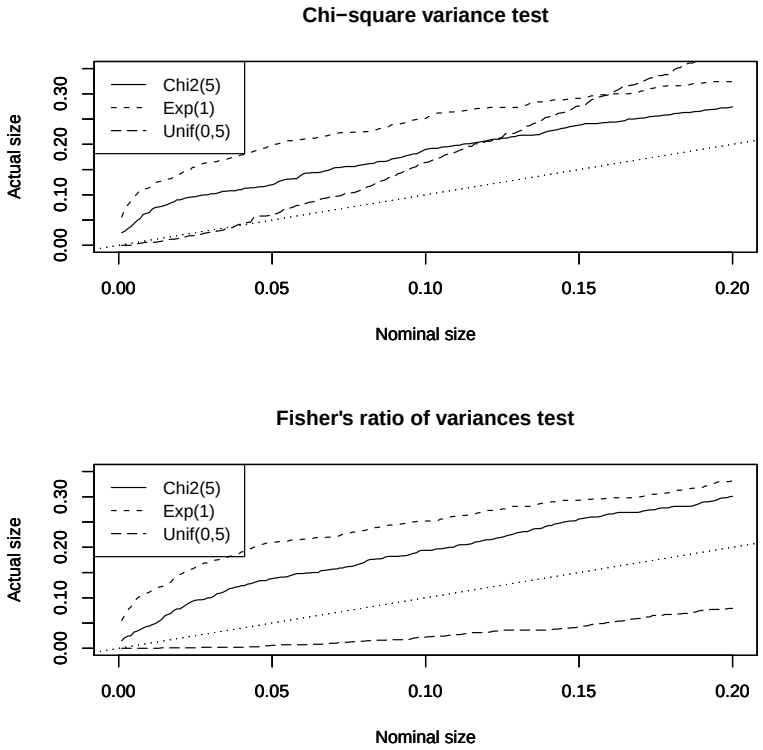


Figure 1: P-value Plots (see Davidson and MacKinnon (1998)) under H_0 of $m = 10000$ replications of test statistics of the chi-square variance test (top) and Fisher's ratio of variances test (bottom) in the large sample Gaussian context. The parameters of the simulation are: $n = n_1 = n_2 = 500$, $Y \overset{d}{=} Y_1 \overset{d}{=} Y_2 \sim \chi^2(5)$ (resp. $\mathcal{E}(1)$, resp. $\mathcal{U}[0,5]$). The dotted lines are 45° lines.

This approach can be used to complete Table 2 for the large sample case, shown in Table 3 for the single variance test only:

Population law	Test statistic	Law
Unknown (n large) with finite 4^{th} moment	$\dfrac{S_n^2 - \sigma_{ref}^2}{\hat{\sigma}_{S_n^2}}$	$\approx \mathcal{N}(0,1)$ asympt.

Table 3: Testing $H_0 : \sigma^2 = \sigma_{ref}^2$ for the large sample case. We let $\hat{\sigma}^2_{S_n^2} = \frac{1}{n(n-1)} \sum_{i=1}^n ((y_i - \overline{Y}_n)^2 - S_n^2)^2$.

The case of a (large sample) test for a difference in scale parameters (possibly weighted by a factor ρ) is also of interest as suggested by the availability of related procedures in R (to compute Ansari-Bradley's and Mood's tests for example). The standard error involved in this test is $\hat{\sigma}_{\hat{\theta}} = \sqrt{\hat{\sigma}^2_{S^2_{n_1}} + \rho^2 \hat{\sigma}^2_{S^2_{n_2}}}$.

The point to be noted here is that this general approach has been extensively used in Coeurjolly et al. (2009) where we end up with a unified approach very similar to the classical t-test from a mathematical point of view. Proofs, which are not very complicated, are provided in the report just cited. The details are not fully expounded here but lead us to propose a more complete, homogeneous teaching framework, with no additional difficulty, to test various parameters such as the mean, the variance, and the difference or ratio of means or variances (for large samples). This approach also allows the direct derivation of asymptotic confidence intervals. Note that Bonnet (2006a) and Bonnet (2006b) use a similar asymptotic approach, with a refinement based on a variance stabilizing transformation, to obtain asymptotic confidence intervals, solely for the single variance and ratio of variances cases. Table 4 gives a summary of the various parameters we can test and the R functions we have implemented to compute the standard error $\hat{\sigma}_{\hat{\theta}}$ of $\hat{\theta}$:

θ	Dataset(s)	$\hat{\sigma}_{\hat{\theta}}$ in R
μ	`y`	`seMean(y)`
σ^2	`y`	`seVar(y)`
$d_\mu = \mu_1 - \rho\mu_2$	`y1, y2`	`seDMean(y1,y2,rho)`
$d_{\sigma^2} = \sigma_1^2 - \rho\sigma_2^2$	`y1, y2`	`seDVar(y1,y2,rho)`
$r_\mu = \mu_1/\mu_2$	`y1, y2`	`seRMean(y1,y2)`
$r_{\sigma^2} = \sigma_1^2/\sigma_2^2$	`y1, y2`	`seRVar(y1,y2)`

Table 4: Various parameters we can test and available R functions to compute standard error $\hat{\sigma}_{\hat{\theta}}$.

These functions can be used in conjunction with (1) to obtain p-values for various tests. For a simple example, if you want to use a sample contained in

	$\mathcal{E}(1)$		$\chi^2(5)$		$\mathcal{U}([0,5])$	
n	χ^2	asymp.	χ^2	asymp.	χ^2	asymp.
30	0.2168	0.2733	0.1278	0.2218	0.0086	0.0801
100	0.2194	0.1765	0.1307	0.1442	0.0061	0.0589
500	0.2157	0.1102	0.1367	0.0928	0.0051	0.0543
1000	0.2153	0.0905	0.1323	0.0787	0.0040	0.0539

Table 5: Type I error in terms of n for the test $H_1 : \sigma^2 < \sigma^2_{ref}$ with $\sigma^2_{ref} = 1$ ($\mathcal{E}(1)$), 10 ($\chi^2(5)$), 25/12 ($\mathcal{U}([0,5])$) based on $m = 10000$ replications.

the vector `y` to test $H_0 : \sigma^2 = 1$, you can use

```
2*pnorm(-abs((var(y)-1)/seVar(y)))
```

This contribution also solves the problem of providing an implemented "robust" (to departure of the i.i.d. large sample distribution from normality) alternative to the chi-square single variance test for large samples. Indeed, we did not find any such procedure in standard statistical software and so it is highly likely that practitioners would incorrectly use a chi-square test on a single variance. It also provides a very simple alternative to the (ratio of variances) Fisher test in large samples. Some other "robust" alternative procedures to the Fisher test in the case of non Gaussian (not necessary large) samples are implemented in R: the Bartlett test (`bartlett.test`), the Fligner test (`fligner.test`) and the Levene test (`levene.test` available in the `lawstat` package). R also provides, through `ansari.test` and `mood.test` functions, Ansari-Bradley's and Mood's two-sample rank-based tests for a difference in scale parameters. The purpose of this paper is not to compare our tests to their competitors in terms of power. We nevertheless conduct two short simulation studies (limited to the probability of Type I error): first for the problem of testing a variance (Table 5), comparing the classical χ^2 single variance test to our procedure, and second for the problem of comparing (the differences d_{σ^2} of) two variances (Tables 6, 7 and 8), comparing the classical Fisher test to our procedure, as well as Ansari-Bradley's test and Mood's test. These simulations were based on the three distributions used earlier in Figure 1. The simulations show that the level α is quite correct (when n increases) for our procedure in the case of testing a single variance and for all three alternative tests (ours, Ansari-Bradley's and Mood's tests) for testing two variances.

	$\mathcal{E}(1)$			
n	$\mathcal{F}$	asympTest	Ansari	Mood
30	0.2827	0.0675	0.0478	0.0497
100	0.3083	0.0500	0.0480	0.0484
500	0.3269	0.0454	0.0484	0.0470
1000	0.3260	0.0526	0.0501	0.0515

Table 6: Type I error for the test $H_1 : \sigma^2_1 \neq \sigma^2_2$ in terms of n for $m = 10000$ replications of the distribution $\mathcal{E}(1)$.

	$\chi^2(5)$			
n	$\mathcal{F}$	asympTest	Ansari	Mood
30	0.1605	0.0676	0.0477	0.0472
100	0.1797	0.0537	0.0516	0.0494
500	0.1911	0.0525	0.0505	0.0498
1000	0.1907	0.0526	0.0503	0.0511

Table 7: Type I error for the test $H_1 : \sigma^2_1 \neq \sigma^2_2$ in terms of n for $m = 10000$ replications of the distribution $\chi^2(5)$.

	$\mathcal{U}([0,5])$			
n	$\mathcal{F}$	asympTest	Ansari	Mood
30	0.0029	0.0652	0.0490	0.0494
100	0.0021	0.0527	0.0490	0.0475
500	0.0024	0.0520	0.0511	0.0511
1000	0.0022	0.0539	0.0528	0.0538

Table 8: Type I error for the test $H_1 : \sigma^2_1 \neq \sigma^2_2$ in terms of n for $m = 10000$ replications of the distribution $\mathcal{U}([0,5])$.

Using asympTest

The R package **asympTest** consists of a main function `asymp.test` and six auxiliary ones designed to compute standard errors of estimates of different parameters, see Table 4. The auxiliary functions will not be the most useful ones for the user, except if he/she wants to compute the confidence interval himself/herself. The function `asymp.test` has been written in the same spirit as the standard R functions `t.test` or `var.test`. The arguments of `asymp.test` and the resulting outputs are also inspired from these functions. In particular, the function `asympt.test` returns an object of class "htest" (which is the general class of test objects in R).

This `asymp.test` function has several arguments, similar to those of the `t.test` function, whose description can be obtained using the command `?asymp.test`.

In order to illustrate this function, let us consider the Digitalis Investigation Group NHLBI Teaching data set (`https://biolincc.nhlbi.nih.gov/teaching/`) which was made available by the NHLBI. Note that statistical processes such as permutations within treatment groups were used to

completely anonymize the data; therefore, inferences derived from the teaching dataset may not be valid.

The DIG Trial was a randomized, double-blind, multicenter trial with more than 300 centers in the United States and Canada participating. The purpose of the trial was to examine the safety and efficacy of Digoxin in treating patients with congestive heart failure in sinus rhythm.

Diastolic BP (`DIABP`, mmHg) is a known risk factor of cardiovascular diseases. In this case, it is desirable to compare the variability of this quantity for placebo (`TRTMT=0`) and treatment (`TRTMT=1`) groups, respectively.

Reading of the data

```
> require(asympTest)
>
> data(DIGdata)
> attach(DIGdata)
> x <- na.omit(DIABP[TRTMT==0])
> y <- na.omit(DIABP[TRTMT==1])
> c(length(x),length(y))
[1] 3400 3395
```

Comparing the two variances

Shapiro-Wilk normality test performed by the function `shapiro.test()` indicates that the two samples seem to be far from the Gaussian distribution. Thus, this should prevent us from using the following Fisher test.

```
> var.test(DIABP ~ TRTMT, data = DIGdata,
+    na.action = na.omit)

        F test to compare two variances

data:  x and y
F = 0.9295, num df = 3399, denom df = 3394
p-value = 0.03328
alternative hypothesis:
   true ratio of variances is not equal to 1
95 percent confidence interval:
 0.8690651 0.9942238
sample estimates:
ratio of variances
          0.929541
```

Instead, let us use our package.

```
> asymp.test(DIABP ~ TRTMT, data = DIGdata,
+    na.action = na.omit, parameter = "dVar")

Two-sample asymptotic diff. of variances test

data:  DIABP by TRTMT
statistic = -1.5272, p-value = 0.1267
alternative hypothesis:
  true diff. of variances is not equal to 0
95 percent confidence interval:
 -21.160491   2.626127
sample estimates:
difference of variances
             -9.267182
```

We can see that `var.test`, not to be used due to the unlikely normality of the data, significantly shows a difference in variances (at a 5% level). We don't obtain the same conclusion with our test.

We can also place ourselves in a fictitious case by generating a sample `x` from a $\mathcal{U}(0;\sqrt{12})$ (i.e. with a true population variance $\sigma^2 = 1$). We then apply both our test and the classical chi-square test to show $H_1 : \sigma^2 > \sigma^2_{ref} = 0.97$.

```
> n <- 1000
> x <- runif(n, max = sqrt(12))
> asymp.test(x, par = "var", alt = "gr",
+    ref = 0.97)

        One-sample asymptotic variance test

data:  x
statistic = 1.753, p-value = 0.0398
alternative hypothesis:
         true variance is greater than 0.97
95 percent confidence interval:
 0.9731491       Inf
sample estimates:
variance
1.021055
> chisq.stat <- (n-1)*var(x)/0.97
> pchisq(chisq.stat, n-1, lower.tail = F)
[1] 0.1207650
```

For the above generated sample `x`, we respectively found the following p-values: 0.0398 and 0.120. In this case, we can thus see that our proposition correctly accepts H_1 (at the 5% level) but not the chi-square single variance test.

Conclusion

This paper has introduced a new package called **asympTest**. This is a contribution to the many R procedures available. It is interesting firstly in the fact that it provides a unified teaching framework to present classical parametric tests (based on the Central Limit Theorem). These tests are made readily available in R through an easy to use function called `asymp.test`. This function resembles `t.test` or `var.test`, so students will not be confused. Secondly, it also makes available in R a robust (to non-normality) alternative to the classical chi-square single variance test. In the future, we also plan to provide tools similar to the `power.t.test` function in the context of large samples.

Bibliography

D. G. Bonnet. Approximate confidence interval for standard deviation of nonnormal distributions. *Computational Statistics & Data Analysis*, 50:775–782, 2006.

D. G. Bonnet. Robust confidence interval for a ratio of standard deviations. *Applied Psychological Measurement*, 30:(5) 432–439, 2006.

G. E. P. Box. Non-normality and tests on variances. *Biometrika*, 40:(3/4) 318–335, 1953.

G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, Belmont, California, 2nd edition, 2001.

J.-F. Coeurjolly, R. Drouilhet, P. Lafaye de Micheaux and J.-F. Robineau. asympTest: an R package for performing parametric statistical tests and confidence intervals based on the central limit theorem. Technical report, URL http://www.citebase.org/abstract?id=oai:arXiv.org:0902.0506.

W. J. Conover, M. E. Johnson and M. M. Johnson. A comparative study of tests for homogeneity of variances with applications to the outer continental shelf bidding data. *Technometrics*, 23:(4) 351–361, 1981.

R. Davidson and J. G. MacKinnon. Graphical methods for investigating the size and power of hypothesis tests. *The Manchester School*, 66:(1) 1–26, 1998.

S. Dean and B. Illowsky. F Distribution and ANOVA: Test of Two Variances. *Connexions*, February 5, 2009. URL http://cnx.org/content/m17075/1.6/

T. Ferguson. A course in large sample theory. *Chapman and Hall*, 1996.

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:(Part II) 179–188, 1935.

R. G. Miller. *Beyond ANOVA, Basics of Applied Statistics*. Texts in Statistical Science Series. Chapman & Hall/CRC, 1997.

C. Ozgur and S. E. Strasser. A study of the statistical inference criteria: Can we agree on when to use z versus t? *Decision Sciences Journal of Innovative Education*, 2:(2) 177–192, 2004.

G. Pan. On a Levene type test for equality of two variances. *Journal of Statistical Computation and Simulation*, 63:59–71, 1999.

M. L. Tiku and A. Akkaya. *Robust Estimation and Hypothesis Testing*. New Age International (P) Ltd, New Delhi, 2004.

Jean-François CoeurJolly
L. Jean Kuntzmann, Grenoble University, France
http://cqls.upmf-grenoble.fr
Jean-Francois.Coeurjolly@upmf-grenoble.fr

Rémy Drouilhet
L. Jean Kuntzmann
Grenoble University
France
http://cqls.upmf-grenoble.fr
Remy.Drouilhet@upmf-grenoble.fr

Pierre Lafaye de Micheaux
Département de Mathématiques et Statistique
Université de Montréal
Canada
http://www.biostatisticien.eu
lafaye@dms.umontreal.ca

Jean-François Robineau
CQLS
http://www.cqls.org
robineau@cqls.fr

asympTest: AN R PACKAGE FOR PERFORMING PARAMETRIC STATISTICAL TESTS AND CONFIDENCE INTERVALS BASED ON THE CENTRAL LIMIT THEOREM

J.-F. Coeurjolly[1], R. Drouilhet[1], P. Lafaye de Micheaux[1] and J.-F. Robineau[2]
[1] SAGAG Team, Laboratory Jean Kuntzmann/ CNRS, Grenoble (France), [2] CQLS, France

February 27, 2009

Abstract

This paper describes an R package implementing large sample tests and confidence intervals (based on the central limit theorem) for various parameters. The one and two sample mean and variance contexts are considered. The statistics for all the tests are expressed in the same form, which facilitates their presentation. In the variance parameter cases, the asymptotic robustness of the classical tests depends on the departure of the data distribution from normality measured in terms of the kurtosis of the distribution.

Keywords: parametric tests and confidence intervals, central limit theorem, R package

1 Introduction

When you are interested in testing a variance parameter for a large sample in the non Gaussian framework, it is not easy to find a test implemented in the standard statistical software. In fact, we could not find one! The only available tool is the chi-square variance test tailor-made to the Gaussian context. This test is however commonly used by practicians even if the Gaussian assumption fails. We applied it to a data set of size $n = 1000$ with an empirical distribution very different from a normal distribution. The p-value of the chi-square variance test with alternative hypothesis $\mathbf{H_1}$: $\sigma^2 < 1$ is 4.79% which leads us to accept the alternative hypothesis at level $\alpha = 5\%$. Is it reasonable to use this test when we know that it cannot be used in the non Gaussian framework because of its sensitivity to departures from normality, *e.g.* Box (1953)?

From a mathematical point of view, one may wonder why no alternative test has been yet implemented since the asymptotic properties of the sample variance are well-known. Things are much easier in the sample mean study because the one sample t-test is known to be robust to departures from normality for large samples, *e.g.* Ozgur and Strasser (2004). This results from a direct application of the central limit theorem. The same remarks are valid when comparing the two-sample t-test (difference of means test) which is robust for large samples and the Fisher test (ratio of variances test) which is not.

In the statistical framework, one may be a simple user, a tool developer, a theoretician or any combination of the above. There are natural interactions between these different communities and it is expected that their knowledge should be shared, above all for tasks that have now become very basic. It is well-known for a theoretician, see *e.g.* Casella and Berger (1990), that a

common method for constructing a large sample test statistic may be based on an estimator that has an asymptotic normal distribution. Suppose we wish to test a hypothesis about a parameter θ, and $\widehat{\theta}_n$ is some estimator of θ based on a sample of size n. If we can prove some form of the central limit theorem to show that as $n \to +\infty$, $(\widehat{\theta}_n - \theta)/\widehat{\sigma}_{\widehat{\theta}} \xrightarrow{d} \mathcal{N}(0,1)$ where $\widehat{\sigma}^2_{\widehat{\theta}}$ is a convergent (in probability) estimate of $Var(\widehat{\theta}_n)$, then one has the basis for an approximate test. This scheme based on the central limit theorem will be called the CLT procedure. We have already specified that we did not find any alternative to the chi-square variance test for testing a variance when the normality assumption fails. On the contrary, the problem of the robustness (to departures from normality) of tests for comparing two (or more) variances has been widely treated in the literature, see *e.g.* Box (1953), Conover et al. (1981), Tiku et al. (1986), Pan (1999) and the references therein. Some alternative procedures to the Fisher test are implemented in `R`: the Bartlett test (`bartlett.test`), the Fligner test (`fligner.test`), the Levene test (`levene.test` available in the `lawstat` package), etc. However to our best knowledge, for large samples, simple alternatives based on the CLT procedure have never been proposed or implemented.

The main objective of this paper is to propose a unified framework, based on the CLT procedure, for large samples to test various parameters such as the mean, the variance, the difference or ratio of means or variances. This approach also allows direct derivation of asymptotic confidence intervals. Tests and confidence intervals are then implemented in our new `R` package, called `asympTest`. This modest contribution also solves the problem of finding a robust (to non-normality) alternative to the chi-square variance test for large samples. It also provides a very simple alternative to the Fisher test. However, note that the purpose of this paper is not to compare our tests to their competitors in terms of power. Finally, a first course of statistical inference usually presents mean tests in both Gaussian and asymptotical frameworks and variance tests restricted to the Gaussian case. The unified approach presented here is is very similar to the classical t-test from a mathematical point of view and gives us the opportunity to propose a more complete teaching framework with no additional difficulty.

The paper is organized as follows. Section 2 deals with the mathematical concepts and describes our main notation. We also propose a mathematical explanation of the reason why the chi-square variance test and the Fisher test are not appropriate even when the sample size is very large. Finally, a general framework is also proposed that allows us to derive some asymptotic statistical tests for the mean, the variance and the difference (and ratio) of means or variances. In Section 3, the `R` package `asympTest` is presented. It notably includes the procedures described in the previous section. Finally, Sections 4, 5 and 6 are devoted to some discussions and the proofs of our results.

2 Mathematical development

2.1 Notation

For one-sample tests, let us denote by $\mathbf{Y} = (Y_1, \ldots, Y_n)$ a sample of n independent and identically distributed random variables with mean μ and variance σ^2. These parameters are classically estimated by

$$\widehat{\mu}(\mathbf{Y}) = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad \widehat{\sigma^2}(\mathbf{Y}) = \frac{1}{n-1}\sum_{i=1}^{n} (Y_i - \widehat{\mu}(\mathbf{Y}))^2 .$$

In the two-sample context, let $\mathbf{Y^{(1)}} = \left(Y_1^{(1)}, \ldots, Y_{n^{(1)}}^{(1)}\right)$ and $\mathbf{Y^{(2)}} = \left(Y_1^{(2)}, \ldots, Y_{n^{(2)}}^{(2)}\right)$ denote two independent samples of $n^{(1)}$ and $n^{(2)}$ random variables with respective means $\mu^{(1)}$ and $\mu^{(2)}$ and variances $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$. We also define the following parameters and their estimated versions by denoting $\mathbf{Y} = \left(\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}\right)$:

- Difference of (weighted) means: $d_\mu = \mu^{(1)} - \rho \times \mu^{(2)}$ ($\rho \in \mathbb{R}$) and $\widehat{d_\mu}\left(\mathbf{Y}\right) = \widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right) - \rho \times \widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)$.
- Difference of (weighted) variances: $d_{\sigma^2} = \sigma^2_{(1)} - \rho \times \sigma^2_{(2)}$ ($\rho \in \mathbb{R}$) and $\widehat{d_{\sigma^2}}\left(\mathbf{Y}\right) = \widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right) - \rho \times \widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)$.
- Ratio of means: $r_\mu = \frac{\mu^{(1)}}{\mu^{(2)}}$ and $\widehat{r_\mu}\left(\mathbf{Y}\right) = \frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)}$.
- Ratio of variances: $r_{\sigma^2} = \frac{\sigma^2_{(1)}}{\sigma^2_{(2)}}$ and $\widehat{r_{\sigma^2}}\left(\mathbf{Y}\right) = \frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}$.

The known parameter ρ is, in our definition of d_μ and d_{σ^2}, intrinsically nonnegative. But note that there is no mathematical problem to deal with negative values of ρ.

In order to compare the Gaussian framework and the general one, we propose to denote by $\mathbf{Y}^G$ (resp. $\mathbf{Y}^{(1),G}$, $\mathbf{Y}^{(2),G}$) a vector (resp. two independent vectors) of n (resp. $n^{(1)}$ and $n^{(2)}$) Gaussian random variables with mean μ (resp. $\mu^{(1)}$ and $\mu^{(2)}$) and variance σ^2 (resp. $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$). In the two-sample context, let us also denote $\mathbf{Y}^G = \left(\mathbf{Y}^{(1),G}, \mathbf{Y}^{(2),G}\right)$.

In the sequel, we will use the notation $:=$ to define some quantity. For some random variable Z and some distribution $\mathcal{L}$, $Z \rightsquigarrow \mathcal{L}$ (resp. $\overset{approx}{\rightsquigarrow} \mathcal{L}$) means that Z follows (resp. approximately follows) the distribution $\mathcal{L}$.

2.2 About the Chi-square and Fisher tests

In this section, we concentrate on parameters σ^2 and r_{σ^2}. The classical statistics of the chi-square test of variance and the Fisher test of ratio of variances are defined by

$$\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}) := (n-1)\frac{\widehat{\sigma^2}\left(\mathbf{Y}\right)}{\sigma^2} \qquad \text{and} \qquad \Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}) := \frac{\widehat{r_{\sigma^2}}\left(\mathbf{Y^{(1)}}, \mathbf{Y^{(2)}}\right)}{r_{\sigma^2}}.$$

The notation $\Lambda^G_{\widehat{\sigma^2},\sigma^2}$ expresses that the statistic is a measure of the departure of $\widehat{\sigma^2}$ from σ^2 in the Gaussian framework. When the data are Gaussian, it is well-known that

$$\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}^G) \rightsquigarrow \chi^2(n-1) \qquad \text{and} \qquad \Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}^G) \rightsquigarrow \mathcal{F}\left(n^{(1)}-1, n^{(2)}-1\right).$$

However, as shown in Fig. 1, both results become untrue (even approximately) under non-normality assumption. The theoretical reason may be explained as follows. Let $n^{(1)} = n$ and assume that there exists $\alpha > 0$ such that $n^{(2)} = \alpha n^{(1)}$. Assume also that Y (resp. $Y^{(1)}$ and $Y^{(2)}$) has a finite kurtosis $k := E((Y-\mu)^4)/Var(Y)$ (resp. $k^{(1)}$ and $k^{(2)}$). Then, as $n \to +\infty$,

$$\frac{\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}) - (n-1)}{\sqrt{n-1}} \quad \xrightarrow{d} \quad \mathcal{N}(0, k-1) \tag{1}$$

$$\sqrt{n}\left(\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}) - 1\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, k^{(1)} - 1 + \frac{k^{(2)}-1}{\alpha}\right). \tag{2}$$

This result is a consequence of a more general one stated in Section 2.3 and proved in Section 6. Equations (1) and (2) lead to the following approximations

$$\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}) \overset{approx}{\rightsquigarrow} \mathcal{N}\left(n-1,(k-1)(n-1)\right) \quad \text{and} \quad \Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}) \overset{approx}{\rightsquigarrow} \mathcal{N}\left(1,\frac{k^{(1)}-1}{n^{(1)}}+\frac{k^{(2)}-1}{n^{(2)}}\right).$$

When data are Gaussian ans when n is large, one obtains the well-known approximations

$$\chi^2(n-1) \simeq \mathcal{N}\left(n-1,2(n-1)\right) \qquad \text{and} \qquad \mathcal{F}\left(n^{(1)},n^{(2)}\right) \simeq \mathcal{N}\left(1,\frac{2}{n^{(1)}}+\frac{2}{n^{(2)}}\right)$$

since $k=k^{(1)}=k^{(2)}=3$ We can underline that $\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y})$ (resp. $\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y})$) and $\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}^G)$ (resp. $\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}^G)$) differ in terms of asymptotical variances. More precisely, the gap between the two frameworks is essentially governed by the kurtosis. Indeed, as $n \to +\infty$

$$\frac{Var(\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}))}{Var(\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}^G))} \to \frac{k-1}{2} \qquad \text{and} \qquad \frac{Var(\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}))}{Var(\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y}^G))} \to \frac{k^{(1)}-1+\frac{k^{(2)}-1}{\alpha}}{2\left(1+\frac{1}{\alpha}\right)}.$$

Tab. 1 proposes the computations of these asymptotic ratios for different distributions. This allows the reader to assess the risk of using the classical statistics $\Lambda^G_{\widehat{\sigma^2},\sigma^2}$ and $\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}$ under the non-normality assumption even when the size of the sample is large.

	$Y \overset{d}{=} Y^{(1)} \overset{d}{=} Y^{(2)} \rightsquigarrow \mathcal{L}$		
Test	$\mathcal{L}=\chi^2(\nu)$	$\mathcal{L}=\mathcal{E}(\lambda)$	$\mathcal{L}=\mathcal{U}([a,b])$
Variance test	$1+\frac{6}{\nu}$	4	$\frac{2}{5}$
Ratio of variances test	$1+\frac{6}{\nu}$	4	$\frac{2}{5}$

Table 1: Ratio of asymptotic variances (non gaussian/gaussian) $\frac{k-1}{2}$ and $\frac{k^{(1)}-1+\frac{k^{(2)}-1}{\alpha}}{2\left(1+\frac{1}{\alpha}\right)}$ in the case where $k=k^{(1)}=k^{(2)}$ and $\alpha=1$.

2.3 Large sample tests based on the central limit theorem

The parameters σ^2, d_μ and d_{σ^2} can be viewed as particular means. Therefore, the idea (widely used in asymptotic theory) is to design asymptotic tests and confidence intervals thanks to the central limit theorem. The variables $\widehat{r_\mu}(\mathbf{Y})-r_\mu$ and $\widehat{r_{\sigma^2}}(\mathbf{Y})-r_{\sigma^2}$ can also be expressed in terms of means to which a central limit theorem can be applied. In order to unify asymptotic results, we define θ as one of the parameters μ, σ^2, d_μ, d_{σ^2}, r_μ and r_{σ^2}. By applying a central limit theorem, the law of large numbers and Slutsky's theorem (see Section 6), one obtains, as $n \to +\infty$,

$$\widehat{\Delta}_{\widehat{\theta},\theta}(\mathbf{Y}) := \frac{\widehat{\theta}(\mathbf{Y})-\theta}{\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})} \xrightarrow{d} \mathcal{N}(0,1), \tag{3}$$

where $\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})$ is the standard error of $\widehat{\theta}(\mathbf{Y})$. The assumptions under which the central limit theorem can be applied and the definition of $\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})$ are stated in Tab. 2. `R` functions have been implemented to evaluate the standard errors of the estimates of θ, see Tab. 3.

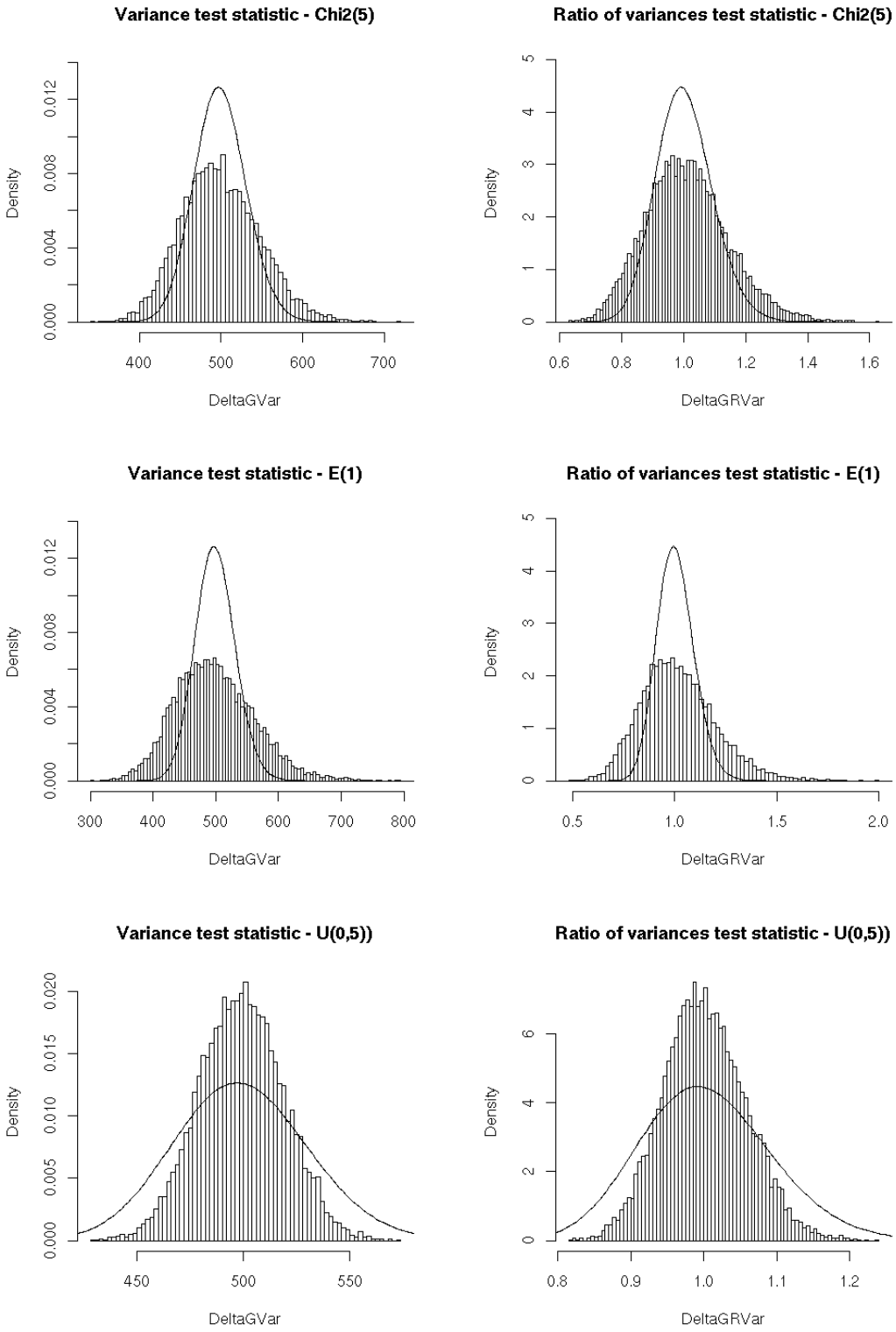


Figure 1: Histograms of $m = 10000$ replications of test statistics of variance test (left) and ratio of variance tests (right) in the Gaussian context. The simulation has been done as follows: $n = n^{(1)} = n^{(2)} = 500$, $Y \overset{d}{=} Y^{(1)} \overset{d}{=} Y^{(2)} \rightsquigarrow \chi^2(5)$ (top), $\mathcal{E}(1)$ (middle) and $\mathcal{U}([0,5])$ (bottom).

θ	Assumptions	$\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})$
μ	$E\left(Y_i^2\right) < +\infty$	$\sqrt{\dfrac{\widehat{\sigma^2}(\mathbf{Y})}{n}}$
σ^2	$E\left(Y_i^4\right) < +\infty$	$\sqrt{\dfrac{\widehat{\sigma^2_{\ddot{Y}}}\left(\ddot{\mathbf{Y}}\right)}{n}}$
d_μ	$E\left(\left(Y_i^{(j)}\right)^2\right) < +\infty,\ j = 1,2$	$\sqrt{\dfrac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{n^{(1)}} + \rho^2 \times \dfrac{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n^{(2)}}}$
d_{σ^2}	$E\left(\left(Y_i^{(j)}\right)^4\right) < +\infty,\ j = 1,2$	$\sqrt{\dfrac{\widehat{\sigma^2_{\ddot{Y}^{(1)}}}\left(\ddot{\mathbf{Y}}^{\mathbf{(1)}}\right)}{n^{(1)}} + \rho^2 \times \dfrac{\widehat{\sigma^2_{\ddot{Y}^{(2)}}}\left(\ddot{\mathbf{Y}}^{\mathbf{(2)}}\right)}{n^{(2)}}}$
r_μ	$\mu^{(2)} \neq 0,\ E\left(\left(Y_i^{(j)}\right)^2\right) < +\infty,\ j = 1,2$	$\dfrac{1}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)}\sqrt{\dfrac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{n^{(1)}} + \widehat{r_\mu}(\mathbf{Y})^2 \times \dfrac{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n^{(2)}}}$
r_{σ^2}	$\sigma^2_{(2)} \neq 0,\ E\left(\left(Y_i^{(j)}\right)^4\right) < +\infty,\ j = 1,2$	$\dfrac{1}{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}\sqrt{\dfrac{\widehat{\sigma^2_{\ddot{Y}^{(1)}}}\left(\ddot{\mathbf{Y}}^{\mathbf{(1)}}\right)}{n^{(1)}} + \widehat{r_{\sigma^2}}(\mathbf{Y})^2 \dfrac{\widehat{\sigma^2_{\ddot{Y}^{(2)}}}\left(\ddot{\mathbf{Y}}^{\mathbf{(2)}}\right)}{n^{(2)}}}$

Table 2: Standard errors of estimates of θ. For the sake of simplicity we denote by $\ddot{Y} := (Y-\mu)^2$, $\ddot{\mathbf{Y}} := (\mathbf{Y} - \widehat{\mu}(\mathbf{Y}))^2$, $\sigma^2_{\ddot{Y}} := Var\left(\ddot{Y}\right)$ and $\widehat{\sigma^2_{\ddot{Y}}}\left(\ddot{\mathbf{Y}}\right) := \frac{1}{n-1}\sum_{i=1}^n \left((Y_i - \widehat{\mu}(\mathbf{Y}))^2 - \widehat{\sigma^2}(\mathbf{Y})\right)^2$.

θ	Dataset(s)	$\widehat{\sigma_{\widehat{\theta}}}(\mathbf{y})$ in `R`
μ	`y`	`seMean(y)`
σ^2	`y`	`seVar(y)`
d_μ	`y1, y2`	`seDMean(y1,y2,rho=1)`
d_{σ^2}	`y1, y2`	`seDVar(y1,y2,rho=1)`
r_μ	`y1, y2`	`seRMean(y1,y2)`
r_{σ^2}	`y1, y2`	`seRVar(y1,y2)`

Table 3: Standard errors of estimates of θ in `R`.

Remark 1 *The figures Fig. 2, Fig. 3 and Fig. 4 allow the reader to illustrate the mathematical result (3).*

Remark 2 *The asymptotic result (3) allows us to easily construct statistical hypothesis tests and confidence intervals, see e.g. Casella and Berger (1990) p. 385 for development.*

Remark 3 *The alternative hypothesis* $\mathbf{H_1}$ *comparing the ratio of means or variances to some reference value* r_0 *may be expressed in terms of the comparison of the weighted differences* $d_\mu := \mu^{(1)} - r_0\mu^{(2)}$ *or* $d_{\sigma^2} := \sigma^2_{(1)} - r_0\sigma^2_{(2)}$ *to 0, the value of* ρ *being fixed to* r_0*.*

3 Using asympTest

The `R` package `asympTest` consists of a main function `asymp.test` and six auxiliary ones designed to compute standard errors of estimates of different parameters, see Tab. 3. The auxiliary functions should not be very useful for the user, except if he/she wants to compute himself/herself the confidence interval. The function `asymp.test` has been written in the same spirit as the `R` functions `t.test` or `var.test`. The arguments of `asymp.test`, its value and the resulting outputs are inspired from the ones of `t.test` or `var.test`. In particular, the function `asympt.test` returns an object of class "htest" (which is the general class of test objects in `R`, see R Development Core Team (2004)).

The main arguments of the function `asymp.test` are:

- `x`: vector of data values.
- `y`: optional vector of data values.
- `parameter`: parameter under testing, must be one of "mean", "var", "dMean", "dVar", "rMean", "rVar".
- `alternative`: alternative hypothesis, must be one of "two.sided" (default), "greater" or "less".
- `reference`: reference value of the parameter under the null hypothesis.
- `conf.level`: confidence level of the interval (default is 0.95). The type-one error is then fixed to 1-`conf.level`.
- `rho`: optional parameter (only used for parameters "dMean" and "dVar") for penalization (or enhancement) of the contribution of the second parameter.

The user may only specify the first letters of the parameter or alternative.

In order to illustrate this function, let us consider the `iris` data available in `R`. This famous (Fisher's or Anderson's Fisher (1935); Anderson (1935)) data set gives the measurements (in centimeters) of the four variables sepal length and width, and petal length and width, for 50 flowers from each species of iris: setosa, versicolor, and virginica.

```
> data(iris)
> attach(iris)
> names(iris)
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"  "Species"
> levels(iris$Species)
[1] "setosa"     "versicolor" "virginica"
```

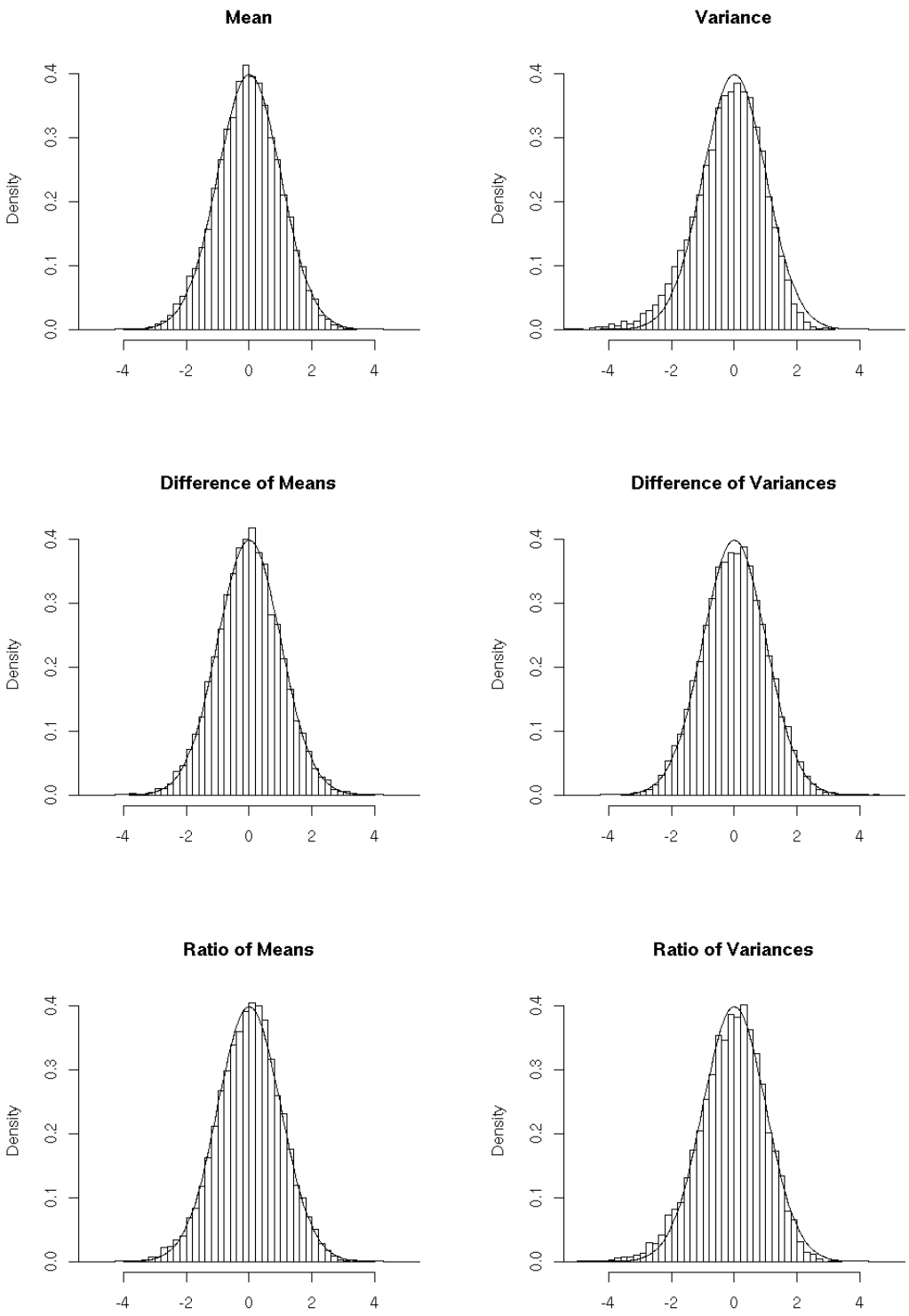


Figure 2: Histograms of $m = 10000$ replications of $\frac{\widehat{\theta}(\mathbf{Y})-\theta}{\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})}$ for $\theta = \mu$, σ^2, d_μ, d_{σ^2}, r_μ and r_{σ^2}. The simulation has been done as follows: $n = n^{(1)} = n^{(2)} = 500$, $Y^{(1)} \rightsquigarrow \chi^2(5)$, $Y^{(2)} \rightsquigarrow \chi^2(5)$.

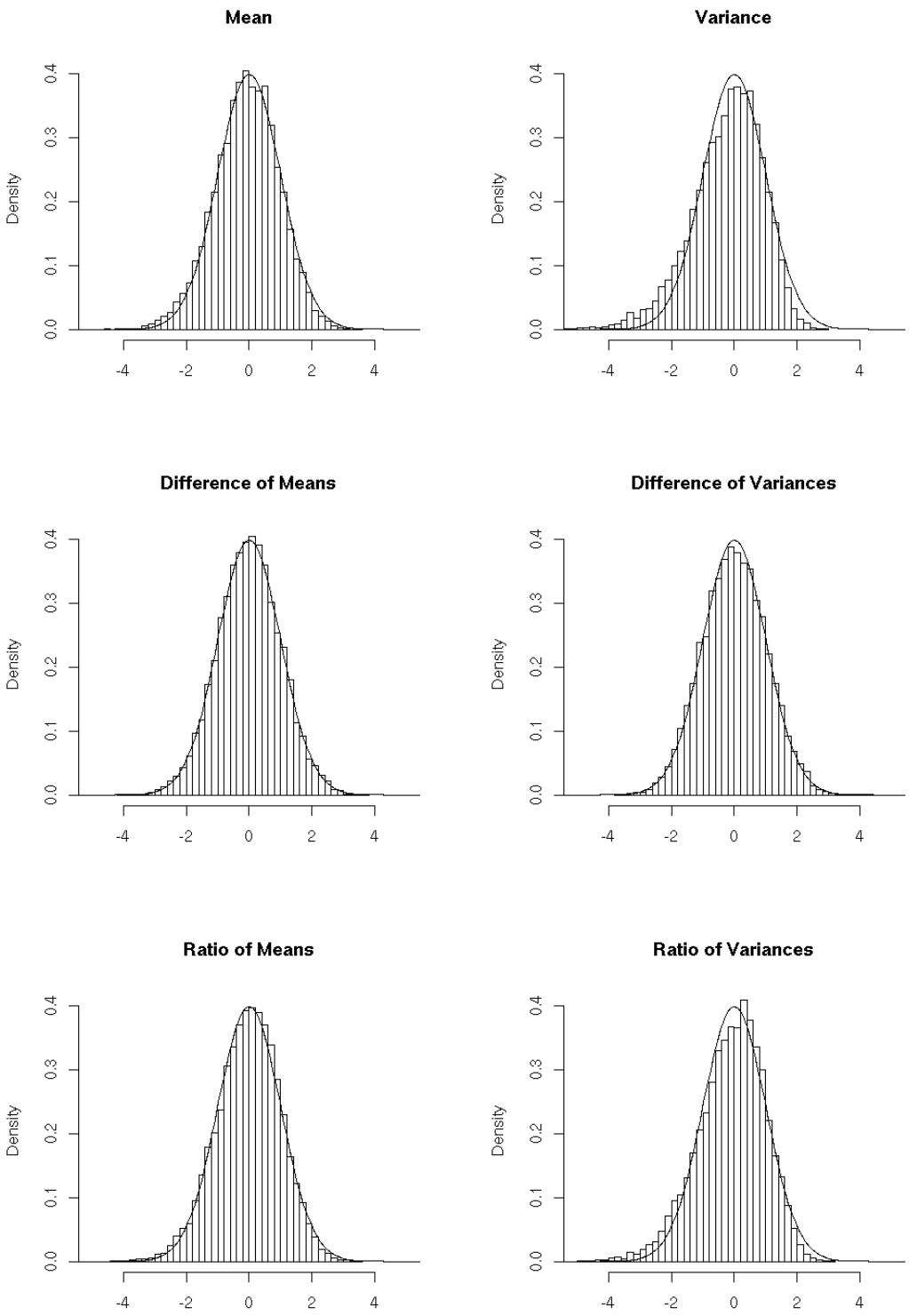


Figure 3: Histograms of $m = 10000$ replications of $\frac{\widehat{\theta}(\mathbf{Y})-\theta}{\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})}$ for $\theta = \mu$, σ^2, d_μ, d_{σ^2}, r_μ and r_{σ^2}. The simulation has been done as follows: $n = n^{(1)} = n^{(2)} = 500$, $Y^{(1)} \rightsquigarrow \mathcal{E}(1)$, $Y^{(2)} \rightsquigarrow \mathcal{E}(1)$.

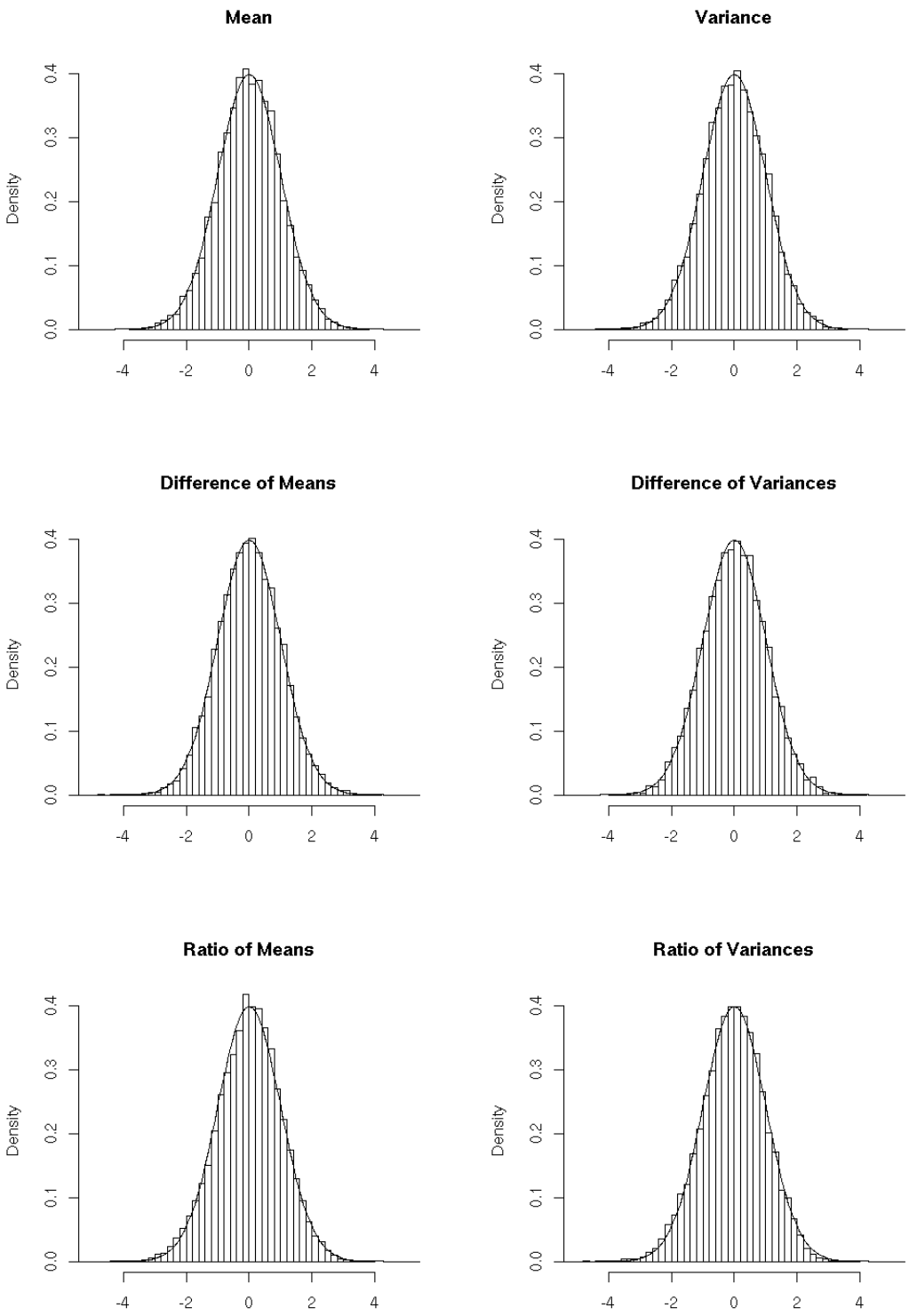


Figure 4: Histograms of $m = 10000$ replications of $\frac{\widehat{\theta}(\mathbf{Y})-\theta}{\widehat{\sigma_{\widehat{\theta}}}(\mathbf{Y})}$ for $\theta = \mu$, σ^2, d_μ, d_{σ^2}, r_μ and r_{σ^2}. The simulation has been done as follows: $n = n^{(1)} = n^{(2)} = 500$, $Y^{(1)} \rightsquigarrow \mathcal{U}([0,5])$, $Y^{(2)} \rightsquigarrow \mathcal{E}(0.5)$.

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
setosa	0.4595	0.2715	0.0548	0.0000
versicolor	0.4647	0.3380	0.1585	0.0273
virginica	0.2583	0.1809	0.1098	0.0870

The following table presents the p-values of all the Shapiro-Wilk normality tests for the different variables and the three species.

Let us concentrate on the variable Petal.Width for which the Gaussian assumption seems to be wrong for each one of the three species. The empirical means and variances are given below:

```
> by(Petal.Width,Species,function(e) c(mean=mean(e),var=var(e)))
Species: setosa
      mean        var
0.24600000 0.01110612
------------------------------------------------------------
Species: versicolor
      mean        var
1.32600000 0.03910612
------------------------------------------------------------
Species: virginica
      mean        var
2.02600000 0.07543265
```

Is the mean petal width of setosa species less than 0.5 ?

```
> require(asympTest)
> asymp.test(Petal.Width[Species=="setosa"],par="mean",alt="l",ref=0.5)

        One-sample asymptotic mean test

data:  Petal.Width[Species == "setosa"]
statistic = -17.0427, p-value < 2.2e-16
alternative hypothesis: true mean is less than 0.5
95 percent confidence interval:
      -Inf 0.2705145
sample estimates:
 mean
0.246
```

Is the mean petal width of virginica species larger than the versicolor one ?

```
> asymp.test(Petal.Width[Species=="virginica"],
+ Petal.Width[Species=="versicolor"],"dMean","g",0)

        Two-sample asymptotic difference of means test

data:  Petal.Width[Species == "virginica"] and Petal.Width[Species == "versicolor"]
statistic = 14.6254, p-value < 2.2e-16
alternative hypothesis: true difference of means is greater than 0
```

```
95 percent confidence interval:
 0.621274      Inf
sample estimates:
difference of means
                0.7
```

Is the mean petal width of virginica species 4 times larger than the setosa one ?

```
> asymp.test(Petal.Width[Species=="virginica"],
+ Petal.Width[Species=="setosa"],"rMean","g",4)

        Two-sample asymptotic ratio of means test

data:  Petal.Width[Species == "virginica"] and Petal.Width[Species == "setosa"]
statistic = 8.0936, p-value = 3.331e-16
alternative hypothesis: true ratio of means is greater than 4
95 percent confidence interval:
 7.374946      Inf
sample estimates:
ratio of means
      8.235772
```

This may also be done via a difference of weighted means test.

```
> asymp.test(Petal.Width[Species=="virginica"],
+ Petal.Width[Species=="setosa"],"dMean","g",0,rho=4)

        Two-sample asymptotic difference of (weighted) means test

data:  Petal.Width[Species == "virginica"] and Petal.Width[Species == "setosa"]
statistic = 14.6447, p-value < 2.2e-16
alternative hypothesis: true difference of (weighted) means is greater than 0
95 percent confidence interval:
 0.9249653       Inf
sample estimates:
difference of (weighted) means
                         1.042
```

4 Type I error risks

4.1 Comparison between classical and asymptotic variance tests

In the context of large samples, two simulation studies are proposed in order to show the lack of reliability of the classical tests for variance parameters compared with the asymptotic tests studied in this paper. For each of the following examples, 10000 simulations of samples of size $n = 1000$ have been performed.

1. Let us consider testing $\mathbf{H_0} : \sigma^2 = 1$ versus $\mathbf{H_1} : \sigma^2 < 1$ with data sampled from distribution $\mathcal{E}(1)$ (i.e., under $\mathbf{H_0}$).

   In the case where $\alpha = 5\%$, the probability of accepting the alternative hypothesis is 8.91% for the asymptotic test and 20.99% for the chi-square test.

	false	true
false	0.7901	0.0000
true	0.1208	0.0891

Table 4: Acceptance of $\mathbf{H_1}$ for the chi-square test (rows) versus the asymptotic test (columns)

2. Let us consider test $\mathbf{H_0} : \sigma^2_{(1)} = \sigma^2_{(2)}$ versus $\mathbf{H_1} : \sigma^2_{(1)} \neq \sigma^2_{(2)}$ with data sampled from distribution $\mathcal{U}([0, 5])$ for both samples (i.e., under $\mathbf{H_0}$).

	false	true
false	0.9498	0.0481
true	0.0000	0.0021

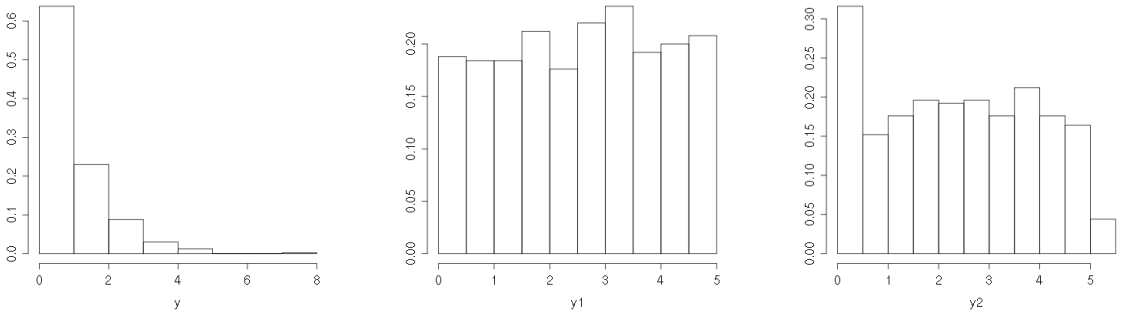
Table 5: Acceptance of $\mathbf{H_1}$ for the Fisher test (rows) versus the asymptotic test (columns)

In the case where $\alpha = 5\%$, the probability of accepting the alternative hypothesis is 5.02% for the asymptotic test and 0.21% for the Fisher test.

In both cases, the probabilities for Type I errors are worse for the classical tests than for the corresponding well-suited asymptotic tests. This is a direct consequence of the previous results summarized by figure Fig. 1.

4.2 Back to the example of the introduction

Now, one may wonder what the consequences of these previous results are for pratical purposes. Let us consider again the example presented in the introduction. In order to illustrate the two-samples case, we propose a second example. In the following `R` outputs, the data of the first (resp. second) example are denoted by `y` (resp. `y1` and `y2`). These samples have a size n=1000 and their empirical distributions, proposed below, do not seem to fit normal distributions.



The following output provides the p-value of the chi-square test for the first example.

```
> pchisq((length(y)-1)*var(y)/1,length(y)-1)
[1] 0.04785152
```

Due to the apparent non normality of the data, one may prefer to apply the corresponding asymptotic test:

```
> asymp.test(y,par="var",alt="l",ref=1)

        One-sample asymptotic variance test

data:  y
statistic = -0.9771, p-value = 0.1643
alternative hypothesis: true variance is less than 1
95 percent confidence interval:
     -Inf 1.070430
sample estimates:
 variance
0.8969455
```

The two decisions do not match, but since the empirical variance is 0.8969455, one may think that σ^2 is slightly inferior to 1. In which case, we have to be cautious because our sample might be of the same kind as the 12.08% of Table 4.

The following output provides the p-value of the Fisher test for the second example.

```
> var.test(y1,y2)

        F test to compare two variances

data:  y1 and y2
F = 0.8874, num df = 499, denom df = 499, p-value = 0.1825
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7444324 1.0578390
sample estimates:
ratio of variances
         0.8874061
```

Due to the apparent non normality of the data, one may prefer to apply the corresponding asymptotic test:

```
> asymp.test(y1,y2,"dVar")

        Two-sample asymptotic difference of variances test

data:  y1 and y2
statistic = -2.0925, p-value = 0.03639
alternative hypothesis: true difference of variances is not equal to 0
95 percent confidence interval:
 -0.49988046 -0.01635305
sample estimates:
difference of variances
             -0.2581168
```

The two decisions do not match, but since the empirical variances are 2.034341 and 2.292458, one may think that $\sigma^2_{(1)}$ and $\sigma^2_{(2)}$ are slightly different. In which case, we have to be cautious because our sample might be of the same kind as the 4.81% of Table 5.

5 Discussion

We have presented an R package implementing large sample tests for various parameters. The interesting point is that each test statistic can be written in the same form, as follows:

$$\widehat{\Delta}_{\widehat{\theta},\theta}(\mathbf{Y}) := \frac{\widehat{\theta}\,(\mathbf{Y}) - \theta}{\widehat{\sigma_{\widehat{\theta}}}\,(\mathbf{Y})}.$$

This form clearly expresses the departure of the estimate from the true parameter, normalized by some quantity measuring the precision of the estimate. This approach is then attractive and easy to present. In the Gaussian framework, the one and two-sample t-tests follow this idea whereas the chi-square variance test and the Fisher test do not. One may wonder if it is possible to embed the classical Gaussian framework within this formalism. More precisely, for an estimate $\widehat{\theta}\,(\mathbf{Y})$ of some parameter θ with standard deviation $\sigma_{\widehat{\theta}} = \sqrt{Var(\widehat{\theta}\,(\mathbf{Y}))}$, let us propose the test statistic

$$\Delta_{\widehat{\theta},\theta}(\mathbf{Y}) := \frac{\widehat{\theta}\,(\mathbf{Y}) - \theta}{\sigma_{\widehat{\theta}}} \text{ or } \frac{\widehat{\theta}\,(\mathbf{Y}) - \theta}{\widetilde{\sigma}_{\widehat{\theta}}} \tag{4}$$

in the case where $\sigma_{\widehat{\theta}}$ or possibly some known asymptotic equivalent $\widetilde{\sigma}_{\widehat{\theta}}$ of $\sigma_{\widehat{\theta}}$ only depends on θ, or

$$\widehat{\Delta}_{\widehat{\theta},\theta}(\mathbf{Y}) := \frac{\widehat{\theta}\,(\mathbf{Y}) - \theta}{\widehat{\sigma_{\widehat{\theta}}}\,(\mathbf{Y})} \tag{5}$$

otherwise, where $\widehat{\sigma_{\widehat{\theta}}}\,(\mathbf{Y})$ is a consistent estimate of $\sigma_{\widehat{\theta}}$.

In the large sample framework, the parameters μ, σ^2, d_μ, r_μ, d_{σ^2} and r_{σ^2} fall into the case of equation (5) and the corresponding statistic approximately follows a $\mathcal{N}(0,1)$ distribution. In this same context, the proportion parameter falls into the case of equation (4) with $\sigma_{\widehat{p}} = \sqrt{p(1-p)/n}$.

In the Gaussian framework, the parameters μ and d_μ leading to the one-sample and two-sample t-tests also fall into the case of equation (5). Let us now concentrate on parameters σ^2 and r_{σ^2} (corresponding to the chi-square variance test and the Fisher test). It is known that the variance of $\widehat{\sigma^2}\left(\mathbf{Y^G}\right)$ and an asymptotic equivalent of the variance of $\widehat{r_{\sigma^2}}\left(\mathbf{Y^G}\right)$ are of the form

$$\sigma^2_{\widehat{\sigma^2}} = \frac{1}{n}\left(\dot{\mu}_4 - \frac{n-3}{n-1}\dot{\mu}_2^2\right) \text{ and } \widetilde{\sigma}^2_{\widehat{r_{\sigma^2}}} = \frac{1}{n^{(1)}}\frac{\dot{\mu}_4^{(1)} - (\dot{\mu}_2^{(1)})^2}{(\dot{\mu}_2^{(2)})^2} + \frac{1}{n^{(2)}} r^2_{\sigma^2}\frac{\dot{\mu}_4^{(2)} - (\dot{\mu}_2^{(2)})^2}{(\dot{\mu}_2^{(2)})^2}$$

where $\dot{\mu}_k = E\Big((Y - E(Y))^k\Big)$ is the $k-$th centered moment of Y (with our notation $\dot{\mu}_2 = \sigma^2$). For Gaussian variables, $\dot{\mu}_4 = 3(\dot{\mu}_2)^2$ which leads to

$$\sigma^2_{\widehat{\sigma^2}} = \frac{2}{n-1}\sigma^4 \text{ and } \widetilde{\sigma}^2_{\widehat{r_{\sigma^2}}} = 2r^2_{\sigma^2}\left(\frac{1}{n^{(1)}} + \frac{1}{n^{(2)}}\right).$$

Now, in order to build a test, let us give the distributions of $\Delta^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G})$ and $\Delta^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G})$ expressed in terms of $\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G})$ and $\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G})$:

$$\Delta^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G}) := \frac{\widehat{\sigma^2}\left(\mathbf{Y^G}\right) - \sigma^2}{\sigma_{\widehat{\sigma^2}}} = \frac{\frac{\Lambda^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G})}{n-1}\sigma^2 - \sigma^2}{\sigma^2\sqrt{2/(n-1)}} \rightsquigarrow \frac{\chi^2(n-1) - (n-1)}{\sqrt{2(n-1)}}$$

and

$$\Delta^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G}) := \frac{\widehat{r_{\sigma^2}}\left(\mathbf{Y^G}\right) - r_{\sigma^2}}{\widetilde{\sigma}_{\widehat{r_{\sigma^2}}}} = \frac{r_{\sigma^2}\Lambda^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G}) - r_{\sigma^2}}{r_{\sigma^2}\sqrt{2/n^{(1)} + 2/n^{(2)}}} \rightsquigarrow \frac{\mathcal{F}(n^{(1)}-1, n^{(2)}-1) - 1}{\sqrt{2/n^{(1)} + 2/n^{(2)}}}.$$

One may propose a name for the previous two free distributions: centered reduced chi-square distribution and centered reduced Fisher distribution respectively denoted by $\chi^2_{cr}(\cdot)$ and $\mathcal{F}_{cr}(\cdot,\cdot)$. If these distributions were implemented such that one may evaluate quantiles and p-values, one could build two new tests directly based on $\Delta^G_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y^G})$ and $\Delta^G_{\widehat{r_{\sigma^2}},r_{\sigma^2}}(\mathbf{Y^G})$. Of course, these two new tests would be strictly equivalent to the classical chi-square variance test and Fisher test and would then fall into the same formalism.

Let us now comment on the concept of robustness in the Gaussian framework. In the particular case of mean hypothesis testing, this robustness is expressed by the fact that $\Delta^G_{\widehat{\mu},\mu}(\mathbf{Y^G})$ is equal to $\Delta_{\widehat{\mu},\mu}(\mathbf{Y^G})$ with the same asymptotic distribution $\mathcal{N}(0,1)$. When considering the two new cases $\theta = \sigma^2$ or $\theta = r_{\sigma^2}$, $\Delta^G_{\widehat{\theta},\theta}(\mathbf{Y^G})$ is no longer equal to $\Delta_{\widehat{\theta},\theta}(\mathbf{Y^G})$ but have at least the same asymptotic distribution $\mathcal{N}(0,1)$. However, one can prove that $\Delta^G_{\widehat{\theta},\theta}(\mathbf{Y^G})$ and $\Delta_{\widehat{\theta},\theta}(\mathbf{Y^G})$ are asymptotically equivalent in probability, which may be viewed as some kind of robustness.

6 Proofs

In this section, we only prove (3). The results (1) and (2) are direct consequences. Recall that for each parameter, some assumptions are needed essentially in order to apply the central limit theorem. They are summarized in Tab. 2.

Parameter μ:

This is a direct application of the central limit theorem.

Parameter σ^2:

By definition

$$\widehat{\sigma^2}(\mathbf{Y}) - \sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\left(Y_i - \overline{Y}\right)^2 - \sigma^2\right) = \frac{1}{n-1}\sum_{i=1}^{n}\left(\left(Y_i-\mu\right)^2 - \sigma^2\right) - \frac{n}{n-1}\left(\overline{Y}-\mu\right)^2.$$

From the CLT, the law of large numbers and Slutsky's Theorem (see *e.g.* Ferguson (1996)), it comes that as $n \to +\infty$

$$\sqrt{n}\left(\overline{Y}-\mu\right)^2 \xrightarrow{\mathbb{P}} 0.$$

Therefore, as $n \to +\infty$,

$$\sqrt{n}\left(\widehat{\sigma^2}(\mathbf{Y}) - \sigma^2\right) \xrightarrow{d} \mathcal{N}(0, Var((Y-\mu)^2)).$$

Since $\sigma_{\widehat{\sigma^2}} := \sqrt{\frac{Var((Y-\mu)^2)}{n}}$ can be consistently estimated by $\widehat{\sigma_{\widehat{\sigma^2}}}(\mathbf{Y}) := \sqrt{\frac{\widehat{\sigma^2_{\ddot{Y}}}(\ddot{\mathbf{Y}})}{n}}$, see Tab. 2, we obtain as $n \to +\infty$

$$\widehat{\Delta}_{\widehat{\sigma^2},\sigma^2}(\mathbf{Y}) := \frac{\widehat{\sigma^2}(\mathbf{Y}) - \sigma^2}{\widehat{\sigma_{\widehat{\sigma^2}}}(\mathbf{Y})} \xrightarrow{d} \mathcal{N}(0,1).$$

Parameter d_μ:

Recall that $n = n^{(1)}$ and $n^{(2)} = \alpha n^{(1)}$. As $n \to +\infty$,

$$\begin{aligned}\sqrt{n}\left(\widehat{d_\mu}(\mathbf{Y}) - d_\mu\right) &= \sqrt{n}\left(\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right) - \mu^{(1)}\right) - \rho \times \frac{\sqrt{n\alpha}}{\sqrt{\alpha}}\left(\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right) - \mu^{(2)}\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{(1)} + \rho^2 \times \frac{\sigma^2_{(2)}}{\alpha}\right).\end{aligned}$$

Since $\sigma_{\widehat{d_\mu}} := \sqrt{\frac{\sigma^2_{(1)} + \rho^2 \times \frac{\sigma^2_{(2)}}{\alpha}}{n}}$ can be consistently estimated by

$$\widehat{\sigma_{\widehat{d_\mu}}}\left(\mathbf{Y}\right) := \sqrt{\frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{n^{(1)}} + \rho^2 \times \frac{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n^{(2)}}},$$

we obtain as $n \to +\infty$

$$\widehat{\Delta}_{\widehat{d_\mu}, d_\mu}(\mathbf{Y}) := \frac{\widehat{d_\mu}\left(\mathbf{Y}\right) - d_\mu}{\widehat{\sigma_{\widehat{d_\mu}}}\left(\mathbf{Y}\right)} \xrightarrow{d} \mathcal{N}(0,1).$$

Parameter d_{σ^2}:

As $n \to +\infty$,

$$\begin{aligned}\sqrt{n}\left(\widehat{d_{\sigma^2}}\left(\mathbf{Y}\right) - d_{\sigma^2}\right) &= \sqrt{n}\left(\widehat{\sigma^2_{(1)}(1)}\left(\mathbf{Y^{(1)}}\right) - \sigma^2_{(1)}\right) - \rho \times \frac{\sqrt{n\alpha}}{\sqrt{\alpha}}\left(\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right) - \sigma^2_{(2)}\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{\ddot{Y^{(1)}}} + \rho^2 \times \frac{\sigma^2_{\ddot{Y^{(2)}}}}{\alpha}\right).\end{aligned}$$

Since $\sigma_{\widehat{d_{\sigma^2}}} := \sqrt{\frac{\sigma^2_{\ddot{Y^{(1)}}} + \rho^2 \frac{\sigma^2_{\ddot{Y^{(2)}}}}{\alpha}}{n}}$ can be consistently estimated by

$$\widehat{\sigma_{\widehat{d_{\sigma^2}}}}\left(\mathbf{Y}\right) := \sqrt{\frac{\widehat{\sigma^2_{\ddot{Y^{(1)}}}}\left(\mathbf{\ddot{Y}^{(1)}}\right)}{n} + \rho^2 \times \frac{\widehat{\sigma^2_{\ddot{Y^{(2)}}}}\left(\mathbf{\ddot{Y}^{(2)}}\right)}{n\alpha}} = \sqrt{\frac{\widehat{\sigma^2_{\ddot{Y^{(1)}}}}\left(\mathbf{\ddot{Y}^{(1)}}\right)}{n^{(1)}} + \frac{\widehat{\sigma^2_{\ddot{Y^{(2)}}}}\left(\mathbf{\ddot{Y}^{(2)}}\right)}{n^{(2)}}},$$

we obtain as $n \to +\infty$

$$\frac{\widehat{d_{\sigma^2}}\left(\mathbf{Y}\right) - d_{\sigma^2}}{\widehat{\sigma_{\widehat{d_{\sigma^2}}}}\left(\mathbf{Y}\right)} \xrightarrow{d} \mathcal{N}(0,1).$$

Parameter r_μ:

Using Slutsky's Theorem, one may assert that as $n \to +\infty$

$$\begin{aligned}\sqrt{n}\left(\widehat{r_\mu}\left(\mathbf{Y}\right) - r_\mu\right) &= \sqrt{n}\left(\frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)} - \frac{\mu^{(1)}}{\mu^{(2)}}\right) \\ &= \sqrt{n}\left(\frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right) - \mu^{(1)}}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)} + \mu^{(1)}\left(\frac{1}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)} - \frac{1}{\mu^{(2)}}\right)\right) \\ &\overset{n\to+\infty}{\sim} \sqrt{n}\left(\frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right) - \mu^{(1)}}{\mu^{(2)}} + \frac{\sqrt{n\alpha}}{\sqrt{\alpha}}\frac{\mu^{(1)}}{(\mu^{(2)})^2}\left(\mu^{(2)} - \widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)\right)\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2_{(1)}}{(\mu^{(2)})^2} + \left(\frac{\mu^{(1)}}{(\mu^{(2)})^2}\right)^2 \frac{\sigma^2_{(2)}}{\alpha}\right).\end{aligned}$$

Since, $\sigma_{\widehat{r_\mu}} := \sqrt{\frac{\frac{\sigma^2_{(1)}}{(\mu^{(2)})^2}+\left(\frac{\mu^{(1)}}{(\mu^{(2)})^2}\right)^2\frac{\sigma^2_{(2)}}{\alpha}}{n}}$ can be consistently estimated by

$$\begin{aligned}\widehat{\sigma_{\widehat{r_\mu}}}(\mathbf{Y}) &:= \sqrt{\frac{\frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)}}{n}+\frac{\left(\frac{\widehat{\mu^{(1)}}\left(\mathbf{Y^{(1)}}\right)}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)^2}\right)^2\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n\alpha}} \\ &= \frac{1}{\widehat{\mu^{(2)}}\left(\mathbf{Y^{(2)}}\right)}\sqrt{\frac{\widehat{\sigma^2_{(1)}}\left(\mathbf{Y^{(1)}}\right)}{n^{(1)}}+\widehat{r_\mu}\left(\mathbf{Y}\right)\frac{\widehat{\sigma^2_{(2)}}\left(\mathbf{Y^{(2)}}\right)}{n^{(2)}}}\end{aligned}$$

we obtain as $n \to +\infty$,

$$\frac{\widehat{r_\mu}\left(\mathbf{Y}\right)-r_\mu}{\widehat{\sigma_{\widehat{r_\mu}}}\left(\mathbf{Y}\right)} \xrightarrow{d} \mathcal{N}(0,1).$$

Parameter r_{σ^2}:

The proof follows the ideas developed for the parameters σ^2, d_{σ^2} and r_μ, and thus is left to the reader.

Acknowledgments

We would like to thank Laurence Pierret for having kindly reread our English. She could in no way be held responsible for any remaining mistakes.

References

E. Anderson. The irises of the gaspe peninsula. *Bulletin of the American Iris Society*, 59:2–5, 1935.

G.E.P. Box. Non-normality and tests on variances. *Biometrika*, 40(3/4):318–335, 1953.

G. Casella and R.L. Berger. *Statistical Inference*. Duxbury press, Belmont, California, 1990.

W.J. Conover, M.E. Johnson, and M.M. Johnson. A comparative study of tests for homogeneity of variances with applications to the outer continental shelf bidding data. *Technometrics*, 23 (4):351–361, 1981.

T.S. Ferguson. *A course in large sample theory*. Chapman and Hall, London, UK, 1996.

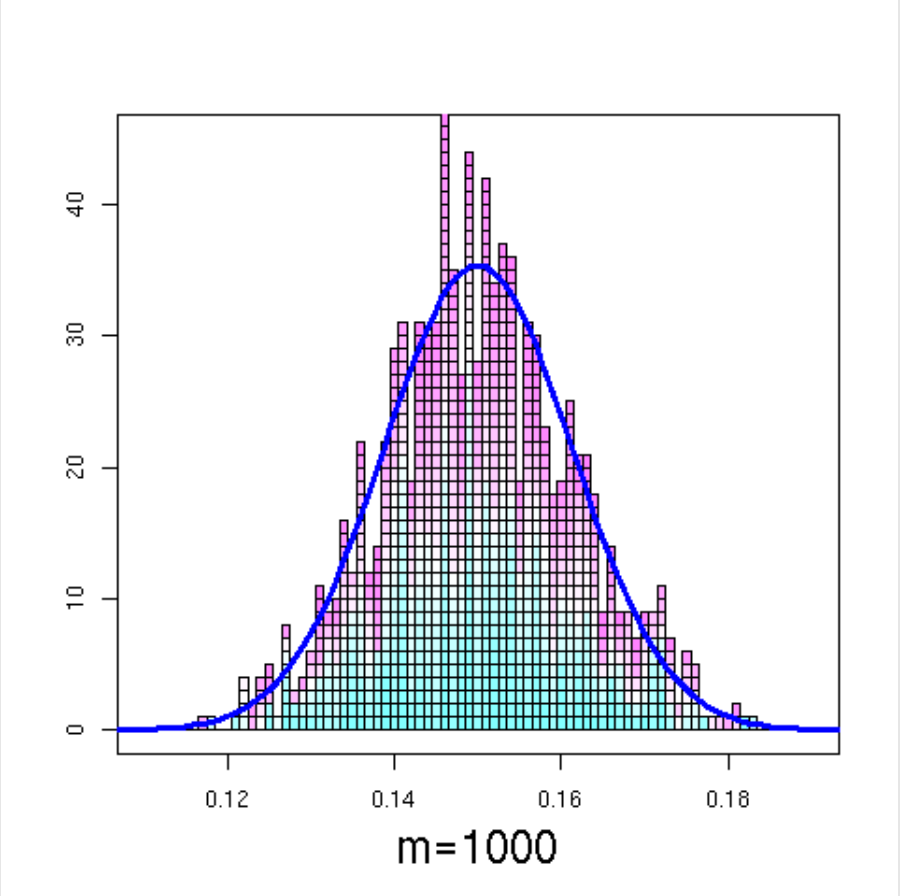
R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7 (Part II):179–188, 1935.

C. Ozgur and S.E. Strasser. A study of the statistical inference criteria: can we agree on when to use z versus t ? *Decision Sciences Journal of Innovative Education*, 2(2):177–192, 2004.

G. Pan. On a levene type test for equality of two variances. *Journal of Statistical Computation and Simulation*, 63:59–71, 1999.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2004. ISBN ISBN 3-900051-00-3. URL `http://www.R-project.org/`.

M.L. Tiku, W.Y. Tan, and N. Balakrishnan. *Robust inference*. Statistics: Textbooks and Monographs, New York, USA, 1986.



[CD05a]
J.-F. Coeurjolly and R. Drouilhet. Approche expérimentale des probabilités comme complément à une approche plus classique, In *37èmes journées de statistiques*, Pau, France, 2005

Estimation de la moyenne : Une approche expérimentale versus une approche classique.

Jean-François Coeurjolly & Rémy Drouilhet

Labsad, BSHM, UPMF, 1251 Avenue Centrale BP 47, 38040 Grenoble Cedex 9
Jean-Francois.Coeurjolly@upmf-grenoble.fr, Remy.Drouilhet@upmf-grenoble.fr

Résumé: Enseignants en deuxième année (de Licence) de sciences économiques dans une filière "non-matheuse", nous avions pour charge d'enseigner les probabilités et la statistique inférentielle pour un volume horaire assez court (20h cours et 20h T.D.). De par nos différentes expériences, nous avons remarqué qu'étant donné le public une approche "classique" trop formalisée posait un certain nombre de difficultés que ce soit sur les concepts probabilistes ou sur la construction et l'interprétation des intervalles de confiance et tests d'hypohèses. Afin d'appréhender ces différents concepts tout en évitant les techniques mathématiques, nous nous sommes orientés vers l'utilisation conjointe d'une approche en simulation (appelée approche expérimentale des probabilités (A.E.P.)) et de logiciels éducatifs permettant la visualisation graphique de ces concepts. Dans cet exposé, nous nous concentrerons uniquement sur l'étude des lois d'échantillonnage puis sur la notion d'intervalle de confiance.

Mots-clés: approche expérimentale des probabilités, approche visuelle, intervalle de confiance.

Abstract: Lecturers in second year of Economics Licence in a "non-mathematical" branch, we were in charge of a course on probability and statistical inference. This lecture was designed to a volume of forty hours. This constraint and our experience convinced us that a "classical" approach is too formal and so not adapted to our public. In order to learn concepts like probability density function, confidence intervals or hypothesis testing, we try to avoid mathematical technics by combining an experimental approach of probability and some educational softwares that allow a graphical visualisation of such concepts. This talk will focus on the following topics : sampling distributions and confidence interval.

1 Introduction et motivations

Dans un séminaire de probabilités ou statistiques, il est courant que l'orateur conclut son exposé par une étude en simulation. Cette étape basée sur l'expérimentation permet entre autres choses de mettre en avant et ainsi de communiquer plus facilement les résultats obtenus en se dégageant de la lourdeur des techniques mathématiques ayant permis de les démontrer. Notre étonnement a été alors le suivant : pourquoi enseigner à un public plutôt rebuté par les mathématiques, les statistiques et les probabilités par une approche classique reposant sur les techniques mathématiques (dont les principales vertues sont de calculer et démontrer) plutôt qu'une approche expérimentale conduisant à une meilleure interprétation des outils mathématiques mis en place ?

Quelques unes des difficultés rencontrées par l'étudiant : (1) Compréhension plus que confuse de la notion de variable aléatoire, (2) Confusion entre variable aléatoire (v.a.) (notée dès que possible en majuscule) et ses réalisations (notées en minuscule), (3) Cadre des v.a. continues bien moins accessibles que celui des v.a. discrètes, (4) Densité de probabilité, un objet graphique très étrange et bien plus difficile qu'un diagramme en baton pour caractériser la loi de probabilité d'une v.a. discrète et ....

Les messages que nous jugeons difficiles à faire passer (notamment à des étudiants non "matheux") : (1) Approximation d'une v.a. discrète (représentation classique de sa loi de probabilité en diagramme en baton) par une v.a. continue (représentation de sa loi de probabilité par une densité de probabilité), (2) Bonne interprétation de la notion d'intervalle de confiance et (3) Interprétation des décisions résultant d'un test d'hypothèses (voir l'exposé intitulé "Construction d'un test d'hypothèses par un approche visuelle et une approche expérimentale des probabilités").

Nos solutions : (1) Appréhension de la notion de v.a. en introduisant une phase expérimentale (plutôt classique) consistant à générer un grand nombre de ses réalisations, (2) Système de notations plus complexes mais plus précises (très utiles en phase d'apprentissage) spécialement adaptées à cette approche expérimentale (3) Représentation des v.a. discrètes et continues par des histogrammes (discrets et continus) définis comme un mur de briques empilées une à une en abscisse en leurs réalisations associées.

Nous estimons que malgré la complexité apparente des notations introduites (impression immédiate), elles nous permettent (dans un deuxième temps) de communiquer plus facilement les messages aux étudiants puisqu'elles décrivent très précisément tous les acteurs de cette approche expérimentale des probabilités.

2 Introduction de l'Approche Expérimentale des Probabilités (A.E.P.)

Pour être le plus direct possible, nous nous limitons ici à présenter l'Approche Expérimentale des Probabilités dans le cadre de l'étude des lois d'échantillonnage. Puisque notre approche pédagogique est basée sur l'utilisation de logiciels éducatifs (pour la plupart développés par nous-mêmes), nous privilégierons dans cet exposé les représentations graphiques que nous montrons aux étudiants.

Notre principal exemple de cours : un industriel veut savoir s'il doit ou pas lancer un nouveau produit (noté produit B) sur le marché. Cette problématique est facile et rapide à appréhender reposant notamment sur l'estimation d'un paramètre défini à partir d'informations relatives à une population finie (ici celle des $N = 2000000$ acheteurs potentiels). L'enjeu de l'industriel est que le **jour J**, il devra dépenser une grosse somme d'argent pour recueillir les choix d'achats de $n = 1000$ individus de son échantillon et qu'avec cette seule information il devra prendre sa décision de lancement ou pas de son produit. L'industriel se fait alors assister

par un *expérimentateur* et un *mathématicien* lui promettant de bien lui faire comprendre la nature des estimations de μ^B dont il disposera le **jour J**.

Dans cet exposé, nous ne traiterons pas directement de test d'hypothèses mais uniquement d'estimation (ponctuelle et par intervalle de confiance) du paramètre, ici le nombre μ^B moyen de produit(s) acheté(s) sur la population totale.

Phase d'expérimentation ou de **simulation** : Malheureusement pour l'industriel, le paramètre d'intérêt μ^B est **inconnu**. Pour anticiper ce qui peut arriver avant la prise de décision finale, il peut simuler toutes les situations possibles dans une phase expérimentale. Pour ce faire, le paramètre d'intérêt μ^B inconnu sera remplacé par un paramètre $\mu^\star$ dont la valeur peut être fixée arbitrairement par l'expérimentateur. La population totale sera alors remplacée par une urne $U^B_{\mu^\star}$ contenant N boules (en remplacement des acheteurs potentiels) dont N_j $(j = 0, ..., J_{max})$ sont numérotées par j (correspondant au nombre de produit(s) B acheté(s)). Notons que la répartition doit être choisie de sorte que la moyenne des numéros de toutes les boules de l'urne soit égale à $\mu^\star$ (dans nos exemples, 0.1 ou 0.15 ou 0.19). La variance quant à elle sera alors notée $\sigma^2_\star$ remplaçant dans la phase expérimentale celle σ^2_B de la vraie population. Construire un échantillon consistera alors à simplement tirer $n = 1000$ boules (avec remise) dans ces urnes.

Un tableau décrivant l'Approche Expérimentale des Probabilités : Soulignons l'utilisation d'un système de notations spécifique et la mise en avant de la chronologie à partir du **jour J**.

↓ **AVANT le jour J**			
Phase expérimentale *Le paramètre à estimer est $\mu^\star$ fixé arbitrairement (par exemple, à 0.19)*			
Avant simulation	$\boldsymbol{\mathcal{E}} = (\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_n)$	$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$	$\widehat{\mu^\star}(\boldsymbol{Y})$
Après simulation	$1^{ère}$ expérience $\boldsymbol{e}_{[1]}$	$\boldsymbol{y}_{[1]}$	$\widehat{\mu^\star}(\boldsymbol{y}_{[\mathbf{1}]})$
	$2^{ème}$ expérience $\boldsymbol{e}_{[2]}$	$\boldsymbol{y}_{[2]}$	$\widehat{\mu^\star}(\boldsymbol{y}_{[\mathbf{2}]})$
	$\vdots$	$\vdots$	$\vdots$
	$m^{ème}$ expérience $\boldsymbol{e}_{[m]}$	$\boldsymbol{y}_{[m]}$	$\widehat{\mu^\star}(\boldsymbol{y}_{[\boldsymbol{m}]})$
	$\vdots$	$\vdots$	$\vdots$
Phase pratique *Le paramètre à estimer est μ^B qui est inconnu*			
Avant pratique	$\boldsymbol{\mathcal{E}} = (\mathcal{E}_1, \mathcal{E}_2, ..., \mathcal{E}_n)$	$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)$	$\widehat{\mu^B}(\boldsymbol{Y})$
↓ **APRES le jour J**			
Après pratique	l'expérience réelle $\boldsymbol{e}$	$\boldsymbol{y}$	$\widehat{\mu^B}(\boldsymbol{y})$

Nous avons ici particularisé nos notations à celle de la problématique de l'industriel mais elles sont facilement adaptables à un cadre général. Nous aimons rassurer nos étudiants en leur affirmant que toutes les notations utilisées (vues comme des "mots" d'un langage mathématique) sont faciles à traduire littéralement (en français). Ainsi, $\widehat{\theta}(\cdot)$ se dit : estimation ou "remplaçant"

du paramètre θ obtenu à partir du jeu de données '$\cdot$'. Les différents jeux de données sont décrits ci-dessous :

→ **Avant le jour J, les futures données :** L'industriel envisage le procédé de construction de ses futures données $\boldsymbol{Y} = (Y_1, \cdots, Y_n)$ constitué des futurs nombres de produit(s) B acheté(s) par les $n = 1000$ individus choisis au hasard (et avec remise) dans la population totale ou dans une des urnes expérimentales. Selon ces deux cas, il obtiendra respectivement comme future estimation soit $\widehat{\mu^B}(\boldsymbol{Y})$ soit $\widehat{\mu^\star}(\boldsymbol{Y})$ obtenue à partir des $Y_1, \cdots, Y_n$:

$$\left.\begin{array}{c} \widehat{\mu^B}(\boldsymbol{Y}) \\ \text{ou} \\ \widehat{\mu^\star}(\boldsymbol{Y}) \end{array}\right\} = \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Il en est de même pour l'estimation de la variance avec $\widehat{\sigma_B^2}(\boldsymbol{Y})$ et $\widehat{\sigma_\star^2}(\boldsymbol{Y})$ futures estimations de σ_B^2 et $\sigma_\star^2$ respectivement.

→ **Avant le jour J, les données simulées :** Après avoir construit son urne expérimentale, l'industriel peut obtenir autant de jeux de données qu'il le souhaite. Il en construit m (ici 10000), le $j^{ème}$ étant noté $\boldsymbol{y_{[j]}}$. Il obtient alors m estimations $(\widehat{\mu^\star}(\boldsymbol{y_{[j]}}))_{j=1,\cdots,m}$ et éventuellement m estimations $(\widehat{\sigma_\star^2}(\boldsymbol{y_{[j]}}))_{j=1,\cdots,m}$.

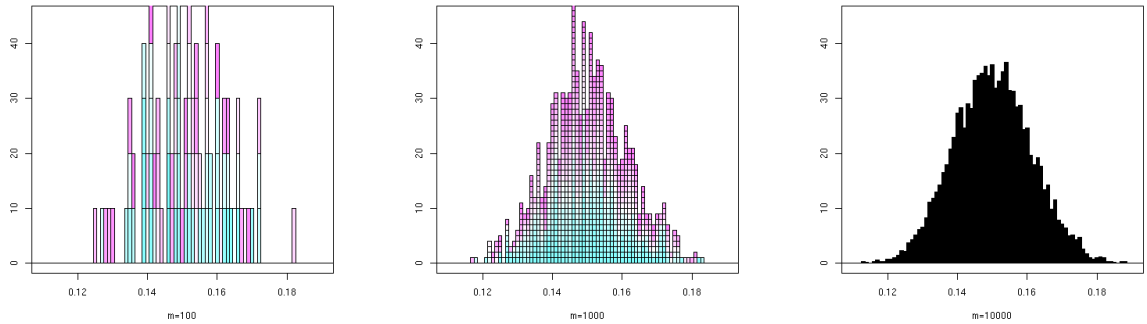
→ **Le jour J, les données réelles :** L'industriel se paie un jeu de données $\boldsymbol{y} = (y_1, \cdots, y_n)$ constitué des nombres de produit(s) B acheté(s) par les $n = 1000$ individus choisis au hasard (et avec remise) dans la population totale. Il en déduit une estimation (réelle) $\widehat{\mu^B}(\boldsymbol{y})$ de μ^B et éventuellement une estimation (réelle) $\widehat{\sigma_B^2}(\boldsymbol{y})$ de σ_B^2.

Après avoir mis en correspondance, le langage mathématique et le langage littéral, nous allons maintenant les mettre en correspondance avec le langage "visuel". En effet, dès que c'est possible, nous avons pour volonté de représenter graphiquement tout objet acteur important dans la phase d'apprentissage. La phase expérimentale facilite la compréhension de la nature aléatoire de l'échantillonnage en proposant une multitude ($m = 10000$ ou plus) de réalisations de la future estimation. Afin de caractériser son comportement aléatoire, il ne reste plus qu'à en faire une représentation graphique de sa répartition. La question est "diagramme en bâtons" ou "histogramme discret" ? Nous optons pour le second qui permettra de visualiser sur un même graphique la loi d'une v.a. discrète ainsi que son approximation par celle d'une v.a. continue.
Histogramme en briques :

- Les m réalisations $\widehat{\mu^\star}(\boldsymbol{y_{[j]}})$ de $\widehat{\mu^\star}(\boldsymbol{Y})$ sont représentées par des briques de surface $1/m$ et de largeur $1/n$ centrées en abscisse en leurs valeurs associées.
- Toutes les briques sont empilées une par une et l'évolution en m de cet empilement laisse apparaître un "mur" de briques de surface totale toujours égale à 1.
- Cette représentation, appelée histogramme (discret), permet de visualiser en un seul coup d'oeil la répartition des $(\widehat{\mu^\star}(\boldsymbol{y_{[j]}}))_{j=1,\cdots,m}$.

Sur une urne particulière, voilà l'évolution en m de l'histogramme en briques.

Evolution des histogrammes discrets



Assez naturellement, nous nous interrogeons sur la meilleure valeur de m pour caractériser le comportement aléatoire (ici) de $\widehat{\mu^\star}(\boldsymbol{Y})$. Les étudiants répondent très rapidement "le plus grand possible" et les plus courageux "une infinité". Ensuite, nous poursuivons notre récit en évoquant que le "matheux" le savait **à l'avance** dès lors toute information relative à l'urne expérimentale lui est fournie. L'étudiant est alors invité à comprendre le résultat du mathématicien dans son propre langage :

$$\widehat{\mu^\star}(\boldsymbol{Y}) \overset{approx.}{\rightsquigarrow} \mathcal{N}(\mu^\star, \frac{\sigma_\star}{n}) \tag{1}$$

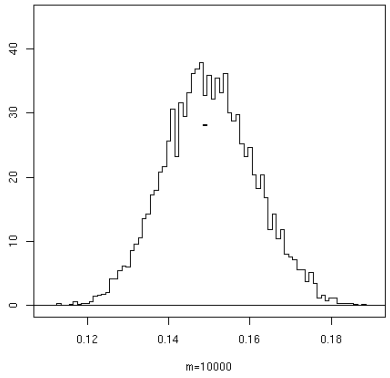
Le tableau et les graphiques ci-dessous décrivent les connaissances de l'expérimentateur et du mathématicien. La courbe "lisse" du dernier graphique exprime la connaissance anticipée du mathématicien décrite en (1). La question que nous posons à l'étudiant est la suivante : à quoi ressemble une brique si m tend vers $+\infty$? Il répond généralement un point. Bien qu'erronée, celle-ci nous satisfait complètement mais nous lui indiquons que sa réponse est juste si n tend lui-aussi vers $+\infty$. Nous poursuivons alors avec l'interprétation de ce résultat. L'industriel s'imagine être le **jour J** dans l'une des situations expérimentales ($\mu^\star$ fixé). et il prend alors conscience que ce qui peut lui arriver **le jour J**, c'est équivalent (ou presque) à :

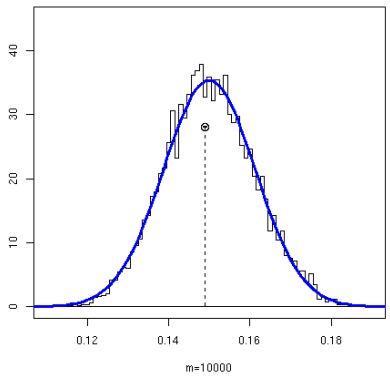
1. Grâce à l'Expérimentateur, choisir (graphique du haut ci-dessous) au hasard une brique (i.e un des m $\widehat{\mu^\star}\left(\boldsymbol{y}_{[j]}\right)$ du tableau ci-dessous)

2. Grâce au Mathématicien, choisir (graphique du bas ci-dessous) au hasard un point sous la "courbe $\mathcal{N}(\mu^\star, \frac{\sigma_\star}{\sqrt{n}})$" associé à son abscisse représentant une réalisation au hasard de $\widehat{\mu^\star}(\boldsymbol{Y})$ choisie parmi une infinité.

Il voit clairement la "courbe $\mathcal{N}(\mu^\star, \frac{\sigma_\star}{\sqrt{n}})$" comme un empilement d'une infinité de briques ("devenues des points") associées à une infinité de réalisations possibles de $\widehat{\mu^\star}(\boldsymbol{Y})$.

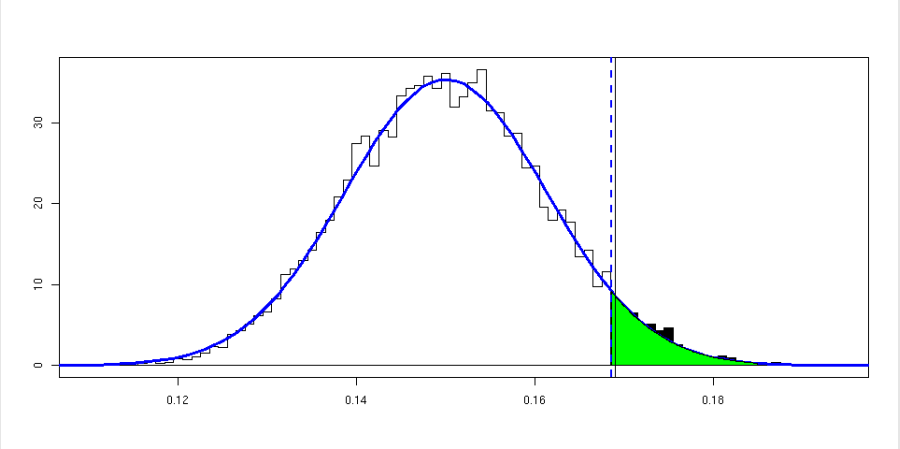
Urne $U^B_{0.15}$	
j	$\widehat{\mu^\star}\left(\boldsymbol{y}_{[j]}\right)$
⋮	⋮
1924	0.147
1925	0.172
1926	0.151
1927	0.132
1928	0.133
1929	0.129
1930	0.146
1931	0.137
1932	0.157
1933	0.143
1934	0.155
1935	0.148
1936	0.151
1937	0.188
1938	0.154
1939	0.152
1940	0.139
⋮	⋮





3 Application à l'intervalle de confiance

Nous ne nous concentrerons pas sur l'aspect mathématique de la construction de l'intervalle de confiance (relativement aisée) mais plutôt sur la difficulté à l'interpréter. En effet, le **jour J** l'industriel disposera d'un intervalle de confiance $[\widetilde{\theta^\star}_{inf}\left(\boldsymbol{y}\right), \widetilde{\theta^\star}_{sup}\left(\boldsymbol{y}\right)]$ (par exemple $[0.113, 0.23]$) obtenu avec 95% de niveau de confiance. La question est alors : quelle est la valeur de la probabilité $\mathbb{P}\left(\mu^B \in [0.113, 0.23]\right)$? Même bien avertis, les étudiants traduisant leur propre compréhension de cette fourchette répondent en coeur : 95%. Ils ne prennent pas conscience que μ^B est ou n'est pas dans l'intervalle obtenue. Pour les éclairer, nous nous en remettons à l'A.E.P. Nous pouvons alors imaginer construire une infinité d'intervalles de confiances $[\widetilde{\theta^\star}_{inf}\left(\boldsymbol{y}_{[1]}\right), \widetilde{\theta^\star}_{sup}\left(\boldsymbol{y}_{[1]}\right)], \cdots, [\widetilde{\theta^\star}_{inf}\left(\boldsymbol{y}_{[m]}\right), \widetilde{\theta^\star}_{sup}\left(\boldsymbol{y}_{[m]}\right)], \cdots$ dont 5% seulement ne contiendront pas μ^B. Le jour J, cela revient alors à en choisir un au hasard parmi cette infinité. Utiliser un tel outil c'est donc parier que nous sommes normalement "chanceux" en tombant sur un parmi les 95% qui contiennent μ^B. Par manque de place, nous ne pouvons faire figurer les représentations graphiques relatives à un de nos logiciels éducatifs illustrant ce contexte.



[CD05b]
J.-F. Coeurjolly and R. Drouilhet. Construction d'un test d'hypothèses par une approche visuelle et une approche expérimentale des probabilités, In *37èmes journées de statistiques*, Pau, France, 2005

Construction d'un test d'hypothèses par une approche visuelle et une approche expérimentale des probabilités

Jean-François Coeurjolly & Rémy Drouilhet

Labsad, BSHM, UPMF, 1251 Avenue Centrale BP 47, 38040 Grenoble Cedex 9
Jean-Francois.Coeurjolly@upmf-grenoble.fr, Remy.Drouilhet@upmf-grenoble.fr

Résumé: Enseignants en deuxième année (de Licence) de sciences économiques dans une filière "non-matheuse", nous avions pour charge d'enseigner les probabilités et la statistique inférentielle pour un volume horaire assez court (20h cours et 20h T.D.). De par nos différentes expériences, nous avons remarqué qu'étant donné le public une approche "classique" trop formalisée posait un certain nombre de difficultés que ce soit sur les concepts probabilistes ou sur la construction et l'interprétation des intervalles de confiance et tests d'hypohèses. Afin d'appréhender ces différents concepts tout en évitant les techniques mathématiques, nous nous sommes orientés vers l'utilisation conjointe d'une approche en simulation (appelée approche expérimentale des probabilités (A.E.P.)) et d'éducatifs permettant la visualisation graphique de ces concepts. Dans cet exposé, on se concentrera uniquement sur le thème "tests d'hypothèses". On considère donc comme prérequis l'interprétation d'une densité de probabilité via l'approche expérimentale (voir l'exposé associé intitulé "Estimation de la moyenne : une approche expérimentale versus une approche plus classique").

Mots-clés: approche expérimentale des probabilités, approche visuelle, tests d'hypothèses.

Abstract: Lecturers in second year of Economics Licence in a "non-mathematical" branch, we were in charge of a course on probability and statistical inference. This lecture was designed to a volume of forty hours. This constraint and our experience convinced us that a "classical" approach is too formal and so not adapted to our public. In order to learn concepts like probability density function, confidence intervals or hypothesis testing, we try to avoid mathematical technics by combining an experimental approach of probability and some educational softwares that allow a graphical visualisation of such concepts. This talk will focus on the topic "hypothesis testing". The interpretation of a probability density function through an experimental approach is required (see the associated talk entitled "Estimation de la moyenne : une approche expérimentale versus une approche plus classique").

1 Quelques problèmes rencontrés par l'étudiant

Sans vouloir être exhaustif, nous listons ici quelques difficultés (rencontrées par l'étudiant) reliées à la construction et la compréhension des tests d'hypothèses : (1) mauvaise compréhension des erreurs de décision mises en jeu dans toute problématique de test d'hypothèses, absence de visualisation de ces erreurs. (2) mauvaise interprétation de la conclusion d'un test d'hypothèses. L'accent étant trop mis sur l'hypothèse H_0, l'étudiant pense qu'étant donnée la formulation

standard d'un test d'hypothèses il peut accepter H_0. (3) confusion entre les différents acteurs en particulier entre le paramètre d'intérêt, son estimateur et son estimation. (4) l'étudiant ne voit pas l'importance de la p−valeur dans un test d'hypothèses et a du mal à percevoir cette quantité comme un risque critique.

2 Tentatives de solutions proposées

Quelques unes des orientations que nous avons choisies : (1) nous pensons qu'il est nécessaire d'éviter, dans un premier temps, la formulation mathématique standard (opposant deux hypothèses) en présentant à l'étudiant une problématique pratique simple permettant une description littérale et sans équivoque des erreurs de décision. (2) utilisation de l'A.E.P. pour donner plus de sens aux surfaces d'erreur de décision. En particulier, on utilisera le message qu'une surface entre deux barres verticales (par exemple) sous la densité de probabilité d'une certaine variable aléatoire correspond à la proportion de réalisations (parmi l'infinité) de cette variable aléatoire comprises entre ces deux barres verticales. (3) il est nécessaire de disposer d'un système de notations complet et adapté au discours de manière à ce que l'accent puisse être mis sur le langage mathématique et non sur les techniques mathématiques; par exemple des notations sont nécessaires pour différencier clairement une variable aléatoire de sa réalisation, ... (4) visualisation des acteurs au cours des différentes étapes de la construction du test nous permettant ainsi plus de raconter une histoire que de construire un outil mathématique. Par ailleurs, une visualisation graphique permet de définir et d'interpréter très naturellement la p−valeur du test.

En combinant ces différents aspects nous espérons que l'étudiant cesse de voir l'outil test d'hypothèses comme une simple recette de cuisine qu'il suffit d'appliquer et qu'il soit plutôt en mesure de comprendre toutes les étapes de la construction d'un test d'hypothèses et les différentes manières d'interpréter les résultats.

3 Exemple de construction d'un test de moyenne (dans un cadre asymptotique)

3.1 Problématique et notations

Un industriel souhaite lancer sur le marché un produit noté produit B. Celui-ci sera jugé rentable s'il est vendu à plus de 300000 exemplaires. On indique à l'industriel que la population ciblée pour ce produit est de taille $N = 2000000$.

Dans toute problématique, nous convions l'étudiant à définir le **paramètre d'intérêt** et **l'assertion d'intérêt** (ce que l'on veut montrer). Ici, le paramètre d'intérêt correspond au nombre moyen de produit B acheté par individu (noté μ^B). On définit également un autre paramètre appelé paramètre d'écart standardisé noté $\delta_{\mu^B,0.15}$ et défini par $\delta_{\mu^B,0.15} = \frac{\mu^B - 0.15}{\sqrt{\sigma_B^2/n}}$ où σ^2 représente la variance des réponses des N individus. Ainsi, le produit B est rentable si $\boxed{\mu^B > 0.15 \Longleftrightarrow \delta_{\mu^B,0.15} > 0}$. Cette dernière expression représentera naturellement l'assertion d'intérêt, que l'on cherchera à **confirmer avec les données sans trop se tromper** car il est

impossible (trop cher) d'évaluer exactement les paramètres μ^B et $\delta_{\mu^B,0.15}$.

On notera $\mathbf{y}$ le jeu de données de l'industriel constitué de $n = 1000$ réponses d'individus choisis au hasard au sein de la population totale . De manière à situer dans le temps l'histoire, on supposera que le jeu de données est recueilli un jour particulier appelé **jour J**. C'est ce jour et uniquement ce jour que la décision ou non de lancer le produit sera prise. Cependant, comme on va le voir de suite, **avant le jour J**, un grand nombre d'informations (en particulier la règle de décision) est disponible. Commençons par noter $\mathbf{Y}$ le vecteur des n variables aléatoires chacune d'entre elles étant associée à la réponse d'un individu choisi au hasard. On peut voir ce vecteur comme un "futur" jeu de données (qui ne s'est pas encore réalisé).

Les paramètres μ^B et $\delta_{\mu^B,0.15}$ étant inconnus, cherchons à définir leur estimateur (remplaçant) calculé chacun à partir du futur échantillon et notés respectivement $\widehat{\mu^B}(\mathbf{Y})$ et $\widehat{\delta_{\mu^B,0.15}}(\mathbf{Y})$ et définis par

$$\widehat{\mu^B}(\mathbf{Y}) = \overline{Y} \quad \text{et} \quad \widehat{\delta_{\mu^B,0.15}}(\mathbf{Y}) = \frac{\widehat{\mu^B}(\mathbf{Y}) - 0.15}{\sqrt{\frac{\widehat{\sigma^2}(\mathbf{Y})}{n}}}$$

où $\widehat{\sigma^2}(\mathbf{Y})$ représente la variance des réponses des n individus calculée à partir du futur échantillon $\mathbf{Y}$ de taille n. Remarquons que le **Jour J**, ces quantités seront notées $\widehat{\mu^B}(\mathbf{y})$ et $\widehat{\delta_{\mu^B,0.15}}(\mathbf{y})$ (que l'on ne pourra pas confondre avec leur estimateur respectif).

Pour la construction du test d'hypothèses, on s'appuiera sur les résultats suivants connus **avant le jour J** :

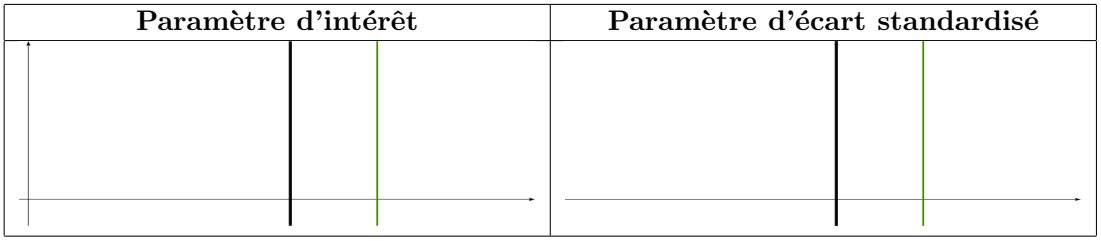
$$\boxed{\widehat{\mu^B}(\mathbf{Y}) \overset{approx.}{\rightsquigarrow} \mathcal{N}(\mu^B, \sqrt{\sigma^2/n})} \text{ et si } \mu^B = 0.15, \boxed{\widehat{\delta_{\mu^B,0.15}}(\mathbf{Y}) \overset{approx.}{\rightsquigarrow} \mathcal{N}(0,1)}.$$

3.2 Les messages des différentes étapes

Durant l'exposé, nous présenterons un éducatif développé en `R` et interfacé en `Tcl/Tk` qui permet de visualiser interactivement les différents acteurs et ainsi construire étape par étape la règle de décision. Les figures ci-après sont des images de cet éducatif. Le texte correspond à une rédaction succinte de ce que nous pouvons dire à l'oral. Précisons qu'en cours, nous fournissons aux étudiants un document avec ces mêmes images et de nombreuses questions leur permettant d'avancer par eux-mêmes sur la construction du test.

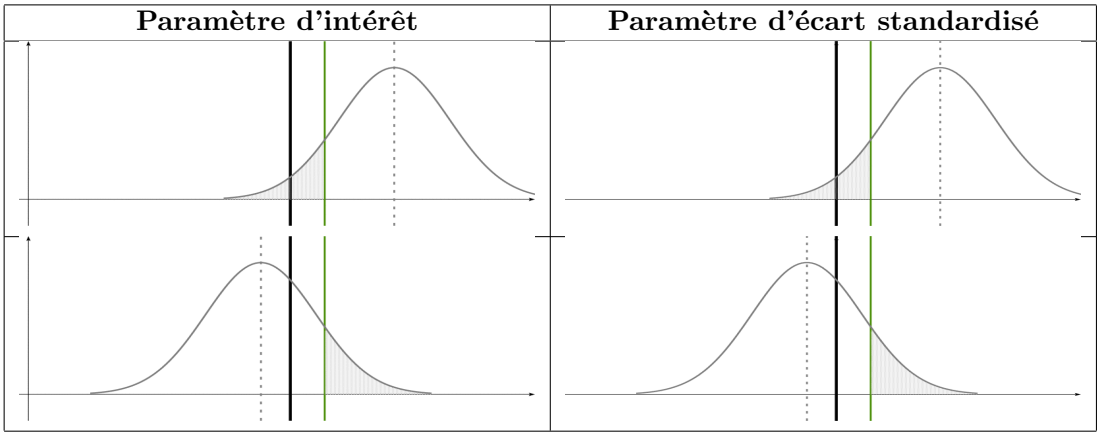
Objectif (avant le jour J)

Commençons par définir l'objectif en traçant le seuil de rentabilité (droite foncée) à savoir 0.15 pour le paramètre d'intérêt 0 pour le paramètre d'écart standardisé. L'objectif de l'industriel est de définir un seuil (droite plus claire) à partir duquel il décidera de lancer le produit B. Nous écrivons cette règle de décision de la manière suivante (pour l'instant) : l'industriel lance le produit B sur le marché si $\boxed{\widehat{\mu^B}(\mathbf{y}) > \mu_{lim}}$ (à partir du paramètre d'intérêt) et $\boxed{\widehat{\delta_{\mu^B,0.15}}(\mathbf{y}) > \delta_{lim}}$ (à partir du paramètre d'écart standardisé). L'industriel est le seul maître dans le choix de ces seuils μ_{lim} et δ_{lim} qu'il aurait naïvement choisis à 0.15 et 0 respectivement.



Erreurs de décision (avant le jour J)

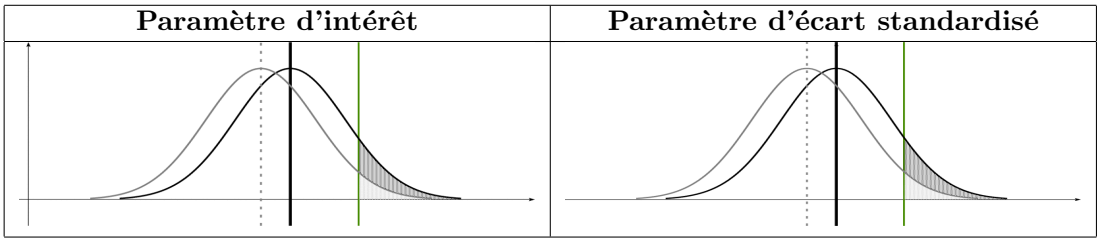
Bien que cela n'ait pas beaucoup d'importance pour le message essentiel qui sortira de cette étape, on fixe la valeur de σ_B.
(1) On commence par prendre conscience que les mauvaises (resp. bonnes) situations pour l'industriel correspondent aux situations pour lesquelles $\mu^B \leq 0.15$ (resp. $\mu^B > 0.15$). A présent, étant donnée une situation fixée (i.e. une droite pointillée correspondant à la valeur d'un μ^B), on connaît (grâce au matheux) la répartition d'une infinité d'estimations du paramètre μ^B; celle-ci est donnée par la courbe "lisse" associée à une $\mathcal{N}(\mu^B, \sqrt{\sigma_B^2/n})$.
(2) Via l'A.E.P., les surfaces grisées correspondent à la proportion parmi l'infinité des estimations possibles si $\mu^B = \ldots$ et $\sigma_B = \ldots$ qui conduisent au vu de la règle de décision à lancer le produit alors que l'on n'aurait pas dû (première situation) et à ne pas lancer le produit alors que l'on aurait dû (deuxième situation). Il s'agit donc de deux surfaces d'erreur de décision que l'on requalifiera littéralement respectivement comme un **risque de devenir pauvre** et un **risque de ne pas devenir riche**.
(3) Parmi ces deux erreurs de décision, la plus grave pour l'industriel correspond au risque de devenir pauvre. La somme des deux risques pouvant aller jusqu'à 100%, on ne peut contrôler les deux risques. Par conséquent, la suite consistera à considérer uniquement les mauvaises situations pour l'indutstriel (i.e. celles pour lesquelles $\mu^B \leq 0.15$).



Pire des situations (avant le jour J)

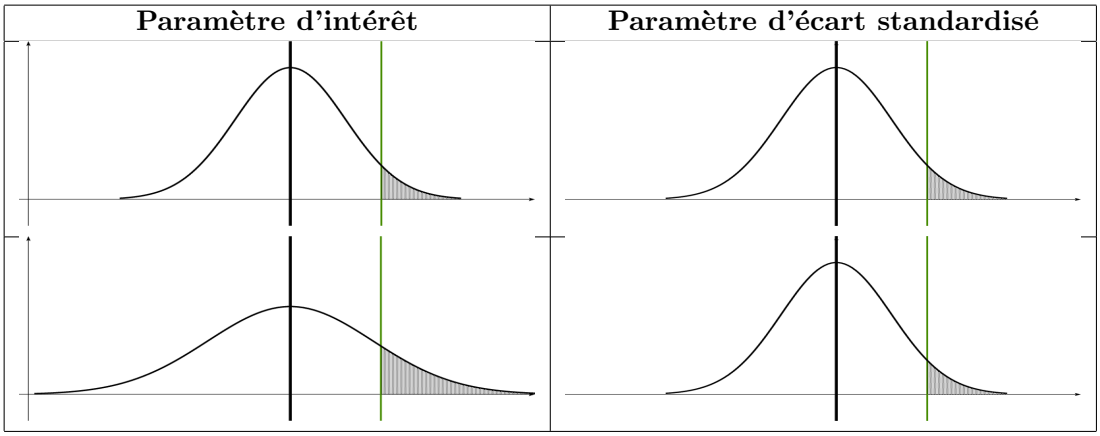
L'étape suivante est facile car très visuelle. Elle consiste à se poser la question : pour une règle de décision fixée, n'y a-t-il pas parmi les mauvaises situations une situation pour laquelle

le risque de devenir pauvre est maximal ? Bien évidemment, cette situation que l'on décide d'appeler **pire des situations** intervient lorsque $\mu^B = 0.15$. On peut alors passer à la suite dès lors que l'on a compris que cette pire des situations ne dépendait pas de la règle de décision que l'on s'était fixée. Ainsi, si on contrôle le risque de devenir pauvre dans la pire des situations, il sera contrôlé pour toutes les mauvaises situations.



Construction d'une règle de décision effective pour un risque de devenir pauvre (maximal) toléré préfixé (avant le jour J)

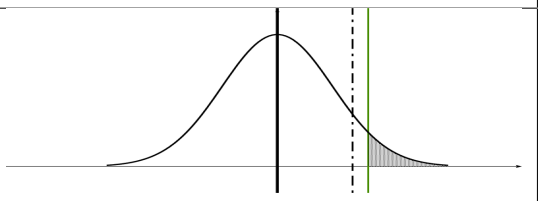
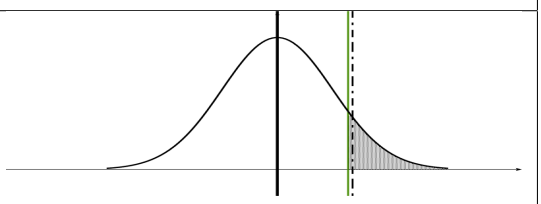
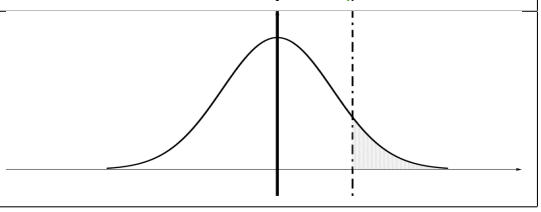
Soit $\alpha = 5\%$, un risque de devenir pauvre toléré par l'industriel. La règle de décision pour ce risque est facile à construire graphiquement (première figure). Le problème n'est malheureusement pas terminé car il faut prendre conscience du résultat du mathématicien énonçant que dans la pire des situations, la répartition de $\widehat{\mu^B}(\mathbf{Y})$ dépend de la variance des réponses des N individus. En revance, la répartition de $\widehat{\delta_{\mu^B,0.15}}(\mathbf{Y})$ ne souffre pas de ce problème (le matheux a tout fait pour!!!). Autrement dit, la règle de décision ne peut être effective qu'en utilisant le paramètre d'écart standardisé.



Application de la règle de décision et p−valeur (après le jour J)

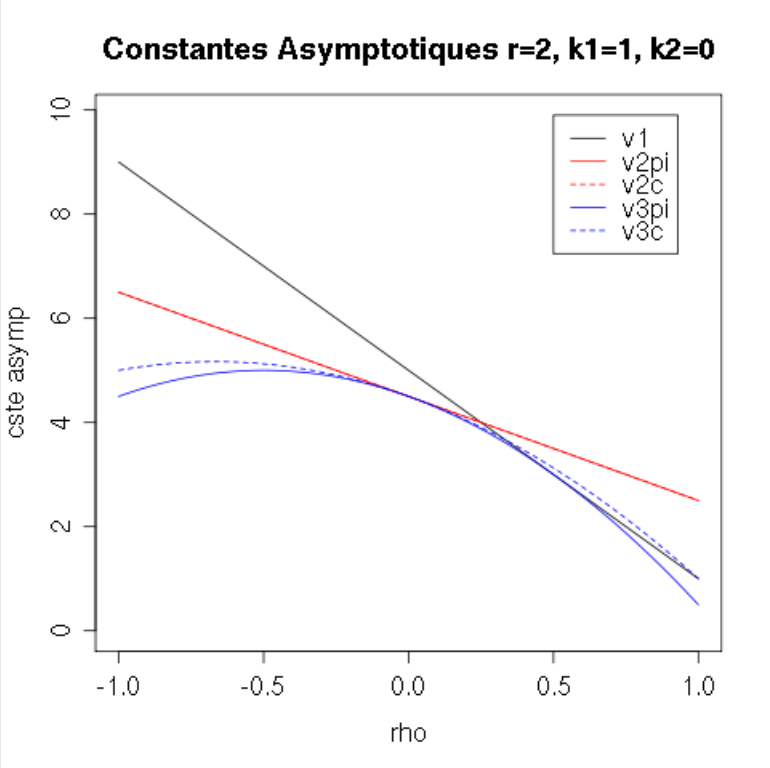
La dernière droite affichée correspond à l'estimation $\widehat{\delta_{\mu^B,0.15}}(\mathbf{y})$. On s'aperçoit donc qu'avec un risque de devenir pauvre (maximal) de 5%, on ne conseille pas l'industriel de lancer son produit sur le marché. Ceci ne veut en aucun prouver que le produit n'est pas rentable puisqu'on ne parviendrait pas à montrer que $\mu^B < 0.15$. Entre parenthèses, avec cette problématique, on comprend en quel sens il n'est donc pas possible d'accepter H_0!!

Revenons à notre problématique, il est évident que si l'industriel souhaite risquer plus ($\alpha = 15\%$, 2-ème figure), on peut lui conseiller de lancer son produit. Avec ces petites manipulations, la question suivante est naturellement : au vu des données, quel est le plus petit risque que l'industriel doit encourir pour lancer son produit sur le marché ? Il suffit de placer la règle de décision au niveau de l'estimation, la surface d'erreur associée correspondrait à ce risque appelé **risque critique** et plus communément **p−valeur** (troisième figure). On espère évidemment que ce risque critique soit raisonnablement petit et en particulier plus petit que le risque toléré par l'industriel (quatrième figure). Autrement dit, on vient de reformuler la règle de décision de manière beaucoup plus simple : l'industriel de lancer le produit si $\boxed{\text{p}-\text{valeur} < \alpha}$.

Paramètre d'intérêt	**Paramètre d'écart standardisé**
	
	
	

Conclusion

Par cette approche, nous pensons répondre à nos attentes. Qui plus est, il nenous faut plus que huit heures pour répondre à une problématique de test et définir la p−valeur. Par ailleurs, l'approche et l'éducatif répondent à d'autres attentes : (1) dans un cadre asymptotique dès lors que le paramètre d'intérêt peut s'écrire comme une moyenne, on traite avec la définition du paramètre d'écart standardisé, de la même manière, un bon nombre de problématiques : variance, différence et rapport de moyennes, différence et rapport de variances,... (2) dès lors que la construction du test est bien comprise le cadre gaussien ne pose plus de problèmes. (3) lorsque le paramètre d'intérêt est une proportion, l'éducatif est également très adapté pour traiter le risque d'erreur de seconde espèce.



[CDD10]
J.-F. Coeurjolly, A. Derbier and R. Drouilhet. Comparaison de différentes stratégies pour estimer une différence de moyennes en présence de données appariées et indépendantes dans un cadre non nécessairement gaussien, *technical report*, 2010.

Comparaison de différentes stratégies pour estimer une différence de moyennes en présence de données appariées et indépendantes dans un cadre non nécessairement gaussien

J.-F. Coeurjolly and R. Drouilhet

May 4, 2009

Abstract

In this paper, we propose different strategies for estimating the difference of means in simultaneous presence of paired data and independent data in a non Gaussian framework. In particular, we consider an imputation of missing data by means or by simple linear regression on the paired data. We compare theoretically and numerically the different estimators. We also prove an asymptotic normality result allowing to construct asymptotic confidence intervals or asymptotic tests.

1 Cadre de travail et notations

On suppose que nos données sont constituées de quatre vecteurs $\boldsymbol{y}^{(1,u)}$, $\boldsymbol{y}^{(2,u)}$, $\boldsymbol{y}^{(1,p)}$ et $\boldsymbol{y}^{(2,p)}$ (de longueur $n^{(1)}$, $n^{(2)}$, $n^{(p)}$ $n^{(p)}$) qui sont des réalisations de quatre vecteurs aléatoires $\boldsymbol{Y}^{(1,u)}$, $\boldsymbol{Y}^{(2,u)}$, $\boldsymbol{Y}^{(1,p)}$ et $\boldsymbol{Y}^{(2,p)}$ où

- $\boldsymbol{Y}^{(1,u)} = \left(Y_1^{(1,u)}, \ldots, Y_{n^{(1)}}^{(1,u)}\right)$ est un vecteur de $n^{(1)}$ réalisations indépendantes d'une variable aléatoire $Y^{(1,u)} \rightsquigarrow \mathcal{L}^{(1)}$ telle que $\boldsymbol{E}(Y^{(1,u)}) = \mu^{(1)}$ et $Var(Y^{(1)}) = \sigma^2_{(1)}$.
- $\boldsymbol{Y}^{(2,u)} = \left(Y_2^{(2,u)}, \ldots, Y_{n^{(1)}}^{(1,u)}\right)$ est un vecteur de $n^{(1)}$ réalisations indépendantes d'une variable aléatoire $Y^{(2,u)} \rightsquigarrow \mathcal{L}^{(2)}$ telle que $\boldsymbol{E}(Y^{(2,u)}) = \mu^{(2)}$ et $Var(Y^{(2)}) = \sigma^2_{(2)}$.
- $\boldsymbol{Y}^{(1,p)} = \left(Y_1^{(1,p)}, \ldots, Y_{n^{(1)}}^{(1,p)}\right)$ et $\boldsymbol{Y}^{(2,p)} = \left(Y_1^{(2,p)}, \ldots, Y_{n^{(1)}}^{(2,p)}\right)$ sont deux vecteurs de $n^{(p)}$ réalisations indépendantes de $Y^{(1,p)}$ (de loi $\mathcal{L}^{(1)}$) et $Y^{(2,p)}$(de loi $\mathcal{L}^{(2)}$) respectivement. Les variables $Y^{(1,p)}$ et $Y^{(2,p)}$ sont éventuellement dépendantes. On notera ρ la corrélation entre ces deux variables.

Les vecteurs $\boldsymbol{Y}^{(1,u)}$ $\boldsymbol{Y}^{(2,u)}$ $\boldsymbol{Y}^{(j,p)}$ sont indépendants (pour $j = 1, 2$).

2 Définition des estimateurs considérés

On s'intéresse à l'estimation de la moyenne $\mu^{(1)} - \mu^{(2)}$. Pour simplifier les notations, on notera $n^{(p)} = n$ et $n^{(1)} := k_1(n) \times n$ et $n^{(2)} := k_2(n) \times n$. Par souci de présentation, nous noterons $k_1(n) = k_1$ et $k_2(n) = k_2$ mais gardons en mémoire que ces quantités pour le moment peuvent éventuellement dépendre de n.

On notera également $\boldsymbol{Y}^{(1,\bullet)} = \left(\boldsymbol{Y}^{(1,p)}, \boldsymbol{Y}^{(1,u)}\right)$ et $\boldsymbol{Y}^{(2,\bullet)} = \left(\boldsymbol{Y}^{(2,p)}, \boldsymbol{Y}^{(2,u)}\right)$. Ainsi,

$$\overline{Y^{(j,\bullet)}} = \frac{n^{(p)}}{n^{(p)}+n^{(j)}}\overline{Y^{(j,p)}} + \frac{n^{(j)}}{n^{(p)}+n^{(p)}}\overline{Y^{(j,u)}} = \frac{1}{1+k_j}\overline{Y^{(j,p)}} + \frac{k_j}{1+k_j}\overline{Y^{(j,u)}}, \text{ pour } j=1,2.$$

L'objectif est de savoir comment estimer au mieux $\theta = \mu^{(1)} - \mu^{(2)}$. Pour cela on va considérer deux stratégies différentes

1. ne pas tenir compte des données indépendantes, i.e. conserver uniquement les données appariées.
2. tenter selon une stratégie particulière de définir des prédictions $\widehat{Y}_i^{(2,u)}$ pour $i = 1, \ldots, n^{(1)}$ et $\widehat{Y}_i^{(1,u)}$ pour $i = 1, \ldots, n^{(2)}$ pour définir l'estimateur de θ par :

$$\frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(\sum_{i=1}^{n^{(p)}}(Y_i^{(1,p)} - Y_i^{(2,p)}) + \sum_{i=1}^{n^{(1)}}(Y_i^{(1,u)} - \widehat{Y}_i^{(2,u)}) + \sum_{i=1}^{n^{(2)}}(\widehat{Y}_i^{(1,u)} - Y_i^{(2,u)})\right)$$

Rentrons un peu plus dans le détail de ces différentes stratégies (en particulier sur les différentes méthodes de prédiction) :

1. Ne pas tenir compte des vecteurs $\boldsymbol{y}^{(1,u)}$ et $\boldsymbol{y}^{(2,u)}$. On a alors l'estimateur

$$\widehat{\theta_1} := \overline{Y^{(1,p)} - Y^{(2,p)}} = \frac{1}{n^{(p)}}\sum_{i=1}^{n^{(p)}} Y_i^{(1,p)} - Y_i^{(2,p)} = \overline{Y^{(1,p)}} - \overline{Y^{(2,p)}}.$$

2. a) Prédiction par la moyenne en utilisant uniquement l'information de l'échantillon apparié: prédire pour $i = 1, \ldots, n^{(1)}$ et pour $i = 1, \ldots, n^{(2)}$ $\widehat{Y}_i^{(2,u)}$ par $\overline{Y^{(2,p)}}$ et $\widehat{Y}_i^{(1,u)}$ par $\overline{Y^{(1,p)}}$. Ceci conduit à l'estimateur

$$\widehat{\theta_2^{\pi}} = \frac{1+k_2}{1+k_1+k_2}\overline{Y^{(1,p)}} + \frac{k_1}{1+k_1+k_2}\overline{Y^{(1,u)}} - \frac{1+k_1}{1+k_1+k_2}\overline{Y^{(2,p)}} - \frac{k_2}{1+k_1+k_2}\overline{Y^{(2,u)}}$$

   b) Prédiction par la moyenne en utilisant le maximum d'information disponible: prédire pour $i = 1, \ldots, n^{(1)}$ et pour $j = 1, \ldots, n^{(2)}$ $\widehat{Y}_i^{(2,u)}$ par $\overline{Y^{(2,p)}}$ et $\widehat{Y}_i^{(1,u)}$ par $\overline{Y^{(1,p)}}$ et définir l'estimateur

$$\widehat{\theta_2^{c}} = \frac{1}{1+k_1}\overline{Y^{(1,p)}} + \frac{k_1}{1+k_1}\overline{Y^{(1,u)}} - \frac{1}{1+k_2}\overline{Y^{(2,p)}} - \frac{k_2}{1+k_2}\overline{Y^{(1,u)}}$$

3. a) Imputation par régression linéaire simple en utilisant uniquement l'information des échantillons appariés
   - pour $i = 1, \ldots, n^{(1)}$ prédire $\widehat{Y}_i^{(2,u)}$ par $\widehat{\beta}_{0,1} + \widehat{\beta}_{1,1}Y_i^{(1,u)}$ où

$$\widehat{\beta}_{1,1} = \frac{cov(\boldsymbol{Y}^{(1,p)}, \boldsymbol{Y}^{(2,p)})}{var(\boldsymbol{Y}^{(1,p)})} \text{ et } \widehat{\beta}_{0,1} = \overline{Y^{(2,p)}} - \widehat{\beta}_{1,1}\overline{Y^{(1,p)}}$$

   - pour $i = 1, \ldots, n^{(2)}$ prédire $\widehat{Y}_i^{(1,u)}$ par $\widehat{\beta}_{0,2} + \widehat{\beta}_{1,2}Y_i^{(2,u)}$ où

$$\widehat{\beta}_{1,2} = \frac{cov(\boldsymbol{Y}^{(1,p)}, \boldsymbol{Y}^{(2,p)})}{var(\boldsymbol{Y}^{(2,p)})} \text{ et } \widehat{\beta}_{0,2} = \overline{Y^{(1,p)}} - \widehat{\beta}_{1,2}\overline{Y^{(2,p)}}$$

On peut alors définir l'estimateur

$$\widehat{\theta_3^{\pi}} = \frac{1+k_1\widehat{\beta}_{1,1}+k_2}{1+k_1+k_2}\overline{Y^{(1,p)}}+\frac{k_1(1-\widehat{\beta}_{1,1})}{1+k_1+k_2}\overline{Y^{(1,u)}}-\frac{1+k_1+k_2\widehat{\beta}_{1,2}}{1+k_1+k_2}\overline{Y^{(2,p)}}-\frac{k_2(1-\widehat{\beta}_{1,2})}{1+k_1+k_2}\overline{Y^{(2,u)}}$$

b) Imputation par régression linéaire simple en utilisant le maximum d'information disponible à chaque fois:

- pour $i = 1, \ldots, n^{(1)}$ prédire $\widehat{Y}_i^{(2,u)}$ par $\widehat{\beta}_{0,1} + \widehat{\beta}_{1,1}Y_i^{(1,u)}$ où

$$\widehat{\beta}_{1,1} = \frac{cov(\boldsymbol{Y}^{(1,p)}, \boldsymbol{Y}^{(2,p)})}{var(\boldsymbol{Y}^{(1,\bullet)})} \text{ et } \widehat{\beta}_{0,1} = \overline{Y^{(2,\bullet)}} - \widehat{\beta}_{1,1}\overline{Y^{(1,\bullet)}}$$

- pour $i = 1, \ldots, n^{(2)}$ prédire $\widehat{Y}_i^{(1,u)}$ par $\widehat{\beta}_{0,2} + \widehat{\beta}_{1,2}Y_i^{(2,u)}$ où

$$\widehat{\beta}_{1,2} = \frac{cov(\boldsymbol{Y}^{(1,p)}, \boldsymbol{Y}^{(2,p)})}{var(\boldsymbol{Y}^{(2,\bullet)})} \text{ et } \widehat{\beta}_{0,2} = \overline{Y^{(1,\bullet)}} - \widehat{\beta}_{1,2}\overline{Y^{(2,\bullet)}}$$

On peut alors définir l'estimateur

$$\begin{aligned}\widehat{\theta_3^{c}} &= \frac{1}{1+k_1}\left(1+\widehat{\beta}_{1,1}\frac{k_1}{1+k_1+k_2}\right)\overline{Y^{(1,p)}} + \frac{k_1}{1+k_1}\left(1-\widehat{\beta}_{1,1}\frac{1}{1+k_1+k_2}\right)\overline{Y^{(1,u)}} \\ &\quad -\frac{1}{1+k_2}\left(1+\widehat{\beta}_{1,2}\frac{k_2}{1+k_1+k_2}\right)\overline{Y^{(2,p)}} - \frac{k_2}{1+k_2}\left(1-\widehat{\beta}_{1,2}\frac{1}{1+k_1+k_2}\right)\overline{Y^{(2,u)}}\end{aligned}$$

3 Variance des estimateurs

On notera $\sigma^2_{(1)} = \sigma^2$ et $\sigma^2_{(2)} = r^2\sigma^2$ pour un $r > 0$. On définit également les quantités

$$\beta_{1,1} = \frac{cov(Y^{(1,p)}, Y^{(2,p)})}{\sigma^2_{(1)}} = \frac{\rho\sigma_{(1)}\sigma_{(2)}}{\sigma^2_{(1)}} = \rho r \text{ et } \beta_{1,2} = \frac{cov(Y^{(1,p)}, Y^{(2,p)})}{\sigma^2_{(2)}} = \frac{\rho\sigma_{(1)}\sigma_{(2)}}{\sigma^2_{(2)}} = \frac{\rho}{r}.$$

Avec ces notations,

$$\beta_{1,1} = \rho r,\ \beta_{1,2} = \frac{\rho}{r} \quad \text{et} \quad \beta_{1,1} \times \beta_{1,2} = \rho^2,\ \frac{\beta_{1,1}}{\beta_{1,2}} = r^2$$

Pour établir la variance des estimateurs, on s'appuie sur les calculs suivants : premièrement

$$Var\left(\left(\overline{Y^{(j,u)}} - \mu^{(j)}\right)\right) = \frac{\sigma^2_{(j)}}{n^{(j)}} = \begin{cases} \frac{\sigma^2}{n} \times \frac{1}{k_1} & \text{pour } j = 1, \\ \frac{\sigma^2}{n} \times \frac{1}{k_2} r^2 = \frac{\sigma^2}{n} \times \frac{1}{k_2}\left(\frac{\beta_{1,1}}{\beta_{1,2}}\right) & \text{pour } j = 2. \end{cases}$$

Ensuite

$$Var\left(\left(\overline{Y^{(j,p)}} - \mu^{(j)}\right)\right) = \frac{\sigma^2_{(j)}}{n^{(p)}} = \begin{cases} \frac{\sigma^2}{n} & \text{pour } j = 1, \\ \frac{\sigma^2}{n} \times r^2 = \frac{\sigma^2}{n} \times \left(\frac{\beta_{1,1}}{\beta_{1,2}}\right) & \text{pour } j = 2. \end{cases}$$

Et enfin

$$Cov\left(\left(\overline{Y^{(1,p)}} - \mu^{(1)}\right), \left(\overline{Y^{(2,p)}} - \mu^{(2)}\right)\right) = \frac{\rho\sigma_{(1)}\sigma_{(2)}}{n^{(p)}} = \frac{\sigma^2}{n}\rho r = \frac{\sigma^2}{n}\beta_{1,1}.$$

En utilisant les résultats précédents on montre la proposition suivante

Proposition 1 *Lorsque* $n \to +\infty$

$$Var(\widehat{\theta}_1) = \frac{\sigma^2}{n} v_1,\ Var(\widehat{\theta}_2^\bullet) = \frac{\sigma^2}{n} v_2^\bullet \text{ et } Var(\widehat{\theta}_3^\bullet) \sim \frac{\sigma^2}{n} v_3^\bullet$$

où les quantités v_1 v_1 *et* $v_j^\bullet$ *pour* $j = 2, 3$ *et* $\bullet = \pi, c$ *sont définies par*

$$\begin{aligned}
v_1 &= 1 + r^2 - 2\rho r = 1 + \frac{\beta_{1,1}}{\beta_{1,2}} - 2\beta_{1,1} \\
v_2^\pi(k_1,k_2) &= \frac{1}{(1+k_1+k_2)^2}\left((1+k_2)^2 + k_1 + r^2((1+k_1)^2+k_2) - 2\rho r(1+k_1)(1+k_2)\right) \\
&= \frac{1}{(1+k_1+k_2)^2}\left((1+k_2)^2 + k_1 + \frac{\beta_{1,1}}{\beta_{1,2}}((1+k_1)^2+k_2) - 2\beta_{1,1}(1+k_1)(1+k_2)\right) \\
v_2^c(k_1,k_2) &= \frac{1}{1+k_1} + \frac{r^2}{1+k_2} - \frac{2\rho r}{(1+k_1)(1+k_2)} = \frac{1}{1+k_1} + \frac{\frac{\beta_{1,1}}{\beta_{1,2}}}{1+k_2} - \frac{2\beta_{1,1}}{(1+k_1)(1+k_2)} \\
v_3^\pi(k_1,k_2) &= \frac{1}{(1+k_1+k_2)^2}\left\{(1+k_1\beta_{1,1}+k_2)^2 + k_1(1-\beta_{1,1})^2 + \frac{\beta_{1,1}}{\beta_{1,2}}\left((1+k_1+k_2\beta_{1,2})^2 + k_2(1-\beta_{1,2})^2\right)\right. \\
&\quad \left. -2\beta_{1,1}(1+k_1\beta_{1,1}+k_2)(1+k_1+k_2\beta_{1,2})\right\} \\
v_3^c(k_1,k_2) &= \left(\frac{1}{1+k_1}\left(1+\beta_{1,1}\frac{k_1}{1+k_1+k_2}\right)\right)^2 + \frac{1}{k_1}\left(\frac{k_1}{1+k_1}\left(1-\beta_{1,1}\frac{1}{1+k_1+k_2}\right)\right)^2 \\
&\quad + r^2\left(\frac{1}{1+k_2}\left(1+\beta_{1,2}\frac{k_2}{1+k_1+k_2}\right)\right)^2 + \frac{r^2}{k_2}\left(\frac{k_2}{1+k_2}\left(1-\beta_{1,2}\frac{1}{1+k_1+k_2}\right)\right)^2 \\
&\quad - \frac{2\rho r}{(1+k_1)(1+k_2)}\left(1+\beta_{1,1}\frac{k_1}{1+k_1+k_2}\right)\left(1+\beta_{1,2}\frac{k_2}{1+k_1+k_2}\right) \\
&= \frac{1}{1+k_1}\left(1+\frac{k_1\beta_{1,1}^2}{(1+k_1+k_2)^2}\right) + \frac{\beta_{1,1}/\beta_{1,2}}{1+k_2}\left(1+\frac{k_2\beta_{1,2}^2}{(1+k_1+k_2)^2}\right) \\
&\quad - \frac{2\beta_{1,1}}{(1+k_1)(1+k_2)}\left(1+\beta_{1,1}\frac{k_1}{1+k_1+k_2}\right)\left(1+\beta_{1,2}\frac{k_2}{1+k_1+k_2}\right)
\end{aligned}$$

4 Comparaison des variances asymptotiques

Soit $k_{1,2} = 1 + k_1 + k_2$ et $k'_{1,2} = (1+k_1)(1+k_2) \times k_{1,2}$

4.1 Comparaison entre imputation par la moyennne et ne rien faire

$$v_1^\pi(k_1,k_2) - v_2^\pi(k_1,k_2) = \frac{1}{k_{1,2}^2}\left(k_1(1+k_1)(1-2\beta_{1,1}) + k_2(1+k_2)\left(\frac{\beta_{1,1}}{\beta_{1,2}} - 2\beta_{1,1}\right) + 2k_1k_2\left(1 + \frac{\beta_{1,1}}{\beta_{1,2}} - \beta_{1,1}\right)\right)$$

pas de généralité vraiment facile à établir mais par exemple pour $k_1 = k_2 = r = 1$ la moyenne est meilleure que ne rien faire si $\rho \leq 4/5$.

$$v_1^c(k_1,k_2) - v_2^c(k_1,k_2) = \frac{k_1}{1+k_1} + \frac{\beta_{1,1}}{\beta_{1,2}}\frac{k_2}{1+k_2} - 2\beta_{1,1}\frac{k_1+k_2+k_1k_2}{(1+k_1)(1+k_2)}$$

pas de généralité vraiment facile à établir. Mais par exemple si $k_1 = k_2 = r = 1$ la moyenne est meilleure que ne rien faire si $\rho \leq 2/3$.

4.2 Comparaison entre imputation par la moyenne et imputation par la régression

$$\begin{aligned}
v_2^\pi(k_1,k_2) - v_3^\pi(k_1,k_2) &= \frac{1}{k_{1,2}^2}\left(-k_1^2\beta_{1,1}^2 - 2k_1k_2\beta_{1,1} - k_1\beta_{1,1}^2 + \frac{\beta_{1,1}}{\beta_{1,2}}(-k_2^2\beta_{1,2}^2 - 2k_1k_2\beta_{1,2} - k_2\beta_{1,2}^2\right.\\
&\quad \left.+2\beta_{1,1}(k_1^2\beta_{1,1} + k_2^2\beta_{1,2} + k_1k_2\beta_{1,1}\beta_{1,2} + k_1\beta_{1,1} + k_2\beta_{1,2})\right)\\
&= \frac{1}{k_{1,2}^2}\left(k_1^2\beta_{1,1}^2 + k_1\beta_{1,1}^2 + k_2^2\beta_{1,1}\beta_{1,2} + k_2\beta_{1,1}\beta_{1,2} + 2k_1k_2(\beta_{1,1}^2\beta_{1,2} - 2\beta_{1,1})\right)\\
&= \frac{1}{k_{1,2}^2}\left(k_1(1+k_1)\beta_{1,1}^2 + k_2(1+k_2)\beta_{1,1}\beta_{1,2} + 2k_1k_2(\beta_{1,1}^2\beta_{1,2} - 2\beta_{1,1})\right)
\end{aligned}$$

Si k_1 ou $k_2 = 0$, la régression est toujours meilleure que la moyenne. Dans le cas contraire, difficile d'établir un résultat général.

Mais par exemple si $k_1 = k_2 = 1 = r$, la moyenne est meilleure que la régression si $\rho \in [0, \sqrt{3}-1]$

$$\begin{aligned}
v_2^c(k_1,k_2) - v_3^c(k_1,k_2) &= -\frac{k_1\beta_{1,1}^2}{(1+k_1)(1+k_1+k_2)^2} - \frac{k_2\beta_{1,1}\beta_{1,2}}{(1+k_2)(1+k_1+k_2)^2}\\
&\quad +\frac{2\beta_{1,1}}{(1+k_1)(1+k_2)}\left(\frac{k_1\beta_{1,1}}{1+k_1+k_2} + \frac{k_2\beta_{1,2}}{1+k_1+k_2} + \frac{k_1k_2\beta_{1,1}\beta_{1,2}}{(1+k_1+k_2)^2}\right)\\
&= \frac{1}{k'_{1,2}}\left(-\beta_{1,1}^2k_1(1+k_2) - \beta_{1,1}\beta_{1,2}k_2(1+k_1) + 2\beta_{1,1}^2k_1(1+k_1+k_2)\right.\\
&\quad \left.+2\beta_{1,1}\beta_{1,2}k_2(1+k_1+k_2) + 2\beta_{1,1}^2\beta_{1,2}k_1k_2\right)\\
&= \frac{1}{k'_{1,2}}\left(\beta_{1,1}^2(2k_1^2 + k_1 + k_1k_2) + \beta_{1,1}\beta_{1,2}(k_2 + 2k_2^2 + k_1k_2) + 2\beta_{1,1}^2\beta_{1,2}k_1k_2\right)\\
&= \frac{1}{k'_{1,2}}\left(\beta_{1,1}^2k_1(1+2k_1) + \beta_{1,1}\beta_{1,2}k_2(1+2k_2) + k_1k_2(\beta_{1,1}^2 + \beta_{1,1}\beta_{1,2} + 2\beta_{1,1}^2\beta_{1,2})\right)\\
&\geq \frac{k_1k_2}{k'_{1,2}}(\rho^2r^2 + \rho^2 + 2\rho^3r) = \frac{k_1k_2}{k'_{1,2}}\rho^2(1 + r^2 + 2\rho r)\\
&\geq \frac{k_1k_2}{k'_{1,2}}\rho^2(1-r)^2 \geq 0
\end{aligned}$$

4.3 Comparaison entre imputation par la régression et ne rien faire

$$\begin{aligned}
v_1 - v_3^\pi(k_1,k_2) &= v_1 - v_2^\pi(k_1,k_2) + v_2^\pi(k_1,k_2) - v_3^\pi(k_1,k_2)\\
&= \frac{1}{k_{1,2}^2}\left(\beta_{1,1}^2 + 1 - 2\beta_{1,1} + k_2(1+k_2)\left(\frac{\beta_{1,1}}{\beta_{1,2}} - 2\beta_{1,1} + \beta_{1,1}\beta_{1,2}\right)\right.\\
&\quad \left.+2k_1k_2\left(1 + \beta_{1,1}^2\beta_{1,2} + \frac{\beta_{1,1}}{\beta_{1,2}} - 3\beta_{1,1}\right)\right)\\
&= \frac{1}{k_{1,2}^2}\left(k_1(1+k_1)(1-\rho r)^2 + k_2(1+k_2)(\rho - r)^2 + 2k_1k_2(\rho^3r - 3\rho r + 1 + r^2)\right)
\end{aligned}$$

Or la fonction $\rho \to \rho^3r - 3\rho r + 1 + r^2$ est décroissante sur $[-1,1]$, donc elle est minorée par $1 + r^2 - 2r = (1-r)^2$. D'où

$$v_1 - v_3^\pi(k_1,k_2) \geq 0.$$

4.4 Comparaison générale dans le cas où k_1 ou $k_2 = 0$

$$v_3^\pi(k_1,0) \leq v_3^c(k_1,0) \leq \left\{ \begin{array}{ll} v_2^\pi(k_1,0) = v_2^c(k_1,0) \leq v_1 & \text{si } \rho \leq \frac{1}{2r} \Leftrightarrow \beta_{1,1} \leq 1/2 \\ v_1 \leq v_2^\pi(k_1,0) = v_2^c(k_1,0) & \text{sinon.} \end{array} \right.$$

et

$$v_3^\pi(0,k_2) \leq v_3^c(0,k_2) \leq \left\{ \begin{array}{ll} v_2^\pi(0,k_2) = v_2^c(0,k_2) \leq v_1 & \text{si } \rho \leq \frac{r}{2} \Leftrightarrow \beta_{1,2} \leq 1/2 \\ v_1 \leq v_2^\pi(0,k_2) = v_2^c(0,k_2) & \text{sinon.} \end{array} \right.$$

Par rapport aux paragraphes précédents, la seule chose qu'il reste à montrer est (pour $k_2 = 0$ par exemple) que $v_3^\pi(k_1,0) \leq v_3^c(k_1,0)$. Or,

$$v_3^c(k_1,0) = \frac{1}{1+k_1}\left(1 + \frac{k_1\beta_{1,1}^2}{(1+k_1)^2}\right) + \frac{\beta_{1,1}}{\beta_{1,2}} - 2\beta_{1,1}\left(\frac{1}{1+k_1} + \frac{k_1\beta_{1,1}}{(1+k_1)^2}\right)$$

et

$$v_3^\pi(k_1,0) = \frac{1+k_1\beta_{1,1}^2}{1+k_1} + \frac{\beta_{1,1}}{\beta_{1,2}} - 2\beta_{1,1}\frac{1+k_1\beta_{1,1}}{1+k_1}.$$

Donc,

$$\begin{array}{rcl} v_3^c(k_1,0) - v_3^\pi(k_1,0) & = & \frac{k_1\beta_{1,1}^2}{(1+k_1)^3}(1-(1+k_1)^2) - 2\frac{k_1\beta_{1,1}^2}{(1+k_1)^2}(1-(1+k_1)) \\ & = & \frac{k_1^2\beta_{1,1}^2}{(1+k_1)^3}\left(-(2+k_1)+2(1+k_1)\right) \\ & = & \frac{k_1^3\beta_{1,1}^2}{(1+k_1)^3} \\ & \geq & 0. \end{array}$$

5 Normalité asymptotique

Commençons par noter que les estimateurs $\widehat{\theta_1}$, $\widehat{\theta_2^\pi}$ $\widehat{\theta_2^c}$, $\widehat{\theta_3^\pi}$ et $\widehat{\theta_3^c}$ du paramètre θ ont tous la forme suivante

$$\widehat{\theta}(\boldsymbol{\alpha}(k_1,k_2)) = \alpha_1(k_1,k_2)\overline{Y^{(1,p)}} + (1-\alpha_1(k_1,k_2))\overline{Y^{(1,u)}} - \alpha_2(k_1,k_2)\overline{Y^{(2,p)}} - (1-\alpha_2(k_1,k_2))\overline{Y^{(2,u)}}$$

où $\boldsymbol{\alpha}(k_1,k_2) = (\alpha_1(k_1,k_2), \alpha_2(k_1,k_2))$ et $\alpha_i(k_1,k_2) \in [0,1]$ $(i = 1,2)$. L'estimateur $\widehat{\theta_1}$ rentre dans ce formalisme en prenant la convention que $\boldsymbol{\alpha}(0,0) = (\alpha_1(0,0), \alpha_2(0,0)) = (1,1)$. Avec ces notations, on a alors

$$\begin{array}{rcl} \widehat{\theta}(\boldsymbol{\alpha}(k_1,k_2)) - \theta & = & \alpha_1(k_1,k_2)\left(\overline{Y^{(1,p)}} - \mu^{(1)}\right) + (1-\alpha_1(k_1,k_2))\left(\overline{Y^{(1,u)}} - \mu^{(1)}\right) \\ & & -\alpha_2(k_1,k_2)\left(\overline{Y^{(2,p)}} - \mu^{(2)}\right) - (1-\alpha_2(k_1,k_2))\left(\overline{Y^{(2,u)}} - \mu^{(2)}\right). \end{array}$$

De même,

$$Var(\widehat{\theta}(\boldsymbol{\alpha}(k_1,k_2))) = \frac{\sigma^2}{n} v(\boldsymbol{\alpha}(k_1,k_2))$$

où

$$v(\boldsymbol{\alpha}(k_1,k_2)) = \alpha_1(k_1,k_2)^2 + \frac{(1-\alpha_1(k_1,k_2))^2}{k_1} + r^2\alpha_2(k_1,k_2)^2 + r^2\frac{(1(\alpha_2(k_1,k_2))^2}{k_2} - 2\rho r\alpha_1(k_1,k_2)\alpha_2(k_1,k_2).$$

Le résultat suivant propose un résultat de normalité asymptotique pour l'estimateur $\widehat{\theta}(\boldsymbol{\alpha}(k_1,k_2))$.

Proposition 2 *Let us recall that we have denoted by* $n^{(p)} = n$, $n^{(1)} = k_1(n) \times n$ *et* $n^{(2)} = k_2(n) \times n$. *Assume for* $i = 1, 2$ *that* $k_i(n) \sim k_i$ *or* $k_i(n) = o(1)$ *as* $n \to +\infty$ *and that* $\alpha_i(k_1, k_2) - \alpha_i(0,0) = \mathcal{O}(k_i(n))$, *then we have the following convergence in distribution*

$$\sqrt{n}\ \frac{\widehat{\theta}(\boldsymbol{\alpha}(k_1,k_2)) \ - \ \theta}{\sigma\sqrt{v(\boldsymbol{\alpha}(k_1,k_2)}} \longrightarrow \mathcal{N}(0,1).$$

Proof. Supposons dans un premier temps que $k_i(n) \sim k_i$, il suffit d'appliquer le théorème central limite aux moyennes empiriques indépendantes $\left(\overline{Y^{(1,u)}} - \mu^{(1)}\right)$, $\left(\overline{Y^{(2,u)}} - \mu^{(2)}\right)$ et $\overline{Z}$ où cette dernière est définie par

$$\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i \text{ et } Z_i = \alpha_1(k_1,k_2)(Y_i^{(1,p)} - \mu^{(1)}) - \alpha_2(k_1,k_2)(Y_i^{(2,p)} - \mu^{(2)}).$$

Si $k_i(n) = k_i = o(1)$, notons

$$\begin{aligned}\sqrt{n}\frac{\widehat{\theta}(\boldsymbol{\alpha}(k_1,k_2)) - \theta}{\sigma\sqrt{v(\boldsymbol{\alpha}(k_1,k_2)}} &= \left(\sqrt{n}\frac{\widehat{\theta}(\boldsymbol{\alpha}(k_1,k_2)) - \theta}{\sigma\sqrt{v(\boldsymbol{\alpha}(k_1,k_2)}} - \sqrt{n}\frac{\widehat{\theta}(\boldsymbol{\alpha}(0,0)) - \theta}{\sigma\sqrt{v(\boldsymbol{\alpha}(k_1,k_2)}}\right) + \sqrt{n}\frac{\widehat{\theta}(\boldsymbol{\alpha}(0,0)) - \theta}{\sigma\sqrt{v(\boldsymbol{\alpha}(k_1,k_2)}} \\ &:= \Delta + U\end{aligned}$$

D'une part lorsque $n \to +\infty$

$$U \sim \sqrt{n}\frac{\widehat{\theta}(\boldsymbol{\alpha}(0,0)) - \theta}{\sigma\sqrt{v(\boldsymbol{\alpha}(0,0))}} = \sqrt{n}\frac{\widehat{\theta_1} - \theta}{\sigma\sqrt{v_1}} \to \mathcal{N}(0,1).$$

D'autre part, on a

$$\Delta = \frac{\sqrt{n}}{\sigma\sqrt{v(\boldsymbol{\alpha}(k_1,k_2))}}\left\{(\alpha_1(k_1,k_2) - \alpha_1(0,0))\left(\overline{Y^{(1,p)}} - \overline{Y^{(1,u)}}\right) - (\alpha_2(k_1,k_2) - \alpha_2(0,0))\left(\overline{Y^{(2,p)}} - \overline{Y^{(2,u)}}\right)\right\}.$$

On peut montrer que

$$\begin{aligned}Var(\Delta) &= \frac{1}{v(\boldsymbol{\alpha}(k_1,k_2)}\Big\{(\alpha_1(k_1,k_2) - \alpha_1(0,0))^2\left(1 + \frac{1}{k_1}\right) + r^2(\alpha_2(k_1,k_2) - \alpha_2(0,0))^2\left(1 + \frac{1}{k_2}\right) \\ &\quad -2\rho r(\alpha_1(k_1,k_2) - \alpha_1(0,0))(\alpha_2(k_1,k_2) - \alpha_2(0,0))\Big\} \\ &= \mathcal{O}(k_1) + \mathcal{O}(k_1) + \mathcal{O}(k_1k_2) = o(1),\end{aligned}$$

lorsque $n \to +\infty$. Ceci montre entre autres que Δ converge en probabilité vers 0 lorsque n tend vers $+\infty$. Le théorème de Slutsky termine la preuve. ■

Pour obtenir un résultat de normalité asymptotique pour $\widehat{\theta_3^{\pi}}$ et $\widehat{\theta_3^{c}}$ rappelons que

$$\widehat{\theta_3^{\pi}} - \widetilde{\theta_3^{\pi}} = \frac{k_1}{1+k_1+k_2}\left(\widehat{\beta}_{1,1} - \beta_{1,1}\right)\left(\overline{Y^{(1,p)}} - \overline{Y^{(1,u)}}\right) - \frac{k_2}{1+k_1+k_2}\left(\widehat{\beta}_{1,2} - \beta_{1,2}\right)\left(\overline{Y^{(2,p)}} - \overline{Y^{(2,u)}}\right)$$

et

$$\begin{aligned}\widehat{\theta_3^{c}} - \widetilde{\theta_3^{c}} &= \frac{k_1}{(1+k_1)(1+k_1+k_2)}\left(\widehat{\beta}_{1,1} - \beta_{1,1}\right)\left(\overline{Y^{(1,p)}} - \overline{Y^{(1,u)}}\right) \\ &\quad -\frac{k_2}{(1+k_2)(1+k_1+k_2)}\left(\widehat{\beta}_{1,2} - \beta_{1,2}\right)\left(\overline{Y^{(2,p)}} - \overline{Y^{(2,u)}}\right)\end{aligned}$$

En utilisant la convergence en probabilité de $\widehat{\beta}_{1,i}$ vers $\beta_{1,i}$ pour $i = 1, 2$ et à nouveau le théorème de Slutsky, on peut montrer que pour $k_1(n)$ et $k_2(n)$ satisfaisant les hypothèses de la proposition précédente que lorsque $n \to +\infty$

$$\sqrt{n}\frac{\widehat{\theta}_3^{\bullet} - \widetilde{\theta}_3^{\bullet}}{\sigma\sqrt{v_3^{\bullet}(k_1, k_2)}} \to 0 \quad (\bullet = \pi, c)$$

en probabilité.

Ainsi, nous obtenons un résultat de normalité asymptotique pour tous les estimateurs développés précédemment, ce qui nous autorise à construire des intervalles de confiance asymptotique ou des tests asymptotiques.

6 Annexe

6.1 Définition des estimateurs

$$\begin{aligned}
\widehat{\theta_2^{\pi}} &= \frac{1}{n^{(1)} + n^{(2)} + n^{(p)}} \left(\sum_{i=1}^{n^{(p)}} (Y_i^{(1,p)} - Y_i^{(2,p)}) + \sum_{i=1}^{n^{(1)}} (Y_i^{(1,u)} - \widehat{Y}_i^{(2,u)}) + \sum_{i=1}^{n^{(2)}} (\widehat{Y}_i^{(1,u)} - Y_i^{(2,u)}) \right) \\
&= \frac{1}{n^{(1)} + n^{(2)} + n^{(p)}} \left(n^{(p)} (\overline{Y^{(1,p)}} - \overline{Y^{(2,p)}}) + n^{(1)} (\overline{Y^{(1,u)}} - \overline{Y^{(2,p)}}) + n^{(2)} (\overline{Y^{(1,p)}} - \overline{Y^{(2,u)}}) \right) \\
&= \frac{1}{n^{(1)} + n^{(2)} + n^{(p)}} \left(n^{(p)} \overline{Y^{(1,p)}} - n^{(p)} \overline{Y^{(2,p)}} + n^{(1)} \overline{Y^{(1,u)}} - n^{(1)} \overline{Y^{(2,p)}} + n^{(2)} \overline{Y^{(1,p)}} - n^{(2)} \overline{Y^{(2,u)}} \right) \\
&= \frac{1 + k_2}{1 + k_1 + k_2} \overline{Y^{(1,p)}} + \frac{k_1}{1 + k_1 + k_2} \overline{Y^{(1,u)}} - \frac{1 + k_1}{1 + k_1 + k_2} \overline{Y^{(2,p)}} - \frac{k_2}{1 + k_1 + k_2} \overline{Y^{(2,u)}}
\end{aligned}$$

$$\begin{aligned}
\widehat{\theta_2^{c}} &= \frac{1}{n^{(1)} + n^{(2)} + n^{(p)}} \left(\sum_{i=1}^{n^{(p)}} (Y_i^{(1,p)} - Y_i^{(2,p)}) + \sum_{i=1}^{n^{(1)}} (Y_i^{(1,u)} - \widehat{Y}_i^{(2,u)}) + \sum_{i=1}^{n^{(2)}} (\widehat{Y}_i^{(1,u)} - Y_i^{(2,u)}) \right) \\
&= \frac{1}{n^{(1)} + n^{(2)} + n^{(p)}} \left(n^{(p)} (\overline{Y^{(1,p)}} - \overline{Y^{(2,p)}}) + n^{(1)} (\overline{Y^{(1,u)}} - \overline{Y^{(2,\bullet)}}) + n^{(2)} (\overline{Y^{(1,\bullet)}} - \overline{Y^{(2,u)}}) \right) \\
&= \frac{1 + \frac{k_2}{1+k_1}}{1 + k_1 + k_2} \overline{Y^{(1,p)}} + \frac{k_1(1 + \frac{k_2}{1+k_1})}{1 + k_1 + k_2} \overline{Y^{(1,u)}} - \frac{1 + \frac{k_1}{1+k_2}}{1 + k_1 + k_2} \overline{Y^{(2,p)}} - \frac{k_2(1 + \frac{k_1}{1+k_2})}{1 + k_1 + k_2} \overline{Y^{(2,u)}} \\
&= \frac{1}{1 + k_1} \overline{Y^{(1,p)}} + \frac{k_1}{1 + k_1} \overline{Y^{(1,u)}} - \frac{1}{1 + k_2} \overline{Y^{(2,p)}} - \frac{k_2}{1 + k_2} \overline{Y^{(1,u)}}
\end{aligned}$$

$$
\begin{aligned}
\widehat{\theta_3^{\pi}} &= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(\sum_{i=1}^{n^{(p)}}(Y_i^{(1,p)}-Y_i^{(2,p)})+\sum_{i=1}^{n^{(1)}}(Y_i^{(1,u)}-\widehat{Y}_i^{(2,u)})+\sum_{i=1}^{n^{(2)}}(\widehat{Y}_i^{(1,u)}-Y_i^{(2,u)})\right)\\
&= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(\sum_{i=1}^{n^{(p)}}(Y_i^{(1,p)}-Y_i^{(2,p)})+\sum_{i=1}^{n^{(1)}}(Y_i^{(1,u)}-\widehat{\beta}_{0,1}-\widehat{\beta}_{1,1}Y_i^{(1,u)})\right.\\
&\qquad\left.+\sum_{i=1}^{n^{(2)}}(\widehat{\beta}_{0,2}+\widehat{\beta}_{1,2}Y_i^{(2,u)}-Y_i^{(2,u)})\right)\\
&= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(\sum_{i=1}^{n^{(p)}}(Y_i^{(1,p)}-Y_i^{(2,p)})+\sum_{i=1}^{n^{(1)}}(Y_i^{(1,u)}-\overline{Y^{(2,p)}}+\widehat{\beta}_{1,1}\overline{Y^{(1,p)}}-\widehat{\beta}_{1,1}Y_i^{(1,u)})\right.\\
&\qquad\left.+\sum_{i=1}^{n^{(2)}}(\overline{Y^{(1,p)}}-\widehat{\beta}_{1,2}\overline{Y^{(2,p)}}+\widehat{\beta}_{1,2}Y_i^{(2,u)}-Y_i^{(2,u)})\right)\\
&= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(n^{(p)}\overline{Y^{(1,p)}}-n^{(p)}\overline{Y^{(2,p)}}+n^{(1)}(1-\widehat{\beta}_{1,1})\overline{Y^{(1,u)}}+n^{(1)}\widehat{\beta}_{1,1}\overline{Y^{(1,p)}}-n^{(1)}\overline{Y^{(2,p)}}\right.\\
&\qquad\left.+n^{(2)}\overline{Y^{(1,p)}}-n^{(2)}(1-\widehat{\beta}_{1,2})\overline{Y^{(2,u)}}-n^{(2)}\widehat{\beta}_{1,2}\overline{Y^{(2,p)}}\right)\\
&= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left((n^{(p)}+n^{()}\widehat{\beta}_{1,1}+n^{(2)})\overline{Y^{(1,p)}}+(n^{(1)}(1-\widehat{\beta}_{1,1})\overline{Y^{(1,u)}}\right.\\
&\left.-(n^{(p)}+n^{()}+n^{(2)}\widehat{\beta}_{1,2})\overline{Y^{(2,p)}}-(n^{(2)}(1-\widehat{\beta}_{1,2})\overline{Y^{(2,u)}}\right)\\
&= \frac{1+k_1\widehat{\beta}_{1,1}+k_2}{1+k_1+k_2}\overline{Y^{(1,p)}}+\frac{k_1(1-\widehat{\beta}_{1,1})}{1+k_1+k_2}\overline{Y^{(1,u)}}-\frac{1+k_1+k_2\widehat{\beta}_{1,2}}{1+k_1+k_2}\overline{Y^{(2,p)}}-\frac{k_2(1-\widehat{\beta}_{1,2})}{1+k_1+k_2}\overline{Y^{(2,u)}}
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\theta}_3^c &= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(\sum_{i=1}^{n^{(p)}}(Y_i^{(1,p)}-Y_i^{(2,p)})+\sum_{i=1}^{n^{(1)}}(Y_i^{(1,u)}-\widehat{Y}_i^{(2,u)})+\sum_{i=1}^{n^{(2)}}(\widehat{Y}_i^{(1,u)}-Y_i^{(2,u)})\right)\\
&= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(\sum_{i=1}^{n^{(p)}}(Y_i^{(1,p)}-Y_i^{(2,p)})+\sum_{i=1}^{n^{(1)}}(Y_i^{(1,u)}-\widehat{\beta}_{0,1}-\widehat{\beta}_{1,1}Y_i^{(1,u)})\right.\\
&\qquad\left.+\sum_{i=1}^{n^{(2)}}(\widehat{\beta}_{0,2}+\widehat{\beta}_{1,2}Y_i^{(2,u)}-Y_i^{(2,u)})\right)\\
&= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(\sum_{i=1}^{n^{(p)}}(Y_i^{(1,p)}-Y_i^{(2,p)})+\sum_{i=1}^{n^{(1)}}(Y_i^{(1,u)}-\overline{Y^{(2,\bullet)}}+\widehat{\beta}_{1,1}\overline{Y^{(1,\bullet)}}-\widehat{\beta}_{1,1}Y_i^{(1,u)})\right.\\
&\qquad\left.+\sum_{i=1}^{n^{(2)}}(\overline{Y^{(1,\bullet)}}-\widehat{\beta}_{1,2}\overline{Y^{(2,\bullet)}}+\widehat{\beta}_{1,2}Y_i^{(2,u)}-Y_i^{(2,u)})\right)\\
&= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(n^{(p)}\overline{Y^{(1,p)}}-n^{(p)}\overline{Y^{(2,p)}}+n^{(1)}(1-\widehat{\beta}_{1,1})\overline{Y^{(1,u)}}+n^{(1)}\widehat{\beta}_{1,1}\overline{Y^{(1,\bullet)}}-n^{(1)}\overline{Y^{(2,\bullet)}}\right.\\
&\qquad\left.+n^{(2)}\overline{Y^{(1,\bullet)}}-n^{(2)}(1-\widehat{\beta}_{1,2})\overline{Y^{(2,u)}}-n^{(2)}\widehat{\beta}_{1,2}\overline{Y^{(2,\bullet)}}\right)\\
&= \frac{1}{n^{(1)}+n^{(2)}+n^{(p)}}\left(\left(n^{(p)}+\widehat{\beta}_{1,1}\frac{n^{(1)}n^{(p)}}{n^{(1)}+n^{(p)}}+\frac{n^{(2)}n^{(p)}}{n^{(1)}+n^{(p)}}\right)\overline{Y^{(1,p)}}\right.\\
&\qquad+\left((1-\widehat{\beta}_{1,1})n^{(1)}+\widehat{\beta}_{1,1}\frac{(n^{(1)})^2}{n^{(1)}+n^{(p)}}+\frac{n^{(1)}n^{(2)}}{n^{(1)}+n^{(p)}}\right)\overline{Y^{(1,u)}}\\
&\qquad-\left(n^{(p)}+\widehat{\beta}_{1,2}\frac{n^{(2)}n^{(p)}}{n^{(2)}+n^{(p)}}+\frac{n^{(1)}n^{(p)}}{n^{(2)}+n^{(p)}}\right)\overline{Y^{(2,p)}}\\
&\qquad\left.-\left((1-\widehat{\beta}_{1,2})n^{(2)}+\widehat{\beta}_{1,2}\frac{(n^{(2)})^2}{n^{(2)}+n^{(p)}}+\frac{n^{(1)}n^{(2)}}{n^{(2)}+n^{(p)}}\right)\overline{Y^{(2,u)}}\right)\\
&= \frac{1}{1+k_1}\left(1+\widehat{\beta}_{1,1}\frac{k_1}{1+k_1+k_2}\right)\overline{Y^{(1,p)}}+\frac{k_1}{1+k_1}\left(1-\widehat{\beta}_{1,1}\frac{1}{1+k_1+k_2}\right)\overline{Y^{(1,u)}}\\
&\quad-\frac{1}{1+k_2}\left(1+\widehat{\beta}_{1,2}\frac{k_2}{1+k_1+k_2}\right)\overline{Y^{(2,p)}}-\frac{k_2}{1+k_2}\left(1-\widehat{\beta}_{1,2}\frac{1}{1+k_1+k_2}\right)\overline{Y^{(2,u)}}
\end{aligned}
$$

Remarquons que

$$
\begin{aligned}
\widehat{\theta_1} - \theta &= \left(\overline{Y^{(1,p)}} - \mu^{(1)}\right) - \left(\overline{Y^{(2,p)}} - \mu^{(2)}\right) \\
\widehat{\theta_2^\pi} - \theta &= \frac{1+k_2}{1+k_1+k_2}\left(\overline{Y^{(1,p)}} - \mu^{(1)}\right) + \frac{k_1}{1+k_1+k_2}\left(\overline{Y^{(1,u)}} - \mu^{(1)}\right) \\
&\quad -\frac{1+k_1}{1+k_1+k_2}\left(\overline{Y^{(2,p)}} - \mu^{(2)}\right) - \frac{k_2}{1+k_1+k_2}\left(\overline{Y^{(2,u)}} - \mu^{(2)}\right) \\
\widehat{\theta_2^c} - \theta &= \frac{1}{1+k_1}\left(\overline{Y^{(1,p)}} - \mu^{(1)}\right) + \frac{k_1}{1+k_1}\left(\overline{Y^{(1,u)}} - \mu^{(1)}\right) - \frac{1}{1+k_2}\left(\overline{Y^{(2,p)}} - \mu^{(2)}\right) - \frac{k_2}{1+k_2}\left(\overline{Y^{(1,u)}} - \mu^{(1)}\right) \\
\widehat{\theta_3^\pi} - \theta &= \frac{1+k_1\widehat{\beta}_{1,1}+k_2}{1+k_1+k_2}\left(\overline{Y^{(1,p)}} - \mu^{(1)}\right) + \frac{k_1(1-\widehat{\beta}_{1,1})}{1+k_1+k_2}\left(\overline{Y^{(1,u)}} - \mu^{(1)}\right) \\
&\quad -\frac{1+k_1+k_2\widehat{\beta}_{1,2}}{1+k_1+k_2}\left(\overline{Y^{(2,p)}} - \mu^{(2)}\right) - \frac{k_2(1-\widehat{\beta}_{1,2})}{1+k_1+k_2}\left(\overline{Y^{(2,u)}} - \mu^{(2)}\right) \\
\widehat{\theta_3^c} - \theta &= \frac{1}{1+k_1}\left(1+\widehat{\beta}_{1,1}\frac{k_1}{1+k_1+k_2}\right)\left(\overline{Y^{(1,p)}} - \mu^{(1)}\right) + \frac{k_1}{1+k_1}\left(1-\widehat{\beta}_{1,1}\frac{1}{1+k_1+k_2}\right)\left(\overline{Y^{(1,u)}} - \mu^{(1)}\right) \\
&\quad -\frac{1}{1+k_2}\left(1+\widehat{\beta}_{1,2}\frac{k_2}{1+k_1+k_2}\right)\left(\overline{Y^{(2,p)}} - \mu^{(2)}\right) - \frac{k_2}{1+k_2}\left(1-\widehat{\beta}_{1,2}\frac{1}{1+k_1+k_2}\right)\left(\overline{Y^{(2,u)}} - \mu^{(2)}\right)
\end{aligned}
$$

signifiant entre autre que les estimateurs $\widehat{\theta_1}$, $\widehat{\theta_2^\pi}$ et $\widehat{\theta_2^c}$ sont des estimateurs sans biais de θ.

6.2 Variances des estimateurs

Ainsi il est facile d'établir dans un premier temps que :

$$
Var(\widehat{\theta_1}) = \frac{\sigma^2}{n}\left(1 + r^2 - 2\rho r\right) = \frac{\sigma^2}{n}\left(1 + \frac{\beta_{1,1}}{\beta_{1,2}} - 2\beta_{1,1}\right) := \frac{\sigma^2}{n} v_1
$$

puis

$$
\begin{aligned}
Var(\widehat{\theta_2^\pi}) &= \frac{\sigma^2}{n}\frac{1}{(1+k_1+k_2)^2}\left((1+k_2)^2 + k_1 + r^2((1+k_1)^2+k_2) - 2\rho r(1+k_1)(1+k_2)\right) \\
&:= \frac{\sigma^2}{n} v_2^\pi(k_1,k_2)
\end{aligned}
$$

et

$$
\begin{aligned}
Var(\widehat{\theta_2^c}) &= \frac{\sigma^2}{n}\left(\frac{1}{(1+k_1)^2} + \left(\frac{k_1}{1+k_1}\right)^2\frac{1}{k_1} + r^2\frac{1}{(1+k_2)^2} + r^2\left(\frac{k_2}{1+k_2}\right)^2\frac{1}{k_2} - \frac{2\rho r}{(1+k_1)(1+k_2)}\right) \\
&= \frac{\sigma^2}{n}\left(\frac{1}{1+k_1} + \frac{r^2}{1+k_2} - \frac{2\rho r}{(1+k_1)(1+k_2)}\right) = \frac{\sigma^2}{n}\left(\frac{1}{1+k_1} + \frac{\frac{\beta_{1,1}}{\beta_{1,2}}}{1+k_2} - \frac{2\beta_{1,1}}{(1+k_1)(1+k_2)}\right) \\
&:= \frac{\sigma^2}{n} v_2^c(k_1,k_2)
\end{aligned}
$$

Let us define $\widetilde{\theta_3^\pi}$ and $\widetilde{\theta_3^c}$ by

$$
\widetilde{\theta_3^\pi} = \frac{1+k_1\beta_{1,1}+k_2}{1+k_1+k_2}\overline{Y^{(1,p)}} + \frac{k_1(1-\beta_{1,1})}{1+k_1+k_2}\overline{Y^{(1,u)}} - \frac{1+k_1+k_2\beta_{1,2}}{1+k_1+k_2}\overline{Y^{(2,p)}} - \frac{k_2(1-\beta_{1,2})}{1+k_1+k_2}\overline{Y^{(2,u)}}
$$

et

$$\begin{aligned}\widetilde{\theta_3^c} &= \frac{1}{1+k_1}\left(1+\beta_{1,1}\frac{k_1}{1+k_1+k_2}\right)\overline{Y^{(1,p)}}+\frac{1}{1+k_1}\left(k_1-\beta_{1,1}\frac{k_1}{1+k_1+k_2}\right)\overline{Y^{(1,u)}}\\ &\quad -\frac{1}{1+k_2}\left(1+\beta_{1,2}\frac{k_2}{1+k_1+k_2}\right)\overline{Y^{(2,p)}}-\frac{1}{1+k_2}\left(k_2-\beta_{1,2}\frac{k_2}{1+k_1+k_2}\right)\overline{Y^{(2,u)}}\end{aligned}$$

Then we obtain,

$$\widehat{\theta_3^\pi}-\widetilde{\theta_3^\pi} = \frac{k_1}{1+k_1+k_2}\left(\widehat{\beta}_{1,1}-\beta_{1,1}\right)\left(\overline{Y^{(1,p)}}-\overline{Y^{(1,u)}}\right)-\frac{k_2}{1+k_1+k_2}\left(\widehat{\beta}_{1,2}-\beta_{1,2}\right)\left(\overline{Y^{(2,p)}}-\overline{Y^{(2,u)}}\right)$$

et

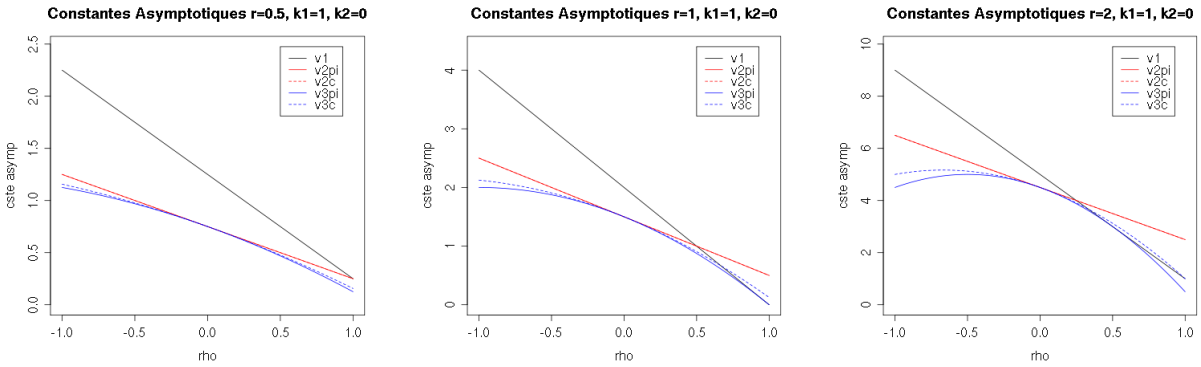
$$\begin{aligned}\widehat{\theta_3^c}-\widetilde{\theta_3^c} &= \frac{k_1}{(1+k_1)(1+k_1+k_2)}\left(\widehat{\beta}_{1,1}-\beta_{1,1}\right)\left(\overline{Y^{(1,p)}}-\overline{Y^{(1,u)}}\right)\\ &\quad -\frac{k_2}{(1+k_2)(1+k_1+k_2)}\left(\widehat{\beta}_{1,2}-\beta_{1,2}\right)\left(\overline{Y^{(2,p)}}-\overline{Y^{(2,u)}}\right)\end{aligned}$$

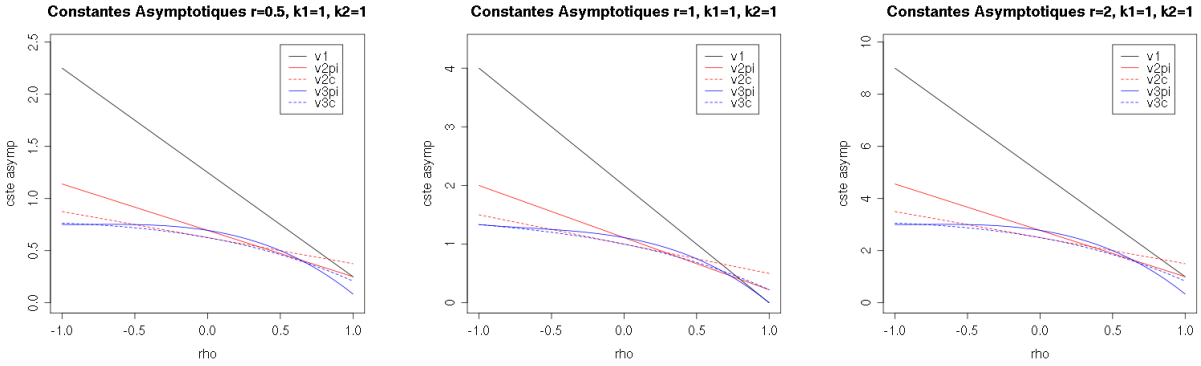
This leads to the almost sure convergence of $\widehat{\theta_3^\pi}-\widetilde{\theta_3^\pi}$ and $\widehat{\theta_3^c}$ -$\widetilde{\theta_3^c}$ towards 0. [DEPEND DE K1 K2 ...A FINIR] Hence,

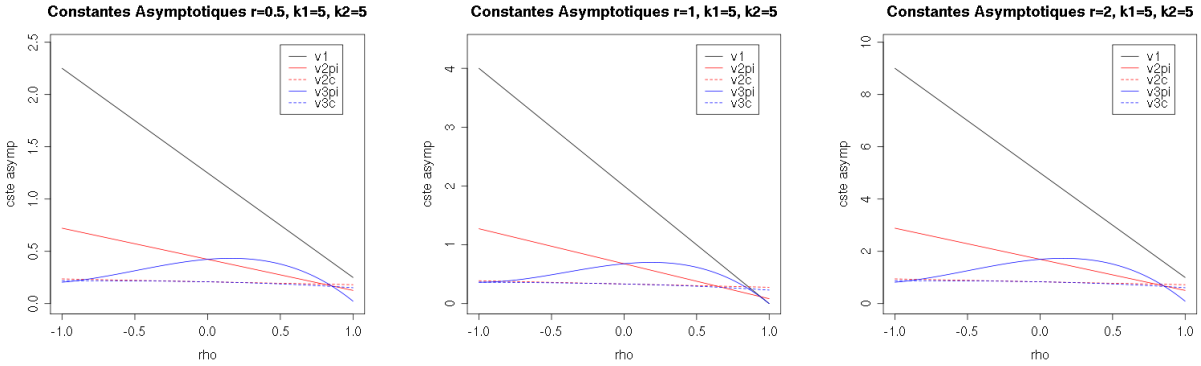
$$\begin{aligned}Var(\widehat{\theta_3^\pi}) &\sim Var(\widetilde{\theta_3^\pi})\\ &= \frac{\sigma^2}{n}\frac{1}{(1+k_1+k_2)^2}\left\{(1+k_1\beta_{1,1}+k_2)^2+k_1(1-\beta_{1,1})^2+r^2\left((1+k_1+k_2\beta_{1,2})^2+k_2(1-\beta_{1,2})^2\right)\right.\\ &\quad \left.-2\rho r(1+k_1\beta_{1,1}+k_2)(1+k_1+k_2\beta_{1,2})\right\}\\ &= \frac{\sigma^2}{n}\frac{1}{(1+k_1+k_2)^2}\left\{(1+k_1\beta_{1,1}+k_2)^2+k_1(1-\beta_{1,1})^2+\frac{\beta_{1,1}}{\beta_{1,2}}\left((1+k_1+k_2\beta_{1,2})^2+k_2(1-\beta_{1,2})^2\right)\right.\\ &\quad \left.-2\beta_{1,1}(1+k_1\beta_{1,1}+k_2)(1+k_1+k_2\beta_{1,2})\right\}\\ &:= \frac{\sigma^2}{n}v_3^\pi(k_1,k_2)\end{aligned}$$

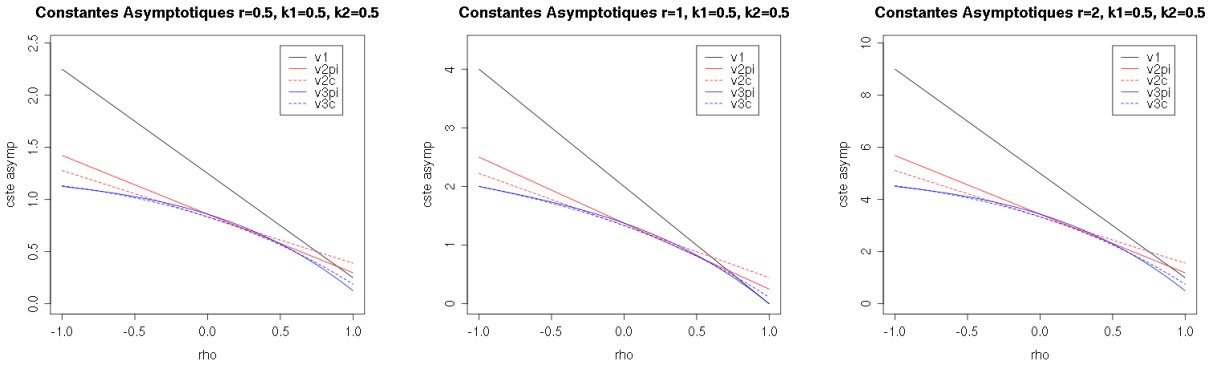
and

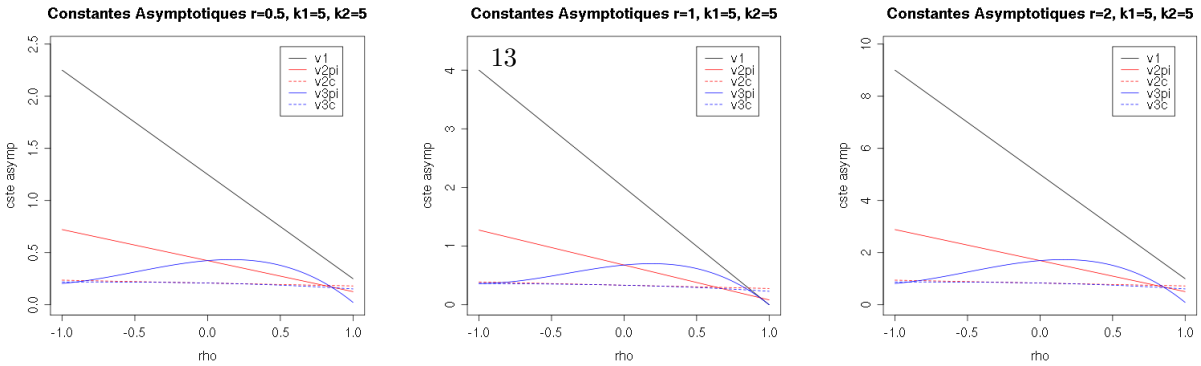
$$\begin{aligned}Var(\widehat{\theta_3^c}) &\sim Var(\widetilde{\theta_3^c})\\ &= \frac{\sigma^2}{n}\left\{\left(\frac{1}{1+k_1}\left(1+\beta_{1,1}\frac{k_1}{1+k_1+k_2}\right)\right)^2+\frac{1}{k_1}\left(\frac{k_1}{1+k_1}\left(1-\beta_{1,1}\frac{1}{1+k_1+k_2}\right)\right)^2\right.\\ &\quad +r^2\left(\frac{1}{1+k_2}\left(1+\beta_{1,2}\frac{k_2}{1+k_1+k_2}\right)\right)^2+\frac{r^2}{k_2}\left(\frac{k_2}{1+k_2}\left(1-\beta_{1,2}\frac{1}{1+k_1+k_2}\right)\right)^2\\ &\quad \left.-\frac{2\rho r}{(1+k_1)(1+k_2)}\left(1+\beta_{1,1}\frac{k_1}{1+k_1+k_2}\right)\left(1+\beta_{1,2}\frac{k_2}{1+k_1+k_2}\right)\right\}\\ &= \frac{\sigma^2}{n}\left\{\frac{1}{1+k_1}\left(1+\frac{k_1\beta_{1,1}^2}{(1+k_1+k_2)^2}\right)+\frac{\beta_{1,1}/\beta_{1,2}}{1+k_2}\left(1+\frac{k_2\beta_{1,2}^2}{(1+k_1+k_2)^2}\right)\right.\\ &\quad \left.-\frac{2\beta_{1,1}}{(1+k_1)(1+k_2)}\left(1+\beta_{1,1}\frac{k_1}{1+k_1+k_2}\right)\left(1+\beta_{1,2}\frac{k_2}{1+k_1+k_2}\right)\right\}\\ &:= \frac{\sigma^2}{n}v_3^c(k_1,k_2)\end{aligned}$$


Constantes Asymptotiques r=0.5, k1=1, k2=0
Constantes Asymptotiques r=1, k1=1, k2=0
Constantes Asymptotiques r=2, k1=1, k2=0
cste asymp
rho
v1
v2pi
v2c
v3pi
v3c


Constantes Asymptotiques r=0.5, k1=1, k2=1
Constantes Asymptotiques r=1, k1=1, k2=1
Constantes Asymptotiques r=2, k1=1, k2=1
cste asymp
rho
v1
v2pi
v2c
v3pi
v3c


Constantes Asymptotiques r=0.5, k1=5, k2=5
Constantes Asymptotiques r=1, k1=5, k2=5
Constantes Asymptotiques r=2, k1=5, k2=5
cste asymp
rho
v1
v2pi
v2c
v3pi
v3c


Constantes Asymptotiques r=0.5, k1=0.5, k2=0.5
Constantes Asymptotiques r=1, k1=0.5, k2=0.5
Constantes Asymptotiques r=2, k1=0.5, k2=0.5
cste asymp
rho
v1
v2pi
v2c
v3pi
v3c


Constantes Asymptotiques r=0.5, k1=5, k2=5
Constantes Asymptotiques r=1, k1=5, k2=5
Constantes Asymptotiques r=2, k1=5, k2=5
cste asymp
rho
v1
v2pi
v2c
v3pi
v3c

Activité scientifique

Les travaux ont été rangés en trois catégories : (i) revues internationales (ii) conférences (inter)nationales et rapports non soumis. (iii) Logiciels sous la forme de package `R`.

Les articles et conférences sont étiquetés par ordre alphabétique des auteurs et rangés par ordre chronologique. Les références générales utilisées dans ce mémoire sont présentées à la fin du document, étiquetées numériquement (pour les différencier) et rangées par ordre alphabétique. .

(i) Articles dans des revues internationales parus ou soumis

[Coe00] J.-F. Coeurjolly. Simulation and identification of the fractional Brownian motion : a bibliographical and comparative study. *J. Stat. Softw.*, 5(7) :1–53, 2000.

[Coe01] J.-F. Coeurjolly. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Infer. Stoch. Process.*, 4(2) :199–227, 2001.

[CI01] J.-F. Coeurjolly and J. Istas. Cramèr-Rao bounds for fractional Brownian motions. *Stat. Probab. Lett.*, 53(4) :435–447, 2001.

[Coe05] J.-F. Coeurjolly. Identification of multifractional Brownian motion. *Bernoulli*, 11(6) :987–1008, 2005.

[CDR07] J.-F. Coeurjolly, R. Drouilhet, and J.-F. Robineau. Normalized information-based divergences. *Problems of Information Transmission*, 43(3) :167–189, 2007.

[Coe08a] J.-F. Coeurjolly. Bahadur representation of sample quantiles for functional of Gaussian dependent sequences under a minimal assumption. *Stat. Probab. Lett.*, 78(15) :2485–2489, 2008.

[Coe08b] J.-F. Coeurjolly. Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. *Annals of Statistics*, 36(3) :1404–1434, 2008.

[BCD08] J.-M. Billiot, J.-F. Coeurjolly, and R. Drouilhet. Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes. *Electronic Journal of Statistics*, 2 :234–264, 2008.

[CDdMR09] J.-F. Coeurjolly, R. Drouilhet, P. Lafaye de Micheaux, and J.-F. Robineau. asympTest : an `R` package for performing parametric statistical tests and confidence intervals based on the central limit theorem. *R Journal*, 2 :26–30, 2009. Long version (19p.) in arxiv : `http://arxiv.org/abs/0902.0506`.

[AC10] S. Achard and J.-F. Coeurjolly. Discrete variations for the fractional Brownian motion in presence of outliers and/or an additive noise. *Statistics Surveys*, 4 :117–147, 2010.

[CD10] J.-F. Coeurjolly and R. Drouilhet. Asymptotic properties of the maximum pseudo-likelihood estimator for stationary Gibbs point processes including the Lennard-Jones model. *Electronic Journal of Statistics*, 4 :677–706, 2010.

[ACLP10] P.O. Amblard, J.-F. Coeurjolly, F. Lavancier, and A. Philippe. Basic properties of the multivariate fractional Brownian motion. *to appear in Bulletin de la Société Mathématique de France, Séminaires et Congrés*, 2010. `http://arxiv.org/abs/1007.0828`.

[CDDL10] J.-F. Coeurjolly, D. Dereudre, R. Drouilhet, and F. Lavancier. Takacs-fiksel method for stationary marked Gibbs point processes. *in revision for Scandinavian Journal of Statistics*, 2010. `http://arxiv.org/abs/1007.3351`.

[CL10] J.-F. Coeurjolly and F. Lavancier. Residuals for stationary marked Gibbs point processes. *submitted*, 2010. `http://arxiv.org/abs/1002.0857`.

[BC10] J.-C. Breton and J.-F. Coeurjolly. Confidence intervals for the Hurst parameter of a fractional Brownian motion based on concentration inequalities. *submitted*, 2010. `http://arxiv.org/abs/0902.0506`.

[CAA10] J.-F. Coeurjolly, P.O. Amblard, and S. Achard. Normalized causal and well-balanced multivariate fractional Brownian motion. *submitted*, 2010. `http://arxiv.org/abs/1007.2109`.

(ii) Conférences (inter)nationales - Revues nationales - Rapports techniques non soumis

[Coe99] J.-F. Coeurjolly. Identification du mouvement Brownien fractionnaire par variations discrètes. In *31èmes journées de statistiques*, pages 479–482, Grenoble, France, 1999. SFdS.

[Coe00] J.-F. Coeurjolly. Identification du mouvement Brownien multifractionnaire. In *32èmes journées de statistiques*, pages 824–827, Fès, Maroc, 2000. SFdS.

[CD04] J.-F. Coeurjolly and R. Drouilhet. L'enseignement de la statistique et des concepts probabilistes par une approche expérimentale comme complément visuel à une approche plus classique. In *Modélisation Stochastique et Statistique, MSS'2004*, Alger, Algérie, 2004. invited conference.

[KES+04a] C. Kretzschmar, S. Ebermeyer, M. Sevin, P. Grousson, R. Drouilhet, and J.-F. Coeurjolly. Enquête panel de suivi des bénéficiaires du dispositif CES. Rapport technique, DRTEFP et DR ANPE, 2004. ISBN 2-11-09782-9.

[KES+04b] C. Kretzschmar, S. Ebermeyer, M. Sevin, P. Grousson, R. Drouilhet, and J.-F. Coeurjolly. Enquête panel de suivi des bénéficiaires du dispositif SIFE. Rapport technique, DRTEFP et DR ANPE, 2004. ISBN 2-11-09783-7.

[KES+04c] C. Kretzschmar, S. Ebermeyer, M. Sevin, P. Grousson, R. Drouilhet, and J.-F. Coeurjolly. Enquête panel de suivi des bénéficiaires du dispositif TRACE. Rapport technique, DRTEFP et DR ANPE, 2004. ISBN 2-11-09782-5.

[Coe05] J.-F. Coeurjolly. L-type estimators of the hölder index of locally self-similar Gaussian processes. In *Second colloqium "Self-similarity and applications"*, Toulouse, France, 2005.

[CD05a] J.-F. Coeurjolly and R. Drouilhet. Approche expérimentale des probabilités comme complément à une approche plus classique. In *37èmes Journées de statistique*, Pau, France, 2005. SFDS. CDROM.

[CD05b] J.-F. Coeurjolly and R. Drouilhet. Construction d'un test d'hypothèses par une approche visuelle et une approche expérimentale des probabilités. In *37èmes journées de statistiques*, Pau, France, 2005. SFdS. CDROM.

[CBD06] J.-F. Coeurjolly, J.-M. Billiot, and R. Drouilhet. Maximum pseudo-likelihood estimator for nearest-neighbour Gibbs point processes. In *6th French-Danish Workshop on Spatial Statistics (invited speaker)*, Skagen, Denmark, 2006.

[Coe07] J.-F. Coeurjolly. Maximum pseudo-likelihood estimator for stationary marked Gibbs point processes. In *Colloque international de Statistique Appliquée pour le Développement en Afrique*, Cotonou, Bénin, 2007. SADA'07.

[CBD07] J.-F. Coeurjolly, J.-M. Billiot, and R. Drouilhet. Maximum pseudo-likelihood for Gibbs point processes. In *39èmes Journées de statistique*, Angers, France, 2007. SFDS. CDROM.

[CBD08] J.-F. Coeurjolly, J.-M. Billiot, and R. Drouilhet. New advances on asymptotic properties for maximum pseudo-likelihood estimator for exponential family of marked Gibbs points processes. In *7th French-Danish Workshop on Spatial Statistics (invited speaker)*, Toulouse, France, 2008.

[Coe09] J.-F. Coeurjolly. Propriétés asymptotiques pour les résidus de processus de Gibbs marqués stationnaires. In *41èmes Journées de statistique*, Bordeaux, France, 2009. SFDS.

[MCLT09] M. Nguile Makao, J.-F. Coeurjolly, B. Liquet, and J.-F. Timsit. Prédiction de la pneumonie nosocomiale à l'aide d'un modèle multi-états. In *41èmes Journées de statistique*, Bordeaux, France, 2009. SFDS. `http://hal.inria.fr/inria-00386790/fr/`.

[HLT+09] T. Hurtut, P.-E. Landes, J. Thollot, Y. Gousseau, R. Drouilhet, and J.-F. Coeurjolly. Appearance-guided synthesis of element arrangements by example. In *7th International Symposium on Non-Photorealistic Animation and Rendering, NPAR 2009, August, 2009*, New Orleans, Etats-Unis, 2009.

[CL09] J.-F. Coeurjolly and F. Lavancier. Asymptotic properties of the residuals process for stationary marked Gibbs point processes. In *Stochastic Processes and their Applications*, Berlin, Germany, 2009.

[CL10] J.-F. Coeurjolly and F. Lavancier. Goodness-of-fit tests for stationary marked Gibbs point processes. Copenhagen, Denmark, 2010. Workshop Spatial Statistics and Image Analysis in Biology.

[AKC+10] S. Achard, S. Kremer, J.-F. Coeurjolly, M. Schenk, F. Renard, C. Ong-Nicolas, J. Namer, V. Mutschler, F. Schneider, and C. Delon-Martin. Global functional disconnections and long-memory alterations in consciousness disorder patients. In *OHBM*, Human Brain Mapping, Barcelona, Spain, 2010.

[CAA10a] J.-F. Coeurjolly, P.O. Amblard, and S. Achard. On multivariate fractional Brownian motion and multivariate fractional Gaussian noise. In *European Signal Processing Conference*, to appear in European Signal Processing Conference, Aalborg, Denmark, 2010. EUSIPCO.

[CAA10b] J.-F. Coeurjolly, P.O. Amblard, and S. Achard. Simulation of self-similar networks. Bristol, U.K., 2010. Workshop Statistical modelling and inference for networks.

[Coe10] J.-F. Coeurjolly. Inférence statistique pour les modèles Gibbsiens. In *Journées MAS de la SMAI*, Bordeaux, France, 2010. (conférencier invité).

[CDD10] J.-F. Coeurjolly, A. Derbier, and R. Drouilhet. Comparaison de différentes stratégies pour estimer une différence de moyennes en présence de données appariées et indépendantes dans un cadre non nécessairement gaussien. *in preparation*, 2010.

(iii) Développement Logiciel

- `SimEstFBM` : ensemble de fonctions `R` sur le problème de simulation et d'identification du mouvement brownien fractionnaire associée à la publication [Coe00] (`http://www.jstatsoft.org/v05/i07`).
- `asympTest` package `R` (disponible sur le CRAN) implémentant dans le cadre asymptotique des tests d'hypothèses paramétriques et intervalles de confiance. (avec R. Drouilhet, P. Lafaye de Micheaux et J.-F. Robineau).
- `dvFBM` : package `R` (disponible sur le CRAN) regroupant un ensemble de procédures permettant d'estimer l'exposant de Hurst d'un bruit gaussien fractionnaire perturbé. (avec S. Achard).
- `CqlsQsa` package `R` pour sélectionner des variables parmi un grand nombre dans un problème de discrimination. Librairie développée dans le cadre de la thèse de J.-Robineau avec ce dernier et Rémy Drouilhet.
- `CqlsTest` package `R` avec utilisation de tcltk (et ses nombreuses librairies) fournissant un logiciel d'initiation à la construction de test d'hypothèses, développé avec (et maintenu par) Rémy Drouilhet . Nous avons utilisé cet outil pendant plusieurs années pour nos propres enseignements.

Les deux dernières librairies ne sont pas disponibles en ligne.

(iv) Encadrement doctoral

- *Jean-François Robineau* (co-encadrée avec Rémy Drouilhet et Catherine Garbay) : "méthodes de sélection de variables (parmi un grand nombre) dans un cadre de discrimination", thèse soutenue en Décembre 2005.
- *Molière Nguile Makao* (co-encadrée avec Benoît Liquet et Jean-François Timsit) : "pneumonie nosocomiale acquise sous ventilation mécanique : prédiction du diagnostic et influence sur le pronostic", soutenance prévue en Novembre 2010.
- *Nadia Morsli* : "Inférence non paramétrique pour les processus ponctuels de Gibbs", thèse débutée en Décembre 2010.

(v) Projets de recherche et activité contractuelle

- Collaboration avec l'entreprise Economie et Humanisme (Mélanie Sévin et Sophie Ebermeyer) et le LABSAD (R. Drouilhet et J.-F. Coeurjolly) avec l'entreprise **Economie et Humanisme** (Mélanie Sévin et Sophie Ebermeyer) (2003) - **Dépouillement d'enquêtes de données de panel** concernant le devenir de travailleurs bénéficiaires de contrats d'insertion.
- Projet MSH-Alpes (2004-06) - **"calcul intensif pour l'analyse spatiale de corpus de données nombreuses"**, ce projet fédère deux équipes issues des laboratoires LabSAD et TIMC (IMAG-UMR CNRS 255). Ce projet souhaite profiter de l'existence de machines en cluster (à la MSH-Alpes) pour explorer l'apport de techniques de parallélisation.

- Contrat de collaboration avec le LIP (Laboratoire Interuniversitaire de Psychologie - Aurélie Derbier) et le LJK (JF Coeurjolly et Rémy Drouilhet) (2007-2010) dans le cadre de la thèse d'Aurélie Derbier - "**Etude épidémiologique des facteurs de risque et de protection du suicide**. Mise en place et évaluation du programme Coping And Support Training au collège et au lycée". Traitements statistiques de questionnaires dans le but de mesurer l'efficacité d'ateliers de prévention du suicide.

- Projet BQR MoDyC (2008-2009) - **Modélisation Dynamique du Cerveau**. Ce projet fédère deux laboratoires le GIPSA-lab (Sophie Achard, Marc Sato et Bertrand Rivet, porteur du projet) et le LJK (JF Coeurjolly et Pierre Lafaye De Micheaux). L'objectif est d'obtenir une description spatiale et temporelle des activations motrices, somatosensorielles et auditives liées à la production et perception des voyelles du Français en utilisant des données d'imagerie par résonance magnétique fonctionnelle. Ce projet a une durée de vie d'une année.

- Projet Réseau National des Systèmes Complexes (2009-2010) - **"Consciousness disorders implications in the brain information network : measure and analysis of functional dynamic connectivity"**. Mêmes acteurs que le projet suivant.

- Projet ANR jeunes chercheurs (2010 -) **"InfoNetComaBrain"**. Méthodes statistiques pour l'étude des réseaux de connectivité fonctionnelle cérébrale, fusion avec la connectivité anatomique. Vers un nouvel outil diagnostique et pronostique pour l'évaluation des désordres de la conscience. Ce projet est à l'interface de plusieurs domaines de compétence : traitement statistique du signal, analyse et visualisation de réseaux complexes, neurosciences. Ces projet fédère le GIPSA-lab (S. Achard, porteur du projet), le LJK (JF Coeurjolly), le Grenoble Institut des Neurosciences (C. Delon-Martin) et le CHU de Strasbourg (S. Kremer et F. Schneider). L'objectif de l'axe dans lequel j'interviens est de développer des modèles multivariés de processus autosimilaires ainsi que des méthodologies permettant d'inférer sur ces modèles.

- Projet IXXI Systèmes Complexes (2010-2012) - **"Sunspot"**. L'objectif de ce projet est la modélisation spatio-temporelle du phénomène des taches solaires. Il engage trois laboratoires : LJK (JF Coeurjolly, porteur du projet), GIPSA-lab (Pierre-Olivier Amblard et Nicolas Le Bihan) et le Laboratoire de Planétologie de Grenoble (Jean Lilenstein). Les difficultés inhérentes à ce sujet proviennent de la nature des données : processus ponctuel spatio-temporel à valeurs sur une sphère.

(vi) Participation à la vie de la recherche

- Participation à l'organisation des journées MAS, Grenoble, (2002).
- Membre des commissions de spécialistes, section 26, de Grenoble 2 et de l'IUT (2002-2006).
- Relecteur pour les revues *Computational Statistics and Data Analysis, ESAIM Probability and Statistics, Journal of Fourier Analysis, Electronic Journal of Statistics, Bernoulli, Statistical Science, Statistics and Computing, Applied Computational and Harmonic Analysis et Stochastic Processes and their Applications.*
- Rapporteur de thèse :
  - Ege Rubak (dirigée par Jesper Møller) "Likelihood Based Inference and Diagnostics for Spatial Data Models", soutenue le 15 octobre 2010 à Aalborg (Danemark).
  - Olaf Kouamo (dirigée par Eric Moulines) "Long memory time series analysis using the wavelet domain", soutenance prévue fin Janvier 2010 à l'ENST d Paris.

- Membre du conseil du Laboratoire Jean Kuntzmann (2008-).
- Co-responsable (avec Jean-Baptiste Durand) du séminaire du Département Statistiques du LJK (Sept. 2009 -).

Bibliographie générale

[1] P. Abry and V. Pipiras. Wavelet-based synthesis of the Rosenblatt process. *Signal Processing*, 86(9) :2326–2339, 2006. 37

[2] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1) :63, 2006. 35, 36

[3] P.O. Amblard, O. Michel, and S. Zozor. On the estimation of the entropy using k-th nearest-neigbours. In *Conference on Mathematics in Signal Processing*, Cirencester, Royaume-Uni, 2009. IMA. 37

[4] P.K. Andersen, Ø Borgan, R.D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer-Verlag, New York, 1993. 301

[5] M.A. Arcones. Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *The Annals of Probability*, pages 2242–2274, 1994. 23, 31

[6] A. Baddeley and J. Møller. Nearest-neighbour Markov point processes and random sets. *International Statistical Review*, pages 89–121, 1989. 296

[7] A. Baddeley, J. Møller, and A.G. Pakes. Properties of residuals for spatial point processes. *Annals of the Institute of Statistical Mathematics*, 60(3) :627–649, 2008. 301, 306

[8] A. Baddeley and R. Turner. Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Australian and New Zealand Journal of Statistics*, 42 :283–322, 2000. 296, 303

[9] A. Baddeley, R. Turner, J. Møller, and M. Hazelton. Residual analysis for spatial point processes. *Journal of the Royal Statistical Society (series B)*, 67 :1–35, 2005. 301, 306

[10] A. J. Baddeley and M. N. M. Van Lieshout. Area-interaction point processes. *Ann. Inst. Statist. Math.*, 47(4) :601–619, 1995. 300

[11] B. Von Bahr and C.G. Esseen. Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *The Annals of Mathematical Statistics*, 36(1) :299–303, 1965. 29

[12] J.-M. Bardet and D. Surgailis. Nonparametric estimation of the local Hurst function of multifractional processes. *preprint*, 2010. 28, 32, 41

[13] S. Baykut, T. Akgül, and S. Ergintav. Estimation of spectral exponent parameter of 1/f process in additive white background noise. *EURASIP Journal on Advances in Signal Processing*, 2007(15) :1–7, 2007. 24

[14] A. Bégyn. Functional limit theorems for generalized quadratic variations of Gaussian processes. *Stochastic Processes and their Applications*, 117(12) :1848–1869, 2007. 28, 32, 41

[15] A. Benassi, S. Cohen, and J. Istas. Identifying the multifractional function of a Gaussian process. *Statistics & Probability Letters*, 39(4) :337–345, 1998. 12, 27

[16] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate : a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 289–300, 1995. 33

[17] K.K. Berthelsen and J. Møller. Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics*, 30(3) :549–564, 2003. 301

[18] E. Bertin, J.-M. Billiot, and R. Drouilhet. Existence of "Nearest-Neighbour" Gibbs Point Models. *Ann. Appl. Probab.*, 31 :895–909, 1999. 295, 296, 297, 298

[19] J. Besag. Spatial interaction and the statistical analysis of lattice system. *Journal of the Royal Statistical Society (series B)*, 26 :192–236, 1974. 301, 302

[20] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds : Ii. *Annals of Statistics*, 33(3) :1225–1259, 2005. 312

[21] J.-M. Billiot. Asymptotic properties of Takacs-Fiksel estimation method for Gibbs point processes. *Statistics*, 30(1) :69–89, 1997. 304

[22] J.-M. Billiot and M. Goulard. An estimation method of the pair potential function for Gibbs point processes on spheres. *Scandinavian journal of statistics*, 28(1) :185–203, 2001. 313

[23] J.-C. Breton, I. Nourdin, and G. Peccati. Exact confidence intervals for the Hurst parameter of a fractional Brownian motion. *Electron. J. Statist.*, 3 :416–425, 2009. 26

[24] P. Breuer and P. Major. Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, 13(3) :425–441, 1983. 21

[25] A. Brouste, J. Istas, and S. Lambert-Lacroix. On Fractional Gaussian Random Fields Simulations. *Journal of Statistical Software*, 23(1) :1–23, 2007. 19

[26] G. Chan and A.T.A. Wood. Simulation of stationary Gaussian vector fields. *Statistics and Computing*, 9(4) :265–268, 1999. 15, 18, 31

[27] A. Chronopoulou, C.A. Tudor, and F. Viens. Variations and Hurst index estimation for a Rosenblatt process using longer filters. *Electronic Journal of Statistics*, 3 :1393–1435, 2009. 37

[28] S. Cohen and J. Istas. An universal estimator of local self-similarity. *ESAIM : Probability and Statistics*, 2002. 19

[29] J.J. Collins and C.J. De Luca. Upright, correlated random walks : A statistical-biomechanics approach to the human postural control system. *Chaos*, 5(1) :57–63, 1994. 11

[30] F. Comets and M. Janzura. A central limit theorem for conditionally centred random fields with an application to Markov fields. *Journal of Applied Probability*, 35(3) :608–621, 1998. 310

[31] P. Craigmile. Simulating a class of stationary Gaussian processes using the Davies–Harte algorithm, with application to long memory processes. *Journal of Time Series Analysis*, 24(5) :505–511, 2003. 17, 31

[32] D. Dacunha-Castelle and M. Duflo. *Exercices de probabilités et statistiques. Tome 1 Problèmes à temps fixe*. Collection Mathémathiques Appliquées pour la maîtrise, Masson, Paris, 1980. 23

[33] D. Dacunha-Castelle and M. Duflo. *Probability and Statistics : Volume 2*. Masson, 1994. 21

[34] R. Dahlhaus. Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17(4) :1749–1766, 1989. 21

[35] D. Dereudre. Gibbs Delaunay tessellations with geometric hardcore conditions. *Journal of Statistical Physics*, 131(1) :127–151, 2008. 295, 298

[36] D. Dereudre. The existence of Quermass-interaction processes for nonlocally stable interaction and nonbounded convex grains. *Adv. Appli. Prob.*, 41(3) :664–681, 2009. 300

[37] D. Dereudre. *Etudes de quelques processus ponctuels gibbsiens en mécanique statistique, géométrie aléatoire et statistique spatiale*. PhD thesis, Univ. Valenciennes, 2010. 299, 300

[38] D. Dereudre, R. Drouilhet, and H.O. Georgii. Existence of Gibbsian point processes with geometry-dependent interactions. *submitted*, 2010. `http://arXiv:1003.2875`. 295, 297, 298, 299, 310

[39] D. Dereudre and F. Lavancier. Campbell equilibrium equation and pseudo-likelihood estimation for non-hereditary Gibbs point processes. *Bernoulli*, 15(4) :1368–1396, 2009. 302, 306

[40] D. Dereudre and F. Lavancier. Practical simulation and estimation for Gibbs Delaunay-voronoï tessellations with geometric hardcore interaction. *to appear in Computational Stat. and Data Analysis*, 2010. 298

[41] G. Didier and V. Pipiras. Integral representations of operator fractional Brownian motion. *To appear in Bernoulli*, 2010. 13, 14, 15

[42] T. Dieker. Simulation of fractional Brownian motion. *Master's thesis, Vrije Universiteit, Amsterdam*, 2002. 17

[43] C.R. Dietrich and G.N. Newsam. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing*, 18(4) :1088–1107, 1997. 17, 31

[44] J.L. Doob. *Stochastic processes*. New York, 1953. 21

[45] P. Doukhan, G. Oppenheim, and M.S. Taqqu. *Theory and applications of long-range dependence*. Birkhauser, 2003. 11, 17, 19

[46] M. Emily and O. François. A statistical approach to estimating the strength of cell-cell interactions under the differential adhesion hypothesis. *Theoretical Biology and Medical Modelling*, 4(1) :37, 2007. 298

[47] M.J. Fadili and E.T. Bullmore. Wavelet-generalized least squares : a new BLU estimator of linear regression models with 1/f errors. *NeuroImage*, 15(1) :217–232, 2002. 34

[48] R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher. The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Current Biology*, 17(24) :2095–2104, 2007. 298

[49] T. Fiksel. Estimation of parameterized pair potentials of marked and non-marked Gibbsian point processes. *Elektron. Inform. Kybernet.*, 20 :270–278, 1984. 304

[50] T. Fiksel. Estimation of interaction potentials of Gibbsian point processes. *Math. Operationsf. Statist. Ser. Statist.*, 19 :77–86, 1988. 304

[51] P. Flandrin. On the spectrum of fractional Brownian motions. *IEEE Transactions on Information Theory*, 35(1) :197–199, 1989. 28, 29

[52] P. Flandrin. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, 38(2) :910–917, 1992. 28, 29

[53] T.R. Fleming and D.P. Harrington. *Counting processes and survival analysis*. Wiley, New York, 1991. 301

[54] G. Fäy, E. Moulines, F. Roueff, and M.S. Taqqu. Estimators of long-memory : Fourier versus wavelets. *Journal of Econometrics*, 151(2) :159–177, 2009. 19

[55] C. Gaetan and X. Guyon. *Modélisation et statistique spatiales*. Springer Verlag, 2008. 293

[56] H.O. Georgii. Canonical and grand canonical Gibbs states for continuum systems. *Communications in Mathematical Physics*, 48(1) :31–51, 1976. 295

[57] H.O. Georgii. *Gibbs measures and Phase Transitions*. de Gruyter, Berlin, 1988. 295

[58] J.K. Ghosh. A new proof of the Bahadur representation of quantiles and an application. *The Annals of Mathematical Statistics*, pages 1957–1961, 1971. 24

[59] L.A. Gil-Alana. Time series modeling of sunspot numbers using long-range cyclical dependence. *Solar Physics*, 257 :371–381, 2009. 311, 312

[60] M. Goulard, A. Särkkä, and P. Grabarnik. Parameter estimation for marked Gibbs point processes through the maximum pseudo-likelihood method. *Scandinavian Journal of Statistics*, 23(3) :365–379, 1996. 303

[61] F. Graner and J.A. Glazier. Simulation of cell-sorting using a two-dimensional extended Potts model. *Phys. Rev. Lett.*, 69 :2013–2016, 1992. 298

[62] C.W.J. Granger. The typical spectral shape of an economic variable. *Econometrica*, 34 :150–161, 1966. 11

[63] X. Guyon. *Random fields on a network. Modeling, Statistics and Applications*. Springer Verlag, New York, 1995. 303

[64] L. Heinrich. Mixing properties of Gibbsian point processes and asymptotic normality of Takacs-Fiksel estimates. *preprint*, 1992. 304

[65] H. Honda. Description of cellular patterns by Dirichlet domain : the two-dimensional case. *Journal of Theoretical Biology*, 72 :523–543, 1978. 298

[66] F. Huang and Y. Ogata. Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics*, 8(3) :510–530, 1999. 309

[67] J. Illian, A. Penttinen H., and Stoyan. *Statistical analysis and modelling of spatial point patterns*. Wiley-Interscience, 2008. 293

[68] J. Istas. Quadratic variations of spherical fractional Brownian motions. *Stochastic Processes and their Applications*, 117(4) :476–486, 2007. 19

[69] J. Istas and G. Lang. Quadratic variations and estimation of the Hölder index of a Gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.*, 33 :407–436, 1997. 19

[70] H. Kokko J. Lindström and E. Ranta. There is nothing new under the sunspots. *Oikos*, 7(3) :565–568, 1996. 311

[71] S. Jaffard. Propriétés des matrices "bien localisées" pres de leur diagonale et quelques applications. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 7(5) :461–476, 1990. 22

[72] J.L. Jensen and H.R. Künsch. On asymptotic normality of pseudo likelihood estimates of pairwise interaction processes. *Ann. Inst. Statist. Math.*, 46 :475–486, 1994. 302, 303, 308

[73] J.L. Jensen and J. Møller. Pseudolikelihood for exponential family models of spatial point processes. *Ann. Appl. Probab.*, 1 :445–461, 1991. 302

[74] R. Olivier J.L. Ballester and F. Baudin. Discovery of the near 158 day periodicity in group sunspot numbers during the eighteenth century. *Astrophysical Journal*, 522 :153–156, 1999. 311

[75] T. Kato and E. Masry. On the spectral density of the wavelet transform of fractional Brownian motion. *Journal of Time Series Analysis*, 20(5) :559–563, 1999. 28, 29

[76] W. S. Kendall, M. N. M. Van Lieshout, and A. J. Baddeley. Quermass-interaction processes conditions for stability. *Adv. Appli. Prob.*, 31 :315–342, 1999. 299

[77] J.T. Kent and A.T.A. Wood. Estimating the fractal dimension of a locally self-similar Gaussian process using increments. *J. Roy. Statist. Soc. Ser. B*, 59 :679–700, 1997. 19

[78] F. Lavancier, A. Philippe, and D. Surgailis. Covariance function of vector self-similar processes. *Statistics & Probability Letters*, 79(23) :2415–2421, 2009. 13, 14

[79] N.A. Lazar. *The statistical analysis of functional MRI data*. Springer Verlag, 2008. 33, 35

[80] E.L. Lehman. *Elements of large-sample theorey*. Springer, 1999. 309

[81] J.E. Lennard-Jones. Cohesion. *Proceedings of the Physical Society*, 43 :461–482, 1931. 299

[82] M. Li and P. Vitányi. *An Introduction to Kolmogorov complexity and its applications*. Springer, New York, 1997. 463

[83] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz, and P. Ciuciu. A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. *Neuroimage*, 41(3) :941–969, 2008. 37

[84] B. Mandelbrot and J. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10 :422–437, 1968. 11

[85] K.V. Mardia and P.E. Jupp. *Directional statistics*. Wiley New York, 2000. 312

[86] S. Mase. Uniform lan condition of planar gibbsian point processes and optimality of maximum likelihood estimators of soft-core potential function. *Probab. Theory and Related Fields*, 92 :51–67, 1992. 301, 309

[87] S. Mase. Consistency of maximum pseudo-likelihood estimator of continuous state space Gibbsian process. *Advances in Applied Probability*, 5 :603–612, 1995. 302

[88] S. Mase. Marked Gibbs processes and asymptotic normality of maximum pseudo-likelihhod estimators. *Math. Nachr.*, 209 :151–169, 2000. 302

[89] K Matthes, J. Kerstan, and J. Mecke. *Infinitely Divisible Point Process*. J. Wiley, 1978. 294

[90] E.W. Maunder. The sun and sun-spots, 1820-1920. *Monthly Notices of the Royal Astronomical Society*, 82 :534–543, 1922. 311

[91] V. Maxim, L. Şendur, J. Fadili, J. Suckling R., Gould, R. Howard, and E. Bullmore. Fractional Gaussian noise, functional MRI and Alzheimer's disease. *NeuroImage*, 25(1) :141–158, 2005. 34

[92] A.I. McLeod and K.W. Hipel. Preservation of the rescaled adjusted range : a reassessment of the Hurst exponent. *Water Ressources Research*, 14(3) :491–508, 1978. 11

[93] J. Møller and K. Berthelsen. Transforming spatial point processes into Poisson processes using random superposition. *submitted for publication*, 2010. 310

[94] J. Møller and K. Helisová. Power diagrams and interaction processes for union of discs. *Adv. Appli. Prob.*, 40(2) :321–347, 2008. 300, 304

[95] J. Møller and R. Waagepetersen. *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, 2003. 293, 296, 301, 306, 309, 310

[96] J. Møller and R. Waagepetersen. Modern statistics for spatial point processes. *Scan. J. Stat.*, 34 :643–684, 2007. 293

[97] X.X. Nguyen and H. Zessin. Ergodic theorems for Spatial Process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 48 :133–158, 1979. 303

[98] I. Nourdin and F.G. Viens. Density formula and concentration inequalities with malliavin calculus. *To appear in Elec. J. Probab.*, 2009. 26, 32

[99] R.F. Peltier and J.L. Véhel. Multifractional Brownian motion : definition and preliminary results. *Rapport de recherche-INRIA*, 1995. 12

[100] X. Pennec. Intrinsic statistics on Riemannian manifolds : Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1) :127–154, 2006. 312

[101] C.J. Preston. *Random fields*. Springer Verlag, 1976. 294, 295

[102] F. Richard and H. Biermé. Estimation of anisotropic Gaussian fields through radon transform. *ESAIM Probab. Stat.*, 12(1) :30–50, 2008. 19

[103] D. Ruelle. *Statistical Mechanics*. Benjamin, New York-Amsterdam, 1969. 295, 299, 303

[104] D. Ruelle. Superstable interactions in classical statistical mechanics. *Comm. Math. Phys.*, 1970. 299, 303

[105] S. S. Thurner, C. Windischberger, E. Moser, and M. Barth. Fractal noise maps reveal human brain activity : a key for unbiased fMRI analysis. *Arxiv preprint cond-mat/0208415*, 2002. 34

[106] G. Samorodnitsky and M.S. Taqqu. *Stable non-Gaussian random processes : stochastic models with infinite variance*. Chapman & Hall/CRC, 1994. 11, 12

[107] R. Serfling. *Approximation theorems of mathematical statistics*. New York : Wiley & Sons, 1980. 23, 24

[108] Haipeng Shen, Zhengyuan Zhu, and Thomas C. M. Lee. Robust estimation of the self-similarity parameter in network traffic using wavelet transform. *Signal Process.*, 87(9) :2111–2124, 2007. 24

[109] R. Srikanth, R. Muralishankar, and A.G. Ramakrishnan. Wavelet-Based Estimation of Hemodynamic Response Function. *Lecture notes in computer science*, pages 1285–1291, 2004. 34

[110] S. Stoev and M.S. Taqqu. How rich is the class of multifractional Brownian motions? *Stochastic Processes and their Applications*, 116(2) :200–221, 2006. 12

[111] S. Stoev, M.S. Taqqu, C. Park, G. Michailidis, and J.S. Marron. LASS : a tool for the local analysis of self-similarity. *Computational statistics & data analysis*, 50(9) :2447–2471, 2006. 22

[112] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic geometry and its applications*. John Wiley and Sons, Chichester, 1995. 300

[113] R. Takacs. Estimator for the pair-potential of a Gibbsian point process. *Math. Operationsf. Statist. Ser. Statist.*, 17 :429–433, 1986. 304

[114] A.H. Tewfik and M. Kim. Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE transactions on information theory*, 38(2) :904–909, 1992. 28

[115] C.A. Tudor. Analysis of the Rosenblatt process. *ESAIM : Probability and Statistics*, 12 :230–257, 2008. 37

[116] D.J. Watts and S.H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393(6684) :440–442, 1998. 35, 36

[117] B. Whitcher, P. Guttorp, and D.B. Percival. Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research*, 105(D11) :941–962, 2000. 35

[118] W. Willinger, M.S. Taqqu, W.E. Leland, and D.V. Wilson. Self-similarity in high speed packet traffic : analysis and modelisation of ethernet traffic measurements. *Statistical Science*, 10 :67–85, 1995. 11

[119] A.T.A. Wood and G. Chan. Simulation of stationary Gaussian processes in [0, 1] d. *Journal of computational and graphical statistics*, 3(4) :409–432, 1994. 17

[120] G.W. Wornell. A Karhunen-Loeve-like expansion for 1/f processes via wavelets. *IEEE Transactions on Information Theory*, 36(4) :859–861, 1990. 28

[121] W.B. Wu. On the Bahadur representation of sample quantiles for dependent sequences. *Annals of statistics*, 33(4) :1934–1963, 2005. 24

[122] E. Zarahn, G.K. Aguirre, and M. d'Esposito. Empirical analyses of BOLD fMRI statistics. *NeuroImage*, 5(3) :179–197, 1997. 34

Résumé

Ce mémoire présente une synthèse de mes activités de recherche depuis mon doctorat. Ces travaux sont organisés en trois parties distinctes. Les deux premières parties ont pour point commun l'inférence statistique de quelques processus stochastiques. Les processus centraux en question sont respectivement le mouvement Brownien fractionnaire (et quelques unes de ses extensions) et les processus ponctuels spatiaux de Gibbs. Comme, nous le verrons par la suite, bien que ces processus soient de nature très différente, ils s'inscrivent dans la modélisation de données dépen- dantes qu'elles soient temporelles ou spatiales. Nos travaux ont pour objectifs communs d'établir des propriétés asymptotiques de méthodes d'estimation ou de méthodes de validation, classiques ou originales. Par ailleurs, une autre similitude est la mise en perspective de ces processus avec des applications faisant intervenir des systèmes complexes (modélisation de signaux issus d'Imagerie par Résonance Magnétique Fonctionnelle et modélisation de taches solaires). La troisième partie, quant à elle, regroupe des thèmes satellites regroupés sous la dénomination contributions à la statistique appliquée.

Mots clés : modélisation, inférence statistique, données dépendantes, processus fractionnaires, processus ponctuels de Gibbs, systèmes complexes.

Abstract

This document is a synthesis of my research activity since my PhD. The contributions are organized in three different parts. The first and second parts are dedicated to the statistical inference of stochastic processes. The main processes we study are the fractional Brownian motion (and some of its extensions) and spatial Gibbs point processes. These processes are of different nature but both are used to model data with strong (temporal or spatial) dependence. My contributions have for common points the study of asymptotic properties of estimation and validation methods. Another common point is the perspective to use these processes in applications dealing with complex systems in particular the modelling of signals derived from Functional Magnetic Resonance Imaging and the modelling of sunspots. Finally, the third part of this document gathers some works under the name contributions to applied statistics.

Keywords : modelling, statistical inference, dependent data, fractional processes, spatial Gibbs point processes, complex systems