Contributions à la statistique des processus et à l'estimation fonctionnelle

Mustapha Rachdi

UNIVERSITÉ PIERRE MENDÈS FRANCE

MÉMOIRE DE PRÉSENTATION DE TRAVAUX EN VUE DE L'OBTENTION DU DIPLÔME D'HABILITATION À DIRIGER DES RECHERCHES

CONTRIBUTIONS À LA STATISTIQUE DES PROCESSUS ET À L'ESTIMATION FONCTIONNELLE

Présenté par

Mustapha RACHDI

le 07 novembre 2006

Commission d'Examen

Paul Deheuvels	Univ. Pierre et Marie Curie, Paris	Examinateur/Président
Michel Delecroix	ENSAI, Rennes	Rapporteur
Peter Hall	Univ. Canberra, Australie	Rapporteur
Dimitris Politis	Univ. de Californie, San Diego, USA	Rapporteur
Philippe Vieu	Univ. Paul Sabatier, Toulouse	Rapporteur
Karim Benhenni	Univ. P. Mendès France, Grenoble	Examinateur
Alain Berlinet	Univ. Montpellier II	Examinateur (excusé)
Jacques Demongeot	Univ. Joseph Fourier, Grenoble	Examinateur
Ali Gannoun	CNAM, Paris	Examinateur
Michel Lejeune	Univ. P. Mendès France, Grenoble	Examinateur

Table des matières

Remerciements

Je tiens à remercier Monsieur le Professeur Paul Deheuvels de l'honneur qu'il m'a fait en acceptant de participer à ce jury et d'en assurer la présidence et ainsi de me faire bénéficier de son expertise internationalement reconnue.

Je voudrais adresser des remerciements particuliers et une reconnaissance totale à Monsieur le Professeur Philippe Vieu pour avoir accepté d'être rapporteur de cette HDR en toute impartialité. Ses encouragements permanents, ses qualités humaines, ses compétences de chercheur de renom ainsi que ses conseils toujours avisés ont eu un rôle éminent tout au long de l'édification de ce travail et durant ma carrière de chercheur.

Je remercie également Monsieur le Professeur Peter Hall et Monsieur le Professeur Dimitris Politis d'avoir bien voulu rapporter sur mes travaux, en y prêtant une attention toute particulière et intéressée. Ils me font également bénéficier de leurs expertises internationalement reconnues.

Je tiens absolument à remercier Monsieur le Professeur Michel Delecroix pour l'intérêt qu'il portait à mes recherches et pour avoir accepté d'être rapporteur sur mes travaux, malgré ses nombreuses occupations.

Professeur Alain Berlinet et Professeur Ali Gannoun ont accepté très gentillement de faire partie du jury. Je les remercie très vivement. Ils m'honorent particulièrement par leur autorités en la matière, leurs renoms, sans oublier leurs extrêmes cordialités.

Je tiens à adresser mes remerciements à mes collègues Docteur Karim Benhenni et Professeur

Michel Lejeune de l'Université Pierre Mendès France et à Monsieur le Professeur Jacques Demongeot de l'Université Joseph Fourier, d'avoir accepté de participer au jury de mon HDR.

Une pensée particulière est adressée à tous mes co-auteurs avec qui j'ai eu des échanges fructueux et également à ceux avec qui nos efforts n'ont pas abouti à des résultats publiables.

Pour finir j'adresse mes remerciements à tous mes collègues des Universités Pierre Mendès France et Joseph Fourier, ainsi qu'à tout le corps administratif.

Introduction générale

Dans ce document de synthèse, notre objectif premier est de présenter nos travaux sur la statistique non paramétrique des processus stochastiques et sur l'estimation fonctionnelle. Plutôt que de vouloir insister sur les détails mathématiques de nos résultats, que l'on pourra toujours retrouver dans les articles correspondants, nous avons choisi de les présenter d'une façon synthétique. Sans prétendre à l'exhaustivité, nous nous sommes attachés à indiquer les articles historiques et à faire un choix de certains articles nous paraîssant les plus intéressants.
Les techniques non paramétriques ont pris une importance de plus en plus grande depuis une trentaine d'années dans la recherche en statistique mathématique. Le nombre toujours croissant d'articles sur ce thème en témoigne. Il faut également signaler que le développement des moyens informatiques et la puissance actuelle de calcul des ordinateurs permettent d'élargir toujours plus le champs d'application de ces méthodes.
Ce document est organisé en respectant des thématiques. En fait, nous avons classifié l'ensemble de nos travaux en six chapitres. Dans chacun de ces chapitres, nous indiquons les travaux concernés avant un bref historique, ensuite nous résumons les principaux résultats, les idées sous-jacentes, et ce qui a motivé ce travail. Nous scindons nos recherches en deux grandes parties : d'abord, l'estimation fonctionnelle et la statistique des processus en dimension finie (chapitres 1, 2, 3 et 4), et puis, l'analyse statistique des données fonctionnelles (chapitre 5). Le dernier chapitre de ce mémoire est le fruit de nos investigations avec l'équipe de Telecom Lille 1 sur la modélisation statistique du canal de transmission à $60\,GHz$ dans les milieux confinés.

L'estimation fonctionnelle en dimension finie est à l'origine de plusieurs investigations, tant sur le plan théorique que sur le plan pratique. Avant les années soixante l'estimation fonctionnelle a été peu étudiée, mis à part quelques articles sur l'estimation par histogramme. Mais depuis les travaux précurseurs de Rosenblatt et Parzen (Cf [180], [215] et [216]) en densité de probabilité et ceux de Watson et Nadaraya (Cf [171] et [242]) en régression, ainsi que ceux de Parzen (Cf [177], [181] et [182]), Brillinger (Cf [31]), Priestley (Cf [192]), en estimation de la densité spectrale, et bien d'autres, ce domaine connaît un grand essort. Depuis, les méthodes non paramétriques se sont donc considérablement développées. Hormis le fait que ces travaux concernaient les propriétés de convergence d'estimateurs non paramétriques, ils ont été généralisés dans de nombreuses directions : observations vectorielles (i.e. à valeurs dans $\mathbb{R}^p$), observations dépendantes (processus mélangeants). D'autres types d'estimateurs

ont aussi été utilisés (δ-suites, estimateurs par ondelettes, ...) voir entre autres [62], [66], [108], [130] et [131]. Les méthodes de pénalisation ont aussi été considérées (Cf entre autres [4], [18] et [186]).
Une présentation des résultats obtenus sur l'estimation fonctionnelle en dimension finie se trouve par exemple dans [21], [24], [64] et [191] pour le cas de l'estimateur à noyau. Par ailleurs, l'analyse spectrale des processus stochastiques représente un outil très important en physique (en particulier, en théorie du signal et de l'image). En effet, la densité spectrale d'un processus représentant la concentration de l'énergie par rapport aux fréquences permet la détection par exemple des phénomènes anormaux. Il est donc très important d'estimer celle-ci. Notons que, depuis les travaux précurseurs de Parzen (Cf [179] et [182]), l'estimation à noyau de la densité spectrale (lissage du périodogramme) a été étudiée sur de nombreuses facettes (Cf [28], [29], [30], [31], [140], [141], [150], [151], [152], [187], [188] et [192], entre autres). Notons aussi que ces estimateurs présentent des avantages certains tant sur le plan théorique, puisqu'ils peuvent atteindre les vitesses optimales de convergence (Cf [3]), que sur le plan pratique puisque leur structure simple permet une programmation aisée.

Dans ce qui suit, nous allons développer le contenu de notre mémoire chapitre par chapitre.

Chapitre 1 : L'analyse de Fourier des données et l'estimation de la densité spectrale pour les processus à temps continu ont prouvé leur utilité en communication (dans le filtrage linéaire et la théorie de prédiction), en sismologie (pour déterminer la nature d'un événement sismique), en océanographie (pour l'étude des ondes océaniques) et dans divers domaines des sciences physiques, médicales .... Dans la statistique des processus à temps continu, les données sont souvent collectées en utilisant un schéma d'échantillonnage. Ceci est motivé par les raisons suivantes : dans la théorie classique de l'estimation spectrale, le périodogramme est calculé à partir des observations de X sur $[0,T]$ où $T>0$, par : $I_T(\lambda) = \int_0^T X(t)\ \exp(-i\,\lambda\,t)\,dt|^2/(2\pi T)$, $\forall \lambda \in \mathbb{R}$. Cet estimateur n'est pas convergeant, mais en le lissant, on construit un estimateur de type noyau $\widehat{\phi}_T$ qui est asymptotiquement consistant sous certaines conditions (Cf [40], [140], [157], [162] et [192]). Dans les applications pratiques, les observations de X ne sont pas obtenues sous une forme analytique, alors l'intégrale $\int_0^T X(t)\ \exp(-i\omega t)\ dt$ ne peut pas être calculée numériquement. Ceci constitue un problème majeur pour calculer le périodogramme. A cette fin, X est observé à des instants $\{t_n\}_{n\in\mathbb{Z}}$ et les observations sont $\{X(t_n), n=1,\ldots,N\}$, avec $N\in\mathbb{N}$. Ces instants d'échantillonnage $\{t_n\}_{n\in\mathbb{Z}}$ sont à déterminer. Dans un premier temps, nous avons choisi de procéder par un échantillonnage périodique du temps. Ce choix introduit le phénomène de repliement ou de recouvrement des ondes (aliasing). Dans ce cas, la densité spectrale $\phi_{\underset{\sim}{X}}$ se calcule à partir de la densité spectrale $\phi_{\tilde{X}}$ correspondante au processus échantillonné $\underset{\sim}{X} = \{X(n/\epsilon)\}_{n\in\mathbb{Z}}$ où $\epsilon>0$ (ici $t_n = n/\epsilon$), si le processus stochastique X est bande-limité c.-à-d. ϕ_X est à support dans $[-\pi\epsilon,\pi\epsilon]$ (Cf [157]). Nous avons construit ensuite l'estimateur de ϕ_X, et établi sa convergence presque complète (p.co.) et sa vitesse de convergence (Cf [197]). La condition sur le support de ϕ_X semble être restrictive. C'est pour cela que dans un deuxième temps, nous avons adopté l'échantillonnage aléatoire qui permet de pallier le phénomène d'aliasing (Cf [30] et [222]). La première difficulté que l'on rencontre est le transfert des conditions de

mélange fort du processus initial X vers le processus échantillonné $\{X_{t_n}\}$. Pour ceci, nous avons établi des résultats sur la mélangeance du processus échantillonné pour divers types de mélangeance de X (Cf [**?**] et [45]). Ensuite, nous avons construit un estimateur de la densité spectrale, puis nous avons établi sa convergence uniforme p.co. et donné sa vitesse de convergence (Cf [198]).
Dans ces précédents travaux, nous nous sommes bien sûr intéressés à l'espace temps. Il s'avère que celui-ci est plus complexe qu'il n'y paraisse, et c'est ce qui a motivé notre intérêt pour les nombres p-adiques. En effet, les applications de l'analyse p-adique à la physique pourraient même aller au-delà des aspects strictement techniques. Des physiciens théoriciens se livrent par exemple à des spéculations sur la structure de l'espace et du temps à très petite échelle. Les lois de la relativité et de la physique quantique semblent indiquer qu'il n'est pas possible de mesurer des longueurs inférieures à une valeur extra-ordinairement petite, appelée longueur de Planck (de l'ordre de 10^{-35} mètre). L'existence d'une distance minimale suggère à certains théoriciens qu'à cette échelle, la structure ultime de l'espace-temps pourrait se décrire non pas en termes de nombres réels, mais en termes de structure p-adique. Pour l'instant, ce ne sont que des études spéculatives, mais il n'est pas exclu qu'elles aboutissent un jour à des conclusions vérifiables par des expériences (Cf [5] et [6]). Pour ces raisons, l'utilisation de l'analyse spectrale pour des processus dont le temps évolue dans le corps des nombres p-adiques a suscité, ces dernières années, beaucoup d'intérêts chez plusieurs mathématiciens. Pour ce type de processus, nous avons construit les premiers estimateurs de la densité spectrale p-adique, et nous avons établi leurs convergences (en moyenne quadratique, presque sûre et leur normalité asymptotique), d'abord pour un échantillonnage déterministe puis pour un échantillonnage aléatoire (Cf [199] et [200]). Comme dans le cas réel, on suppose que la mesure spectrale soit la résultante de deux mesures : l'une est absolument continue par rapport à la mesure de Haar (mesure de Haar sur Q_p) et l'autre est discrète. Pour estimer, dans ce cas, la densité de la partie continue, nous avons adapté la méthode de double fenêtres. Nous avons estimé la densité de la partie continue, ainsi que la densité aux points où la mesure spectrale présente des atomes (Cf [195]).
Cette étude est incomplète sans l'extrapolation aux champs aléatoires. Nous avons donc généralisé, dans [204], la méthode de double fenêtres aux processus multidimensionnels stationnaires (mais à temps réel) dont la mesure spectrale est mixte c.-à-d. qui s'écrit comme la somme d'une mesure absolument continue, d'une mesure discrète d'ordre fini et de mesures continues sur diverses droites.

Chapitre 2 : Dans le cadre de l'estimation fonctionnelle par la méthode du noyau, on a affaire à un paramètre dit de lissage ou largeur de fenêtre. Ce paramètre joue un rôle crucial dans le comportement asymptotique de l'estimateur. En particulier la vitesse de convergence de l'estimateur en dépend. Des méthodes de choix automatique et optimal de ce paramètre de lissage sont mises en place : validation-croisée, plug-in (méthode de type injection), bootstrap (méthode de type ré-échantillonnage), ... On essaie donc de sélectionner un paramètre de lissage optimal, c.-à-d. qui rend la mesure d'erreur de l'estimateur à noyau minimale :

$$\frac{d(h_0)}{\inf(d(h))} \to 1$$

selon divers modes de convergence, où $d(h)$ est une mesure d'erreur de l'estimateur à noyau construit avec un paramètre de lissage h, et où h_0 est une valeur de ce paramètre directement calculable à partir des données.
Dans ce chapitre, nous nous sommes donc intéressés aux aspects théorique et pratique de ce problème. Les résultats qui y sont présentés concernent essentiellement les estimateurs à noyau de la densité spectrale obtenus par un lissage du périodogramme. Ces résultats sont donnés pour divers types d'erreurs quadratiques et divers modes de convergence. D'abord, en construisant un critère de choix du paramètre de lissage pour un processus à temps continu (avec ou sans hypothèse de mélangeance) (Cf [196] et [207]). Ensuite, nous avons étendu les résultats obtenus au cas des processus à temps discret (toujours avec ou sans hypothèse de mélangeance) (Cf chapitre 3 de [193] et [194]). Plus précisément, nous avons proposé, dans chaque cas, un critère de sélection de la largeur de fenêtre basé sur la validation croisée, puis nous avons établi des résultats d'optimalité asymptotique de la fenêtre sélectionnée.
De plus, pour que notre contribution dans ce domaine puisse bénéficier d'un champ d'applications plus vaste, il était naturel de penser à des prolongements possibles de ces résultas au cadre multidimensionnel. Dans [201], nous avons construit un critère de choix de la largeur de fenêtre spectrale pour les champs aléatoires, puis, nous avons montré son optimalité asymptotique.

Chapitre 3 : Dans la réalité, on est confronté, le plus souvent, au traitement de problèmes pour lesquels les observations sont entachées d'erreurs. L'origine des erreurs peut provenir, entre autres, de l'appareil de mesure, de la lecture des observations ou de l'échelle utilisée. Dans les problèmes statistiques, il est important de comprendre comment l'estimation peut être affectée quand on dispose seulement de données entachées d'erreurs au lieu des données brutes (non accessibles). Ceci a stimulé notre intérêt pour ce sujet. Nous nous sommes donc intéressés à cette problématique. En fait, notre contribution se situe plus particulièrement dans l'estimation fonctionnelle quand les données sont, soit entachées d'erreurs additives (c.-à-d. un modèle de déconvolution), soit quantifiées. Le modèle de déconvolution (c.-à-d. on observe : $Z = X + erreur$) existe dans plusieurs domaines scientifiques qui utilisent des mesures qui sont accompagnées par des erreurs, et a été suffisamment étudié. Pour une revue bibliographique, nous renvoyons à [161] pour les applications, à [153] et [154] pour l'estimation de la densité de probabilité de X dans les problèmes de déconvolution, à [17] et [233] pour l'estimation du mode et du mode conditionnel quand l'observation ne présente pas d'erreurs et à [126] quand les données sont entachées d'erreurs.
Le modèle de déconvolution peut être rencontré en microfluométrie, en électrophorèse, en biostatistique, et en d'autres domaines dans lesquels les mesures ne peuvent pas être observées directement. Par exemple, dans l'étude de la maladie de SIDA, Z peut être le temps à partir d'un certain instant jusqu'au moment de l'infection, et X est la période d'incubation (le temps entre l'occurence de l'infection jusqu'au moment de l'apparition des symptômes). D'autres situations pratiques utilisant des modèles de déconvolution peuvent être trouvées dans [160].
Dans [202], nous avons proposé deux estimateurs du mode de X et nous avons étudié leurs convergences en moyenne quadratique, ainsi que leurs normalités asymptotiques. Le premier

estimateur n'est autre que le mode empirique, et le deuxième repose sur la même idée que dans [233] c.-à-d. que sa construction utilise l'estimateur à noyau de la dérivée première de la densité de probabilité plutôt que d'utiliser un estimateur de la densité de probabilité elle-même. Ensuite, sous des conditions de mélange fort, nous avons montré la convergence presque sûre du mode empirique multivarié pour un processus stationnaire entaché d'erreurs (Cf [203]).
Dans le même état d'esprit, nous nous sommes intéressés aux données quantifiées. En fait, la forme la plus connue de la quantification est l'erreur d'arrondi, qui se produit dans tous les systèmes numériques (Cf [100]). Un quantificateur général approxime une valeur observée par la plus proche valeur parmi un nombre fini de valeurs représentatives. En fait, le problème de quantification se pose le plus souvent dans les domaines de communication, dans la théorie de l'information et du signal (Cf [7], [39], [100], [144] et [158], entre autres). Nous considérons donc, le problème d'estimation de la courbe de croissance pour des données quantifiées et corrélées. Afin que l'estimateur construit soit consistant, nous supposons disposer d'observations répetées. Nous avons établi le comportement asymptotique de l'estimateur construit à partir des données quantifiées et fait des comparaisons avec le cas de données non quantifiées. Nous avons fourni également l'expression de la largeur de fenêtre optimale, et nous avons comparé à travers un exemple les performances des deux estimateurs construits à partir d'observations bruitées ou non (Cf [12] et [14]).

Chapitre 4 : Remarquons que les processus stationnaires de second ordre réels comme complexes ont joué un rôle important dans la modélisation de plusieurs phénomènes. Cependant, dans la nature, tous les phénomènes ne possèdent pas nécessairement cette propriété, d'où leurs intérêts pour certains probabilistes et statisticiens. Il est devenu donc légitime d'essayer de s'affranchir de l'hypothèse de stationnarité. Pour ceci, notre intérêt s'est dirigé, d'abord, vers l'étude d'une classe de processus non stationnaires dont le temps suit une échelle logarithmique. Bien que ces processus sont non stationnaires au sens classique : par rapport à la loi additive ($I\!E(X(t+s)X(s)) = \rho(t)$), ils sont tout de même stationnaires par rapport à la loi multiplicative ($I\!E(X(t \times s)X(s)) = \rho(t)$). De plus, ces processus sont rencontrés dans la pratique en tant que processus à temps continu. Dans le but de répondre à plusieurs questions posées dans la pratique (quand on utilise ce type de processus), nous avons développé dans [94], la méthode d'échantillonnage aléatoire pour estimer la densité spectrale (transformée de Mellin des covariances) correspondante, donné sa vitesse de convergence et vérifié la pertinence de l'estimateur par des simulations. Ensuite, nous avons généralisé ces résultats aux processus φ-réductibles. Ce sont des processus qui ont subi une déformation dans le temps par l'exercice de la bijection φ (Cf [95] : ces résultats ne font pas partie de ce mémoire).
Dans la lignée de ce sujet et de ce chapitre, nous avons généralisé, dans [96], cette étude aux processus de second ordre indexés par un semi-groupe régulier. Ces processus ont une fonction de covariance définie positive, que nous avons prolongé à une fonction de covariance définie positive sur un groupe incluant le semi-groupe en question. Nous avons prouvé donc que si le noyau est invariant par translation, alors, notre processus admet une représentation spectrale.

Remarquons que dans le chapitre 3 (Cf [14]), nous avons considéré l'estimation de la fonction de croissance quand les données sont quantifiées et que le processus d'erreur est stationnaire au sens large. Partant de [85], nous avons essayé de s'affranchir de l'hypothèse de stationnarité, ce qui a motivé l'étude du problème d'estimation de la courbe de croissance moyenne quand le processus d'erreur est non stationnaire. La non stationnarité mentionnée ici est générale c.-à-d. nous n'imposons aucune hypothèse sur la forme de la fonction d'autocorrélation du processus d'erreur. Ceci nous amène à affirmer que ce travail généralise tous les travaux considérant des processus d'erreurs non stationnaires mais admettant des formes particulières de la fonction de covariance (Cf [13]).

Chapitre 5 : Les données fonctionnelles ont fait leur apparition dans plusieurs domaines de la statistique appliquée (médecine, environnement, chimie, économie , ...). Il est donc de plus en plus fréquent de travailler avec ce type de données. D'un point de vue technique, un échantillon de données fonctionnelles peut être rencontré dans beaucoup de problèmes statistiques (classification, discrimination, études longitudinales, prévision, régression, ...). Ainsi, un vrai défi se pose aux statisticiens pour construire des méthodes statistiques afin de traiter un tel type de données. Ce champ de la statistique moderne est devenu populaire grâce au livre [209] et est généralement connu sous le nom de *Analyse Statistique des Données Fonctionnelles*. Le lecteur peut trouver dans [78] une vue d'ensemble sur des problématiques et des avancées récentes liées à ce domaine important de la statistique moderne.
Dans cette partie, nous nous intéressons à un problème spécifique faisant intervenir des données fonctionnelles : la prévision d'une variable scalaire Y étant donnée une certaine variable fonctionnelle X. En d'autres termes, la question est d'estimer : $r(.) = I\!E(Y|X = .)$, quand X est une variable aléatoire à valeurs dans un espace de dimension infinie. L'objet à estimer est l'opérateur r. L'estimation de l'opérateur r a été traitée par plusieurs auteurs durant la dernière décennie, et la plupart de ces travaux concernait les modèles linéaires pour l'opérateur r. Le domaine de la statistique dit *Modèle Linéaire Fonctionnel* est devenu populaire grâce aux ouvrages [209] et [210] et le lecteur trouvera les plus récentes avancées pour ce type de modèle dans [42]. Comme dans le cas non fonctionnel, c'est-à-dire dans le cas où X est une variable réelle multivariée (mais de dimension finie), il y a un véritable défi pour la relaxation de l'hypothèse de linéarité. De plus, ce problème est souvent plus important dans le cadre fonctionnel car il est souvent impossible d'avoir à disposition des outils graphiques pour vérifier l'exactitude d'un tel genre d'hypothèse. Dans [81] quelques travaux ont été lancés dans cette direction, en construisant des modèles non paramétriques pour l'opérateur r, et en n'exigeant qu'une hypothèse sur la régularité de l'opérateur r. Par analogie avec la terminologie utilisée dans le cas non fonctionnel, ce champ de la statistique est appelé actuellement *Estimation Opératorielle* ou *Estimation non paramétrique pour données fonctionnelles*. Les résultats obtenus dans [82] montrent qu'une version fonctionnelle de l'estimateur à noyau peut être utilisée pour construire un estimateur non paramétrique pour l'opérateur r. En plus, que ces estimateurs opératoriels admettent de bonnes propriétés asymptotiques, le fléau de dimension est contrôlé par le biais de considérations convenables sur la concentration de la loi de la variable fonctionnelle X dans des petites boules. Le lecteur trouvera dans la monographie [83], une discussion sur les méthodes non paramétriques

relatives aux données fonctionnelles. Mais, comme c'est le cas avec plusieurs estimateurs non paramétriques, que ce soit dans le cas non fonctionnel ou dans le cas fonctionnel considéré ici, il existe un paramètre de lissage qui intervient dans la construction des estimateurs et qui doit être sélectionné convenablement afin d'assurer de bonnes performances dans la pratique (Cf [81], pour une application sur des données réelles en chimie). Notre approche pour le choix du paramètre de lissage sera basée sur la procédure de validation croisée, qui est connue dans la littérature, et qui donne des réponses très satisfaisantes pour ce type de problématique, (Cf [116] pour le cas non fonctionnel et [25], [47], [120] et [211] pour le traitement des données longitudinales).
Le but principal du paragraphe 5.1.1 de ce chapitre est de présenter une procédure de choix global (GCV) et automatique du paramètre de lissage, et d'établir ensuite son optimalité asymptotique dans un sens quadratique (Cf [205] et [206]). Pour des raisons techniques, nous avons établi des extensions fonctionnelles de résultats obtenus dans le cas de dimension finie (Cf [112], [133], [147] et [231]). De plus, nous pensons que les résultats que nous avons établi dans le paragraphe 5.1.2, ne sont pas seulement important pour la résolution du problème de choix du paramètre de lissage, mais ils seront aussi, utiles pour d'autres développements futurs.
Par ailleurs, il était naturel de penser à une version locale de la procédure de validation croisée fonctionnelle. Nous avons donc introduit et étudié dans le paragraphe 5.1.3, une version locale (LCV) de la méthode de validation croisée (Cf [8] et [9]). Par suite, pour un souci d'utilité en pratique de ces procédures, nous avons étudié les résultats théoriques obtenus, d'abord sur des données simulées (Cf paragraphe 5.1.4), et ensuite, sur des données réelles (Cf paragraphe 5.1.5). Aussi, nous avons comparé le LCV et le GCV en fonction du choix de la semi-norme d'une part, et du rapport signal/bruit, d'autre part.
Une fois réglés ces problèmes de choix de la largeur de fenêtre, nous nous sommes intéressés au problème de dépendance entre les données. Plus spécialement, au processus d'erreur du modèle de régression fonctionnel. En effet, dans le paragraphe 5.2, nous avons supposé que le processus d'erreur est à longue mémoire. Pour une discussion plus détaillée sur les séries temporelles à longue mémoire et leurs applications, on peut se référer par exemple à [123], à [164] pour les modèles économétriques et à [227] pour les études environnementales ou climatiques.
Dans le paragraphe 5.2, nous avons exposé nos résultats sur la convergence forte (ponctuelle et uniforme) de l'estimateur à noyau de l'opérateur r quand le processus d'erreur est à longue mémoire (Cf [11]). Le fait que cette étude soit spécifique au processus d'erreur à longue mémoire, nous a incité à considérer, dans [10], le problème d'estimation de l'opérateur de régression r quand les erreurs sont corrélées (processus d'erreur n'est pas nécessairement à longue mémoire). Dans le paragraphe 5.3, nous nous sommes intéressés à l'estimation de r quand la variable explicative est fonctionnelle-déterministe (fixed-design). Nous y'avons étudié la performance de l'estimateur à noyau de r en termes de la convergence en moyenne quadratique et de la convergence uniforme presque complète. En particulier, nous avons considéré dans ce contexte, également, le cas de processus d'erreur à longue mémoire. Finalement, comme ce sera le cas tout au long de ce mémoire, une étude par simulations est conduite afin de confirmer l'aspect applicatif de nos résultats théoriques pour plusieurs types

de variables explicatives fonctionnelles et plusieurs types de processus d'erreurs.

Chapitre 6 : Nous avons choisi d'introduire, dans cette synthèse, des réalisations en statistique appliquée qui sont le fruit d'une de nos collaborations avec des non matheux sur la modélisation de situations concrètes. En effet, nous avons collaboré avec des chercheurs de l'ENIC[1] sur la modélisation statistique et électronique du canal de transmission dans les milieux confinés (salle de cours, passages souterrains, tunnels, ...). L'option choisie est de trouver des fréquences disponibles dans la gamme millimétrique. Les bandes passantes disponibles sont en effet plus importantes et permettent d'atteindre des débits plus élevés. Dans ce cadre, la bande autour de 60 Ghz a été identifiée comme une bande possible pour les transmissions à très haut débit et, qui plus est, sans licence. Un autre avantage de cette bande est la forte atténuation des ondes et, en particulier, l'absorption due à l'oxygène. Si cette particularité peut augmenter la complexité des systèmes à mettre en oeuvre pour assurer des qualités de service satisfaisantes, elle permet un très bon confinement des ondes, simplifiant de ce fait la planification en fréquence et réduisant nettement la pollution électromagnétique.
Pour ceci, après des tests statistiques sur la corrélation ou non de plusieurs composantes aléatoires des signaux (amplitudes, phases, retards, réponses impulsionnelles, dépendance des trajets, ...), nous avons proposé des modéles statistiques du canal de transmission. Afin de se détacher de toutes les hypothèses et spéculations antérieures, qui nous paraissent non fondées (Cf [219] et [227]), nous avons opté dans [49] pour une modélisation non paramétrique. En effet, les tests statistiques préliminaires nous ont convaincu de la distribution uniforme des phases sur $(0, 2\pi)$ et de la régression des amplitudes sur les retards. Par contre, la modélisation des retards par un processus de Poisson nous a paru médiocre et fait l'objet d'une étude en cours. Ceci nous a incité à mieux comprendre le phénomène et à trouver un modèle plus convenable au canal. En effet, en considérant les mécanismes responsables des réflexions dans un milieu confiné à 60 GHz et la méthodologie de sondage du canal, nous avons présenté un modèle multitrajets de la composante aléatoire basée sur le concept de groupe d'ondes (Cf [145]). Notre approche consiste à modéliser un faisceau par une succession de réflexions sur des obstacles dont les surfaces sont approchées par un ensemble de facettes planes. Ce modèle permet de tenir compte des différents modes de diffractions telles que les réflexions diffuses ou spéculaires. De plus, en considérant que les caractéristiques de propagation varient de manière significative en environnement indoor, nous avons développé, dans [48], un modèle englobant les évanouissements à petite et à moyenne échelles ainsi qu'un modèle de canal numérique appelé modèle K en référence aux lois suivies par ses coefficients (Cf [228] and [219]).

Pour rendre la lecture de ce mémoire plus facile, celui-ci est composé principalement de six chapitres (décrits ci-dessus), puis d'un paragraphe sur les travaux en cours et les prolongements possibles et immédiats de nos travaux et enfin de la bibliographie générale. Dans le document qui accompagne ce mémoire, nous avons complilé la majorité de nos travaux de recherche.

[1] Ecole Nationale des Ingénieurs en Communication (Telecom Lille 1)

Chapitre 1

Estimation de la densité spectrale

L'analyse de Fourier des données et l'estimation de la densité spectrale pour les processus à temps continu ont prouvé leur utilité en communication (dans le filtrage linéaire et la théorie de prédiction, [54]), en sismologie (pour déterminer la nature d'un événement sismique, [43]), en océanographie (pour l'étude des ondes océaniques, [165]), et dans divers domaines des sciences physiques, médicales, ...

Dans ce chapitre, nous nous intéressons au problème d'estimation de la densité spectrale. Il se divise en trois paragraphes. Le premier paragraphe traite de l'estimation forte de la densité spectrale pour les processus à temps continu (dans $\mathbb{R}$) : on effectue deux types d'échantillonnage (périodique et aléatoire). Dans le second paragraphe, nous mettons en place les premiers résultats de l'analyse spectrale des processus à temps p-adique (dans Q_p) stationnaires. En effet, dans ce paragraphe, nous définissons un certain nombre d'outils permettant de faire de l'estimation spectrale quand le temps évolue dans un corps abélien localement compact muni de la mesure de Haar : d'abord à partir d'un échantillonnage certain, ensuite à partir d'un échantillonnage aléatoire du temps, et enfin nous estimons la densité de la partie continue quand la mesure spectrale p-adique se présente comme la somme d'une partie continue et d'une partie atomique. Cette étude ne sera pas complète, si l'on ne s'intéresse pas à l'estimation spectrale pour les champs aléatoires stationnaires. A cette dernière problématique est dévoué le troisième paragraphe de ce chapitre c.-à-d. à l'analyse spectrale d'un champ aléatoire stationnaire de second ordre dont la mesure spectrale est mixte c.-à-d. qui s'écrit comme la somme d'une mesure absolument continue, d'une mesure discrète d'ordre fini et de mesures continues sur diverses droites.

1.1 Estimation forte de la densité spectrale d'un processus à temps continu

Soit $X = \{X(t)\}_{t\in\mathbb{R}}$ un processus réel, de second ordre, centré, strictement stationnaire de fonction d'autocovariance continue $c_X^{(2)} \in L^1(\mathbb{R})$ et de densité spectrale :

$$\phi_X(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} c_X^{(2)}(t)\exp(-i\,t\lambda)\,dt.$$

1.1.1 Echantillonnage périodique du temps

Comme c'est déjâ mentionné plus haut, dans la pratique, il n'est pas possible d'observer X sur un intervalle continu de temps. Il est donc logique de construire un estimateur à partir d'un échantillon pris à des instants discrets. L'échantillonnage le plus naïf est l'échantillonnage périodique $t_n = n/\epsilon$, $n \in \mathbb{Z}$ où $\epsilon > 0$. Le processus échantillonné $\widetilde{X} = \{X(n/\epsilon)\}_{n\in\mathbb{Z}}$ est stationnaire, centré, de covariances : $c_{\tilde{X}}^{(2)}(n) = c_X^{(2)}(n/\epsilon)$ pour $n \in \mathbb{Z}$. Le problème qui se pose quand on effectue ce type d'échantillonnage est que l'on estime, non pas la densité spectrale ϕ_X (du processus X), mais celle associée à $\widetilde{X}$: $\phi_{\tilde{X}}(\lambda) = \epsilon\sum_{k=-\infty}^{+\infty}\phi_X(\epsilon\lambda + 2\pi k\epsilon)$ qui est une fonction 2π-périodique. Ce phénomène appelé le phénomène de repliement des ondes (communément appelé *phénomène d'aliasing*) a été largement étudié pour les processus de second ordre (Cf [139], [140], [141], [149], [150] et [222], entre autres). Pour pallier ce phénomène d'aliasing, nous supposons que la densité spectrale ϕ_X est à support dans l'intervalle $\mathcal{D} = (-\pi\epsilon, \pi\epsilon)$ (Cf [222]).
Soit W une fonction réelle, paire et satisfaisant

$$W \in L^1(\mathbb{R}),\ \int_{-\infty}^{+\infty} W(\lambda)\,d\lambda = 1 \text{ et } \lambda^2\,W(\lambda) \in L^1(\mathbb{R}) \tag{1.1.1}$$

On définit la fenêtre spectrale W_N par : $W_N(\lambda) = M_N W(M_N\,\lambda)$, où $(M_N)_{n\in\mathbb{N}}$ est une suite d'entiers positifs satisfaisant : $M_N \to \infty$ et $M_N/N \to 0$ quand $N \to \infty$.
Usuellement, la densité spectrale $\phi_{\tilde{X}}$ est estimée à partir des observations $\{X(n/\epsilon), n = 1,\ldots,N\}$ par une statistique empirique qu'on appelle le périodogramme (Cf [192]) :

$$\widehat{I}_N(u) = \frac{1}{2\pi N}\left|\sum_{k=1}^{N} X(k/\epsilon)\ \exp(-i\,ku)\right|^2.$$

Cet estimateur est généralement non consistant malgré qu'il soit asymptotiquement sans biais. Pour le rendre consistant, on procède par un lissage de celui-ci. Donc, on estime $\phi_X(\lambda)$ par :

$$\widehat{\phi}_X(\lambda) = \frac{1}{\epsilon}\widehat{\phi}_{\tilde{X}}\left(\frac{\lambda}{\epsilon}\right) \text{ où } \widehat{\phi}_{\tilde{X}}(\lambda) = \int_{-\infty}^{+\infty} W_N(\lambda - u)\widehat{I}_N(u)\,du$$

qui peut s'écrire sous la forme suivante, convenable pour les applications pratiques :

$$\widehat{\phi}_X(\lambda) = \frac{1}{2\pi N\epsilon}\left\{\sum_{k=1}^{N} X^2(k/\epsilon) + 2\sum_{n=1}^{N-1} W\left(n/M_N\right)\cos(n\lambda/\epsilon)\sum_{k=1}^{N-n} X\left((k+n)/\epsilon\right)\,X\left(k/\epsilon\right)\right\}$$

Les propriétés statistiques de cet estimateur s'obtiennent à partir des résultats classiques sur l'estimation de la densité spectrale pour les processus à temps discret (Cf [30] et [192]). Sous l'hypothèse sur l'α-mélangeance suivante :

$$\exists 0 < s < 1,\ (N - M_N)^{(1-\gamma+2s)/2}\,\alpha((N - M_N)^{1-s}) \text{ avec } 1 - s/2 < \gamma < 4s - 1 \qquad (1.1.2)$$

est le terme général d'une série convergente, si la fonction de covariance vérifie : $t\,c_X^{(2)}(t) \in L^1(\mathbb{R})$ et $\sum_{n=1}^{+\infty} n|c_X^{(2)}(n/\epsilon)| < \infty$, et le paramètre de lissage vérifie $M_N\,N^{\gamma-1} \to 0$, alors nous obtenons sans vitesse de convergence la convergence uniforme presque complète de $\widehat{\phi}_X(\lambda)$ vers $\phi_X(\lambda)$. En revanche, pour obtenir des vitesses de convergence de l'ordre de $(\ln(N)/N)^s$ pour $0 < s < 1/10$, nous remplaçons l'hypothèse (1.1.2) par : il existe $\kappa > 0$ et $\mu \in [0, 1[$ tel que $\alpha(k) \leq \kappa\mu^k$, $k \geq 1$ (Cf [21]).

1.1.2 Echantillonnage poissonien du temps

Dans le paragraphe précédent, nous avons effectué un échantillonnage périodique. Pour pallier le phénomène d'aliasing, nous avons supposé que le support de ϕ_X est dans $\mathcal{D}$. Il faut se rendre à l'évidence que cette condition est assez restrictive. Une solution alternative qui permet de résoudre à la fois les problèmes d'échantillonnage et d'aliasing est l'échantillonnage aléatoire du temps. En effet, on considère le processus échantillonné $\check{X} = \{X(t_n)\}_{n\in\mathbb{Z}}$ associé au processus X, en échantillonnant le temps par la suite aléatoire $\{t_n\}_{n\in\mathbb{N}}$. Dans toute la suite, les instants d'échantillonnage sont supposés être indépendants du processus X et gérés par un processus de Poisson sur $[0, +\infty)$, tel que : $t_0 = 0$ et $t_n = t_{n-1} + \theta_n$, $n \in \mathbb{N}^*$, où $\{\theta_n\}_{n\in\mathbb{N}^*}$ est une suite de v.a. positives, i.i.d., et de loi $F(x) = 1 - e^{-\beta x}$, où $\beta > 0$ est l'intensité moyenne d'échantillonnage (supposée connue). Nous renvoyons le lecteur à [101] pour l'estimation de ce paramètre.
Le principe d'estimation de ϕ_X consiste au développement de celle-ci dans une base orthonormée, et ensuite, en l'estimation des coefficients de ϕ_X dans cette base. Pour ceci, la suite des covariances $c_{\check{X}}^{(2)}(n)$ du processus échantillonné $\check{X}$ est donnée par :

$$c_{\check{X}}^{(2)}(n) = I\!E\{X(t_{k+n})\,X(t_k)\} = \int_0^{+\infty} f_n(t)\,c_X^{(2)}(t)\,dt \text{ (Cf [157])}$$

où $\forall n \in \mathbb{N}^*$, $f_n(t) = \beta(\beta t)^{n-1}/(n-1)!\ e^{-\beta t}1_{\mathbb{R}_+}(t)$, est la densité de probabilité de $(t_{k+n} - t_k)$ (f_n ne dépend pas de k). Si l'on note par $\{g_n\}_{n\in\mathbb{N}}$ le système orthonormalisé du système complet $\{f_n(t)\}_{n\in\mathbb{N}}$ dans $L^2(\mathbb{R})$, alors la densité spectrale ϕ_X s'écrit :

$$\phi_X(t) = \sum_{n=1}^{+\infty} a_n\,\mathcal{G}_n(\lambda) \in L^2(\mathbb{R}) \text{ où } \mathcal{G}_n(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp(-it\lambda)g_n(|t|)\,dt$$

Soit N un entier naturel. Considérons l'échantillon $\{X(t_k),\ k = 1, \ldots, N\}$ de $\check{X}$. On estime donc $\phi_X(\lambda)$ par l'estimateur tronqué :

$$\widehat{\phi}_X(\lambda) = \sum_{n=1}^{M_N} W_n(N)\widehat{a}_n(N)\mathcal{G}_n(\lambda) \text{ où } \widehat{a}_n(N) = \sqrt{\frac{2}{\beta}}\frac{1}{N}\sum_{k=1}^{n}(-2)^{k-1}C_{n-1}^{k-1}\sum_{j=1}^{N-k} X(t_j)\,X(t_{j+k})1_{1\leq k\leq N}$$

où $W_n(N) = W(\exp(n\,\varsigma)/N^b)$ avec $\varsigma > \ln(3)$ et $0 < b < \varsigma/(2\ln(3))$: le choix de la forme de W_n est motivé par des applications pratiques (Cf [162] et [163]). Un autre choix (sous quelques conditions supplémentaires sur W) n'influencera pas nos résultats.
Donc, si le processus X est α-mélangeant (avec $t\alpha(t) \in L^1(\mathbb{R})$), et admettant des trajectoires presque sûrement continues, alors sous des hypothèses supplémentaires sur le noyau et la largeur de fenêtre spectrale, nous montrons–en utilisant le résultat de [45]–la convergence uniforme presque complète de $\widehat{\phi}_X$ vers ϕ_X. Si de plus le processus X est géométriquement fortement mélangeant et $M_N = (\ln(N)/N)^{-\nu}$ pour $\nu > 4/25$, alors quand $\sum_{n=M_N+1}^{+\infty} |a_n| = O(N^{-r})$, $r > 0$ (condition satisfaite par une large classe de processus tels que les processus gaussiens), nous obtenons la convergence uniforme presque complète avec une vitesse de l'ordre de $(\ln(N)/N)^{2s/(2s+1)}$, avec $0 < s < \inf(\nu/(2-2\nu), 2r)$.

1.2 Processus stationnaires indexés par le corps des nombres p-adiques

Dans cette partie de notre mémoire, nous nous intéressons aux processus stochastiques à valeurs réelles mais dont le temps évolue dans un groupe abélien localement compact. En particulier, nous sommes intéressés par le groupe des nombres dits p-adiques. Les principales raisons de cet intérêt sont les suivantes : en physique mathématique, on utilisait jusqu'à aujourd'hui les nombres réels et complexes, car les coordonnées de l'espace-temps admettent une bonne description en termes de nombres réels.
Dans le but de répondre à de nombreuses questions en physique, un intérêt croissant a été porté récemment aux nombres $p-$adiques, qui ont trouvés des applications dans la théorie des *supercordes* (théorie qui utilise des distances très petites, de l'ordre de la constante de Planck), où il n'y a pas de raison de supposer que les représentations mathématiques habituelles de l'espace-temps restent valables.
Il semble que les nombres $p-$adiques trouveront des applications, non seulement, en physique mathématique, et en particulier dans la théorie des *cordes* et la théorie des corps, mais aussi dans d'autres domaines scientifiques dans lesquels il existe des comportements fractaux et de structures hiérarchiques, par exemple dans les théories des turbulences, des systèmes dynamiques, de la physique statistique et en d'autres domaines scientifiques (Cf [87], [132], [235], [236], [238] et [240]).
Notre apport dans cette direction est fondamental. En effet, il s'agit des premiers résultats dans ce domaine après ceux dans [32]. Nous pensons que nos travaux sur l'estimation de la densité spectrale p-adique feront l'objet d'une bonne base de recherche pour toute personne voulant appliquer ou tout simplement s'investir dans un tel domaine de recherche.

Vu que les nombres p-adiques ne sont pas très populaires (à nos jours) sauf pour les spécialistes en théorie des nombres et les physiciens (mécanique quantique par exemple, [237] et [239]), nous pensons qu'un aperçu sur les nombres p-adiques s'impose.

Les nombres p-adiques

Soit p un nombre premier. La norme $|.|_p$ sur le corps des nombres rationnels $\mathbb{Q}$ est définie par :

$$\forall x \in \mathbb{Q}, \quad |x|_p = \begin{cases} p^{-\nu(x)} & \text{si } x = p^{\nu(x)}\, a/b, \text{ où } p \text{ ne divise ni } a \text{ ni } b. \\ 0 & \text{si } x = 0 \end{cases}$$

où $\nu(x) \in \mathbb{Z}$, désigne l'évaluation du nombre p-adique x.
L'application $|.|_p$ est une norme sur $\mathbb{Q}$ appelée norme p-adique. La complétion de $\mathbb{Q}$ relativement à cette norme est notée Q_p, et est appelée *le corps des nombres p-adiques.*
On s'est posé la question naturelle suivante : pour p un nombre premier fixé, existe-t-il seulement Q_p comme complétion du corps $\mathbb{Q}$? La réponse fût immédiate et négative grâce au théorème d'Ostrowski c.-à-d. *les normes Euclidiennes et les normes p-adiques (p est un nombre premier), sont les seules normes non équivalentes sur* $\mathbb{Q}$.
Soit $x \in Q_p,\ (x \neq 0)$. Le nombre p-adique x peut être représenté d'une manière unique sous la forme (représentation d'Hansel) :

$$x = p^{\nu(x)} \sum_{j=0}^{\infty} a_j p^j \text{ où } \ 0 \leq a_j < p, \quad a_0 > 0, \quad j = 0, 1, 2, ... \tag{1.2.1}$$

où la série (1.2.1) est convergente relativement à la norme p-adique $|.|_p$.

Définition 1.2.1. La partie fractionnaire d'un nombre p-adique x, notée $\langle x \rangle_p$ ou $\langle x \rangle$ est définie par :

$$\langle x \rangle = \begin{cases} 0 & \text{si } \nu(x) \geq 0, \\ p^{\nu(x)} \sum_{i=0}^{-\nu(x)-1} a_i\, p^i & \text{si } \nu(x) < 0. \end{cases}$$

On remarque que pour tout $x \in Q_p$, si $\nu(x) < 0$ alors $0 < \langle x \rangle < 1$.

Dans toute la suite, on notera par $U_n(x_0)$, la boule p-adique de centre x_0 et de rayon p^n, c.-à-d. :

$$U_n(x_0) = \{x \in Q_p \,:\, |x - x_0|_p \leq p^n\}$$

et on notera également $U_n = U_n(0)$ la boule p-adique de centre 0 et de rayon p^n. Nous avons immédiatement les propriétés suivantes :

1. $U_n(x_0)$ est compacte, ouverte dans Q_p.
2. si $x_1 \in U_n(x_0)$ alors $U_n(x_1) = U_n(x_0)$: tout point de la boule $U_n(x_0)$ est centre de celle-ci.
3. si $U_n(x_0) \cap U_m(x_1) \neq \emptyset$ et $m \leq n$ alors $U_m(x_1) \subset U_n(x_0)$: deux boules p-adiques sont soit disjointes soit l'une est incluse dans l'autre.

$(Q_p, +, \times)$ est complet, séparable, localement compact et totalement discontinu.

Caractères du groupe $(Q_p, +)$ et analyse de Fourier sur Q_p

$(Q_p, +)$ est un groupe abélien localement compact. D'après le théorème de Haar, il existe une mesure μ positive sur Q_p, unique à une constante près et qui vérifie :

$$\forall a \in Q_p,\ d(t+a) = dt,\ d(at) = |a|_p\, dt \text{ et } \mu(\mathbb{Z}_p) = 1 \text{ où } \mathbb{Z}_p = U_0.$$

Pour tout $A \in \mathcal{B}_{Q_p}$ (tribu borélienne sur Q_p), $\mu(A)$ désigne la mesure de Haar de A (cette mesure est explicitée dans [122], pages 202-203).
Les caractères γ de Q_p sont les applications $\gamma : (Q_p, +) \longrightarrow (\mathbb{C}, \times)$, continues et qui vérifient :

$$|\gamma(t)| = \sqrt{\gamma(t)\gamma(-t)} = 1 \text{ et } \forall t, s \in Q_p,\ \gamma(t+s) = \gamma(t)\gamma(s)$$

A partir de [92] et [122] (pages 400-402), nous obtenons ce qui suit :

$$\forall \gamma \in \widehat{Q}_p,\ \exists \gamma \in Q_p \text{ tel que } \forall t \in Q_p\ :\ \gamma(t) = \exp(2i\pi\langle \gamma t \rangle),$$

où $\langle \gamma t \rangle$ est la partie fractionnaire du nombre p-adique γt et où $\widehat{Q}_p$ désigne le groupe dual du groupe $(Q_p, +)$.
Ceci permet de définir la transformée de Fourier de $f : Q_p \longrightarrow \mathbb{R}$, qu'on note par $\mathcal{F}f$:

$$\forall u \in Q_p, \quad \mathcal{F}f(u) = \int_{Q_p} f(x) \exp(2i\pi\langle ux \rangle)\, dx.$$

La transformée de Fourier est définie pour toute fonction absolument intégrable c.-à-d. $\forall f \in L^1(Q_p)$. Si en plus $f \in L^2(Q_p)$, alors nous avons la transformée de Fourier inverse :

$$f(x) = c \int_{Q_p} \exp(-2i\pi\langle ux \rangle)\, \mathcal{F}f(u) du, \text{ où } c \text{ est une constante positive.}$$

Comme dans le cas réel, la formule de Plancherel est définie par :

$$\int_{Q_p} |f(x)|^2 dx = c \int_{Q_p} |\mathcal{F}f(u)|^2 du.$$

Exemple 1.2.1.

1. $D_n(\lambda) = \int_{U_n} \exp(-2i\pi\langle \lambda t \rangle)\, dt$ correspond au *noyau de Dirichlet* p-adique. Après calculs :
$$D_n(\lambda) = \begin{cases} p^n & \text{si } |\lambda|_p \leq p^{-n} \\ 0 & \text{ailleurs} \end{cases}$$
2. $\mathcal{F}\delta = 1$ et $\mathcal{F}1 = \delta$ où δ est la mesure de Dirac p-adique.

Les transformées de Fourier usuelles sont données dans [92] et [234].

1.2.1 Estimation de la densité spectrale pour un processus p-adique

Soit $X = \{X(t)\}_{t\in Q_p}$ un processus stochastique à valeurs réelles et à temps p-adique. Nous supposons que X est stationnaire de second ordre, centré, continu et de fonction d'autocovariance continue et absolument intégrable (dans $L^1(Q_p)$) où :

$$c_X^{(2)}(u) = \mathrm{cum}\{X(t+u), X(t)\} \text{ pour tout } t,\ u \in Q_p.$$

La fonction d'autocovariance $c_X^{(2)}$ est semi-définie positive et continue ; alors à partir du théorème de Bochner, il existe une mesure F_X à variation totale bornée sur Q_p, telle que :

$$c_X^{(2)}(u) = \int_{Q_p} \exp(2i\pi\langle ux\rangle)\, dF_X(x),$$

F_X est dite la mesure spectrale de X, et est déterminée de façon unique à partir de $c_X^{(2)}$. Comme $c_X^{(2)} \in L^1(Q_p)$, la densité spectrale p-adique ϕ_X est donc définie par :

$$\phi_X(\lambda) = \int_{Q_p} c_X^{(2)}(t) \exp(-2i\pi\langle t\lambda\rangle)\, dt,\ \forall\lambda \in Q_p.$$

Pour estimer $\phi_X(\lambda)$, on doit se donner un échantillon dans Q_p. La question qui se pose immédiatement est : comment choisir cet échantillon ? Pour répondre à cette question, la première difficulté à laquelle nous sommes confrontés est le fait que $(Q_p, +)$ n'est pas totalement ordonné. Pour résoudre ce problème, on se sert d'une propriété topologique sur les nombres p-adiques : les boules $(U_n)_{n\in\mathbb{N}}$ forment une partition de Q_p et $\lim_{n\to+\infty} U_n = Q_p$. C'est pour cette raison que dans la section 1.2.2 nous utiliserons un échantillonnage certain (les observations seront prises dans U_n) et dans la section 1.2.3, nous procéderons par un échantillonnage aléatoire, en généralisant celui de la section 1.1.2.

1.2.2 A partir d'un échantillonnage certain

Nous supposons observer le processus X sur la boule U_n c.-à-d. nous disposons des valeurs $X(t)$, $t \in U_n$. A partir de ces observations, nous calculons la transformée finie de Fourier et puis le périodogramme p-adique :

$$I_{U_n}(\lambda) = \int_{Q_p} X(t)\ \exp(-2i\pi\langle(t-s)\lambda\rangle)\, dtds.$$

On suppose que le processus X est stationnaire jusqu'à l'ordre 4, et la fonction cumulante d'ordre quatre

$$c_X^{(4)}(u_1, u_2, u_3) = \mathrm{cum}\{X(t+u_1), X(t+u_2), X(t+u_3), X(t)\} \in L^1(Q_p) \tag{1.2.2}$$

$$\text{Il existe } n_0 \in \mathbb{N} \text{ tel que, pour } n \geq n_0, \int_{F_n} |c_X^{(2)}(t)|dt \leq \frac{\mathcal{C}}{p^n},\ \text{où } F_n = Q_p \setminus U_n \tag{1.2.3}$$

où $\mathcal{C}$ désigne une constante positive générique.
Nous avons établi sous les hypothèses (1.2.2) et (1.2.3), des résultats semblables à ceux connus classiquement sur le périodogramme, à savoir :

$$I\!E[I_{U_n}(\lambda)] = \phi_X(\lambda) + O\left(\frac{1}{p^n}\right), \text{ où } O(.) \text{ est uniforme en } \lambda$$

et

$$\lim_{n\to+\infty} \text{var}[I_{U_n}(\lambda)] = \phi_X^2(\lambda) + \phi_X^2(0).$$

Pour obtenir un estimateur consistant (asymptotiquement) de la densité spectrale p-adique, nous procédons par un lissage du périodogramme. Pour ceci, soit $(M_n)_{n\in\mathbb{N}}$ une suite de nombres rationnels, dont les termes sont des puissances de p, et telle que :

$$\lim_{n\to+\infty} M_n = 0 \quad \text{et} \quad \lim_{n\to+\infty} p^n M_n = +\infty. \tag{1.2.4}$$

Considérons la fenêtre spectrale $W_n(\lambda) = W(\lambda M_n)/M_n$ où $W : Q_p \longrightarrow \mathbb{R}$, est une fonction continue, positive et paire, et qui vérifie :

$$W \in L^\infty(Q_p) \cap L^1(Q_p) \text{ et } \int_{Q_p} W(\lambda)d\lambda = 1. \tag{1.2.5}$$

Le lissage du périodogramme conduit donc à l'estimateur de la densité spectrale p-adique suivant :

$$\widehat{\phi}_{X,n}(\lambda) = \int_{Q_p} W_n(\lambda - u) I_{U_n}(u)du.$$

Sous (1.2.2) et (1.2.3), nous obtenons donc la convergence en moyenne quadratique de $\widehat{\phi}_{X,n}$:

$$\forall \lambda \in Q_p,\ I\!E\left(\widehat{\phi}_{X,n}(\lambda)\right) = \int_{Q_p} W_n(\lambda - u)\phi_X(u)du + O\left(\frac{1}{p^n}\right)$$

$$p^n M_n \text{cov}\left(\widehat{\phi}_{X,n}(\lambda_1), \widehat{\phi}_{X,n}(\lambda_2)\right) = \left(\delta(\lambda_1 + \lambda_2) + \delta(\lambda_1 - \lambda_2)\right)\phi_X^2(\lambda_1) \int_{Q_p} W^2(u)du + O(M_n), \tag{1.2.6}$$

$\forall \lambda_1,\ \lambda_2 \in Q_p$, où δ désigne ici le symbole de Kronecher p-adique.
Pour obtenir la convergence uniforme presque complète de $\phi_X(\lambda)$, nous avons introduit des hypothèses sur la suite $\{M_n\}_{n\in\mathbb{N}} : M_n = p^{-n\alpha}$ avec $0 < \alpha < 1$. Mais, nous n'avons pas donné un ordre de grandeur de la vitesse de convergence. En revanche, sous des hypothèses sur le mélange fort de X et en utilisant des inégalités de grandes déviations, nous pensons qu'on puisse obtenir ces vitesses de convergence par les techniques classiques. Par ailleurs, en s'inspirant des résultats de [140], nous avons obtenu la normalité asymptotique de ϕ_X c.-à-d. pour tout $\lambda_1, \ldots, \lambda_r$, r nombres p-adiques ($r > 3$), on a $\left\{\sqrt{(p^n M_n)}\left[\phi_X(\lambda_i) - I\!E\{\phi_X(\lambda_i)\}\right]\right\}_{i=1}^r$ sont conjointement asymptotiquement normalement distribuées de moyenne 0 et de variance donnée par (1.2.6).

1.2.3 A partir d'un échantillonnage aléatoire

Soit $(X(\tau_k)_{k\in\mathbb{Z}}$ le processus échantillonné de X, où τ_k est une suite de variables aléatoires à valeurs p-adiques. D'après [53] et [129], il existe un processus de comptage, noté $\mathcal{N}$, indépendant de X, qui est associé à la suite $(\tau_k)_{k\in\mathbb{Z}}$ de variables aléatoires à valeurs dans Q_p. Le processus stochastique $\mathcal{N}$ est défini par :

$$\begin{array}{rccl} \mathcal{N}: & \mathcal{B}_{Q_p}\times\Omega & \longrightarrow & \mathbb{N} \\ & (A,\omega) & \longmapsto & \mathcal{N}(A,\omega)=\sum\limits_{k\in\mathbb{Z}} 1_A\left(\tau_k(\omega)\right). \end{array}$$

Pour tout $A\in\mathcal{B}_{Q_p}$, la variable aléatoire $\mathcal{N}$ suit une loi de Poisson de paramètre $\Lambda(A)=\beta\mu(A)$ où μ est la mesure de Haar sur Q_p. Comme dans la section 1.1.2, nous supposons connue l'intensité moyenne β et nous renvoyons le lecteur à [101] pour l'estimation de ce paramètre.
A partir de $\mathcal{N}$ et τ_k, on défnit le processus échantillonné comme dans le cas réel par :

Définition 1.2.2. Le processus échantillonné Z est défini par :

$$Z(A)=\int_A X(t)\mathcal{N}(dt)=\sum_{k\in\mathbb{Z}} X(\tau_k)1_A(\tau_k)=\sum_{\tau_k\in A} X(\tau_k),\quad \forall A\in\mathcal{B}_{Q_p}.$$

Le processus à accroissements Z est stationnaire de second ordre. Nous démontrons la relation générale entre Z et X suivante :

$$\phi_Z(\lambda) \;=\; \beta c_X^{(2)}(0)+\beta^2\phi_X(\lambda). \tag{1.2.7}$$

où $\phi_Z(\lambda)$ désigne la densité spectrale de Z.
A partir de (1.2.7), on pose :

$$\widehat{\phi}_X(\lambda)=\frac{1}{\beta^2}\left[\widehat{\phi}_Z(\lambda)-\beta\widehat{c}_X^{(2)}(0)\right]$$

où

$$\widehat{c}_X^{(2)}(0)=(\beta p^n)^{-1}\int_{U_n} X^2(t)\mathcal{N}(dt) \text{ et } \widehat{\phi}_Z(\lambda)=\int_{Q_p} W_n(\lambda-u)I_Z(u)du$$

avec

$$\widehat{I}_Z(\lambda)=\frac{1}{p^n}\left|\int_{U_n}\exp(-2i\pi\langle\lambda t\rangle)\,X(t)\mathcal{N}(dt)\right|^2.$$

Sous des conditions de sommabilité des cumulants d'ordre quatre, nous avons obtenu la convergence en moyenne quadratique de $\widehat{\phi}_X(\lambda)$ à partir de celles des estimateurs $\widehat{\phi}_Z(\lambda)$ et $\widehat{c}^{(2)}(0)$. Nous avons établi aussi la convergence presque complète (sans vitesse de convergence) de cet estimateur sous des conditions supplémentaires sur la largeur de fenêtre. Par ailleurs, il nous semble très facile d'obtenir un ordre de grandeur de la vitesse de convergence, en ajoutant des hypothèses de mélange fort sur le processus X.
Par ailleurs, en s'inspirant de [140], nous avons démontré la normalité asymptotique de $\widehat{\phi}_X(\lambda)$. En effet, si le processus stochastique X est stationnaire jusqu'à l'ordre k ($k\geq 3$), et si :

– $\int_{Q_p^{k-1}}(1+|u_j|)\left|c_X^{(k)}(u_1,\ldots,u_{k-1})\right| du_1\ldots du_{k-1} < +\infty$, pour $j=1,\ldots,k-1$, $k=2,\ldots,K$ (pour un entier $K\geq 2$ fixé), où $c_X^{(k)}$ est le cumulant d'ordre k du processus stochastique X.
– $\int_{Q_p^{k-1}}(1+|u_j|)\left|c_{\mathcal{N}}^{(k)}(u_1,\ldots,u_{k-1})\right| du_1\ldots du_{k-1} < +\infty$, pour $j=1,\ldots,k-1$,

où

$$c_{\mathcal{N}}^{(k)}(t_2-t_1,\ldots,t_k-t_1)\ dt_1\ldots dt_k = \mathrm{cum}(\mathcal{N}(t_1+dt_1),\ldots,\mathcal{N}(t_k+dt_k)),\ \text{pour des } t_j \text{ distincts,}$$

et si $M_n = O(p^{-n\alpha})$ pour $0<\alpha<1$, alors pour tout $\lambda_1,\ldots.\lambda_r \in Q_p$, les variables $\left\{\sqrt{(p^nM_n)}\{\widehat{\phi}_X(\lambda_i)-I\!E\{\widehat{\phi}_X(\lambda_i)\}\right\}_{i=1}^r$ sont conjointement asymptotiquement gaussiennes de moyenne 0 et de covariances d'ordre $O\left(1/p^nM_n\right)$.

1.2.4 Quand la mesure spectrale est mixte

Rappelons que les résultats que nous avons énoncés dans les paragraphes précédents sont obtenus lorsque la mesure spectrale est absolument continue par rapport à la mesure de Lebesgue (dans le cas réel) ou par rapport à la mesure de Haar (dans le cas p-adique). Dans ce paragraphe, nous nous intéressons à un cas plus général c.-à-d. quand la mesure spectrale $dF_X(t)$ possède des atomes :

$$dF_X(t) = \phi_X(t)dt + \sum_{m=1}^{q} a_m\delta_{\lambda_m},$$

où, pour tout $m\in\{1,\ldots,q\}$, $\lambda_m\in Q_p$ et a_m est un nombre réel positif.
Soit $X=\{X(t)\}_{t\in Q_p}$ un processus p-adique, à valeurs réelles, de second ordre, faiblement stationnaire, centré, de fonction d'autocovariance $c_X^{(2)}$.
Supposons que ϕ_X est continue, bornée et que pour tout $m\in\{1,\ldots,q\}$, a_m et λ_m sont connus.
Nous savons que

$$\forall t\in Q_p,\ c_X^{(2)}(t) = \int_{Q_p}\exp(2i\pi\langle tu\rangle)\,dF_X(u) \quad = \quad \mathcal{F}f(t) + \sum_{m=1}^{q} a_m\exp(2i\pi\langle t\lambda_m\rangle),$$

A partir des observations $\{X(t)\}_{t\in U_n}$ prises sur la boule $U_n=B(0,p^n)$, on défini le périodogramme $\widehat{I}_X(\lambda)$ par

$$\widehat{I}_X(\lambda) = \frac{1}{p^n}\int_{U_n^2} X(t)X(s)\exp(-2i\pi\langle(t-s)\lambda\rangle)\,dtds.$$

Dans notre situation, le périodogramme n'est pas convergent (il n'est même pas asymptotiquement sans biais)

$$I\!E(\widehat{I}_X(\lambda)) = \int_{U_n}\exp(-2i\pi\langle t\lambda\rangle)\,\mathcal{F}f(t)dt + \sum_{m=1}^{q} a_mD_n(\lambda-\lambda_m).$$

Donc, on est confronté à un problème plus délicat que dans le cas où la mesure spectrale est absolument continue par rapport à la mesure de Haar. Pour pallier cette difficulté on adopte la méthode de double fenêtres.

Méthode de double fenêtres

Nous généralisons et utilisons la méthode de double fenêtres qui a été introduite et appliquée dans [192] aux processus harmoniques, mais pour un exemple simple de fenêtres spectrales. Ensuite, dans [218], on a appliqué cette méthode aux processus stationnaires α-stables. L'intérêt de cette méthode est qu'elle permet de trouver un estimateur asymptotiquement sans biais et consistant de $\phi_X(\lambda)$, quand λ parcourt Q_p. Pour généraliser cette méthode, on introduit deux fenêtres spectrales $W_n^{(i)},\ i \in \{1,2\}$ telles que : $W_n^{(i)}(\lambda) = W(\lambda M_n^{(i)})/M_n^{(i)}$, où W satisfait les conditions (1.2.5) et où pour chaque $i = 1,\ 2$, $(M_n^{(i)})_{n\in\mathbb{N}}$ est une suite de nombres rationnels dont les termes sont des puissances de p et telle que :

$$\lim_{n\to+\infty} M_n^{(i)} = 0 \text{ et } \lim_{n\to+\infty} {M_n^{(i)}}^2 p^n = +\infty,\ M_n^{(1)} \leq M_n^{(2)},\ \lim_{n\to+\infty} M_n^{(1)}/M_n^{(2)} = 0$$

On exige de plus que les fenêtres spectrales vérifient :

$$\forall \varepsilon > 0,\ W_n^{(i)}(x) \to 0 \text{ uniformément } \forall x \in Q_p \,/\, |x|_p > \varepsilon, \text{ quand } n \to +\infty \text{ et } \forall i \in \{1,2\},$$

$$\exists\ c \in \mathbb{R} \text{ telle que } W_n^{(2)}(x) - c\, W_n^{(1)}(x) = 0\ ,\ \forall x \in B(0, p^{-n}),$$

ce qui implique que $c = M_n^{(1)}/M_n^{(2)}$. On construit donc deux estimateurs de ϕ_X :

$$\widehat{\phi}_X^{(1)}(\lambda) = \int_{Q_p} W_n^{(1)}(\lambda - u)\widehat{I}_X(u)du \ \text{ et } \ \widehat{\phi}_X^{(2)}(\lambda) = \int_{Q_p} W_n^{(2)}(\lambda - u)\widehat{I}_X(u)du,$$

pour en déduire l'estimateur suivant :

$$\widehat{\phi}_X(\lambda) = \begin{cases} \widehat{\phi}_X^{(1)}(\lambda) & \text{si } \forall m \in \{1,\ldots,q\},\ \lambda \neq \lambda_m \\ \frac{1}{1-c}\left(\widehat{\phi}_X^{(2)}(\lambda) - c\,\widehat{\phi}_X^{(1)}(\lambda)\right) & \text{si } \exists m \in \{1,\ldots,q\},\ \text{tel que } \lambda = \lambda_m. \end{cases}$$

Donc, si les cumulants d'ordre quatre de X sont absolument intégrables, nous obtenons d'une part

$$I\!E\left(\widehat{\phi}_X(\lambda)\right) = \phi_X(\lambda) + O\left(1/M_n^{(1)} + 1/M_n^{(2)}\right),\ \forall \lambda \in Q_p,$$

et d'autre part, si il existe $n_0 \in \mathbb{N}$ tel que, pour tout $n \geq n_0$, $\int_{F_n} |\mathcal{F}\phi_X(t)|\, dt \leq \mathcal{C}/p^n$, où $F_n = Q_p \setminus U_n$, alors

$$\text{var}\ (\widehat{\phi}_X(\lambda)) = \begin{cases} O\left(1/p^n M_n^{(1)}\right) & \text{si } \forall\, m \in \{1,\ldots,q\},\ \lambda \neq \lambda_m \\ O\left(1/(p^n M_n^{(1)})\right) + O\left(1/(p^n M_n^{(2)})\right) & \text{si } \exists\, m_0 \in \{1,\ldots,q\}, \lambda = \lambda_{m_0}, \end{cases}$$

Donc l'estimateur est asymptotiquement convergent en moyenne quadratique.

Par ailleurs, si les coefficients a_i pour $i = 1, \ldots, m$ sont inconnus, nous proposons comme estimateur asymptotiquement sans biais, le périodogramme modifié :

$$\widehat{I}_n(\lambda) = \frac{\widehat{I}_X(\lambda)}{\Delta_n(0)},$$

où $\Delta_n(\lambda) = p^{-n}|D_n(\lambda)|^2$ désigne le noyau p-adique de Fejer (Cf [32]), avec $D_n(\lambda) = \int_{U_n} \exp(-2i\pi\langle\lambda t\rangle)dt$ (noyau p-adique de Dirichlet). Soit λ un nombre p-adique quelconque, nous obtenons :

$$\lim_{n\to+\infty} I\!E\{\widehat{I}_n(\lambda)\} = a_i 1_{\{\exists i\in\{1,\ldots,q\},\, \lambda=\lambda_i\}}$$

1.3 Analyse spectrale des champs aléatoires stationnaires à spectre mixte

Cette étude ne sera pas complète si l'on ne s'intéresse pas à l'estimation spectrale pour des champs aléatoires. L'analyse spectrale des champs aléatoires stationnaires a été largement étudiée dans la littérature (Cf la bibliographie de [37] et [192]). Pour montrer l'intérêt indéniable du sujet, nous nous restreignons ici à une application océanographique précise : la détermination de la longueur d'onde et de la direction de la houle sur une image radar ou optique. Habituellement, une transformation de Fourier à deux dimensions de l'image permet de mettre en évidence la structure spatiale des vagues (orientation, fréquence), même s'il ne s'agit pas exactement du spectre véritable de la houle puisque les relations entre le coefficient de rétrodiffusion et la hauteur des vagues ne font pas encore l'objet d'une théorie bien établie. L'estimation de la densité spectrale de puissance de l'intensité de l'image souffre néanmoins de certaines limitations (Cf [124]).
Malgré le fait que le modèle étudié ici, est plus général que celui du paragraphe 1.2.4, les contextes restent très différents.

Estimation de la densité spectrale mixte

Nous considérons un champ aléatoire stationnaire et centré $X = \{X_{t,s}\}_{(t,s)\in\mathbb{Z}^2}$ dont la fonction d'autocovariance est $c_X^{(2)}(t_1,t_2) = I\!E(X_{s_1+t_1,s_2+t_2}X_{s_1,s_2})$. Supposons que

$$\begin{aligned} c_X^{(2)}(t_1,t_2) &= \int\int_{-\pi}^{\pi} \exp(i(t_1u_1+t_2u_2))\,\phi(u_1,u_2)du_1du_2 + \sum_{m=1}^{q} a'_m \exp(i(t_1\omega_{1m}+t_2\omega_{2m})) \\ &+ \sum_{i=1}^{q'} \int \phi_i(u)\exp(i(t_1u+t_2(a_iu+b_i)))\,du \end{aligned}$$

Le but de ce travail, est d'estimer la densité ϕ en tout point. Pour y parvenir, nous utilisons le périodogramme :

$$\widehat{I}_X(\lambda_1,\lambda_2)=\frac{1}{(2\pi)^2MN}\left|\sum_{s=1}^{N}\sum_{u=1}^{M}X_{s,u}\exp(-is\lambda_1-iu\lambda_2)\right|^2$$

Le lissage de ce périodogramme passe par deux phases. D'abord, un lissage classique pour estimer $\phi(\lambda_1,\lambda_2)$ avec $(\lambda_1,\lambda_2)\in\mathcal{A}=\{(\lambda_1,\lambda_2)\in]-\pi,\pi[^2:\ \lambda_1\neq\pm\omega_{1j},\ \lambda_2\neq\pm\omega_{2j}$ et $\lambda_2-b_i\pm a_i\lambda_1\notin 2\pi\mathbb{Z}$ pour $1\leq i\leq q'$ et $1\leq j\leq q\}$, et puis, le lissage par la méthode de double fenêtres quand $(\lambda_1,\lambda_2)\notin\mathcal{A}$.
Considérons l'estimateur suivant

$$\widehat{\phi}(\lambda_1,\lambda_2)=\begin{cases}\widehat{\phi}_1(\lambda_1,\lambda_2) & \text{si } (\lambda_1,\lambda_2)\in\mathcal{A},\\ \widehat{\phi}_2(\lambda_1,\lambda_2) & \text{sinon.}\end{cases}$$

où $\widehat{\phi}_1$ et $\widehat{\phi}_2$ sont définis de la manière suivante :

$$\widehat{\phi}_1(\lambda_1,\lambda_2)=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}W_N^{(1)}(\lambda_1-u_1)\times W_M^{(1)}(\lambda_2-u_2)\widehat{I}_X(u_1,u_2)du_1du_2.$$

zt

$$\begin{aligned}\widehat{\phi}_2(\lambda_1,\lambda_2) &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left(\frac{W_N^{(2)}(\lambda_1-u_1)-cW_N^{(1)}(\lambda_1-u_1)}{1-c}\right)\\ &\times \left(\frac{W_M^{(2)}(\lambda_2-u_2)-c'W_M^{(1)}(\lambda_2-u_2)}{1-c'}\right)\widehat{I}_X(u_1,u_2)du_1du_2.\end{aligned}$$

Les fenêtres spectrales $W_N^{(i)}$ (resp. $W_M^{(i)}$) pour $i=1,2$ et la constante c (resp. c') sont choisis de la même façon que dans la section 1.2.4.
Nous donnons deux hypothèses sur la fonction ϕ que nous utiliserons pour étudier la vitesse de convergence du biais et de la variance de $\widehat{\phi}(\lambda_1,\lambda_2)$:

$$|\phi(\lambda_1-u_1,\lambda_2-u_2)-\phi(\lambda_1,\lambda_2)|\leq\mathcal{C}||(u_1,u_2)||^\gamma \text{ où } 0<\gamma\leq 1 \tag{1.3.1}$$

$$\left|\phi(\lambda_1-u_1,\lambda_2-u_2)-\phi(\lambda_1,\lambda_2)-\frac{\partial\phi}{\partial x}(\lambda_1,\lambda_2).u_1-\frac{\partial\phi}{\partial y}(\lambda_1,\lambda_2).u_2\right|\leq\mathcal{C}||(u_1,u_2)||^\gamma \text{ où } 1\leq\gamma\leq 2, \tag{1.3.2}$$

où $\mathcal{C}$ est une constante positive générique.
Nous montrons donc que $\widehat{\phi}(\lambda_1,\lambda_2)$ est asymptotiquement sans biais :

- si $(\lambda_1,\lambda_2)\in\mathcal{A}$, alors la vitesse de convergence du biais est de l'ordre de $O\left(1/(M_N^{(1)})^\gamma+1/(L_M^{(1)})^\gamma\right)$ si ϕ satisfait (1.3.1) et de l'ordre de $O\left(1/M_N^{(1)}+1/L_M^{(1)}\right)$ si ϕ satisfait (1.3.2).

- si $(\lambda_1, \lambda_2) \notin \mathcal{A}$, alors la vitesse de convergence du biais est de l'ordre de $O(U_{N,M} + S_N + S_M)$ où

$$U_{N,M} = \begin{cases} O\left((L_M^{(1)})^2/M + 1/N\right) & \text{si } \lambda_2 - a_i\lambda_1 - b_i \in 2\pi\mathbb{Z} \\ O\left(1/N + 1/M\right) & \text{sinon,} \end{cases}$$

$$S_N = \begin{cases} O\left((M_N^{(1)})^2/N\right) & \text{si } \lambda_1 = w_{1j} \\ O\left(1/N\right) & \text{sinon} \end{cases} \quad \text{et } S'_M = \begin{cases} O\left((L_M^{(1)})^2/M\right) & \text{si } \lambda_2 = w_{2j} \\ O\left(1/M\right) & \text{sinon} \end{cases}$$

Le résultat suivant montre que notre estimateur est également consistant

$$\text{var}\left(\widehat{\phi}(\lambda_1, \lambda_2)\right) = \begin{cases} O(1/M) & \text{si } (\lambda_1, \lambda_2) \notin \mathcal{A} \\ O\Big(1/MN + R_{N,M} + T_{N,M}\Big) & \text{si } (\lambda_1, \lambda_2) \in \mathcal{A} \end{cases}$$

où

$$T_{N,M} = \begin{cases} (M_N^{(1)})^2/MN & \text{si } \lambda_1 = \pm\omega_{1j} \text{ et } \lambda_2 \neq \pm\omega_{2j} \\ (L_M^{(1)})^2/MN & \text{si } \lambda_1 \neq \pm\omega_{1j} \text{ et } \lambda_2 = \pm\omega_{2j} \\ (L_M^{(1)} M_N^{(1)})^2/MN & \text{si } (\lambda_1, \lambda_2) = (\pm\omega_{1j}, \pm\omega_{2j}) \\ 1/MN & \text{si } \lambda_1 \neq \pm\omega_{1j} \text{ et } \lambda_2 \neq \pm\omega_{2j} \end{cases}$$

et

$$R_{N,M} = \begin{cases} 1/M & \text{si } \lambda_2 \pm a_i\lambda_1 - b_i \notin 2\pi\mathbb{Z} \\ (L_M^{(1)})^2/M & \text{sinon} \end{cases}$$

Chapitre 2

Choix d'estimateurs : sélection du paramètre de lissage

Depuis une trentaine d'années, et en particulier depuis les articles précurseurs de Stone (Cf [226]) en densité et de Härdle et Marron (Cf [116]) en régression, l'étude des méthodes automatiques de sélection du paramètre de lissage est devenu un champ de recherche particulièrement Privilégié. D'un point de vue théorique, c'est le comportement asymptotique convenable de ce paramètre qui peut permettre d'atteindre des vitesses optimales de convergence. D'un point de vue pratique, c'est ce paramètre qui va déterminer le degré de lissage de la courbe estimée et donc la qualité de l'estimation. Dans les deux cas, il est nécessaire d'opter pour un paramètre de lissage qui réalise l'équilibre entre les effets de biais et de dispersion.
De nombreux articles ont été consacrés à ce sujet, comme en témoignent les revues bibliographiques dans [148] et [176] en densité ainsi que dans [232] en régression. Ces bibliographies laissent apparaître trois catégories de méthodes de sélection du paramètre de lissage : les méthodes de type validation-croisée, les méthodes de type ré-échantillonnage et les méthodes de type injection.
Les méthodes de type validation-croisée, inspirées des techniques de choix de modèle en estimation paramétrique, consiste à introduire des critères pour estimer l'erreur d'estimation et à prendre ensuite comme paramètre de lissage celui qui minimise un tel critère. Les méthodes de ré-echantillonnage consistent à utiliser les techniques de Bootstrap pour estimer la distribution de l'erreur d'estimation, estimation à partir de laquelle on peut obtenir un paramètre de lissage optimal. Les méthodes de type Injection consistent à partir directement d'un résultat de type (2.1.1) dans lequel les constantes sont précisées, d'estimer ensuite ces constantes de manière non paramétrique et finalement d'injecter ces constantes estimées dans l'expression de l'erreur d'estimation afin d'en déduire un paramètre de lissage optimal.
Les études qui se sont développées dans ce domaine se limitent à la recherche de paramètres

de lissage optimaux au sens d'une erreur de type L^2. Les seules exceptions que nous connaissions sont la méthode de double noyau (Cf [65]) qui a pour but de choisir une largeur de fenêtre adaptée à la norme L^1, et la méthode basée sur le nombre de points d'inflexion (Cf [52]) qui a pour but de choisir une largeur de fenêtre adaptée à la norme L^∞. Ces deux méthodes ne concernent que le problème d'estimation de la densité, et à notre connaissance aucun résultat d'optimalité les concernant n'est encore établi. Par erreur de type L^2 nous entendons aussi bien Erreur Quadratique Moyenne Intégrée (MISE) qu'Erreur Quadratique Intégrée (ISE) ou qu'Erreur Quadratique Empirique (ASE). Dans [147] on a montré que ces trois versions d'erreurs quadratiques sont équivalentes pour de très nombreux estimateurs non paramétriques. Dans [231], on a étendu ces équivalences au cas d'échantillons dépendants.
Malgré le nombre relativement important d'articles sur le sujet, ce domaine de recherche est loin d'être clos et de nombreuses questions restent d'actualité (Cf perspectives à la page 91). Pour l'heure, notre apport dans le domaine s'est concentré sur l'étude des techniques de validation croisée pour les estimateurs à noyau de la densité spectrale définis comme lissage du périodogramme (Cf [192] par exemple), c.-à-d. il existe une fonction paire $W(x) \in L^1(\mathbb{R})$, tel que

$$\widehat{\phi}_X(\lambda) = \int_{-\pi}^{\pi} W_T(u-\lambda)\, \widehat{I}_T(u) du \text{ où } \widehat{I}_T(u) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} X_t \exp(itu) \right|^2$$

est le périodogramme associé aux observations $X_1, \ldots, X_T$ du processus X et $W_T(u) = W(u/h)/h$.
Ces techniques sont en fait celles qui jusqu'ici ont été les plus utilisées, comme en témoigne la revue bibliographique dans [33]. Notons que ces techniques sont aussi utilisées pour d'autres estimateurs fonctionnels. On se rapportera à [63], [172], [241] et à [243] pour ce qui concerne les estimateurs Splines, ainsi qu'à [68] pour ce qui concerne l'estimation par les polynômes locaux.

2.1 Choix de la largeur de fenêtre spectrale : cas de processus à temps discret

Soit $X = \{X_t\}_{t\in\mathbb{Z}}$ un processus réel, faiblement stationnaire, centré et α-mélangeant (Cf [67]). La densité spectrale de X peut s'écrire sous la forme :

$$\phi_X(\lambda) = \frac{1}{2\pi} \sum_{t=-\infty}^{+\infty} c_X^{(2)}(t) \cos(\lambda t).$$

La méthode d'estimation développée ici ne nécessite aucune hypothèse de type paramétrique sur la densité spectrale ϕ_X, mais seulement des conditions de régularité. Il faut noter que nos résultats restent valables quand les observations constituant l'échantillon sont assujetties à diverses conditions de mélange (Cf [20], [21] et [104]).

Dans ce qui suit, nous allons construire une méthode de sélection basée sur les idées de validation croisée, et ensuite nous allons montrer qu'elle fournisse une estimation asymptotiquement optimale de ϕ_X.
Critère de validation croisée : la règle de sélection que nous proposons ici est motivée par les mêmes considérations intuitives que dans [118] et [220].
Pour ceci, notons $N = 2T$ et $\omega_j = 2\pi j/N$. Nous définissons la version validée croisée de $\widehat{\phi}_X$ par

$$\widehat{\phi}_X^{-j}(\lambda) = \int_{-\pi}^{\pi} W_T(u-\lambda)\, \widehat{I}_T^{-j}(u)du$$

avec, pour tout $j = 1, \ldots, N$,

$$\widehat{I}_T^{-j}(\omega) = \begin{cases} \widehat{I}_T(\omega) & \text{pour } \omega \notin (\omega_{2j-1}, \omega_{2j+1}) \cup (\omega_{N-(2j+1)}, \omega_{N-(2j-1)}) \\ \theta_{1,\omega}\widehat{I}_T(\omega_{2j-1}) + \theta_{2,\omega}\widehat{I}_T(\omega_{2j+1}) & \text{pour } \omega \in (\omega_{2j-1}, \omega_{2j+1}) \\ \widehat{I}_T^{-j}(-\omega) & \text{pour } \omega \in (\omega_{N-(2j+1)}, \omega_{N-(2j-1)}), \end{cases}$$

donc $\widehat{I}_T^{-j}(\omega_j)$ est identique à $\widehat{I}_T(\omega)$ sauf sur $(\omega_{2j-1}, \omega_{2j+1})$ et $(\omega_{N-(2j+1)}, \omega_{N-(2j-1)})$, sur lesquels $\widehat{I}_T^{-j}(\omega)$ interpole les points $\widehat{I}_T(\omega_{2j-1})$ et $\widehat{I}_T(\omega_{2j+1})$.
Si $\omega \notin (0, 2\pi)$, on définit $\widehat{I}_T^{-j}(\omega) = \widehat{I}_T^{-j}(\omega + 2m\pi)$ où m est un entier relatif choisi de telle sorte que $\omega + 2m\pi \in (0, 2\pi)$.
Le critère de validation croisée considéré ici est défini par

$$CV(h) = \int_{-\pi}^{\pi} \widehat{\phi}_X^2(\lambda)d\lambda - \frac{2}{N}\sum_{j=1}^{N} \widehat{\phi}_X^{-j}(\omega_j)\widehat{I}_T(\omega_j)$$

et la largeur de fenêtre sélectionnée est celle qui minimise ce critère c.-à-d.,

$$\widehat{h} = \arg\inf_{h \in H_T} CV(h) \text{ où } H_T = (AT^{-a}, BT^{-b}) \text{ avec } 0 < a < b < 1/5 + \varepsilon \text{ et } 0 < A < B < \infty.$$

Nous montrons donc sous des conditions sur les cumulants d'ordre deux et sous des hypothèses générales sur le mélange fort de X et sur le noyau W, l'optimalité asymptotique du paramètre sélectionné par notre critère en ce sens :

$$\frac{ISE(\widehat{h})}{\inf\limits_{h \in H_T} ISE(h)} \to 1, \text{ presque sûrement,} \tag{2.1.1}$$

2.2 Choix de la largeur de fenêtre spectrale : cas de processus à temps continu

Cette section se scinde en deux parties. La première partie (n'a pas fait l'objet d'une publication dans une revue) concerne le choix de la largeur de fenêtre spectrale quand le processus étudié est à temps continu sans passer par un échantillonnage du temps. Alors que dans la

deuxième partie, nous avons procédé par un échantillonnage aléatoire du temps (qui permet de pallier le phénomène de repliement des ondes) afin de rendre la méthode applicable dans la pratique.

2.2.1 Sans échantillonnage du temps

Soit $X = \{X(t)\}_{t\in\mathbb{R}}$ un processus stochastique réel à temps continu, faiblement stationnaire, centré et α-mélangeant de densité spectrale ϕ_X :

$$\forall\lambda \in \mathbb{R},\ \phi_X(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} c_X^{(2)}(u)\exp(i\lambda u)\,du.$$

Nous nous proposons de construire, d'abord, un critère non paramétrique de choix du paramètre de lissage sans passer par un échantillonnage du temps, et nous montrons que sa minimisation est asymptotiquement équivalente à celle de l'erreur quadratique intégrée. Pour ceci, supposons qu'on dispose d'un enregistrement $\{X(t),\ t \in [0,T]\}$ où T est un réel positif, et considérons l'estimateur à noyau $\widehat{\phi}_X$ de ϕ_X défini par

$$\widehat{\phi}_X(x) \quad = \quad \int_{-\infty}^{+\infty} I_T^\star(u)W_T(\lambda - u)du,$$

où $I_T^\star(u)$ est le périodogramme défini par :

$$I_T^\star(\lambda) \quad = \quad \frac{1}{2\pi}\int_{-T}^{T}\left(\frac{1}{T}\int_0^{T-u} X(s+u)X(s)ds\right)\exp(iu\lambda)\,du$$

Critère de choix du paramètre de lissage : notons $\omega_j = 2\pi j/T$. On pose

$$CV(h) = \int_{-\pi}^{\pi}\widehat{\phi}_X^2(\lambda)d\lambda - \frac{2\pi}{T}\sum_{j=1}^{T}\widehat{\phi}_X^{-j}(\omega_j)I_T^\star(\omega_j)$$

où $\widehat{\phi}_X^{-j}(\lambda)$ est la version validée croisée de $\widehat{\phi}_X(\lambda)$, définie par

$$\widehat{\phi}_X^{-j}(\lambda) = \int_{-\infty}^{+\infty} I_T^{\star\star^{-j}}(u)W_T(u-\lambda)du,$$

avec

$$I_T^{\star\star^{-j}}(\omega) = \begin{cases} I_T^\star(\omega) & \text{pour } \omega \notin \Lambda_j \\ \theta_{1,\omega}I_T^\star(\omega_{j-1}) + \theta_{2,\omega}I_T^\star(\omega_{j+1}) & \text{ailleurs} \end{cases}$$

où $\Lambda_j = (\omega_{j-1}, \omega_{j+1})$, $\theta_{1,\omega} = (\omega - \omega_{j-1})/(\omega_{j+1} - \omega_{j-1})$ et $\theta_{2,\omega} = 1 - \theta_{1,\omega}$.
Sous des conditions générales sur le mélange fort de X et sur la fenêtre spectrale, nous montrons l'optimalité asymptotique du paramêtre $\widehat{h} = \arg\min CV(h)$ au même sens qu'en (2.1.1).

2.2.2 Avec échantillonnage du temps

Dans ce paragraphe, nous contruisons et étudions un critère de choix du paramètre de lissage, quand on estime la densité spectrale pour un processus à temps continu à partir d'observations prises à des instants discrets.
Soit $X = \{X(t)\}_{t\in\mathbb{R}}$ un processus réel de second ordre, stationnaire au sens large et centré. D'après le paragraphe 1.2.3, nous estimons la densité spectrale de X en passant par l'estimation de la densité spectrale du processus échantillonné Z et de la variance de X. Ceci, nous conduit à estimer $\phi_X(\lambda)$ par :

$$\widehat{\phi}_X(\lambda) = \frac{1}{\beta^2}\left(\widehat{\phi}_Z(\lambda) - \frac{\beta}{2\pi}\widehat{c}_X^{(2)}(0)\right),$$

où $$\widehat{\phi}_Z(\lambda) = \int_{\infty}^{+\infty} W_T(\lambda-u) I_T^{\star\star}(u)du \text{ et } \widehat{c}_X^{(2)}(0) = \frac{1}{\beta T}\int_0^T X^2(t)d\mathcal{N}(t)$$

avec $\mathcal{N}$ est le processus de comptage associé aux instants d'échantillonnage et

$$I_T^{\star\star}(\lambda) = \frac{1}{2\pi T}\left|\int_0^T \exp(-iu\lambda)\, X(u)d\mathcal{N}(u)\right|^2.$$

En établissant un résultat du type

$$MISE(h) = C_1/Th + C_2h^4 + o\left(1/Th + h^4\right),$$

où C_1 et C_2 ne dépendent pas de h mais de ϕ_X et ϕ_X'', on rejoint la même problématique sur le choix de la largeur de fenêtre comme dans le cadre de la fonction de régression ou de la densité de probabilité.
Critère de choix du paramètre de lissage : comme $\widehat{I}_T(\lambda)$ n'est pas 2π-périodique et l'estimateur n'est pas construit de la même façon que dans le paragraphe précédent, nous proposons la version ci-dessous, en nous inspirant entre autres de [118] sur les critères de choix de la largeur de fenêtre pour l'estimation de la densité de probabilité pour des données dépendantes, et de [220] sur l'estimation de la fonction de hasard.
Pour ceci, pour $j = 1, \ldots, T$, considérons

$$I_T^{\star\star^{-j}}(\omega) = \begin{cases} I_T^{\star\star}(\omega) & \text{si} \quad \omega \notin \Lambda_j \\ \theta_{1,\omega} I_T^{\star\star}(\omega_{j-1}) + \theta_{2,\omega} I_T^{\star\star}(\omega_{j+1}) & \text{si} \quad \omega \in \Lambda_j \end{cases}$$

où $\theta_{1,\omega} = 1 - (\omega - \omega_{j-1})/(\omega_{j+1} + \omega_{j-1})$ et $\theta_{2,\omega} = 1 - \theta_{1,\omega}$. Nous en déduisons que $I_T^{\star\star^{-j}}(\omega)$ ne coincide jamais avec $\widehat{I}_T(\omega_j)$, et nous définissons l'estimateur validé croisé associé à Z par

$$\widehat{\phi}_Z^{-j}(\lambda) = \int_{\mathbb{R}} I_T^{\star\star^{-j}}(\omega) W_T(\lambda-\omega)d\omega.$$

En fait, $\widehat{\phi}_Z^{-j}(\omega_j)$ ne dépend pas de $\widehat{I}_T(\omega_j)$. Nous établissons donc notre critère de la manière suivante :

$$CV(h) = \int_0^{2\pi} \widehat{\phi}_X^2(x)dx - \frac{2}{T}\sum_{j=1}^{T} \widehat{\phi}_X^{-j}(\omega_j)\widetilde{I}_T(\omega_j)$$

où

$$\widehat{\phi}_X^{-j}(\lambda) = \beta^{-2}\left(\widehat{\phi}_Z^{-j}(\lambda) - \frac{\beta}{2\pi}\widehat{c}_X^{(2)}(0)\right) \text{ et } \widetilde{I}_T(\lambda) = \beta^{-2}\left(I_T^{\star\star}(\lambda) - \frac{\beta}{2\pi}\widehat{c}_X^{(2)}(0)\right)$$

est le périodogramme modifié associé à $\{X(\tau_k)\}_{k=1}^{\mathcal{N}(T)}$. Finalement, nous choisissons la largeur de fenêtre par : $\widehat{h} = \arg\min_{h\in H_T} CV(h)$ où $H_T = (A\ T^{-1/5},\ B\ T^{-1/5})$, avec $0 < A < B < \infty$. Par suite, nous établissons des résultats semblables à ceux établis dans [101] dans le cadre de l'estimation de l'intensité moyenne d'un processus ponctuel. En fait, sous des hypothèses spécifiques, nous montrons que

$$|I\!E\{CV(h) - ISE(h)\}| = O\left(\frac{1}{Th}\right) \text{ et } \mathrm{var}\{CV(h)\} = O\left(\frac{1}{Th}\right)$$

et puis

$$I\!E\left\{(CV(h) - MISE(h))^2\right\} = O\left(\frac{1}{Th}\right).$$

De plus, nous montrons que la largeur de fenêtre $\widehat{h}$ obtenue par validation croisée, est asymptotiquement optimale au sens (2.1.1), mais avec un mode de convergence faible (en probabilité).

2.3 Choix optimal de la largeur de fenêtre spectrale : cas de champ aléatoire

Considérons un champ aléatoire de second ordre stationnaire et centré, $X = \{X_{n_1,n_2} : n_1, n_2 \in \mathbb{Z}\}$, ayant des cumulants bornés jusqu'à l'ordre quatre et dont la fonction d'autocovariance est absolument sommable. L'estimation de la densité spectrale ϕ_X d'un tel processus a été largement étudiée (Cf [192] pour une revue bibliographique). Un estimateur convergeant en moyenne quadratique vers ϕ_X est donné dans [192].
A partir de l'échantillon $\{X_{n_1,n_2},\ n_1 = 1,\ldots,N_1,\ n_2 = 1,\ldots,N_2\}$ de X, on construit le périodogramme :

$$\widehat{I}_{N_1,N_2}(u_1,u_2) = \frac{1}{(2\pi)^2 N_1 N_2}\left|\sum_{n_1=1}^{N_1}\sum_{n_2=1}^{N_2} X_{n_1,n_2}\exp(-iu_1n_1 - iu_2n_2)\right|^2.$$

Comme dans le cas unidimensionnel, on montre que cet estimateur empirique n'est pas consistant. Pour le rendre consistant, on le lisse avec deux fenêtres spectrales. Pour ceci, soient $W_{N_1}^{(1)}$ et $W_{N_2}^{(2)}$, telles que :

$$W_{N_1}^{(1)}(x) = M_{N_1}^{(1)} W\left(xM_{N_1}^{(1)}\right) \text{ et } W_{N_2}^{(2)}(x) = M_{N_2}^{(2)} W\left(xM_{N_2}^{(2)}\right)$$

où W est une fonction continue, positive, paire, nulle en dehors de l'intervalle $[-1,1]$ et telle que :

$$\int_{-1}^{1} W(x)dx = 1,\ M_{N_i}^{(i)} \to +\infty \text{ et } \frac{M_{N_i}^{(i)}}{N_i} \to 0 \text{ pour } i = 1,2.$$

En convolant le périodogramme avec ces deux fenêtres spectrales, nous obtenons :

$$\widehat{\phi}_X(u_1,u_2)=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}W_{N_1}^{(1)}(u_1-x_1)W_{N_2}^{(2)}(u_2-x_2)\widehat{I}_{N_1,N_2}(x_1,x_2)dx_1dx_2.$$

Nous montrons que $\widehat{\phi}_X(u_1,u_2)$ est un estimateur asymptotiquement sans biais de $\phi_X(u_1,u_2)$ et consistant, avec :

$$\mathrm{var}\left(\widehat{\phi}_X(u_1,u_2)\right)=O\left(\frac{M_{N_1}^{(1)}M_{N_2}^{(2)}}{N_1N_2}\right).$$

Il est clair que le choix de $M_{N_1}^{(1)}$ et $M_{N_2}^{(2)}$ joue un rôle important (au niveau de la vitesse de convergence). Notons par $h_1=1/M_{N_1}^{(1)}$ et $h_2=1/M_{N_2}^{(2)}$ les largeurs des deux fenêtres spectrales. Nous cherchons donc un critère $CV(h_1,h_2)$ nous permettant de sélectionner un couple (h_1,h_2) minimisant l'erreur quadratique intégrée. Pour ceci, notons $\omega_j=2\pi j/N_1$, $\omega_{j'}=2\pi j'/N_2$, $\overline{N}_1=[(N_1-1)/2]$ et $\overline{N}_2=[(N_2-1)/2]$ où $[x]$ est la partie entière de x. nous considérons l'estimateur validé croisé suivant :

$$\widehat{\phi}_X^{-j,-j'}(x_1,x_2)=\int_0^{2\pi}\int_0^{2\pi}\widehat{I}_{N_1,N_2}^{-j,-j'}(u_1,u_2)W_{N_1}^{(1)}(x_1-u_1)W_{N_2}^{(2)}(x_2-u_2)du_1du_2$$

avec

$$\widehat{I}_{N_1,N_2}^{-j,-j'}(u_1,u_2)=\begin{cases}\widehat{I}_{N_1,N_2}(u_1,u_2) & \text{si } (u_1,u_2)\notin\Lambda_{j,j'}\\ \theta_1(u_1,u_2)\widehat{I}_{N_1,N_2}(\omega_{j-1},\omega_{j'-1})+\theta_2(u_1,u_2)\widehat{I}_{N_1,N_2}(\omega_{j+1},\omega_{j'-1}) & \\ +\theta_3(u_1,u_2)\widehat{I}_{N_1,N_2}(\omega_{j-1},\omega_{j'+1})+\theta_4(u_1,u_2)\widehat{I}_{N_1,N_2}(\omega_{j+1},\omega_{j'+1}) & \text{sinon}\end{cases}$$

où $\Lambda_{j,j'}=(\omega_{j-1},\omega_{j+1})\times(\omega_{j'-1},\omega_{j'+1})$.
La construction de $\widehat{I}_{N_1,N_2}^{-j,-j'}(u_1,u_2)$ lorsque $(u_1,u_2)\in\Lambda_{j,j'}$ est faite comme si $\widehat{I}_{N_1,N_2}$ était bilinéaire. Dans ce cas

$$\theta_1(u_1,u_2)=\alpha\beta,\ \theta_2(u_1,u_2)=(1-\alpha)\beta,\ \theta_3(u_1,u_2)=\alpha(1-\beta)\text{ et }\theta_4(u_1,u_2)=(1-\alpha)(1-\beta)$$

où $\alpha=(u_1-\omega_{j+1})/(\omega_{j-1}-\omega_{j+1})$ et $\beta=(u_2-\omega_{j'+1})/(\omega_{j'-1}-\omega_{j'+1})$.
Nous montrons que : $I\!E\left(\widehat{\phi}_X^{-j,-j'}(\lambda_1,\lambda_2)-\widehat{\phi}_X(\lambda_1,\lambda_2)\right)=O\left(1/N_1N_2\right)$. Suite à ce résultat, nous établissons notre critère, noté CV, qui est défini par :

$$CV(h_1,h_2)=\int\int_0^{2\pi}\widehat{\phi}_X^2(\lambda_1,\lambda_2)\rho(\lambda_1,\lambda_2)d\lambda_1d\lambda_2-\frac{2}{N_1N_2}\sum_{j=1}^{\overline{N}_1}\sum_{j'=1}^{\overline{N}_2}\widehat{\phi}_X^{-j,-j'}(\omega_j,\omega_{j'})\widehat{I}_{N_1,N_2}(\omega_j,\omega_{j'})\rho(\omega_j,\omega_{j'})$$

où ρ est une fonction de poids. Les largeurs de fenêtres spectrales seront choisies par :

$$(\widehat{h}_1,\widehat{h}_2)=\arg\min_{(h_1,h_2)}CV(h_1,h_2)$$

Nous montrons, par la suite, qu'en moyenne, quand N_1 et N_2 sont assez grands, le critère $CV(h_1, h_2)$ est approximativement égal à l'erreur quadratique intégrée $ISE(h_1, h_2)$ et que la variance de $CV(h_1, h_2)$ est asymptotiquement nulle :

$$|I\!E\{CV(h_1, h_2) - ISE(h_1, h_2)\}| = O\left(\frac{1}{N_1 N_2}\right) \text{ et } \operatorname{var}\{CV(h_1, h_2)\} = O\left(\frac{1}{N_1 N_2 h_1 h_2}\right).$$

De plus, le couple des largeurs des fenêtres spectrales $(\widehat{h}_1, \widehat{h}_2)$, qui minimise CV ci-dessus, est asymptotiquement optimal :

$$\frac{ISE(\widehat{h}_1, \widehat{h}_2)}{ISE(\widehat{\widehat{h}}_1, \widehat{\widehat{h}}_2)} \to 1 \text{ en probabilité, quand } N_1, N_2 \to +\infty$$

où $(\widehat{h}_1, \widehat{h}_2) = \arg \min_{(h_1,h_2)} CV(h_1, h_2)$ et $(\widehat{\widehat{h}}_1, \widehat{\widehat{h}}_2) = \arg \min_{(h_1,h_2)} ISE(h_1, h_2)$

2.3.1 Simulations

Dans ce paragraphe, nous étudions sur des données simulées, la performance de la méthode de validation croisée appliquée à l'estimateur à noyau de la densité spectrale pour un processus à temps discret, continu et discret bidimensionnel (champ aléatoire) stationnaires.
Processus à temps discret. Considérons la superposition suivante de sinusoïdes à coefficients aléatoires

$$X_t = \sum_{j=1}^{n/2} \left(A_j \cos(\omega_j t) + B_j \sin(\omega_j t)\right),$$

où $A_1, \ldots, A_{n/2}$ et $B_1, \ldots, B_{n/2}$ sont indépendantes avec A_j et $B_j \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \theta_j^2)$. Sa densité spectrale est donnée par $\phi_X(\omega_j) = \theta_j^2/2$, où $\omega_j = 2\pi j/n$ (Cf [192], pages 147, 241) ; donc nous remarquons que $\{X_t\}_{t\in\mathbb{Z}}$ fournit une approximation d'un processus ayant comme densité spectrale :

$$\phi_X(\omega_j) = \frac{n\theta_j^2}{4\pi} \text{ pour tout } j = 1, \ldots, n/2.$$

Donc par des choix propres de θ_j, nous pouvons *approximer* des processus avec des densités spectrales ayant la forme désirée. Ici, nous prenons $\theta_j = \sin(2\pi j/n+1)$ pour $j = 1, \ldots, n/2$.

Processus à temps continu. Soit $X = \{X(t)\}_{t\in\mathbb{R}^+}$ un processus gaussien, centré, de fonction de covariance et de densité spectrale données par

$$c_X^{(2)}(t) = \exp(-|t|) \text{ et } \phi_X(\lambda) = \frac{1}{\pi(1+\lambda^2)}$$

La méthode de simulation de ce processus, reposera sur une représentation, qui fait l'objet de la proposition suivante (Cf [77]).

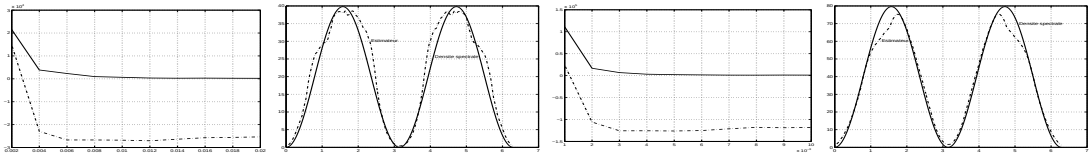

FIG. 2.1 – De gauche vers la droite : Le critère CV et ISE puis ϕ_X et $\widehat{\phi}_X$ calculé avec $\hat{h}$ pour $n = 500$. Ensuite, Le critère CV et ISE puis ϕ_X et $\widehat{\phi}_X$ calculé avec $\hat{h}$ pour $n = 10000$.

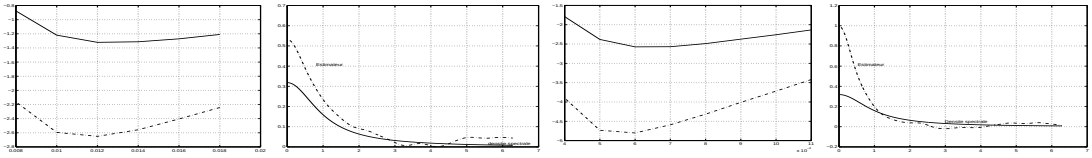

FIG. 2.2 – De gauche vers la droite : Le critère CV et ISE puis ϕ_X et $\widehat{\phi}_X$ calculé avec $\hat{h}$ pour $n = 500$. Ensuite, Le critère CV et ISE puis ϕ_X et $\widehat{\phi}_X$ calculé avec $\hat{h}$ pour $n = 10000$.

Proposition 2.3.1. *Pour tout couple* $(t, t'), (t > t' > 0)$*, on a*

$$X(t) = \exp(-(t-t'))X(t') + \left[1 + \exp(-2(t-t'))\right]^{1/2} Z_{t,t'}$$

Champ aléatoire. Nous reprenons la méthode utilisée pour simuler un processus bidimensionnel de Gauss Markov centré dont la fonction de covariance et la densité spectrale sont définies par

$$R(n_1, n_2) = \mathrm{e}^{-\sqrt{n_1^2+n_2^2}} \text{ et } \phi_X(x_1, x_2) = \frac{1}{\pi(1 + x_1^2 + x_2^2)}.$$

Nous remarquons que nos critères pour les processus à temps discret, continu ou pour les champs aléatoires fournissent des résultats très satisfaisant comme en atteste les figures FIG. 2.1, FIG. 2.2 et FIG. 2.3.

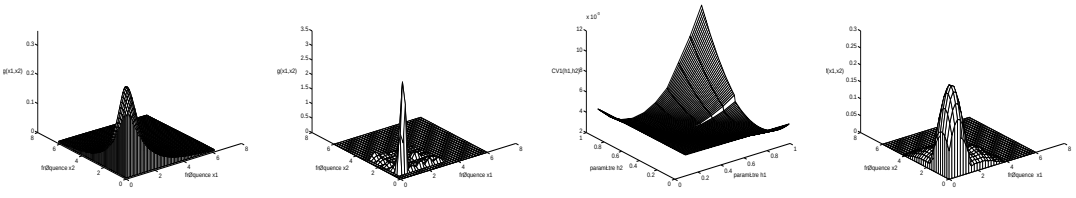

FIG. 2.3 – De gauche vers la droite : la densité spectrale ϕ_X, $\widehat{\phi}_X$ calculé pour un h choisi au hasard, le critère CV et puis $\widehat{\phi}_X$ calculé avec $(\hat{h}_1, \hat{h}_2)$.

Chapitre 3

Estimation fonctionnelle quand les observations sont entachées d'erreurs

Les données que l'on collecte sont souvent sujettes à des erreurs dûes aux manipulations des appareils de mesure, aux conditions météorologiques, aux questionnaires erronés, ou autres. Les problèmes engendrés par ces erreurs sont souvent ignorés par les statisticiens lors du développement des méthodes pour l'inférence statistique. Dans la majorité des méthodes statistiques standards on traite numériquement les données statistiques comme si elles étaient des observations réelles exactes. Il est donc important de comprendre comment l'estimation peut être affectée quand on n'a accès qu'aux données entachées d'erreurs.
Pour toutes ces raisons confondues, nous nous sommes intéressés à ce type de problématiques. Notre apport dans ce domaine touche à deux sujets particuliers. D'abord, l'estimation du mode d'une densité de probabilité pour les modèles de déconvolution c.-à-d. quand les données sont entachées d'une erreur additive (Cf paragraphe 3.1), ensuite, l'estimation de la courbe de croissance quand les données sont quantifiées et répétées et les erreurs sont corrélées (Cf paragraphe 3.2).

3.1 Estimation du mode d'une densité de probabilité quand les observations sont entachées d'erreurs

Soit $Y_1, \ldots, Y_n$ un échantillon de variables aléatoires i.i.d de densité de probabilité f_Y. Les observations sont de la forme

$$Y_i = X_i + \varepsilon_i, \forall\, i \in \{1, \ldots, n\},$$

et les erreurs $\{\varepsilon_i\}_{i=1,\ldots,n}$ sont des variables aléatoires i.i.d., indépendantes de $\{X_i\}_{i=1,\ldots,n}$. On suppose que la densité de probabilité $f_\varepsilon(x)$ des erreurs est connue. Ce modèle peut trouver

des applications dans plusieurs domaines dans lesquels les mesures (les données) ne peuvent pas être observées directement. Par exemple, concernant la maladie de SIDA, Y peut être considéré comme la durée entre l'instant du début et l'instant auquel l'infection se produit, et X est la période d'incubation (la durée entre le moment de l'infection et le moment d'apparition des symptômes). D'autres aspects pratiques des problèmes de déconvolution peuvent être trouvés dans [160].
Parmi les travaux sur l'estimation de f_X, on peut citer entre autres [44], [65], [72], [73], [143], [153], [154], [161], [224], [225] et [247] (se référer à leurs bibliographies pour la littérature).
L'estimation du mode a été considérée par plusieurs auteurs (Cf [46], [69], [70], [102], [105], [136], [179] et [233]). Dans [168] et [213] on a estimé les pics d'une régression, et dans [17] et [175] on a estimé le mode conditionnel.
Notre contribution dans ce domaine se situe dans l'étude de deux estimateurs $\widehat{\theta}_{n,1}$ et $\widehat{\theta}_{n,2}$ du mode θ de f_X et dans l'établissement de leurs convergences en termes de convergence en moyenne quadratique et normalité asymptotique. Avant d'énoncer nos résultats, il faut remarquer que pour contrôler la vitesse de convergence de l'estimateur de la densité de probabilité (Cf [72]), deux importants cas sur la fonction caractéristique φ_ε de ε, seront considérés.

- φ_ε est *exponentiellement décroissante* ou super lisse d'ordre β, à l'infini c.-à-d.

$$a_0|x|^{\beta_0}\exp(-a|x|^\beta) \leq |\varphi_\varepsilon(x)| \leq a_1|x|^{\beta_1}\exp(-a|x|^\beta) \text{ quand } x \to \infty, \tag{3.1.1}$$

  où a, a_0, a_1, β sont des constantes positives et β_0, β_1 sont des réels quelconques.
- φ_ε est *géométriquement décroissante* ou ordinairement lisse d'ordre β, à l'infini c.-à-d.

$$d_0|x|^{-\beta} \leq |\varphi_\varepsilon(x)| \leq d_1|x|^{-\beta} \text{ quand } x \to \infty, \tag{3.1.2}$$

  où d_0, d_1, β sont des constantes positives.

Pour définir nos estimateurs, soit K une densité de probabilité paire et bornée, dont la fonction caractéristique est $\varphi_K(t)$. Notons par $\widehat{\varphi}(t) = 1/n\sum_{k=1}^n \exp(i\,t\,Y_k)$ la fonction caractéristique empirique de $\{Y_k\}_{k=1,\ldots,n}$.
On estime $f_X^{(j)}(x)$ la dérivée $j$ème de $f_X(x)$ par :

$$\widehat{f}_X^{(j)}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp(-i\ t\,x)(-i\,t)^j\varphi_K(t\,h_n)\frac{\widehat{\varphi}(t)}{\varphi_\varepsilon(t)}dt = \frac{1}{n\,h_n^{j+1}}\sum_{k=1}^n \widetilde{K}_n^{(j)}\left(\frac{x-Y_k}{h_n}\right)$$

où $\widetilde{K}_n^{(j)}(.)$ est la $j$ème dérivée du noyau de déconvolution :

$$\widetilde{K}_n(t) = 1/2\pi\int_{\mathbb{R}} \exp(-itu)\varphi_K(u)/\varphi_\varepsilon(u/h_n)du.$$

Premier estimateur : soit $\mathcal{D} = (a, b)$. Le premier estimateur $\widehat{\theta}_{n,1}$ du mode que nous considérons est le mode empirique (Cf [179]). Cet estimateur du mode est défini en supposant que $\widehat{\theta}_{n,1}$ est l'unique mode de $\widehat{f}_X$ sur $\mathcal{D}$.
Sous des conditions générales sur le noyau et la largeur de fenêtre, nous montrons la convergence en moyenne quadratique de $\widehat{\theta}_{n,1}$ vers θ avec une vitesse de convergence de l'ordre de

$(\ln n)^{-2/\beta}$ sous (3.1.1) et de l'ordre de $n^{-2/(5+2\,\beta)}$ sous (3.1.2).

Deuxième estimateur : le deuxième estimateur du mode $\widehat{\theta}_{n,2}$ consiste en l'estimation de la dérivée première $f_X^{(1)}$ de la densité f_X (Cf [233]). Cet estimateur est défini, en supposant que :

$$\exists\, \widehat{\theta}_{n,2} \in \mathcal{D} \text{ tel que } 0 = \widehat{f}_X^{(1)}(\widehat{\theta}_{n,2}) < |\widehat{f}_X^{(1)}(x)|, \text{ pour tout } x \in \mathcal{D},\ x \neq \widehat{\theta}_{n,2}.$$

De façon similaire, sous des conditions générales sur le noyau et sur la largeur de fenêtre, nous montrons la convergence en moyenne quadratique de $\widehat{\theta}_{n,2}$ vers θ avec une vitesse de convergence de l'ordre de $(\ln n)^{-2/\beta}$ sous (3.1.1) et de l'ordre de $n^{-2/(5+2\,\beta)}$ sous (3.1.2).

Normalité asymptotique des deux estimateurs : si f_X admet une dérivée seconde continue, alors

$$f_X^{(1)}(\theta) = 0\,,\ f_X^{(2)}(\theta) < 0.$$

De façon similaire, si la dérivée seconde de l'estimateur $\widehat{f}_X$ existe, alors

$$\widehat{\theta}_{n,k} \in \mathbb{R} \text{ tel que } \widehat{f}_X^{(1)}(\theta_{n,k}) = 0 \text{ et } \widehat{f}_X^{(2)}(\widehat{\theta}_{n,k}) < 0,$$

où $\widehat{\theta}_{n,k}$ est le mode de $\widehat{f}_X$ pour $k = 1, 2$.
Alors, sous des hypothèses techniques, pour $k = 1,\ 2,\ \widehat{\theta}_{n,k}$ est asymptotiquement gaussien (quand $n \to \infty$) :

$$\begin{aligned}
\left(n\, h_n^{3+2\beta}\right)^{1/2} \left(\widehat{\theta}_{n,k} - \theta\right) &\to \mathcal{N}\left(0, \left\{ f_X^{(1)}(\theta)/(f_X^{(2)}(\theta))^2 \right\} J\right), \text{ sous (3.1.2)} \\
\frac{\left(\widehat{\theta}_{n,k} - \theta\right)}{\sigma(n)} &\to \mathcal{N}(0,1)\,, \text{ sous (3.1.1)}
\end{aligned}$$

où $J = 1/2\pi|c|^2 \int_{\mathbb{R}} |u|^{2(1+\beta)}\ |\varphi_K(u)|^2\ du$ et $\sigma(n)^2 = \mathrm{var}(\widehat{f}_X^{(1)}(\theta))$.

3.2 Estimation de la régression pour des données quantifiées et des erreurs corrélées

La forme la plus connue de la *quantification* (quantization) est l'erreur d'arrondi, qui se produit dans tous les systèmes numériques. Un *quantificateur* (quantizer) général approxime une valeur observée par la plus proche valeur parmi un nombre fini de valeurs représentatives.
Le problème de quantification est considéré le plus souvent dans les domaines de communication et dans la théorie de l'information et du signal. Dans la théorie de communication, dans [144] et [158] on a montré comment choisir le meilleur système de quantification en minimisant l'erreur quadratique moyenne. Une vue d'ensemble sur la théorie et les techniques de quantification est donnée dans [100] et des applications en statistiques peuvent être trouvées dans [137] et [138]. Dans [39] on a fourni le quantificateur asymptotique optimal en

utilisant les pourcentiles d'une densité de probabilité qui minimise asymptotiquement l'erreur quadratique moyenne d'une classe de fonctions de densités positives. Dans [7] et [35] on a considéré l'approximation d'une intégrale aléatoire à partir d'observations quantifées quand le processus à intégrer admet une covariance possédant une structure qui se comporte comme la covariance de Wiener sur la diagonale.
Le modèle de courbe de croissance est très utilisé en biostatistique (étude de la croissance des animaux, plantes, ...), en recherche médicale, en études épidémiologiques, et a été considéré par plusieurs auteurs (Cf [91] et [214]). Dans [185] on a étudié l'estimation de la surface en dessous de la fonction de croissance, qui est connue comme étant la courbe de concentration par rapport au temps en recherche pharmacologique, et est basé sur la concentration du médicament par rapport aux différentes zones d'un comprimé ou de l'organisme. Par ailleurs, le modèle de régression non paramétrique avec des erreurs corrélées a été considéré par plusieurs auteurs (Cf [142], [221], [244], [245] et [246], entre autres). Dans ces travaux on a considéré différentes constructions de l'estimateur à noyau de la régression pour améliorer l'estimateur à noyau standard considéré dans [25] et [90], et dans [84] quand des observations corrélées sont introduites. Dans [248] on trouve une très bonne synthèse sur les méthodes paramétriques et non paramétriques qui permettent de modéliser les courbes de croissance avec des mesures répétées et des données longitudinales.
Nous nous intéressons donc à l'estimation de la courbe moyenne de croissance quand la variable explicative est déterministe (fixed-design). Nous répétons m fois l'expérience. Dans chacune de ces expériences n mesures de la réponse sont obtenues :

$$Y_j(x_i) = f(x_i) + \varepsilon_j(x_i) \text{ pour } j = 1, \ldots, m \text{ et } i = 1, \ldots, n,$$

où f est la fonction de croissance inconnue et $\varepsilon = \{\varepsilon_j\}_{j\in\mathbb{N}}$ est le processus d'erreurs.
Les points d'échantillonnage $\{x_i,\, i = 1, \ldots, n\}$ sont équidistants, mais d'autres types d'échantillonnages comme l'échantillonnage régulier déterministe (non uniforme) ou l'échantillonnage aléatoire peuvent être considérés (Cf [38]). Bien que les mesures répétées soient naturellement disponibles dans les applications pratiques, elles permettent d'obtenir un estimateur de f asymptotiquement consistant (Cf les commentaires dans [113] et [119]).
Nous nous intéressons donc à estimer de façon consistante f, d'abord, à partir des observations $\{Y_j(x_i) : i = 1, \ldots, n \text{ et } j = 1, \ldots, m\}$ quand les erreurs sont corrélées, et ensuite, à partir des observations quantifiées $\{Q(Y_j(x_i)) : i = 1, \ldots, n \text{ et } j = 1, \ldots, m\}$, où Q est la fonction de quantification et les x_i sont des constantes connues telles que : $0 \leq x_1 < x_2 < \cdots < x_n \leq 1$.
Le processus d'erreur ε est supposé centré, gaussien et faiblement stationnaire de fonction de covariance :

$$cov(\varepsilon_j(x), \varepsilon_l(y)) = \begin{cases} \rho(x-y) & \text{si } j = l \\ 0 & \text{si } j \neq l, \end{cases}$$

où ρ est une fonction paire telle que $|\rho(u)| \leq 1$ pour tout $u \in [-1, 1]$. De plus, on suppose que la fonction d'autocovariance vérifie la condition de Hölder d'ordre $\alpha > 0$. Donc ρ peut

être décomposée autour de 0 comme suit :

$$\rho(t) = \begin{cases} \rho(0) - \lambda|t|^{\alpha} + o(|t|^{\alpha}) & \text{quand } 0 < \alpha < 2 \\ \rho(0) + \frac{|t|^2}{2}\rho''(0) + o(t^2) & \text{quand } \alpha \geq 2, \end{cases} \tag{3.2.1}$$

pour $\lambda > 0$ et $\rho''(0) \neq 0$.
Notre apport dans cette direction se situe dans l'établissement des performances asymptotiques en termes de l'erreur quadratique moyenne intégrée de l'estimateur de la courbe de croissance moyenne construit à partir d'observations quantifiées, et la détermination de la largeur de fenêtre asymptotiquement optimale qui va dépendre de la régularité du processus à travers les paramètres α et λ, le nombre de répétitions m et le nombre de niveaux de quantification N. Nous étudions, également, à travers un exemple de simulation le comportement de la performance de l'estimateur en termes des paramètres n, m et N. Nous comparons aussi la performance de l'estimateur quantifié et l'estimateur contruit à partir de données non quantifiées.

Estimation de la courbe de croissance moyenne quand les erreurs sont corrélées. L'estimateur de f basé sur les observations $\{Y_j(x_i),\ i = 1, \ldots, n,\ j = 1, \ldots, m\}$, quand les x_i sont des constantes connues, telles que $0 \leq x_1 < x_2 < \cdots < x_n \leq 1$, avec $\max_i |x_i - x_{i-1}| = O(1/n)$, est donné quand $0 < x < 1$ (Cf [90] et [119]), par :

$$\widehat{f_h}(x) = \frac{1}{n}\sum_{i=1}^{n} W_{h,i}(x)\ \bar{Y}(x_i),$$

où

$$\bar{Y}(x) = \frac{1}{m}\sum_{j=1}^{m} Y_j(x) \text{ et } W_{h,i}(x) = n\int_{m_{i-1}}^{m_i} K_h(x-u)\,du,$$

et les points moyens $\{m_i,\ i = 0, \ldots, n\}$ sont définis par

$$m_0 = 0,\ m_i = (x_i + x_{i+1})/2,\ \text{quand } i = 1, \ldots, n-1 \text{ et } m_n = 1,$$

avec $K_h(x) = 1/h\,K(x/h)$ où le noyau K est une fonction satisfaisant les hypothèses suivantes :

$$K \text{ est une fonction paire, de Lipshitz et de support } [-1,1] \text{ avec } \int_{-1}^{1} K(v)\,dv = 1,$$

et $h = h(n,m)$ est la paramètre de lissage vérifiant $h \geq 0$ et $\lim_{n,m\to+\infty} h = 0$ quand $m/n = O(1)$ (Cf [119], pour les commentaires).
Si f est deux fois continûment dérivable sur $[0,1]$ avec $f''(x) \neq 0$, alors nous obtenons

$$\mathbb{E}\left(\widehat{f_h}(x) - f(x)\right)^2 =$$

$$\begin{cases} \frac{1}{m}\left(\rho(0) - \lambda h^{\alpha}\mathcal{C}_K(\alpha)\right) + \frac{h^4}{4}(f''(x))^2 d_K^2 + O\left(\frac{1}{n^{\alpha}m} + \frac{h^2}{n}\right) + o\left(h^4 + \frac{h^{\alpha}}{m}\right) & \text{si } 0 < \alpha < 2 \\ \frac{1}{m}\left(\rho(0) + \frac{h^2}{2}\rho''(0)\mathcal{C}_K(2)\right) + \frac{h^4}{4}(f''(x))^2 d_K^2 + O\left(\frac{1}{n^2m} + \frac{h^2}{n}\right) + o\left(h^4 + \frac{h^2}{m}\right) & \text{si } \alpha \geq 2, \end{cases}$$

où

$$d_K = \int_{-1}^{1} u^2 K(u)\, du \text{ et } \mathcal{C}_K(\alpha) = \int_{-1}^{1}\int_{-1}^{1} |u-v|^\alpha K(u)\, K(v)\, du\, dv.$$

De plus, cette erreur quadratique moyenne est minimale si le paramètre de lissage est :

$$h^* = \begin{cases} \left(\frac{\lambda\alpha\mathcal{C}_K(\alpha)}{d_K^2(f''(x))^2}\right)^{\frac{1}{4-\alpha}} m^{-\frac{1}{4-\alpha}} & \text{si } 0<\alpha<2 \\ \\ \left(\frac{-2\rho''(0)}{d_K(f''(x))^2}\right)^{\frac{1}{2}} m^{-\frac{1}{2}} & \text{si } \alpha \geq 2. \end{cases}$$

Remarque 3.2.1.

1. Si $\alpha = 1$, alors (comme pour le processus d'erreur d'Ornstein-Uhlenbeck) nous avons $\lambda = \rho'(0-)$, et le paramètre de lissage optimal est :

$$h^* = \left(\frac{\rho'(0-)\mathcal{C}_K(1)}{d_K^2(f''(x))^2}\right)^{\frac{1}{3}} m^{-\frac{1}{3}}.$$

   Celui-ci correspond à la largeur de fenêtre obtenue dans [119].

2. Quand le processus d'erreur est non stationnaire (comme le processus de Wiener) et de fonction d'autocovariance $R(s,t) = \sigma^2 \min(s,t)$, On montre que la largeur de fenêtre optimale est de la forme :

$$h^* = \left(\frac{\sigma^2\mathcal{C}_K(1)}{2d_K^2(f''(x))^2}\right)^{\frac{1}{3}} m^{-\frac{1}{3}}.$$

3. En utilisant le paramètre de lissage h^*, nous obtenons :

$$MSE(h^*) \sim \begin{cases} \frac{1}{m}\rho(0) - \frac{A'(\alpha)}{m^{\beta(\alpha)}} & \text{quand } 0<\alpha<2 \\ \\ \frac{1}{m}\rho(0) - \frac{C_0}{m^2} & \text{quand } \alpha \geq 2, \end{cases}$$

   pour des constantes positives $A(\alpha)$ et C_0 où $1 < \beta(\alpha) = 4/(4-\alpha) < 2$.

Estimation de la courbe moyenne de croissance à partir d'observations quantifiées et des erreurs corrélées. Le système de quantification est déterminé par les niveaux de quantification $z_1 < z_2 < \ldots < z_N$ et par les bornes des intervalles $y_1 < y_2 < \ldots < y_N$. Les niveaux de quantification sont les pourcentiles d'une fonction de densité de probabilité continue p_Q donnée par (Cf [39]) :

$$\int_{z_{k-1}}^{z_k} p_Q(t)\, dt = \frac{1}{N},\ k = 2, \ldots, N,$$

et les intervalles de quantification sont définis par :

$$y_k = \frac{(z_{k-1} + z_k)}{2},\ k = 2, \ldots, N.$$

Donc, la fonction de quantification est définie par :

$$Q(y) = z_k \text{ quand } y_k < y < y_{k+1},$$

où $-\infty = y_1 < z_1 < y_2 < z_2 < \ldots < y_N < z_N < y_{N+1} = +\infty$, (Cf [100], [144] et [158]).
Nous considérons l'estimateur de f construit à partir des observations bruitées $\{Q(Y_j(x_i)) : i = 1, \ldots, n \text{ et } j = 1, \ldots, m\}$:

$$\widehat{f}_{Q,h}(x) = \frac{1}{n} \sum_{i=1}^{n} W_{h,i}(x) \bar{Z}(x_i),$$

avec

$$\bar{Z}(x) = \frac{1}{m} \sum_{j=1}^{m} Q(Y_j(x)),$$

Nous supposons que le processus d'erreur $\varepsilon = \{\varepsilon_j\}_{j \in \mathbb{N}}$ est centré, gaussien et faiblement stationnaire dont la fonction d'autocovariance ρ qui satisfait l'hypothèse (3.2.1). Il est prouvé que le processus de quantification $Q(Y(x))$ admet une fonction d'autocovariance ρ_Q qui peut être développée au voisinage de 0 comme suit :

$$\rho_Q(t) = \begin{cases} \rho_Q(0) - \lambda_N |t|^{\alpha/2} + o(|t|^{\alpha/2}) & \text{si } 0 < \alpha < 2 \\ \rho_Q(0) - \beta_N |t| + o(|t|) & \text{si } \alpha \geq 2, \end{cases} \tag{3.2.2}$$

où

$$\lambda_N = \left(\frac{\alpha \sqrt{\lambda}}{2\sqrt{\pi \rho(0)}} \right) B_N \text{ et } \beta_N = \left(-\frac{\rho''(0)}{2\pi \rho(0)} \right)^{1/2} B_N,$$

avec

$$B_N = \frac{1}{\sqrt{2\pi}} \sum_{k=2}^{N} (z_k - z_{k-1})^2 \exp(-y_k^2/2).$$

Les niveaux de quantification optimaux z_k^*, $k = 1, \ldots, N$ correspondent aux pourcentiles de la densité optimale asymptotique :

$$p_Q^*(t) = \frac{1}{\sqrt{6\pi}} \exp\left(-\frac{t^2}{6} \right), t \in \mathbb{R}.$$

Si l'on suppose que f est deux fois dérivable sur $[0,1]$ avec $f''(x) \neq 0$, alors la MSE de l'estimateur $\widehat{f}_{Q,h}$ est donnée, quand n, m et $N \to +\infty$, par :

$$\begin{aligned}
&I\!E\left(\widehat{f}_{Q,h}(x) - f(x)\right)^2 \\
&= \begin{cases}
\frac{1}{m}\left(\rho_Q(0) - \frac{3}{2}\frac{h^{\alpha/2}}{N}\,\alpha\sqrt{\frac{\lambda}{\rho(0)}}\,\mathcal{C}_K(\alpha/2)\right) + \frac{h^4}{4}d_K^2(f''(x))^2 \\
+O\left(\frac{1}{Nmn^{\alpha/2}} + \frac{h^2}{N} + \frac{1}{nN} + \frac{1}{N^2}\right) + o\left(h^4 + \frac{h^{\alpha/2}}{Nm}\right) & \text{si } 0 < \alpha < 2 \\
\\
\frac{1}{m}\left(\rho_Q(0) - 3\frac{h}{N}\sqrt{-\frac{\rho''(0)}{2\rho(0)}}\mathcal{C}_K(1)\right) + \frac{h^4}{4}d_K^2(f''(x))^2 \\
+O\left(\frac{1}{Nmn} + \frac{h^2}{N} + \frac{1}{nN} + \frac{1}{N^2}\right) + o\left(h^4 + \frac{h}{Nm}\right) & \text{si } \alpha \geq 2,
\end{cases}
\end{aligned}$$

De plus, la MSE est asymptotiquement minimale si l'on considère le paramètre de lissage suivant :

$$h_Q^* = \begin{cases}
\left(\frac{3\alpha^2\sqrt{\lambda/\rho(0)}\,\mathcal{C}_K(\alpha/2)}{4d_K^2(f''(x))^2}\right)^{2/(8-\alpha)} (mN)^{-2/(8-\alpha)} & \text{si } 0 < \alpha < 2 \\
\\
\left(\frac{3\sqrt{-\rho''(0)/\rho(0)}\mathcal{C}_K(1)}{d_K^2(f''(x))^2}\right)^{1/3} (mN)^{-1/3} & \text{si } \alpha \geq 2.
\end{cases}$$

Remarque 3.2.2.

1. En utilisant la largeur de fenêtre h_Q^*, nous obtenons :

$$MSE(h_Q^*) \sim \begin{cases}
\frac{1}{m}\rho_Q(0) - \frac{A'(\alpha)}{(mN)^{\eta(\alpha)}} & \text{si } 0 < \alpha < 2 \\
\\
\frac{1}{m}\rho_Q(0) - \frac{C_0'}{(mN)^{4/3}} & \text{si } \alpha \geq 2,
\end{cases}$$

   où $A'(\alpha)$ et C_0' sont des constantes positives et $1 < \eta(\alpha) = 8/(8-\alpha) < 4/3$.
   La vitesse du second terme de MSE, c.-à-d. $MSE - \rho_Q(0)/m$, aura une vitesse de convergence plus faible que lorsque les données ne sont pas quantifiées si, et seulement si, $m > N^{2(4-\alpha)/\alpha}$ où $0 < \alpha < 2$ et $m > N^2$ avec $\alpha \geq 2$.

2. La largeur de fenêtre optimale h_Q^* (pour les données quantifiées) peut être comparée avec la largeur h^* (pour les données non quantifiées). On peut remarquer que $h_Q^* < h^*$ si, et seulement si, les niveaux de quantitification sont tels que $N > A''(\alpha)m^{\gamma(\alpha)}$ où $0 < \alpha < 2$ et $N > C_0''m^{1/2}$ avec $\alpha \geq 2$, et où $\gamma(\alpha) = \alpha/(2(4-\alpha)) < 1/2$ avec $A''(\alpha)$ et C_0'' sont des constantes positives.

Simulations. Nous présentons un exemple de simulation afin de rendre plus clair les résultats théoriques que nous venons d'établir. Les données sont générées à partir de la courbe de croissance quadratique : $f(x) = x^2$, $0 < x < 1$ et à partir de différentes valeurs de la taille

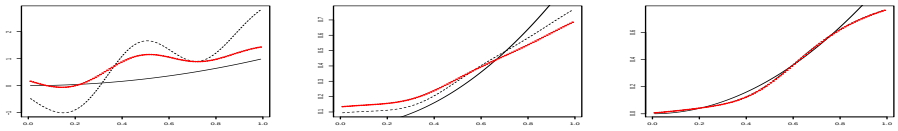

FIG. 3.1 – La courbe de croissance quadratique f en ligne continue, l'estimateur $\widehat{f}_h$ en ligne discontinue et l'estimateur $\widehat{f}_{Q,h}$ en petits cercles quand $n = 100$, $N = 15$ et $m = 1, 10, 50$.

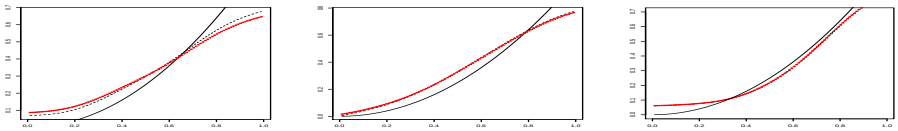

FIG. 3.2 – La courbe de croissance quadratique f en ligne continue, l'estimateur $\widehat{f}_h$ en ligne discontinue et l'estimateur $\widehat{f}_{Q,h}$ en petits cercles quand $n = 100$, $m = 50$ et $N = 3, 10, 50$

de l'échantillon n, du nombre de répétitions m et du nombre de niveaux de quantification N. Les points d'échantillonnage sont tels que $x_i = i/(n+1)$, $i = 1, \ldots, n$.
Nous considérons comme fonction d'autocovariance ρ du processus d'erreur :

$$\rho(t,s) = \exp(-\lambda|t-s|^{\alpha}),\ 0 < \alpha < 2 \text{ et } \lambda > 0.$$

Les données sont générées à partir de la fonction f et du processus d'erreur suivant :

$$\varepsilon_i = 0,75\,\varepsilon_{i-1} + Z_i,\ \text{où } Z_i \text{ sont iid } \sim \mathcal{N}(0,1), \lambda = -\ln(0,75) \text{ et } \alpha = 1.$$

On remarque à partir des figures FIG. 3.1, que dans le cas d'absence de répétitions ($m = 1$), les deux estimateurs admettent de mauvaises performances pour une taille d'échantillon $n = 100$. Quand on dispose d'un nombre satisfaisant de répétitions ($m = 10$, $m = 50$), les deux estimateurs admettent de bonnes performances. De plus, quand le nombre de répétitions m augmente, ces estimateurs deviennent de plus en plus meilleurs, comme nous l'avions établi. Ensuite, nous comparons les performances des deux estimateurs, sur 100 simulations, en augmentant le niveau de quantification N et en fixant $(n, m) = (100, 50)$. On remarque à partir des figures FIG. 3.2, que lorsque N est petit l'estimateur $\widehat{f}_h$ admet une performance meilleure que l'estimateur quantifié $\widehat{f}_{Q,h}$. Cependant, les deux estimateurs admettent les mêmes performances quand $N \approx 10$. De plus, quand N est très grand, les intervalles de quantification sont tellement petits que les observations quantifiées s'approchent de plus en plus des observations non quantifiées. C'est pour cela que la différence entre les deux estimateurs devient très minime. Il serait donc intéressant d'étudier le compromis entre le nombre d'observations et le nombre de niveaux de quantification. Par ailleurs, il faut noter que ces estimateurs souffrent des problèmes aux bords. Pour pallier ce problème on peut utiliser des estimateurs par polynômes locaux (Cf [74] et [75]). Manifestement, les mêmes conclusions restent valables à propos du problème de quantification si des estimateurs plus stables sont utilisés.

Chapitre 4

Estimation fonctionnelle à partir de processus non stationnaires

Dans nos précédents travaux traitant de l'estimation de la densité spectrale (resp. de la courbe de croissance), nous avons supposé que le processus (resp. le processus d'erreur) étudié est stationnaire. Alors que, la pertinence de cette hypothèse est discutable. Nous avons traité cette problématique en essayant de s'affranchir de cette condition. Ceci a conduit à la réalisation des travaux résumés dans ce chapitre. Ce chapitre se scinde donc en deux grands paragraphes : *Estimation spectrale pour des processsus non stationnaires* et *Estimation de la courbe de croissance quand le processus d'erreur n'est pas stationnaire*. Dans chaque problématique, des simulations ont été conduites afin de reconfirmer l'exactitude des résultats théoriques sur des données simulées.

4.1 Estimation spectrale pour les processus M-stationnaires à partir d'un échantillonnage aléatoire

Pour s'affranchir de l'hypothèse de stationnarité, nous nous sommes intéressés à une classe de processus non staionnaires. Cette classe contient les processus dits multiplicativement stationnaires ou M-stationnaires. Bien que ces processus ne soient pas stationnaires par rapport à la loi additive, ils sont, tout de même, stationnaires par rapport la loi multiplicative. Cette propriété garantit au moins une structure similaire à celle vérifiée par les processus stationnaires, spécialement la représentation spectrale.
Le concept de stationnarité sous une loi de composition interne (dans un groupe) a été intensivement considéré (Cf [111]). Il peut aussi se généraliser à une opération dans un semi-groupe (Cf [97]). Les processus M-stationnaires sont définis et existent naturellement en tant que processus à temps continu (Cf [99]). Cependant, aucun avantage ne peut être

tiré de cette généralisation de la notion de stationnarité sans passer par un échantillonnage du temps. Un aspect particulier de notre étude est de permettre à ces processus d'être utilisables en pratique, en estimant leur densité spectrale. Précisément, nous adoptons une technique d'échantillonnage aléatoire, et montrons que les schémas utilisés dans [150] et [163] peuvent être utilisés ici. Nous avons donc obtenu la convergence en moyenne quadratique de l'estimateur construit. Nous avons testé la validité de cette estimation sur un processus de Gauss-Markov M-stationnaire.

Définition 4.1.1. Un processus stochastique $X = \{X(t),\, t > 0\}$ est dit faiblement multiplicativement stationnaire (ou M-stationnaire), si :

$$\forall t, \tau > 0,\ I\!E(X(t)) = \mu \text{ est indépendant de } t \text{ et sa variance } \mathrm{var}(X(t)) < +\infty$$

$$\text{et } I\!E((X(t) - \mu)(X(t.\tau) - \mu)) = c_X^{(2)}(\tau) \text{ existe et indépendante de } t.$$

Au sens classique, les processus M-stationnaires sont non stationnaires, mais, à chaque processus M-stationnaire correspond un processus stationnaire. Ils sont donc dits *processus réductibles* (Cf [99]) :
Soit $Y = \{Y(u),\, u \in \mathbb{R}\}$ un processus stochastique défini par $Y(u) = X(t)$ où $t = \exp(u)$. Alors

$$X \text{ est M-stationnaire si, et seulement si, } Y \text{ est stationnaire.}$$

Le processus Y est appelé processus dual stationnaire de X.
Notons par $c_Y^{(2)}$ la fonction d'autocovariance de Y. Alors, pour tout $\tau > 0$, nous avons

$$c_X^{(2)}(\tau) = c_X^{(2)}(\tau^{-1}) = c_Y^{(2)}(\ln(\tau)).$$

Avant de continuer, nous donnons quelques exemples de processus M-stationnaires.

Exemple 4.1.1.

- Considérons le processus X, défini pour $t > 0$ par : $X(t) = a\cos(\beta \ln(t) + \omega)$, où a est une constante et ω est une variable aléatoire de loi uniforme sur $[0, 2\pi]$. Alors X est M-stationnaire, centré et de fonction d'autocovariance $c_X^{(2)}(\tau) = A^2\cos(\beta \ln(\tau))/2$ pour tout $\tau > 0$.
- Soit le processus X, défini pour $t > 0$ par : $X(t) = \sum_{i=1}^{N} (A_i \cos(a_i \ln t) + B_i \sin(a_i \ln t))$, où pour chaque i et j, A_i et B_i sont des variables aléatoires centrées et non corrélées et telles que, pour tout $i \in \{1, \ldots, n\}$, $I\!E(A_i^2) = I\!E(B_i^2)$. Alors X est un processus M-stationnaire, centré, avec : $c_X^{(2)}(\tau) = \sum_{i=1}^{N} I\!E\{A_i^2\}\cos(a_i \ln(\tau))$.
- Un processus ε est dit un M-bruit blanc si son processus dual est un bruit blanc. Un M-bruit blanc vérifie donc : $I\!E(\varepsilon(t)) = 0$ et $I\!E(\varepsilon(t)\varepsilon(t\tau)) = c\,\delta(\ln(\tau))$, où δ est le symbole de Kronecker.
- Considérons le processus X, défini pour tout $t > 0$ par : $X(t) = \int_0^t H(1/u)\varepsilon(u)d\ln(u)$, où
  – $\int_0^{+\infty} |H(t)|^2\, d\ln t > \infty$ et $H(t) = 0$ pour $0 < t < 1$.
  – ε est un M-bruit blanc.
  Donc X est un processus M-stationnaire appelé *processus M-linéaire.*

- Le processus d'Euler est un cas particulier des processus M-linéares, parfois appelé aussi processus à longue mémoire, où

$$H(t) = \begin{cases} \sum_{i=1}^{M} \sum_{j=1}^{m_i} c_{ij} (\ln t)^j t^{-a_i} & \text{pour } t \geq 1 \\ 0 & \text{pour } 0 < t < 1, \end{cases}$$

où c_{ij} et a_i sont des nombres complexes de parties réelles positives. Nous avons donc

$$c_X^{(2)}(\tau) = \sum_{i=1}^{M} \sum_{j=1}^{m_i} b_{ij} (\ln \tau)^j \tau^{-a_i},$$

où les constantes b_{ij} sont déterminées de façon unique à partir des constantes a_i et c_{ij}. L'entier $k = \sum_{i=1}^{M}(m_i + 1)$ est appelé *ordre du processus*. Notons que la place des processus d'Euler dans la théorie des processus M-stationnaires est équivalente à celle des processus autoregressifs dans la classe des processus stationnaires classiques.

Soit X un processus M-stationnaire. La densité M-spectrale ϕ_X de X est définie par la transformée de Mellin, si elle existe, de l'autocovariance $c_X^{(2)}$:

$$\phi_X(x) = \int_0^{+\infty} \tau^{-2i\pi x} c_X^{(2)}(\tau) d\ln(\tau) = \int_{-\infty}^{+\infty} e^{-2i\pi xu} c_Y^{(2)}(u) du, \tag{4.1.1}$$

Nous avons supposé tacitement l'existence de la transformée de Mellin dans (4.1.1). Nous utilisons cette remarque, également, pour la transformée inverse (Cf [166] et [167]). A partir de (4.1.1), il suit d'abord que $\phi_X(x) = \phi_Y(x)$, où ϕ_Y est la densité spectrale de Y.
Dans la suite, nous utilisons le terme densité spectrale au lieu de densité M-spectrale, et nous supposons que le processus X est centré, M-stationnaire jusqu'à l'ordre quatre.

Remarque 4.1.1. Soit X un processus M-stationnaire et Y son processus dual. Par la relation entre X et Y, nous obtenons

$$c_X^{(2)}(\tau) = c_X^{(2)}(\tau^{-1}) \text{ et } c_X^{(2)}(\tau) = c_Y^{(2)}(\ln(\tau)).$$

Si l'on suppose que $c_X^{(2)}$ est continue et que $\tau^{-\kappa} c_X^{(2)}(\tau) \in L^{1/\kappa}(\mathbb{R}_+)$ pour $\kappa = 1/2$ ou 1, alors $c_Y^{(2)}$ est continue, absolument intégrable et de carré intégrable.

Echantillonnage du temps et estimation de la densité spectrale. Pour échantillonner le temps, nous considérons le schéma $\{t_n\}_{n\in\mathbb{N}}$ (Cf chapitre 1) :

$$t_0 = 1 \text{ et } t_n = t_{n-1}\beta_n \tag{4.1.2}$$

où $\{\beta_n\}_{n\in\mathbb{N}}$ est une suite de variables aléatoires de loi de Pareto, et de densité

$$f_{\beta_n}(x) = \beta x^{-\beta-1} 1_{[1,+\infty[}(x), ; \beta > 0.$$

Considérons donc le schéma d'échantillonnage $\tau_n = \ln(t_n)$ pour tout $n \in \mathbb{N}$. La séquence $\{\tau_n\}_{n\in\mathbb{N}}$ vérifie $\tau_0 = 0$ et $\tau_n = \tau_{n-1} + \alpha_n$ pour $n \in \mathbb{N}^*$ où $\alpha_n = \ln(\beta_n)$ est une suite de variables aléatoires de loi exponentielle de paramètre $\beta > 0$, et de densité

$$f_{\alpha_n}(x) = \beta \exp(-\beta x) 1_{[0,+\infty[}(x).$$

La fonction d'autocovariance du processus échantillonné $\{X(t_n)\}_{n\in\mathbb{N}}$:

$$\rho_{X,n} = I\!E(X(t_{k+n})X(t_k)) = I\!E_{\{t_n\}} \circ I\!E_X(X(t_{k+n})X(t_k)) = \int_0^{+\infty} c_X^{(2)}(t) f_{\alpha_n}(t) dt.$$

Après quelques manipulations algébriques, on montre que la densité spectrale se décompose de la façon suivante :

$$\phi_X(\lambda) = \sum_{n=1}^{+\infty} a_n \mathcal{G}_{X,n}(\lambda) \text{ dans } L^2(\mathbb{R}),$$

où

$$\mathcal{G}_{X,n}(\lambda) = (-1)^{n-1} \frac{\sqrt{2\beta}}{\pi(\lambda^2+\beta^2)} \cos\left((2n-1)\arctan\left(\frac{\lambda}{\beta}\right)\right),\ a_n = \sum_{k=1}^{n} \theta_{n,k}\rho_{X,n} \text{ et } \theta_{n,k} = \mathcal{C}_{n-1}^{k-1}\sqrt{2/\beta}(-2)^{k-1}$$

En disposant des observations $X(t_k)$, $k = 1, \ldots, N$ pour $N > 0$, on estime $\rho_{X,n}$ par

$$\widehat{\rho_{X,n}}(N) = \begin{cases} \frac{1}{N}\sum_{k=1}^{N-n} X(t_{k+n})X(t_k) & \text{si} \quad 1 \le n < N \\ 0 & \text{sinon.} \end{cases}$$

et puis, on estime a_n par : $\widehat{a}_n(N) = \sum_{k=1}^{n} \theta_{n,k}\widehat{\rho_{X,n}}(N)$. Finalement, on estime la densité spectrale par

$$\widehat{\phi}_X(\lambda) = \sum_{n=1}^{M_N} \gamma_n(N)\widehat{a}_n(N)\mathcal{G}_{X,n}(\lambda),$$

où $\gamma_n(N) = g\left(\exp(\alpha n)/N^b\right)$, où $\alpha > \ln(3)$, $0 < b\alpha/(2\ln(3))$, et où la fonction réelle g est Lipschitz d'ordre 1.
Sous des conditions générales sur la largeur de fenêtre M_N et sur la fonction cumulante du processus X, si la fonction d'autocovariance est r-fois différentiable, et de dérivées successives absolument continues, alors, l'erreur quadratique moyenne et l'erreur quadratique moyenne intégrée de notre estimateur sont données par :

$$I\!E\left(\widehat{\phi}_X(\lambda) - \phi_X(\lambda)\right)^2 = O\left(\frac{1}{\ln(N)}\right)^{r-2} \text{ et } I\!E\left(\int_{-\infty}^{+\infty} |\widehat{\phi}_X(\lambda) - \phi_X(\lambda)|^2 d\lambda\right) = O\left(\frac{1}{\ln(N)}\right)^{r-1}.$$

Simulations. Nous considérons un processus X gaussien M-stationnaire, centré, de fonction de covariance $c_X^{(2)}(t) = \exp(-\ln(t))$ et de densité spectrale $\phi_X(\lambda) = 1/\pi(1+\lambda^2)$ (Basse-bande d'un processus de Gauss-Markov). Nous utiliserons la fenêtre spectrale de Parzen (Cf [178]). L'estimation a été faite pour de nombreuses valeurs de l'intensité moyenne β afin

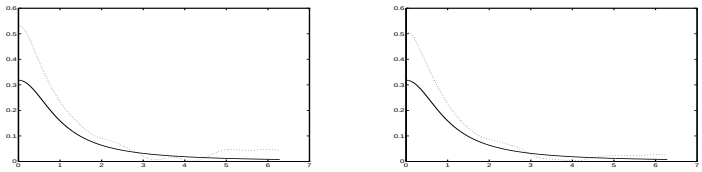

FIG. 4.1 – La densité spectrale en ligne continue et l'estimateur en pointillés pour le processus de Gauss-Markov pour $\beta = 1/\pi$: à gauche $N = 500$ et à droite $N = 1000$

de tester son influence sur la performance de l'estimateur. Pour chaque valeur de l'intensité, l'estimateur a été évalué pour plusieurs valeurs de M_N. Les instants d'échantillonnage $\{t_k\}_{k=1,...,N}$ et les observations $\{X(t_k)\}_{k=1,...,N}$ ont été générés récurcivement comme suit. Les instants $\{t_k\}_{k=1,...,N}$ sont donnés dans une forme récurcive. Un échantillon des valeurs de $\{\alpha_k\}_{k=1,...,N}$ a été obtenu par : $\alpha_k = -\ln(\theta_k)/\beta$, $k = 1, \ldots, N$, où les variables aléatoires $\{\theta_k\}_{k=1,...,N}$ suivent une loi uniforme sur $[0, 1]$. Les observations $\{X(t_k)\}_{k=1,...,N}$ sont obtenues via la représentation dynamique du processus X (Cf [77]) :

$$X(t) = \exp(-(t-\tau))X(\tau) + \sqrt{2}\int_{\tau}^{t} \exp(-(t-s))dW(s),\ t > \tau \geq 0,$$

où $\{W(t),\ t \geq 0\}$ est un processus de Wiener, centré et de covariance $c_X^{(2)}(t, s) = \min(t, s)$. Donc, on voit aisément que :

$$X(t_0) = 0 \text{ et } X(t_{k+1}) = \exp(-(t_{k+1}-t_k))X(t_k)+[1-\exp(-2(t_{k+1}-t_k))]^{1/2}\,\zeta_{k+1},\ k = 0, 1, ...$$

avec $\zeta_k \sim \mathcal{N}(0, 1)$. Les résultats de notre estimation sont présentés dans les figures FIG. 4.1. On peut remarquer que l'estimation est assez bonne, mais peut être améliorée par un choix judicieux et automatique du paramètre de lissage (Cf [194] et [196]). Les résultats des simulations indiquent que lorsque la taille de l'échantillon est assez grande l'estimation devient de plus en plus intéressante.
Cette technique n'est pas destinée uniquement aux processus M-stationnaires. Elle peut être prolongée (Cf [94]), à d'autres formes de stationnarités, où il y a une déformation du temps (Cf [183] et [184]).

4.2 Représentation spectrale des processus indexés par un semi-groupe via extension à un groupe

Les processus stochastiques de second ordre sont souvent dits (faiblement) non stationnaires s'ils ne sont pas stationnaires par rapport à la loi additive de l'espace temps. Ces processus peuvent être stationnaires par rapport à d'autres lois binaires si leurs fonctions d'autocovariance est invariante par translation (par rapport à cette loi binaire). Depuis longtemps, les opérations rapportant à une autre structure de groupe ont été étudiées dans [111]. Des

opérations rapportant à un semi-groupe avec une structure d'involution ont été étudiées dans [97], en employant l'analyse harmonique. La notion de stationarité à partir d'une déformation bijective de l'espace temps a été présentée dans [103] pour une application sur les conditions ambiantes et développée ensuite dans [184] et [183] et dans [184] et [94] pour des processus à temps continu. Ces différentes approches partagent le même but, à savoir, ramener la fonction de covariance à deux variables à une fonction d'une seule variable et obtenir une représentation spectrale pour la fonction de covariance. Par suite, par le théorème de Karhunen (Cf [128]) on obtient une représentation spectrale du processus lui-même. Dans ce paragraphe, nous supposons que l'espace temps est muni d'une structure de semi-groupe régulier (sans involution). Par conséquent, aucune des approches ci-dessus ne peut être appliquée à notre cas. Mais nous pouvons définir une relation d'équivalence sur l'espace-temps pour que l'ensemble quotient associé soit un groupe. Puis, nous montrons que n'importe quel noyau défini positif et invariant par translation sur le semi-groupe régulier, peut être prolongé à un noyau défini positif sur cet ensemble quotient. Un processus serait donc stationnaire pour la structure de semi-groupe si sa fonction de covariance (qui est un noyau défini positif) est invariante par translation. Par conséquent, la représentation spectrale sur le groupe quotient permet une représentation spectrale pour le processus. Comme application, nous définissons la stationnarité et la stationnarité multiplicative pour des processus à temps discret indexés par l'ensemble $\mathbb{N}$. Nous prolongeons ainsi la notion de stationnarité multiplicative présentée pour les processus à temps-continu dans [99]. La représentation spectrale peut être dérivée dans le cas continu par la transformée de Mellin à partir du groupe $(\mathbb{R}_+^*, \times)$ ou par la réductibilité à la stationnarité habituelle par une transformation logarithmique. L'ensemble $\mathbb{N}^*$ est seulement un semi-groupe pour la loi multiplicative, par conséquent aucune représentation directe ne peut être employée. Mais, avec la méthode présentée ici, $\mathbb{N}$ est naturellement identifié à un sous-ensemble quotient $\mathbb{Q}_+^*$ qui a une structure de groupe.

Prolongement des noyaux définis positifs :

D'un semi-groupe régulier à un groupe. Soit $(\mathcal{T}, \cdot)$ un semi-groupe, non vide et d'élément neutre e. Nous supposons que $\mathcal{T}$ est régulier : $\forall u,\ s,\ t \in \mathcal{T}$,

$$s \cdot u = t \cdot u \implies s = t. \tag{4.2.1}$$

Notons que si $(\mathcal{T}, \cdot)$ est un sous-ensemble d'un certain groupe $(\mathcal{T}, \cdot)$, alors $\mathcal{T}$ est régulier, puisque $s \cdot u = t \cdot u$ implique $s \cdot u \cdot u^{-1} = t \cdot u \cdot u^{-1}$ ou $s = t$ où u^{-1} est l'élément inverse de u dans $\mathcal{T}$.
Si (4.2.1) est satisfaite, la relation binaire $\mathcal{R}$ sur $\mathcal{T} \times \mathcal{T}$:

$$(s_1, s_2)\mathcal{R}(t_1, t_2) \Leftrightarrow s_1 \cdot t_2 = s_2 \cdot t_1$$

est une relation d'équivalence. Notons par $\mathbb{G}$ le groupe quotient $\mathbb{G} = (\mathcal{T} \times \mathcal{T})/\mathcal{R}$ d'éléments (ou classes) $\mathbf{s} = [s_1, s_2]$. Le monoïde $(\mathbb{G}, \bullet)$ est donc un groupe pour la loi $\bullet$ définie par :

$$\mathbf{s} \bullet \mathbf{t} = [s_1, s_2] \bullet [t_1, t_2] = [s_1 \cdot t_1, s_2 \cdot t_2]$$

Son élément neutre est $\mathbf{e} = [e, e]$ et comme $(s_1 \cdot s_2, s_2 \cdot s_1)\mathcal{R}(e, e)$, l'élément inverse de $\mathbf{s} = [s_1, s_2]$ est $\mathbf{s}^{-1} = [s_2, s_1]$.
Nous simplifierons les notations en remplaçant : st par $s \cdot t$ et $\mathbf{st}$ par $\mathbf{s} \bullet \mathbf{t}$.

Des noyaux invariants semi-définis positifs à des fonctions définies positives. Un noyau $r : \mathcal{T} \times \mathcal{T} \to \mathbb{C}$, défini sur un semi-groupe $(\mathcal{T}, \cdot)$ est dit invariant par translation si

$$r(su, tu) = r(s, t), \quad \text{pour tout} \quad s, t \text{ et } u \quad \text{dans} \quad \mathcal{T}.$$

Le noyau est dit défini semi-positif (ou p.d.) si :

$$\sum_{j,k=1}^{n} \alpha_j \overline{\alpha}_k r(t_j, t_k) \geq 0, \quad \text{pour tout } n \in \mathbb{N},\ \alpha_j \in \mathbb{C},\ t_j \in \mathbb{G},\ j = 1, \ldots, n.$$

Si $(\mathcal{T}, \cdot)$ est un groupe, un noyau invariant par translation prend la forme d'une fonction à une seule variable par :

$$r(s, t) = r(e, t \cdot s^{-1}) = R(t \cdot s^{-1})$$

La fonction R est dite définie positive si, et seulement si, son noyau associé r est défini positif. Nous obtenons donc le résultat suivant :
Si le semi-groupe $(\mathcal{T}, \cdot)$ est régulier, notons son extension par $(\mathbb{G}, \bullet)$,

$$\mathbf{R}(\mathbf{s}) = r(s_1, s_2) \quad \text{si} \quad \mathbf{s} = [s_1, s_2]. \tag{4.2.2}$$

et

$$\mathbf{r}(\mathbf{s}, \mathbf{t}) = \mathbf{R}(\mathbf{s} \bullet \mathbf{t}^{-1}).$$

Les fonctions $\mathbf{R}$ et $\mathbf{r}$ admettent les propriétés suivantes :

- si r satisfait (4.2.2), alors $\mathbf{R}$ (ou $\mathbf{r}$) est bien défini, c.-à-d. la valeur $\mathbf{R}(\mathbf{s})$ ne dépend pas du représentant de la classe $\mathbf{s}$.
- $\mathbf{r}$ est invariant par translation.
- si r est un noyau p.d. sur $(\mathcal{T}, \cdot)$, alors $\mathbf{r}$ est un noyau p.d. sur $(\mathbb{G}, \bullet)$.

Représentation spectrale sur les semi-groupes réguliers. Soit $X = \{X_t\}_{t \in \mathcal{T}}$ un processus stochastique défini sur un semi-groupe régulier $\mathcal{T}$, centré et de fonction de covariance

$$r(s, t) = I\!E(X_s \bar{X}_t).$$

Cette fonction est un noyau p.d. sur $\mathcal{T} \times \mathcal{T}$. Le processus est stationnaire si r satisfait la propriété de translation invariante (4.2.2). Donc, à partir de (4.2.2), il existe une fonction définie positive $\mathbf{R}$ sur $\mathbb{G}$ telle que $r(s, t) = \mathbf{R}([s, t])$.
Notons que le morphisme $s \to \mathbf{s} = [s, e]$, qui est injectif, identifie $\mathcal{T}$ comme un sous ensemble de $\mathbb{G}$ et que

$$r(s, t) = \mathbf{R}([s, e] \bullet [t, e]^{-1}).$$

Si $\mathbf{R}$ est continue, puisqu'elle est définie positive sur le groupe $\mathbb{G}$ alors le théorème de la représentation intégrale de Bochner implique (Cf [217]) :

$$R(\mathbf{s} \bullet \mathbf{t}^{-1}) = \int_{\widetilde{\mathbb{G}}} \rho(\mathbf{s})\overline{\rho}(\mathbf{t}) d\mu(\rho), \quad \mathbf{s}, \mathbf{t} \in \mathbb{G}, \tag{4.2.3}$$

pour une mesure de Radon positive μ définie sur l'ensemble $\widetilde{\mathbb{G}}$ (des caractères ρ de $\mathbb{G}$). Donc r écrite comme une fonction d'une variable et (4.2.3) conduit à

$$r(s,t) = \int_{\widetilde{\mathbb{G}}} \rho(\mathbf{s})\bar{\rho}(\mathbf{t}) d\mu(\rho), \quad s, t \in \mathcal{T}, \tag{4.2.4}$$

Ensuite, par le théorème de Karhunen (Cf [128]), une représentation spectrale du processus est possible. Précisément, il existe un processus de second ordre Z à accroissements indépendants et de base μ sur $\widetilde{\mathbb{G}}$, tel que :

$$X_t = \int_{\widetilde{\mathbb{G}}} \rho(\mathbf{t}) Z(d\rho), \quad t \in \mathcal{T}.$$

Exemples. Les exemples suivants illustrent bien l'utilité des résultats obtenus.

- *Processus stationnaires sur* $\mathbb{N}$. Soit $\mathcal{T} = \mathbb{N}$ muni de la loi additive habituelle. Soit X un processus stochastique indexé par $\mathcal{T}$ et de fonction de covariance r :

  $$r(m,n) = r(m+l, n+l), \quad m, n, l \in \mathbb{N}$$

  D'après (4.2.2), il existe une fonction définie positive $\mathbf{R}$ sur l'extension $\mathbb{G} = \mathbb{Z}$ telle que

  $$r(m,n) = \mathbf{R}(m-n), \quad m, n \in \mathbb{N}.$$

  Les caractères de $(\mathbb{Z}, +)$ sont les fonctions $\rho(t) = \exp(it\lambda)$ pour $\lambda \in [-\pi, \pi]$. Nous obtenons donc

  $$r(m,n) = \int_{-\pi}^{\pi} \exp(i\lambda(m-n)) d\mu(\lambda)$$

  et la représentation spectrale pour les processus indexés par les entiers

  $$X_n = \int_{-\pi}^{\pi} \exp(i\lambda n) Z(d\lambda), n \in \mathbb{N}.$$

- *Processus multiplicativement stationnaires sur* $\mathbb{N}^*$. De façon similaire, nous considérons le produit usuel sur $\mathcal{T} = \mathbb{N}^*$. Pour n'importe quelle fonction de covariance vérifiant

  $$r(m,n) = r(ml, nl), \quad m, n, l \in \mathbb{N}^*$$

  il existe une fonction définie positive $\mathbf{R}$ sur l'extension $\mathbb{G} = \mathbb{Q}_+$ telle que

  $$r(m,n) = \mathbf{R}(m/n), \quad m, n \in \mathbb{N}^*.$$

Les caractères du groupe $(\mathbb{Q}_+, \cdot)$ sont bien connus et ont la forme $\rho(t) = t^{i\lambda}$ pour $\lambda \in \mathbb{R}$. Donc, en identifiant le groupe des caractères $\widetilde{\mathbb{Q}}_+^*$ à $\mathbb{R}$, nous obtenons

$$r(m,n) = \int_{\mathbb{R}} (m/n)^{i\lambda}\, d\mu(\lambda).$$

Par suite, la fonction de covariance est écrite sous la forme de transformée de Mellin. Elle permet de prolonger aux processus à temps discret la majorité des résultats obtenus dans [99] concernant les processus à temps continu.

Généralement, ni la régularité du semi-groupe, ni la propriété d'invariance de la fonction de covariance ne sont nécessaires ni suffisantes pour que le processus ou la covariance admette une représentation spectrale. Par exemple, le mouvement brownien à temps dans $\mathcal{T} = [0,1]$ admet la fonction de covariance $r(s,t) = s \wedge t$. L'opération binaire $s.t = s \wedge t$ équipe $\mathcal{T}$ d'une structure de semi-groupe non régulier. Clairement, r ne vérifie pas (4.2.2). Cependant, en considérant le semi-groupe $(\mathcal{T}, \wedge)$ muni de l'identité comme involution (Cf [97]), on obtient une représentation spectrale de r et par la suite une représentation de X.

4.3 Estimation non paramétrique de la courbe de croissance moyenne pour un processus d'erreur non stationnaire

Soit le problème d'estimation non paramétrique de la courbe de croissance moyenne pour des données répétées et avec un processus d'erreur non stationnaire. Nous considérons m réplications de l'expérience, chacune comportant n mesures de la réponse :

$$Y_j(x_i) = f(x_i) + \varepsilon_j(x_i) \text{ pour } j = 1, \ldots, m \text{ et } i = 1, \ldots, n$$

où f est la courbe de croissance moyenne (inconnue) et ε est le processus d'erreur.
Notons que ce problème a été abondamment étudié, mais pour des formes particulières de la fonction d'autocovariance de ε. D'autres modèles tenant en compte les effets individuels ont été considérés dans [25] et [174], connus respectivement comme modèles additifs à deux ou trois étapes, avec des structures de covariance non stationnaires mais particulières.
Pour ce même problème, nous avons étudié l'estimation de la fonction de croissance avec un processus d'erreur stationnaire (Cf [13], pour la bibliographie). Nous nous proposons donc de s'affranchir de l'hypothèse de stationnarité. Pour ce fait, nous avons choisi de généraliser les travaux existant en ne spécifiant aucune forme particulière de la fonction d'autocovariance du processus d'erreur. Nous avons obtenu également la fenêtre optimale. Cette fenêtre dépend de la singularité de la fonction d'autocovariance sur la diagonale. A titre d'indication, cette généralisation comporte les processus d'Ornstein-Uhlenbeck, et d'autres classes spécifiques de processus non stationnaires dont la structure de covariance est paramétrique telle que celle considérée dans [84] et prolongée dans [173] pour les données non équilibrées.

Estimation de la courbe moyenne de croissance. Supposons que nous disposions des observations $\{Y_j(x_i),\ i=1,\ldots,n,\ j=1,\ldots,m\}$. Nous estimons f par :

$$\widehat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} W_{h,i}(x)\ \bar{Y}(x_i)$$

où x_i, $\bar{Y}$, $W_{h,i}$ et les points moyens $\{m_i,\ i=1,\ldots,n\}$ sont définis comme dans le paragraphe 3.2. Pour la suite nous aurons besoin des hypothèses suivantes :

(i) La fonction d'autocovariance ρ existe et est continue sur le carré-unité $[0,1]^2$.

(ii) $\rho(x,y)$ est dérivable à droite et à gauche de la diagonale $x=y$:

$$\rho^{(0,1)}(x,x^-) = \lim_{y\nearrow x}\frac{\partial\rho}{\partial y}(x,y) \text{ et } \rho^{(0,1)}(x,x^+) = \lim_{y\searrow x}\frac{\partial\rho}{\partial y}(x,y)$$

existent et sont continues.
La fonction de saut sur la diagonale $\alpha(x) = \rho^{(0,1)}(x,x^-) - \rho^{(0,1)}(x,x^+)$ est supposée continue et non identiquement nulle.

(iii) $\rho(x,y)$ est supposée admettre des dérivées partielles mixtes jusqu'à l'ordre deux en dehors de la diagonale $x\neq y$ dans le carré unité et vérifie $\sup_{0\le x\neq y\le 1}\left|\rho^{(i,j)}(x,y)\right| < \infty$ pour tout entier i, j tel que $0\le i+j\le 2$.

(iv) La fonction de covariance ρ admet au moins 2 dérivées partielles mixtes continues :

$$\rho^{(i,j)}(x,x)\neq 0,\ i+j=2,\ \forall x\in[0,1].$$

Dans ce cas, la fonction de saut α est identiquement nulle sur $[0,1]$.

Si la fonction de covariance vérifie les hypothèses (i), (ii) et (iii) et f est deux fois continûment dérivable sur $[0,1]$ avec $f''(x)\neq 0$ pour $0<x<1$, et $m/n = O(1)$, alors quand $n,m\to+\infty$

$$I\!E\left(\widehat{f}_h(x) - f(x)\right)^2 = \frac{1}{m}\left(\rho(x,x) - \frac{1}{2}\alpha(x)C_K h\right) + \frac{h^4}{4}d_K^2(f''(x))^2 + O\left(\frac{1}{mn}+\frac{h^2}{n}\right) + o\left(h^4+\frac{h}{m}\right)$$

où $C_K = \int_{-1}^1\int_{-1}^1 |u-v|K(u)K(v)du\,dv = 2\int_{-1}^1\int_u^1 (v-u)K(u)K(v)du\,dv$ et $d_K = \int u^2K(u)\,du$.
L'erreur quadratique moyenne est minimale en prenant la largeur de fenêtre locale suivante :

$$h_x^* = (\alpha(x)C_K)^{1/3}\left(2d_K^2(f''(x))^2\right)^{-1/3} m^{-1/3}$$

La largeur de fenêtre globale asymptotiquement optimale peut être obtenue en utilisant une mesure d'erreur d'estimation globale tel que l'erreur quadratique moyenne intégrée :

$$h^* = \left(C_K\int_0^1\alpha(x)dx\right)^{1/3}\left(2d_K^2\int_0^1(f''(x))^2dx\right)^{-1/3} m^{-1/3}.$$

Les processus d'erreur suivants vérifient les hypothèses (i), (ii) et (iii). Une classe générale de fonctions d'autocovariances avec des fonctions de saut constantes peut aussi être trouvée.

Exemple 4.3.1.

1. Soit ε le processus d'erreur de Wiener d'autocovariance $\rho(x,y) = \sigma^2 \min(x,y)$. La fonction de saut $\alpha(x) = \sigma^2 > 0$ et $\rho^{(i,j)}(x,y) = 0$, pour tout entier i, j tel que $i+j = 2$ et $x \neq y$. Dans ce cas $h_x^* = \left(\sigma^2 C_K\right)^{1/3} \left(2d_K^2 (f''(x))^2\right)^{-1/3} m^{-1/3}$.
2. Si le processus d'erreur ε est d'Uhlenbeck-Ornstein d'autocovariance stationnaire $\rho(x,y) = \sigma^2 \exp(-\lambda|x-y|/2)$ pour $\sigma > 0$ et $\lambda > 0$. La fonction de saut est constante $\alpha(x) = 2\sigma^2\lambda$.
   Pour un processus d'erreur stationnaire d'autocovariance $\rho(x,y) = \rho(x-y)$ tel que le processus d'Uhlenbeck-Ornstein, la fonction de saut : $\alpha(x) = \rho^{(1)}(0^-) - \rho^{(1)}(0^+) = 2\rho^{(1)}(0^-)$ est constante. Alors $h_x^* = \left(\rho^{(1)}(0^-) C_K d_K^{-2} (f''(x))^{-2}\right)^{1/3} m^{-1/3}$, ce qui correspond à la largeur de fenêtre donnée dans [119].
3. Considérons une transformation de l'échelle de temps, qui peut produire une autocovariance non stationnaire de la forme :

   $$\rho(x,y) = \sigma^2 \rho^{|x^\lambda - y^\lambda|/\lambda} \text{ pour } (x,y) \in [0,1]^2,\ \sigma^2 > 0,\ 0 < \rho < 1 \text{ et } \lambda > 0$$

   (Cf [84] et [174]).
   En particulier quand $\lambda = 1$, nous obtenons un processus d'erreur d'Uhlenbeck-Ornstein de fonction d'autocovariance stationnaire avec $\alpha(x) = -2\sigma^2 \ln(\rho)$. Par contre, pour $\lambda \neq 1$, la fonction d'autocovariance est non stationnaire et la fonction de saut est $\alpha(x) = -2\sigma^2 \ln(\rho) x^{\lambda-1}$. Dans ce cas $h_x^* = \left(-x^{\lambda-1} \ln(\rho) C_K \sigma^2 d_K^{-2} (f''(x))^{-2}\right)^{1/3} m^{-1/3}$, ce qui correspond à la largeur de fenêtre obtenue dans [84].

Nous avons obtenu également l'expression asymptotique de l'erreur quadratique moyenne pour des processus d'erreur plus lisses. En effet, si l'on suppose que ρ vérifie l'hypothèse (iv) et que f est deux fois continûment dérivable sur $[0,1]$ avec $f''(x) \neq 0$, pour $0 < x < 1$, alors quand n, $m \to +\infty$ et $m/n = O(1)$:

$$I\!E\left(\widehat{f}_h(x) - f(x)\right)^2 = \frac{1}{m}\left(\rho(x,x) + \rho^{(0,2)}(x,x)\, d_K h^2\right) + \frac{h^4}{4} d_K^2 (f''(x))^2 + O\left(\frac{1}{mn} + \frac{h^2}{n}\right) + o\left(h^4 + \frac{h^2}{m}\right)$$

La largeur de fenêtre asymptotiquement optimale est donnée par :

$$h_x^* = \left(-2\rho^{(0,2)}(x,x) d_K^{-1} (f''(x))^{-2}\right)^{1/2} m^{-1/2}$$

En particulier, si le processus d'erreur est stationnaire, tel que $\rho^{(0,2)}(x,x) = \rho''(0)$ alors $h_x^* = (-2\rho''(0))^{1/2} \left(d_K (f''(x))^2\right)^{-1/2} m^{-1/2}$.
La largeur de fenêtre globale asymptotiquement optimale est :

$$h^* = \left(2\int_0^1 \rho^{(0,2)}(x,x)dx\right)^{1/2} \left(d_K \int_0^1 (f''(x))^2 dx\right)^{-1/2} m^{-1/2}.$$

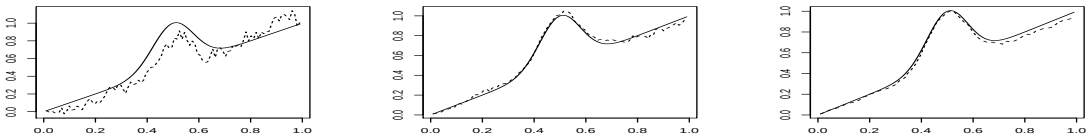

FIG. 4.2 – La courbe de croissance moyenne f en ligne continue, l'estimateur $\widehat{f}_h$ en ligne discontinue pour $n = 100$, $m \in \{1, 10, 50\}$ avec 2 comme variance du processus d'erreur

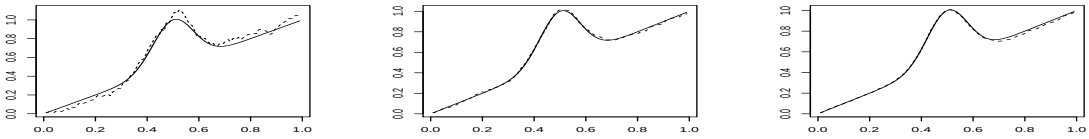

FIG. 4.3 – La courbe de croissance moyenne f en ligne continue, l'estimateur $\widehat{f}_h$ en ligne discontinue pour $n = 100$, $m \in \{1, 10, 50\}$ avec processus d'erreur brownien fractionnaire

Simulations. Considérons le processus d'erreur admettant la fonction d'autocovariance statonnaire :

$$\rho(x, y) = (1 + \alpha|x - y|)\, \exp(-\alpha|x - y|),\ \alpha > 0$$

et la densité spectrale $\varphi(\lambda) = 2\alpha^3/(\pi(\alpha^2 + \lambda^2)^2)$. Cette fonction d'autocovariance vérifie l'hypothèse (iv) avec : $\rho''(0) = -\alpha^2 \neq 0$ et $h_x^* = \left(2\alpha^2 d_K^{-1}(f''(x))^{-2}\right)^{1/2} m^{-1/2}$.
La fonction de croissance considérée ici est celle étudiée dans [84] :

$$f(x) = x - 0.5\exp(-80(x - 0.5)^2).$$

Les graphiques FIG. 4.2 et FIG. 4.3 illustrent bien la performance de l'estimateur $\widehat{f}_h$ suivant le nombre de réplications m.

Chapitre 5

Estimation de la régression pour des données fonctionnelles

Les données fonctionnelles ont fait leur apparition dans plusieurs domaines de la statistique appliquée (médecine, environnement, ...). Il est donc de plus en plus fréquent de travailler avec ce type de données. D'un point de vue technique, un échantillon de données fonctionnelles peut être rencontré dans de nombreux problèmes statistiques (classification, discrimination, études longitudinales, prévision, ...). Ainsi, c'est devenu un vrai défi pour les statisticiens de construire des procédures statistiques permettant de traiter ce type de données. Ce champ de la statistique moderne est devenu populaire grâce aux livres [22], [83], [209] et [210], et est généralement connu sous le nom d'*Analyse Statistique des Données Fonctionnelles*. Le lecteur peut trouver dans [82] une vue d'ensemble sur des problématiques et des avancées récentes liées à ce domaine important de la statistique moderne.
Dans ce chapitre, nous nous intéressons à des problèmes spécifiques faisant intervenir les données fonctionnelles : la prévision d'une variable réponse scalaire Y étant donnée une certaine variable fonctionnelle X. En d'autres termes, la question est d'estimer l'opérateur de régression $r(.) = I\!E(Y|X = .)$ quand X est une variable aléatoire à valeurs dans un espace de dimension éventuellement finie. L'estimation de l'opérateur r a été traitée par plusieurs auteurs durant la dernière décennie, et la plupart de ces travaux concernaient les modèles linéaires (r est linéaire). D'ailleurs, ce domaine de la statistique qui est connu sous le nom de *Modèles Linéaires Fonctionnels* est devenu populaire grâce à l'ouvrage [209] (Cf, [42] pour les plus récentes avancées et une revue bibliographique sur ce sujet). Comme dans le cas non fonctionnel c.-à-d. lorsque la variable explicative est réelle multivariée (mais de dimension finie), il y a une réelle difficulté pour s'affranchir de l'hypothèse de linéarité de l'opérateur r. De plus, ce problème est particulièrement important dans le cadre fonctionnel, car il est souvent impossible de disposer d'outils graphiques permettant de visionner l'exactitude (ou la pertinence) de ce type d'hypothèses. Dans [82], on a montré qu'une approche non

paramétrique du problème est envisageable, en introduisant des estimateurs à noyau *opératoriels* et en exigeant qu'une condition de régularité (condition de Hölder) sur l'opérateur r. Par analogie avec la terminologie utilisée dans le cadre non fonctionnel (dimension finie), ce champ de la statistique est actuellement appelé *Estimation Opératorielle* ou *Estimation Non paramétrique pour Données Fonctionnelles*. Pour d'autres références bibliographiques, le lecteur peut consulter les travaux : [11], [19], [23], [34], [76], [80], [107], [109], [121], [127], [134], [135], [170], [208] et [211], entre autres.
En plus, que ces estimateurs opératoriels admettent de bonnes propriétés asymptotiques, la réduction de dimension est contrôlée par le biais de considérations convenables sur les probabilités des petites boules. Le lecteur trouvera dans la monographie [83], une longue discussion sur les méthodes non paramétriques relatives aux données fonctionnelles. Mais, comme c'est le cas avec plusieurs estimateurs non paramétriques, que ce soit dans le cas non fonctionnel ou dans le cas fonctionnel considéré ici, il existe un paramètre de lissage qui intervient dans la construction des estimateurs et qui doit être sélectionné convenablement afin d'assurer de bonnes performances dans la pratique (Cf [81], pour une application sur des données réelles en chimie). Notre approche pour le choix du paramètre de lissage sera basée sur la procédure de validation croisée, qui est connue dans la littérature, et qui donne des réponses très satisfaisantes à ce type de problématique, (Cf [116] pour le cas non fonctionnel et [25], [47], [120] et [211] pour le traitement des données longitudinales).
Le but principal du paragraphe 5.1.1 de ce chapitre est de présenter une procédure de choix global (GCV) et automatique du paramètre de lissage, et d'établir ensuite, son optimalité asymptotique dans un sens quadratique. Au vu du travail technique qu'a nécessité la démonstration de ce résultat, nous nous sommes vus obligés de regarder le lien entre les différentes mesures quadratiques. Les résultats de cette investigation sont résumés dans le paragraphe 5.1.2. Dans ce paragraphe, nous avons établi des extensions fonctionnelles de résultats obtenus dans le cas de dimension finie (Cf [112], [133] et [147]). De plus, nous pensons que ces résultats ne sont pas seulement important pour la résolution du problème de choix du paramètre de lissage, mais ils seront aussi, utiles pour d'autres développements futurs.
Par ailleurs, il était naturel de penser à une version locale de la procédure de validation croisée fonctionnelle. Nous avons donc introduit et étudié dans le paragraphe 5.1.3, une version locale (LCV) de la méthode de validation croisée. Par suite, pour un souci de praticité de ces procédures, nous avons étudié les résultats théoriques obtenus, d'abord sur des données simulées (Cf paragraphe 5.1.4), et ensuite, sur des données réelles (Cf paragraphe 5.1.5). Aussi, nous avons comparé le LCV et le GCV en fonction du choix de la semi-norme d'une part, et du rapport signal-bruit, d'autre part.
Une fois réglés ces problèmes de choix de la largeur de fenêtre, nous nous sommes intéressés au problème de dépendance entre les données. Plus spécialement, au processus d'erreur du modèle de régression fonctionnelle. En effet, dans le paragraphe 5.2, nous avons supposé que le processus d'erreur est à longue mémoire. En fait, ces processus sont d'un grand intérêt pour la modélisation de la structure de corrélation des processus, pour lesquels la dépendance faible habituelle (mélanges, ...) n'est pas satisfaite. De plus, dans plusieurs applications statistiques, la corrélation entre les observations décroît plus lentement vers zéro

pour les processus à longue mémoire que pour les modèles classiques (les modèles ARMA, par exemple). Par ailleurs, les modèles à longue mémoire ont été introduits dans [146] pour les mouvements browniens fractionnaires, et plusieurs développements dans la théorie des processus ont suivi (Cf [26], [27], [98] et [125]). Une bonne introduction sur les processus dépendants est donnée dans [15] et [16], entre autres. Pour une discussion plus détaillée sur les séries temporelles à longue mémoire et leurs applications, on peut se référer par exemple à [123] et à [164] pour les modèles économétriques et à [227] pour les études environnementales ou climatiques.
Fréquemment, on dit qu'un processus stationnaire $\varepsilon = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ est à longue mémoire si, pour $\gamma \in]0,1]$, et $\mathcal{C} > 0$, sa fonction d'autocorrélation vérifie :

$$|f_\varepsilon(j)| \sim \mathcal{C}\,|j|^{-\gamma} \text{ quand } |j| \to +\infty \tag{5.0.1}$$

D'autres définitions équivalentes à celle-ci introduisant la densité spectrale $\phi_X(\lambda)$ peuvent être utilisées, c.-à-d. pour $\gamma \in]0,1]$ et $\mathcal{C}(\phi_X) > 0$, ϕ_X vérifie :

$$\phi_X(\lambda) \sim \mathcal{C}(\phi_X)\,\lambda^{-1-\gamma} \text{ quand } \lambda \to 0 \text{ (Cf [50])}.$$

Dans le cas de dimension finie, le problème d'estimation de la régression quand le processus d'erreur est à longue mémoire, a été largement étudié (Cf [51], [93], [106], [155] et [212], entre autres). Dans le cas fonctionnel (dimension éventuellement finie), le problème d'estimation de l'opérateur de régression r pour des données mélangeantes a été étudié dans [82]. Dans [82], on a établi les vitesses de convergence presque complète pour une version fonctionnelle de l'estimateur à noyau. Par ailleurs, dans [1], on a utilisé une méthodologie par ré-échantillonnage pour estimer un opérateur linéaire pour données et réponses fonctionnelles. Récemment, dans [156], on a prouvé la normalité asymptotique de l'estimateur à noyau pour des données fonctionnelles fortement mélangeantes.
Dans le paragraphe 5.2, nous avons étudié la convergence forte (ponctuelle et uniforme) de l'estimateur à noyau de l'opérateur r quand le processus d'erreur est à longue mémoire. Le fait que cette étude soit spécifique au processus d'erreur à longue mémoire, nous a incité à considérer, dans le paragraphe 5.3, le problème d'estimation de l'opérateur de régression r quand les erreurs sont corrélées. En fait, dans le paragraphe 5.3, nous nous sommes intéressés à l'estimation de r quand la variable explicative est fonctionnelle-déterministe (fixed-design). Nous avons donc étudié la performance de l'estimateur à noyau de r en termes de la convergence en moyenne quadratique et de la convergence uniforme presque complète. En particulier, nous avons considéré dans ce contexte le cas de processus d'erreur à longue mémoire. Finalement, comme c'est le cas tout au long de ce mémoire, une étude par simulations a été conduite afin de confirmer l'aspect applicatif de nos résultats théoriques pour plusieurs types de variables explicatives fonctionnelles (aléatoires ou déterministes) et plusieurs processus d'erreurs.

Préliminaires. Soit (E,d) un espace semi-métrique de dimension éventuellement finie. Soit $(X_1,Y_1),\ldots,(X_n,Y_n)$ un échantillon de réalisations de la variable aléatoire (X,Y) à valeurs dans $E\times\mathbb{R}$. Suivant [82], le problème d'estimation de l'opérateur fonctionnel r peut

être traité par les techniques d'estimation à noyau. Précisément, on a introduit l'estimateur opératoriel suivant :

$$\widehat{r}_h(x) = \frac{\sum_{i=1}^n Y_i K\left(h^{-1} d(x, X_i)\right)}{\sum_{i=1}^n K\left(h^{-1} d(x, X_i)\right)} \tag{5.0.2}$$

où K est une fonction réelle définie sur $\mathbb{R}_+$ et $h = h(n) \in \mathbb{R}_+$ est le paramètre de lissage (largeur de fenêtre) : $\lim_{n\to+\infty} h = 0$.
Nous utiliserons la notation suivante :

$$\widehat{r}_h(x) = \frac{\widehat{r}_{1,h}(x)}{\widehat{r}_{2,h}(x)}$$

où

$$\widehat{r}_{2,h}(x) = \frac{1}{n\, I\!E(\Delta_1(x))} \sum_{i=1}^n \Delta_i(x) \text{ avec } \Delta_i(x) = K\left(\frac{d(x, X_i)}{h_n}\right)$$

et

$$\widehat{r}_{1,h}(x) = \frac{1}{n\, I\!E(\Delta_1(x))} \sum_{i=1}^n \Gamma_i(x) \text{ avec } \Gamma_i(x) = Y_i\, \Delta_i(x)$$

Récemment, les propriétés asymptotiques de cet estimateur ont été étudiées dans plusieurs articles. Précisément, dans [82], on a établi sa convergence presque complète, puis dans [156] on a établi sa normalité asymptotique et dans [79] on a obtenu la forme asymptotique de son erreur quadratique moyenne. Dans tous ces articles, il est montré que le paramètre de lissage joue un rôle crucial dans la vitesse de convergence de $\widehat{r}_h$. Ce fait n'est point surprenant puisqu'il apparaît dans tous les problèmes de lissage non paramétriques (pas nécessairement fonctionnel).
Pour toute la suite de ce chapitre, nous présentons les hypothèses auquelles nous ferons, éventuellement, appel dans l'établissement de nos résultats.

Hypothèses

Fonction de poids. Nous supposons que W est une fonction positive bornée et à support dans un compact $\mathcal{S}$ d'intérieur non vide, et tel que

$$\mathcal{S} \subset \cup_{l=1}^{l_n} \mathcal{B}_l \text{ où } \mathcal{B}_l,\ l = 1, \ldots, l_n \text{ sont des boules dans l'espace } E, \quad \text{de même rayon } r_n \text{ avec } l_n = r_n^{-\zeta},\ \zeta > 0. \tag{5.0.3}$$

Fonction de poids locale. Nous supposons que, $\forall x \in E$, la fonction de poids :

$$W_x \text{ est une fonction positive bornée et à support compact dans } \mathcal{B}(x, w) \tag{5.0.4}$$

où $\mathcal{B}(x, w)$ désigne la boule fermée de centre la courbe x et de rayon $w = h^\gamma$, $0 < \gamma < 1$.

Largeur de fenêtre. Nous supposons que H_n est un ensemble fini de paramètres h ($h \to 0$), tel que :

$$\text{il existe } \tau_0 > 0, \text{ tel que } card(H_n) = O\left(n^{\tau_0}\right) \tag{5.0.5}$$

Le noyau. Nous supposons que le noyau K est de type II (Cf les détails dans [83]), c.-à-d. K est à support compact inclus dans $[0,1]$, dérivable sur $(0,1)$ et tel que :

$$\text{Il existe } -\infty < C_2 < C_1 < 0 \text{ tel que } C_2 \leq K' \leq C_1. \tag{5.0.6}$$

ou

$$\begin{aligned} &K \text{ est strictement décroissant sur } (0,1) \text{ et } \exists C_3,\, C_4 > 0 \text{ tel que} \\ &C_3\, 1_{(0,1)}(z) \leq K(z) \leq C_4\, 1_{(0,1)}(z) \text{ pour tout } z \in \mathbb{R}. \end{aligned} \tag{5.0.7}$$

Concentration de X. Nous supposons que la loi de probabilité de $d(x,X)$ peut être développée autour de 0 comme suit :

$$\mathbb{P}\left(d(x,X_i) \leq s\right) = \mathcal{C}_x\, \varphi(s) + o(\varphi(s)),\ \text{pour tout } x \in \mathcal{S}. \tag{5.0.8}$$

et satisfait la condition : $\sup_{x\in\mathcal{S}} \mathcal{C}_x < \infty,\ \varphi(s) > 0, \forall s$, et

$$\lim_{t\to 0} \varphi(t) = 0 \text{ et } \lim_{n\to+\infty} n\,\varphi(h) = +\infty \tag{5.0.9}$$

et pour tout $h \in H_n$:

$$\begin{aligned} &\exists \tau > 0 \text{ tel que } \phi(h) = O(n^{-\tau}) \text{ et } \forall x \in E,\, \forall s \in [0,1], \exists\, l_{0,x}(s) > 0 \\ &\text{tel que} \lim_{t\to 0} \frac{\varphi(st)}{\varphi(s)} = l_{0,x}(s). \end{aligned} \tag{5.0.10}$$

Concentration de (X_i, X_j) : nous supposons que la loi de probabilité du couple (X_i, X_j) peut être écrite sous la forme :

$$C_5\, \psi(h) \leq \mathbb{P}\left((X_i,X_j) \in \mathcal{B}(x,h) \times \mathcal{B}(x,h)\right) \leq \ C_6\, \psi(h),\ \text{pour } x \in E \tag{5.0.11}$$

pour toute fonction positive ψ telle que $\psi(h)/\varphi(h)^2$ est bornée, et C_5 et C_6 sont des constantes positives.

Opérateur de régression. Le modèle statistique pour l'opérateur de régression fonctionnel r est non paramétrique en ce sens que nous supposons seulement l'hypothèse de régularité suivante :

$$\exists\, 0 < \mathcal{C} < \infty,\, \exists\, \beta > 0,\ \text{tel que } \forall x, y \in E\ :\ |r(x) - r(y)| \leq \mathcal{C}\, d(x,y)^{\beta}. \tag{5.0.12}$$

Moments conditionnels. Nous supposons l'hypothèse de bornage habituelle suivante :

$$\forall k \in \mathbb{N}^*, \exists \mathcal{C}_k\ > 0 \text{ tel que } \mathbb{E}\left(|Y|^k | X = x\right) \leq \mathcal{C}_k\ \text{pour tout } x \in E. \tag{5.0.13}$$

$$\begin{aligned} &\exists \sigma > 0 \text{ tel que } \mathbb{E}\left(Y^2 | X = x\right) = \sigma(x) \geq \sigma > 0 \text{ pour tout } x \in E \\ &\text{et } \sigma(x) \text{ est continue.} \end{aligned} \tag{5.0.14}$$

Condition de mélange : nous supposons que les paires (Y_i, Y_j) sont mélangeantes, pour lesquelles le coefficient de mélange vérifie :

$$\exists\, q > 2,\ \alpha(n) \leq C_7\, n^{-q},\ \text{où } C_7 \text{ est une constante positive.} \tag{5.0.15}$$

Remarquons que notre jeu d'hypothèses est relativement peu restrictif, puisque les conditions sur Y, W, K et H_n sont les mêmes qu'en dimension finie (Cf pour plus de détails [116])). Ces conditions sur le noyau K sont utilisées pour simplifier la présentation des démonstrations, mais des extensions au cas de noyaux continus peuvent être obtenues facilement en suivant les mêmes idées que dans le Lemme 4.4 dans [83].
La seule hypothèse qui mérite une discussion est la condition de concentration (5.0.8). Cette condition est classiquement utilisée pour contrôler la concentration de la loi de la variable X sur les petites boules d'une part, et le fléau de dimension qui peut se produire dans notre cadre (dimension infinie) d'autre part. Ce n'est pas notre objectif de développer ces points ici, mais le lecteur peut trouver dans [83] comment la fonction $\varphi(.)$ admet une influence sur les vitesses de convergence des estimateurs et, plus particulièrement, comment le choix d'une bonne semi-métrique peut réduire cette influence de dimension. Dans [83], il est expliqué, également, comment la récente théorie des probabilités des petites boules et des processus stochastiques de dimension infinie permet de voir que la condition (5.0.8) est satisfaite (avec des expressions précises de la concentration $\varphi(.)$) par une large classe de processus stochastiques, comme les processus de diffusion (incluant les processus d'Ornstein-Uhlenbeck standards) et plusieurs autres processus gaussiens (incluant les mouvements browniens fractionnaires standards).

5.1 Choix de la largeur de fenêtre

Dans cette section, nous allons présenter, d'abord, une version globale, puis locale du critère de choix du paramètre de lissage et comparer, ensuite, les deux critères : local (LCV) et global (GCV), sur des données simulées, puis sur des données réelles.

5.1.1 Critère de choix global de la largeur de fenêtre

Rappelons l'erreur quadratique moyenne et l'erreur quadratique moyenne integrée :

$$ASE(h) = n^{-1} \sum_{j=1}^{n} \left(\widehat{r}_h(X_j) - r(X_j)\right)^2 W(X_j) \text{ et } MISE(h) = \int \mathbb{E}\left(\widehat{r}_h(x) - r(x)\right)^2 W(x)\, dF(x),$$

où W est une fonction de poids positive (connue) et dF désigne la mesure de répartition. En s'inspirant de [116], nous introduisons le critère :

$$GCV(h) = n^{-1} \sum_{j=1}^{n} \left(Y_j - \widehat{r}_h^j(X_j)\right)^2 W(X_j) \tag{5.1.1}$$

où $\widehat{r}_h^j(x)$ est l'estimateur *validé croisé* de $r(x)$ donné par :

$$\widehat{r}_h^j(x) = \frac{\sum_{\substack{i=1\\ i\neq j}}^{n} Y_i K\left(h^{-1}D(x,X_i)\right)}{\sum_{\substack{i=1\\ i\neq j}}^{n} K\left(h^{-1}D(x,X_i)\right)} \tag{5.1.2}$$

Le critère GCV est connu sous le nom de *critère de validation croisée* et son minimiseur $h_0 = \arg\min_{h\in H_n} GCV(h)$ est appelé *la largeur de fenêtre validée croisée*. Nous montrons donc que, sous les hypothèses (5.0.3), (5.0.5), (5.0.6), (5.0.8), (5.0.10), (5.0.12), (5.0.13), (5.0.14), le critère de sélection du paramètre de lissage qui consiste à choisir h_0 dans un ensemble d'intérêt $H_n \subset \mathbb{R}_+$, minimisant $GCV(h)$ est asymptotiquement optimal relativement aux distances $\widetilde{d} = ASE$ ou $MISE$, au sens suivant :

$$\frac{\widetilde{d}(\widehat{r}_{h_0}, r)}{\inf\limits_{h\in H_n} \widetilde{d}(\widehat{r}_h, r)} \to 1, \text{ presque sûrement (p.s.), quand } n \to +\infty$$

5.1.2 Equivalence asymptotique entre les mesures quadratiques

Dans ce paragraphe, nous établissons trois résultats concernant l'équivalence asymptotique entre plusieurs mesures de risque. Ces résultats jouent un rôle important, d'abord dans la preuve de notre principal résultat, et puis nous assurent l'équivalence entre les mesures de risque. Nous pensons que ces résultats permettront des développements futurs dans ce domaine.
Sous les hypothèses (5.0.3), (5.0.5), (5.0.6), (5.0.8), (5.0.10), (5.0.12), (5.0.13) et (5.0.14), nous avons

$$\sup_{h\in H_n} \left| \frac{MISE(h) - ISE(h)}{MISE(h)} \right| \to 0, \text{ p.co., quand } n \to \infty. \tag{5.1.3}$$

et

$$\sup_{h\in H_n} \left| \frac{ASE(h) - ISE(h)}{ISE(h)} \right| \to 0, \text{ p.co., quand } n \to \infty. \tag{5.1.4}$$

Si de plus une des trois conditions suivantes est satisfaite : nh^k est bornée, pour $k \geq 2$, ou $\phi(h)n^{1/2}h^{2\beta} < 1$, ou $h \leq \mathcal{C}\, n^{-\kappa}$ pour $\kappa \geq 1/4\beta$, alors, on a

$$\sup_{h\in H_n} \left| \frac{MISE(h) - ASE(h)}{MISE(h)} \right| \to 0, \text{ p.co., quand } n \to \infty. \tag{5.1.5}$$

Remarque 5.1.1. De (5.1.3), (5.1.4) et (5.1.5), on peut établir des résultats généraux semblables à ceux obtenus dans le cas réel (Cf [231]) :

$$\sup_{h\in H_n} \left| \frac{\widetilde{d}(\widehat{r}_h(x), r(x)) - \widetilde{d}(\widehat{r}_h(x), r(x))}{\widetilde{d}(\widehat{r}_h(x), r(x))} \right| \to 0, \ p.s. \text{ quand } n \to \infty.$$

où $\widetilde{d} \in \{ASE, ISE, MISE\}$.

5.1.3 Critère de choix local de la largeur de fenêtre

Il est bien connu, en dimension finie, qu'autour d'un pic de l'opérateur de régression, le biais de l'estimateur $\widehat{r}_h$ est particulièrement significatif. C'est pour cette raison, que dans un tel cas, il serait meilleur de choisir une petite largeur de fenêtre afin de réduire ce biais. Autrement dit, il faut trouver une valeur assez grande de la largeur de fenêtre qui permet de réduire la variance de $\widehat{r}_h$ sans augmenter son biais. Il est aussi clair, en dimension infinie, que la concentration de la loi de la variable explicative X aura une influence sur le choix d'un paramètre de lissage approprié (variance de l'estimateur augmente quand la concentration de la loi de X décroît, et qui est aussi le cas quand h décroît (Cf condition (5.0.8)). De plus, dans les zones où X admet une petite concentration de sa loi, la largeur de fenêtre devrait être prise suffisamment grande afin de permettre d'inclure un grand nombre de courbes (données) dans l'estimation. Tel argument est développé pour les estimateurs splines dans [223]. Dans des travaux comme [114] et [169], on a traité le problème de localisation adaptative du paramètre de lissage, en utilisant des techniques assez différentes de la notre, mais ceci concerne spécialement l'estimation de la régression pour des données déterministes (fixed-design). Dans [230], on a proposé une procédure de choix local du paramètre de lissage dans le cas général, mais en dimension finie, ce qui a inspiré ce travail.
Il faut noter que le problème de sélection du paramètre de lissage, dans le cas de dimension infinie, est plus compliqué qu'en dimension finie. En particulier, on n'a pas à disposition le *nuage de points* comme outil graphique pour explorer la relation entre les variables explicatives et la variabe réponse scalaire. De ce fait, il est très difficile d'avoir une idée sur la relation entre les dites variables. C'est pourquoi, pour une telle relation, plusieurs domaines de l'espace avec plusieurs (basse/haute) concentrations peuvent apparaître, alors qu'ils n'apparaîssent pas dans l'échantillon de données fonctionnelles (voir, par exemple, les courbes simulées dans la section 5.1.4).
La méthode de sélection du paramètre de lissage h relativement à la courbe x consiste dans un choix de h_x qui minimise, sur un sous-ensemble particulier H_n de paramètres de lissage (à spécifier), le critère de validation croisée suivant :

$$LCV_x(h) = n^{-1} \sum_{j=1}^{n} (Y_j - \widehat{r}_h^j(X_j))^2 \; W_x(X_j), \tag{5.1.6}$$

où $\widehat{r}_h^j$ est l'estimateur validé croisé de r, défini par (5.1.2).
La largeur de fenêtre correspondante à la courbe $x \in E$ est donc obtenue par :

$$h_x = \arg \min_{h \in H_n} LCV_x(h),$$

Avant d'énoncer notre résultat d'optimalité asymptotique de h_x, il est important de noter que la seule condition qui va changer par rapport au GCV est celle concernant la fonction de poids W_x. Cette condition est similaire à celle considérée dans [230]. Elle permet d'avoir un nombre d'observations maximum au voisinage de la courbe x. Par ailleurs, les conditions sur les moments, sur la régularité de l'opérateur de régression r et sur la concentration de

X restent les mêmes.

Nous avons donc le résultat d'optimalité asymptotique suivant. Sous les hypothèses (5.0.4), (5.0.5), (5.0.6), (5.0.8), (5.0.10), (5.0.12), (5.0.13) et (5.0.14), la largeur de fenêtre h_x qui minimise le critère de choix local de validation-croisée LCV_x est asymptotiquement optimal : pour tout $\widetilde{d} \in \{\widetilde{d}_{A,x}, \widetilde{d}_{I,x}, \widetilde{d}_x\}$, nous avons presque sûrement :

$$\frac{d(h_x)}{\inf\limits_{h\in H_n}(d(h))} \to 1 \text{ quand } n \to \infty$$

où $\widetilde{d}_{A,x}$, $\widetilde{d}_{I,x}$, $\widetilde{d}_x$ désignent des versions locales de ASE, ISE et MSE (obtenues en remplaçant W par W_x).
Le cas particulier où $W_x(z) = W(z)$, $\forall x \in E$ et $\forall n \in \mathbb{N}$, conduit au critère de choix global de la largeur de fenêtre GCV défini par (5.1.1).

5.1.4 Simulations

Nous nous proposons d'explorer plusieurs situations susceptibles d'être rencontrées dans le cadre fonctionnel. Ceci, nous permettra de montrer, sur des données simulées, qu'une largeur de fenêtre locale permet d'obtenir de meilleurs résultats qu'une largeur de fenêtre globale. Pour ceci, nous simulons trois types de courbes :

Simulation 1 $\longrightarrow$ courbes "lisses",
Simulation 2 $\longrightarrow$ courbes "bruitées",
Simulation 3 $\longrightarrow$ courbes "moches" ou irrégulières.

Dans chacune de ces trois situations, nous considérons un modèle de régression fonctionnelle non paramétrique hétéroscédastique :

$$Y_i = r(X_i) + \sigma_{1,\varepsilon}\varepsilon_i,\ i \in \mathcal{S}_1 = \{1, \ldots, 200\},$$

et

$$Y_i = r(X_i) + \sigma_{2,\varepsilon}\varepsilon_i,\ i \in \mathcal{S}_2 = \{201, \ldots, 400\}, \tag{5.1.7}$$

où les erreurs ε_i sont des variables aléatoires indépendantes suivant une loi $\mathcal{N}(0,1)$, et $\sigma_{1,\varepsilon}$ et $\sigma_{2,\varepsilon}$ sont contrôlées en considérant plusieurs valeurs du rapport signal/bruit (SNR) :

$$\frac{\sigma^2_{1,\varepsilon}}{\frac{1}{200}\sum_{i\in\mathcal{S}_1}(r(X_i) - \overline{r(X)}_{\mathcal{S}_1})^2} = \frac{\sigma^2_{2,\varepsilon}}{\frac{1}{200}\sum_{i\in\mathcal{S}_2}(r(X_i) - \overline{r(X)}_{\mathcal{S}_2})^2} = 5\%;\ 10\%;\ 20\%;\ 30\%,$$

où $\overline{r(X)}_{\mathcal{S}_k}$ désigne la moyenne des $r(X_i)$ sur $\mathcal{S}_k$ pour $k = 1, 2$. Comme nous allons le remarquer plus tard, l'hétéroscédasticité constitue un bon exemple d'étude de la structure fonctionnelle. Ceci n'est pas visible sur les tracés des courbes. De plus, ce type de données simulées montre clairement l'avantage d'utiliser la sélection d'une largeur de fenêtre locale plutôt que globale.
Dans la suite de ce paragraphe, nous noterons par r^{LCV} (resp. par r^{GCV}) l'estimateur à noyau de l'opérateur r calculé avec la largeur de fenêtre locale (resp. globale) sélectionnée par LCV (resp. par GCV).

Courbes lisses. Le premier échantillon de courbes simulées consiste en la construction de deux ensembles de courbes lisses :

$$X_i(t) = 3a_i(t-0.5)^2 + b_i,\ \forall t \in [0,1], \forall i \in \mathcal{S}_1,\ \text{où } a_i \sim \mathcal{N}(-3, 0.5),$$

et le second ensemble de courbes est défini comme suit :

$$X_i(t) = 3a_i(t-0.5)^2 + b_i,\ \forall t \in [0,1], \forall i \in \mathcal{S}_2,\ \text{où } a_i \sim \mathcal{N}(4, 3).$$

Les 400 variables aléatoires réelles b_i sont i.i.d. et suivent la loi $\mathcal{N}(0,3)$. Maintenant, nous disposons de 400 échantillons de courbes lisses. De plus, les courbes (c.-à-d. les X_i) sont discrétisées sur la même grille. Cette grille est constituée de 100 mesures équidistantes dans $[0,1]$. Pour compléter cette première simulation, nous définissons l'opérateur de régression r_1 par :

$$r_1(x) = \text{sign}(x'(1) - x'(0))\sqrt{\int_0^1 [x''(t)]^2\, dt},$$

dans tel cas $r_1(X_i) = 6\, a_i$, $i = 1, \ldots, 400$. Il suffit donc de simuler 400 valeurs des erreurs en fixant le rapport signal/bruit (comme c'est mentionné plus haut). Ceci a pour but de calculer les réponses scalaires correspondantes (c.-à-d. $Y_i = r_1(X_i) + \varepsilon_i$). Pour préciser nos idées, le graphique de gauche de la figure FIG. 5.1 montre 100 courbes régulières, alors que dans celui de droite on a tracé les reponses Y_i en fonction des $r_1(X_i)$. Comme on peut le remarquer, la forme des courbes est très régulière/lisse. Concernant la structure de l'échantillon fonctionnel, le graphique de droite montre que la distribution des paires $(r_1(X_i), Y_i)$ est très concentrée (petite variabilité) pour les valeurs négatives. Ceci est due essentiellement à la distribution fixe des coefficients a_i (petite variabilité pour des valeurs négatives et grande variabilité pour celles qui sont positives). Ce type d'hétéroscédasticité dans les données est très intéressant pour la comparaison des comportements, à travers leurs performances de prédiction, de l'estimateur à noyau utilisant une largeur de fenêtre locale et celui utilisant une largeur de fenêtre globale à travers leurs performances de prédiction. Pour ceci, soient $\mathcal{I}_0 = \{1, \ldots, 100, 201, \ldots, 300\}$, $\mathcal{I}_1 = \{101, \ldots, 200\}$ et $\mathcal{I}_2 = \{301, \ldots, 400\}$ trois ensembles d'indices. Nous prenons $\mathcal{L} = \{(X_i, Y_i)\}_{i \in \mathcal{I}_0}$ comme un échantillon d'apprentissage et les deux échantillons tests suivants :

$$\mathcal{T}_1 = \{(X_i, Y_i)\}_{i \in \mathcal{I}_1} \text{ et } \mathcal{T}_2 = \{(X_i, Y_i)\}_{i \in \mathcal{I}_2}.$$

Le premier échantillon test $\mathcal{T}_1$ contient 100 données fonctionnelles correspondant à une situation de "basse variabilité". Le deuxième échantillon test (c.-à-d. $\mathcal{T}_2$) est construit à partir de 100 données fonctionnelles faisant partie de la zone de "haute variabilité". Finalement, l'échantillon d'apprentissage (c.-à-d. $\mathcal{L}$) contient 200 données fonctionnelles, dont 100 de chaque situation (c.-à-d. basse/haute variabilité). L'échantillon d'apprentissage permet de calculer les largeurs de fenêtre globale et locale à travers une procédure de validation-croisée et les estimateurs à noyau fonctionnels correspondants r_1^{GCV} et r_1^{LCV} de r_1. La performance en termes de prévision est évaluée en calculant l'erreur quadratique moyenne relative (rmse)

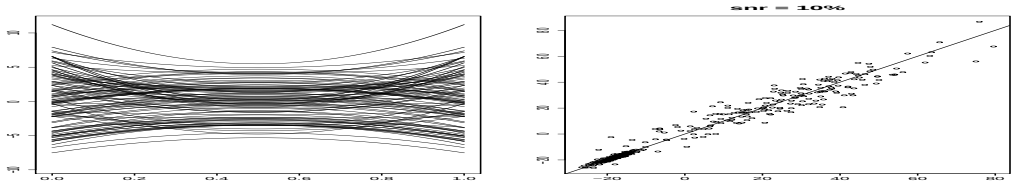

FIG. 5.1 – A gauche : 100 courbes régulières (c.-à-d. les X_i) et à droite : les réponses observées Y_i en fonction des réponses $r_1(X_i)$

Echantillons tests	$\mathcal{T}_1$ (basse variabilité)				$\mathcal{T}_2$ (haute variabilité)			
SNR	5%	10%	20%	30%	5%	10%	20%	30%
rmse_{LCV}	0.06	0.10	0.20	0.30	0.06	0.11	0.15	0.17
rmse_{GCV}	0.73	0.60	0.62	0.76	0.06	0.10	0.13	0.17

TAB. 5.1 – Courbes lisses : erreurs quadratiques moyennes relatives correspondant à plusieurs valeurs d'échantillons test et rapports signal/bruit.

sur l'échantillon test :

$$rmse_{GCV} = \frac{\sum_{i\in\mathcal{I}_k}\left(Y_i - r_1^{GCV}(X_i)\right)^2}{\sum_{i\in\mathcal{I}_k}(Y_i-\overline{Y})^2},$$

et, de façon similaire,

$$rmse_{LCV} = \frac{\sum_{i\in\mathcal{I}_k}\left(Y_i - r_1^{LCV}(X_i)\right)^2}{\sum_{i\in\mathcal{I}_k}(Y_i-\overline{Y})^2}, \text{ pour } k=1,2.$$

Il faut noter que la forme de l'opérateur de régression (c.-à-d. r_1) et le lissage des courbes nous conduisent à utiliser des semi-métriques basées sur les dérivées secondes (Cf pour plus de détails l'ouvrage [83]). Notons aussi que le calcul de cette semi-métrique est basé sur l'approximation de chaque courbe par des B-splines[1]. Les résultats sont donc résumés dans la table TAB. 5.1.4. Il apparaît que le choix de la largeur de fenêtre locale permet d'obtenir une très bonne qualité de prévision dans les zones où il y a une basse variabilité (c.-à-d. par rapport à $\mathcal{T}_1$) alors que les choix global et local conduisent à des performances similaires quand il s'agit de la zone de haute variabilité (c.-à-d. par rapport à $\mathcal{T}_2$). On s'interroge alors sur ces conclusions. Clairement, un choix global de la largeur de fenêtre conduit à un sur-lissage des données quand il s'agit de basse variabilité, alors qu'un choix local de la largeur de fenêtre permet de bien adapter les données dans les deux zones. La figure FIG.

[1]Les programmes (codes) et le mode d'emploi pour leur usage pratique sont disponibles sur le site Web *http ://www.lsp.ups-tlse.fr/staph/npfda/*

5.2 met l'accent sur ce fait. Elle présente les réponses observées en fonction des réponses prédites dans les deux cas (basse/haute variabilité de l'échantillon test) quand le SNR=10%. Clairement la prévision locale (c.-à-d. prévision quand on choisi localement le paramètre de lissage) est plus adaptative que la prévision globale (Cf les 1er et 3ème graphiques de FIG. 5.2). La deuxième idée concerne l'aspect fonctionnel des courbes. En effet, dans les situations pratiques, nous ne connaissons pas l'opérateur de régression r_1. En outre, on n'a aucun outil standard pour visualiser les courbes (c.-à-d. les X_i) en fonction des réponses correspondantes (c.-à-d. les Y_i). Par conséquent, nous ne disposons d'aucun outil pour se faire une idée sur le lien entre les courbes et les réponses. Autrement dit, dans la pratique, le graphique de droite de la figure FIG. 5.1 n'est pas disponible. Ainsi, il est impossible de savoir si un genre d'hétéroscédasticité se produit ou non dans la structure de l'ensemble des données fonctionnelles. En regardant notre échantillon de courbes lisses (Cf graphique de gauche de FIG.5.1), aucun effet d'hétéroscédasticité n'apparaît, alors qu'il existe un lien entre les réponses et les courbes explicatives (c.-à-d. les coefficients a_i). C'est la différence principale avec le cas standard (quand la variable explicative X est une variable aléatoire réelle et non pas une courbe aléatoire) où il est facile de visualiser le nuage de points et d'identifier quelques zones avec diverses concentrations (si elles existent) qui suggèrent l'utilisation de largeurs de fenêtres locales.

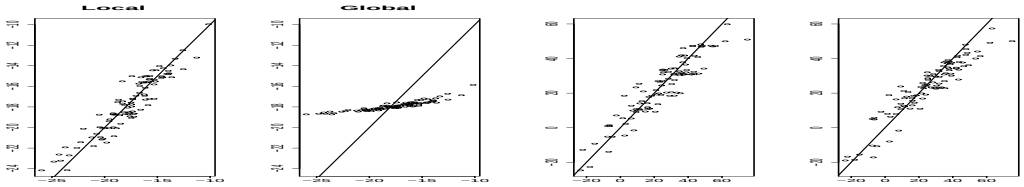

FIG. 5.2 – Courbes lisses (SNR=10%) : les réponses Y_i en fonction de leurs prédicteurs. Les deux graphiques de gauche (resp. de droite) concernent le cas de basse variabilité correspondant à $\mathcal{T}_1$ (resp. de haute variabilité correspondant à $\mathcal{T}_2$). Le 1er et le 3ème graphiques (resp. le 2ème et le 4ème graphiques) sont obtenus avec une largeur de fenêtre locale (resp. globale).

Courbes bruitées. Nous considérons, une nouvelle fois, les courbes que nous avons simulé précédemment, mais en y ajoutant un bruit blanc gaussien (Cf FIG. 5.3) :

$$X_i^*(t) \;=\; X_i(t) + \eta_{i,t}, \text{ avec } \eta_{i,t} \sim \mathcal{N}(0, 0.5) \text{ pour } i = 1, \ldots, 400.$$

Le cadre, ici, est un peu différent que dans le paragraphe précédent. En effet, nous disposons de l'échantillon fonctionnel $\{(X_i^*, Y_i)\}_{i=1,\ldots,400}$ (à la place de $\{(X_i, Y_i)\}_{i=1,\ldots,400}$ considéré précédemment). Donc, nous avons à prédire les réponses Y_i à partir des variables explicatives X_i^* (c.-à-d. nous avons à estimer r_1). Nous considérons les mêmes échantillons tests et d'apprentissage que précédemment (en remplaçant seulement les X_i par les X_i^*) : $\mathcal{L}^* = \{(X_i^*, Y_i)\}_{i\in\mathcal{I}_0}$ est l'échantillon d'apprentissage, $\mathcal{T}_1^* = \{(X_i^*, Y_i)\}_{i\in\mathcal{I}_1}$ et $\mathcal{T}_2^* = \{(X_i^*, Y_i)\}_{i\in\mathcal{I}_2}$ sont les deux

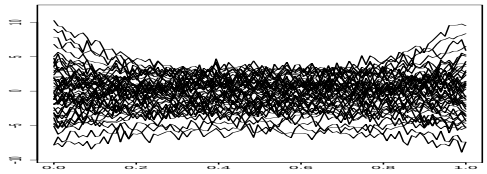

FIG. 5.3 – Un échantillon de 50 courbes bruitées.

échantillons tests (basse/haute variabilité). Comme dans ce qui précède, on utilise la semi-métrique basée sur les dérivées secondes. Mais, comme les courbes sont bruitées, nous les lisserons en réduisant, seulement, la taille de la base des B-splines. Les résultats qui sont résumés dans la table TAB. 5.1.4 montrent que le choix local de la largeur de fenêtre fourni des prévisions bien meilleures dans la zone de basse variabilité (c.-à-d. par rapport à $\mathcal{T}_1^*$) que celles obtenues par un choix global de la largeur de fenêtre. Nous traçons également

Echantillons tests	$\mathcal{T}_1^*$ (basse variabilité)				$\mathcal{T}_2^*$ (haute variabilité)			
SNR	5%	10%	20%	30%	5%	10%	20%	30%
rmse_{LCV}	0.30	0.34	0.44	0.45	0.06	0.11	0.19	0.28
rmse_{GCV}	0.68	0.73	0.75	0.76	0.08	0.11	0.22	0.27

TAB. 5.2 – Courbes bruitées : erreurs quadratiques moyennes relatives en fonction de plusieurs échantillons tests et valeurs du rapport signal/bruit.

les réponses observées (c.-à-d. les Y_i) en fonction de leurs prédicteurs quand SNR=10% (Cf FIG. 5.4). Nous remarquons donc l'occurence des mêmes comportements que lorsqu'il s'agit de courbes lisses.

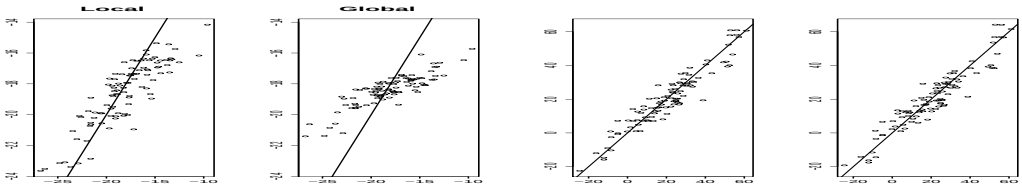

FIG. 5.4 – Courbes bruitées (SNR=10%) : les réponses Y_i en fonction de leurs prédicteurs. Les deux graphiques de gauche (resp. de droite) concernent le cas de basse variabilité correspondant à $\mathcal{T}_1$* (resp. de haute variabilité correspondant à $\mathcal{T}_2$*). Le 1er et le 3ème graphiques (resp. le 2ème et le 4ème graphiques) sont obtenus avec une largeur de fenêtre locale (resp. globale).

Courbes irrégulières. Le troisième et dernier ensemble de données simulées concerne des courbes irrégulières. La différence avec les données bruitées est que les réponses sont reliées directement aux courbes elles-mêmes et non à une version lissée de celles-ci (cf paragraphe précédent). Plus précisémment, on considère 400 courbes irrégulières (Cf FIG.5.5) :

$$\widetilde{X}_i(t) \;=\; a_i \sin(4(b_i - t)) + b_i + \eta_{i,t},$$

où les coefficients a_i, b_i, et η_i sont définis comme dans le paragraphe précédent. De plus, on défini un nouveau opérateur de régression r_2 par :

$$r_2(x) \;=\; \int_0^1 \frac{dt}{1 + |x(t)|},$$

Les réponses (c.-à-d. les $\widetilde{Y}_i$) sont construits à partir de r_2 dans le même sens que les simulations précédentes. Le graphique de gauche de la figure FIG. 5.5 montre un échantillon de 50 courbes irrégulières, alors que le graphique de droite montre les réponses $\widetilde{Y}_i$ en fonction de $r_2(\widetilde{X}_i)$. Nous considérons également $\widetilde{\mathcal{L}} = \{(\widetilde{X}_i, \widetilde{Y}_i)\}_{i\in\mathcal{I}_0}$ comme échantillon d'apprentissage, et $\widetilde{\mathcal{T}}_1 = \{(\widetilde{X}_i, \widetilde{Y}_i)\}_{i\in\mathcal{I}_1}$ et $\widetilde{\mathcal{T}}_2 = \{(\widetilde{X}_i, \widetilde{Y}_i)\}_{i\in\mathcal{I}_2}$ comme échantillons tests (basse/haute variabilité). A partir de la forme des courbes, la semi-métrique basée sur l'analyse en composante principale fonctionnelle est utilisée (Cf [83], pour plus de détails). Les résultats sont résumés dans la table TAB. 5.1.4. Dans les deux situations, le choix local de la largeur de fenêtre

Echantillons tests	$\widetilde{\mathcal{T}}_1$ (basse variabilité)				$\widetilde{\mathcal{T}}_2$ (haute variabilité)			
SNR	5%	10%	20%	30%	5%	10%	20%	30%
rmse_{LCV}	0.17	0.27	0.31	0.36	0.20	0.22	0.32	0.42
rmse_{GCV}	0.29	0.37	0.42	0.45	0.43	0.45	0.51	0.56

TAB. 5.3 – Erreurs quadratiques moyennes relatives correspondant à plusieurs échantillons tests et rapports signal/bruit.

donne des résultats meilleurs que le choix global (c.-à-d. quand les échantillons test sont $\mathcal{T}_1$ et $\mathcal{T}_2$). Encore une fois, on trace les réponses observées (c.-à-d. les $\widetilde{Y}_i$) en fonction de leurs prédicteurs quand SNR=10% (Cf FIG. 5.6) ce qui montre le bon comportement de l'estimateur fonctionnel à noyau obtenu à partir d'un choix local de la largeur de fenêtre.

5.1.5 Application sur des données réelles : courbes spectrométriques

Les bons résultats des simulations obtenus nous ont encouragés à utiliser systématiquement une largeur de fenêtre locale (dans le cadre fonctionnel). Ceci, sera confirmé dans l'étude suivante sur des données réelles. L'ensemble des données dites spectrométriques (cf

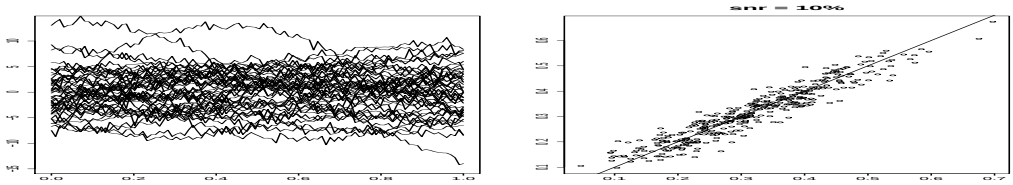

FIG. 5.5 – Graphique de gauche : 50 courbes irrégulières (c.-à-d. les $\widetilde{X}_i$) et Graphique de droite : $\widetilde{Y}_i$ en fonction de $r_2(\widetilde{X}_i)$

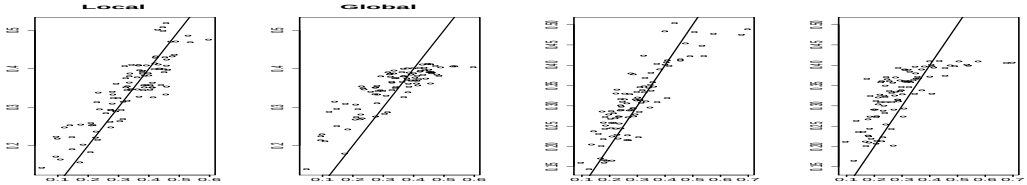

FIG. 5.6 – Courbes bruitées (SNR=10%) : les réponses $\widetilde{Y}_i$ en fonction de leurs prédicteurs. Les deux graphiques de gauche (resp. de droite) concernent le cas de basse variabilité correspondant à $\widetilde{\mathcal{T}}_1$ (resp. de haute variabilité correspondant à $\widetilde{\mathcal{T}}_2$). Le 1er et le 3ème graphiques (resp. le 2ème et le 4ème graphiques) sont obtenus avec une largeur de fenêtre locale (resp. globale).

figure FIG. 5.7) a l'avantage d'être bien connu dans la littérature sur les données fonctionnelles. Il s'agit en fait de 250 morceaux de viande hachés finement. Pour chaque unité i sur les 215 bouts de viande, on observe une courbe spectrométrique x_i qui correspond à la mesure d'absorbance à 100 longueurs d'ondes (c.-à-d. $x_i = (\xi_i(\lambda_1), \ldots, \xi_i(\lambda_{100}))$). De plus, pour chaque unité i, nous avons sa contenance de gras y_i pour $i = 1, \ldots, 215$, obtenue par un procédé chimique. Etant donné une nouvelle courbe spectrométrique x, notre principal objectif est de prédire de façon optimale la contenance de gras correspondante $\widehat{y}$. En effet, obtenir une courbe spectrométrique revient moins cher (en termes de temps et de coût) que de mettre en place le procédé chimique nécessaire à la détermination du pourcentage de gras. C'est pour cela que c'est un défi économique important de prédire la contenance de gras à partir de la courbe spectrométrique. Dans le but de comparer la performance de notre prévision, nous construisons l'échantillon d'apprentissage avec 160 courbes spectrométriques (et les réponses correspondantes) et nous calculons les erreurs quadratiques moyennes relatives avec les 55 courbes spectrométriques restant. Les résultats sont présentés dans la figure FIG. 5.8. De plus, on obtient $\text{rmse}_{LCV} = 1.89$ et $\text{rmse}_{GCV} = 5.37$. Nous remarquons, encore une fois, que le verdict est sans appel pour favoriser un choix local de la largeur de fenêtre. Ce qui améliore significativement la qualité de prévision par rapport à un choix global de la largeur. En conclusion, il est trivial de penser qu'un choix local de la largeur de fenêtre conduit à un bon comportement de l'estimateur fonctionnel à noyau. Ceci est valable des deux points de vue, théorique et pratique. En fait, cette partie met la lumière sur la nécessité d'utiliser un critère de choix local de la largeur de fenêtre dans le cadre fonctionnel (c.-à-d. quand la variable explicative est une variable fonctionnelle). La raison principale vient du fait que les variables fonctionnelles admettent souvent une structure

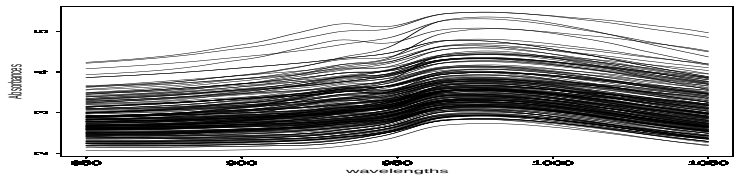

FIG. 5.7 – Les courbes spectrométriques

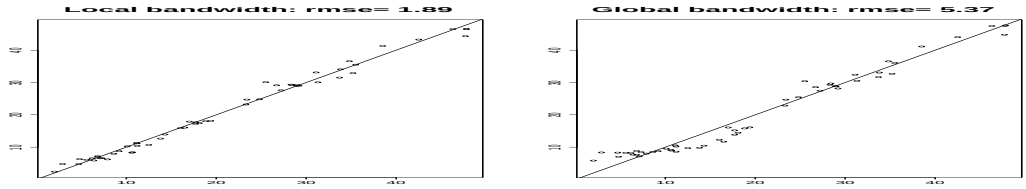

FIG. 5.8 – Données spectrométriques : réponses observées en fonction des réponses prédites

complexe. De plus, la partie informative de l'ensemble des données fonctionnelles peut être fortement concentrée dans divers secteurs de l'espace et clairsemée ailleurs. En particulier, comme c'est remarqué dans le paragraphe Simulations, l'hétéroscédasticité n'apparaît pas fondamentalement à cause de la forme des courbes elles-mêmes. La deuxième difficulté dans l'analyse des données fonctionnelles est due au manque d'outils graphiques pour explorer le lien éventuel entre des données fonctionnelles et des réponses scalaires, ce qui n'est pas le cas dans le cadre multivarié (c.-à-d. quand la variable explicative est de dimension finie, on peut visualiser un nuage de points). Ainsi, afin de prévoir une telle structure en dimension élevée, les méthodes adaptatives sont préférables, comme le choix local de la largeur de fenêtre. Manifestemment, le gain en employant le paramètre local est beaucoup plus important dans le cadre fonctionnel qu'en dimension finie.

5.2 Estimation forte de l'opérateur de régression pour des données fonctionnelles et des processus d'erreur à longue mémoire

Dans ce paragraphe, nous nous intéressons à la structure de covariance du processusd'erreur. En fait, nous supposons que le processus d'erreur ε est un processus de second ordre faiblement stationnaire, centré et unitaire, et tel que :

$$I\!E\left(\varepsilon_i \varepsilon_j\right) = f_\varepsilon(i-j) \text{ pour } i,\, j \in \mathbb{Z}$$

La plupart des modèles standards de séries temporelles qui ne supposent pas la sommabilité de la suite des covariances $(f_\varepsilon(i))_{i\in\mathbb{Z}}$ ont induit les processus à longue mémoire (Cf [71]) :

$$\sum_{i=-\infty}^{+\infty} |f_\varepsilon(i)| = +\infty.$$

Nous estimons l'opérateur de régression $r(x) = I\!E\,(Y|X = x)$ pour $x \in E$ comme précédemment par $\widehat{r}_h(x)$ (Cf expression (5.0.2)).
D'abord, sous la condition (5.0.12), nous avons :

$$I\!E(\widehat{r}_{1,h}(x)) - r(x) = O\left(h^\beta\right), \text{ quand } n \to +\infty$$

Ensuite, sous les conditions (5.0.7), (5.0.8), (5.0.9), (5.0.11), (5.0.12) et (5.0.13), si le processus d'erreur est à longue mémoire (fonction de covariance qui vérifie (5.0.1)), alors : $\widehat{r}_{1,h}(x)$ converge vers $I\!E(\widehat{r}_{1,h}(x))$ et $\widehat{r}_{2,h}(x)$ converge vers 1, presque complétement, quand $n \to +\infty$. Ceci nous conduit à :

$$\widehat{r}_h(x) \to r(x), \text{p.co., quand } n \to +\infty \text{ pour tout } x \in S.$$

Par ailleurs, si l'on remplace la condition (5.0.7) par (5.0.6), et si en plus de l'hypothèse (5.0.3), on dispose de la condition suivante (Cf [83] pour plus de détailles) :

$$\forall \epsilon_1 > 0, \exists\, C_3 > 0 \text{ et } \epsilon < \epsilon_1 \text{ tel que } \int_0^\epsilon \varphi(u)\, du > C_3\, \epsilon\, \varphi(\epsilon),$$

alors, nous obtenons :

$$\sup_{x\in S} |\widehat{r}_h(x) - r(x)| \to 0, \text{p.co., quand } n \to +\infty$$

5.3 Données déterministes fonctionnelles : estimation de l'opérateur de régression pour des données dépendantes

Ce paragraphe se situe dans la continuité du précédent à une particularité près. En effet, nous nous intéressons au modèle $Y = r(x) + \varepsilon$ où la réponse Y est une variable aléatoire réelle, la variable explicative x est une variable fonctionnelle déterministe (Functional Fixed-Design) et ε est un processus stochastique réel de second ordre et stationnaire.
Parmi les motivations qui nous ont incités à considérer ce sujet, il en est une à savoir que les données déterministes ont un sens dans beaucoup de situations pratiques dans lesquelles on contrôle l'entrée x (signal d'un message, la concentration d'un catalyseur dans une réaction chimique, ...) et qui produit une sortie aléatoire Y affectée par un bruit. En particulier, la chimiométrie offre un large domaine d'applicabilité des outils de la régression fonctionnelle sous des données déterministes fonctionnelles (Cf [1] et [86]).
Pour les données fonctionnelles aléatoires, on a fourni dans [41] des résultats consistants

dans le cas de la régression linéaire avec réponse scalaire. Dans le cadre de données fonctionnelles déterministes et des réponses fonctionnelles et aléatoires, on a construit dans [1] des estimateurs consistants quand l'opérateur de régression est linéaire (pour des données fonctionnelles bien choisies).
Dans ce paragraphe, nous nous intéressons au modèle de régression non linéaire quand la variable explicative est fonctionnelle déterministe et la réponse est une variable aléatoire réelle et quand le processus d'erreur $\varepsilon = \{\varepsilon_t\}_{t\in\mathbb{Z}}$ est stationnaire de second ordre.
plus précisément, notre but est d'estimer l'opérateur r à partir de l'échantillon (x_1, Y_1), ..., (x_n, Y_n) de (x, Y) avec $I\!E|Y| < \infty$. Pour ceci, nous utilisons l'estimateur à noyau $\widehat{r}_h(x)$ donné par l'expression (5.0.2).
Sous les hypothèses (5.0.7), (5.0.12) et (5.0.13), l'estimateur $\widehat{r}_h$ converge en moyenne quadratique, c.-à-d. (quand $n \to +\infty$) :

$$I\!E\left(\widehat{r}_h(x) - r(x)\right)^2 = O\left(h^{2\beta}\right) + O\left(\frac{1}{n_x(h)}\right) + O\left(\frac{s_n}{n_x(h)}\right) \text{ où } s_n = \sum_{k=1}^{n} |f_\varepsilon(k)|$$

et $n_x(h) = card\{x_i : x_i \in \mathcal{B}(x,h)\} = \sum_{i=1}^{n} 1_{\mathcal{B}(x,h)}(x_i)$.
En particulier, si le processus d'erreur est à longue mémoire, alors nous obtenons :

$$MSE(x) = O\left(h^{2\beta}\right) + O\left(\frac{n^{1-\gamma}}{n_x(h)}\right), \text{ quand } n \to +\infty$$

Par ailleurs, dans le cas particulier où les réponses sont non corrélées, c.-à-d. dans notre cas, si les erreurs $\varepsilon_1, \ldots, \varepsilon_n$ sont non corrélées, alors

$$MSE(x) = O\left(h^{2\beta}\right) + O\left(\frac{1}{n_x(h)}\right), \text{ quand } n \to +\infty.$$

On rejoint donc le résultat obtenu en dimension finie (Cf [106]). De plus, sous des conditions supplémentaires, nous avons obtenu, facilement, les vitesses de convergence de $\widehat{r}_h$ et la forme exacte des constantes de MSE comme dans [79] (ceci fera l'objet d'un chapitre de la thèse de doctorat de S. Hedli-Griche).

5.3.1 Simulations

Nous testons les performances de nos résultats théoriques sur des données simulées. Pour ne pas alourdir la réaction de ce paragraphe, nous nous'y restreignons, simplement, aux données déterministes fonctionnelles. Nous considérons le modèle de régression $Y = r(x) + \varepsilon$ pour deux types de processus d'erreur : Ornstein-Uhlenbeck (OU) et FARIMA (voir Fig. 5.9). Nous générons deux courbes $(x_k^{(j)}(t))_{t\in[0,1]}$ pour $j \in \{1, 2\}$ pour la taille de l'échantillon $n = 300$ et le nombre de répétitions $m = 100$. Les courbes $(x_k^{(1)}(t))$ sont des courbes quadratiques et $(x_k^{(2)}(t))$ sont des courbes irrégulières. Les échantillons des courbes $(x_k^{(1)}(t))$ et $(x_k^{(2)}(t))$ sont simulés à partir des familles suivantes (Cf Fig. 5.10) :

$$x_k^{(1)}(t) = k\,t^2 \text{ et } x_k^{(2)}(t) = k\,t^2 \sin(kt) \text{ pour } k = 1, ...,$$

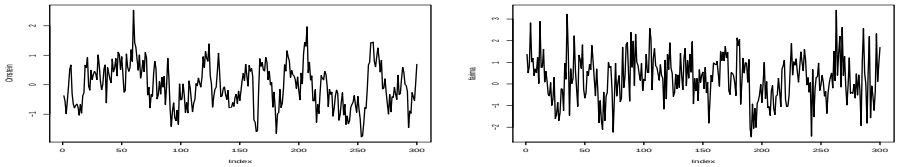

FIG. 5.9 – De gauche vers la droite : le processus d'Ornstein-Uhlenbeck (OU), FARIMA(0.1,0.3,0.1)

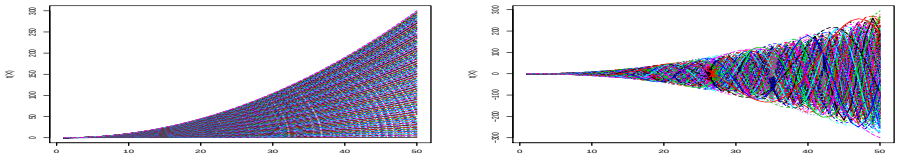

FIG. 5.10 – De gauche vers la droite : 300 courbes (données) de $x_k^{(j)}$, $j = 1, 2$.

Les observations $Y_i^{(j)}$ pour $j = 1, 2$ et $i = 1, \ldots, n$, ont été générées à partir d'observations identiquement distribuées des données $x_i^{(j)}$. Nous prenons l'opérateur non linéaire $r^{(j)}$ pour le $j$ème modèle pour $j = 1, 2$ comme suit :

$$r^{(1)}(x) = \int_0^1 |x'(t)| \log |x'(t)| dt \text{ et } r^{(2)}(x) = \int_0^1 [(x(t)']^2 dt$$

Le but de nos simulations et de retrouver nos résultats théoriques en terme de convergence de $\widehat{r}_h$ pour les données $x_i^{(j)}$, $j = 1, 2$ quand on considère différents types de processus d'erreurs et de semi-métriques.
Nous avons utilisé les deux semi-métriques SMD et SMPCA pour traiter ces données. Les résultats sont donnés dans les figures FIG. 5.11, FIG. 5.12, FIG. 5.13 et FIG. 5.14) pour différents types de mémoires -mémoire courte- (processus d'Ornstein-Uhlenbeck standards) et -mémoire longue- (processus FARIMA). Il est clair que lorsque les courbes sont régulières, la semi-métrique SMD est plus appropriée que SMPCA. Ceci est confirmé, également, par les valeurs des RMSE données dans la table TAB. 5.4.

Modèles	Modèle 1		Modèle 2	
Processus d'erreur	OU	FARIMA	OU	FARIMA
RMSE(SMD)	0.0001453992	0.0004322320	0.4280282	0.4280282
RMSE(SMPCA)	0.0001484537	0.00044402760	0.00886394	0.008863943

TAB. 5.4 – Modèle 1 : processus d'erreur OU (resp. FARIMA(0.1,0.3,0.1)) on calcule RMSE pour les semi-métriques SMD et SMPCA

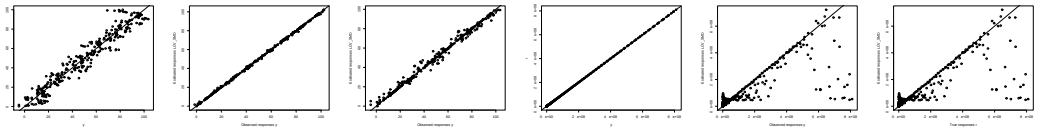

FIG. 5.11 – Modèles 1 et 2 avec SMD(OU) : de gauche vers la droite, pour chaque modèle les graphiques successifs représentent : (y, r), $(y, \widehat{r}_h)$, $(\widehat{r}_h, r)$

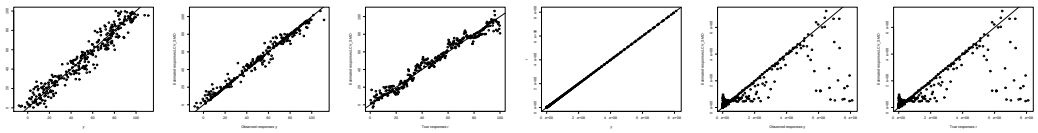

FIG. 5.12 – Modèles 1 et 2 avec SMD(FARIMA) : de gauche vers la droite, pour chaque modèle les graphiques successifs représentent : (y, r), $(y, \widehat{r}_h)$, $(\widehat{r}_h, r)$

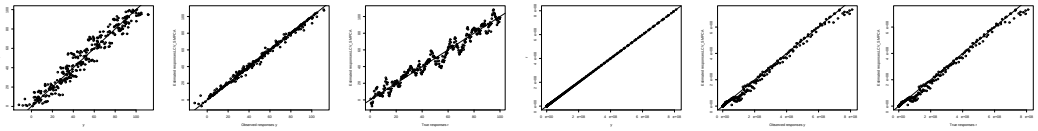

FIG. 5.13 – Modèles 1 et 2 avec SMPCA(OU) : de gauche vers la droite, pour chaque modèle les graphiques successifs représentent : (y, r), $(y, \widehat{r}_h)$, $(\widehat{r}_h, r)$

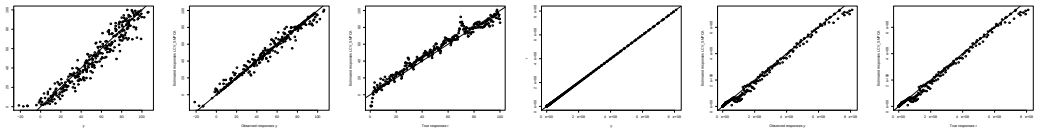

FIG. 5.14 – Modèles 1 et 2 avec SMPCA(FARIMA) : de gauche vers la droite, pour chaque modèle les graphiques successifs représentent : (y, r), $(y, \widehat{r}_h)$, $(\widehat{r}_h, r)$

Chapitre 6

Modélisation du canal de transmission à 60 Ghz dans les milieux confinés

Les communications mobiles tendent de plus en plus vers une globalisation des services et une augmentation des débits transmis. Nous nous intéressons aux communications intra-bâtiments pour une prochaine génération de communications mobiles. Il nous faut donc viser des débits très importants (plusieurs centaines de Mbits par seconde) et une souplesse d'utilisation, aussi bien pour la gestion des utilisateurs que pour la gestion des services.
L'option choisie est de trouver des fréquences disponibles dans la gamme millimétrique. Les bandes passantes disponibles sont en effet plus importantes et permettent d'atteindre des débits plus élevés. Dans ce cadre, la bande autour de 60 Ghz a été identifiée comme une bande possible pour les transmissions à très haut débit et, qui plus est, sans licence. Un autre avantage de cette bande est la forte atténuation des ondes, en particulier, l'absorption due à l'oxygène. Si cette particularité peut augmenter la complexité des systèmes à mettre en oeuvre pour assurer des qualités de service satisfaisantes, elle permet un très bon confinement des ondes, simplifiant de ce fait la planification en fréquence et réduisant nettement la pollution électromagnétique.

Pour rendre ce chapitre facile à lire, nous avons opté pour la descritpion des méthodes utilisées et des résultats obtenus. Pour les détails techniques, nous renvoyons à [48] et [49]. Une vue globale de ce projet et les prolongements possibles peuvent être trouvés sur le site internet : http ://www.iemn.univ-lille1.fr/recherches/circuits.htm#ancre3

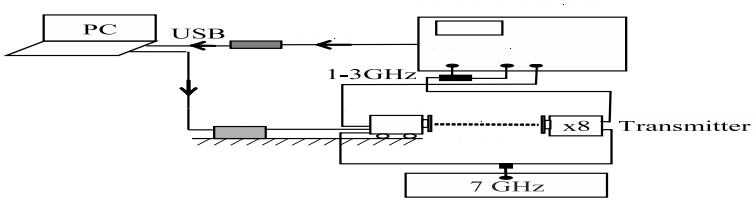

FIG. 6.1 – Principe du sondeur de canal : un signal à 7 GHz est généré et multiplié par 8. A ce signal, l'analyseur de réseau ajoute un autre signal dont la fréquence va de 1 à 3 GHz. La fonction de transfert est mesurée sur une bande allant de 57 à 59 GHz

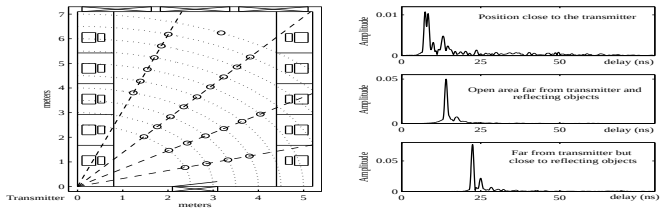

FIG. 6.2 – Salle des mesures.

Canal de transmission. Le sondeur de canal à 60 GHz développé à l'IEMN[1] est un sondeur fréquentiel large bande reposant sur une analyse vectorielle de la fonction de transfert par transposition de fréquence. Pour cela deux têtes d'émission et de réception hétérodynes ont été développées en intégration monolithique avec des fréquences RF de 57 à 59 GHz et des fréquences intermédiaires de 1 à 3 GHz. Un analyseur de réseau vectoriel permet après calibrage la mesure vectorielle de la fonction de transfert du canal et, après transformée de Fourier inverse, la détermination de sa réponse impulsionnelle. La résolution temporelle de ce sondeur est de 0.5 ns pour une durée de réponse impulsionnelle de 800 ns typiquement, ce qui le rend particulièrement bien adapté à l'analyse des caractéristiques d'un environnement confiné. Le but de nos mesures étant de caractériser les phénomènes de propagation, non seulement à grande échelle mais également à petite échelle (ce qui correspond à mesurer les signaux reçus à des emplacements séparés d'une distance inférieure à la longueur d'onde). Un système de positionnement précis sur lequel le module récepteur est fixé a été mis en place. La distance entre deux mesures adjacentes est paramétrisable par pas de 1 mm sur une longueur totale d'environ 50 cm. L'environnement des mesures est une salle informatique de taille moyenne (Cf FIG. 6.2).

[1]Institut d'Electronique de Microélectronique et de Nanotechnologie (Lille)

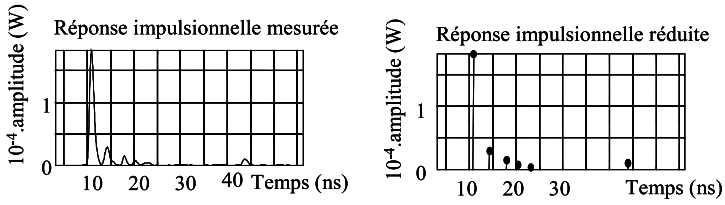

FIG. 6.3 – Réduction de la réponse impulsionnelle

6.0.2 Modélisation statistique des amplitudes, des retards et des impulsions

Nous considérons le canal stationnaire et nous le modélisons par la réponse impulsionnelle d'un filtre linéaire :

$$h(t) = \sum_{k=1}^{N} \beta_k \delta(t - \tau_k) \exp(j\theta_k)$$

Cette réponse impulsionnelle est caractérisée par quatre variables aléatoires :

- N : nombre de trajets.
- $\{\theta_k\}_{k=0,\ldots,N}$: phase de chaque trajet.
- $\{\tau_k\}_{k=0,\ldots,N}$: les retards de chaque trajet. τ_0 n'est pas une variable aléatoire car il est toujours pris égal à 0.
- $\{a_k\}_{k=0,\ldots,N}$: amplitude de chaque trajet dont la valeur dépend du retard.

Dans un premier temps nous ramenons la réponse impulsionnelle mesurée à une séquence de mots (Figure FIG. 6.3) correspondant à tous les maximums locaux rencontrés (qui sont au plus 100 fois plus faible que le pic principal). Chaque mot est défini par le retard du maximum, son amplitude et sa phase. Cette réduction ne donne pas nécessairement la réponse impulsionnelle exacte. Par exemple, certains trajets peuvent être masqués. Cependant, notre objectif est de définir un modèle représentant le comportement global du canal dans la pièce et non d'être le plus précis possible sur chaque réponse impulsionnelle. L'approximation faite lors de la réduction ne présente aucun inconvénient majeur.
Notre modèle se base sur l'étude statistique des quatre variables aléatoires décrivant le canal. Dans [227], on a développé un modèle similaire pour un environnement extérieur à 800 MHz et dans [219] pour un environnement intra-bâtiment à 1.5 GHz.
Nous avons choisi de ne faire aucune hypothèse a priori, mais nous avons débuté l'étude de chaque variable par des estimations non paramétriques, et vérifié l'existence ou non, de corrélations entre les différentes variables. Nous avons utilisé l'estimateur à noyau, pour estimer la densité de probabilité des phases $\{\theta_k\}$. En utilisant un test de Kolmogrov-Smirnov, nous avons confirmé l'hypothèse, qui semble visuellement acceptable, que les phases sont uniformément distribuées sur l'intervalle $[0, 2\pi)$.

Indépendance entre les phases, les amplitudes et les retards. Nous utilisons le test de Hoeffding pour évaluer l'indépendance des phases d'une part et des amplitudes et des

retards de l'autre part. A l'issue de ce test statistique, nous pouvons affirmer que les phases sont indépendantes des retards et des amplitudes.
Un test de Spearman a été, également, effectué pour montrer l'existence d'une corrélation entre les retards et les amplitudes. Par conséquent, nous allons tout d'abord estimer la distribution des retards, puis représenter la corrélation entre les amplitudes et les retards.

Distribution des retards. On considère souvent que les retards suivent une loi de Poisson. Cependant, dans [228] on a proposé une loi de Poisson modifiée pour modéliser les retards. Quand un trajet arrive, le taux d'arrivée moyen est augmenté d'un facteur κ pour une durée Δ secondes. Le facteur κ et la durée Δ peuvent être choisis comme fonctions du retard. Si $\kappa = 1$ et $\Delta = 0$ on retombe sur une loi de Poisson classique. Sinon, l'arrivée d'un trajet induit l'augmentation de la probabilité d'un autre trajet dans les secondes suivantes. La médiocrité de ce modèle a conduit à proposer dans [219], un mélange de deux distributions de Poisson : les trajets arrivent en paquets ("clusters") : les temps d'arrivée des paquets suivent une loi de Poisson et les temps d'arrivée des trajets dans chaque paquet suivent une autre loi de Poisson. Pour notre étude, nous avons choisi de ne pas faire d'hypothèse a priori et donc d'utiliser une estimation non paramétrique de la loi d'arrivée des retards. Comme le laissent supposer les figures dans [49], de la densité de probabilité des retards estimée par un estimateur à noyau et un test de Kolmogorov-Smirnov le confirme : une loi de Poisson ne permet pas de bien modéliser la distribution des retards. Nous avons donc décidé de conserver le modèle d'estimation non paramétrique. Nous pouvons alors estimer les puissances de chaque trajet par une régression non linéaire en fonction des retards.

Modélisation des amplitudes. A partir des tests effectués sur les données, il paraît évident qu'il y a une relation (qui est fondamentale) entre les amplitudes et les retards. Sans hypothèse préalable sur cette relation, nous avons procédé par l'estimation d'une régression non linéaire entre ces deux variables. Pour ceci, nous avons utilisé une estimation à noyau de cette régression (Cf [36] et [117]), dont la forme reflète exactement l'atténuation des amplitudes quand les retards augmentent (Cf [49]). Par contre, pour l'étude de la distribution du bruit, nous avons calculé, d'abord, les résidus $\{e_k\}$, et nous avons alors justifié l'approximation de la loi de ces résidus par une distribution gaussienne, centrée dont la variance dépend des retards.
Par ailleurs, la distribution du nombre de trajets ne pose pas de problème notable. Elle est estimée, simplement, en calculant :
Probabilité d'avoir k trajets = Nombre de réponses impulsionnelles ayant k trajets / Nombre total de réponses impulsionnelles

6.0.3 Modélisation des lignes à retards en utilisant les lois K

En considérant les mécanismes responsables des réflexions dans un milieu confiné à 60 GHz et la méthodologie de sondage du canal, nous présentons un modèle multitrajets de la composante aléatoire basée sur le concept de groupe d'ondes (Cf [145]). Notre approche consiste à modéliser un faisceau par une succession de réflexions sur des obstacles dont les surfaces

sont approchées par un ensemble de facettes planes. Ce modèle permet de tenir compte des différents modes de diffractions telles que les réflexions diffuses ou spéculaires. De plus, en considérant que les caractéristiques de propagation varient de manière significative en environnement indoor, nous développons un modèle englobant les évanouissements à petite et à moyenne échelle ainsi qu'un modèle de canal numérique appelé modèle K en référence aux lois suivies par ses coefficients (Cf [219] et [228]).

Modèle statistique multitrajets. Dans cette étude, nous considérons que l'onde électromagnétique émise est diffractée sur différents objets de l'environnement. La diffraction ayant ici le sens de champ modifié par la présence d'un obstacle et comprend tous les types d'interactions possibles. A haute fréquence, le champ diffracté ne dépend pas, en un point d'observation donné, du champ en tout point de la surface de l'environnement mais seulement du champ aux voisinages de certains points de diffractions. Chacun d'eux ayant une contribution qui dépend uniquement des propriétés électromagnétiques et de la géométrie locale de la surface diffractante. Cette dernière peut être approchée au premier ordre par un ensemble de surfaces planes appelées facettes. L'onde incidente crée à la surface de la facette un champ électromagnétique dont une partie rayonne vers le récepteur. Pour le récepteur, les facettes peuvent être considérées comme un réseau d'antennes fictives dont les diagrammes de rayonnement sont fonctions de leurs tailles, orientations et des propriétés de conduction du milieu. L'ensemble des ondes issues d'un même voisinage est appelé groupe d'ondes ou faisceau et est caractérisé par une orientation relative à la disposition de la surface de l'obstacle. Par ce principe de localisation, le problème global de la diffraction d'un champ incident est ramené à un ensemble de contributions distinctes et le canal consiste en une distribution spatiale de faisceaux ou groupe d'ondes. Cette approche rejoint ainsi celles développées dans [219] et [228], pour la modélisation statistique de la propagation électromagnétique en milieu urbain et indoor respectivement. Le profil temporel de canal est donné par :

$$h(t) = \sum_{k=1}^{\infty} h_k(t),$$

où $h_k(t)$ est le profil du faisceau k dont l'étalement temporel est noté τ_k :

$$h_k(t) = \sum_{l=1}^{\infty} h_{k,l}(t) = \sum_{l=1}^{\infty} \beta_{k,l} \exp(j(2\pi f_c \tau_{k,l} + \theta_{k,l})\ \delta(t - T_k - \tau_{k,l})\ 1_{[0,\tau_k]}(t - T_k - \tau_{k,l})$$

où $1_{[0,\tau_k]}(t)$ étant la fonction indicatrice sur $[0, \tau_k]$, $\delta(.)$ l'impulsion de Dirac, T_k est l'instant d'arrivée de la première onde du groupe k, $\beta_{k,l}$, $\theta_{k,l}$ et $\tau_{k,l}$ sont respectivement l'amplitude, la phase et l'instant d'arrivée de l'onde l du faisceau k. Le temps τ_k est mesuré depuis l'instant initial T_k du groupe k.
Le profil du canal s'écrit en conséquence :

$$h(t) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \beta_{k,l} \exp(j(2\pi f_c \tau_{k,l} + \theta_{k,l}))\, \delta(t - T_k - \tau_{k,l})\, 1_{[0,\tau_k]}(t - T_k - \tau_{k,l})$$

En considérant les phénomènes physiques de propagation indoor, nous supposons vérifiées les hypothèses suivantes :

- Les distances facettes-récepteur dépendent de la géométrie de la surface mais aussi de l'orientation de la surface diffractante par rapport au récepteur. Les facettes "visibles", c'est à dire de contribution non nulle au récepteur, sont réparties aléatoirement dans le voisinage. Les instants d'arrivée $\{\tau_{k,l},\ l \text{ entier }\}$ des ondes de chaque faisceau k forme un processus de Poisson de paramètre λ_k sur une durée τ_k petite par rapport au temps symbole. λ_k et τ_k caractérisent respectivement la densité temporelle d'ondes et la dispersion temporelle du faisceau. Cette dernière dépend de la taille et de l'orientation du voisinage de diffraction par rapport au récepteur. Chaque faisceau est paramétré par un nombre moyen d'ondes $\lambda_k\tau_k$ grand, une onde moyenne d'amplitude β_k et de phase θ_k. En considérant homogènes les voisinages de diffraction et indépendantes les facettes, les ondes résultantes sont modélisées par des variables aléatoires complexes indépendantes et identiquement distribuées. Remarquons qu'elles ne sont pas nécessairement centrées et circulaires puisque les trajets facettes-récepteur étant proches leurs phases ne varient pas nécessairement de plusieurs longueurs d'onde (5 mm dans le cas d'une émission à 60 GHz).
- Nous supposons de plus vérifiée l'hypothèse, dont l'initiative revient à [228] et [229], que les instants d'arrivée de chaque faisceau forment une séquence temporelle de Poisson de taux moyen d'arrivée Λ et que les amplitudes et phases des faisceaux sont des variables aléatoires indépendantes et identiquement distribuées.
- La densité temporelle d'onde λ_k, la dispersion temporelle τ_k, l'amplitude β_k et la phase θ_k des faisceaux sont décrits par quatre processus aléatoires variant dans le temps. Le canal est en conséquence modélisé par un processus non stationnaire. La non-stationnarité de $\{\beta_k\}$, $\{\theta_k\}$, $\{T_k\}$ et $\{\tau_k\}$ est cependant lente comparée au débit utile, nous supposons ainsi leurs variations négligeables sur une durée symbole.

Ainsi, les hypothèses développées tiennent compte de faisceaux de dispersions et de densités temporelles variables.

Modèle d'évanouissement à petite échelle. Considérons le champ émis de phase nulle et d'amplitude unité, le champ reçu E en un site s est la somme des ondes issues des trajets composants le canal :

$$E(s) = \sum_{k=1}^{\infty} B_k \text{ avec } B_k = \sum_{l=1}^{N_k} \beta_{k,l} \exp(2\pi f_c(T_k + \tau_{k,l}) + \theta_{k,l})$$

Commençons par déterminer la loi des B_k. Chaque faisceau comporte un nombre aléatoire d'ondes indépendantes et identiquement distribuées. Suivant les réflexions des faisceaux, les phases des ondes sont de distribution quelconque entre 0 et 2π et leurs amplitudes sont variables. Ces hypothèses permettent de tenir compte de situations différentes. Considérons un trajet comportant une réflexion sur un obstacle incliné par rapport à l'onde incidente, les distances facettes-récepteur varient de plusieurs longueurs d'onde et leur phase est alors de distribution uniforme. Si cette hypothèse n'est plus vérifiée, certaines valeurs entre 0

et 2π sont privilégiées et la densité de la phase est en conséquence multimodale sur le support $[0, 2\pi]$. A l'extrême, si les ondes sont en phase, l'onde résultante est constructive et la réflexion est dite spéculaire. Dans ce cas, l'ensemble des facettes du voisinage de diffraction agit comme une antenne dont la taille est celle du voisinage. Le nombre d'ondes par faisceau étant distribué suivant une loi de Poisson dont le paramètre est grand et le faisceau k est une variable aléatoire complexe gaussienne dont la moyenne est $\beta_k \exp(j\theta_k)$ et la variance est donnée par :

$$\sigma_k^2 = \frac{\tau_k}{2}\beta_k^2\lambda_k.$$

En conséquence, chaque faisceau k est paramétré par un nombre moyen d'ondes $\lambda_k\tau_k$, un gain β_k et une phase θ_k. La variation de longueur des faisceaux étant beaucoup plus grande que la longueur d'onde émise, leurs phases sont uniformément distribuées sur $[0, 2\pi]$. En considérant de plus que le nombre de faisceaux moyen par temps symbole (ΛT) est grand, $E(s)$ converge vers une gaussienne circulaire, centrée et de variance

$$\sigma^2 = \Lambda T\sigma_B^2$$

avec σ_B^2 l'espérance de la variance suivant l'amplitude et la phase des faisceaux. Remarquons que σ_B^2 et σ^2 sont respectivement les puissances moyennes des faisceaux et du canal.
Jusqu'ici, nous avons modélisé le canal en considérant l'émetteur et le récepteur à une position fixe et avons retrouvé le modèle d'évanouissement à petite échelle connu sous le nom de modèle de Rayleigh en référence à la distribution de l'amplitude de $E(s)$. Ses paramètres $\{\Lambda, \sigma_B^2\}$ sont déterministes, leurs valeurs dépendent en effet de l'environnement de propagation pour une configuration précise. Considérons à présent le voisinage du récepteur. Puisque les ondes d'un même faisceau dépendent de l'orientation des facettes modélisant la surface de diffraction, sa densité temporelle d'ondes λ_k ainsi que son étalement temporel τ_k sont variables suivant la position du récepteur dans le voisinage. Il en est de même pour Λ et σ_B^2. Pour tenir compte de toutes ces variations, σ^2 est supposée aléatoire. Cette hypothèse supplémentaire permet de tenir compte des diffractions suivantes : si Λ est variable, le processus de Poisson à l'origine des instants d'arrivée des faisceaux est un processus de Poisson composé. Les durées entre les arrivées successives des faisceaux ne sont plus indépendantes. Le processus permet de prendre en compte les groupes ("clusters") de faisceaux présents dans les profils temporels.
En supposant σ_B^2 variable, on tient compte de plus de la variation de la puissance des faisceaux. Par cette hypothèse, les instants d'arrivée des ondes de chaque faisceau forment aussi un processus de Poisson composé et la dispersion temporelle des faisceaux peut être considérée comme aléatoire. Les ondes de chaque faisceau arrivent alors plus ou moins groupées au récepteur et les voisinages diffractant d'orientation et de formes variées.
Ainsi, supposer σ^2 aléatoire permet de rendre plus général le modèle statistique puisqu'il tient compte d'un grand nombre de modes de diffraction.
Parce que la variable σ^2 est cachée, la modélisation de sa distribution est délicate. Le choix de sa distribution doit principalement tenir compte de son support positif et de la souplesse de la forme de sa densité de probabilité de manière à être adaptée à différents environnements de propagation. A ce titre, nous proposons une approche identique à celle développée

en imagerie radar. La variable σ^2 est supposée distribuée suivant la loi Gamma

$$f_{\sigma^2}(u) = \frac{u^{\alpha-1}}{\gamma^\alpha \Gamma(\alpha)} \exp\left(-\frac{u}{\gamma}\right),\ u \geq 0$$

On obtient alors pour chaque coefficient la loi K (Cf [228]).

$$f_{|E|}(x) = \frac{4x^\alpha}{\gamma^{\frac{\alpha+1}{2}}\Gamma(\alpha)} K_{\alpha-1}\left(\frac{2}{\sqrt{\gamma}}x\right),\ x \in [0, +\infty[$$

avec $K_{\alpha-1}$ est la fonction de Bessel modifiée de seconde espèce, et γ est le paramètre d'échelle. La forme de la densité K dépend uniquement du paramètre $\alpha(n)$. Remarquons que la loi K converge pour α grand vers la loi de Rayleigh.

Réponse impulsionnelle numérique du canal. Considérons à présent un système de communication numérique de durée symbole T. Le filtre numérique modélisant le canal multitrajets est construit en intégrant $h(t)$ sur chaque temps symbole. Puisque les obstacles de l'environnement sont confinés dans une pièce, les instants d'arrivée des ondes ainsi que leurs densités temporelles sont fonctions du temps. Pour des systèmes de communication large bande ($>$ 100 MHz), la durée symbole est de l'ordre de quelques dizaines de nanosecondes. Le filtre numérique modélisant le canal est un vecteur aléatoire dont les variables $h(n)$ sont associées à chaque intervalle de temps $[nT, (n+1)T]$. Chaque variable $h(n)$ est la somme des faisceaux dont les instants d'arrivée sont compris entre nT et $(n+1)T$. En reprenant les hypothèses décrites dans le paragraphe précédent, le filtre numérique modélisant le canal multitrajets est un vecteur aléatoire de loi K et de coefficient $\{\alpha(n), \gamma(n)\}$.

Estimation de la loi de σ^2. Nous proposons d'estimer les paramètres de la loi K par la méthode des moments. A partir des moments d'ordre 1 et 2, l'estimateur du paramètre de forme et du paramètre d'échelle sont obtenus en résolvant

$$\begin{aligned} \widehat{\alpha} &= \min_\alpha \left(\frac{\pi}{4}\frac{\mu_2}{\mu_1}\alpha\Gamma^2\left(\alpha+\frac{1}{2}\right) - \Gamma^2(\alpha)\right) \\ \widehat{\gamma} &= \frac{\mu_2}{\widehat{\alpha}} \end{aligned}$$

Sur les 30 emplacements, nous obtenons pour α une valeur moyenne de $1,08$ et un écart-type de 0.075 alors que pour γ, l'estimation moyenne vaut $1.3\,10^{-6}$ et son écart type $2.11\,10^{-6}$. En première approximation, α est estimé à 1 quelque soit l'emplacement, ce qui signifie que la loi suivie par σ^2 est exponentielle de paramètres γ fonction de l'emplacement considéré. La densité du champ électrique de la composante aléatoire s'écrit alors :

$$f_{|E|}(x) = \frac{4}{\sqrt{\gamma}} K_0\left(\frac{2}{\sqrt{\gamma}}x\right),\ x \in [0, +\infty[$$

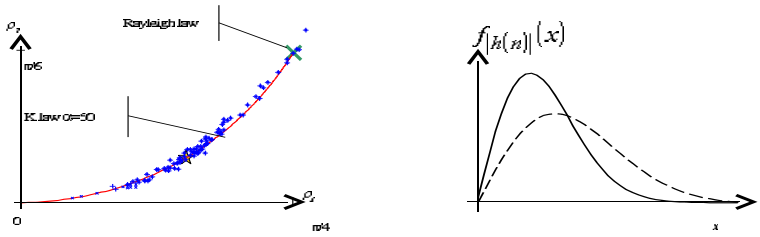

FIG. 6.4 – A gauche : loi K (ligne continue), loi de Rayleigh (une croix à la pointe de la courbe) et coefficients de la réponse impulsionnelle (étoiles) dans le graphe (ρ_1, ρ_2). A droite : Comparaison entre la densité de Rayleigh (courbe en dessous) et la densité de probabilité de la loi K pour $\alpha = 50$ (courbe en dessus), avec les mêmes variances

Remarquons que pour α grand, la loi de l'amplitude du champ électrique converge vers la loi de Rayleigh, ce que confirme son coefficient de variation qui tend vers 0. Néanmoins, la taille des échantillons étant petite (30 valeurs par emplacements), la précision des estimateurs est discutable.

Modèle numérique de canal sélectif en fréquence. Afin d'analyser le modèle de canal numérique de loi K et de le comparer à celui de Rayleigh, nous introduisons deux paramètres fonctions des moments d'ordre 1, 2 et 3 des amplitudes des $|h(n)|$. Ces paramètres, notés ρ_1 et ρ_2, sont invariants par rapport au facteur d'échelle de $\sigma^2(n)$, ils ne dépendent que de la forme de sa densité $\alpha(n)$:

$$\rho_1 = \frac{\mu_1^2}{\mu_2} \text{ et } \rho_2 = \frac{\mu_1^3}{\mu_3}$$

où μ_1, μ_2 et μ_3 sont respectivement les moments d'ordre 1, 2 et 3 de $|h(n)|$:

$$\begin{aligned}
\mu_1 &= I\!E[|h(n)|] = \sqrt{\frac{\pi\gamma(n)}{2}}\frac{\Gamma(\alpha(n)+\frac{1}{2})}{\Gamma(\alpha(n))} \\
\mu_2 &= I\!E[|h(n)|^2] = 2\alpha(n)\gamma(n) \\
\mu_3 &= I\!E[|h(n)|^3] = 3\sqrt{\frac{\pi\gamma(n)^3}{2}}\frac{\Gamma(\alpha(n)+\frac{3}{2})}{\Gamma(\alpha(n))}
\end{aligned}$$

Pour chaque coefficient $|h(n)|$, l'estimation du couple (ρ_1, ρ_2), notée $(\widehat{\rho}_1(n), \widehat{\rho}_2(n))$, est obtenue à l'aide des moments empiriques. Les couples $(\widehat{\rho}_1(n), \widehat{\rho}_2(n))$ sont représentés dans le graphique (ρ_1, ρ_2) (Cf graphique de gauche dans FIG. 6.4) par des points dont la position est comparée à la courbe K.
Le graphique FIG. 6.4 montre l'adéquation des données expérimentales à la ligne K. La plupart des coefficients sont en effet proches de cette courbe. Dans ce graphe, la loi de Rayleigh est représentée par une croix (à la pointe de la courbe). Les points $(\widehat{\rho}_1(n), \widehat{\rho}_2(n))$ étant distants de celui-ci, la distribution de Rayleigh n'est pas adaptée pour modéliser la plupart des coefficients de la réponse impulsionnelle du canal. A titre d'illustration, nous proposons de comparer la densité K à celle de Rayleigh estimée sur l'ensemble de coefficients et des

emplacements. La loi de Rayleigh ayant un coefficient de variation constant, sa moyenne conditionne sa variance. Grâce à son paramètre de forme, la loi K permet de s'affranchir de cette relation améliorant les capacités d'ajustement de sa distribution à celle des données expérimentales. Sur la figure de droite de FIG. 6.4, on remarque une nette différence et en particulier une plus grande dissymétrie de la densité K par rapport à celle de Rayleigh.