



HAL
open science

Sommes et extrêmes en physique statistique et traitement du signal : ruptures de convergences, effets de taille finie et représentation matricielle

Florian Angeletti

► **To cite this version:**

Florian Angeletti. Sommes et extrêmes en physique statistique et traitement du signal : ruptures de convergences, effets de taille finie et représentation matricielle. Autre [cond-mat.other]. Ecole normale supérieure de lyon - ENS LYON, 2012. Français. NNT : 2012ENSL0776 . tel-00779703

HAL Id: tel-00779703

<https://theses.hal.science/tel-00779703>

Submitted on 22 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

en vue d'obtenir le grade de

Docteur de l'École normale supérieure de Lyon - Université de Lyon
Spécialité : Physique

Laboratoire de Physique de l'ENS Lyon
École doctorale PHAST

présentée et soutenue publiquement le 6 décembre 2012 par

Florian Angeletti

Sommes et extrêmes
en physique statistique et traitement du signal :
Ruptures de convergences, effets de taille finie et
représentation matricielle

Sous la direction de :

Eric Bertin et Patrice Abry

Jury

M. Bernard Derrida,	Président du Jury
M. Satya Majumdar,	Rapporteur
M. Hugo Touchette,	Rapporteur
M. Jean-Yves Tournieret,	Rapporteur
M. Eric Bertin,	Directeur de thèse
M. Patrice Abry,	Codirecteur de thèse

Résumé

Cette thèse s'est développée à l'interface entre physique statistique et traitement statistique du signal, afin d'allier les perspectives de ces deux disciplines sur les problèmes de sommes et maxima de variables aléatoires. Nous avons exploré trois axes d'études qui mènent à s'éloigner des conditions classiques (i.i.d.) : l'importance des événements rares, le couplage avec la taille du système, et la corrélation. Combinés, ces trois axes mènent à des situations dans lesquelles les théorèmes de convergence classiques sont mis en défaut. Pour mieux comprendre l'effet du couplage avec la taille du système, nous avons étudié le comportement de la somme et du maximum de variables aléatoires indépendantes élevées à une puissance dépendante de la taille du signal. Dans le cas du maximum, nous avons mis en évidence l'apparition de lois limites non standards. Dans le cas de la somme, nous nous sommes intéressés au lien entre effet de linéarisation et transition vitreuse en physique statistique. Grâce à ce lien, nous avons pu définir une notion d'ordre critique des moments, montrant que, pour un processus multifractal, celui-ci ne dépend pas de la résolution du signal. Parallèlement, nous avons construit et étudié, théoriquement et numériquement, les performances d'un estimateur de cet ordre critique pour une classe de variables aléatoires indépendantes. Pour mieux cerner l'effet de la corrélation sur le maximum et la somme de variables aléatoires, nous nous sommes inspirés de la physique statistique pour construire une classe de variable aléatoires dont la probabilité jointe peut s'écrire comme un produit de matrices. Après une étude détaillée de ses propriétés statistiques, qui a montré la présence potentielle de corrélation à longue portée, nous avons proposé pour ces variables une méthode de synthèse en réussissant à reformuler le problème en termes de modèles à chaîne de Markov cachée. Enfin, nous concluons sur une analyse en profondeur du comportement limite de leur somme et de leur maximum.

Abstract

This thesis has grown at the interface between statistical physics and signal processing, combining the perspectives of both disciplines to study the issues of sums and maxima of random variables. Three main axes, venturing beyond the classical (i.i.d) conditions, have been explored : The importance of rare events, the coupling between the behavior of individual random variable and the size of the system, and correlation. Together, these three axes have led us to situations where classical convergence theorems are no longer valid. To improve our understanding of the impact of the coupling with the system size, we have studied the behavior of the sum and the maximum of independent random variables raised to a power depending of the size of the signal. In the case of the maximum, we have brought to light non standard limit laws. In the case of the sum, we have studied the link between linearisation effect and glass transition in statistical physics. Following this link, we have defined a critical moment order such that for a multifractal process, this critical order does not depend on the signal resolution. Similarly, a critical moment estimator has been designed and studied theoretically and numerically for a class of independent random variables. To gain some intuition on the impact of correlation on the maximum or sum of random variables, following insights from statistical physics, we have constructed a class of random variables where the joint distribution probability can be expressed as a matrix product. After a detailed study of its statistical properties, showing that these variables can exhibit long range correlations, we have managed to recast this model into the framework of Hidden Markov Chain models, enabling us to design a synthesis procedure. Finally, we conclude by an in-depth study of the limit behavior of the sum and maximum of these random variables.

Table des matières

Résumé	iii
Abstract	iv
Table des matières	v
Introduction	1
Introduction	1
I De la somme aux extrêmes	9
Introduction	11
Introduction	11
1 Sommes et extrêmes	13
1 Sommes	13
2 Maximum	16
3 Rupture de convergence pour les lois de puissance	18
3.1 Convergence logarithmique	18
3.2 Modèle REM dégénéré et transition vitreuse	20
2 Renormalisation des extrêmes	23
1 Définition du problème	23
2 Approche par renormalisation	24
2.1 Transformation de renormalisation et conditions de normalisation	24
2.2 Équation aux dérivées partielles du flot	25
2.3 Distributions stables	26
2.4 Cas $Q = 0$: redécouverte des distributions limites standards .	26
2.5 Cas $Q \neq 0$: émergence de distributions stables non standards	27
3 Domaine d'attraction des lois non standards	30
3.1 Comportement asymptotique des paramètres de forçage	30
3.2 Modifier le domaine d'attraction par le biais de $q(s)$	32
4 Approche alternative	36
4.1 Facteurs de normalisation du maximum transformé	36
4.2 Exposants q_n divergents	37
4.3 Exposant q_n s'annulant à l'infini	38
4.4 Exposants q_n de limite finie	38

5	Conclusion	39
II	Ordre critique et transitions de phases	41
	Introduction	43
	Introduction	43
3	Classes de variables aléatoires	51
1	Classe A : Variables i.i.d.	51
1.1	Type II : Lois irrégulières de moments finis	51
1.2	Interprétation	53
1.3	Conséquences	53
1.4	Exposant local de lois de puissance	53
1.5	Exemples	54
2	Classe B : processus multifractals	55
2.1	Cascades de Poisson composées	56
2.2	Variation d'échelle et propriétés multifractales	57
2.3	Propriétés de grandes déviations	58
3	Approche unifiée	60
4	Ordre critique	61
1	Effet de concentration	61
1.1	Classe A-II	61
1.2	Classe B	63
2	Effet de taille finie	64
2.1	Région accessible à taille finie	64
2.2	Moments tronqués	66
3	Transition de phase et moment critique	68
3.1	Frontière critique	68
3.2	Ordre critique pour la classe A-II	69
3.3	Ordre critique pour la classe B	71
4	Comportement asymptotique et REM	72
4.1	Classe A-II	72
4.2	Classe B	74
5	Rapport de participation en analyse multifractale	76
5	Estimation	79
1	Estimation de θ	79
1.1	Définition	79
1.2	Performances : analyse théorique	80
1.3	Performances : analyse numérique	82
2	Estimation de $\rho_l(y_m(n))$	85
2.1	Définition	85
2.2	Performances : analyse théorique	86
2.3	Performances : analyse numérique	86
3	Estimation de q^*	86

3.1	Définition	86
3.2	Performances : analyse numérique	89
4	Séries temporelles corrélées	89
4.1	Analyse théorique	91
4.2	Impact des dépendances sur \hat{q}^*	92
4.3	Estimateurs modifiés	93
Conclusions et discussion		97
Conclusions		97
5	Classe A	97
6	Classe B	98
III Représentation matricielle		101
Introduction		103
Introduction		103
6 Caractérisation théorique		107
1	Probabilité jointe et produit de matrices	107
1.1	Définition et remarques générales	107
1.2	Marginale, moments et dépendance	108
1.3	Condition de stationnarité	109
2	Structure de dépendance	111
2.1	Matrice de structure diagonalisable	111
2.2	Matrice de structure non diagonalisable	112
7 Synthèse		115
1	Motivation	115
2	Modèle par chaîne de Markov cachée	116
3	Conception pratique	119
4	Illustrations	121
4.1	Série temporelle stationnaire	121
4.2	Vecteur aléatoire	125
4.3	Exemple en dimension infinie	129
8 Sommes et extrêmes revisités		133
1	Motivation	133
2	Équations matricielles	133
3	Modèle à chaîne de Markov cachée	135
4	Chaînes irréductibles	137
4.1	Chaînes homogènes	137
4.2	Chaînes non homogènes	138
5	Chaîne réductible simple à deux états	139
5.1	Maximum	141
5.2	Lois limites pour le maximum	142
5.3	Normalisation de la somme	144

5.4	Lois limites pour la somme	145
5.5	Rupture de la loi des grands nombres	146
6	Chaînes réductibles simples à d états	146
6.1	Définitions	146
6.2	Séparation des transitions	147
6.3	Distribution limite le long d'un chemin irréversible	149
7	Cas général	151
8	Conclusion	156
 Conclusion générale		161
Appendices		165
A Fonctions lentement variables		167
B REM (<i>Random Energy Model</i>)		169
C Convergence du maximum transformé		173
1	Domaine Gumbel	174
2	Domaine Weibull	175
3	Domaine Fréchet	175
D Évaluation par la méthode du col de $\mathbb{E}[X^q]$		177
E Moments tronqués et estimateur des moments		179
1	Classe A	179
2	classe B	183
F Estimation de θ		187
1	Construction	187
2	Consistance	188
3	Normalité asymptotique	188
Liste de publications		191
Bibliographie		193

Introduction

La théorie des probabilités a parcouru un long chemin depuis sa modeste origine au sein du "Liber de ludo aleae" de Jérôme Cardan [29]. Si les jeux de hasard furent longtemps leur principale application, au fil des siècles, les probabilités et le hasard se sont forgés une place de choix au sein des sciences. Ce fut un long cheminement ponctué de tiraillements philosophiques sur la place du hasard et du déterminisme dans le monde scientifique. Ainsi Laplace, déterministe acharné, pensait que sa théorie des probabilités [37] ne serait qu'un palliatif temporaire aux lacunes scientifiques de son temps. L'aurore du vingtième siècle avec la mécanique quantique, la théorie du chaos ou la physique statistique révéla que cette conviction était déraisonnablement optimiste.

Peut-être le hasard n'existe-t-il pas en-soi. Il n'est pas sûr que cette querelle puisse jamais être tranchée. Néanmoins, notre ignorance et les limites afférentes à notre esprit humain sont, elles, bien réelles. De l'acceptation de ces limites jaillit le hasard et la théorie des probabilités, donnant naissance à un des concepts scientifiques les plus féconds et ce dans presque tous les domaines de la science. Que ce soit en physique théorique avec la physique statistique, en traitement de l'information avec les applications du traitement statistique du signal, mais aussi dans les sciences sociales à travers les innombrables usages des statistiques, le hasard a définitivement gagné sa place dans les sciences modernes.

Physique statistique et traitement statistique du signal

La théorie des probabilités a grandi en symbiose avec de nombreux autres champs scientifiques tels que la physique statistique, le traitement du signal, les statistiques. Malheureusement, du fait du cloisonnement des disciplines scientifiques, ces relations se sont souvent développées de façon ponctuelles et limitées. D'un côté, la physique statistique s'efforce à réduire la complexité inhérente à la description de systèmes physiques en dérivant des observables macroscopiques simples à partir d'états microscopiques inaccessibles à l'observation. De l'autre, le traitement du signal s'attelle à extraire des informations significatives à partir de signaux complexes et bruités. Les similitudes entre ces deux approches sont visibles. Nonobstant la différence de langage, la physique statistique et le traitement du signal se retrouvent confrontés à une même difficulté : apprivoiser le hasard pour obtenir des résultats quasi-déterministes. À partir de cette nécessité commune, les techniques adoptées et les problèmes considérés divergent considérablement. Mais cette diversité de méthodes et d'approches crée justement une occasion d'enrichir les deux domaines.

L'interface entre physique statistique et traitement du signal est trop vaste pour une simple thèse. Je me concentrerai donc sur un sujet aussi fondamental que fas-

cinant : les questions de convergences de sommes et maxima de variables aléatoires et leurs applications. Les questions liées à la convergence des sommes de variables aléatoires furent parmi les premières à intéresser les mathématiciens et les parieurs. Un des résultats des plus naturels de la théorie de la probabilité, formulé dès 1713 par Jacques Bernoulli [16], est la loi des grands nombres. Elle affirme simplement que si on observe la moyenne de n variables aléatoires indépendantes identiquement distribuées (i.i.d.) et de moyenne finie,

$$\frac{1}{n} \sum_{i=1}^n X_i, \quad (1)$$

plus le nombre n d'échantillons augmente, plus grande est la probabilité que S se rapproche de son espérance théorique $\mathbb{E}[X]$. Cette loi peut sembler évidente, sa démonstration l'est moins ([18, 45]) et requiert que la loi des variables aléatoires X_i soit suffisamment *régulière* afin que l'espérance théorique $\mathbb{E}[X_i]$ soit finie. Cette loi des grands nombres est un concept fondamental tant en physique statistique qu'en traitement statistique du signal. L'hypothèse de base de la physique statistique n'est-elle pas que l'on peut décrire précisément le comportement macroscopique d'un système à partir du comportement microscopique moyen ? En d'autres termes, la puissance descriptive de la physique statistique repose en grande partie sur la loi des grands nombres. Parallèlement, en traitement statistique du signal, l'idée même de l'estimation de statistiques est bâtie sur la possibilité d'obtenir une mesure quasi-déterministe à partir d'un grand nombre d'observations d'une variable aléatoire. L'estimateur des moments (pour des variables aléatoires positives)

$$S(n, q) = \frac{1}{n} \sum_{i=1}^n X_i^q \quad (2)$$

est ainsi un exemple célèbre d'estimateur dont la consistance est une conséquence directe de la loi des grands nombres.

Les raffinements de la loi des grands nombres sous la forme du théorème central limite, ou de la loi des logarithmes itérés [45], ont des conséquences plus retentissantes encore. L'omniprésence des lois gaussiennes tant en physique statistique qu'en traitement du signal peut s'interpréter comme une conséquence directe de ce théorème et du fait que nombre d'observables physiques ou statistiques sont le résultat d'agrégations additives d'un grand nombre de phénomènes "microscopiques" sous-jacents. Ces résultats plus fins permettent également de quantifier les vitesses de convergences des estimateurs de moments en statistiques.

Ces questions de sommes de variables aléatoires prennent une tournure encore plus intéressante dans le contexte de la physique statistique des systèmes désordonnés. Dans la physique statistique des systèmes à l'équilibre, un des objets les plus fondamentaux est la fonction de partition Z [43, 93]

$$Z(\beta) = \sum_{s \in \Omega} e^{-\beta E_s}, \quad (3)$$

où la somme porte sur l'ensemble des micro-états s . Cette fonction de partition dépend de l'inverse de la température β et des énergies E_s associées aux micro-états s de l'univers Ω . Il s'agit d'une fonction déterministe qui permet d'étudier le comportement des observables macroscopiques sous l'effet des fluctuations thermiques

et d'en déduire certaines propriétés physiques du système. Deux grandeurs dérivées de Z sont particulièrement dignes d'intérêts, l'énergie libre F

$$F = -\frac{\ln Z}{\beta} \quad (4)$$

qui caractérise la distribution des états dans un système canonique (à température et nombre de particule fixés) et l'entropie S

$$S = k_B \beta^2 \partial_\beta F \quad (5)$$

qui constitue une mesure de la dispersion de la distribution des états, i.e., une mesure du désordre d'information associée à cette distribution [43, 93].

Dans le cadre des systèmes désordonnés, on rajoute un second degré de fluctuations "gelées" en sus des fluctuations thermiques [104, 86]. Ce second niveau d'aléa permet d'étudier l'effet des fluctuations structurelles des milieux complexes que sont les verres de spin. Dans un tel système désordonné, les énergies des configurations ne sont plus déterministes, la fonction de partition Z devient donc une somme de variables aléatoires qui présente de troublantes similitudes avec l'estimateur des moments. Si on pose $X = \exp Y$, l'estimateur des moments peut se réécrire comme

$$S(n, q) = \frac{1}{n} \sum_{i=1}^n e^{qY_i}. \quad (6)$$

Il existe donc une analogie formelle entre l'estimateur des moments S et la fonction de partition Z d'un système désordonné. Dans cette analogie, l'ordre des moments q joue un rôle similaire à la température inverse β .

Pour poursuivre cette analogie, un des modèles les plus célèbres de la physique des systèmes désordonnés, le *Random Energy Model* (REM [40, 86]) s'est avéré être particulièrement fructueux. Ce modèle est une illustration simple du scénario de champ moyen pour la transition vitreuse dans les verres de spin [40] ou les liquides en surfusion [71, 26]. Dans ce modèle, toutes les configurations s possèdent une énergie indépendante E_s , issues de la même distribution. Ces énergies sont gelées, i.e., elles n'évoluent pas en fonction du temps. Dans la version originelle du modèle [40], ces variables E_s sont tirées à partir d'une distribution gaussienne. Un tel choix mène à des statistiques log-normales pour les variables $X_k = \exp(-\beta E_k)$. Le REM a été cependant étendu à d'autres distributions, telle que les lois log-quasi-exponentielles étirées, obtenues en choisissant une distribution Y_k dont la queue est de la forme $p(y) \underset{y \rightarrow +\infty}{\sim} \exp(-cy^\rho)$, avec $\rho > 1$ et $c > 0$ [27]. L'intérêt principal du REM provient de l'existence d'une transition vitreuse à température finie, et de la facilité avec laquelle cette transition peut être étudiée analytiquement [40]. Or l'existence de cette transition de phase vitreuse implique que la loi des grands nombres ne s'applique pas dans le cas du REM. Cela est dû à la combinaison de deux effets différents. Premièrement, dans le cas du REM, la distribution des énergies dépend de la taille du système. De plus, la distribution des variables X_k est *irrégulière* : la fonction caractéristique d'une loi log-normale n'est pas analytique en 0, ce qui crée un certain nombre de complications. Notamment, ces lois *irrégulières* ne peuvent pas toujours être entièrement caractérisée par leurs moments entiers [8]. Dans ces circonstances, les hypothèses sur lesquelles reposent la loi des grands nombres ne sont plus vérifiées,

et la transition vitreuse du *REM* démontre qu'il est possible de transgresser alors la loi des grands nombres. Si on revient à l'estimateur des moments S , cela signifie que si on couple la forme des distributions avec la taille n du signal, la consistance de l'estimateur n'est plus assurée : il est possible que l'estimateur converge vers une valeur différente du moment théorique. Une possibilité simple pour réaliser ce couplage est de considérer que l'ordre des moments q que l'on étudie dépend de la taille du signal n . On retrouve alors le même type de dépendance avec la taille du signal que dans le *REM*. Ben Arous et ses collaborateurs [10] ont montré qu'il est nécessaire de fortement modifier le théorème central limite dans ces conditions et, comme le suggérait l'analogie avec le *REM*, la loi des grands nombres ne demeure valide que pour certaines formes de $q(n)$.

Il s'agit là d'une limite fondamentale de la loi des grands nombres ou encore du théorème central limite. Ces théorèmes requièrent que les variables aléatoires considérées soient régulières, indépendantes et identiquement distribuées. On pourrait penser qu'il s'agit d'hypothèses raisonnables. Si nous vivions dans un univers parfaitement stationnaire, sans interactions, sans événements exceptionnels, tel serait le cas. Heureusement, notre univers n'est ni si terne ni si raisonnable. Sans corrélation, il est impossible de considérer des systèmes physiques en interaction. De même, supposer que des signaux réels ne sont jamais corrélés est une affirmation fort audacieuse. On retrouve aussi nombre de systèmes en physique statistique dont les distributions de probabilité font intervenir des lois de puissance, sans même parler des problèmes d'estimation des lois de puissance en traitement du signal ; on ne peut donc se satisfaire des lois régulières. Bien que le caractère identiquement distribué puisse paraître plus consensuel, il entraîne pourtant un nécessaire découplage entre les distributions des variables aléatoires et la taille n du système considéré.

Bien sûr, il existe des extensions des résultats de convergence dans certains cas spécifiques et contrôlés, mais il s'agit encore d'exceptions plutôt que de résultats généraux [18, 63]. Ces contraintes limitent considérablement l'application de ces résultats aux systèmes réels. Cela d'autant plus que ces difficultés apparaissent rarement seules. Par exemple, dans le cadre de l'analyse multifractale [82, 49], on se retrouve confronté à des variables aléatoires fortement corrélées, irrégulières, et dont le comportement dépend de manière régulière de l'échelle d'analyse. Similairement en physique statistique hors d'équilibre, les systèmes considérés sont souvent non homogènes et fortement corrélés. Les modèles de physique des systèmes désordonnés comme le *REM* présentent, eux, un couplage non trivial avec la taille du système (dans l'approximation de champ moyen) associé avec des lois irrégulières.

Problèmes et contributions

Il existe donc de nombreuses situations, physiquement et statistiquement intéressantes, dans lesquelles les résultats classiques de la théorie des probabilités s'avèrent insuffisants. Réciproquement, ces problèmes de convergences, qui peuvent amener à violer la loi des grands nombres, sont des signes clairs de l'existence de phénomènes non triviaux et potentiellement intéressants. Dans cette thèse, nous distinguerons trois causes potentielles de ces problèmes de convergence de la somme : l'importance des événements rares, le couplage avec la taille du système, et, plus classiquement, la corrélation. Ces trois sources de difficultés mais aussi de richesse de comportement

seront les trois axes qui structurent cette thèse. Afin de pouvoir mener à bien nos études, nous essaierons d'introduire parcimonieusement ces complications dans les problèmes que nous étudierons, tout en essayant de préserver le plus possible les structures mathématiques qui ont rendus possibles l'étude exhaustive des problèmes de convergence de la somme et du maximum dans le cas i.i.d..

De la somme aux extrêmes

Si on s'intéresse aux événements rares et aux lois irrégulières, le maximum

$$M(n) = \max\{X_1, \dots, X_n\} \quad (7)$$

est une quantité naturelle à étudier, qui sera le protagoniste de notre partie I. En effet, pour des lois irrégulières, les événements extrêmes jouent un rôle non négligeable dans le comportement de la somme S (cf. chapitre 1). De plus, sommes et maxima de variables aléatoires partagent une structure mathématique commune qui permet bien souvent de transférer des raisonnements d'un cas à l'autre. Il est ainsi peu surprenant que l'on retrouve des résultats comparables lorsqu'on s'intéresse à la convergence du maximum d'un jeu de variables aléatoires. Comme dans le cas des sommes de variables aléatoires, il existe une caractérisation impressionnante du comportement asymptotique de ce maximum de variables aléatoires [59, 53]. Les lois limites associées à ces maxima appartiennent à une famille restreinte de distributions appelées lois des extrêmes généralisées. Nous ferons brièvement le point sur ces résultats de convergence dans le chapitre 1, afin de présenter les outils mathématiques communément utilisés pour l'étude des sommes et des maxima de variables aléatoires. En nous inspirant de l'analogie avec le REM, nous montrerons que l'équivalent de l'énergie libre par spin, la log-moyenne

$$\frac{\ln S(n, q = 1)}{\ln n} \quad (8)$$

est fort utile pour résumer le comportement de la convergence de la somme et du maximum de lois de puissance. Cependant, dans le cas du maximum, ce comportement de convergence demeure très régulier, quelles que soient les variables aléatoires considérées. Il est donc intéressant d'utiliser notre second axe de questionnement et d'introduire un couplage avec la taille du système. Dans le cas de la somme, nous avons naturellement considéré le cas de sommes de variables élevées à une puissance dépendant de la taille du signal. La généralisation au maximum se fait aisément

$$M(n, q) = \max\{X_1^q, \dots, X_n^q\} = \max\{X_1, \dots, X_n\}^q. \quad (9)$$

Il est même plus facile d'étudier le comportement du maximum élevé à une puissance $M(n, q)$. En effet, le maximum et la puissance commutent entre eux, ce qui n'est pas le cas pour la somme $S(n, q)$. L'élévation à la puissance préserve donc la structure mathématique associée au maximum, ce qui facilite l'étude des lois limites du maximum.

Bogachev et ses collaborateurs ont montré [23] qu'il est alors possible d'accéder à des lois limites non standard dans certaines situations. En suivant une méthode par renormalisation développée pour le cas classique [61, 62, 17], nous avons pu classifier

l'ensemble des lois limites non-standards et détailler le mécanisme de transition entre les lois standards et non-standards dans le chapitre 2. De façon très intéressante, cette convergence vers des lois limites non indépendantes dépend du rapport entre l'ordre de perturbation q_n et un ordre critique q_n^* [3].

Ordre critique et estimateur des moments

L'élévation des variables aléatoires à une puissance dépendant de la taille du signal peut donc faire apparaître des comportements de convergence non-standards dans le cas du maximum. Si on revient au cas de la somme, les résultats de Ben Arous [10] confirme l'existence d'un effet similaire lorsqu'on s'intéresse à la somme de variable aléatoires i.i.d. élevées à une puissance dépendant de la taille du système pour des variables aléatoires suivant des lois irrégulières. Nous intéresserons donc plus en détail dans la partie II à ce changement de comportement, que l'on peut assimiler à une transition vitreuse si on poursuit l'analogie avec le REM. Cela revient à étudier le comportement de S le long d'une courbe $(n, q(n))$. L'interaction entre somme et puissance est moins harmonieuse, il est donc plus difficile de dériver un comportement exact. Cependant, inspiré par la physique du REM, nous montrerons dans le chapitre 4 qu'il apparaît une compétition entre deux effets distincts : un effet de taille finie et un effet de concentration. La compétition entre ces deux effets fait apparaître un ordre critique q^* qui décrit jusqu'à quel ordre des moments l'estimateur converge effectivement vers le moment théorique.

Au-delà de cet ordre critique, l'estimateur ne converge plus vers la valeur théorique, et il s'avère que le comportement typique de l'estimateur des moments est correctement décrit par la valeur d'un moment tronqué

$$T(n, q) = \mathbb{E} \left[\begin{cases} X & \text{si } X < x_n^\dagger \\ 0 & \text{sinon} \end{cases} \right] \quad (10)$$

où x_n^\dagger correspond à la limite supérieure du domaine pouvant être exploré à partir du nombre fini d'observations disponibles. Ce résultat est assez intuitif physiquement, et sa traduction dans le langage mathématique est assez élégante, puisque celle-ci entrelace le comportement de la log-moyenne avec les moments tronqués :

$$\frac{\ln S(n, q(n))}{\ln n} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \lim_{n \rightarrow +\infty} \frac{\ln T(n, q(n))}{\ln n} \quad (11)$$

La log-moyenne confirme son utilité pour une estimation grossière mais très robuste du comportement de la somme de variables aléatoires. À partir de là, en poursuivant l'analogie avec le modèle REM, on peut calculer quantitativement la position du moment critique au-delà duquel l'estimation des moments théoriques est impossible. Il est presque fascinant que ce moment critique soit l'analogue exact de la température critique de transition vitreuse du REM.

On peut aller plus encore dans l'application de cette analogie entre physique des systèmes désordonnés et estimation des moments. Il est ainsi possible de généraliser cette analyse de l'ordre critique des moments au cas multifractal. Cela nous a permis d'obtenir une explication physique de l'effet de linéarisation en l'analyse multifractale. Celui-ci résulte d'une compétition tripartite entre la corrélation, le couplage

avec l'échelle d'analyse et le caractère irrégulier des lois impliquées. Le résultat de ce conflit d'intérêts acharné est l'apparition d'un ordre critique intrinsèque, qui ne dépend que des caractéristiques statistiques du processus multifractal. Au-delà de cet ordre critique, il est impossible d'évaluer correctement le moment théorique avec l'estimateur des moments quelle que soit l'échelle d'analyse [4].

De plus, il s'avère que dans le cas indépendant, cet ordre critique peut être estimé empiriquement à partir d'observations de données réelles. Intrigué par la possibilité de mesurer cet analogue de la température de transition vitreuse dans des signaux réels, j'ai construit un estimateur pratique de ce moment critique. Le chapitre 5 développe en détail l'analyse des performances de cet estimateur. Celui-ci a été testé sur des signaux synthétiques, soit parfaitement indépendants, soit corrélés avec une corrélation exponentielle [1]. Les performances ouvrent des perspectives encourageantes d'applications pratiques.

Représentation matricielle

Le traitement de la corrélation dans la partie II demeure empirique, que ce soit dans l'analyse des performances de l'estimateur du moment critique ou l'approche heuristique de l'effet de la corrélation sur l'effet de taille finie dans le cadre de l'analyse multifractale. Cela n'est pas forcément surprenant, l'ensemble des lois non indépendantes est vaste, et étudier simultanément le jeu d'interactions entre corrélation, couplage à taille finie et lois irrégulières est une entreprise trop ambitieuse. Afin de progresser plus posément, nous présenterons dans la partie III un modèle synthétique de variables aléatoires corrélées.

Puisque nous nous intéressons aux problèmes de convergence dans les sommes et les extrêmes, il est intéressant de considérer des modèles qui préservent le plus possible les structures algébriques des lois sommes et extrêmes, cela afin de pouvoir adapter facilement les idées développées dans le cas indépendant. Ces contraintes sont assez floues, cependant il s'avère que la mécanique quantique [47, 7] et la physique des systèmes hors d'équilibre ont déjà développé un modèle répondant à ces contraintes en utilisant une représentation matricielle des probabilités jointes. Un des exemples les plus canoniques de ces représentations en physique hors d'équilibre est le modèle ASEP (*Asymmetric Simple Exclusion Process*, [41, 81, 22]). Cette forme de loi préserve un certain nombre de propriétés structurelles des lois indépendantes à cause de sa structure de produit, tout en introduisant de la corrélation à cause des propriétés de non-commutativité des matrices. Ces propriétés théoriques permettent de calculer aisément la plupart des grandeurs statistiques classiques. Ces solutions exactes ont permis d'obtenir une description très précise de certains modèles de physique théorique. Cependant, ces approches, très spécifiques aux problèmes physiques étudiés, n'avaient pas encore été complétées par une approche plus générique centrée sur l'analyse des propriétés statistiques de ce type de distribution. Nous profiterons donc du chapitre 6 pour établir une caractérisation détaillée des lois à représentation matricielle de dimension finie. Nous montrerons ainsi qu'il existe deux classes distinctes de représentations matricielles associées à des structures de corrélations très différentes.

Si les propriétés théoriques de ces lois à représentation matricielle sont assez directes à dériver, il est toutefois plus difficile d'en déduire immédiatement une mé-

thode de synthèse. Cependant, il est possible de reformuler ces lois à représentations matricielles comme des modèles à chaîne de Markov cachée particuliers. À partir de cette reformulation, j'ai pu construire une méthode de synthèse efficace et flexible pour ces variables aléatoires, détaillée dans le chapitre 7. Cet algorithme de synthèse permet notamment de choisir de façon quasi-indépendante les lois marginales, la forme de la structure de la corrélation, avec cependant une contrainte forte sur la forme de la corrélation [2].

Pour conclure, nous reviendrons au comportement asymptotique de la somme et du maximum au chapitre 8, profitant pleinement de l'intuition obtenue dans la construction d'un algorithme de synthèse. Nous décrirons notamment une rupture de la loi des grands nombres dans le cadre de ces représentations matricielles. Cette rupture est due à l'influence simultanée de la corrélation, du caractère non stationnaire et d'un couplage avec la taille du signal.

Première partie
De la somme aux extrêmes

Introduction

La statistique des extrêmes, c'est-à-dire la statistique des plus grandes valeurs dans un jeu de variables aléatoires, a attiré bien des attentions dans de nombreux domaines, de la théorie des probabilités à la physique – où des interactions fécondes avec les systèmes désordonnés [27, 30, 51, 52] ainsi que les marches aléatoires et les fluctuations d'interfaces ont récemment fleuri [60, 75, 78]– en passant par l'hydrologie [70], la finance [45, 76] ou l'ingénierie [59]. Pour des variables aléatoires indépendantes et identiquement distribuées, les distributions asymptotiques sont connues depuis longtemps [48, 55, 59, 53]. Suivant la forme de la queue de la distribution parente (i.e., la distribution utilisée pour générer le jeu de variables aléatoires), émergent trois distributions différentes. Pour les distributions parentes dont la queue décroît plus rapidement que toute loi de puissance, la distribution limite est la célèbre loi de Gumbel (loi qui apparaît d'ailleurs dans d'autres applications en dehors du domaine des valeurs extrêmes [34]). Si la distribution parente décroît de manière similaire à une loi de puissance à l'infini, on obtient une distribution de Fréchet, au contraire si cette décroissance s'avère algébrique près d'une borne supérieure, apparaît alors une distribution de Weibull.

Sommes et extrêmes

Ces résultats sont fort semblables aux résultats de convergence de la somme de variables aléatoires vers les lois stables. Cette similitude s'explique en grande partie par l'existence d'une structure mathématique commune entre somme et maximum de variables aléatoires. Cependant, cette mise en miroir n'est pas parfaite et il existe des différences irrémédiables entre somme et maximum. Nous profiterons donc de l'ouverture de cette thèse pour rappeler brièvement les différents résultats de convergence associés aux maximum et somme de variables aléatoires afin de mettre en lumière ces similitudes et différences. Nous nous pencherons plus en détail sur le comportement des facteurs de normalisation de la somme et du maximum lorsqu'on s'intéresse aux lois de puissance. Dans le cas de la somme, ces facteurs de normalisation présentent une rupture de comportement très intéressante qui survient lorsque l'espérance de la loi de puissance devient infinie. Nous rappellerons notamment que ce changement de comportement peut être interprété comme une transition vitreuse d'un modèle REM (Random Energy Model [40]) dégénéré similaire au modèle de pièges [25]. Ce phénomène de transition vitreuse dégénérée est cependant spécifique à la somme et n'existe pas lorsqu'on étudie le maximum. Il est donc tentant d'enquêter plus avant sur cette intrigante différence entre somme et maximum. On peut notamment se demander s'il est possible de provoquer une telle transition vitreuse dans le maximum en introduisant une perturbation du maximum dépendante de la

taille du système.

Extrêmes et transformation en puissance

Or, il a été récemment montré par Ben Arous et ses collaborateurs [10] que la distribution limite de sommes de variables aléatoires peut être modifiée en élevant les termes de la somme à une puissance divergeant avec le nombre de termes. De telles perturbations puisent leur motivation au sein de la physique des systèmes désordonnés. En effet de telles sommes peuvent être interprétées comme la fonction partition du REM (non-dégénéré cette fois), un des systèmes désordonnés les plus simples –qui a déjà donné lieu à des développements récents en lien avec la statistique des extrêmes [30, 51, 52].

Il est donc naturel de s'enquérir si de telles transformations, c'est-à-dire l'élévation d'une variable aléatoire à une puissance dépendante de la taille du signal, ne peuvent donner naissance à des distributions limites non classiques dans le cas des extrêmes. En termes du modèle REM, cela revient à considérer la statistique du maximum des poids de Boltzmann (qui, sommés, donnent la fonction de partition). Un exemple physique plus concret serait sans doute la statistique des temps de piégeage des particules dans un modèle de pièges [25], dans lequel les particules sont piégées dans des puits d'énergie et ne peuvent s'échapper que par activation thermique. Ces temps extrêmes sont connus pour jouer un rôle important dans ce contexte, et, dans la limite des basses températures et des grands nombres de pièges, la statistique du temps de piégeage maximum pourrait dépendre de l'interaction entre ces deux limites.

Ce problème du maximum transformé par une puissance fut récemment abordé dans la littérature mathématique [23], à la suite des travaux de Ben Arous et ses coauteurs sur le problème des sommes, obtenant des résultats précis à propos de la transition entre le domaine d'attraction des lois Gumbel et Fréchet pour une classe particulière de distributions. Dans un autre registre, ce problème est également lié avec la question de l'existence de distributions limites alternatives pour des procédures de normalisation utilisant des fonctions de puissance [98, 88, 57, 100, 28].

Dans le chapitre 2, nous abordons le problème général de la distribution limite du maximum d'un jeu de variables aléatoires élevées à une puissance dépendant de la taille de l'échantillon. Pour cela, nous généralisons l'approche par groupe de renormalisation introduite récemment pour traiter les effets de taille finie dans le cas de la statistique des extrêmes classiques [61, 62, 17]. Pour des exposants variant avec la taille du système, nous avons découvert des distributions limites non standards qui s'avèrent dériver des lois standards par un changement exponentiel de variables. Nous clarifierons cette surprenante relation par la suite, grâce à un argument simple reposant sur le comportement des facteurs de normalisation.

Chapitre 1

Des sommes aux extrêmes : mise en miroir

1 Sommes

Les sommes de variables aléatoires constituent une des quantités les plus naturelles à étudier dans le cadre de la théorie des probabilités. Après tout, les prémisses de la théorie abstraite des probabilités naquirent autour de la question de la convergence des sommes de variables aléatoires de type Bernoulli vers une loi gaussienne [38]. Nombre de théorèmes furent élaborés pour répondre à cette question originelle ouvrant la voie au large panorama de la théorie actuelle.

Trois résultats fondamentaux joueront un rôle particulièrement important au fil de cette thèse. En premier lieu, la loi forte des grands nombres établit ce qui pourrait sembler être une évidence : la moyenne empirique de variables aléatoires de même loi tend presque sûrement vers l'espérance de ces variables aléatoires lorsque le nombre d'observations croît.

Théorème 1. *Loi forte des grands nombres.*

Soit un ensemble de variables aléatoires i.i.d. $(X_i)_{i \in \mathbb{N}}$, si $|\mathbb{E}[X_i]| < +\infty$ alors

$$S_n \equiv S(n, q = 1) = \frac{\sum_{i=1}^n X_i}{n} \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[X]. \quad (\text{I.1})$$

La preuve de ce résultat remarquable s'appuie sur trois points-clefs : premièrement, l'existence de $\mathbb{E}[X]$ limite les possibilités d'apparition de valeurs atypiques de X . Deuxièmement, en l'absence de corrélation, une fois ces valeurs atypiques sous contrôle, il est possible d'utiliser l'inégalité de Bienaymé-Tchebychev pour quantifier plus précisément la valeur *typique* de $S(n, q = 1)$. Un troisième point important réside dans l'interdépendance des valeurs de $S(n, q = 1)$ pour des valeurs de n proches l'une de l'autre.

Un des raffinements potentiels de cette loi forte des grands nombres s'exprime au sein du théorème central limite qui détaille le comportement de la somme autour de sa valeur moyenne, ce au prix d'une condition légèrement plus forte.

Théorème 2. *Théorème central limite.*

Soit un ensemble de variables aléatoires i.i.d. X_i telles que $\mathbb{E}[X_i^2] < +\infty$. Alors

$$\frac{\sum_{i=1}^n X_i - n\mathbb{E}[X_i]}{\sqrt{n\text{Var}[X_i]}} \xrightarrow{\mathcal{D}} \mathcal{N}_{0,1}. \quad (\text{I.2})$$

Un des outils très utiles pour démontrer ce théorème central limite est la fonction caractéristique ϕ

$$\phi_X(z) = \mathbb{E} \left[e^{-izX} \right]. \quad (\text{I.3})$$

Cette fonction caractéristique possède quatre propriétés extrêmement intéressantes lorsqu'on étudie les propriétés de convergence de la somme. En premier lieu, elle détermine de manière unique la loi des variables aléatoires. Si on a $\phi_X = \phi_Y$ alors X et Y sont de même loi. Ensuite pour deux variables aléatoires i.i.d. X et Y ,

$$\phi_{X+Y}(z) = \phi_X(z)\phi_Y(z). \quad (\text{I.4})$$

L'application $X \mapsto \phi_X$ est donc un morphisme¹ entre l'addition de variables aléatoires et la multiplication de fonctions caractéristiques. De plus, le théorème de continuité de Lévy montre que ce morphisme se comporte bien avec le passage à la limite. Si on considère une suite de variables aléatoires X_n dont les fonctions caractéristiques convergent simplement alors les variables aléatoires convergent bien en distribution, i.e.,

$$\left(\forall z, \lim_{n \rightarrow +\infty} \phi_{X_n}(z) = \phi_{X_l}(z) \right) \Leftrightarrow X_n \xrightarrow{\mathcal{D}} X_l. \quad (\text{I.5})$$

Pour finir, la fonction caractéristique est également compatible avec la normalisation

$$\begin{aligned} \phi_{aX}(z) &= \phi_X(az) \\ \phi_{X+b}(z) &= \phi_X(z)e^{-ibz} \end{aligned} \quad (\text{I.6})$$

En utilisant ces quatre propriétés, il est bien souvent plus simple d'établir les résultats de convergence dans l'espace des fonctions caractéristiques avant d'utiliser l'équivalence (I.5) pour revenir dans l'espace des distributions.

Par exemple, le théorème central limite peut se dériver simplement en utilisant

$$\phi_X(z) = 1 - i\mathbb{E}[X]z - \mathbb{E}[X^2] \frac{z^2}{2} + o(z^2), \quad z \rightarrow 0. \quad (\text{I.7})$$

Cela entraîne que pour la variable normalisée $W = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X])}{\sqrt{n\text{Var}[X]}}$

$$\begin{aligned} \phi_W(z) &= \left(e^{i \frac{\mathbb{E}[X]}{\sqrt{n\text{Var}[X]}} z} \phi_X \left(\frac{z}{\sqrt{n\text{Var}[X]}} \right) \right)^n \\ &= \left(1 - \frac{z^2}{2n} + o\left(\frac{1}{n}\right) \right)^n \\ &\underset{n \rightarrow +\infty}{\sim} e^{-\frac{z^2}{2}}. \end{aligned} \quad (\text{I.8})$$

La fonction caractéristique permet également de caractériser exhaustivement l'ensemble des lois stables pour l'addition. C'est-à-dire l'ensemble des lois \mathcal{L} telles que pour tout couple de variables i.i.d. X et Y de loi \mathcal{L} , il existe des facteurs de normalisation (a, b) tels que $a(X + Y - b)$ soit de loi \mathcal{L} .

1. Plus précisément un morphisme de monoïde, puisque l'addition de variables aléatoires n'est pas inversible.

En résolvant l'équation fonctionnelle associée à cette condition [45] on peut montrer que les lois stables \mathcal{S}_α appartiennent à une famille de lois à 4 paramètres

$$\phi_{\mathcal{S}_\alpha}(z) = \exp \left(i\gamma t - c|z|^\alpha \left(1 - i\beta \operatorname{sign}(t) \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \text{si } \alpha \neq 1 \\ -\frac{2}{\pi} \ln|t| & \text{si } \alpha = 1 \end{cases} \right) \right) \quad (\text{I.9})$$

avec $0 < \alpha \leq 2$, $-1 < \beta < 1$, $c > 0$ et $\gamma \in \mathbb{R}$. Malheureusement, sauf cas particuliers [99], il n'existe pas d'expression analytique simple des densités de probabilités de ces lois.

Ces lois stables sont les seules lois limites possibles de la somme de variable aléatoires i.i.d. [56]. En utilisant ces lois stables, il est alors possible de généraliser le théorème central limite aux lois dont la variance n'est pas finie et la queue de la distribution se comporte comme une loi de puissance. Pour quantifier ce comportement en loi de puissance, il est utile de faire appel à la notion de fonction lentement variable. On dit qu'une fonction L est lentement variable si et seulement si

$$\forall t > 0, \lim_{x \rightarrow +\infty} \frac{L(tx)}{L(x)} = 1. \quad (\text{I.10})$$

Un exemple représentatif de fonction lentement variable est le logarithme $\ln x$. On peut remarquer que si L est une fonction lentement variable alors

$$\forall \varepsilon > 0, \begin{cases} \lim_{x \rightarrow +\infty} L(x)x^\varepsilon = +\infty, \\ \lim_{x \rightarrow +\infty} L(x)x^{-\varepsilon} = 0. \end{cases} \quad (\text{I.11})$$

Les fonctions lentement variables varient donc plus lentement que toute fonction de puissance. Elles possèdent de nombreuses autres propriétés qui sont décrites plus en détail dans l'annexe A. Puisque nous sommes intéressés uniquement par les propriétés générales des fonctions lentement variables, nous utilisons la plupart du temps la même notation L pour n'importe quelle fonction lentement variable.

Dans le cas de sommes de variables aléatoires, si on considère un ensemble de variables aléatoires X_i de cumulative $F(x) \equiv \mathbb{P}(X \leq x)$ telle que

$$\begin{aligned} 1 - F(x) &= \frac{c_1 + o(x)}{x^\alpha} L(x), \\ F(-x) &= \frac{c_2 + o(x)}{x^\alpha} L(x) \end{aligned} \quad (\text{I.12})$$

où $\alpha \in [0, 2)$, $c_1, c_2 \in \mathbb{R}^+$, $c_1 + c_2 > 0$ et L est une fonction lentement variable², il existe alors une suite de normalisation (a_n, b_n) telle que

$$\frac{S_n - b_n}{a_n} \xrightarrow{\mathcal{D}} \mathcal{S}_\alpha. \quad (\text{I.13})$$

2. La condition $1 - F(x) = (c_1 + o(x))x^{-\alpha}L(x)$ correspond au fait que $1 - F(x)$ soit une fonction régulièrement variable d'indice $-\alpha$ [19]. Cependant, l'essentiel des propriétés des fonctions régulièrement variables peut être formulé à partir des fonctions lentement variables. Nous nous contenterons donc par la suite des fonctions lentement variables.

où \mathcal{S}_α est une loi stable de paramètre $\beta = \frac{c_1 - c_2}{c_1 + c_2}$, $c = 1$, $\gamma = 0$. On peut de plus choisir

$$\begin{cases} b_n = \begin{cases} \mathbb{E}[X_i] & \text{si } \alpha > 1, \\ 0 & \text{sinon} \end{cases} \\ a_n = F^{-1}\left(1 - \frac{1}{n}\right) = L_2(n)n^{1/\alpha}, \end{cases} \quad (\text{I.14})$$

avec L_2 une fonction lentement variable dépendant de L . On peut faire ici deux remarques assez intéressantes sur ce résultat classique. Premièrement, pour $\alpha < 1$, l'espérance $\mathbb{E}[X]$ devient infinie, la loi des grands nombres n'est plus valide et l'estimateur de la moyenne $S(n, q = 1)$ croît comme $n^{1/\alpha}$ (donc plus rapidement qu'une croissance linéaire). Deuxièmement, contrairement au cas à variance finie, le comportement du paramètre de normalisation a_n dépend fortement du paramètre α . Nous reviendrons plus en détail sur ce changement de comportement par la suite.

2 Maximum

De façon intéressante, on retrouve un comportement assez similaire à celui des sommes dans le cas des maxima, avec quelques différences notables. Par exemple, si dans le cas des sommes de variables aléatoires, l'espace le plus naturel pour étudier la convergence est celui des fonctions caractéristiques, dans le cas des maxima, il est préférable de s'intéresser à l'espace des fonctions cumulatives F . On retrouve alors les quatre propriétés fondamentales que nous avons évoquées pour les fonctions caractéristiques.

Premièrement, la fonction cumulative d'une variable aléatoire X caractérise complètement la loi de cette variable. Deuxièmement, la fonction cumulative est également un morphisme : Soit deux variables i.i.d. X et Y , on a alors

$$F_{\max\{X,Y\}}(x) = F_X(x)F_Y(x). \quad (\text{I.15})$$

La fonction cumulative se comporte également bien vis-à-vis du passage à la limite, si on considère une suite de variables aléatoires X_n dont la fonction cumulative converge simplement alors les variables aléatoires convergent bien en distribution

$$\left(\forall x, F_{X_l} \text{ continue en } x, \lim_{n \rightarrow +\infty} F_{X_n}(x) = F_{X_l}(x)\right) \Leftrightarrow X_n \xrightarrow{\mathcal{D}} X_l. \quad (\text{I.16})$$

La fonction cumulative permet également de prendre facilement en compte les normalisations

$$\begin{aligned} F_{aX}(x) &= F_X\left(\frac{x}{a}\right), \\ F_{X+b}(x) &= F_X(x - b). \end{aligned} \quad (\text{I.17})$$

On peut mettre en miroir les équations (I.15), (I.16) et (I.17) avec les équations (I.4), (I.5) et (I.6) correspondantes pour la fonction caractéristique. L'analogie est presque parfaite exceptée que les termes de normalisation décrits dans (I.17) et (I.6) prennent une forme différente et que la fonction caractéristique est une fonction de $\mathbb{C} \rightarrow \mathbb{C}$ tandis que la fonction cumulative est une fonction de $\mathbb{R} \rightarrow \mathbb{R}$. Ces différences sont suffisantes pour obtenir un comportement asymptotique distinct du comportement

asymptotique de la somme mais les structures de convergence restent remarquablement similaires.

Par exemple, le maximum est plus sensible que la somme au comportement de la queue de la distribution, il n'y a donc pas d'équivalent simple à la loi forte des grands nombres. Cependant, on retrouve des résultats de convergence similaires à ceux du théorème de convergence vers les lois stables.

Dans le cas du maximum, on peut exprimer les lois max-stables \mathcal{F}_ζ sous la forme suivante

$$\mathcal{F}_\zeta(y) = \begin{cases} e^{-(1+\zeta y)^{-1/\zeta}} & \text{si } \zeta > 0 \quad (y > -1/\zeta), \\ e^{-e^{-y}} & \text{si } \zeta = 0, \\ e^{-(1+\zeta y)^{-1/\zeta}} & \text{si } \zeta < 0 \quad (y < -1/\zeta). \end{cases} \quad (\text{I.18})$$

Il y a donc trois type de lois limites distinctes : les lois de Fréchet $\zeta > 0$, les lois Gumbel $\zeta = 0$ et les lois Weibull $\zeta < 0$. Chacun de ce type de loi est associé à un domaine d'attraction spécifique.

Domaine d'attraction Fréchet :

Les lois ayant une queue en loi de puissance convergent vers les lois de types Fréchet. Plus précisément, si la fonction cumulative complémentaire $\overline{F}(x) = 1 - F(x)$ vérifie

$$\overline{F}(x) = L(x)x^{-\frac{1}{\zeta}}, \quad \zeta > 0 \quad (\text{I.19})$$

avec L une fonction lentement variable, alors le maximum normalisé converge vers F_ζ , c'est-à-dire une loi de Fréchet :

$$\frac{M(n)}{a_n} \equiv \frac{\max\{X_1, \dots, X_n\}}{a_n} \xrightarrow{\mathcal{D}} F_\zeta. \quad (\text{I.20})$$

De plus, on peut choisir

$$a_n = \overline{F}^{-1}\left(\frac{1}{n}\right) \quad (\text{I.21})$$

ce qui signifie que a_n croît en

$$a_n = L_2(n)n^\zeta. \quad (\text{I.22})$$

Domaine d'attraction Weibull :

Similairement, si on considère des variables aléatoires à support borné supérieurement, avec $x_f = \sup\{x, \overline{F}(x) > 0\}$, telles que

$$F(x_f - x) = L\left(\frac{1}{x}\right)x^{-\frac{1}{\zeta}}, \quad \zeta < 0, \quad (\text{I.23})$$

alors $\max\{X_1, \dots, X_n\}$ converge vers F_ζ , une loi Weibull :

$$\frac{M(n) - x_f}{a_n} \xrightarrow{\mathcal{D}} F_\zeta. \quad (\text{I.24})$$

On peut alors choisir les a_n comme

$$a_n = x_f - \overline{F}^{-1}\left(\frac{1}{n}\right) = L_2(n)n^\zeta. \quad (\text{I.25})$$

Domaine d'attraction Gumbel :

Le cas Gumbel est plus complexe et rassemble essentiellement des lois de "type" exponentiel. Plus précisément, le maximum d'un ensemble de variables aléatoires de cumulative F converge vers la distribution Gumbel si et seulement si \overline{F} est asymptotiquement équivalent à une fonction de Von Mises. Les fonctions de Von Mises correspondent à une classe de fonctions rapidement décroissantes pouvant être représentées par

$$\overline{F}(x) = c \exp\left(-\int_{x_0}^x \frac{dt}{g(t)}\right), \quad c > 0, \quad x_0 < x_f \quad (I.26)$$

avec g une fonction positive, absolument continue vis-à-vis de la mesure de Lebesgue et vérifiant $\lim_{x \rightarrow x_f} g'(x) = 0$ [45]. Avec cette condition, on peut montrer que

$$\frac{M(n) - b_n}{a_n} \xrightarrow{\mathcal{D}} F_{\zeta=0}. \quad (I.27)$$

avec (a_n, b_n) vérifiant

$$\begin{cases} b_n = \overline{F}^{-1}\left(\frac{1}{n}\right), \\ a_n = \frac{\int_{b_n}^{x_f} \overline{F}(t) dt}{\overline{F}(b_n)} = n \int_{b_n}^{x_f} \overline{F}(t) dt. \end{cases} \quad (I.28)$$

On retrouve donc trois domaines de lois délimités soit par un comportement en loi de puissance (convergence vers une loi de Fréchet), soit par l'existence d'un point d'arrêt et un comportement algébrique dans son voisinage (loi Weibull) ou bien par une décroissance "quasi-exponentielle" (loi Gumbel).

D'un certain point de vue, ces trois classes de lois établissent une classification assez lisse du comportement asymptotique des fonctions cumulatives et on peut remarquer que contrairement au cas de la somme, on ne retrouve pas la position singulière des lois normales.

3 Rupture de convergence pour les lois de puissance

3.1 Convergence logarithmique

La similarité des résultats de convergence entre somme et maximum est frappante. Par contraste, les points de divergence entre ces deux processus aléatoires n'en deviennent que plus intrigant. Notamment, nous avons déjà remarqué que dans le cas de la somme de lois de puissance ($\overline{F}_X = L(x)x^{-1/\rho}$), il apparaît une rupture de comportement remarquable lorsqu'on se trouve à la frontière $\rho = 1$. En effet, pour $\rho < 1$, on sait que³

$$\lim_{n \rightarrow +\infty} S(n, q = 1) \stackrel{\text{p.s}}{=} \mathbb{E}[X]. \quad (I.29)$$

3. Les fonctions lentement variables croissant moins vite que toutes fonctions de puissance, on a $\exists C, \varepsilon, \overline{F}_X(x) < Cx^{\varepsilon-1/\rho}$. Par conséquent, $\mathbb{E}[X] = \int_{-\infty}^{+\infty} \overline{F}x dx < +\infty$.

Au contraire, pour $\rho > 1$, nous avons (cf. l'Eq. (I.13))

$$n^{1-\rho} S(n, q = 1) \xrightarrow{\mathcal{D}} \mathcal{S}_{1/\rho}, \quad (\text{I.30})$$

où $\mathcal{S}_{1/\rho}$ est une loi stable de paramètre $1/\rho$ (cf. définition (I.9)).

Pour obtenir une vue plus synthétique, il est intéressant de considérer une moyenne logarithmique $\ln S / \ln n$. Dans le cas $\rho < 1$, la loi forte des grands nombres entraîne immédiatement que

$$\lim_{n \rightarrow +\infty} \frac{\ln S(n, q = 1)}{\ln n} \stackrel{\text{p.s.}}{=} 0. \quad (\text{I.31})$$

De manière similaire, il est possible d'utiliser la convergence en distribution de X vers une loi stable et le lemme de Borel-Cantelli, pour montrer que pour $\rho > 1$ ⁴

$$\lim_{n \rightarrow +\infty} \frac{\ln S(n, q = 1)}{\ln n} \stackrel{\text{p.s.}}{=} \rho - 1. \quad (\text{I.32})$$

Si on réunit les deux résultats précédents, on obtient que

$$\lim_{n \rightarrow +\infty} \frac{\ln S(n, q = 1)}{\ln n} \stackrel{\text{p.s.}}{=} \begin{cases} 0 & \text{si } \rho \leq 1, \\ \rho - 1 & \text{si } \rho > 1. \end{cases} \quad (\text{I.33})$$

Le changement de comportement est donc criant entre $\rho < 1$ et $\rho > 1$. On peut apporter une première explication à ce changement de comportement en considérant le log-maximum $\frac{\ln(M(n)/n)}{\ln n}$. Le terme $M(n)/n$ est utilisé ici pour faciliter la comparaison avec $S(n, q = 1)$. La convergence des lois de puissance vers des lois de Fréchet implique (pour des raisons similaires à celles invoquées pour l'Eq. (I.32)) que, quel que soit ρ ,

$$\lim_{n \rightarrow +\infty} \frac{\ln(M(n)/n)}{\ln n} \stackrel{\text{p.s.}}{=} \rho - 1. \quad (\text{I.34})$$

On a donc un comportement régulier en ρ dans le cas du maximum. Cependant, on peut remarquer que les propriétés du maximum et de la somme impliquent que, pour des variables X positives, $M/n < S$. Or si cette inégalité est amplement respectée pour $\rho < 1$, pour $\rho > 1$, les résultats précédents et la figure 1.1 montrent que pour $n \rightarrow +\infty$

$$\frac{\ln(M(n)/n)}{\ln n} \sim \frac{\ln S(n, q = 1)}{\ln n}. \quad (\text{I.35})$$

Le maximum M et la somme nS sont donc du même ordre de grandeur lorsque $\rho > 1$. On peut même aller plus loin en utilisant un résultat de Teugels [20, 73] qui montre notamment que pour $\rho > 1$.

$$\mathbb{E} \left[\frac{nS(n, q = 1)}{M(n)} \right] \rightarrow \frac{\rho}{\rho - 1}. \quad (\text{I.36})$$

Lorsque $\rho > 1$, les valeurs extrêmes jouent un rôle significatif dans le comportement de la somme, ce qui n'est pas le cas pour $\rho < 1$. Le changement de comportement de $\ln S / \ln n$ peut donc s'expliquer par ce rôle accru des valeurs extrêmes dans la somme.

4. La convergence en probabilité est immédiate, puisque $\mathbb{P}(|\ln S / \ln n - (\rho - 1)| > \epsilon) = \mathbb{P}(\ln(n^{1-\rho} S) > \epsilon \ln n) \xrightarrow{n \rightarrow +\infty} 0$. La convergence presque sûre nécessite de considérer une sous-suite $n_k = 2^k$, et le fait que pour des variables positives $S(2n, q = 1) > S(n, q = 1)/2$.

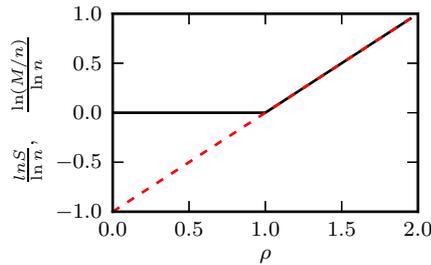


FIGURE 1.1 – Comportement asymptotique de $\ln S / \ln n$ (en trait plein noir) et $\ln(M/n) / \ln n$ (en pointillé rouge) en fonction de ρ . Si le comportement asymptotique de $\ln M$ est régulier, ce n'est pas le cas de $\ln S / \ln n$ qui présente un changement de comportement pour $\rho = 1$. Pour $\rho > 1$, les valeurs extrêmes jouent un rôle dominant dans la somme, $\ln(M/n)$ et $\ln S$ sont donc similaires. Au contraire, pour $\rho < 1$, la somme est dominée par la contribution des termes non extrémaux, son comportement diffère donc de celui de $\ln M$. Ce changement de comportement est bien retranscrit par $\ln S / \ln n$.

3.2 Modèle REM dégénéré et transition vitreuse

On peut proposer une interprétation plus physique de ce changement de comportement. Si on observe attentivement la figure 1.1, on peut trouver que le changement de comportement à $\rho = 1$ ressemble à une transition de phase. En suivant cette intuition, on peut réinterpréter la somme nS comme une fonction de partition d'un modèle à énergie aléatoire particulier. On considère une fonction de partition de la forme

$$Z(\beta) = \sum_{s=1}^{2^N} e^{-\beta E_s}. \quad (\text{I.37})$$

Les énergies E_s sont des variables gelées dans une configuration aléatoire. La fonction de partition $Z(\beta)$ est alors une somme de variables aléatoires. Si on suppose que les variables E_s sont distribuées suivant une distribution exponentielle, les variables $\exp(-\beta E_s)$ sont, elles, des lois de puissance d'exposant $1/\beta$. On a alors une analogie formelle entre $Z(\beta)$ et $nS(n, q = 1)$ pour $\rho = \beta$. Il s'agit d'une forme dégénérée du REM où les énergies suivent une distribution exponentielle indépendante de la taille du système –on retombe donc dans ce cas sur le modèle de pièges. Si on poursuit l'analogie avec la physique, il est alors naturel d'étudier l'énergie libre par spin f définie par

$$f = -\frac{\ln Z}{\beta N}. \quad (\text{I.38})$$

L'analogue de f dans le cas de la somme est alors $\rho(1 + \ln S(n, q = 1) / \ln n)$. Tenant compte du fait que $n = 2^N$, nous savons déjà à travers l'Eq. (I.33) que

$$\lim_n f \stackrel{\text{p.s.}}{=} \begin{cases} -\frac{\ln 2}{\beta} & \text{si } \beta \leq 1, \\ -\ln 2 & \text{si } \beta > 1. \end{cases} \quad (\text{I.39})$$

Cela entraîne notamment que l'entropie par spin $s = \beta^2 \partial_\beta f$ est dans la limite $n \rightarrow +\infty$

$$s \stackrel{\text{p.s.}}{=} \begin{cases} \ln 2 & \text{si } \beta \leq 1, \\ 0 & \text{si } \beta > 1. \end{cases} \quad (\text{I.40})$$

Le changement de comportement en $\rho \equiv \beta = 1$ correspond donc à une transition vitreuse dégénérée dans laquelle on passe brutalement d’une phase gelée à entropie par spin nulle à une phase haute température, avec une discontinuité de l’entropie par spin. Il est intéressant de noter que cette transition vitreuse apparaît dès lors que la loi des grands nombres est brisée et n’est pas sensible à la perte de la convergence vers les lois normales pour $\rho > 1/2$. Cela explique pour partie la régularité de comportement du maximum de lois de puissance puisqu’il n’existe pas d’analogie à la loi des grands nombres lorsqu’on s’intéresse au maximum. Face à cette singularité, une question émerge naturellement : peut-on observer un tel changement de comportement pour le logarithme du maximum $\ln M / \ln n$? Pour des variables aléatoires i.i.d., nous savons déjà que cela est impossible, il faut donc s’affranchir de cette contrainte i.i.d.. Une piste potentielle réside dans le modèle REM, pour lequel il existe un couplage entre la taille du système et la forme des distributions. Ben Arous et ses collaborateurs ont ainsi montré dans [10], qu’il est possible de modifier les distributions limites de la somme de variables aléatoires en les élevant à une puissance dépendant de la taille du signal. Il a été récemment montré [23] qu’il est également possible d’induire une transition entre le domaine Gumbel et le domaine Fréchet en élevant les variables aléatoires à une puissance dépendante de la taille du signal. Si on s’intéresse au logarithme du maximum, on peut remarquer immédiatement que

$$\frac{\ln M(n)^{q(n)}}{\ln n} = q(n) \frac{\ln M(n)}{\ln n}. \quad (\text{I.41})$$

L’introduction d’une puissance variant avec la taille n du système modifie donc manifestement le comportement du maximum, au moins pour les variables appartenant au domaine d’attraction de type Fréchet. On peut se demander quelles sont les conséquences de ce changement de comportement sur les domaines d’attraction et les lois limites du maximum, ou s’il est possible d’observer d’autres transitions entre différents domaines d’attractions que celle décrites dans [23]. Dans le chapitre 2, nous proposons une analyse exhaustive de ces phénomènes qui repose sur une approche par groupe de renormalisation complétée par une approche algébrique. Nous ferons notamment la lumière sur l’existence d’une fonction puissance critique q_n^* qui caractérise la dynamique du comportement du maximum M en fonction de $q(n)$. Cette puissance critique du maximum fait écho à la puissance critique découverte par Ben Arous et ses collaborateurs [10] dans le cas de la somme de variables aléatoires de type exponentielle. Nous reviendrons plus en détail sur cette question de puissances critiques pour la somme de variables aléatoires dans la partie II –en nous concentrant sur ses applications potentielles en traitement du signal.

Chapitre 2

Extrêmes, transformations de puissance et renormalisation

1 Définition du problème

Nous partons d'un jeu de variables i.i.d. $(W_i)_{i=1,\dots,n}$ dont nous considérons le maximum M_n^W :

$$M_n^W = \max\{W_1, \dots, W_n\}. \quad (\text{I.42})$$

Comme rappelé au chapitre 1, les théorèmes classiques de la statistique des extrêmes décrivent en détail la convergence asymptotique de M_n^W en fonction de la forme de la queue de la distribution de probabilité de W . Nous souhaitons nous enquérir si une transformation, elle-même fonction de n , des variables W_i peut induire l'émergence de distributions asymptotiques différentes des distributions classiques. Nous sommes particulièrement intéressés par le cas des transformations de puissance, de la forme $U_{i,n} = W_i^{q_n}$ où q_n est fonction de n . Nous commençons cependant par une classe plus générale de transformation. Pour ce faire, nous considérons une variable transformée

$$U_{i,n} = \omega_n^{-1}(W_i), \quad i = 1, \dots, n, \quad (\text{I.43})$$

où ω_n est une fonction bijective croissante¹. La distribution cumulative de $U_{i,n}$ s'exprime alors comme :

$$F_{U,n}(u) = F_W(\omega_n(u)). \quad (\text{I.44})$$

Dans le cas des transformations de puissance, cette définition de $\omega_n(u)$ entraîne que

$$\omega_n(u) = u^{1/q_n}, \quad u > 0. \quad (\text{I.45})$$

Cette transformation s'inspire de l'étude [10] concernant le comportement des sommes d'exponentielles de variables aléatoires, et notamment la perte du théorème central limite en présence d'exposants de puissance croissant rapidement.

D'un côté, si ω_n varie suffisamment lentement avec n , on s'attend à ce que la transformation ω_n n'altère pas la distribution limite du maximum. D'un autre côté, pour des transformations bien choisies, il devrait être possible d'atteindre de nouveaux

1. Il est également possible de considérer des transformations bijectives décroissantes au prix de changements mineurs.

types de distributions limites. Considérons donc le maximum de la variable transformée $M_n^U = \max\{U_{1,n}, \dots, U_{n,n}\}$. Sa distribution cumulative $F_n(m)$ peut s'écrire en fonction de la cumulative F_W de la variable W

$$F_n(m) = F_W(\omega_n(m))^n. \quad (\text{I.46})$$

Dans la section suivante, nous concevrons une formulation par groupe de renormalisation de l'Eq. (I.46), qui nous permettra de dériver directement l'ensemble des distributions limites.

2 Approche par renormalisation

2.1 Transformation de renormalisation et conditions de normalisation

Suivant la voie pavée par les articles [62, 17], nous introduisons les fonctions $g_n(m) = -\ln[-\ln F_n(m)]$ ainsi que $g_W(w) = -\ln[-\ln F_W(w)]$, et réécrivons l'Eq. (I.46) sous la forme

$$g_n(m) = g_W(\omega_n(m)) - \ln n. \quad (\text{I.47})$$

Comme dans le cas des théorèmes de convergence standard, il est nécessaire, afin de converger, de normaliser le maximum M_n^U à travers $X_n = (M_n^U - b_n)/a_n$, où a_n et b_n sont choisis afin de vérifier certaines conditions arbitraires (par exemple, fixer la valeur des deux premiers moments de X_n). De plus, il est pratique de considérer n comme une variable réelle positive plutôt qu'entière et de définir $s = \ln n$. Nous supposons par conséquent que $\omega_n(m)$ peut être étendu de manière C^1 aux valeurs entières de n , et nous définissons la fonction $\omega(m, s) = \omega_{e^s}(m)$. En combinant ces résultats, l'Eq. (I.47) se transforme en une équation d'évolution de la variable $X_s \equiv X_{e^s}$

$$g(x, s) = g_W(\omega(a(s)x + b(s), s)) - s, \quad (\text{I.48})$$

avec

$$F_{X_s}(x) = \exp[-\exp(-g(x, s))]. \quad (\text{I.49})$$

Afin de déterminer $a(s)$ et $b(s)$, il est nécessaire d'imposer des conditions de normalisation sur $g(x, s)$. De telles contraintes sont assez arbitraires et peuvent varier selon que l'on préfère une approche pragmatique ou plus théorique. Dans une approche pragmatique, il peut être pratique de fixer certains moments de la distribution (par exemple les deux premiers). Au contraire, il s'avère que pour des applications plus théoriques, il est plus facile de fixer la valeur de $g(x, s)$ et sa dérivée en un point donné x . Ici, nous choisissons les mêmes conditions que celles appliquées dans [17], c'est-à-dire²

$$\begin{aligned} g(0, s) &= 0, \\ \partial_x g(0, s) &= 1. \end{aligned} \quad (\text{I.50})$$

Injectées dans l'équation (I.48), ces conditions impliquent que

$$g_W(\omega(b(s), s)) = s, \quad (\text{I.51})$$

$$a(s) = b'(s) + \frac{\partial_s \omega(b(s), s)}{\partial_m \omega(b(s), s)}, \quad (\text{I.52})$$

2. À travers ce chapitre, nous utilisons les notations $\partial_x \equiv \partial/\partial x$ et $\partial_s \equiv \partial/\partial s$.

avec $b'(s)$ la dérivée de $b(s)$ et $\partial_m \omega$ la dérivée selon le premier argument de $\omega(m, s)$. Pour simplifier l'expression de g_W , nous contractons la transformation $\omega(x, s)$ et l'opération de normalisation en une unique transformation $\mathcal{T}(x, s)$:

$$\mathcal{T}(x, s) = \omega(a(s)x + b(s), s), \quad (\text{I.53})$$

ce qui permet de simplifier l'Eq. (I.48) en

$$g(x, s) = g_W(\mathcal{T}(x, s)) - s. \quad (\text{I.54})$$

2.2 Équation aux dérivées partielles du flot

L'équation fonctionnelle (I.54) peut être convertie en une équation aux dérivées partielles. La première étape est de dériver $g(x, s)$ selon x et s :

$$\partial_x g(x, s) = g'_W(\mathcal{T}(x, s)) \partial_x \mathcal{T}(x, s), \quad (\text{I.55})$$

$$\partial_s g(x, s) = g'_W(\mathcal{T}(x, s)) \partial_s \mathcal{T}(x, s) - 1. \quad (\text{I.56})$$

En réinjectant l'Eq. (I.55) dans l'Eq. (I.56) pour éliminer g'_W , on obtient

$$\partial_s g(x, s) = \mathcal{U}(x, s) \partial_x g(x, s) - 1, \quad (\text{I.57})$$

où nous avons défini

$$\mathcal{U}(x, s) = \frac{\partial_s \mathcal{T}(x, s)}{\partial_x \mathcal{T}(x, s)}. \quad (\text{I.58})$$

La fonction $\mathcal{U}(x, s)$ peut être explicitée grâce aux Eqs. (I.53) et (I.58). Nous commençons par calculer $\partial_x \mathcal{T}$ et $\partial_s \mathcal{T}$ (pour alléger les notations, nous éludons les dépendances explicites en s de $a(s)$ et $b(s)$) :

$$\partial_x \mathcal{T}(x, s) = a \partial_m \omega(ax + b, s), \quad (\text{I.59})$$

$$\partial_s \mathcal{T}(x, s) = (a'x + b') \partial_m \omega(ax + b, s) + \partial_s \omega(ax + b, s). \quad (\text{I.60})$$

Nous obtenons au final

$$\mathcal{U}(x, s) = \frac{a'}{a} x + \frac{b'}{a} + \frac{1}{a} \frac{\partial_s \omega(ax + b, s)}{\partial_m \omega(ax + b, s)}. \quad (\text{I.61})$$

À partir de ce point, nous nous concentrons sur le cas des transformations en puissance $\omega(m, s) = m^{1/q(s)}$, avec $m > 0$. Nous avons

$$\partial_m \omega(m, s) = \frac{\omega(m, s)}{q(s)m}, \quad (\text{I.62})$$

$$\partial_s \omega(m, s) = -\frac{q'(s) \ln m}{q(s)^2} \omega(m, s), \quad (\text{I.63})$$

et $\mathcal{U}(x, s)$ s'écrit

$$\mathcal{U}(x, s) = \frac{a'}{a} x + \frac{b'}{a} - \frac{q'}{q} \left(x + \frac{b}{a} \right) \ln(ax + b). \quad (\text{I.64})$$

En prenant en compte les conditions de normalisation (I.52), qui s'écrivent dans le cas d'une transformation de puissance $\omega(m, s)$,

$$a = b' - \frac{q'}{q} b \ln b, \quad (\text{I.65})$$

il est possible de réécrire l'Eq. (I.64) en

$$\mathcal{U}(x, s) = \left(\frac{a}{b} + \partial_s \ln \frac{a}{b} \right) x + 1 - \frac{q'}{q} \left(x + \frac{b}{a} \right) \ln \left(\frac{a}{b} x + 1 \right). \quad (\text{I.66})$$

Si l'on définit

$$\lambda(s) = \frac{a(s)}{b(s)}, \quad \delta(s) = \partial_s \ln \lambda(s), \quad \gamma(s) = \lambda(s) + \delta(s), \quad Q(s) = \frac{q'(s)}{q(s)}, \quad (\text{I.67})$$

il est possible de réécrire l'Eq. (I.66) sous la forme plus compacte

$$\mathcal{U}(x, s) = 1 + \gamma(s)x - Q(s) (\lambda(s)x + 1) \frac{\ln(\lambda(s)x + 1)}{\lambda(s)}. \quad (\text{I.68})$$

Cette équation est une extension de celle obtenue dans [17], avec un terme supplémentaire $Q(s) (\lambda(s)x + 1) \frac{\ln(\lambda(s)x + 1)}{\lambda(s)}$ due à la transformation de puissance. On peut d'ailleurs retrouver l'équation du flot de [17] en posant $Q(s) = 0$.

2.3 Distributions stables

Dans le cas d'une solution stationnaire, c'est-à-dire une solution de l'Eq. (I.57) vérifiant $\partial_s g(x, s) = 0$, on peut dériver selon s l'Eq. (I.57) pour obtenir

$$\begin{aligned} \partial_s [\mathcal{U}(s, x) \partial_x g(s, x)] &= 0 \\ (\partial_s \mathcal{U}(s, x)) \partial_x g(s, x) &= 0 \\ \partial_s \mathcal{U}(s, x) &= 0. \end{aligned} \quad (\text{I.69})$$

Une solution stationnaire $g(x)$ n'apparaît donc qu'à la condition que $\mathcal{U}(x, s)$ soit indépendant de s , c'est-à-dire $\mathcal{U}(x, s) = \mathcal{U}(x)$. Dans ce cas, cette solution stationnaire $g(x)$ est obtenue en intégrant l'équation différentielle

$$g'(x) = \frac{1}{\mathcal{U}(x)}, \quad (\text{I.70})$$

avec la condition $g(0) = 0$.

Nous recherchons donc des conditions sous lesquelles $\mathcal{U}(x, s)$ devient indépendant de s .

2.4 Cas $Q = 0$: redécouverte des distributions limites standards

Dans le cas $Q = 0$, nous avons

$$\mathcal{U}(x, s) = 1 + \gamma(s)x. \quad (\text{I.71})$$

La condition $\partial_s \mathcal{U} = 0$ entraîne que $\gamma(s)$ doit être égale à une constante γ . Par conséquent,

$$g'(x) = \frac{1}{1 + \gamma x}. \quad (\text{I.72})$$

Si on prend en compte les conditions de normalisation (I.50), on obtient la fonction point fixe

$$g(x) = \frac{1}{\gamma} \ln(1 + \gamma x). \quad (\text{I.73})$$

En reformulant ces résultats en terme de distribution cumulative $F(x) = \exp[-\exp(-g(x))]$, on retrouve les distributions limites classiques déjà décrites dans l'Eq. (I.18),

$$F(x) = \mathcal{F}_\gamma(x) = \exp\left[-(1 + \gamma x)^{-\frac{1}{\gamma}}\right], \quad 1 + \gamma x > 0, \quad (\text{I.74})$$

où γ joue le rôle du paramètre ζ . Ainsi nous retrouvons les distributions limites classiques dans le cas où la puissance $q(s)$ est une constante, c'est-à-dire $q(s) = q_0$. Ce résultat était attendu : si X appartient au domaine d'attraction \mathcal{F}_ζ , X^{q_0} appartient soit au domaine d'attraction de \mathcal{F}_{ζ/q_0} si $\zeta > 0$, soit au domaine d'attraction de \mathcal{F}_ζ si $\zeta \leq 0$.

2.5 Cas $Q \neq 0$: émergence de distributions stables non standards

Nous nous intéressons maintenant au cas $Q \neq 0$. En dérivant l'Eq. (I.68) par rapport à s , on obtient une somme de fonctions linéairement indépendantes

$$\begin{aligned} \partial_s \mathcal{U}(x, s) = & \gamma'(s)x - Q'(s) (\lambda(s)x + 1) \frac{\ln(\lambda(s)x + 1)}{\lambda(s)} \\ & - Q(s) \left(x - \frac{\lambda'(s)}{\lambda(s)^2} \right) \ln(\lambda(s)x + 1) \\ & + Q(s) \frac{\lambda'(s)}{\lambda(s)} x \end{aligned} \quad (\text{I.75})$$

Observant l'Eq. (I.75), il est clair que $\mathcal{U}(x, s)$ n'est indépendant de s que si $Q(s)$ et $\lambda(s)$ sont constants, $Q(s) = Q$ et $\lambda(s) = \lambda > 0$. Par suite, les solutions stationnaires ne peuvent exister que si $q(s)$ est de la forme $q(s) = K e^{Qs}$, avec $K > 0$ une constante réelle. Injectant Q et λ dans l'Eq. (I.66), $\mathcal{U}(x)$ prend la forme

$$\mathcal{U}(x) = 1 + \lambda x - \frac{Q}{\lambda} (1 + \lambda x) \ln(1 + \lambda x). \quad (\text{I.76})$$

La combinaison de l'Eq. (I.76) avec l'Eq. (I.70) mène à

$$g'(x) = \frac{1}{(1 + \lambda x) \left(1 - \frac{Q}{\lambda} \ln(1 + \lambda x)\right)}. \quad (\text{I.77})$$

Par définition, $g(x)$ doit être une fonction croissante de x , ce qui signifie que $g'(x) \geq 0$, ce qui implique que x appartient à un intervalle restreint de valeurs, $x_{\min} < x < x_{\max}$. Les bornes x_{\min} and x_{\max} sont déterminées par les conditions $1 + \gamma x > 0$ et $1 - \frac{Q}{\lambda} \ln(1 + \lambda x) > 0$. Si $Q > 0$, on trouve

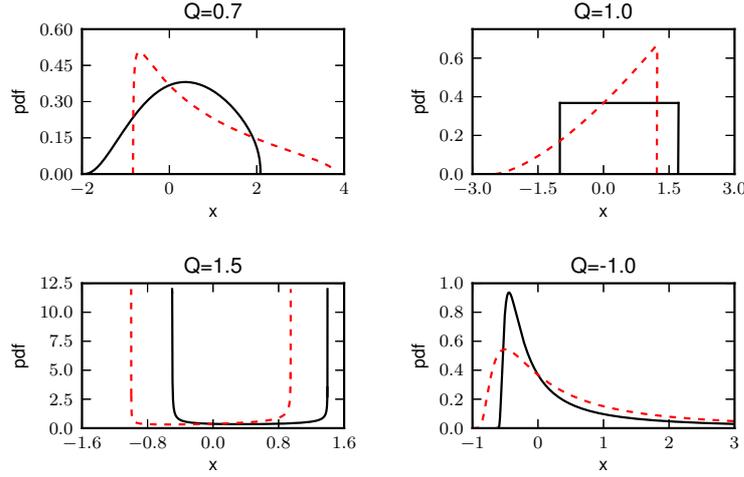


FIGURE 2.1 – Densité de probabilité $p_{\lambda,Q}(x)$ associée à la distribution des extrêmes généralisées $\mathcal{F}_{\lambda,Q}(x)$. En haut à gauche : $Q = 0.7$, ligne noire pleine $\lambda = 0.5$, ligne rouge pointillé $\lambda = 1.2$. En haut à droite : $Q = 1$, ligne noire pleine $\lambda = 1$, ligne rouge pointillé $\lambda = 0.4$. En bas à gauche : $Q = 1.5$, ligne noire pleine $\lambda = 2$, ligne rouge pointillé $\lambda = 1$. En bas à droite : $Q = -1$, ligne noire pleine $\lambda = 1$, ligne rouge pointillé $\lambda = 0.01$.

$$x_{\min} = -\frac{1}{\lambda}, \quad x_{\max} = \frac{1}{\lambda} (e^{\lambda/Q} - 1), \quad \lambda > 0. \quad (\text{I.78})$$

Parallèlement, pour $Q < 0$,

$$x_{\min} = \frac{1}{\lambda} (e^{\lambda/Q} - 1), \quad x_{\max} = +\infty, \quad \lambda > 0. \quad (\text{I.79})$$

Retournant à l'Eq. (I.77), cette équation peut être intégrée pour obtenir

$$g(x) = -\frac{1}{Q} \ln \left(1 - \frac{Q}{\gamma} \ln(1 + \gamma x) \right), \quad x_{\min} < x < x_{\max}, \quad (\text{I.80})$$

qui prend également en compte la condition $g(0) = 0$. Les distributions cumulatives correspondantes $\mathcal{F}_{\lambda,Q}(x)$ s'écrivent

$$\mathcal{F}_{\lambda,Q}(x) = \exp \left[- \left(1 - \frac{Q}{\lambda} \ln(1 + \lambda x) \right)^{1/Q} \right], \quad x_{\min} < x < x_{\max}. \quad (\text{I.81})$$

Ces distributions généralisent les distributions des extrêmes standards; pour $Q = 1$, cette expression se réduit à une loi de puissance sur l'intervalle $[x_{\min}, x_{\max}]$. On peut noter que l'expression (I.81) de $\mathcal{F}_{\lambda,Q}(x)$ converge, dans la limite $Q \rightarrow 0$, vers la distribution standard $\mathcal{F}_{\lambda}(x)$ donné par l'Eq. (I.74). Parallèlement, il est intéressant que la limite, lorsque $\lambda \rightarrow 0$, de la cumulative $\mathcal{F}_{\lambda,Q}$ soit la distribution limite standard \mathcal{F}_{-Q} .

Remarquablement, ces distributions des extrêmes généralisées sont directement reliées aux distributions standards. En termes de cumulative, l'expression (I.81) implique que

$$\mathcal{F}_{\lambda,Q}(x) = \mathcal{F}_{-Q}(\ln(1 + \lambda x)/\lambda). \quad (\text{I.82})$$

À ce sujet, en posant $\zeta = -Q$, l'Eq. (I.82) implique que si Z est une variable aléatoire de cumulative \mathcal{F}_ζ , alors la cumulative de $[\exp(\lambda Z) - 1]/\lambda$ n'est autre que $\mathcal{F}_{\lambda, -\zeta}$. On peut noter que l'Eq. (I.82) est également valide pour $Q = 0$, et correspond à une relation bien connue entre les distributions Gumbel et Weibull.

Il est intéressant de constater que pour $Q > 0$ ces distributions sont équivalentes, à une transformation affine près, aux distributions non standards obtenues dans [28] en utilisant des transformations de puissances comme terme de normalisation. Cependant, il faut noter que l'essence de notre approche est fondamentalement différente de celle de [28]. Dans cet article, les puissances q_n sont utilisées comme un paramètre ajustable de normalisation afin d'assurer la convergence vers une limite non dégénérée. Dans notre approche, nous considérons les puissances q_n comme une fonction donnée a priori, et étudions les limites non dégénérées obtenus à travers une normalisation affine classique. Il est donc a priori hautement non trivial que les deux mêmes ensembles de distributions non standard aient émergé de ces deux approches.

Il est possible de dériver les expressions des densités de probabilités $p_{\lambda, Q}(x)$ de ces distributions des extrêmes généralisées, ce qui donne

$$p_{\lambda, Q}(x) = \frac{1}{1 + \lambda x} \left[1 - \frac{Q}{\lambda} \ln(1 + \lambda x) \right]^{\frac{1}{Q}-1} \exp \left[- \left(1 - \frac{Q}{\lambda} \ln(1 + \lambda x) \right)^{1/Q} \right]. \quad (\text{I.83})$$

Comme illustré sur la Fig. (2.1), cette famille de densité de probabilité a un comportement non trivial près des bornes. Pour $Q > 0$, le support de $p_{\lambda, Q}$ est $(-1/\lambda, [\exp(\lambda/Q) - 1]/\lambda)$. Si nous considérons le comportement asymptotique dans la limite $x \equiv -1/\lambda + \epsilon$, avec $\epsilon \rightarrow 0$, on obtient

$$p_{\lambda, Q} \left(-\frac{1}{\lambda} + \epsilon \right) \approx \frac{1}{\epsilon} \left[\frac{Q}{\lambda} |\ln \epsilon| \right]^{\frac{1}{Q}-1} e^{|\frac{Q}{\lambda} \ln \epsilon|^{1/Q}}. \quad (\text{I.84})$$

Nous avons donc un changement de comportement à $Q = 1$ où la densité de probabilité correspond à une loi de puissance d'exposant $1/\lambda - 1$. Pour $Q > 1$, $p_{\lambda, Q}(x)$ diverge lorsque $x \rightarrow -1/\lambda$ plus rapidement que toute fonction de puissance, tandis que pour $Q < 1$, $p_{\lambda, Q}(x)$ et ses dérivées convergent vers 0.

Similairement, dans le voisinage de la borne supérieure $x \equiv [\exp(\lambda/Q) - 1]/\lambda - \epsilon$, nous avons

$$p_{\lambda, Q}(x_{\max} - \epsilon) \approx e^{-\lambda/Q} \left(Q e^{-\lambda/Q} \epsilon \right)^{\frac{1}{Q}-1}, \quad (\text{I.85})$$

et une singularité apparaît donc à la borne supérieure pour $Q > 1$.

À l'opposé, pour $Q < 0$, $(\exp(\lambda/Q) - 1)/\lambda$ devient la borne inférieure de la distribution, et poser $x \equiv x_{\min} + \epsilon$ avec $\epsilon \rightarrow 0$ mène à

$$p_{\lambda, Q}(x_{\min} + \epsilon) \approx e^{-\lambda/Q} \left(|Q| e^{-\lambda/Q} \epsilon \right)^{\frac{1}{Q}-1} \exp \left[\left(|Q| e^{-\lambda/Q} \epsilon \right)^{\frac{1}{Q}} \right]. \quad (\text{I.86})$$

Puisque $Q < 0$, le terme exponentiel est prépondérant et la densité de probabilité et ses dérivées à tout ordre convergent vers 0 au voisinage de x_{\min} . De l'autre côté, pour $x \rightarrow +\infty$, $F(x)$ se comporte comme l'exponentielle d'une puissance de logarithme :

$$F(x) \underset{x \rightarrow +\infty}{\approx} \exp \left[- \left| \frac{Q}{\lambda} \ln x \right|^{-\frac{1}{Q}} \right]. \quad (\text{I.87})$$

Il y a donc quatre comportements asymptotiques distincts pour les fonctions $p_{\lambda, Q}(x)$ en fonction de la valeur de Q ($Q < 0$, $0 < Q < 1$, $Q = 1$, $Q > 1$).

3 Domaine d'attraction des lois non standards

Après avoir découvert les lois limites non standards pour la distribution des extrêmes en section 2, nous nous intéressons désormais à leurs domaines d'attraction. Pour ce faire, nous utilisons l'équation aux dérivées partielles du flot pour développer une description heuristique de ces domaines d'attraction. Cette description procure une bonne approximation des domaines d'attraction standard et donnera une bonne intuition sur les domaines non standards. Enfin, ces résultats heuristiques sont vérifiés indépendamment dans la section 4 en suivant une approche indépendante qui met en lumière la relation profonde entre les lois standards et non standards.

Bien que l'équation stationnaire résultant de (I.70) et (I.76) ne puisse pas être utilisée directement pour décrire la dynamique de l'équation générale (I.57), elle met en avant l'existence de deux importants paramètres de contrôle dans le cas stationnaire : Q et λ .

Si nous retournons à l'équation générale du flot (I.57) et à l'expression générale de \mathcal{U} décrite par l'Eq. (I.68), ces deux variables peuvent être réinterprétées comme des termes de forçage extérieur. À un niveau heuristique, la convergence de l'Eq. (I.57) devrait être dictée par le comportement asymptotique de ces paramètres. Notamment, si $\gamma(s)$ et $Q(s)$ convergent simultanément vers leurs limites finies respectives Γ et \mathcal{Q} , il est raisonnable de s'attendre à ce que le maximum transformé converge vers la loi limite $\mathcal{F}_{\Gamma, \mathcal{Q}}$. En corollaire, pour des lois appartenant au domaine d'attraction classique de paramètre ζ et dans le cas $\mathcal{Q} = 0$, on devrait obtenir $\Gamma = \zeta$. Cela devrait donner lieu à un cas test intéressant pour notre argument heuristique.

3.1 Comportement asymptotique des paramètres de forçage

Si on considère la transformation $x \mapsto x^{q(s)}$ comme une action d'un agent externe sur les variables aléatoires X_i , on peut interpréter les termes en $Q(s)$ induits par cette transformation dans l'équation du flot comme des termes de forçage extérieur. Cependant, l'interprétation de $\lambda(s)$ et $\delta(s)$ est bien plus floue. La définition de λ nous dit que

$$\begin{aligned} \lambda(s) &= \partial_s \ln b(s) - q'(s) \ln b(s) \\ &= \frac{q(s)}{g_W^{-1}(s) g'_W(g_W^{-1}(s))} \\ &\equiv q(s) \lambda_0(s), \end{aligned} \tag{I.88}$$

Il est donc possible de factoriser λ entre un facteur $q(s)$ et un terme $\lambda_0(s)$ ne dépendant que de la fonction cumulative de la distribution parente. Avec le changement de variables $s = g(x)$ et la relation $g(x) = -\ln(-\ln F(x))$, l'équation précédente devient

$$\lambda_0(g(x)) = \frac{|F(x) \ln F(x)|}{x F'(x)}. \tag{I.89}$$

Il devrait être noté que $\lambda_0(s) > 0$ grâce aux propriétés de la fonction cumulative. De plus, si on s'intéresse au comportement asymptotique, on peut simplifier plus avant l'équation précédente en

$$\lambda_0(g(x)) \underset{x \rightarrow x_F}{\sim} -\frac{\bar{F}(x)}{x \bar{F}'(x)}. \tag{I.90}$$

La variable $\lambda_0(s)$ correspond à l'inverse de l'exposant local de la fonction cumulative complémentaire \overline{F} au point $x = g(s)$. Si $\lambda_0(s)$ admet une limite finie et non nulle Λ_0 pour $s \rightarrow +\infty$, il existe alors une fonction lentement variable normalisée³ L telle que

$$\overline{F}(x) = L(x)x^{-1/\Lambda_0}. \quad (\text{I.91})$$

En d'autres termes, le paramètre Λ_0 décrit le comportement en loi de puissance de la queue de la distribution parente. Puisque pour $Q(s) = 0$, $\lambda(s)$ et $\lambda_0(s)$ sont confondues, les lois appartenant aux domaines d'attraction classiques nous procurent des exemples pertinents du comportement asymptotique de $\lambda_0(s)$. En particulier, si la loi appartient au domaine Fréchet de paramètre ζ , il existe une fonction lentement variable L telle que

$$\overline{F}(x) = L(x)x^{-\frac{1}{\zeta}}. \quad (\text{I.92})$$

Avec l'hypothèse de régularité que L est normalisée, cela entraîne que $\Lambda_0 = \zeta$. De manière opposée, pour une loi appartenant au domaine Weibull, par définition, il existe une fonction lentement variable L telle que

$$\overline{F}(x) = (x_F - x)^{-1/\zeta} L((x_F - x)^{-1}) \quad (\text{I.93})$$

où x_F est le point terminal de la distribution (c'est-à-dire $F(x_F) = 1$). Sous l'hypothèse que L est normalisée, nous avons

$$\lambda_0(g(x)) \underset{x \rightarrow x_F}{\sim} \frac{\zeta}{x_F} (x_F - x). \quad (\text{I.94})$$

Le domaine Gumbel est plus difficile à caractériser. Cependant, les lois en exponentielle étirée

$$\overline{F}(x) = \exp(-x^\alpha) \quad (\text{I.95})$$

constituent souvent un sous-ensemble intéressant de ce domaine. Pour de telles lois, un court calcul mène à

$$\lambda_0(g(x)) \underset{x \rightarrow +\infty}{\sim} \frac{x^{-\alpha}}{\alpha}. \quad (\text{I.96})$$

Dans le cas général, une loi appartient au domaine d'attraction Gumbel si et seulement si $1/\overline{F}(x)$ est une fonction variant rapidement :

$$\forall t > 1, \lim_{x \rightarrow +\infty} \frac{\overline{F}(x)}{\overline{F}(tx)} = +\infty. \quad (\text{I.97})$$

En utilisant la représentation intégrale des fonctions variant rapidement [19, p. 178], il est possible de montrer, sous une hypothèse de régularité, que pour toutes les lois du domaine Gumbel, $\lambda_0(s) \rightarrow 0$. Par conséquent, pour les lois des domaines Weibull et Gumbel, nous avons $\Lambda_0 = 0$. Pour résumer, dans le cas des lois appartenant au domaine standard de paramètre ζ , on a $\Lambda_0 = \max(0, \zeta)$.

De manière similaire, $\delta(s)$ peut être décomposée en

$$\delta(s) = Q(s) + \delta_0(s), \quad \delta_0(s) = \partial_s \ln \lambda_0(s). \quad (\text{I.98})$$

³. Une fonction lentement variable est dite normalisée si $\lim_{x \rightarrow +\infty} xL'(x)/L(x) = 0$ [19, p. 15] (cf. annexe A)

Une conséquence intéressante de l'équation précédente est que les comportements asymptotiques de $\delta_0(s)$ et $\lambda_0(s)$ sont entrelacés. Si nous supposons que $\delta_0(s)$ et $\lambda_0(s)$ admettent simultanément une limite finie lorsque s tend vers $+\infty$, si $\lim_{s \rightarrow \infty} \lambda_0(s) > 0$, alors nécessairement $\lim_{s \rightarrow \infty} \delta_0(s) = 0$. Cette situation correspond au domaine Fréchet. Au contraire, si $\lim_{s \rightarrow \infty} \lambda_0(s) = 0$ (une situation typique pour les domaines Gumbel ou Weibull), et si $\delta_0(s)$ admet une limite finie Δ_0 lorsque $s \rightarrow \infty$, alors $\Delta_0 < 0$. Après quelques mornes calculs, on peut étendre l'expression (I.98) de $\delta_0(s)$ en

$$\begin{aligned} \delta_0(g(x)) &= |\ln F(x)| - \frac{F''(x)|\ln F(x)|}{F'(x)^2} - 1 - \lambda_0(g(x)) \\ &\underset{s \rightarrow +\infty}{\sim} \frac{\overline{F}(x)\overline{F}''(x)}{\overline{F}'(x)^2} - 1 - \lambda_0(g(x)). \end{aligned} \quad (\text{I.99})$$

Pour continuer notre étude des domaines classiques, nous considérons une loi du domaine Weibull de paramètre ζ de fonction lentement variable L . Sous l'hypothèse de régularité que L et L' soient toutes deux des fonctions variant lentement normalisées, on peut vérifier que $\Delta_0 = \zeta$. Le même type de calcul pour les lois en exponentielles étirées (un exemple type de loi de domaine Gumbel) mène à

$$\delta_0(g(x)) \underset{x \rightarrow +\infty}{\sim} -x^{-\alpha}, \quad (\text{I.100})$$

soit, $\Delta_0 = 0$. En d'autres termes, les limites Λ_0 et Δ_0 sont respectivement la partie positive et la partie négative de $\Gamma_0 = \lim_{s \rightarrow \infty} (\delta_0(s) + \lambda_0(s))$, plus précisément $\Lambda_0 = \max(0, \Gamma_0)$ et $\Delta_0 = \min(0, \Gamma_0)$.

Pour une loi appartenant au domaine d'attraction standard de paramètre ζ , nous avons, comme attendu, $\Gamma_0 = \zeta$. Ainsi, pour $Q = 0$, nous retrouvons les résultats classiques sur la forme des domaines d'attractions (avec toutefois des hypothèses de régularité plus forte sur \overline{F}). À partir de là, il est facile d'étendre ces résultats au cas non standard ($Q \neq 0$).

En réunissant les Eqs. (I.88) et (I.98), $\gamma(s)$ peut être réécrit comme

$$\gamma(s) = \delta_0(s) + Q(s) + q(s)\lambda_0(s). \quad (\text{I.101})$$

Cette expression met en exergue un jeu d'interactions très intéressant entre la queue de la distribution et la transformation de puissance. Le comportement de la queue de la distribution est responsable de la présence du paramètre δ_0 tandis que $Q(s)$ dérive directement du choix de $q(s)$. Le terme $q(s)\lambda_0(s)$, lui, représente au contraire l'interaction entre les deux effets.

3.2 Modifier le domaine d'attraction par le biais de $q(s)$

Si nous considérons une loi parente fixée, les fonctions $\delta_0(s)$ et $\lambda_0(s)$ sont également fixées. Le seul paramètre libre restant est alors $q(s)$. Dans cette situation, la limite Γ est déterminée par le comportement asymptotique de $\lambda(s) = q(s)\lambda_0(s)$. Introduisons donc la limite $\Lambda \equiv \lim_{s \rightarrow \infty} \lambda(s)$, dans l'hypothèse que cette limite existe. Si nous définissons une fonction puissance caractéristique

$$q^*(s) = \frac{1}{\lambda_0(s)}, \quad (\text{I.102})$$

nous obtenons

$$\Lambda = \lim_{s \rightarrow \infty} \frac{q(s)}{q^*(s)}. \quad (\text{I.103})$$

En d'autres termes, en prenant la limite $s \rightarrow \infty$ de l'Eq. (I.101), on s'aperçoit que Γ ne dépend que de Δ_0 , \mathcal{Q} et du ratio limite entre $q(s)$ et la fonction puissance caractéristique $q^*(s)$. À partir de là, deux cas différents apparaissent. D'un côté, un cas particulier émerge si $\lim_{s \rightarrow +\infty} q(s)/q^*(s) = 0$. En effet, si $\lim_{s \rightarrow \infty} \lambda(s) = 0$, l'expression de $\mathcal{U}(x, s)$ peut être linéarisée en

$$\mathcal{U}(x, s) \approx 1 + \gamma(s)x - Q(s)x. \quad (\text{I.104})$$

Par suite, si $\gamma(s)$ et $Q(s)$ admettent des limites finies Γ et \mathcal{Q} , nous obtenons une équation aux dérivées partielles correspondant à la distribution limite standard avec $\zeta = \Delta_0$. Cela implique que les domaines Weibull et Gumbel ne sont pas altérés par de telles transformations. Cependant, pour une loi appartenant au domaine Fréchet, de paramètre ζ , nous avons $\Lambda_0 = \zeta$ et $\Delta = 0$. Cela signifie que $\Lambda = 0$ est uniquement possible si $\lim_s q(s) = 0$ et $Q < 0$. En d'autres termes, une puissance décroissante déstabilise l'ensemble du domaine Fréchet et toute loi appartenant à ce domaine converge vers une distribution Gumbel une fois exposée à une transformation de puissance dont les exposants tendent vers 0.

D'un autre côté, des distributions non standards apparaissent lorsque $q(s) \sim \Lambda q^*(s)$. Dans ce cas, nous avons

$$\Gamma = \Delta_0 + \mathcal{Q} + \Lambda. \quad (\text{I.105})$$

De plus, la relation $Q(s) = q'(s)/q(s)$ entraîne

$$Q(s) = -\delta_0(s) + \frac{\lambda'(s)}{\lambda(s)}. \quad (\text{I.106})$$

Donc, si nous supposons que $\lim_{s \rightarrow +\infty} \lambda'(s)/\lambda(s) = 0$, nous obtenons

$$\begin{cases} \mathcal{Q} = -\Delta_0, \\ \Gamma = \Lambda. \end{cases} \quad (\text{I.107})$$

Ainsi les transformations de puissance de la forme $q(s) \sim \lambda q^*(s)$ conduisent vers les lois limites non standards $\mathcal{F}_{\lambda, -\Delta_0}$. De plus, il est possible de calculer une représentation analytique de q_n^* en définissant le terme d'erreur $\epsilon(s) = Q(s) + \Delta_0$. Inclure l'Eq. (I.67) mène alors à l'équation différentielle exacte

$$\partial_s \ln q^*(s) = -\Delta_0 + \epsilon(s). \quad (\text{I.108})$$

Résoudre cette équation donne

$$q_n^* = L^*(n)n^{-\Delta_0}, \quad (\text{I.109})$$

$$L^*(n) = \exp\left(\int_0^n \frac{\epsilon(\ln t)}{t} dt\right). \quad (\text{I.110})$$

Le facteur $L^*(n)$ correspond à un terme correctif dépendant de la structure fine de convergence de $\delta_0(s)$ et $\lambda(s)$. De plus, puisque $\epsilon(\ln n) \rightarrow 0$, l'Eq. (I.110) correspond

exactement à la représentation intégrale des fonctions variant lentement normalisées (cf. annexe A). En d'autres termes, $L^*(n)$ est une correction variant lentement du comportement en puissance de q_n^* , du moins si $\Delta_0 \neq 0$. Si $\Delta_0 = 0$, ce terme devient prépondérant. En particulier, pour les distributions de type Fréchet, $\Lambda_0 = \zeta$ implique que $q_n^* \approx 1/\zeta$. D'un autre côté, pour le domaine Gumbel, $\lambda_0(s) \rightarrow 0$ implique $L^*(n) \rightarrow +\infty$. Par exemple, dans le cas des lois en exponentielle étirée, nous avons $\delta_0(s) = \epsilon(s) \sim \frac{1}{s}$, par conséquent $q_n \approx K \ln n$ ($K > 0$). Le calcul de $\lambda_0(s)$ conduit à l'expression plus précise

$$q_n^* \approx \alpha \ln n. \quad (\text{I.111})$$

Par conséquent, l'usage d'une transformation de puissance appropriée ($q_n = \Lambda L^*(n)n^{|\zeta|}$) peut induire une loi appartenant au domaine Weibull à converger vers une distribution non standard $\mathcal{F}_{\lambda, -\Delta_0}$. D'un autre côté, même une transformation de puissance relativement lente ($q_n = \Lambda \alpha \ln n$ pour les lois en exponentielle de puissance) peut suffire pour induire une loi du domaine Gumbel à converger vers une loi limite de type Fréchet. Ce résultat est d'ailleurs parfaitement cohérent avec les résultats plus précis (mais plus spécifiques) obtenues dans [23] concernant la transition entre les domaines Gumbel et Fréchet.

Cependant, ne considérer que des lois parentes originaires des domaines standards permet d'obtenir uniquement des distributions non standards avec un \mathcal{Q} positif. Les domaines d'attraction non standards avec un paramètre $\mathcal{Q} < 0$ correspondent aux lois parentes dont la queue est trop lourde pour appartenir aux domaines standards mais qui peuvent tout de même être "normalisées" par des transformations de puissance dont l'exposant tend vers 0. Un exemple de ces lois serait typiquement les lois en puissances de logarithme avec

$$\overline{F}(x) = 1 - L(\ln x)(\ln x)^{-\frac{1}{\alpha}}, \quad (\text{I.112})$$

où L et sa dérivée L' sont des fonctions variant lentement normalisées avec $\alpha > 0$. Un court calcul montre que pour ces lois $\lambda_0(g(x)) \sim \alpha \ln x$ et $\Delta_0 = \alpha$. Par conséquent ces lois appartiennent aux domaines d'attraction de $\mathcal{F}_{\lambda, -\alpha}$ avec $\alpha > 0$ et $\lambda > 0$. La fonction de puissance caractéristique associée est une loi de puissance décroissante $q_n^* = L^*(n)n^{-\alpha}$. En allant plus loin, exposée à une transformation de puissance dont l'exposant décroît plus vite que $n^{-\alpha}$, ces lois de puissance logarithmique convergent vers une distribution de Fréchet de paramètre α .

Pour conclure, notre analyse heuristique a mis en lumière l'existence de transitions entre les domaines d'attraction standards et non standards lorsqu'on modifie la dépendance en n de la transformation de puissance $q(n)$. Nous avons également montré qu'il est possible d'utiliser ces transformations de puissances pour "normaliser" des lois au-delà des domaines d'attraction (un exemple typique étant les lois puissances de logarithme) de manière à les faire converger vers des distributions non dégénérées. Cet ensemble de transitions est résumé dans la Table 2.1. Une représentation plus compacte de ces transitions peut être obtenue en utilisant le ratio q_n/q_n^* comme clef de la classification des distributions limites comme le montre la Table 2.2.

Loi parente	Dépendance fonctionnelle de q_n				
	$q_n = L(n)n^Q, \quad Q < 0$	λ	$L_d(n)$	$L(n)n^Q, \quad Q > 0$	
Dom. Weibull $\zeta < 0, L^*(n)$	\mathcal{F}_ζ			$ \zeta , \lambda L^*$ $\mathcal{F}_{\lambda, Q}$	-
Dom. Gumbel $L^*(n)$	\mathcal{F}_0			$0, \lambda L^*$ \mathcal{F}_λ	-
Dom. Fréchet $\zeta > 0$	\mathcal{F}_0		$0, \lambda$ $\mathcal{F}_{\lambda\zeta}$		-
Puiss. log $\alpha > 0, L^*(n)$	\mathcal{F}_α	$-\alpha, \lambda L^*$ $F_{\lambda, - \zeta }$			-

TABLE 2.1 – Classification des distributions limites en fonction des dépendances fonctionnelles de l'exposant q_n . La position des transitions est décrite par le couple (Q, L) . Par exemple, pour les lois appartenant au domaine Weibull, la transition vers les domaines non standard occure pour $Q = |\zeta|$ et $L = \lambda L^*$. La fonction variant lentement $L^*(n)$ est une caractéristique de loi parente –cf l'Eq. (I.109). La fonction $L_d(n)$ est une fonction lentement variable tendant vers l'infini avec n . Le signe '-' signifie qu'aucune loi limite (non dégénérée) n'émerge.

Loi parente	Limite du ratio q_n/q_n^*		
	$q_n/q_n^* \rightarrow 0$	$q_n/q_n^* \rightarrow \Lambda > 0$	$q_n/q_n^* \rightarrow +\infty$
Dom. Fréchet, $\zeta > 0$	Gumbel \mathcal{F}_0	Fréchet \mathcal{F}_Λ	-
Dom. Gumbel, \mathcal{F}_0	Gumbel \mathcal{F}_0	Fréchet \mathcal{F}_Λ	-
Dom. Weibull, $\zeta < 0$	Weibull \mathcal{F}_ζ	Non standard $\mathcal{F}_{\Lambda, \zeta }$	-
Puiss. log, $\alpha > 0$	Fréchet \mathcal{F}_α	Non standard $\mathcal{F}_{\Lambda, -\alpha}$	-

TABLE 2.2 – Classification de la distribution limite en fonction de la limite du ratio q_n/q_n^* lorsque $n \rightarrow \infty$. Le signe '-' signifie qu'aucune loi limite (non dégénérée) n'émerge.

4 Approche alternative de la caractérisation des domaines d'attraction non standards

4.1 Facteurs de normalisation du maximum transformé

Dans la section précédente, des arguments heuristiques ont montré que seules des transitions très spécifiques sont possibles entre les domaines d'attraction standards et non standards. Dans le but de vérifier ces résultats, cette section présente une approche alternative de l'étude du maximum transformé dans laquelle nous essayerons de rapprocher les comportements de convergence des variables transformées et non transformées.

Pour ce faire, considérons une variable aléatoire W appartenant au domaine d'attraction de la loi limite \mathcal{F}_ζ . Par définition du domaine d'attraction, cela signifie qu'il existe une suite de normalisation (α_n, β_n) telle que

$$\forall x, \quad F_W(\alpha_n x + \beta_n)^n \xrightarrow{n \rightarrow +\infty} \mathcal{F}_\zeta(x). \quad (\text{I.113})$$

Comme nous l'avions rappelé dans le chapitre I, ces séquences de normalisation sont bien connues et des expressions analytiques sont disponibles dans la littérature [45, 53].

De manière similaire, la variable transformée par ω , M_n^U , converge en distribution si et seulement si il existe une fonction \mathcal{F} et une séquence (a_n, b_n) telles que

$$\forall x, \quad F_W(\omega_n(a_n x + b_n))^n \xrightarrow{n \rightarrow +\infty} \mathcal{F}(x). \quad (\text{I.114})$$

Cependant, il s'agit d'une généralisation du problème précédent, pour laquelle on ne connaît ni de conditions de convergence générales ni les suites de normalisation associées. Nous proposons dans la section suivante d'exploiter la remarquable similarité entre les deux équations précédentes pour obtenir ces conditions de convergence pour les transformations de puissance.

En premier lieu, on peut remarquer que la seule différence entre ces deux équations repose sur le terme $(\alpha_n x + \beta_n)$ dans l'Eq. (I.113) qui devient $\omega_n(a_n x + b_n)$ dans l'Eq. (I.114). Cette ressemblance suggère qu'un moyen simple d'obtenir la convergence en distribution du maximum transformé est d'exploiter notre connaissance des facteurs de normalisation de la distribution originale. S'il existe une suite de normalisation (a_n, b_n) telle que

$$\forall x, \quad \omega_n(a_n x + b_n) \underset{n \rightarrow +\infty}{\sim} \alpha_n \nu(x) + \beta_n, \quad (\text{I.115})$$

et si nous supposons de plus que

$$\lim_{n \rightarrow +\infty} F_W(\omega(a_n x + b_n))^n = \lim_{n \rightarrow +\infty} F_W(\alpha_n \nu(x) + \beta_n)^n, \quad (\text{I.116})$$

on obtient

$$\tilde{\mathcal{F}}(x) = \mathcal{F}_\zeta(\nu(x)). \quad (\text{I.117})$$

Par conséquent, si les conditions (I.116) et (I.115) sont vérifiées, il est possible de lier la distribution limite transformée et la distribution limite originale.

La condition (I.116) correspond à un problème de convergence assez technique. Nous supposons pour l'instant que ces conditions sont vérifiées. Nous montrons néanmoins dans l'annexe C que cette condition est bel et bien vérifiée avec notre choix de (a_n, b_n) .

La condition (I.115) est plus intéressante et définit à travers $\nu(x)$ le type de la transition. Pour une transformation de puissance, celle-ci peut être réécrite

$$\forall x, \quad (a_n x + b_n)^{1/q_n} \underset{n \rightarrow +\infty}{\sim} \alpha_n \nu(x) + \beta_n. \quad (\text{I.118})$$

Si on considère la Table 2.2, quatre transitions distinctes devraient être possibles. Pour les lois du domaine Gumbel ou Weibull, on devrait avoir soit $\nu(x) = x$ si $\lim_{n \rightarrow \infty} q_n/q_n^* = 0$ soit $\nu(x) = \ln(1 + \Lambda x)/\Lambda$ si $\lim_{n \rightarrow \infty} q_n/q_n^* = \Lambda > 0$. Parallèlement, pour le domaine de Fréchet de paramètre ζ , des transitions devraient apparaître pour des exposants q_n convergeant vers une valeur finie, associés à $\nu(x) = [(1 + \Lambda x)^{\zeta/\Lambda} - 1]/\zeta$ (avec $\Lambda = \lim_{n \rightarrow \infty} q_n/\zeta$). Enfin, pour des exposants q_n tendant vers 0, ces transitions devraient être associées à une fonction $\nu(x) = [\exp(\zeta x) - 1]/\zeta$.

Par suite, en s'appuyant sur les résultats obtenus en section 3, il est naturel d'étudier séparément le comportement de $(a_n x + b_n)^{1/q_n}$ lorsque q_n diverge, s'annule, ou converge asymptotiquement.

4.2 Exposants q_n divergents

Dans le cas où $\lim_{n \rightarrow \infty} q_n = +\infty$, l'expression $(a_n x + b_n)^{1/q_n}$ présente deux comportements asymptotiques distincts, ce qui conduit à l'existence de deux régimes de convergence différents. En premier lieu, si $a_n/b_n \xrightarrow{n \rightarrow +\infty} 0$, on obtient le comportement asymptotique suivant

$$(a_n x + b_n)^{1/q_n} \underset{n \rightarrow +\infty}{\sim} b_n^{1/q_n} \left(1 + \frac{a_n}{b_n q_n} x \right). \quad (\text{I.119})$$

La condition Eq. (I.118) est alors vérifiée si

$$\begin{cases} b_n = \beta_n^{q_n}, \\ a_n = b_n q_n \frac{\alpha_n}{\beta_n}. \end{cases} \quad (\text{I.120})$$

Ce choix de (a_n, b_n) est compatible avec l'hypothèse que $a_n/b_n \xrightarrow{n \rightarrow +\infty} 0$ si

$$q_n \frac{\alpha_n}{\beta_n} \xrightarrow{n \rightarrow +\infty} 0. \quad (\text{I.121})$$

Le terme α_n/β_n correspond au paramètre $\lambda_0(n)$ défini dans la section 3. Par suite, en utilisant la définition de l'exposant caractéristique $q_n^* = 1/\lambda_0(n)$ introduite dans l'Eq. (I.102), le résultat précédent énonce que si q_n est négligeable par rapport à q_n^* (c'est-à-dire que $\lim_{n \rightarrow \infty} q_n/q_n^* = 0$) alors le maximum transformé converge vers la même distribution limite que le maximum non transformé. Il s'agit bien du comportement attendu à partir des résultats de la section 3.

Le second comportement asymptotique survient lorsque que a_n/b_n est essentiellement constant. En factorisant b_n , on obtient alors

$$(a_n x + b_n)^{1/q_n} \underset{n \rightarrow +\infty}{\sim} b_n^{1/q_n} \left(1 + \frac{\ln(1 + \frac{a_n x}{b_n})}{q_n} \right). \quad (\text{I.122})$$

La condition (I.118) est vérifiée si

$$\begin{cases} b_n = (\beta_n)^{q_n}, \\ a_n = \lambda b_n, \\ \frac{q_n}{q_n^*} \underset{n \rightarrow +\infty}{\rightarrow} \lambda. \end{cases} \quad (\text{I.123})$$

Dans ce cas, $\nu(x) = (1 + \lambda \ln x)/\lambda$ implique que X_n converge vers les lois non standards. Plus précisément, si q_n est asymptotiquement équivalent à λq_n^* alors le maximum converge vers la distribution $\mathcal{F}_{\lambda, -\zeta}$. Une fois encore, nous retrouvons exactement les résultats de la section 3 pour les domaines Weibull et Gumbel.

4.3 Exposant q_n s'annulant à l'infini

La seconde transition intéressante apparaît lorsque q_n converge vers 0 à l'infini. À partir de la Table 2.2, nous savons qu'un terme exponentiel devrait apparaître dans $(a_n x + b_n)^{1/q_n}$. La manière la plus simple d'obtenir un tel terme est de supposer que $a_n = b_n q_n \zeta$. Nous avons alors

$$\begin{aligned} (a_n x + b_n)^{1/q_n} &= (b_n)^{\frac{1}{q_n}} (1 + q_n \zeta x)^{\frac{1}{q_n}} \\ &\underset{n \rightarrow +\infty}{\sim} (b_n)^{\frac{1}{q_n}} e^{\zeta x}. \end{aligned} \quad (\text{I.124})$$

En substituant $\nu(x) = [\exp(\zeta x) - 1]/\zeta$ dans l'équation, on peut satisfaire la condition Eq. (I.118) sous l'hypothèse que

$$\begin{cases} b_n = (\beta_n)^{q_n}, \\ \frac{\beta_n}{\alpha_n} \equiv q_n^* \underset{n \rightarrow +\infty}{\rightarrow} \frac{1}{\zeta}. \end{cases} \quad (\text{I.125})$$

Il est donc possible de passer du domaine Fréchet au domaine Gumbel en utilisant n'importe quelle transformation de puissance d'exposant convergeant vers 0, ce qui conforte notre analyse heuristique de la section précédente. Cela confirme le fait que le domaine Fréchet est très instable vis-à-vis d'une perturbation en fonction de puissance. N'importe quelle transformation de puissance s'annulant à l'infini est suffisante pour altérer la distribution limite d'une loi de type Fréchet et la faire ainsi converger vers la distribution Gumbel.

4.4 Exposants q_n de limite finie

Dans le cas d'une puissance convergeant vers une limite finie, on s'attend à observer uniquement des transitions entre des domaines Fréchet de paramètre ζ distincts. En tenant compte des possibles transformations affines, on peut définir sans perte de généralité $a_n = \lambda b_n$ et $b_n = \beta_n^{q_n}$. On obtient alors,

$$(a_n x + b_n)^{1/q_n} = \beta_n (1 + \lambda x)^{1/q_n}, \quad (\text{I.126})$$

et le choix $\nu(x) = [(1 + \lambda x)^{\zeta/\lambda} - 1]/\zeta$ mène à la condition

$$\begin{cases} q_n^* \xrightarrow{n \rightarrow +\infty} \frac{1}{\zeta}, \\ \frac{q_n}{q_n^*} \xrightarrow{n \rightarrow +\infty} \lambda. \end{cases} \quad (\text{I.127})$$

Nous observons, comme attendu, une transition entre le domaine Fréchet de paramètre ζ et le domaine Fréchet de paramètre λ . Si le choix de $\nu(x)$ peut apparaître assez arbitraire, il faut noter qu'il s'agit simplement d'une conséquence de la représentation choisie pour les lois limites Fréchet. D'autres choix de représentation ($F(x) = \exp(-x^{-1/\zeta})$) mènent à l'expression $\nu(x) = x^{\zeta/\lambda}$ qui s'avère bien plus simple. Cependant, il est justement remarquable que notre méthode par renormalisation nous ait permis de mettre en lumière cette transition malgré cette complexité apparente.

Avec cette transition entre les domaines Fréchet, nous avons retrouvé l'ensemble des transitions décrites par la Table 2.2 entre les domaines standards d'attraction et nos domaines non standards. Pour ce faire, nous avons utilisé seulement les intuitions obtenues par notre analyse de l'équation aux dérivées partielles du flot de renormalisation, combinées avec des arguments élémentaires sur les coefficients de normalisation. L'annexe C montre que ces arguments sont suffisants pour développer une preuve formelle de la convergence du maximum transformé.

5 Conclusion

Dans ce chapitre, l'approche par renormalisation du problème du maximum développé dans [61, 62, 17] fut étendue au cas où les variables W_i sont soumises à une transformation ω_n dépendante de la taille du système. La réduction de ce problème du maximum en une équation aux dérivées partielles s'avère une généralisation directe du cas standard et mène, dans le cas des transformations de puissance $U_{i,n} = W_i^{q_n}$, à une classification élégante des distributions limites.

Cette classification a permis de mettre en lumière l'existence de lois max-stables non standards reflétant les lois max-stables standards. Ces nouvelles lois limites sont étroitement liées aux lois limites standards. Cependant le comportement de l'équation aux dérivées partielles décrivant l'évolution de la distribution du maximum est plus complexe et met en jeu d'intrigantes interactions entre le taux de croissance de la transformation de puissance et la queue de la distribution.

Ces interactions furent l'objet d'une analyse plus en profondeur à travers l'étude du comportement asymptotique des paramètres de forçage apparaissant dans l'équation aux dérivées partielles du flot. Par un argument heuristique, il fut montré qu'il est possible de retrouver une légère approximation des domaines standards. De plus, le même argument nous conduit à une description intéressante des domaines d'attraction des lois limites non standards, révélant l'existence de transitions spécifiques entre les lois classiques et leurs reflets non standard. Ces transitions ne se révèlent qu'en variant la dépendance fonctionnelle q_n de la transformation, et elles sont notamment associées à des exposants caractéristiques q_n^* . Si $q_n/q_n^* \rightarrow 0$, la transformation de puissance est trop lente pour influencer la convergence du maximum vers sa

loi limite standard. Au contraire, si $q_n \sim \lambda q_n^*$ la convergence est altérée et la distribution converge vers une loi limite non standard. Faisant usage de l'intuition âprement gagnée grâce à l'équation du flot, il fut alors possible de confirmer l'existence de ces transitions grâce à une approche plus directe, basée sur l'étude des paramètres de normalisation. Cette approche nous permet d'ailleurs de détailler les mécanismes sous-jacents de ces transitions.

Deuxième partie

Ordre critique et transitions de phases

Introduction

Motivation

L'estimation des moments d'un ordre donné à partir d'un nombre fini d'observations d'une série temporelle $\{X_k, k = 1, \dots, n\}$ peut sembler être un problème à la fois *simple* et *naturel*. Dans les conditions standards, avec peu d'informations connues (ou conjecturées) a priori sur les données, l'estimateur classique des moments

$$S(n, q) = \frac{1}{n} \sum_{k=1}^n X_k^q \quad (\text{II.1})$$

s'impose comme un choix évident (du moins pour des variables aléatoires X_k positives). La théorie des probabilités fournit aux praticiens de nombreux résultats théoriques fort utiles concernant les performances de cet estimateur, notamment dans le cas asymptotique pour une durée d'observation tendant vers l'infini, c'est à dire $n \rightarrow +\infty$ (cf. chapitre I ou e.g., [18]). La loi forte des grands nombres, lorsqu'elle est applicable, montre que l'estimateur des moments est consistant : $\lim_{n \rightarrow +\infty} S(n, q) \stackrel{\text{p.s.}}{=} \mathbb{E}[X^q]$. Le théorème central limite quantifie plus précisément le comportement asymptotique. Cependant de tels résultats peuvent être interprétés comme la solution d'une situation très abstraite dans laquelle l'observateur s'intéresse à un seul ordre statistique et dispose d'une observation de durée potentiellement infinie. En pratique, la situation *naturelle* est souvent presque complètement opposée à cette formulation théorique : Les expérimentateurs doivent fournir des efforts considérables pour obtenir une observation de taille n suffisamment grande et se demandent donc naturellement jusqu'à quel ordre des moments il est possible d'obtenir une estimation correcte, étant donnée la taille finie des données observées $\{X_k, k = 1, \dots, n\}$. L'opposition dialectique entre les résultats théoriques classiques et ces questions plus pragmatiques est esquissée par la Fig. 2.2. Cette différence entre résultat théorique et questionnement pratique résulte pour partie du fait que les résultats théoriques requièrent que l'ordre des moments q soit complètement découplé de la taille du système n .

Or pour une observation de taille finie, il est bien connu qu'il est impossible d'estimer correctement les moments à tout ordre, même dans le cas où tous les moments sont finis ($\mathbb{E}[X^q] < +\infty$). En effet, pour n fixé,

$$S(n, q) \underset{q \rightarrow +\infty}{\sim} \frac{1}{n} \max_{k=1, \dots, n} X_k^q \neq \mathbb{E}[X^q]. \quad (\text{II.2})$$

Ce comportement correspond à un effet de *linéarisation*. Une fois n fixé, la courbe expérimentale de $\ln S(n, q)$ en fonction de q tend nécessairement vers une droite

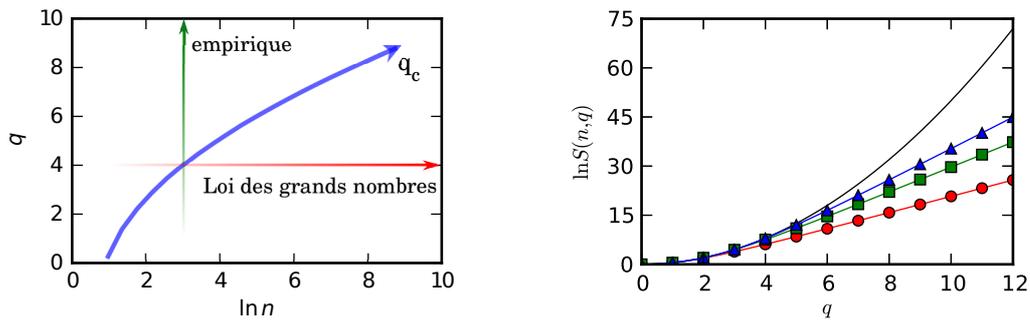


FIGURE 2.2 – **Plan (n, q) et effet de linéarisation.** À gauche, du point de vue théorique, la performance de l'estimateur des moments $S(n, q)$ est caractérisée, pour un q fixé, dans la limite $n \rightarrow +\infty$, i.e., le long d'une ligne horizontale dans le plan (n, q) . En pratique, cependant, le problème expérimental consiste à déterminer quel est le plus grand ordre q des moments que l'on peut évaluer correctement à partir d'une observation de longueur finie n donnée. Le praticien se déplace donc le long d'une ligne verticale dans le plan (n, q) . À droite, empiriquement, dans la limite $q \rightarrow +\infty$, pour un n fixé, $\ln S(n, q)$ se comporte nécessairement comme une fonction linéaire de q . Ceci est illustré ici par une distribution log-normale pour trois valeurs différentes de n ($n = 10^2$ (cercle rouge), $n = 10^3$ (carré vert), $n = 10^6$ (triangle bleu)); la ligne noire pleine montre la valeur théorique $\ln \mathbb{E}[X^q]$. Ces trois courbes montrent que la pente asymptotique et le point de divergence avec $\ln \mathbb{E}[X^q]$ dépendent tous deux de n . Cette partie a pour but de définir et estimer de façon pratique cet ordre critique $q^*(n)$ à partir duquel l'estimation commence à s'écarter de la moyenne d'ensemble $\mathbb{E}[X^q]$.

dans la limite $q \rightarrow +\infty$

$$\lim_{q \rightarrow +\infty} \ln S(n, q) = q \ln \max_{k=1, \dots, n} X_k - \ln n. \quad (\text{II.3})$$

Cet effet de linéarisation traduit donc une différence de comportement de la log-moyenne qui fait écho à la rupture de convergence que nous avons évoquée au chapitre I dans le cas des lois de puissance. Pour ces lois, ce changement de comportement apparaît dès lors que les moments théoriques divergent, ce qui n'est pas le cas ici. Cependant, lorsque cet effet de linéarisation apparaît, la valeur estimée des moments s'écarte de la valeur théorique ; sur ce point, on retrouve un parallèle fort intéressant avec les lois de puissance, si ce n'est que cette séparation ne se produit pas pour un ordre fixe mais apparaît lorsqu'on couple l'ordre des moments avec la taille du système.

Une question vient donc naturellement à l'esprit, question qui sera au cœur de cette partie II : quelle est la région des ordres $q \leq q^*$ au sein de laquelle les moments peuvent être correctement estimés, ce à partir d'une seule observation de taille finie n ? De façon surprenante, cette question semble avoir éveillé relativement peu d'intérêt dans la littérature [68]. Dans l'analyse de données réelles, il s'agit pourtant d'une question d'un intérêt évident. Cette question joue également un rôle vital dans l'analyse des invariances d'échelle plus connue sous le nom d'analyse multifractale. Dans celle-ci, les propriétés de covariance avec l'échelle sont mesurées à travers la dépendance selon l'échelle d'analyse a des moments d'ordre q de quantités multirésolutions calculées à partir des observations $\{X_k, k = 1, \dots, n\}$. On retrouve donc la même problématique que dans le cas i.i.d. mais pour une classe de processus aléatoires très différente.

Quand les X_k sont des variables aléatoires "régulières" (définies par le fait que leur fonction caractéristique est analytique en 0), et pour des observations i.i.d., une borne est donnée dans [68] :

$$q^*(n) \propto \frac{\ln n}{\ln \ln n}. \quad (\text{II.4})$$

Bien qu'il s'agisse d'un résultat très intéressant, son usage demeure limité en pratique par sa restriction aux observations i.i.d. de lois régulières mais aussi et surtout par le fait qu'il s'agit d'un résultat asymptotique dont le préfacteur est difficile à déterminer en pratique.

On peut noter que ce sous-ensemble des lois régulières correspond aux lois dont la queue décroît suffisamment rapidement pour que la fonction génératrice des moments g soit définie sur un voisinage de 0

$$\exists \epsilon > 0, g(\epsilon) \equiv \mathbb{E} [e^{\epsilon X}] < +\infty. \quad (\text{II.5})$$

En effet, la fonction génératrice des moments est directement reliée à la fonction caractéristique

$$\phi(z) = \mathbb{E} [e^{-izX}], \quad (\text{II.6})$$

par $g(t) = \phi(it)$. La fonction génératrice des moments correspond donc à la restriction de la fonction caractéristique sur l'axe imaginaire. De plus, on peut montrer que la fonction caractéristique est toujours analytique sur une bande horizontale [77] et que ses singularités se trouvent nécessairement sur l'axe imaginaire. Par conséquent, la fonction génératrice est définie sur un voisinage de 0 si et seulement si la fonction

caractéristique est analytique en 0. Cela implique notamment que tous les moments d'une loi régulière sont finis et pour k entier

$$\mathbb{E} [X^k] = g^{(k)}(0). \quad (\text{II.7})$$

Un exemple standard de ces fonctions régulières sont les variables aléatoires gaussiennes $\mathcal{N}_{\mu,\sigma}$ dont la fonction génératrice des moments $g(q) = \exp(\mu q + q^2\sigma^2/2)$ est définie sur l'ensemble de l'axe réel.

Il est donc intéressant d'essayer de généraliser le type de résultats décrits par l'Eq. (II.4) pour d'autres classes de variables aléatoires.

On peut commencer par étudier le cas des variables aléatoires i.i.d. que nous appellerons classe A. Dans ce contexte, un contrepoint aux lois régulières sont les lois de puissances et assimilées dont les moments ne sont finis que jusqu'à un ordre limite q_c . On peut cependant trouver un juste milieu entre ces deux classes de lois, en considérant des lois irrégulières dont tous les moments sont finis. Un exemple classique de loi de ce type est la loi log-normale.

On peut donc établir la classification suivante des lois aléatoires en fonction de leur comportement asymptotique :

type I Les lois régulières dont la fonction caractéristique est analytique en 0.

type II Les lois irrégulières dont tous les moments sont finis.

type III Les lois irrégulières dont les moments sont infinis au-delà d'un ordre limite q_c .

Il s'agit d'une déclinaison relativement naturelle des types de lois que l'on peut rencontrer dans le cadre de notre classe A. La sous-classe A-II, faisant le lien entre les lois régulières et les lois de puissance, retiendra l'essentiel de notre attention.

Comme nous l'avions évoqué plus haut, le phénomène de linéarisation des moments joue également un rôle majeur dans la phénoménologie de l'analyse multifractale. Cependant, dans ce contexte, les variables aléatoires étudiées sont fortement corrélées, et diffèrent notablement des variables i.i.d. de la classe A. Ces processus multifractals définissent donc une deuxième classe : la classe B. Pour cette classe de processus aléatoires, il est courant de rencontrer des lois dont les moments ne sont pas tous finis, nous nous focaliserons donc sur la sous-classe B-III.

Il faut cependant noter que si cette classification permet de clarifier les différentes formes de lois que nous allons étudier par la suite, les types ainsi définis sont trop généraux pour admettre une représentation pratique pour les calculs théoriques. Nous choisirons donc par la suite des sous-types plus spécifiques mais qui en capturent l'essence. Ceux-ci seront décrits en détail dans le chapitre 3.

Analogie avec la physique des systèmes désordonnée

L'étude de l'effet de linéarisation que nous proposons repose sur une analogie formelle entre estimateur des moments et fonction de partition du REM [40].

Le REM peut être représenté comme un modèle de spins, bien que la notion de spin n'y joue pas de rôle réellement significatif. On considère un système de N spins

tel que chacune des 2^N configurations des spins est associée à une énergie aléatoire E_i qui est tirée aléatoirement avec la loi

$$P(E) = \frac{1}{N\pi J^2} \exp\left(-\frac{E^2}{NJ^2}\right). \quad (\text{II.8})$$

Les énergies E_i sont i.i.d. de moyenne nulle mais leur variance dépend de la taille du système N . La fonction du partition du REM est alors décrite par

$$Z = \sum_{i=1}^{2^N} e^{-\beta E_i}. \quad (\text{II.9})$$

La fonction de partition Z est alors visiblement une somme de variables aléatoires. Si on pose $X_i \equiv \exp(-E_i)$, on obtient

$$Z = \sum_{i=1}^{2^N} X_i^\beta. \quad (\text{II.10})$$

Les variables X_i sont ici des lois log-normales. Il y a donc une analogie formelle entre $Z(N, \beta)$ et $2^N S(2^N, \beta)$. Pour aller plus loin, on peut réécrire la fonction de partition Z en fonction de la densité d'énergie $\rho(E)$,

$$\rho(E) = \sum_{i=1}^{2^N} \delta(E_i - E), \quad (\text{II.11})$$

sous la forme

$$Z = \int \rho(E) e^{-\beta E} dE. \quad (\text{II.12})$$

La densité $\rho(E)$ est également une variable aléatoire (plus précisément ρ est un processus ponctuel), mais ces fluctuations sont faibles lorsque N augmente si $\rho(E) > 1$, on peut donc supposer que $\rho(E) \approx \mathbb{E}[\rho(E)] = 2^N P(E)$. Cependant, lorsque $\mathbb{E}[\rho(E)] \ll 1$, pour une réalisation donnée, il n'y a typiquement aucune configuration d'énergie E , d'où $\rho(E) \approx 0$. On peut donc considérer que la fonction de partition typique du système peut se réduire à

$$Z_{\text{typ}} = \int_{E_{\min}}^{E_{\max}} \mathbb{E}[\rho(E)] e^{-\beta E} dE \quad (\text{II.13})$$

où $[E_{\min}, E_{\max}]$ représente le domaine où $\mathbb{E}[\rho(E)] > 1$. Cette réduction du domaine des énergies fait alors apparaître deux situations bien distinctes séparées par une température de transition vitreuse

$$T_g = \frac{J}{2\sqrt{\ln 2}}. \quad (\text{II.14})$$

En-deçà de cette température T_g , le système est gelé et son entropie par spin s'annule. Plus précisément, l'entropie par spin du système peut s'écrire

$$S(T) = \begin{cases} \ln 2 - \frac{J^2}{4T^2} & \text{si } T > T_g, \\ 0 & \text{si } T < T_g. \end{cases} \quad (\text{II.15})$$

La physique sous-tendant cette transition vitreuse dans le REM est assez simple. Au-dessus de la température critique, l'activation thermique est efficace et un grand nombre de configurations microscopiques sont explorées : le système est dans un état liquide. En-deçà de la température critique, l'activation thermique ne joue plus un rôle significatif et le système est gelé dans les quelques états de plus basses énergies. En conséquence, son entropie par degré de liberté s'annule (plus de détails sur le REM sont rappelés brièvement dans l'annexe B). Ce caractère non trivial vient du fait que l'entropie microcanonique $S(E)$ du système s'annule avec une pente finie à l'énergie minimale E_{\min} .

Trois points-clefs concourent alors à souligner l'intérêt d'une analogie entre REM et estimation des moments : le REM et l'estimation de moments impliquent tous deux l'évaluation de sommes de variables aléatoires élevées à une certaine puissance (qui constitue un paramètre de contrôle du problème) ; dans les deux cas, ces variables aléatoires appartiennent aux types II ou III ; enfin par construction, le REM introduit un couplage entre la taille du système et les distributions, or nous nous intéressons ici au comportement asymptotique de l'estimateur des moments dans le plan (n, q) . Au-delà des limites $q \rightarrow +\infty$ à n fixé et $n \rightarrow +\infty$ à q fixé, il est nécessaire d'introduire un couplage entre n et q si nous souhaitons étudier le comportement asymptotique de $S(n, q)$ le long de courbes (n, q) plus complexes. En poursuivant l'analogie, il est alors naturel de donner un rôle central à l'analogue de l'énergie libre par spin qu'est la log-moyenne $\ln S / \ln n$.

Cette démarche s'inscrit dans une féconde lignée d'analogies entre modèles de physique statistique et analyse multifractale qui ont permis de nombreuses avancées après la contribution déterminante de Parisi et Frisch [50] et les développements reportés dans [9]. On peut noter que l'effet de linéarisation a déjà été étudié à travers le prisme de modèles de physique statistique. Par exemple, les modèles proposés dans [92, 91] font appel à la notion physique d'*onde progressive* [79, 30].

Contributions : Vers un ordre critique

Cette seconde partie de ma thèse a pour but d'explorer jusqu'à quel point ces arguments de physique statistique, qui interviennent dans l'étude du REM pour expliquer la transition de phase vers une phase à entropie nulle, nous permettent de comprendre l'effet de linéarisation dans l'estimation des moments que ce soit dans le cas i.i.d. ou dans le cadre de l'analyse multifractale. Cette contribution complète et enrichit donc les connexions entre analyse multifractale et physique statistique [50, 97, 9, 49].

Dans ce contexte, le but premier de cette partie II est de définir un ordre critique $q^*(n)$ jusqu'auquel il est possible d'estimer correctement tous les moments d'ordre $q \leq q^*(n)$ à partir d'une observation de taille n finie de variables aléatoires positives X_k , supposées être de la forme $X_k = \exp Y_k$. Nous précisons dans le chapitre 3 les classes de variables Y_k que nous considérons.

La définition de cet ordre critique repose sur une argumentation heuristique détaillée et argumentée, empruntée à l'analyse du REM [40, 86]. Celle-ci est cependant fortement soutenue, dans le cas i.i.d., par une analyse théorique complète conduite par Ben Arous et ses collaborateurs [10] ; cette construction est détaillée dans le chapitre 4. Nous essaierons dans ce chapitre de mettre en avant les nombreux points

communs entre le cas i.i.d. et le cas multifractal. Nous sonderons également les limites de l'analogie entre REM et processus multifractal en analysant numériquement les rapports de participation associés à un processus multifractal.

Enfin, dans le chapitre 5, nous développerons une procédure pour l'estimation pratique de $q^*(n)$ pour des observations de taille finie dans le cas i.i.d.. Les performances de cet estimateur seront étudiées théoriquement et numériquement. Pour mieux résister à l'épreuve des données réelles, nous proposons quelques améliorations de cette procédure d'estimation pour tenir compte de la présence d'observations interdépendantes. Les performances de ces estimateurs modifiés sont étudiées de manière détaillée en utilisant des simulations de Monte-Carlo (cf. Section 4). Le code source implémentant ces procédures en MATLAB et la documentation afférente sont disponibles sur demande.

Chapitre 3

Classes de variables aléatoires

Nous allons commencer par définir précisément les types de variables aléatoires que nous comptons étudier. Comme annoncé dans l'introduction, nous nous intéressons à deux classes très différentes de processus aléatoires : d'un côté, nous souhaitons étudier l'effet de linéarisation pour des variables i.i.d. irrégulières de moments finis : la classe A-II. De l'autre côté, les processus multifractals donnent un exemple très intéressant de variables corrélées pour lesquelles l'effet de linéarisation joue un rôle majeur dans l'estimation pratique ; ils nous fourniront donc un exemple de lois de la classe B.

1 Classe A : Variables i.i.d.

1.1 Type II : Lois irrégulières de moments finis

Log-normale :

Un exemple typique est la loi log-normale. Des variables aléatoires i.i.d. de lois log-normales X peuvent être générées facilement en prenant l'exponentielle de variables aléatoires i.i.d. gaussiennes : $X = \exp Y$ avec $Y \stackrel{\mathcal{D}}{=} \mathcal{N}_{\mu, \sigma}$. La densité de probabilité de X est alors

$$p_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right). \quad (\text{II.16})$$

On peut aussi remarquer que la fonction cumulative de Y ($F_Y \equiv \mathbb{P}(Y < y)$) vérifie $\ln \overline{F}(y) \approx -y^2/2$ pour $y \rightarrow +\infty$.

Dans le présent chapitre, en s'inspirant des extensions rencontrées naturellement à la fois dans le contexte du REM [40, 86] ou du multifractal (cf. e.g., [108]) mais aussi dans le cadre théorique développé par Ben Arous et ses collaborateurs [10], nous considérons une classe de variables plus large en généralisant le cas log-normal : tout d'abord, le comportement en loi de puissance y^2 est étendu au cas y^ρ . Ensuite, nous considérons un comportement en quasi-loi de puissance, en exigeant uniquement des propriétés asymptotiques qui capturent l'essence mathématique de ce comportement en loi de puissance sans fixer exactement la forme de la densité de probabilités, dans l'esprit de [27].

Lois log-quasi exponentielles étirées :

Pour ce faire, nous définissons le sous-type des lois log-quasi exponentielles étirées. Ces lois doivent s'exprimer sous la forme

$$X = \exp Y, \tag{II.17}$$

où la fonction cumulative de Y , F_Y est supposée être une fonction continue :

$$\bar{F}(y) = e^{-\Psi(y)}. \tag{II.18}$$

De plus, on requiert que la fonction Ψ soit une fonction régulièrement variable d'indice ρ [19], c'est-à-dire qu'elle vérifie :

$$\Psi(y) = L(y)y^\rho, \text{ avec } \rho > 1, \tag{II.19}$$

où L est une fonction lentement variable (cf. annexe A).

De plus, pour des raisons techniques que nous rendrons plus claires dans le chapitre 4, nous nous restreignons au cas des fonctions $\Psi(y)$ "deux fois normalisées" (cf. l'annexe A). En d'autres termes, nous imposons que

$$\lim_{y \rightarrow +\infty} \frac{yL'(y)}{L(y)} = 0, \tag{II.20}$$

$$\lim_{y \rightarrow +\infty} \frac{y^2L''(y)}{L(y)} = 0. \tag{II.21}$$

Ces conditions imposent une certaine forme de régularité de la dérivée de L . En particulier, l'équation (II.20) exclut les fonctions variant lentement oscillant de plus en plus rapidement. Ainsi on peut vérifier qu'avec la condition (II.20), Ψ vérifie

$$\lim_{y \rightarrow +\infty} \frac{y\Psi'(y)}{\Psi(y)} = \rho. \tag{II.22}$$

De manière similaire, l'équation (II.21) régule les oscillations potentielles de L'' vis-à-vis de L . De plus, si on s'intéresse au logarithme de la densité de probabilité

$$\phi(y) = -\ln p_Y(y), \tag{II.23}$$

on peut vérifier que Ψ et ϕ sont reliées par

$$\phi(y) = \Psi(y) - \ln \Psi'(y). \tag{II.24}$$

À partir de là, on peut montrer que si $\Psi(y) = y^\rho L(y)$ avec L une fonction lentement variable deux fois normalisée alors il existe une fonction L_2 lentement variable normalisée telle que $\phi(y) = y^\rho L_2(y)$. Cela entraîne notamment que :

$$\lim_{y \rightarrow +\infty} \frac{y\phi'(y)}{\phi(y)} = \rho. \tag{II.25}$$

Dans le cadre de ces lois quasi-exponentielles étirées, la fonction cumulative et la densité de probabilité ont donc un comportement assez semblable et il est nécessaire d'imposer que $L(y) = x^{-\rho}\Psi(y)$ soit deux fois normalisée afin que ϕ soit suffisamment régulière.

1.2 Interprétation

L'équation (II.19) combinée avec les conditions (II.20) et (II.21) contrôle le comportement de la queue positive de la distribution Y et définit donc une large classe de variables aléatoires X , dont tous les moments sont finis, mais dont la fonction caractéristique n'est pas analytique en 0. L'intérêt majeur de cette classe de variables aléatoires provient de sa position intermédiaire entre d'une part les lois de type I, dont tous les moments sont finis et dont la queue décroît exponentiellement rapidement (ce qui implique que leur fonction caractéristique est analytique en 0, par exemple la valeur absolue d'une gaussienne) et les lois de type III, dont les moments ne sont finis que jusqu'à un ordre donné (un exemple étant les distributions de Pareto). Le paramètre ρ joue un rôle clef dans l'étude de cette région intermédiaire des lois de type II : de manière stricte, ces trois types de lois sont mutuellement exclusifs. Cependant, à partir de la classe des lois log-quasi-exponentielles étirées, il est possible d'approcher des lois de type I en faisant tendre ρ vers $+\infty$ et réciproquement lorsque ρ tend vers 1, le comportement des lois log-quasi-exponentielles étirées se rapproche de celui des lois de type III.

1.3 Conséquences

À partir des définitions (II.18) et (II.19), il est possible de déduire un certain nombre de propriétés des variables Y (et leurs conséquences sur X). La fonction de densité de probabilité p_Y de Y peut s'écrire :

$$p_Y(y) = \Psi'(y)e^{-\Psi(y)}. \quad (\text{II.26})$$

Tous les moments de X sont finis et peuvent s'écrire :

$$\mathbb{E}[X^q] = \int_{-\infty}^{+\infty} \Psi'(y)e^{qy-\Psi(y)} dy = \int_{-\infty}^{+\infty} \Psi'(y)e^{qy-y^\rho L(y)} dy. \quad (\text{II.27})$$

Néanmoins, la fonction génératrice des moments de X ,

$$\mathbb{E}[e^{tX}] = \int_0^{+\infty} \Psi'(y)e^{t \exp(y) - L(y)y^\rho} dy, \quad (\text{II.28})$$

est finie uniquement sur \mathbb{R}^- et la fonction caractéristique de X n'est donc pas analytique en 0 (voir [77] pour des conséquences plus détaillées de cette non-analyticité). Il est intéressant de remarquer que ces propriétés entraînent que les variables aléatoires Y sont de type I. Leur fonction génératrice est en effet définie sur \mathbb{R} car

$$\mathbb{E}[e^{qY}] = \mathbb{E}[X^q]. \quad (\text{II.29})$$

On a donc généré des variables aléatoires de type II en exponentiant un sous-ensemble de lois de type I.

1.4 Exposant local de lois de puissance

Dans la définition (II.19), le paramètre ρ est naturellement désigné sous le nom d'exposant de loi de puissance. Cependant, dans la suite de notre étude, un exposant

local joue un rôle crucial. Nous parlons d'exposant *local* pour désigner le comportement pré-asymptotique de la loi, c'est à dire pour un y fini :

$$\rho_l(y) = \frac{\partial \ln \Psi(y)}{\partial \ln y} = \frac{y\Psi'(y)}{\Psi(y)}. \quad (\text{II.30})$$

On peut vérifier que les conditions (II.20) et (II.21) impliquent :

$$\lim_{y \rightarrow +\infty} \rho_l(y) = \rho, \quad (\text{II.31})$$

$$\lim_{y \rightarrow +\infty} \rho'_l(y) = 0. \quad (\text{II.32})$$

1.5 Exemples

Afin de disposer d'une variété intéressante de lois tests, nous définissons la famille des lois log-gamma étirées $\Gamma_{\alpha,\beta}^\rho$. Ces lois sont définies par

$$Y = \Gamma_{\alpha,\beta}^{1/\rho}, \quad (\text{II.33})$$

où $\Gamma_{\alpha,\beta}$ est une variable de loi gamma de paramètre de forme α et de paramètre d'échelle β

$$p_{\Gamma_{\alpha,\beta}}(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}}, \quad (\text{II.34})$$

où $\Gamma(\alpha) = \int_{-\infty}^{+\infty} t^{\alpha-1} e^{-t} dt$ est la fonction gamma [102]. On peut vérifier que ces lois log-gamma étirées appartiennent bien à la classe des lois log-quasi-exponentielles étirées. Dans le cas $\rho = 1$ et $\beta = 1$,

$$\Psi(y) = -\ln(\gamma_n(\alpha, y)) \quad (\text{II.35})$$

où $\gamma_n(\alpha, y)$ est la fonction gamma incomplète inférieure normalisée [102]

$$\gamma_n(\alpha, y) = \frac{\int_0^y t^{\alpha-1} e^{-t} dt}{\Gamma(\alpha)}. \quad (\text{II.36})$$

Asymptotiquement, pour $y \rightarrow +\infty$, $\gamma_n(\alpha, y) \approx y^\alpha e^{-y}$, on peut donc en déduire que

$$\Psi(y) = yL(y), \quad (\text{II.37})$$

avec L une fonction lentement variable. Les lois $\Gamma_{\alpha,\beta}$ forment donc une famille de loi quasi-exponentielle étirée de paramètre $\rho = 1$. À partir de cette famille, il est facile de générer une famille plus générale explorant l'ensemble des valeurs de ρ . En effet, si une variable Y suit une loi de type quasi-exponentielle étirée de paramètre $\rho = 1$, alors $\beta Y^{1/\rho}$ appartient à la classe des lois de paramètre ρ . Les lois gamma étirées $\Gamma_{\alpha,\beta}^\rho$ correspondent donc à des lois gamma $\Gamma_{\alpha,\beta}$ qui ont été "étirées" afin de donner naissance à des lois quasi-exponentielles de paramètre ρ .

Dans les analyses numériques conduites dans le chapitre 5, nous utilisons deux sous-familles particulièrement utiles de lois paramétrées par ρ

- Les distributions log-Weibull correspondent au cas $\alpha = 1$ et $\beta = 1$. Elles vérifient

$$F(y) = 1 - e^{-y^\rho}, \quad \text{avec } \rho > 1, \quad y > 0. \quad (\text{II.38})$$

Cela implique notamment que leur exposant de puissance local est une constante

$$\forall y, \quad \rho_l(y) \equiv \rho. \quad (\text{II.39})$$

- Nous définissons les lois log-gamma conjuguées par $\alpha = \rho$ et $\beta = 1$. Leur densité de probabilité s'exprime simplement

$$p_Y(y) = \frac{1}{2\Gamma(1 + \frac{1}{\rho})} e^{-|y|^\rho}. \quad (\text{II.40})$$

Cela implique que leur fonction cumulative complémentaire prend la forme

$$F_Y(y) = \frac{1}{2\Gamma(1 + 1/\rho)} \int_{-\infty}^y e^{-|t|^\rho} dt, \quad \text{avec } \rho > 1. \quad (\text{II.41})$$

- La distribution log-normale correspond à un cas spécial des lois log-gamma conjuguées, avec $\rho = 2$ et $\beta = \sqrt{1/2}$.

2 Classe B : processus multifractals

L'analyse multifractale est aujourd'hui considérée comme un outil canonique pour l'étude des propriétés de mise à l'échelle et des fluctuations de la régularité des séries temporelles (ou des champs de dimension d) [101, 67, 108, 5]. En pratique, cette analyse multifractale revient essentiellement au calcul des moyennes temporelles ou spatiales (des $q^{\text{ème}}$ moments) de $X(a, t)$, quantités dépendantes du temps et de l'échelle. Des exemples typiques de telles quantités sont les accroissements $X(t+a) - X(t)$ du signal X [49, 50], les coefficients en ondelettes [9] ou les coefficients d'ondelettes dominants [108]. Cela donne naissance aux fameuses fonctions de structure $S(a, n, q) = \frac{1}{n} \sum_{k=1}^n |X(a, t_k)|^q$. Nonobstant la corrélation des variables X , ces fonctions de structure correspondent exactement aux estimateurs des moments des quantités multirésolutions $X(a, t_k)$ avec toutefois une dépendance jointe sur l'échelle d'analyse a et la localisation (temporelle ou spatiale) t . En pratique, l'analyse multifractale suppose que ces fonctions de structure se comportent comme des lois de puissance vis à vis de l'échelle d'analyse a sur un intervalle $a_m < a < a_M$, avec $a_M/a_m \gg 1$,

$$S(a, n, q) \simeq S_0(q) a^{\zeta(q)}, \quad (\text{II.42})$$

et cette analyse se concentre sur l'estimation de l'exposant d'échelle $\zeta(q)$. Cet exposant $\zeta(q)$ est une fonction concave de la statistique d'ordre q qui ressemble fortement à la log-moyenne puisque

$$\zeta(q) = \lim_{a \rightarrow 0} \frac{\ln S(a, n(a), q)}{\ln a}. \quad (\text{II.43})$$

Il a été observé et justifié que l'exposant $\zeta(q)$ se comporte nécessairement comme une fonction linéaire de q au-delà d'une certaine valeur – voir [89, 90] pour les descriptions originales du phénomène, [96] pour son analyse dans le cas des cascades multiplicatives de Mandelbrot, et [74, 6] (ou respectivement [11]) pour des contributions plus récentes plus orientées traitement du signal (ou respectivement analyse statistique) dans le cadre de l'analyse multifractale de réalisations des trajectoires de processus stochastiques. Prenant exemple sur [74], nous nous référerons à cet effet sous le nom d'*effet de linéarisation* en analyse multifractale, faisant ainsi le parallèle avec l'effet de linéarisation apparaissant dans le cas i.i.d..

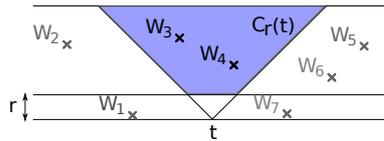


FIGURE 3.1 – Construction de la cascade de Poisson composée

2.1 Cascades de Poisson composées

La quintessence des exemples de processus multifractals sont les célèbres cascades multiplicatives de Mandelbrot (voir par exemple [82, 49, 101] pour des revues). Cependant, pour notre étude, nous utiliserons plutôt les cascades de Poisson composées plus récemment introduite par Barral et Mandelbrot dans [13] (voir aussi [31]), parce que leurs propriétés statistiques sont plus faciles à manier de manière tant pratique que théorique : leurs accroissements sont stationnaires et caractérisés par une invariance d'échelle continue, i.e., l'Eq. (II.42) est valide pour un intervalle continu d'échelles $a \in [a_m, a_M]$. Pour ces raisons, ces cascades sont désormais considérées comme le processus multifractal de référence. La cascade $Q_r(t)$ correspond à un produit de variables aléatoires positives i.i.d. W_i , appelées multiplicateurs et associées à un nuage de points (t_i, r_i) distribués aléatoirement sur un rectangle

$$I_{r,L} = \left\{ (t', r') : r \leq r' \leq 1, -\frac{1}{2} \leq t' \leq L + \frac{1}{2} \right\}. \quad (\text{II.44})$$

Plus précisément, la cascade est définie pour $r > 0$ par

$$Q_r(t) = B_r(t) \prod_{(t_i, r_i) \in \mathcal{C}_r(t)} W_i, \quad (\text{II.45})$$

où seuls les multiplicateurs associés aux points appartenant au cône

$$\mathcal{C}_r(t) = \left\{ (t', r') : r \leq r' \leq 1, t - \frac{r'}{2} \leq t' \leq t + \frac{r'}{2} \right\} \quad (\text{II.46})$$

sont pris en compte comme représenté sur la figure 3.1. Les points (t_i, r_i) sont tirés à partir d'un processus de Poisson de mesure d'intensité $dm(r, t)$ sur le rectangle $I_{r,L}$. Le paramètre $B_r(t)$ est une constante de normalisation telle que $\mathbb{E}[Q_r(t)] = 1$.

Il a été montré que cette cascade vérifie la relation fondamentale :

$$\mathbb{E}[Q_r(t)^q] = \exp[-\varphi(q) m(\mathcal{C}_r(t))], \quad q > -1 \quad (\text{II.47})$$

où $m(\mathcal{C}_r(t)) = \int_{\mathcal{C}_r(t)} dm(r', t')$ correspond à la mesure du cône $\mathcal{C}_r(t)$, et où $\varphi(q)$ est définie par

$$\varphi(q) = (1 - \mathbb{E}[W^q]) - q(1 - \mathbb{E}[W]). \quad (\text{II.48})$$

Pour des raisons de simplicité, nous ne considérons que les cas où $\varphi(q)$ est une fonction concave régulière, dont un exemple typique est le cas log-normal $\varphi(q) = cq(1 - q)$, avec une constante $c > 0$. Le mouvement de Poisson composé, $Z(t)$, est alors obtenu en intégrant la cascade, $Q_r(t)$, selon l'axe temporel, et en prenant ensuite la limite $r \rightarrow 0$:

$$Z(t) = \lim_{r \rightarrow 0} \int_0^t Q_r(s) ds. \quad (\text{II.49})$$

Cette définition donne naissance à un processus bien défini à condition que $\varphi(1^-) \geq -1$ [13].

2.2 Variation d'échelle et propriétés multifractales

Les accroissements $X(a, t)$ du mouvement $Z(t)$, définis par

$$X(a, t) = Z(t + a) - Z(t), \quad (\text{II.50})$$

avec $a > 0$, sont positifs, du fait de la positivité de la cascade $Q_r(t)$ (cf. l'Eq. (II.49)). Si la mesure d'intensité du processus de Poisson se factorise sous la forme $dm(r, t) = g(r)drdt$, les accroissements $X(a, t)$ correspondent alors à un processus stationnaire [31], ce qui signifie que toutes les propriétés statistiques ne dépendent pas du temps t . De façon intéressante, il a été également montré que les moments $\mathbb{E}[X(a, t)^q]$ ne sont finis que pour $-1 < q < q_c$, où q_c est déterminé par [12] :

$$q_c = \sup \{q \geq 1 : q + \varphi(q) - 1 \geq 0\}. \quad (\text{II.51})$$

On s'attend donc à ce que la probabilité $\mathbb{P}(X(a, t) \geq x)$ se comporte asymptotiquement comme $\mathbb{P}(X(a, t) \geq x) \sim x^{-q_c}$ lorsque $x \rightarrow +\infty$, et par conséquent les variables $X(a, t)$ ont des queues lourdes.

De plus, si $g(r)dr = c[dr/r^2 + \delta_{\{1\}}(dr)]$ (comme proposé dans [12]), où $\delta_{\{1\}}(dr)$ dénote une masse ponctuelle à $r = 1$, la loi infiniment divisible sous-tendant la construction de $Z(t)$ induit les propriétés d'invariance d'échelle suivantes

$$\mathbb{E}[X(a, t)^q] = C_q a^{\lambda(q)}, \quad (\text{II.52})$$

pour $-1 < q < q_c$, avec $\lambda(q) = q + \varphi(q)$ [12, 31]. Il faut noter que (II.52) est valide pour tout a dans l'intervalle $0 < a < L$.

Le spectre multifractal $D(h)$ n'est autre que la dimension de Hausdorff de l'ensemble des points t de la ligne réelle qui possèdent le même exposant de singularité h (aussi appelé exposant de Hölder)

$$X(a, t) \simeq ca^{h(t)}, \quad a \rightarrow 0. \quad (\text{II.53})$$

La fonction $D(h)$ procure donc une description *globale* des fluctuations *locales* d'une réalisation de la trajectoire $X(t)$. Pour une introduction détaillée à l'analyse multifractale, le lecteur est invité à se référer par exemple à [67].

À partir des résultats obtenus dans [13], il peut être inféré que le spectre multifractal $D(h)$ du mouvement de Poisson composé peut être dérivé de la transformée de Legendre concave de $\lambda(q)$

$$\lambda^*(h) = \inf_q \{qh - \lambda(q)\}, \quad (\text{II.54})$$

et peut donc être exprimé comme :

$$D(h) = \begin{cases} 1 + \lambda^*(h), & \text{si } 1 + \lambda^*(h) \geq 0, \\ -\infty, & \text{sinon.} \end{cases} \quad (\text{II.55})$$

Il est également intéressant de quantifier la structure de dépendance de $X(a, t)$. Il a été montré que la fonction de corrélation à deux points de $X(a, t)$ prend la forme suivante [107] :

$$\mathbb{E}[X(a, t)X(a, t + s)] = \sigma^2 (|s + a|^{\lambda(2)} + |s - a|^{\lambda(2)} - 2|s|^{\lambda(2)}), \quad (\text{II.56})$$

où σ^2 est la constante

$$\sigma^2 = \frac{1}{\lambda(2)(\lambda(2) - 1)}. \quad (\text{II.57})$$

Cette fonction de corrélation à deux points peut être réécrite sous la forme

$$\mathbb{E}[X(a, t)X(a, t + s)] = a^{\lambda(2)} f\left(\frac{s}{a}\right) \quad (\text{II.58})$$

où

$$f(u) = \sigma^2 \left(|u + 1|^{\lambda(2)} + |u - 1|^{\lambda(2)} - 2|u|^{\lambda(2)} \right). \quad (\text{II.59})$$

L'équation (II.58) montre que les variables $X(a, t)$ sont corrélées avec une échelle de temps caractéristique a . Ce résultat s'avérera fort utile dans le chapitre suivant à la section 3.3. Cependant, il est nécessaire de rappeler que ce temps de corrélation a apparaissant dans la fonction de corrélation à deux points de $X(a, t)$ est un effet induit par la procédure de "mesure" elle-même. En d'autres termes, c'est une conséquence du fait que $X(a, t)$ correspond à un incrément du signal $Z(t)$ à l'échelle a . Le signal $Z(t)$, lui, est invariant d'échelle et ne possède donc en aucune manière une échelle de temps caractéristique.

2.3 Propriétés de grandes déviations

La statistique des incréments a été caractérisée par les moments $\mathbb{E}[X(a, t)^q]$, décrit par l'Eq. (II.52). Il est également possible de caractériser cette statistique par sa densité de probabilité. Pour des raisons qui se clarifieront dans les sections suivantes, il est pratique de considérer la variable aléatoire $H_a(t)$ définie comme

$$H_a(t) = \frac{\ln X(a, t)}{\ln a}. \quad (\text{II.60})$$

H_a est directement reliée à la variable Y_a définie de manière similaire au cas i.i.d. par

$$Y_a(t) = -|\ln a| H_a(t). \quad (\text{II.61})$$

Il faut noter que dans l'Eq. (II.53), $H_a(t)$ correspond, dans la limite des petites échelles $a \rightarrow 0$, à l'exposant de Hölder $h(t)$ [108]. La densité de probabilité des variables $H_a(t)$, pour un t donné, est notée $p_a(h)$. Cette densité ne dépend pas du temps, à cause de la stationnarité du processus $X(a, t)$.

Nous souhaitons montrer que $p_a(h)$ obéit à un principe de grande déviation dans la limite $a \rightarrow 0$, plus précisément

$$p_a(h) \approx e^{-|\ln a| \psi(h)}. \quad (\text{II.62})$$

Une manière classique de dériver un principe de grande déviation est le théorème de Gärtner-Ellis [105, 54, 44], celui-ci permettant également de déterminer une expression explicite de $\psi(h)$. En premier lieu, nous définissons

$$\mu(q) = \lim_{a \rightarrow 0} \frac{1}{|\ln a|} \ln \mathbb{E} \left[e^{q|\ln a| h_a(t)} \right]. \quad (\text{II.63})$$

La fonction $\mu(q)$ peut être calculée à partir de l'Eq. (II.52), menant à

$$\mu(q) = -\lambda(-q), \quad -q_c < q < 1. \quad (\text{II.64})$$

À partir des propriétés de $\lambda(q)$, on peut déduire que $\mu(q)$ est une fonction convexe régulière. Si on suppose l'existence de la fonction de grande déviation $\psi(h)$ introduite dans l'équation (II.62),¹ le théorème de Gärtner-Ellis donne l'expression suivante

$$\psi(h) = \sup_q \{qh - \mu(q)\}. \quad (\text{II.65})$$

L'existence d'un moment limite q_c implique que l'équation précédente n'est valide que pour $h > h_c$ avec h_c tel que

$$h_c = \mu'(-q_c) = \lambda'(q_c). \quad (\text{II.66})$$

En utilisant l'Eq. (II.65) et la propriété $\lambda(q_c) = 1$ résultant de l'Eq. (II.51), h_c peut être caractérisé par

$$\psi(h_c) = 1 - q_c h_c, \quad (\text{II.67})$$

une propriété dont nous ferons usage par la suite.

Il faut noter que $\psi(h)$ est la transformée de Legendre convexe de $\mu(q)$, qui est plus employée que la transformée de Legendre concave apparaissant dans l'Eq. (II.54). En utilisant l'Eq. (II.64), les deux transformées de Legendre peuvent être reliées de la façon suivante :

$$\begin{aligned} \psi(h) &= \sup_q \{qh + \lambda(-q)\} \\ &= \sup_{q'} \{-q'h + \lambda(q')\} \\ &= -\inf_{q'} \{q'h - \lambda(q')\}, \end{aligned} \quad (\text{II.68})$$

avec $q' = -q$, menant à

$$\psi(h) = -\lambda^*(h), \quad (\text{II.69})$$

ou de façon équivalente, $\psi(h) = 1 - D(h)$, pourvu que $D(h) \geq 0$. Finalement, nous remarquons que le comportement aux grandes déviations de $p_a(h)$ pour $a \rightarrow 0$ peut être réécrit comme

$$p_a(h) \approx a^{\psi(h)}, \quad (\text{II.70})$$

ce qui s'accorde parfaitement avec le formalisme multifractal thermodynamique utilisé pour les applications pratiques de l'analyse multifractale, qui repose sur l'hypothèse heuristique que $p_a(h) \sim a^{1-D(h)}$ [50, 9, 49].

Nous pouvons également observer que pour $h < h_c$, λ^* développe une branche linéaire

$$\lambda^*(h) = h q_c - 1. \quad (\text{II.71})$$

Cette expression peut différer de la fonction de grande déviation ψ , mais procure au moins l'enveloppe convexe de ψ , qui est cohérente avec la divergence des moments de X pour $q \geq q_c$: en effet, si $p_a(h) \approx a^{-h q_c}$, alors $\mathbb{P}(X(a, t) \geq x) \sim x^{-q_c}$.

Si on revient maintenant à la variable Y que nous avons utilisée dans le cas i.i.d., on peut noter que

$$Y = -|\ln a| H_a. \quad (\text{II.72})$$

1. Si $\mu(q)$ était fini pour tout réel q , le théorème de Gärtner-Ellis impliquerait directement l'existence de ψ . Dans le cas présent, μ n'est fini que pour $-q_c < q < 1$, nous ne pouvons que conjecturer l'existence de ψ .

Par suite, on peut relier Ψ et ψ par

$$P_Y(y) = \Psi'(y)e^{\Psi(y)} = \frac{p_h(y/\ln a)}{|\ln a|}. \quad (\text{II.73})$$

Soit

$$\Psi(y) + \ln \Psi'(y) = -|\ln a|\psi(y/\ln a) - \ln |\ln a|. \quad (\text{II.74})$$

La fonction ψ représente donc une forme invariante d'échelle de Ψ dans le cas multifractal. La variable H_a permet donc d'intégrer implicitement la dépendance de Ψ en fonction de a .

On peut ainsi distinguer trois différences majeures entre la classe A-II et la classe B. Premièrement, la présence d'un ordre limite q_c signifie que le comportement asymptotique de la classe B correspond au type III. Deuxièmement, cette classe B est également fortement non i.i.d., les corrélations sont à longue portée. Troisièmement, les lois des accroissements $X(a, t)$ dépendent de l'échelle d'analyse a .

3 Approche unifiée

Si on revient au problème de la somme de variables aléatoires, ces différences signifient que dans la classe B, l'échelle d'analyse a constitue un paramètre vital du problème et éclipse partiellement le nombre d'observations n ; au contraire dans la classe A, l'échelle d'analyse a n'a aucun sens physique et seul le nombre d'observations n importe. Il y a donc une certaine symétrie de comportement entre la classe A et la classe B, ce malgré leur différence, et il est possible d'exploiter cette symétrie pour étudier un problème unifié. Il suffit pour cela de considérer que pour la classe A, l'échelle d'analyse est un paramètre muet : les observations $X_a(t)$ ne sont que des variables aléatoires i.i.d., ce quels que soient a et t . On peut alors définir un estimateur des moments à l'échelle d'analyse a par

$$S(a, n, q) = \frac{1}{n} \sum_{k=1}^n X_a(t_k)^q. \quad (\text{II.75})$$

Pour la classe B, on retrouve bien la définition de la fonction de structure tandis que pour la classe A l'équation se simplifie en une somme classique de variables aléatoires i.i.d.. Il est donc possible au prix de cette légère complication d'étudier simultanément le comportement de l'estimateur des moments pour les classes A et B. Tel sera l'objet du chapitre 4.

Chapitre 4

Ordre critique de l'estimateur des moments

Notre but dans ce chapitre est de dévoiler un changement de comportement pour l'estimateur des moments $S(a, n, q)$, pour les deux classes A et B de variables aléatoires, lorsqu'on couple l'ordre des moments q avec la taille n ou l'échelle du système a . Dans le cas i.i.d., cela revient essentiellement à coupler q et n et donc étudier le comportement de S le long d'une courbe dans le plan (q, n) . Le cas multifractal est plus complexe et fait intervenir principalement le couplage entre n et a . Néanmoins, dans les deux cas, nous définirons un ordre critique q^* au-delà duquel l'estimateur des moments $S(a, n, q)$ n'estime plus correctement la moyenne d'ensemble $\mathbb{E}[X(a, t)^q]$, et nous étudierons comment q^* se comporte asymptotiquement lorsque $n \rightarrow +\infty$.

Notre analyse repose sur une compétition entre deux effets : un effet de concentration de la contribution des moments s'oppose à un effet de taille finie associé à une modélisation par troncature des moments. Une telle analyse s'inspire fortement des analyses similaires conduites en physique statistique, notamment dans le cadre de l'étude du REM (*Random Energy Model*) [40], et nous retrouverons des résultats fort semblables à ceux d'une transition vitreuse.

1 Effet de concentration : Contribution dominante aux moments

Si on réécrit les moments X en fonction des variables $Y = \ln X$ et de la fonction Ψ définie à l'Eq. (II.18), on obtient

$$\mathbb{E}[X^q] = \int_{-\infty}^{+\infty} e^{qy - \Psi(y) + \ln \Psi'(y)} dy. \quad (\text{II.76})$$

Cette expression présente de fortes similitudes avec les expressions intégrales utilisées par la méthode du col [36].

1.1 Classe A-II

Dans le cas i.i.d., Ψ est essentiellement une fonction de puissance pour la classe A-II. On peut donc adapter la démonstration utilisée pour la méthode du col pour évaluer l'intégrale précédente (la preuve technique est reportée en annexe D) :

$$\mathbb{E}[X^q] \underset{q \rightarrow +\infty}{\sim} \sqrt{\frac{2\pi}{\Psi''(y_m) + (\ln \Psi')''(y_m)}} \exp(qy_m - \Psi(y_m) + \ln \Psi'(y_m)), \quad (\text{II.77})$$

où y_m est déterminée par

$$q - \Psi'(y_m) + \frac{\Psi''}{\Psi'}(y_m) = 0. \quad (\text{II.78})$$

Les conditions (II.20) et (II.21) –imposées dans la sous-section 1.1 – impliquent que Ψ' est croissante pour q suffisamment grand et que

$$\lim_{y \rightarrow +\infty} \frac{\Psi''(y)}{\Psi'(y)} = 0. \quad (\text{II.79})$$

En combinant ces résultats avec l' Eq. (II.78), nous avons (en sélectionnant éventuellement la solution minimale)

$$\Psi'(y_m) \approx q. \quad (\text{II.80})$$

Qualitativement, y_m dénote la valeur de Y d'où provient l'essentiel de la contribution au moment d'ordre q , $\mathbb{E}[X^q]$. De plus, lorsque q augmente cette contribution devient relativement de plus en plus importante, on peut donc parler d'un effet de concentration des moments : Pour q suffisamment grand, seul les observations dans le voisinage de y_m contribue de manière significative au moment (théorique) d'ordre q . L'Eq. (II.78) définit implicitement une fonction $y_m(q)$, qui dépend des détails de la distribution de Y . On peut cependant remarquer que puisque $\Psi(y) = y^\rho L(y)$ avec L lentement variable et normalisée, on peut en déduire que $\Psi'(y) = y^{\rho-1} L_2(y)$. On peut donc résoudre l'équation pour quantifier l'évolution de y_m en fonction de q . Puisque $\rho > 1$, il existe une fonction lentement variable L_3 telle que

$$y_m(q) = q^{\frac{1}{\rho-1}} L_3(q). \quad (\text{II.81})$$

Ce point y_m se comporte donc essentiellement, pour $q \rightarrow +\infty$, comme une fonction de puissance d'exposant $1/(\rho - 1)$.

Type I

Dans les cas de lois de type I, il est plus simple d'utiliser les variables X et $\Psi_X = -\ln \mathbb{P}(X > x)$. L'équation (II.76) devient alors

$$\mathbb{E}[X^q] = \int_0^{+\infty} e^{q \ln x - \Psi_X(x) + \ln \Psi'_X(x)} dx. \quad (\text{II.82})$$

La fonction Ψ_X vérifie ici $\Psi_X(x) = x^\rho L(x)$. Il est donc possible d'appliquer une approximation de type méthode du col pour calculer l'intégrale précédente (cf. l'annexe D pour les détails techniques). Le point de concentration est alors défini par

$$x_m \left(\Psi' + \frac{\Psi''}{\Psi'} \right) (x_m) = q. \quad (\text{II.83})$$

Utilisant le fait que $\Psi''(x) \sim \Psi'(x)/x$, l'expression précédente se simplifie en

$$x_m \Psi'(x_m) \approx q. \quad (\text{II.84})$$

On peut donc expliciter la dépendance de x_m en q sous la forme

$$x_m = q^{\frac{1}{\rho}} L(q). \quad (\text{II.85})$$

Si nous revenons temporairement à $y = \ln x$ pour faire la comparaison avec les lois de type II, nous obtenons

$$y_m = \frac{\ln q}{\rho} + \ln L(q). \quad (\text{II.86})$$

Le point de concentration se déplace donc plus lentement que dans le cas des lois de type II, à cause de la plus grande vitesse de décroissance des queues des distributions.

1.2 Classe B

On peut également appliquer une méthode du col dans le cas d'un processus X multifractal. Dans ce cas, l'expression sous forme de fonction de grande déviation de p_h implique que

$$\Psi(y) - \ln \Psi'(y) = |\ln a| \psi(h). \quad (\text{II.87})$$

Il est donc naturel de faire le changement de variable $Y = -|\ln a|H$ dans l'équation (II.76), ce qui donne

$$\mathbb{E} [X(a, t)^q] \approx \int_{-\infty}^{+\infty} e^{-|\ln a|[qh + \psi(h)]} dh. \quad (\text{II.88})$$

si on ne tient pas compte du terme non exponentiel.

Dans la limite $a \rightarrow 0$, une évaluation par la méthode du point col montre que la contribution dominante à l'intégrale est localisée en $h = h_m$ défini par

$$\psi'(h_m) = -q. \quad (\text{II.89})$$

Au signe près et en remplaçant h par y , on retrouve la même expression que dans l'Eq. (II.80). Le moment $\mathbb{E} [X(a, t)^q]$ s'écrit

$$\mathbb{E} [X(a, t)^q] \approx \sqrt{\frac{2\pi}{|\ln a| |\psi''(h_m)|}} e^{-|\ln a| [qh_m + \psi(h_m)]}. \quad (\text{II.90})$$

Pour $h > h_c$, $\psi(h) = -\lambda^*(h)$ et h_m vérifie :

$$(\lambda^*)'(h_m) = q. \quad (\text{II.91})$$

Les propriétés de la transformée de Legendre impliquent que :

$$qh_m - \lambda^*(h_m) = \lambda(q) \quad (\text{II.92})$$

et

$$h_m = \lambda'(q), \quad (\text{II.93})$$

ce qui définit implicitement h_m comme une fonction de q . On retrouve donc également un effet de concentration des moments dans le cas multifractal. Cependant,

sa cause diffère. Dans le cas i.i.d., il était nécessaire d'utiliser la forme particulière de la queue de la distribution. Dans le cas multifractal, nous n'avons utilisé finalement que la dépendance en a de la fonction Ψ . Cette différence est plus visible si on s'intéresse à y_m . La relation $Y_a = -|\ln a|H_a$ implique que

$$y_m(a, q) = -|\ln a|h_m(q). \quad (\text{II.94})$$

Dans le cas multifractal, nous avons donc suffisamment d'information sur le comportement de Ψ_a en fonction de a pour pouvoir décrire la dépendance en a de y_m . Mais contrairement au cas i.i.d., nous n'avons pas assez d'information sur Ψ pour décrire la dépendance en q de y_m .

2 Effet de taille finie : moments tronqués et estimateur des moments

Considérons maintenant une observation de taille finie n . Ici, puisque H et Y ne diffèrent que par une constante multiplicative, il est naturel d'unifier le cas i.i.d. et le multifractal en se concentrant sur les valeurs observées de Y $\{Y_k, k \in 1 \dots n\}$. Ces échantillons se trouvent entre le minimum empirique $m_k = \min_k \{Y_k\}$ et le maximum empirique $M_k = \max_k \{Y_k\}$. Une observation de taille finie n'explore donc qu'une région bornée du support de Y .

Quand $n \rightarrow +\infty$, les bornes de cette région s'étendent naturellement pour couvrir progressivement tout le domaine de définition de la loi de Y . Cependant, pour des n finis, il est vraisemblable que cette région bornée affecte les propriétés de l'estimateur des moments pour un q donné, si l'information nécessaire pour évaluer ce moment (qualitativement y_m ou h_m) se trouve au-delà des frontières de cette région. Cette section a pour but de détailler et quantifier le raisonnement précédent et en tirer les conséquences adéquates concernant les propriétés de l'estimateur des moments.

2.1 Région accessible à taille finie

Définissons donc en premier lieu une frontière au-delà de laquelle la probabilité d'observer un échantillon Y_k est faible. Pour quantifier cette région des valeurs typiquement disponible au sein de cet ensemble d'observations, une idée consiste à fixer un seuil $y_\tau^\dagger(a, n)$ tel que

$$\mathbb{P}(\forall k = 1, \dots, n, Y_k < y_\tau^\dagger(a, n)) = e^{-\tau}. \quad (\text{II.95})$$

La notation $\mathbb{P}(\forall k = 1, \dots, n, Y_k < y_\tau^\dagger(a, n))$ dénote la probabilité que toutes les variables aléatoires Y_1, \dots, Y_n soient plus petites que la valeur seuil $y_\tau^\dagger(a, n)$. La constante positive τ est une valeur arbitraire quantifiant la "dureté" de la frontière y_τ^\dagger . Idéalement, τ ne devrait jouer qu'un rôle négligeable dans la limite $n \rightarrow +\infty$.

Classe A-II

Si les variables aléatoires $\{Y\}$ sont indépendantes, nous avons

$$\begin{aligned} \mathbb{P}(\forall k = 1, \dots, n, Y_k < y_\tau^\dagger) &= e^{-\tau} \\ \Leftrightarrow F(y_\tau^\dagger)^n &= e^{-\tau} \end{aligned}$$

où $F(y) = \mathbb{P}(Y < y)$ est la fonction cumulative. En passant au logarithme et utilisant $\ln F = -\overline{F} + o(\overline{F})$, l'expression précédente devient

$$\begin{aligned} -n\overline{F}(y_\tau^\dagger) + o(\overline{F}(y_\tau^\dagger)) &= -\tau \\ \Leftrightarrow \overline{F}(y_\tau^\dagger) &= \frac{\tau}{n} + o\left(\frac{\tau}{n}\right) \\ \Leftrightarrow \overline{F}(y_\tau^\dagger) &= \frac{\tau}{n} + o\left(\frac{1}{n}\right) \end{aligned}$$

L'équation (II.95) se simplifie donc pour les grands n , en :

$$\overline{F}(y_\tau^\dagger) = \frac{\tau}{n}. \quad (\text{II.96})$$

Pour les types de variables aléatoires considérées ici (cf. les définitions. (II.18) et (II.19)), il est plus lisible de prendre le logarithme de l'équation précédente pour obtenir

$$\Psi(y_\tau^\dagger(n)) = \ln n - \ln \tau. \quad (\text{II.97})$$

On peut inverser cette équation pour détailler l'évolution de y^\dagger en fonction de τ et n . Utilisant le fait que $\Psi(y) = y^\rho L(y)$, on obtient

$$y_\tau^\dagger(n) = (\ln n)^{1/\rho} \left(1 - \frac{\ln \tau}{\ln n}\right)^{1/\rho} L_2\left(\ln n \left(1 - \frac{\ln \tau}{\ln n}\right)\right). \quad (\text{II.98})$$

On peut voir que dans la limite $n \rightarrow +\infty$, à τ constant, les termes en τ de l'expression précédente s'évanouissent :

$$y_\tau^\dagger(n) \sim (\ln n)^{1/\rho} L_2(\ln n) \quad (\text{II.99})$$

Cela indique que τ joue bien un rôle négligeable et qu'il est naturel de considérer simplement $y^\dagger(n)$, défini par :

$$\Psi(y^\dagger(n)) = \ln n, \quad (\text{II.100})$$

ce qui quantifie la borne supérieure de la queue de Y qu'il est réellement possible d'observer pour une observation typique de taille finie n . Notons que $y^\dagger(n)$ ne dépend que de n et des détails de la queue de la distribution de Y .

Classe A-I

On peut appliquer le même raisonnement aux lois de type I, en considérant le maximum des variables aléatoires X plutôt que celui des variables Y . Il suffit d'utiliser dans ce cas les variables X et $\Psi_X = -\ln \mathbb{P}(X < x)$. On obtient alors

$$\Psi_X(x^\dagger) = \ln n. \quad (\text{II.101})$$

Cela implique que

$$x^\dagger = (\ln n)^{\frac{1}{\rho}} L(\ln n). \quad (\text{II.102})$$

On retrouve encore une fois une croissance logarithmiquement plus lente de y^\dagger pour les lois de type I comparée aux lois de type II.

Classe B

Pour un processus multifractal comme la cascade de Poisson composée les variables $\{Y_a\}$ sont fortement dépendantes. Cependant nous pouvons postuler qu'il existe un nombre effectif d'échantillons indépendants $n_{\text{eff}}(a, n) < n$ tel que

$$\mathbb{P}(y < y_\tau^\dagger) = \frac{\tau}{n_{\text{eff}}(a, n)}, \quad (\text{II.103})$$

ou de façon équivalente,

$$\Psi(y_\tau^\dagger) = \ln n_{\text{eff}}(a, n) - \ln \tau. \quad (\text{II.104})$$

En utilisant cette notion de nombre effectif d'échantillons indépendants, il est donc possible de réutiliser les résultats du cas i.i.d. en remplaçant n par n_{eff} . Ce nombre effectif d'échantillons dépend en principe de n et a . Cependant, pour une échelle a donnée, la quantité d'information est limitée, il existe donc un nombre effectif maximum d'échantillons $n_{\text{eff}}^{\text{max}}(a) = \sup_n n_{\text{eff}}(n, a)$. Par la suite, nous supposons que n est choisi de telle sorte que $n_{\text{eff}}(n, a) = n_{\text{eff}}^{\text{max}}(a)$: n est alors suffisamment grand pour exploiter toute l'information à l'échelle a .

Pour expliciter les calculs, il est plus pratique de revenir aux variables H_a . Nous allons donc chercher une expression explicite de $h^\dagger = -y^\dagger/|\ln a|$.

En utilisant la forme de grande déviation de l'Eq. (II.62) dans l'Eq. (II.104), on obtient :

$$\ln \left(\frac{n_{\text{eff}}(a)}{\tau} \right) = - \ln \left(\int_{-\infty}^{h^\dagger(a)} e^{-|\ln a| \psi(h)} dh \right). \quad (\text{II.105})$$

Étant donné que ψ est une fonction décroissante de h sur cet intervalle, un argument de type méthode du col équivaut à évaluer l'intégrale sur sa borne (supérieure) :

$$\ln \left(\frac{n_{\text{eff}}(a)}{\tau} \right) \simeq |\ln a| \psi(h^\dagger(a)). \quad (\text{II.106})$$

Le seuil $h^\dagger(a)$ est alors déterminé par l'équation implicite

$$\psi(h^\dagger(a)) = \frac{1}{|\ln a|} (\ln n_{\text{eff}}(a) - \ln \tau). \quad (\text{II.107})$$

Le paramètre arbitraire τ disparaît bien dans la limite $a \rightarrow 0$ pourvu que $n_{\text{eff}}(a)$ croisse bien lorsque a décroît. L'Eq. (II.107) est le pendant de l'équation (II.100). On retrouve le même type de comportement, si ce n'est l'apparition du paramètre a dans l'Eq. (II.107).

2.2 Moments tronqués

L'analyse à taille finie précédente dit essentiellement que, pour une observation donnée de taille finie n , le domaine d'existence de Y , \mathbb{R}_+ doit être séparé en deux sous-régions distinctes :

- $(-\infty, y^\dagger)$, la région accessible qui contient la plupart des échantillons observés,¹

1. Pour des moments positifs, les valeurs positives de Y jouent un rôle dominant, on peut donc se focaliser sur la frontière supérieure y^\dagger . Cependant, une extension naturelle aux moments négatifs seraient d'introduire la contrepartie inférieure y_-^\dagger de y^\dagger qui jouerait un rôle exactement symétrique.

- $[y^\dagger, +\infty)$, une région inaccessible, au sein de laquelle il n'y a qu'une très faible probabilité d'observer un échantillon.

On peut donner un sens formel à cette idée de région inaccessible en utilisant le lemme de Borel-Cantelli. Ce lemme de Borel-Cantelli nous dit que pour une suite d'événements A_n , si

$$\sum_k \mathbb{P}(A_k) < \infty \quad (\text{II.108})$$

alors presque sûrement les événements A_n ne se produisent au plus qu'un nombre fini de fois. Ici, les événements qui nous intéressent sont l'apparition d'un point Y dans la région inaccessible. C'est-à-dire

$$A_n = \{\max\{Y_1, \dots, Y_n\} > y_\tau^\dagger\}. \quad (\text{II.109})$$

Par définition de y_τ^\dagger , nous avons

$$\mathbb{P}(A_k) = \exp(-\tau). \quad (\text{II.110})$$

Il est donc nécessaire de faire lentement diminuer τ en fonction de n si nous souhaitons que la somme des $\mathbb{P}(A_k)$ soit finie. Si on choisit une suite τ_n , telle que

$$\sum_k \mathbb{P}(A_n) = \sum_k \exp(-\tau_n) < +\infty, \quad (\text{II.111})$$

alors, presque sûrement, tous les points Y_k se trouveront dans la région accessible, sauf pour un nombre fini de valeur de n . Cela implique que pour n suffisamment grand, nous avons

$$S(n, q) \stackrel{\text{p.s.}}{=} \frac{1}{n} \sum_{k=1}^n e^{qY_k} \chi_{]-\infty, y^\dagger[}(Y_k) \quad (\text{II.112})$$

où χ_E est une fonction caractéristique d'ensemble :

$$\chi_E(y) = \begin{cases} 1 & \text{si } y \in E, \\ 0 & \text{sinon.} \end{cases} \quad (\text{II.113})$$

L'équation (II.112) signifie qu'il est parfaitement légitime d'ignorer les valeurs atypiques de Y dans l'estimateur des moments. Cela nous mène naturellement à définir des moments tronqués :

$$T(a, n, q) = \int_0^{y^\dagger(a, n)} e^{qy - \Psi(y) + \ln \Psi'(y)} dy. \quad (\text{II.114})$$

Ces moments tronqués donnent une estimation heuristique des valeurs typiques des moments, estimation dont on a retiré toute contribution de termes atypiques. Il ne faut pas oublier que ces moments tronqués dépendent en principe du choix arbitraire de τ_n dans la définition de y^\dagger . Cette dépendance n'a cependant aucune conséquence dans la suite de nos raisonnements sur le moment tronqué $T(a, n, q)$. On montrera en effet (cf. Eq. (II.99)) que pour des τ_n raisonnables $y_{\tau_n}^\dagger \rightarrow y_1^\dagger$. L'équation (II.112) indique l'existence d'un lien clair entre T et S . En raffinant ce résultat², il est possible de formaliser ce lien et d'obtenir des résultats de convergence presque sure

2. Notamment en choisissant plus parcimonieusement les A_k et les τ_k associés

de $\ln S / \ln n$ vers $\ln T / \ln n$. La preuve, inspirée par [95], repose essentiellement sur un partitionnement intelligent de la région accessible et une application des inégalités de Markov et Bienaymé-Tchebitchev. La démonstration reste toutefois assez technique et nous l'avons donc reportée à l'annexe E.

En résumé, on peut montrer que pour la classe A-II si on choisit une fonction $q(n)$ qui vérifie

$$\exists \epsilon > 0, \quad q(n) \ll (\ln n)^{2-1/\rho-\epsilon} \equiv q_M(n), \quad (\text{II.115})$$

alors

$$\lim_{n \rightarrow +\infty} \frac{\ln S(n, q(n))}{|\ln n|} \stackrel{\text{p.s.}}{=} \lim_{n \rightarrow +\infty} \frac{\ln T(n, q(n))}{|\ln n|}. \quad (\text{II.116})$$

La fonction $q_M(n)$ représente une borne pour la vitesse de croissance de $q(n)$. Au-delà de cette borne, $q(n)$ croît trop vite pour assurer la convergence presque sûre. Cependant, cette borne n'est guère restrictive en pratique.³

Parallèlement pour la classe B, au prix d'une conjecture réaliste sur la structure de dépendance des $X(a, t)$, on peut également montrer que pour q constant

$$\lim_{a \rightarrow 0} \frac{\ln S(a, q)}{|\ln a|} \stackrel{\text{p.s.}}{=} \lim_{a \rightarrow 0} \frac{T(a, q)}{|\ln a|}. \quad (\text{II.117})$$

Ce résultat signifie essentiellement qu'asymptotiquement le logarithme de l'estimateur des moments converge vers le logarithme du moment tronqué, à une condition de normalisation près. Il s'agit d'un résultat assez peu surprenant mais suffisamment robuste pour nous procurer des informations importantes sur le comportement de S en fonction de $q(n)$. On peut d'ailleurs remarquer que pour q constant, la loi des grands nombres implique que l'équation (II.116) devient très sobrement

$$\lim_{n \rightarrow +\infty} \frac{\ln S(n, q = \text{const})}{|\ln n|} \stackrel{\text{p.s.}}{=} 0. \quad (\text{II.118})$$

L'équation (II.116) n'est donc intéressante, pour des lois de type II, que lorsque q et n sont couplés.

3 Transition de phase et moment critique

3.1 Frontière critique

La combinaison de l'évaluation par contribution dominante aux moments (cf. Section 1) et des troncatures dues à la taille finie des observations (cf. Section 2) donne naissance à deux situations fort distinctes en fonction de la position relative de $y_m(q)$ (respectivement $h_m(a, q)$) et $y^\dagger(n)$ (respectivement $h^\dagger(a, n)$).

Si $y^\dagger(a, n) > y_m(a, q)$, le point de concentration des moments y_m appartient à la région accessible avec n observations. Les moments tronqués T , et par conséquent l'estimateur des moments S , représentent correctement la moyenne d'ensemble

3. Elle permet en effet une croissance proche de $(\ln n)^2$ qui est déjà plus rapide que la vitesse de croissance de l'ordre critique pour les lois de type I.

$\mathbb{E}[X^q]$, qui peut donc être évaluée (en utilisant des méthodes du col légèrement modifiées si besoin) par

$$\begin{cases} \text{Classe A-II : } & \frac{\ln T(n,q)}{\ln n} \underset{n \rightarrow +\infty}{\sim} \frac{q(n)y_m(q(n)) - \Psi(y_m(q(n)))}{\ln n}, \\ \text{Classe B : } & \frac{\ln T(a,q)}{|\ln a|} \underset{a \rightarrow 0}{\sim} -(qh_m(q) + \psi(h_m(q))). \end{cases} \quad (\text{II.119})$$

ce qui peut être comparé avec l'Eq. (II.77) pour vérifier que dans ce cas $\ln T \approx \ln \mathbb{E}[X^q]$.

Au contraire, si $y^\dagger(a, n) < y_m(a, q)$, y_m appartient à la région inaccessible avec n observations. On ne dispose donc pas des informations nécessaires pour évaluer correctement le moment théorique à partir des observations. De plus, la contribution dominante au moment tronqué n'est plus localisée en $y_m(a, q)$ mais sur la borne $y^\dagger(a, n)$ du domaine accessible. Dans ce cas, $\ln T(a, n, q)$ devient (en usant encore une fois de la méthode du col pour $a \rightarrow 0$ ou $n \rightarrow +\infty$) :

$$\begin{cases} \text{Classe A-II : } & \frac{\ln T(n,q)}{\ln n} \sim \frac{q(n)y^\dagger(n) - \Psi(y^\dagger(n)) + \ln \Psi'(y^\dagger(n))}{\ln n}, \\ \text{Classe B : } & \frac{\ln T(a,q)}{|\ln a|} \underset{a \rightarrow 0}{\sim} -(qh^\dagger(a, n) + \psi(h^\dagger(a, n))). \end{cases} \quad (\text{II.120})$$

On observe donc que, pour des a suffisamment petits ou des n suffisamment grands, S subit une "transition de phase" lorsqu'on fait varier q . Cette transition a lieu à l'ordre critique $q^*(a, n)$, défini par

$$y^\dagger(a, n) = y_m(a, q). \quad (\text{II.121})$$

Cette égalité définit implicitement une frontière divisant le plan (n, q) (ou le plan (a, q)) en deux. Ces deux régions distinctes correspondent à un comportement différent des moments tronqués (dans le cadre théorique du REM, il s'agit d'une transition de phase entre une phase vitreuse et une phase "liquide"). Il faut également noter que l'Eq. (II.121) peut aussi être interprétée comme une définition du nombre minimal d'échantillons indépendants $n_a^*(q)$ nécessaire pour évaluer correctement le moment d'ordre q . De façon intéressante, l'Eq. (II.120) révèle un comportement linéaire en q de $\ln S_n(a, q)$ lorsque $q \geq q_a^*$, expliquant donc l'effet de linéarisation reporté dans [89, 96, 74, 6].

Les résultats précédents nous ont permis de définir un ordre critique q^* à travers un mécanisme commun au cas i.i.d. et multifractal. Le comportement de q^* en fonction des paramètres n ou a dépend cependant fortement du type de processus, il est donc plus pratique d'étudier séparément le comportement de q^* pour les classes A et B.

3.2 Ordre critique pour la classe A-II

Commençons donc par le cas i.i.d.. La combinaison de l'équation (II.121) avec les équations (II.78) et (II.100) mène à :

$$\frac{q^*}{\ln n} = \frac{1}{y^\dagger} \frac{y^\dagger \Psi'(y^\dagger)}{\Psi(y^\dagger)} - \frac{\Psi''(y^\dagger)}{\Psi(y^\dagger) \Psi'(y^\dagger)}. \quad (\text{II.122})$$

Sous les conditions décrites en (II.20) et (II.21), on peut montrer que le deuxième terme du membre de droite de l'équation plus haut est négligeable par rapport au

premier, en utilisant notamment que

$$\psi''(x)/\psi'(x) \underset{x \rightarrow +\infty}{\sim} \frac{\rho - 1}{x}. \quad (\text{II.123})$$

De plus, dans le membre de droite de l'équation (II.122), on peut reconnaître l'exposant de queue local ρ_l :

$$\frac{y^\dagger \Psi'(y^\dagger)}{\Psi(y^\dagger)} = \rho_l(y^\dagger(n)). \quad (\text{II.124})$$

Si on définit de plus

$$\theta(n) = \frac{\ln n}{y^\dagger(n)}, \quad (\text{II.125})$$

$q^*(n)$ peut être approché par :

$$q^*(n) \simeq \rho_l(y^\dagger(n))\theta(n), \quad (\text{II.126})$$

ce qui se révélera particulièrement utile pour l'estimation pratique de $q^*(n)$.

Un corollaire intéressant est que le taux de croissance de q^* en fonction de n et ρ peut être évalué. L'équation (II.19) implique l'existence d'une fonction variant lentement L_2 , directement reliée à L (cf. [19]), telle que :

$$\Psi^{-1}(\ln n) = L_2(\ln n)(\ln n)^{1/\rho}. \quad (\text{II.127})$$

Si on injecte ce résultat dans l'Eq. (II.126), on obtient

$$q^*(n) = \frac{\rho}{L_2(\ln n)}(\ln n)^{1-\frac{1}{\rho}}. \quad (\text{II.128})$$

Lorsque $\rho \rightarrow +\infty$, cette expression rappelle fortement l'Eq. (II.4), évoquée dans l'introduction de cette partie et valide pour la classe régulière de variables aléatoires. Cependant, lorsque ρ diminue et s'approche de 1, cet ordre critique s'accroît bien plus lentement en fonction de n . Par exemple, dans le cas log-normal, $\rho = 2$ implique que $q^*(n) \propto \sqrt{\ln n}$ et on peut montrer que $L_2(x)$ est asymptotiquement égale à

$$L_2(x) \sim \frac{\sqrt{2} \operatorname{erf}^{-1}(1 - 2e^{-x})}{\sqrt{x}}. \quad (\text{II.129})$$

Il est intéressant de noter qu'il est possible d'utiliser l'équation (II.126) pour calculer exactement $q^*(n)$.

Ordre critique pour les lois de type I

On peut appliquer la méthode précédente aux lois de type I, afin d'obtenir une expression de leur ordre critique des moments. En faisant le rapport entre les équations (II.84) et (II.101), on obtient

$$\frac{q^*}{\ln n} = \frac{x^\dagger \phi'(x^\dagger)}{\phi(x^\dagger)} \rightarrow \rho. \quad (\text{II.130})$$

Cela signifie que q^* est asymptotiquement équivalent à

$$q^* \sim \rho \ln n. \quad (\text{II.131})$$

On retrouve le même ordre de grandeur que l'équation (II.4) dérivée dans [68]. Il est intéressant de noter que le facteur correctif en $\ln \ln n$ de l'équation (II.4) n'apparaît pas ici. Ce n'est pas forcément surprenant étant donné que nous ne nous intéressons pas aux mêmes conditions de convergence : nous avons mis l'accent sur la convergence de $\ln S / \ln n$ vers $\ln \mathbb{E}[X^q] / \ln n$, alors que l'article [68] s'intéresse plus à la convergence de $S(n, q)$. Il est donc logique d'obtenir des contraintes moins fortes dans notre cas.

3.3 Ordre critique pour la classe B

Nous nous enquérons à présent du comportement asymptotique de q^* dans le cas multifractal dans la limite $n_{\text{eff}}(a) \rightarrow \infty$ (ou $a \rightarrow 0$). En pratique, cette limite est atteinte en considérant successivement des résolutions δt de plus en plus petites. Nous avons vu dans l'équation (II.58) que les variables aléatoires $X(a, t)$ sont corrélées sur un temps caractéristique a , $h_a(t)$ doit donc être sujet à la même forme de corrélation. Une estimation naturelle de $n_{\text{eff}}(a)$ est donc

$$n_{\text{eff}}(a) = \frac{L}{a}, \quad (\text{II.132})$$

où L est la longueur totale du signal.

Tout d'abord, nous observons que les équations (II.107) et (II.132) impliquent

$$\psi(h^\dagger(a, n)) \rightarrow 1 \quad (\text{II.133})$$

lorsque $a \rightarrow 0$. Par conséquent, $h^\dagger(a, n)$ converge dans la limite $a \rightarrow 0$ vers une limite finie h_0^\dagger , indépendante de n , et uniquement déterminée par $\psi(h_0^\dagger) = 1$ ce qui, dans notre contexte multifractal, peut être réécrit comme :

$$D(h_0^\dagger) = 0. \quad (\text{II.134})$$

Ce résultat est particulièrement intéressant du point de vue de l'analyse multifractale, et nous discuterons de son interprétation dans la section 4.2.

Les équations (II.92) et (II.93) reposent implicitement sur l'hypothèse que $h_0^\dagger > h_c$. Discutons donc brièvement de cette hypothèse. En utilisant l'Eq. (II.67) et le fait que ψ est une fonction décroissante de h_c à h_0^\dagger , on peut s'apercevoir que la condition $h_0^\dagger > h_c$ équivaut à

$$1 < 1 - h_c q_c. \quad (\text{II.135})$$

Par suite, la propriété $\lambda'(q_c) = q_c < 0$, condition qui est vérifiée pour tout cas intéressant, implique $h_0^\dagger > h_c$. Cela confirme donc la validité des Eqs. (II.92) et (II.93). La combinaison de ces équations avec les Eqs. (II.106) et (II.121) mène à la relation :

$$\frac{\ln(n_{\text{eff}}(a)/\tau)}{\ln a} = q^*(a)\lambda'(q^*(a)) - \lambda(q^*(a)). \quad (\text{II.136})$$

Avec $n_{\text{eff}}(a) = L/a$, nous obtenons

$$\frac{\ln(L/\tau)}{\ln a} - 1 = q^*(a)\lambda'(q^*(a)) - \lambda(q^*(a)). \quad (\text{II.137})$$

Dans la limite $a \rightarrow 0$, l'Eq. (II.137) définit une limite finie q^* de l'ordre critique :

$$0 = 1 + q^*\lambda'(q^*) - \lambda(q^*). \quad (\text{II.138})$$

On peut remarquer que le coefficient numérique 1 correspond à la limite du rapport $\ln n_{\text{eff}} / |\ln a|$. Si le nombre effectif d'échantillons variait en $\ln n_{\text{eff}} \approx -\alpha \ln a$, on aurait

$$0 = \alpha + q^* \lambda'(q^*) - \lambda(q^*), \quad (\text{II.139})$$

ce qui correspond au résultat obtenu par [11] en considérant un processus multifractal dans lequel le nombre d'échelles intégrales considérées dépend de l'échelle a d'analyse.

La comparaison des Eqs. (II.134) et (II.138) montre de plus immédiatement que :

$$h_0^\dagger = \lambda'(q^*). \quad (\text{II.140})$$

L'ordre critique q^* est donc indépendant de a pour a suffisamment petit. Il est important de noter que $q^* \neq q_c$. En utilisant les Eqs. (II.138) et (II.51), on peut montrer aisément que $q^* < q_c$. Par conséquent, l'existence d'un ordre critique jusqu'auquel $S_n(a, q)$ évalue précisément la moyenne d'ensemble n'est pas relié à la finitude des moments de $|X(a, t)|$ mais intervient pour des valeurs de q bien inférieures à q_c .

Pour des processus multifractals tels que le mouvement de Poisson composé, on vient de montrer que l'estimateur des moments (ou fonction de structure) $S(a, q)$ ne converge pas vers la moyenne d'ensemble $\mathbb{E}[X(a, t)^q]$ ce quel que soit la taille d'analyse a , si $q^* < q < q_c$.

L'ordre critique q^* et l'exposant de Hölder critique h_0^\dagger sont donc indépendants du nombre d'échantillons disponibles n . Par conséquent, accroître n (à travers une diminution de la période d'échantillonnage δt) ne permet pas d'améliorer l'estimation de $S(n, q > q^*)$. De plus, l'Eq. (II.137) montre qu'en pratique l'ordre critique effectif ne varie que peu avec n ou a . Il est à noter que pour certaines formes spécifiques de $\lambda(q)$, il est possible que la solution de l'Eq. (II.138) soit $q^* = +\infty$, ce qui peut être compris soit comme une absence d'un effet de linéarisation dans ces cas ou (interprétation qui a notre préférence) comme un rejet à l'infini de l'effet de linéarisation, effet général, pour ces cas particuliers.

4 Comportement asymptotique de l'estimateur des moments et lien avec le REM

4.1 Classe A-II

La convergence presque sûre établie dans l'Eq. (II.116), démontrée quelles que soient les valeurs relatives de $y_m(q)$ et $y^\dagger(n)$, indique que la transition de phase du modèle REM se transfère également à l'estimateur des moments. Si on réunit les résultats précédents, notamment l'Eq. (II.119), cela signifie que dans la limite $q \rightarrow +\infty$, $n \rightarrow +\infty$, pour une borne $q_M(n)$ vérifiant la condition (II.115), on a le résultat suivant :

$$\frac{\ln S(n, q)}{\ln n} \underset{\text{p.s.}}{\sim} \begin{cases} \frac{q y_m(q) - \Psi(y_m(q))}{\ln n} & q \leq q^*(n), \\ \left(\rho_l(y^\dagger(n))^{\frac{q}{q^*}} - 1 \right) & q^*(n) \leq q \leq q_M(n). \end{cases} \quad (\text{II.141})$$

Si on s'intéresse au comportement de S par rapport à $\mathbb{E}[X^q]$, on peut réécrire le résultat précédent sous la forme approchée

$$\ln S(n, q) \approx \begin{cases} \ln \mathbb{E}[X^q] & q \leq q^*(n), \\ \ln n \left(\rho_l (y^\dagger(n))^{\frac{q}{q^*}} - 1 \right) & q^*(n) \leq q \leq q_M(n). \end{cases} \quad (\text{II.142})$$

Ainsi, pour $q \leq q^*(n)$, l'estimateur des moments estime correctement la moyenne d'ensemble $\mathbb{E}[X^q]$, tandis que $\ln S(n, q)$ se comporte linéairement en q lorsque $q \geq q^*(n)$. On observe donc une rupture de la loi des grands nombres, que l'on peut caractériser à travers le comportement du logarithme de l'estimateur des moments normalisé par $\ln n$, qui n'est autre que l'analogue formel de l'énergie libre du REM. L'effet de linéarisation est donc l'analogue d'une transition vitreuse.

On peut obtenir un résultat plus précis si on suit une courbe $q(n) = tq^*(n)$, en utilisant

$$\begin{aligned} \ln n &= \Psi(y^\dagger), \\ q &= \Psi'(y_m) = \lambda \Psi'(y^\dagger). \end{aligned} \quad (\text{II.143})$$

La seconde ligne de l'équation (II.143) implique que pour $t < 1$

$$\frac{y_m}{y^\dagger} \sim t^{\frac{1}{\rho-1}}. \quad (\text{II.144})$$

On peut alors transformer la première ligne de l'expression (II.142)

$$\begin{aligned} \frac{qy_m - \Psi(y_m)}{\ln n} &= \frac{y_m \Psi'(y_m) - \Psi(y_m)}{\Psi(y^\dagger)} \\ &\sim \left(\frac{y_m}{y^\dagger} \right)^\rho \frac{y^\dagger \Psi'(y^\dagger) - \Psi(y^\dagger)}{\Psi(y^\dagger)} \\ &\sim t^{\frac{\rho}{\rho-1}} \left(\frac{y^\dagger \Psi'(y^\dagger)}{\Psi(y^\dagger)} - 1 \right) \\ &\sim t^{\frac{\rho}{\rho-1}} (\rho_l - 1), \end{aligned} \quad (\text{II.145})$$

ce qui donne donc de manière plus synthétique

$$\lim_{n \rightarrow +\infty} \frac{\ln S(n, tq^*(n))}{\ln n} = \begin{cases} (\rho_l - 1)t^{\frac{\rho}{\rho-1}} & t \leq 1, \\ \rho_l t - 1 & t > 1. \end{cases} \quad (\text{II.146})$$

Cette transition de phase est illustrée par la figure 4.1 obtenue par simulations numériques. Nous avons utilisé un grand nombre r de réalisations indépendantes de séries de variables aléatoires i.i.d. log-normales X pour calculer une valeur moyenne de $\ln S(n, q)$, pour différentes valeurs de n et q . Cette moyenne est ensuite comparée à $\ln \mathbb{E}[X^q]$. Cette figure valide clairement le fait que pour $q < q^*(n)$, $\ln S(n, q)$ correspond à $\ln \mathbb{E}[X^q]$, alors que $\ln S(n, q)$ devient significativement plus petit que $\ln \mathbb{E}[X^q]$ pour $q > q^*(n)$ et se comporte linéairement en q . Les résultats obtenus pour différents n peuvent être superposés en traçant $\ln S(n, q)$ et $\ln \mathbb{E}[X^q]$ comme des fonctions de $q/q^*(n)$ plutôt qu'en tant que fonction de q , où $q^*(n)$ est obtenue en résolvant numériquement l'équation (II.121) pour chaque n .

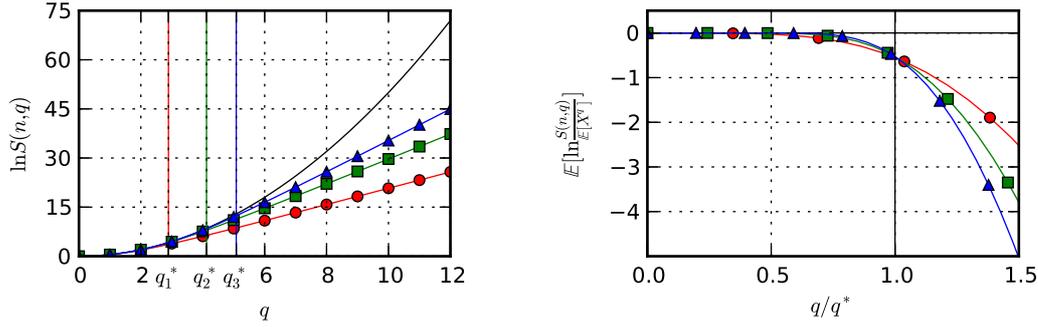


FIGURE 4.1 – **Prédictions quantitatives de l'effet de linéarisation.** À gauche, $\mathbb{E}[\ln S(n, q)]$ en fonction de q , avec les valeurs prédites de q_c , pour trois différentes valeurs de n : $n = 10^2$ (cercle rouge), $n = 10^3$ (carré vert), $n = 10^6$ (triangle bleu); la ligne noire pleine montre l'espérance théorique $\ln \mathbb{E}[X^q]$. À droite, $\mathbb{E}[\ln(S(n, q)/\mathbb{E}[X^q])]$ en fonction q/q^* , pour les 3 mêmes valeurs de n . $\mathbb{E}[\ln S(n, q)]$ est obtenue par des simulations de Monte-Carlo, pour des distributions log-normales. Ces graphiques valident clairement la signification pratique du moment critique $q^*(n)$.

4.2 Classe B

Dans le cas multifractal, si nous assimilons $S(a, q)$ avec $T(a, q)$ et combinons les équations (II.119) et (II.120), nous obtenons une fonction de structure empirique $S(a, n, q)$ qui se comporte typiquement comme une loi de puissance de l'échelle d'analyse a lorsque $a \rightarrow 0$, plus précisément $S(a, n, q) \sim S_0(q) a^{\zeta_e(q)}$, avec $\zeta_e(q)$ un exposant d'échelle empirique. Plus formellement, nous pouvons définir l'exposant d'échelle empirique $\zeta_e(q)$ comme la variable aléatoire :

$$\zeta_e(q) = \lim_{a \rightarrow 0} \frac{S(a, n, q)}{\ln a}. \quad (\text{II.147})$$

L'Eq. (II.117) implique alors que $\zeta_e(q)$ est presque sûrement égal à sa valeur moyenne $\mathbb{E}[\zeta_e(q)] \equiv \zeta(q)$:

$$\forall q \in]-1, q_c[, \quad \zeta_e(q) \stackrel{\text{p.s.}}{=} \zeta(q). \quad (\text{II.148})$$

La valeur de $\zeta(q)$ peut être exprimée, faisant usage des équations (II.119) et (II.120), par

$$\zeta(q) = \begin{cases} \lambda(q), & -1 < q \leq q^*, \\ 1 + q\lambda'(q^*), & q > q^*. \end{cases} \quad (\text{II.149})$$

Il s'agit ni plus ni moins que du transfert de l'équation (II.142) au cas de la classe B. On retrouve donc la "transition vitreuse" que nous observons dans le cas i.i.d..

On peut justifier cette analogie entre transition vitreuse et effet de linéarisation en étendant l'analogie avec le REM. Étant donné la correspondance entre, d'un côté, la fonction de partition Z du REM et la fonction de structure de l'analyse multifractale, et entre, d'un autre côté, la température inverse β et l'ordre des moments q , on peut établir que l'analogue de l'entropie par degré de liberté (qui est nulle dans la phase basse température du REM $\beta > \beta_g \equiv T_g^{-1}$) est la quantité $q\zeta'(q) - \zeta(q) + 1$, qui est effectivement nulle dans le régime linéaire c'est-à-dire pour $q > q^*$. La Table 4.1

MF	REM
$S(a, n, q) \approx \frac{1}{n_{\text{eff}}} \sum_{j=1}^{n_{\text{eff}}} X(a, t_{k_j})^q$	$Z/2^N = \frac{1}{2^N} \sum_{j=1}^{2^N} e^{-\beta E_j}$
$\ln n_{\text{eff}} \sim -\ln a$	$\ln n = N \ln 2$
q	$\beta = T^{-1}$
$\ln X(a, t_{k_j})$	$-E_j$
$h_a(t_{k_j})$	$E_j/N \ln 2$
$-q^{-1} \ln(n_{\text{eff}} S(a, q))$	$F = -T \ln Z$
$q\zeta'(q) - \zeta(q) + 1$	$S/(N \ln 2)$
q^*	β_g
h_0^\dagger	$\epsilon^\dagger / \ln 2$
h_m	$\epsilon_m / \ln 2$

TABLE 4.1 – **MF vs. REM.** Correspondance entre les quantités définies en analyse multifractale (MF) et dans le REM, valide dans la limite des petits a et des grands N . Afin d’interpréter $S(a, q)$ comme une somme de variables (presque) indépendantes, une séquence de n_{eff} instants $\{t_{k_j}, j = 1, \dots, n_{\text{eff}}\}$ est extraite du jeu complet $\{t_k, k = 1, \dots, n\}$.

ébauche la correspondance entre les quantités apparaissant dans le REM et en analyse multifractale (voir l’annexe B pour la définition des notations utilisées dans le REM).

Les propriétés plus haut, qui apparaissent comme une conséquence de l’Eq. (II.134), peuvent être interprétées en suivant un chemin parallèle aux arguments du REM (voir l’annexe B). Dans un échantillon donné, le nombre de points indépendants possédant un exposant de singularité h croît comme $n_{\text{eff}}(a) e^{-\psi(h)|\ln a|}$, lorsque $a \rightarrow 0$. Si $n_{\text{eff}}(a) = L/a$, ce nombre croît donc comme $e^{(1-\psi(h))|\ln a|}$. Cela signifie que pour une réalisation donnée, il y aura un grand nombre de points avec un exposant de singularité h lorsque $1 - \psi(h) > 0$ (ce qui correspond à $D(h) > 0$); au contraire, il n’y aura aucun point de singularité h dans une réalisation typique si $1 - \psi(h) < 0$, et ce quelle que soit la durée d’observation L et l’échelle d’analyse a . La valeur h_0^\dagger , telle $D(h_0^\dagger) = 0$ (cf. l’Eq. (II.134)), peut donc être interprétée de manière simple dans ce cadre théorique.

Il est intéressant de remarquer que les interprétations reportées plus haut en termes de h comparé à h_0^\dagger et de l’entropie avait déjà été envisagée par B. Mandelbrot dans une série d’articles déterminants dédiée aux aspects pratiques de l’analyse multifractale, les plus marquants parmi ceux-ci étant e.g. [83, 85].

L’analyse théorique de l’effet de linéarisation dans la présente contribution à partir d’arguments inspirés du REM en physique statistique est similaire (et conforte donc pleinement) la conjecture formulée dans [74, 6], qui provient de l’interprétation en termes de valeurs extrêmes et de régularité locale d’observations empiriques obtenues de l’application du formalisme multifractal à des simulations numériques du mouvement de Poisson composé et d’autres processus multifractals apparentés.

Des analyses théoriques complémentaires [96] ainsi que des analyses numériques (voir section 5 plus bas) menées pour des processus multifractals distincts des MPC

(non détaillées ici) suggèrent que ces résultats obtenus ici pour les MPC sont valides pour une classe bien plus large de processus multifractal (cf., e.g., [74]).

5 Rapport de participation en analyse multifractale

Dans l'étude des systèmes de type REM, ou plus généralement, de la phase gelée des verres de spins, la condensation des configurations explorées au sein d'un petit sous-ensemble est classiquement mesurée grâce à un outil autant théorique que numérique appelé *rapport de participation* [86, 87]. Dans cette section, nous mettons à l'œuvre cet outil pour analyser davantage l'*effet de linéarisation* dans le contexte de l'analyse multifractale.

La définition du rapport de participation $\rho(a, q, p)$ est adaptée à partir de sa formulation classique en physique statistique pour le contexte des processus multifractals, avec $n_a = L/a$:

$$\rho(a, q, p) = \frac{\sum_{k=1}^{n_a} |X(a, ka)|^{qp}}{(\sum_{k=1}^{n_{\text{eff}}} |X(a, ka)|^q)^p}. \quad (\text{II.150})$$

En poursuivant l'analogie avec l'analyse du REM, on peut supposer que, dans la phase vitreuse $q > q^*$, le rapport de participation n'est pas *auto-moyennant*, ce qui signifie qu'il dépend explicitement de l'observation (ou réalisation) du processus $Z(t)$ et par extension de ses accroissements $X(a, t)$, ceci même dans la limite $a \rightarrow 0$. Dans cette limite, son espérance satisfait l'expression fermée suivante pour tous $p > 1$ et $q > 0$ [86, 87] :

$$\lim_{a \rightarrow 0} \mathbb{E} [\rho(a, q, p)] = \begin{cases} 0 & \text{si } q < q^*, \\ \frac{\Gamma(p-q^*/q)}{\Gamma(p)\Gamma(1-q^*/q)} & \text{sinon .} \end{cases} \quad (\text{II.151})$$

Le comportement attendu rappelé par l'Eq. (II.151) est désormais testé numériquement, dans le contexte des processus MPC, grâce à des simulations de Monte-Carlo. Le rapport $\rho(a, q, p)$, défini au-dessus par l'Eq. (II.150), est calculé pour 500 réalisations indépendantes de MPC de longueur 2^{22} , pour $q = 1, \dots, 15$, $p = 2, 4, 5$ et $a = 2^j$, avec $j = 1, \dots, 18$. La moyenne d'ensemble $\mathbb{E} [\rho(a, q, p)]$ est estimée par la moyenne $\mathbb{E} [\rho(a, q, p)]_{MC}$ de $\rho(a, q, p)$ sur ces réalisations indépendantes. La valeur attendue de $\mathbb{E} [\rho(a, q, p)]$, selon l'Eq. (II.151), et la valeur estimée $\mathbb{E} [\rho(a, q, p)]_{MC}$ sont comparées dans la Fig. 4.2. Pour $q \ll q^*$, $\mathbb{E} [\rho(a, q, p)]_{MC} \simeq \mathbb{E} [\rho(a, q, p)] \equiv 0$, pour tout $p > 1$ et $a > 0$. Pour $q \gg q^*$, $\mathbb{E} [\rho(a, q, p)]_{MC}$ s'éloigne indubitablement de 0, pour tout $p > 1$ et $a > 0$, et suit de plus une dépendance en q et p , qui correspond globalement aux valeurs attendues $\mathbb{E} [\rho(a, q, p)]$ décrites par l'Eq. (II.151). La transition des valeurs nulles aux valeurs non nulles de $\mathbb{E} [\rho(a, q, p)]_{MC}$ a lieu typiquement pour des valeurs de q proche de q^* , tel que calculé théoriquement par l'Eq. (II.138).

Cette correspondance entre la valeur attendue $\mathbb{E} [\rho(a, q, p)]$, selon l'Eq. (II.151) et $\mathbb{E} [\rho(a, q, p)]_{MC}$ n'est pas parfaite. Ce désaccord provient probablement de sources variées. D'un côté, les énergies $\{E_k, k = 1, \dots, n\}$ du REM et les $\{\ln T(a, t_k), k = 1, \dots, n_a\}$ du MPC, bien que toutes deux à queues lourdes n'ont pas exactement la même distribution. De l'autre, la dérivation des résultats théoriques de l'Eq. (II.151) repose sur une hypothèse d'indépendance exacte des $\{E_k, k = 1, \dots, n\}$, alors que

les $\{\ln T(a, k), k = 1, \dots, n_a\}$ sont encore significativement corrélés, ainsi que prédit par l'Eq. (II.56), ce qui pourrait affecter la moyenne d'ensemble asymptotique. Il est à noter que ces résultats ne sont montrés que pour une échelle arbitraire $a = 4$, simplement parce que les résultats précédents sont identiques à toutes les échelles.

Nous considérons que ces observations empiriques fournissent des résultats satisfaisant, confortant numériquement l'analyse théorique des effets de linéarisation observés dans l'analyse à partir d'arguments de type REM.

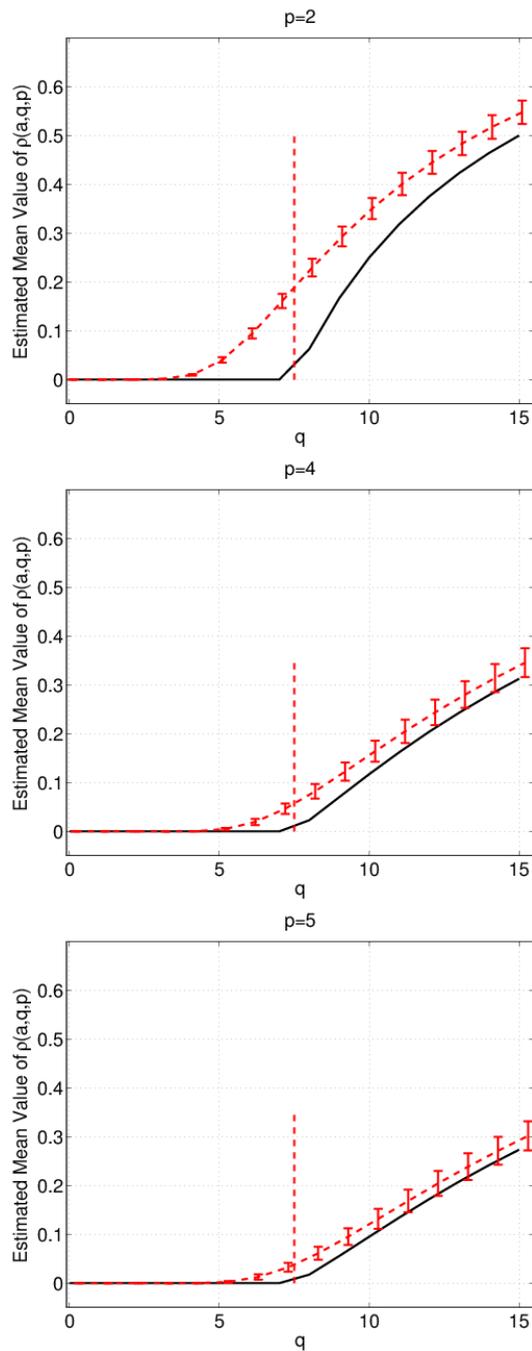


FIGURE 4.2 – **Rapport de participation.** Ligne noire pleine : $\mathbb{E}[\rho(a, q, p)]$, voir l'Eq. (II.151); ligne rouge en pointillé : Moyenne de $\rho(a, q, p)$ sur des simulations Monte-Carlo, avec l'intervalle de confiance à 95% associé; ligne rouge verticale et en pointillé : position du q^* critique calculé à partir de l'Eq. (II.138). En haut : $p = 2$, au milieu $p = 4$, en bas $p = 5$. Toutes les figures correspondent à $a = 4$.

Chapitre 5

Estimation de l'ordre critique

Le but de ce chapitre est de définir une procédure d'estimation de $q^*(n)$ –pour les variables de la classe A-II– qui peut être utilisée en pratique sur des données réelles à partir d'un nombre fini d'observations $\{X_1, \dots, X_n\}$. Nous étudierons également les performances statistiques de cet estimateur, à la fois théoriquement et numériquement. Dans la section 3, nous avons montré (cf. l'Eq. (II.126)) que $q^*(n)$ peut être naturellement scindé en deux composantes $\theta(n)$ et $\rho_l(y^\dagger(n))$. Il est donc naturel de construire une procédure d'estimation séparée pour chacune de ces deux composantes et de conduire leur étude de façon indépendante dans les sections 1 et 2 respectivement. Nous recomposerons enfin l'estimateur q^* comme un produit de ces deux estimateurs dans la section 3. Il faut noter que, dans ce chapitre, nous fixerons le paramètre arbitraire τ à la valeur $\tau = 1$.

1 Estimation de θ

1.1 Définition

Un calcul direct montre que $y^\dagger(n) = \Psi^{-1}(\ln n) = \overline{F}_Y^{-1}(\frac{1}{n})$, ce qui implique que la définition de θ (cf. l'Eq. II.125) peut être réécrite en terme du $(1 - 1/n)$ -ème quantile de la distribution de Y :

$$\theta = \frac{\ln n}{\overline{F}_Y^{-1}(\frac{1}{n})}. \quad (\text{II.152})$$

L'estimation de θ peut par conséquent être essentiellement interprétée comme le problème de l'estimation du $(1 - \frac{1}{n})$ -ème quantile de Y .

Classiquement, le p -ème quantile est estimé, pour une observation de taille n , en utilisant les statistiques d'ordre de rang $[pn]$ [18] (où $[pn]$ est l'arrondi entier de pn). Dans le présent contexte, cela revient à utiliser la plus grande valeur observée

$$M_{Y,n} = \max\{Y_1, \dots, Y_n\}, \quad (\text{II.153})$$

comme une estimation du $(1 - \frac{1}{n})$ -ième quantile.

Pour la classe de variables aléatoires Y considérée ici, $M_{Y,n}$ appartient au domaine d'attraction de la loi Gumbel [45, 59, 53]. Cela implique qu'il existe une séquence de $a_n \geq 0$ telle que $[M_{Y,n} - \overline{F}_Y^{-1}(1 - \frac{1}{n})]/a_n$ converge en distribution vers une loi Gumbel

de fonction cumulative $F(x) = \exp(-e^{-x})$. Or, l'espérance $\mathbb{E}[G]$ d'une loi Gumbel est non nulle. Plus précisément, elle est égale à la constante d'Euler-Mascheroni [102]

$$\mathbb{E}[G] = \gamma \approx 0.5772 \quad (\text{II.154})$$

Par conséquent, si on utilise directement le maximum comme estimateur du $(1 - \frac{1}{n})$ -ième quantile, on obtient

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\frac{M_{Y,n} - \overline{F}_Y^{-1} \left(\frac{1}{n} \right)}{a_n} \right] = \gamma. \quad (\text{II.155})$$

De plus, pour une loi exponentielle de paramètre $\lambda = 1$, on a $a_n = 1$. Cela indique clairement que $M_{Y,n}$ est un estimateur biaisé de $\overline{F}_Y^{-1}(\frac{1}{n})$ pour des échantillons de taille finie et qu'il n'est pas consistant dans le cas général.

Pour dépasser cette limitation, il est possible de raffiner la procédure d'estimation en utilisant non pas uniquement $M_{Y,n}$, mais les k valeurs les plus grandes au sein de l'observation $\{Y_1, \dots, Y_n\}$. Définissons donc $\{Y_{i,n}, i = 1, \dots, n\}$ la liste ordonnée (par ordre décroissant) des observations $\{Y_k, k = 1, \dots, n\}$ et soit $k_\theta(n)$ une fonction de n satisfaisant (pour des raisons qui apparaîtront plus tard) :

$$\lim_{n \rightarrow +\infty} \frac{k_\theta(n)}{n} = 0. \quad (\text{II.156})$$

Nous définissons une collection paramétrée d'estimateur de θ par (avec $k \equiv k_\theta(n)$) :

$$\hat{\theta}^{(k)} = \frac{\ln n}{\Omega_k}, \quad (\text{II.157})$$

où Ω_k consiste en une combinaison linéaire des statistiques d'ordre :

$$\Omega_k = \sum_{i=1}^k \alpha_i Y_{i,n}, \text{ avec } \sum_{i=1}^k \alpha_i \equiv 1. \quad (\text{II.158})$$

De manière triviale, le cas $k = 1$ revient à utiliser uniquement $M_{Y,n}$.

1.2 Performances : analyse théorique

Pour étudier la performance de $\hat{\theta}^{(k)}$, il est nécessaire d'analyser les propriétés statistiques de Ω_k , qui peut être réécrit en

$$\Omega_k = \overline{F}^{-1} \left(\frac{1}{n} \right) + a_n \sum_{i=0}^k \alpha_i U_i, \quad (\text{II.159})$$

où

$$U_i = \frac{Y_{i,n} - \overline{F}^{-1} \left(\frac{1}{n} \right)}{a_n}. \quad (\text{II.160})$$

Dans la limite $n \rightarrow +\infty$, les $\{U_i, i = 1, \dots, k = k_\theta(n)\}$ convergent en distribution ($\xrightarrow{\mathcal{D}}$) vers le vecteur aléatoire :

$$(U_1, \dots, U_k) \xrightarrow{\mathcal{D}} \Lambda^{(k)} \equiv (G_1, \dots, G_k), \quad (\text{II.161})$$

dont la probabilité jointe s'écrit [45, 53] :

$$p_{\Lambda_k}(g_1, \dots, g_k) = \chi_{g_1 > \dots > g_k} \exp \left(-e^{-g_k} - \sum_i g_i \right). \quad (\text{II.162})$$

Le changement de variable $\Lambda^{(k)} \equiv \{G_1, \dots, G_k\} \rightarrow \Lambda'^{(k)} \equiv \{\{\Delta_i = i(G_i - G_{i+1}), i = 1, \dots, k-1\}, G_k\}$ conduit à :

$$p_{\Lambda'_k}(\delta_1, \dots, \delta_{k-1}, g_k) = \exp \left(-e^{-g_k} - kg_k - \sum_{i < k} \delta_i \right), \quad (\text{II.163})$$

ce qui montre qu'asymptotiquement ($n \rightarrow +\infty$), les $\{\Delta_i, i = 1, \dots, k-1\}$ sont indépendants.

À partir de ces définitions, on peut dériver un certain nombre de propriétés de Ω_k , pour $k \geq 2$. Ces propriétés sont présentées par la suite.

Proposition 1. *À condition qu'il soit appliqué à un vecteur aléatoire de distribution p_{Λ_k} , l'estimateur Ω_k de variance minimale sous la contrainte d'être non biaisé à n fini (i.e., $\mathbb{E}[\Omega_k] = \overline{F}^{-1}(\frac{1}{n})$ ou, de manière équivalente, $\mathbb{E}[\sum_{i=1}^k \alpha_i G_i] \equiv 0$) est obtenu en choisissant les $\{\alpha_i, i = 1, \dots, k\}$ tels que :*

$$\begin{cases} \forall i, i \neq k & \alpha_i = \frac{\sum_{l=1}^{k-1} \frac{1}{l} - \gamma}{k-1} \\ & \alpha_k = 1 - (k-1)\alpha_1 \end{cases} \quad (\text{II.164})$$

Cette proposition résulte presque directement des propriétés suivantes (dont les preuves sont reportées en annexe F)

$$\mathbb{E}[G_k] = \gamma - \sum_{j=1}^{k-1} \frac{1}{j}, \quad (\text{II.165})$$

$$\text{Var}[G_k] = \frac{\pi^2}{6} - \sum_{j=1}^{k-1} \frac{1}{j^2}, \quad (\text{II.166})$$

$$\mathbb{E}[\Delta_j] = 1, \quad (\text{II.167})$$

$$\text{Var}[\Delta_j] = 1. \quad (\text{II.168})$$

Proposition 2. *Lorsque $n \rightarrow +\infty$, $k \rightarrow +\infty$ avec $\frac{k}{n} \rightarrow 0$, Ω_k est un estimateur consistant de $\overline{F}^{-1}(1/n)$.*

Il s'agit d'une conséquence directe de la combinaison de la proposition 1 ci-dessus avec le fait que pour toutes les variables aléatoires appartenant à la classe étudiée ici, le vecteur aléatoire $\{U_i, i = 1, \dots, k\}$ converge en distribution vers $\Lambda^{(k)}$. Une preuve plus détaillée est disponible en annexe F.

Proposition 3. *Lorsque $n \rightarrow +\infty$, $k \rightarrow +\infty$ avec $\frac{k}{n} \rightarrow 0$, Ω_k est asymptotiquement normal.*

Lorsque Ω_k est appliqué à un vecteur aléatoire dont la densité est exactement p_{Λ_k} , Ω_k peut être scindé en

$$\Omega_k = G_k + \alpha_k \sum_i \Delta_i. \quad (\text{II.169})$$

En utilisant le fait que le vecteur $\{\Delta_i, i = 1, \dots, k-1\}$ est composé de variables aléatoires i.i.d. de variance unitaire, on obtient

$$\frac{\Omega_k + \mathbb{E}[G_k] - G_k}{(\zeta(1; k-1) - \gamma)^2} \xrightarrow{\mathcal{D}} \mathcal{N}_{0,1} \text{ avec } \zeta(s; n) = \sum_{k=1}^n \frac{1}{k^s}. \quad (\text{II.170})$$

et on a la convergence en probabilité suivante

$$\mathbb{E}[G_k] - G_k \xrightarrow{\mathbb{P}} 0. \quad (\text{II.171})$$

Cela implique :

$$\frac{\Omega_k}{(\zeta(1; k-1) - \gamma)^2} \xrightarrow{\mathcal{D}} \mathcal{N}_{0,1}. \quad (\text{II.172})$$

Cela montre que Ω_k est normal pour des n finis lorsqu'on l'applique à un vecteur aléatoire dont la loi est exactement p_{Λ_k} . La normalité asymptotique dans le cas général s'ensuit en utilisant la convergence en distribution de $\{U_i, i = 1, \dots, k = k_\theta(n)\}$ vers $\Lambda^{(k)}$. Comme précédemment, une preuve détaillée est développée en annexe F.

La consistance de $\theta^{(k)}$ est une conséquence directe de celle de Ω_k . Cependant, il faut noter que $\theta^{(k)}$ n'est pas asymptotiquement normal.

Ces résultats analytiques sont obtenus pour des choix de $k_\theta(n)$ vérifiant la contrainte décrite en Eq. (II.156). Le choix précis de $k_\theta(n)$ est cependant une question délicate qui peut être reformulée comme un cas classique de compromis biais-variance : de grandes valeurs de k devraient fournir une variance plus faible alors que de faibles valeurs de k sont nécessaires pour s'assurer de la convergence des statistiques d'ordre vers la loi des extrêmes ($\Lambda^{(k)}$) et donc un biais faible. Ce compromis est étudié ici grâce à des simulations de Monte-Carlo.

1.3 Performances : analyse numérique

Nous avons mené des simulations numériques en utilisant deux classes d'exemples, d'un côté les lois log-Weibull, de l'autre les lois log-gamma conjuguées, familles décrites dans la section 3, pour différentes valeurs de ρ – qui caractérise le comportement de la queue de la distribution (i.e. $\ln \bar{F}_Y(y) \approx -y^\rho$), et différentes tailles d'observation n . Pour l'ensemble des simulations, 500 réalisations indépendantes d'observations de n variables aléatoires i.i.d. sont générées.

À partir de celle-ci, le biais (relatif), la variance et l'erreur quadratique moyenne (EQM) de $\theta^{(k)}$ sont évalués et reportés dans la Fig. 5.1. Pour ces deux types de lois, il apparaît que, comme prévu par l'analyse théorique ci-dessus, augmenter k_θ de 1 à 2 donne lieu à une diminution significative du biais et par suite de l'EQM. Pour $k_\theta = 2$, $\alpha_1 \approx \alpha_2 \approx \frac{1}{2}$, une simple moyenne arithmétique de la première et de la seconde plus grande valeur observée donne donc de bien meilleurs résultats que le seul maximum $M_{Y,n}$. Ce biais ne décroît que légèrement lorsqu'on augmente k_θ au-delà

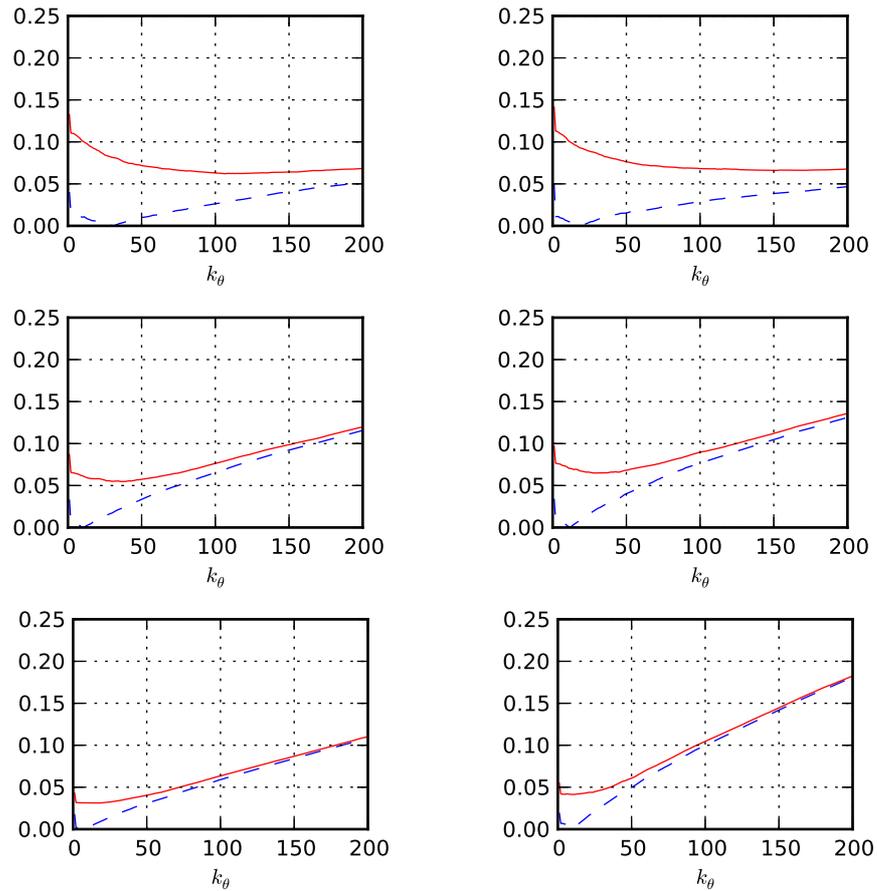


FIGURE 5.1 – $\hat{\theta}^{(k_\theta)}$ en fonction de k_θ : biais relatif et EQMs. À gauche : distribution log-Weibull ; à droite : lois log-gamma conjuguées. De haut en bas : $\rho = 1.2, 2, 4$, $n = 1000$. Lignes pointillées : biais relatif ; ligne pleine : EQM relative. Ces graphiques montrent essentiellement le bénéfice dû à l'accroissement de k_θ de 1 à 2. Ils montrent également que k_θ peut être augmenté davantage avec un effet positif tant que $k_\theta \ll n$. Ceci s'avère vrai pour tous les ρ .

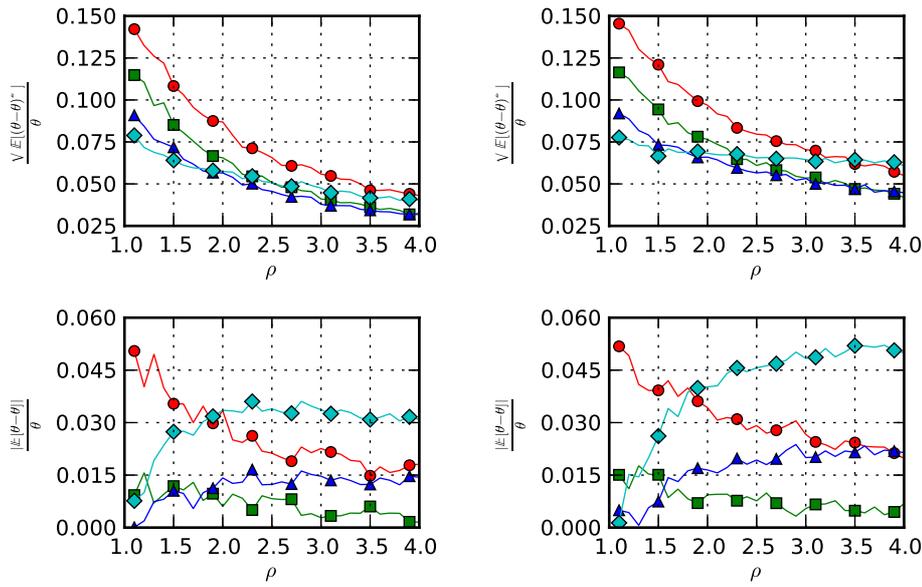


FIGURE 5.2 – $\hat{\theta}^{(k_\theta)}$ en fonction de ρ : biais et EQM relatifs. À gauche : distribution log-Weibull ; à droite : lois log-gamma conjuguées. En haut : EQM relative ; en bas : biais relatif. $n = 1000$. 'o' rouge : $k_\theta = 1$; '□' vert : $k_\theta = 2$; '△' bleu : $k_\theta = 25$; '◇' cyan : $k_\theta = 50$. Ces graphiques montrent une fois encore les bénéfices liés à l'augmentation de k_θ de 1 et 2, et l'existence d'un k_θ optimal dépendant essentiellement de n mais aussi légèrement de ρ .

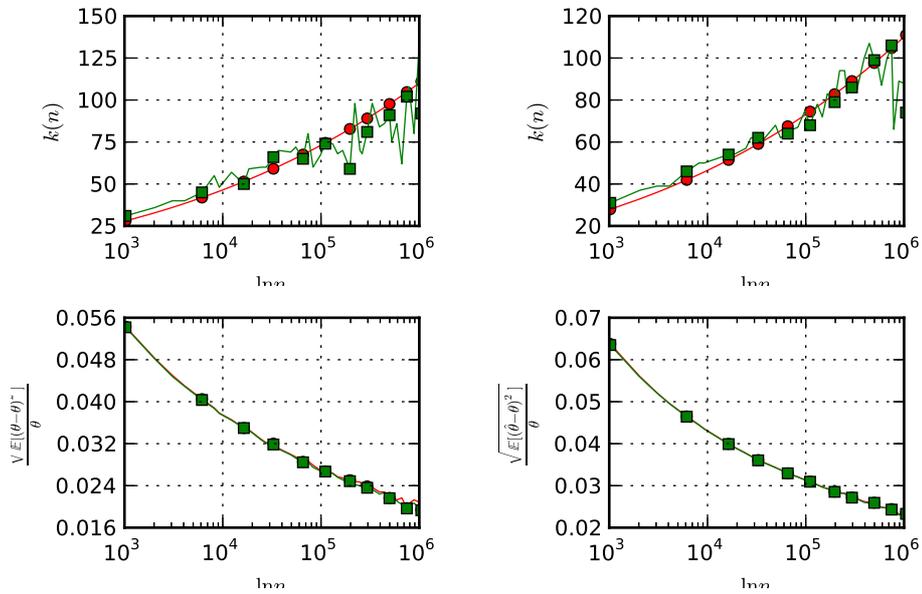


FIGURE 5.3 – $\hat{\theta}^{(k_\theta)}$: $k_{\theta,opt}$ et EQMs en fonction de $\ln n$. À gauche : distribution log-Weibull ; À droite : loi log-gamma conjuguée. En haut : $k_{\theta,opt}$, en bas : EQM. 'o' rouge : $k_\theta = \exp(\sqrt{1.6 \ln n})$; '□' vert : $k_{\theta,opt}$: EQM minimale pour $k_\theta < 1000$. $\rho = 2$. Ces courbes montrent l'adéquation des formules empiriques pour $k_{\theta,opt}(n)$.

de 2, avant de finalement augmenter à nouveau lorsque k_θ devient non négligeable par rapport à n . L'erreur quadratique moyenne, combinant biais et variance, montre l'intérêt d'utiliser des k_θ plus grand que 2 mais vérifiant $k_\theta \ll n$, ce quel que soit ρ . Ce bénéfice bien que modéré est indubitablement présent lorsqu'on compare k_θ à $k_\theta = 2$ ou aux grandes valeurs k_θ (pour lesquelles k_θ est trop proche de n). On peut également noter que plus ρ s'approche de 1, plus grandes sont les valeurs de k_θ minimisant l'erreur quadratique moyenne.

En complément, la Fig. 5.2 reporte les erreurs quadratiques moyennes relatives (à gauche) accompagnées du biais relatif (à droite) en fonction de ρ , pour différentes valeurs de k_θ . En plus de confirmer que le bénéfice dû à l'usage de $k_\theta = 2$, ou légèrement plus grand, par rapport à $k_\theta = 1$ existe quel que soit ρ , cette figure montre également que le biais et l'erreur quadratique moyenne demeurent à peu près constants pour $\rho \geq 2$, tandis qu'ils s'accroissent significativement (dans la plupart des cas) lorsque $\rho \rightarrow 1^+$.

Pour finir, la Fig. 5.3 compare la valeur optimale $k_{\theta,opt}$ de k_θ (i.e., la valeur de k_θ associée à la plus faible EQM) à des paramétrisations empiriques potentielles de k_θ en fonction de n : $k_\theta(n) = 10 \ln n - 40$, $k_\theta(n) = \exp(\sqrt{1.6 \ln n})$. Bien que totalement empiriques, ces expressions donnent des résultats corrects pour les deux types de distributions utilisées et ce sur un large intervalle de valeurs de ρ ($\rho \in [1.5, 3]$) et de n , et peuvent donc servir comme valeur par défaut pratique pour l'estimateur.

2 Estimation de $\rho_l(y_m(n))$

2.1 Définition

Par simplicité, $\rho_l(y^\dagger(n))$ est abrégé dans la suite en ρ_E . D'après sa définition (cf. Eq. (II.124)), ρ_E peut être interprété comme l'exposant local mesurant le comportement de la fonction Ψ (cf. Eq. (II.18)) autour de la plus grande observation disponible $y^\dagger(n)$. Cela suggère d'utiliser directement la cumulative empirique afin d'estimer ρ_E . En effet, pour un intervalle suffisamment petit autour de y^\dagger , on a $\partial_{\ln y} h(y^\dagger) = \rho_E$, ce qui nous permet d'écrire :

$$\ln(-\ln \overline{F}_Y(y^\dagger)) \approx \rho_E \ln y^\dagger + \beta. \quad (\text{II.173})$$

Si on considère la liste ordonnée des statistiques d'ordre $Y_{i,n}$, le $(1 - \frac{i}{n})$ -ième quantile peut être approché par $Y_{i,n}$, par suite :

$$\ln\left(-\ln \frac{i}{n}\right) \approx \rho_E \ln Y_{i,n} + \beta. \quad (\text{II.174})$$

Cela suggère que ρ_E peut être estimé à l'aide d'un ajustement par moindres carrés (non pondérés) à partir des points $Y_{i,n}$ pour $i \leq k_\rho$, où k_ρ est suffisamment petit comparé à n :

$$\hat{\rho}_E = \frac{C(\ln y, \ln(\ln n - \ln i))}{C(\ln y, \ln y)}, \quad (\text{II.175})$$

avec $C(x, y) = \overline{xy} - \overline{x} \overline{y}$ où $\overline{x} = \frac{1}{k} \sum_{i=1}^{k_\rho} x_i$.

2.2 Performances : analyse théorique

De nouveau, le choix de k_ρ est un problème délicat qui peut se réinterpréter dans l’idiome usuel d’un compromis biais-variance : parce que ρ_E est un exposant local, qui peut converger lentement vers ρ , il est nécessaire de l’estimer au sein d’un intervalle étroit autour de y^\dagger , ce qui argue pour le choix d’un petit k_ρ ; au contraire, de grands k_ρ diminuent la variance de l’estimateur, au prix d’une augmentation du biais.

Cela suggère de choisir k_ρ de telle sorte que :

$$\lim_{n \rightarrow +\infty} \frac{k_\rho(n)}{n} = 0. \quad (\text{II.176})$$

2.3 Performances : analyse numérique

Les performances de $\hat{\rho}_E$ ainsi que les conséquences en termes de compromis biais-variance du choix de $k_\rho(n)$ sont étudiées numériquement à travers des simulations de Monte-Carlo, menées de la même manière que dans la section 1.3.

Le biais et l’EQM relatifs sont reportés dans la Fig. 5.4, en fonction k_ρ . On peut observer que pour les distributions log-Weibull, le biais est négligeable et plus k_ρ augmente, plus l’EQM décroît. Cela provient du fait que pour ces lois $\rho_E(n) \equiv \rho, \forall n$, ce qui implique que $\hat{\rho}_E$ est non-biaisé quel que soit k_ρ . Par conséquent augmenter k_ρ diminue la variance à coût nul. Cependant, il s’agit d’un cas particulier qui s’éloigne fortement du cas général, qui est mieux modélisé ici par le second exemple, i.e., les distributions log-gamma conjuguées. Pour ce deuxième exemple, la Fig. 5.4 indique que le choix optimal de $k_\rho(n)$ dépend de la valeur ρ et doit être choisi bien plus grand pour $\rho \rightarrow 1$.

Le biais et l’erreur quadratique moyenne relatifs sont reportés sur la Fig. 5.5, en fonction de ρ . Cette figure montre que pour les distributions log-Weibull, augmenter k_ρ procure un bénéfice uniforme en ρ . Au contraire, pour les lois log-gamma conjuguées, la valeur optimale de k_ρ augmente lorsque $\rho \rightarrow 1$.

Pour finir, la Fig. 5.6 compare $k_{\rho,opt}$ (i.e., k_ρ associé à l’EQM minimale) à des choix empiriques potentiels pour l’estimation $k_\rho(n) : k_\rho(n) = 50 \ln n, k_\rho(n) = 8n^{1/3}$. Comme précédemment, ces formules bien qu’empiriques donnent des repères pratiques pour l’utilisation de l’estimateur.

3 Estimation de q^*

3.1 Définition

À partir de $\hat{\theta}^{(k_\theta)}$ et $\hat{\rho}_E^{(k_\rho)}$ définis plus haut, on peut proposer un estimateur de q^* construit à partir de leur produit :

$$\hat{q}^{*(k_\theta, k_\rho)} = \hat{\theta}^{(k_\theta)} \hat{\rho}_E^{(k_\rho)}. \quad (\text{II.177})$$

Pour étudier les performances de \hat{q}^* , son biais $\mathcal{B}[\hat{q}^*] = \mathbb{E}[\hat{q}^* - q^*]$ et sa variance $\text{Var}[\hat{q}^*]$ peuvent être mis sous la forme :

$$\mathcal{B}[\hat{q}^*] = \text{Cov}[\hat{\theta}, \hat{\rho}_E] + \mathcal{B}[\hat{\theta}](\rho_E + \mathcal{B}[\hat{\rho}_E]) + \mathcal{B}[\hat{\rho}_E](\theta + \mathcal{B}[\hat{\theta}]), \quad (\text{II.178})$$

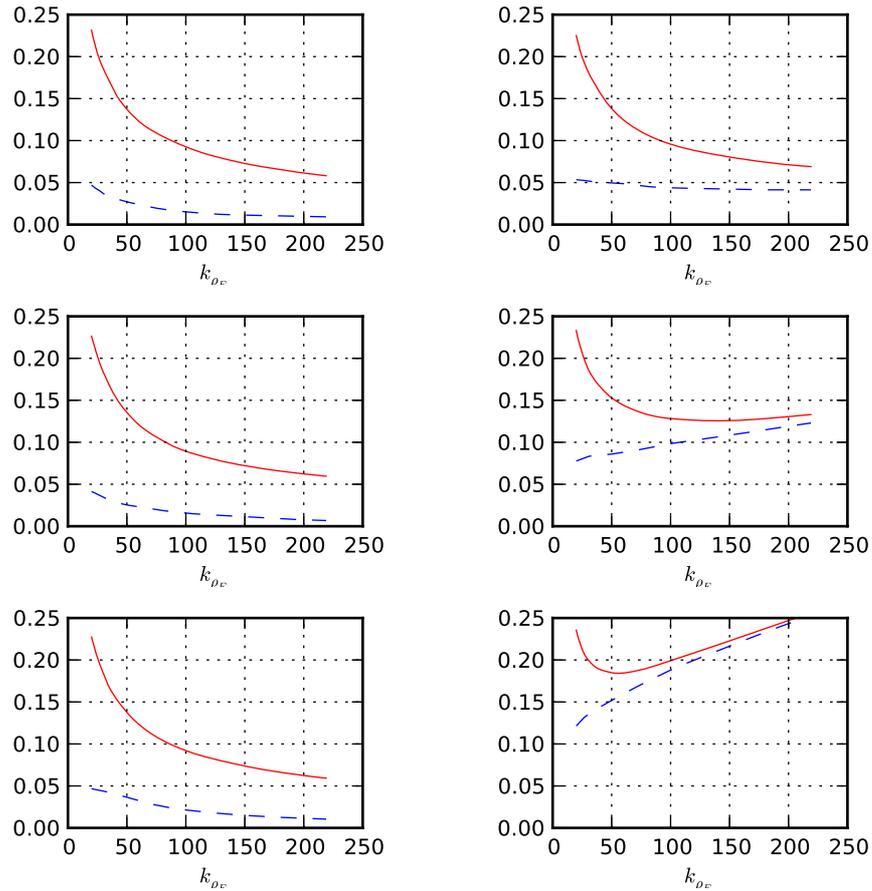


FIGURE 5.4 – $\hat{\rho}_E^{(k_\rho)}$ en fonction de k_ρ : biais et EQM relatifs. À gauche : distribution log-Weibull ; à droite : distribution log-gamma conjuguée. De haut en bas : $\rho = 1.2, 2, 4$, $n = 1000$. Ligne bleue pointillée : biais relatifs ; ligne rouge pleine : EQMs relatifs. Pour les distributions log-weibull, parce que $\rho_E(n) \equiv \rho$, augmenter k_ρ a toujours un effet bénéfique en faisant seulement diminuer la variance. Cependant, en général, il existe un $k_{\rho,opt}$ optimal qui dépend de n et ρ .

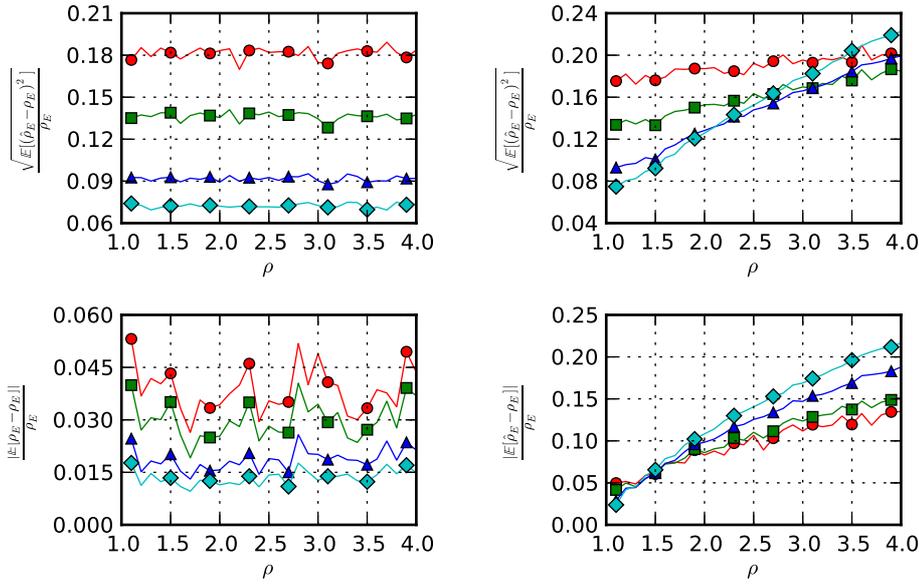


FIGURE 5.5 – $\hat{\rho}_E^{(k_\rho)}$ en fonction de ρ : biais et EQM relatifs. À gauche : distribution log-Weibull ; à droite : distribution log-gamma conjuguée. 'o' rouge : $k_\rho = 30$; '□' vert : $k_\rho = 50$; '△' bleu : $k_\rho = 100$; '◇' cyan : $k_\rho = 200$.

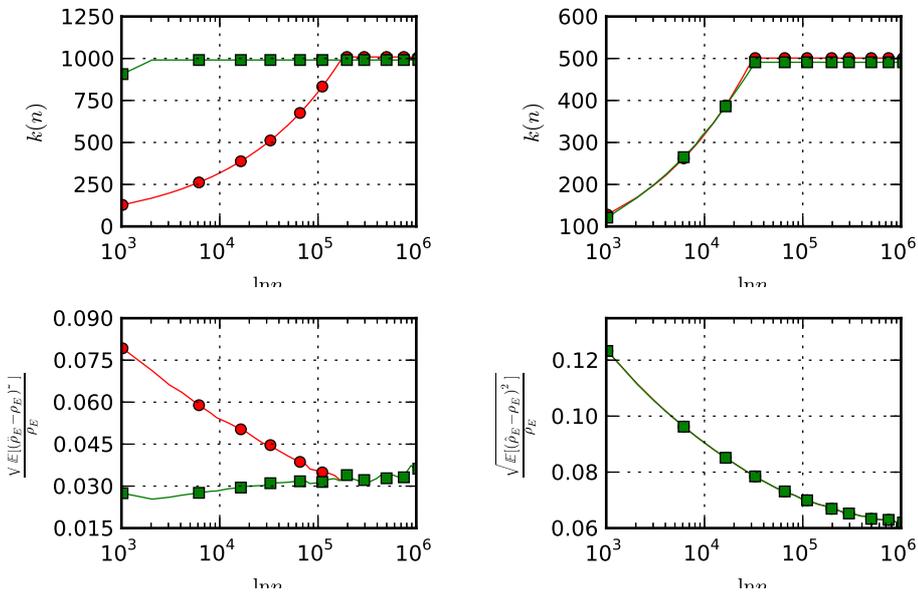


FIGURE 5.6 – $\hat{\rho}_E^{(k_\rho)}$: $k_{\rho,opt}$ et EQMs en fonction de $\ln n$. À gauche : distribution log-Weibull ; à droite : distribution log-gamma conjuguée. En haut : $k_{\rho,opt}$, en bas : EQM. $\rho = 2$. 'o' rouge : $k_\rho(n) = 8n^{0.4}$; '□' vert : $k_{\rho,opt}$: EQM minimale pour $k_\rho < 1000$. $\rho = 2$. Ces courbes montrent l'adéquation des formules empiriques proposées avec $k_{\rho,opt}(n)$.

et

$$\begin{aligned}
 \text{Var} [\hat{q}^*] &= \text{Var} [\hat{\theta}] \text{Var} [\hat{\rho}_E] \\
 &+ \text{Cov} [\hat{\theta}^2, \hat{\rho}_E^2] - \text{Cov} [\hat{\theta}, \hat{\rho}_E]^2 \\
 &+ \text{Var} [\hat{\theta}] (\rho_E + \mathcal{B}[\rho_E])^2 + \text{Var} [\hat{\rho}_E] (\theta + \mathcal{B}[\theta])^2 \\
 &- 2\text{Cov} [\hat{\theta}, \hat{\rho}_E] (\rho_E + \mathcal{B}[\rho_E])(\theta + \mathcal{B}[\theta]).
 \end{aligned} \tag{II.179}$$

De plus, si on suppose que les biais relatifs sont négligeables, on obtient :

$$\begin{aligned}
 \frac{\mathcal{B}[\hat{q}^*]}{q^*} &\approx \frac{\mathcal{B}[\hat{\theta}]}{\theta} + \frac{\mathcal{B}[\hat{\rho}_E]}{\rho_E} + \frac{\text{Cov} [\hat{\theta}, \hat{\rho}_E]}{\rho_E \theta}. \\
 \frac{\text{Var} [\hat{q}^*]}{(q^*)^2} &\approx \frac{\text{Var} [\hat{\theta}] \text{Var} [\hat{\rho}_E]}{(\rho_E \theta)^2} \\
 &+ \frac{\text{Cov} [\hat{\theta}^2, \hat{\rho}_E^2] - \text{Cov} [\hat{\theta}, \hat{\rho}_E]^2}{(\rho_E \theta)^2} \\
 &+ \frac{\text{Var} [\hat{\theta}]}{\theta^2} + \frac{\text{Var} [\hat{\rho}_E]}{\rho_E^2} - 2 \frac{\text{Cov} [\hat{\theta}, \hat{\rho}_E]}{\rho_E \theta}.
 \end{aligned} \tag{II.180}$$

$$\tag{II.181}$$

Ces calculs montrent que les performances de \hat{q}^* dépendent significativement de la covariance entre les valeurs estimées $\hat{\theta}$ et $\hat{\rho}$. Des simulations numériques permettent d'enquêter plus en détail sur cette covariance.

3.2 Performances : analyse numérique

Pour analyser plus avant cet estimateur, nous utilisons des simulations de Monte-Carlo, fort semblables en esprit à celles présentées précédemment, utilisant les choix empiriques détaillés plus haut pour $k_\theta(n)$ et $k_\rho(n)$. La Fig. 5.7 montre que la covariance $\text{Cov} [\hat{\theta}, \hat{\rho}_E]$ reste assez élevée et positive, quels que soient ρ et n , par conséquent celle-ci contribue significativement au biais et à la variance de \hat{q}^* . De plus, les résultats reportés dans la Fig. 5.8 montrent essentiellement que l'EQM pour \hat{q}^* varie entre 5% et 25%, suivant la distribution de Y et la valeur de ρ , avec des valeurs de l'EQM plus grandes lorsque $\rho \rightarrow 1$. Bien que ces valeurs de l'EQM puissent sembler élevées, \hat{q}^* procure néanmoins aux expérimentateurs une procédure empirique qui mène à un ordre de grandeur satisfaisant du véritable q^* : dans les applications, il est rare d'avoir besoin d'une valeur précise de q^* , alors que la connaissance d'un ordre de grandeur estimé pour q^* s'avère souvent vitale dès qu'il est nécessaire d'interpréter les données.

4 Séries temporelles corrélées

Jusqu'à présent, nous n'avons traité que les cas d'observations $\{X_1, \dots, X_n\}$ constituées d'échantillons indépendants entre eux, alors même que nous nous intéressons principalement à l'analyse des séries temporelles stationnaires, pour lesquelles il est plus vraisemblable, en pratique, que ces échantillons soient corrélés. Le but de la section suivante est d'analyser l'impact de la dépendance au sein des observations

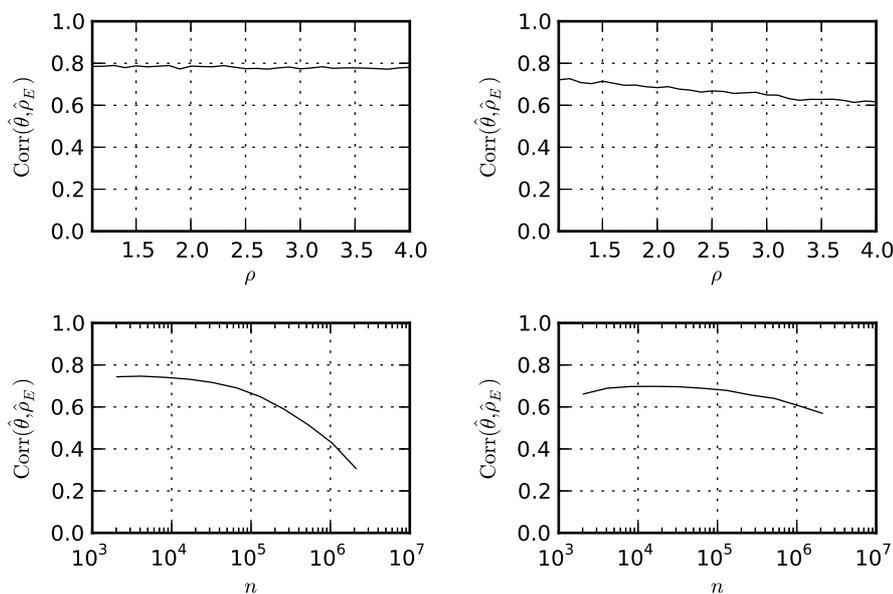


FIGURE 5.7 – **Corrélation entre $\hat{\theta}^{(k_{\theta, opt})}$ et $\hat{\rho}_E^{(k_{\rho, opt})}$** . À gauche : distribution log-Weibull ; à droite : distribution log-gamma conjuguée. En haut : en fonction de ρ , $n = 1000$. En bas : en fonction de n , $\rho = 2$.

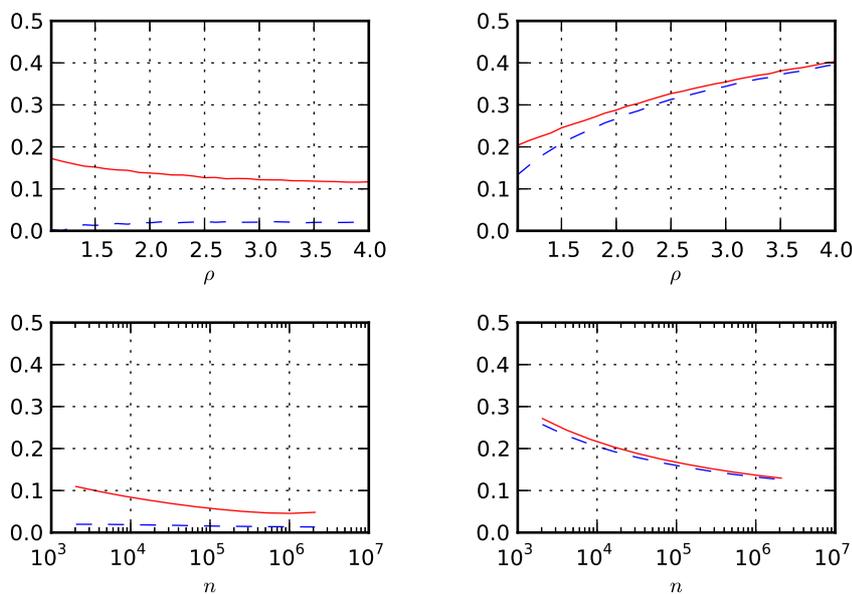


FIGURE 5.8 – \hat{q}^* : **biais relatif (ligne bleu tiretée) et EQMS (ligne rouge pleine)**. À gauche : distribution log-Weibull ; à droite : distribution log-gamma conjuguée. En haut : en fonction de ρ , $n = 1000$. En bas : en fonction de n , $\rho = 2$.

sur les définitions et l'estimation de q^* . Nous proposerons également des modifications simples et élémentaires de \hat{q}_c afin de s'accommoder de la présence de telles dépendances.

4.1 Analyse théorique

Dans le chapitre 4 plus haut, le moment critique émerge à partir d'un équilibre entre deux facteurs en opposition : la contribution dominante au moment opposée à l'effet de taille finie. Si la définition et l'évaluation du premier ne dépendent que de la distribution marginale des données et ne sont que peu affectées par la dépendance, le second facteur est bien plus influencé par les corrélations. En effet, la région accessible pour une taille finie fut définie par :

$$\mathbb{P}(\forall k = 1, \dots, n, Y_k < y^\dagger) = \frac{1}{e}. \quad (\text{II.182})$$

Pour des données indépendantes, cette équation se simplifie en $\Psi(y^\dagger) = \ln n$. Pour une série temporelle corrélée, il est naturel de s'attendre à ce que y_T^\dagger puisse être définie à travers un nombre effectif d'échantillons n_{eff} comme nous l'avons fait dans le cas multifractal :

$$\Psi(y_T^\dagger) = \ln n_{\text{eff}}. \quad (\text{II.183})$$

À partir de cette modification, l'analyse conduite en section 3.3 suggère de définir un ordre critique des moments en présence de dépendances q^{*T} comme :

$$q^{*T}(n) = q^*(n_T^*). \quad (\text{II.184})$$

Comme dans le cas indépendant, q^{*T} peut être divisé en deux termes,

$$q^{*T}(n) = \rho_E^T(n)\theta^T(n). \quad (\text{II.185})$$

En réinjectant y_T^\dagger dans la définition de ρ_E^T et θ^T on obtient

$$\rho_E^T(n) = \rho_E(n_T^*), \quad \theta^T(n) = \theta(n_T^*). \quad (\text{II.186})$$

Appelons τ la longueur de corrélation de la série temporelle analysée, définie comme $\tau = \int_0^{+\infty} tC(t)dt / \int_0^{+\infty} C(t)dt$ où $C(t) = EX_0X_t - \mathbb{E}[X_0]^2$ est la fonction d'autocovariance de la série. Si τ est bien définie, il est naturel de penser que n_{eff} se comporte essentiellement comme $n^* \propto \frac{n}{\tau}$, avec également, comme condition naturelle, $\lim_{\tau \rightarrow 0} n_{\text{eff}}(n, \tau) = n$, ce qui suggère la forme phénoménologique :

$$n_{\text{eff}} = \frac{n}{1 + \kappa\tau}, \quad (\text{II.187})$$

avec κ une constante positive devant être déterminée. Cela mène à :

$$q^{*T}(n) = q^*(\tau, n) = q^*(n/(1 + \kappa\tau)). \quad (\text{II.188})$$

4.2 Impact des dépendances sur \hat{q}^*

Afin d'étudier l'impact de la dépendance sur \hat{q}^* (cf. la définition (II.177)), l'estimateur est appliqué sur un grand nombre de réalisations indépendantes de séries temporelles stationnaires, de longueur n , de distributions marginales (log-Weibull ou log-gamma conjuguées) et fonctions de covariance $C(t)$ prescrites a priori. Ces séries sont synthétisées numériquement en utilisant une méthode décrite dans [66, 64] basée sur un développement en polynôme d'Hermite et les matrices circulantes enveloppantes (*circulant embedded matrix*). Pour mieux comprendre l'impact de la covariance, nous analyserons le cas où $C(t)$ prend la forme simple :

$$C(t) = e^{-\frac{|t|}{\tau}}, \quad (\text{II.189})$$

avec τ une constante ajustable, correspondant à la longueur de corrélation.

Les résultats, reportés dans la Fig. 5.9, montrent en premier lieu un fort désaccord entre \hat{q}^* et $q^*(\tau, n)$ et en second lieu et de manière cruciale, qu'aucune opération de mise à échelle du type proposé par l'Eq. (II.188) ne peut corriger ce désaccord. En effet, une mise à échelle comme celle décrite par l'Eq. (II.187) correspond uniquement à des translations horizontales de \hat{q}^* . De plus, si on suppose que $n_{\text{eff}} < n$, seules des translations à gauche sont possibles, ce qui rend impossible de superposer toutes les courbes $\hat{q}_c(\tau, n)$ par une mise à échelle.

L'estimateur \hat{q}^* ne converge donc pas vers $q^*(\tau, n)$. Pour analyser plus avant la cause de cette incohérence, étudions séparément $\hat{\theta}^{(k_\theta)}$ et $\hat{\rho}_E^{(k_\rho)}$.

La Fig. 5.10 obtenue à partir des mêmes simulations de Monte-Carlo, avec $k_\theta = 1$, indique qu'il existe un κ (≈ 0.08 ici) tel que $\mathbb{E} \left[\frac{\hat{\theta}}{\ln n} \right] \approx \frac{\theta(n_{\text{eff}})}{\ln n_{\text{eff}}}$ (assez logiquement $\hat{\theta} = \ln n_{\text{eff}}/\Omega$ doit être utilisé à la place de $\hat{\theta} = \ln n/\Omega$, comme $\ln n_{\text{eff}}$ par définition ne varie pas avec τ .) Cela signifie que Ω_1 exhibe le comportement attendu, $\mathbb{E} [\Omega_1(n, \tau)] = \mathbb{E} [\Omega_1(n_{\text{eff}})]$. Cependant, augmenter k_θ de 1 à 2 et au-delà ne réduit pas le biais comme dans le cas indépendant mais au contraire l'augmente significativement. Cela peut facilement être expliqué, puisqu'il est fort probable que la seconde plus grande valeur soit très proche de la plus grande valeur et fortement corrélée avec elle, et de ce fait n'apporte aucune réduction du biais. Au contraire, cette corrélation empêche la convergence des statistiques d'ordre vers leur limite statistique et compromet la consistance de Ω .

Cette corrélation significative des statistiques d'ordre explique également les faibles performances de $\hat{\rho}_E$. Comme illustré par la Fig. 5.11, $\hat{\rho}_E$ ne converge pas vers $\rho_E(\tau, n)$ et une mise à l'échelle n'est pas suffisante pour expliquer ce désaccord. Une interprétation potentielle est que $\hat{\rho}_E$ est fortement lié à la convergence du $k^{\text{ième}}$ ordre statistique vers le $(1 - k/n)$ quantile. Les corrélations des échantillons génèrent des corrélations parmi les statistiques d'ordre parce que les échantillons corrélés ont tendance à se regrouper par clusters en terme de rang. La corrélation des échantillons se transfère donc aux statistiques d'ordre, ce qui entrave drastiquement la convergence des statistiques d'ordre vers le quantile.

Par conséquent, appliqué à des séries temporelles stationnaires corrélées, \hat{q}^* , comme défini en Section 3, ne peut converger vers q^{*T} . Cependant, il est possible de modifier simplement l'estimateur pour tenir compte efficacement des corrélations. Ces modifications consistent à imposer qu'une fois la plus grande valeur sélectionnée, la seconde plus grande valeur doit être choisie suffisamment éloignée de la première,

la troisième valeur maximale est ensuite prise suffisamment loin des précédentes et ainsi de suite.

L'estimateur modifié \hat{q}^{*T} consiste ainsi essentiellement à l'application de l'estimateur original \hat{q}^* à ces statistiques d'ordre suffisamment éloignés.

4.3 Estimateurs modifiés

Une manière efficace d'implémenter l'idée décrite plus haut est de construire l'ensemble des échantillons effectivement indépendants étape par étape. Considérons l'observation originale $\mathcal{P}_1 = X_1, \dots, X_n$. En premier lieu, le maximum $X_{k_1} = \max\{\mathcal{P}_1\}$ est sélectionné. Ensuite, tous les échantillons proches de X_{k_1} en temps sont supprimés. En utilisant la distance (paramétrée par une constante arbitraire β)

$$d_\beta(X_i, X_j) = \max(|j - i|, \beta \text{Card}\{k, \min(X_i, X_j) < X_k < \max(X_j, X_i)\}), \quad (\text{II.190})$$

l'ensemble \mathcal{P}_1 est criblé en éliminant les échantillons appartenant à la boule de centre X_{k_1} et de rayon s . Un nouvel ensemble $\mathcal{P}_2 = \mathcal{P}_1 \setminus B_{d_\beta}(X_{k_1}, s)$ est défini et la procédure est répétée jusqu'à obtenir l'ensemble des points effectivement indépendants $\mathcal{P}_e = \{X_{k_1}, \dots, X_{k_l}\}$. Les estimateurs modifiés $\hat{\theta}^T$ et $\hat{\rho}_E^T$ sont alors définis par l'application des estimateurs originaux $\hat{\theta}$ ou $\hat{\rho}_E$ à \mathcal{P}_e .

Des simulations de Monte-Carlo conduites comme décrit précédemment suggèrent que le seuil devrait être choisi suivant la formule empirique :

$$s = \alpha\tau, \quad (\text{II.191})$$

avec le choix de $\alpha \approx 0.01$, comme illustré par la Fig. 5.12.

Ces simulations de Monte-Carlo montrent également que l'estimation de θ une fois corrigée (en supposant la valeur de n_{eff} connue) mène au résultat attendu $\theta(n_{\text{eff}})$ (cf. la Fig. 5.13) et q^* peut donc être estimé de façon très satisfaisante (cf. la Fig. 5.14).

Ces performances remarquables requièrent cependant une connaissance a priori de τ , et les paramètres α et κ furent fournis par des simulations de Monte-Carlo. Ces paramètres ne sont pas connus pour des données réelles et nécessiteraient une estimation indépendante. De telles estimations vont au-delà du cadre de cette thèse. Elles pourraient se baser sur l'estimation spectrale ou d'autres techniques classiques. Cette question est encore ouverte et pourrait faire l'objet d'éventuelles futures contributions. Cependant, afin de conclure cette étude, la Fig. 5.15 reporte les espérances et écart-types de \hat{q}_c^T , appliqués à des lois log-normales corrélées exponentiellement avec $\tau = 100$, mais pour lesquelles la valeur de τ utilisée dans \hat{q}^{*T} est volontairement erronée. De façon satisfaisante, la Fig. 5.15 ne montre qu'une faible sensibilité à la valeur de τ utilisée, ce notamment lorsque la valeur de τ est surestimée. Cela signifie que, en pratique, une estimation grossière de la véritable longueur de corrélation pourrait s'avérer suffisante pour obtenir des performances adéquates de \hat{q}^* .

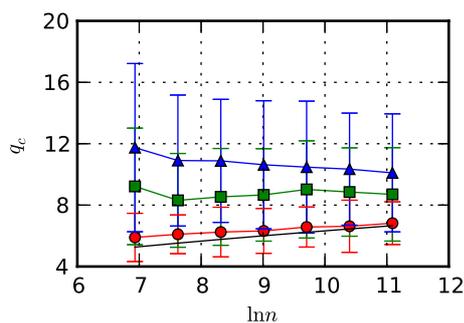


FIGURE 5.9 – $\hat{q}_c^{(k_\theta, k_\rho)}$: **Impact des dépendances.** Appliqué à des observations corrélées, \hat{q}^* est perturbé par un biais structurel qui ne peut être corrigé par une mise à l'échelle (de la forme suggérée par l'Eq. (II.188)). $(k_\theta, k_\rho) = (1, 100)$. Ligne noire : véritable q^* . Ligne avec marqueurs : $\hat{q}_c^{(k_\theta, k_\rho)}$, pour $\tau = 10$ (cercles rouges), $\tau = 50$ (carrés verts), $\tau = 100$ (triangles bleus).

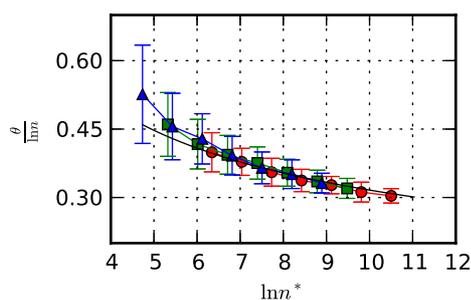


FIGURE 5.10 – $\hat{\theta}$: **Impact des dépendances.** Mise à l'échelle de $\frac{\theta}{\ln n}$ en fonction de $n^* = n/(1 + \kappa\tau)$, pour $\tau \in \{10, 50, 100\}$ (cercles rouges $\tau = 10$, carrés verts $\tau = 50$, triangles bleus $\tau = 100$). La ligne noire correspond à la valeur théorique $\theta(n)/\ln n$.

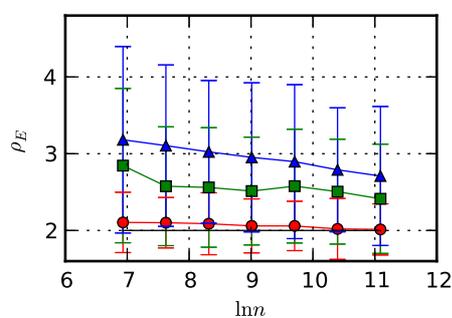


FIGURE 5.11 – $\hat{\rho}_E$: **Impact des dépendances.** Ligne noire : valeur théorique, cercles rouges $\tau = 10$, carrés verts $\tau = 50$, triangles bleus $\tau = 100$.

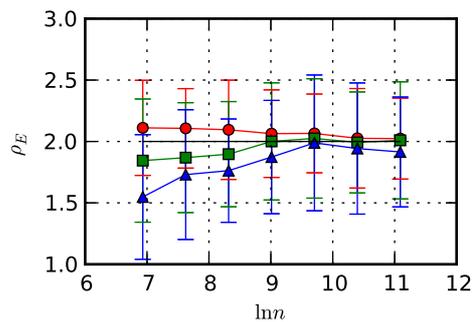


FIGURE 5.12 – **Estimateur modifié** $\hat{\rho}_E^T$. $k_\rho = 100$. Ligne noire : valeur théorique, cercles rouges $\tau = 10$, carrés verts $\tau = 50$, triangles bleus $\tau = 100$.

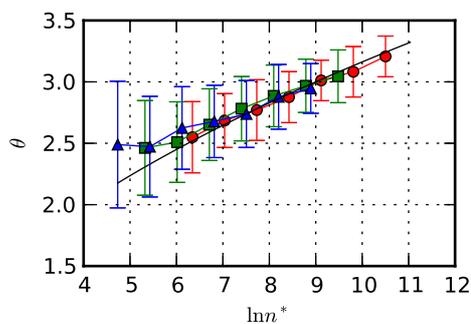


FIGURE 5.13 – **Estimation de $\theta(n^*)$ avec l'estimateur corrigé** $\hat{\theta}^T$. $k_\theta = 10$. Ligne noire : valeur théorique, cercles rouges $\tau = 10$, carrés verts $\tau = 50$, triangles bleus $\tau = 100$.

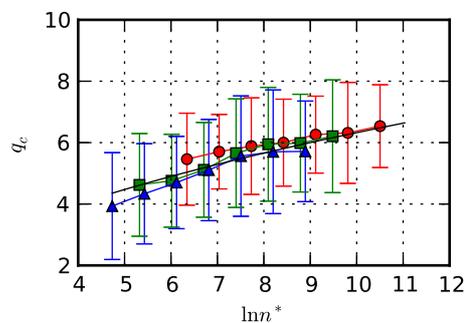


FIGURE 5.14 – **Estimation de $q^*(n^*)$ avec l'estimateur corrigé**. Ligne noire : valeur théorique, cercles rouges $\tau = 10$, carrés verts $\tau = 50$, triangles bleus $\tau = 100$.

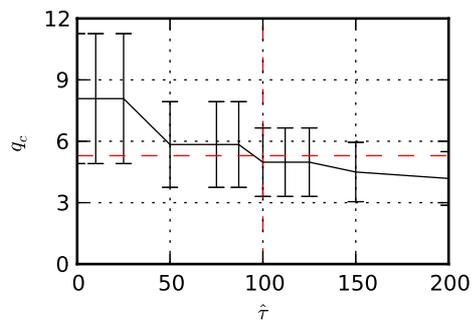


FIGURE 5.15 – \hat{q}_c^T : **Influence du choix de τ .** Les lignes noires pleines montrent la valeur moyenne et l'écart-type de \hat{q}_c^T pour différents a priori sur τ . La ligne verticale en tireté à $\tau = 100$ correspond à la véritable longueur de corrélation des données. La ligne horizontale en pointillé montre la valeur théorique de $q^*(n, \tau)$.

Conclusions et discussion

Dans ce chapitre, nous avons discuté de la question de la définition et de l'estimation d'un ordre critique q^* jusqu'auquel il est possible d'estimer correctement les moments théoriques à partir d'une observation de taille donnée. Notre travail s'est concentré sur deux classes très distinctes de lois aléatoires. En premier lieu, nous avons étudié les processus i.i.d. (classe A) de type II. C'est-à-dire des lois dont tous les moments sont finis mais dont la queue varie suffisamment lentement pour que leur fonction caractéristique ne soit pas analytique en 0 (telle que par exemple, les distributions log-normales). Ce cas apparaît dans certaines applications modernes, telles que par exemple le REM en physique statistique, ou bien les cascades multiplicatives inspirées par la turbulence hydrodynamique. En contrepoint, nous nous sommes intéressés à la classe des processus multifractals et plus particulièrement aux accroissements d'un mouvement de Poisson composé, choisi en tant que représentant simple des processus multifractals. Pour ces deux classes de processus, on voit apparaître un effet de linéarisation : au-delà d'un certain ordre critique, moments théoriques et estimateur des moments ne coïncident plus.

Nous avons montré qu'il est possible de donner une définition formelle, quantitative et unifiée pour cet ordre critique en combinant des effets de concentration et de taille finie qui apparaissent pour ces classes de processus. Cependant le comportement de l'ordre critique s'est avéré très spécifique à la classe considérée.

5 Classe A

Dans la classe A, le paramètre ρ quantifiant le poids de la queue de la distribution joue un rôle prépondérant dans la croissance de l'ordre critique q^* en fonction de la taille de l'observation. Le cas plus standard des lois ayant une queue gaussienne ou en loi de puissance peut d'ailleurs être retrouvé en prenant respectivement les limites $\rho \rightarrow +\infty$ et $\rho \rightarrow 1$.

Nous avons défini un estimateur qui peut être appliqué sur une réalisation unique d'un jeu de données réelles. Nous avons mené une analyse théorique et numérique, s'appuyant sur des simulations de Monte-Carlo. Cet estimateur procure aux praticiens un outil prêt à l'emploi. Des procédures MATLAB implémentant ces estimateurs sont disponibles sur requête.

La question de la dépendance au sein de ces observations a été abordée. Nous avons montré que celle-ci dégrade la performance de l'estimateur. Cependant, nous avons proposé une modification simple de la procédure d'estimation pour prendre en compte la dépendance. L'efficacité de cette modification a été éprouvée par des simulations de Monte-Carlo. L'utilisation de cette procédure modifiée requiert seulement une estimation grossière d'une longueur de corrélation typique parmi les observa-

tions, qui peut être obtenue en utilisant des méthodes classiques de traitement du signal (telle que l'estimation spectrale).

La définition et la validation d'une procédure originale et systématique intégrant l'estimation de la longueur de corrélation typique au sein de l'estimation de q^* est à l'étude et sera peut-être l'objet d'une future contribution.

6 Classe B

Dans la Classe B, cet ordre critique est entièrement défini par les paramètres du processus et ne dépend pas du nombre d'échantillons présents dans l'observation : Augmenter la taille de l'échantillon (en diminuant la période d'échantillonnage δt) ne permet pas d'accroître l'ordre maximal des moments pouvant être évalués. Cet ordre critique n'est pas relié non plus à la divergence des moments au-delà d'un ordre limite q_c . De plus, pour $q \geq q^*$, l'estimateur des moments exhibe un comportement en loi de puissance en fonction de l'échelle, avec un exposant d'échelle qui varie cependant linéairement en q . L'ordre critique q^* et la pente de ce comportement linéaire sont prédits de manière quantitative. Ces prédictions sont obtenues en adaptant aux processus multifractals des arguments de physique statistique employés dans l'analyse du REM.

La raison pour laquelle augmenter la taille d'échantillonnage ne permet pas de calculer les moments pour $q > q^*$ peut être comprise comme l'analogie d'une propriété de *non auto-moyennage* des verres de spins de type REM. Ce travail rappelle par certains côtés l'analogie qui mène, dans le contexte de l'analyse multifractale, du formalisme thermodynamique au formalisme multifractal, analogie couramment utilisée pour mesurer le spectre multifractal à partir de données empiriques [50, 9].

Cette contribution peut ainsi être interprétée comme un effort supplémentaire pour rendre explicite la féconde connexion entre les formalismes thermodynamiques et multifractals, dans l'esprit de e.g., [50, 97, 9, 49], avec une emphase spécifique sur l'alliance de différentes perspectives sur l'effet de linéarisation : celle de l'estimation statistique basée sur les réalisations de trajectoires de processus stochastiques, celle de la physique statistique et celle de l'analyse fonctionnelle des régularités locales.

Des simulations de Monte-Carlo, basées sur la synthèse numérique de réalisations indépendantes de trajectoires du mouvement de Poisson composé et l'estimation de rapports de participation –un outil classique de la physique statistique de la matière condensée– ont permis de confirmer ces prédictions. Ces dites prédictions, basées sur des arguments de physique statistique de type REM, sont parfaitement cohérentes avec celles proposées dans [74, 6], reposant sur l'analyse de la statistique des extrêmes du formalisme multifractal. Elles procurent une interprétation de la raison pour laquelle les moyennes temporelles ne convergent pas vers les moyennes d'ensemble.

L'analyse poursuivie ici peut être appliquée de manière presque directe, *mutatis mutandis*, à d'autres processus multifractals. C'est notamment le cas pour le mouvement brownien fractionnaire à temps multifractal [84], qui est obtenu en subordonnant le mouvement de Poisson composé à un mouvement brownien classique. Ce processus constitue un modèle très attractif pour restituer les propriétés multifractales des données expérimentales. Des simulations de Monte-Carlo (non présentées ici) semblables à celles décrites dans la Section 5, de ce mouvement fractionnaire en

temps multifractal, mènent à des résultats totalement cohérents avec les conclusions obtenues pour l'analyse du mouvement de Poisson composé.

De plus, les analyses conduites ici peuvent être naturellement étendues à des quantités multirésolutions autres que les accroissements. Nous avons réalisé des simulations Monte-Carlo sur le mouvement de Poisson composé et le mouvement brownien fractionnaire à temps multifractal en utilisant des accroissements d'ordre p (c'est-à-dire, des accroissements d'accroissements d'accroissements ...) ainsi que des coefficients en ondelettes (calculés à partir d'ondelettes mères avec un nombre varié de moments nuls, voir e.g., [80]). Ces simulations mènent toutes à des conclusions comparables.

Troisième partie
Représentation matricielle

Introduction

Motivation

Dans la partie II, nous avons pu constater qu'on peut observer des ruptures de convergence si on s'intéresse au comportement asymptotique de l'estimateur des moments $S(n, q)$ le long de certaines courbes (n, q) pour des variables indépendantes. De manière similaire, nous avons pu montrer que l'effet de linéarisation de l'analyse multifractale résulte d'une interaction non triviale entre l'échelle d'analyse, la structure de dépendance et la distribution des accroissements. Cependant, le traitement de la corrélation dans cette partie II demeure très empirique du fait de la relative complexité des modèles multifractals considérés. Il est donc tentant de considérer des modèles plus simples pour lesquels il est possible d'obtenir des formes exactes du maximum ou de la somme de variables aléatoires corrélées. Pour ce faire, il est nécessaire de construire un modèle qui préserve autant que possible les structures algébriques qui nous ont permis d'étudier les comportements de la somme et des extrêmes dans le cas indépendant.

Une idée est de s'inspirer d'une classe de solutions exactes développées dans le cadre de la physique quantique et de la physique statistique des systèmes hors d'équilibre qui repose sur le *matrix product ansatz*. Un des exemples les plus connus en physique hors d'équilibre est la classe des modèles ASEP (pour *Asymmetric Simple Exclusion Processes* [41, 81, 22]) qui décrivent la diffusion (asymétrique) de particules sur un réseau unidimensionnel avec exclusion (au plus une particule par site). Ces modèles reposent sur une généralisation de la forme en produit de la densité de probabilité jointe $P(x_1, \dots, x_n)$ sous la forme d'un produit matriciel $P(x_1, \dots, x_n) \propto \prod_{k=1}^n R_d(x_k)$. Cette forme de loi préserve un certain nombre de propriétés structurelles des lois indépendantes grâce à sa structure de produit, tout en introduisant de la corrélation du fait des propriétés de non-commutativité des matrices. Les propriétés statistiques du modèle sont ainsi directement liées aux propriétés algébriques des matrices R_d . Ce lien entre propriétés algébriques et propriétés statistiques a permis d'étudier en détail les propriétés statistiques du modèle ASEP [41, 81] ainsi que bien d'autres modèles statistiques [72, 14, 46], faisant du *matrix product ansatz* un outil important de la physique des systèmes hors d'équilibres.

Étude théorique des systèmes à représentation matricielle

Nous proposons cependant d'opérer un renversement de perspective. Plutôt que d'étudier un système physique particulier à travers le prisme du *matrix product ansatz*, nous nous proposons d'étudier le *matrix product ansatz* afin de déterminer les propriétés statistiques générales des modèles qu'il peut décrire. Une telle étude détaillée des propriétés statistiques de ces modèles à représentation matricielle ouvre

également des perspectives intéressantes en traitement du signal, en donnant un exemple concret de vecteurs aléatoires dont l'auto-corrélation mais également les ordres statistiques supérieurs sont finement contrôlés.

Une approche exhaustive dans le cas général est difficile. Cependant dans les cas des représentations de dimension finie, les propriétés algébriques des matrices R_d sont plus régulières et le chapitre 6 décrit de manière détaillée les propriétés statistiques de ces systèmes à représentation matricielle. Il est intéressant de noter qu'il apparaît deux sous-classes de modèles dont les propriétés statistiques diffèrent fortement. La première classe décrit des modèles, essentiellement stationnaires, dont la corrélation à courte portée est associée à la classe des matrices diagonalisables ; tandis que la seconde sous-classe, intrinsèquement non stationnaire, avec des corrélations à longue portée est associée aux matrices non diagonalisables.

Synthèse

Une fois les propriétés statistiques de ces lois à représentation matricielle dérivées, il est naturel de se demander s'il est possible de générer efficacement des variables aléatoires synthétiques associées à ces lois. Si la représentation sous forme de produit de matrices de la probabilité jointe facilite le calcul des propriétés statistiques de ces lois, elle ne donne cependant guère d'indication sur une méthode de synthèse. Cependant, il est possible de reformuler ces lois à représentation matricielle comme des modèles à chaîne de Markov cachée particuliers, détaillés dans le chapitre 7. Cette reformulation mène à une méthode de synthèse efficace et flexible pour ces lois. Cependant une telle méthode de synthèse n'est pas suffisante pour pouvoir utiliser directement ces lois à représentation matricielle dans des applications en traitement du signal. En effet, il est souvent nécessaire en traitement du signal d'imposer des contraintes sur la distribution marginale, la corrélation ou d'autres ordres statistiques. Ces contraintes ne sont guère naturelles pour les lois à représentation matricielles. Afin de pouvoir utiliser ces lois en pratique, nous proposons des sous-classes particulières de lois à représentation matricielle pour lesquelles nous avons développé des algorithmes de résolution de ces contraintes.

Sommes et extrêmes

Une fois ces problèmes pratiques de synthèse résolus, il est intéressant de revenir à notre questionnement initial : peut-on facilement prévoir le comportement asymptotique du maximum ou de la somme de variables aléatoires à représentation matricielle. Un des intérêts de la structure en produit de matrices est qu'elle préserve une bonne partie de la structure algébrique associée au maximum ou la somme de lois indépendantes. Le chapitre 8 suit donc le raisonnement classique pour obtenir une description des lois limites sous forme d'équations fonctionnelles matricielles. Ces équations sont cependant plus complexes à résoudre que dans le cas scalaire, nous nous proposons donc d'analyser le comportement asymptotique de la somme et du maximum en utilisant la reformulation sous forme de modèle à chaîne de Markov cachée développée dans le chapitre 7. De manière intéressante, on voit alors apparaître des lois limites non standards lorsque la chaîne de Markov cachée présente des transitions irréversibles. En particulier, dans le cas de la somme, on peut montrer que, sous certaines conditions, la loi des grands nombres n'est plus vérifiée

à cause de la présence de corrélations à l'échelle du signal.

Chapitre 6

Caractérisation théorique des propriétés statistiques

1 Probabilité jointe et produit de matrices

1.1 Définition et remarques générales

Inspiré par la physique statistique hors d'équilibre et notamment le modèle ASEP (cf. e.g. [41, 81, 22]), nous proposons un cadre général pour l'analyse et la synthèse de vecteur aléatoire \underline{X}_n de taille n dont la probabilité jointe est prescrite a priori. Notre méthode s'appuie essentiellement sur une représentation matricielle de la probabilité jointe $\mathbb{P}(\underline{X}_n)$ que nous allons maintenant introduire. Cette représentation matricielle repose sur l'alliance de quatre éléments principaux. En premier lieu, pour faciliter notre étude, nous nous limitons (sauf mention contraire) aux matrices de dimension finie d . Le second élément est une matrice de probabilité $\mathcal{P}(x)$ de dimension $d \times d$ telle que pour tout indice (i, j) , $\mathcal{P}_{i,j}(x)$ soit une densité de probabilité classique. Adjoignons à cette matrice de probabilité une matrice de structure \mathcal{E} de dimension $d \times d$, non nulle et de coefficients positifs. La signification de ce nom de matrice de structure apparaîtra plus clairement par la suite. À partir de ces deux matrices complémentaires, nous définissons la matrice de représentation $R_d(x) = \mathcal{E} \circ \mathcal{P}(x)$. Ici, \circ dénote la multiplication coefficients par coefficients. En d'autres termes,

$$\forall(i, j), \quad R_d(x)_{i,j} = \mathcal{E}_{i,j} \mathcal{P}_{i,j}(x). \quad (\text{III.1})$$

De plus, la matrice $\mathcal{P}_{i,j}$ étant une matrice de densité de probabilité, la matrice \mathcal{E} n'est autre que l'intégrale de $R_d(x)$,

$$\int_{-\infty}^{+\infty} R_d(x) dx = \mathcal{E}. \quad (\text{III.2})$$

Une fois la matrice de représentation construite, il ne reste plus qu'à faire le lien entre notre représentation matricielle et une densité de probabilité. Ce lien se fera à travers une forme linéaire \mathcal{L} sur l'espace des matrices. Utilisant la dualité entre forme linéaire et vecteur en dimension finie, il est souvent plus agréable d'utiliser la matrice duale \mathcal{A} (de dimension $d \times d$) de \mathcal{L} définie par

$$\forall M, \mathcal{L}(M) = \text{tr}(\mathcal{A}^T M). \quad (\text{III.3})$$

Pour des raisons qui deviendront apparentes dès le prochain paragraphe, nous nous restreignons aux matrices \mathcal{A} non nulles et à coefficients positifs. Munis de ces définitions, nous nous intéressons aux vecteurs aléatoires \underline{X} dont la probabilité jointe prend la forme :

$$\mathbb{P}_{\underline{X}}(x_1, \dots, x_n) = \frac{\mathcal{L}(\prod_{k=1}^n R_d(x_k))}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.4})$$

Si $\mathcal{L}(\mathcal{E}^n)$ est non nulle, la positivité de \mathcal{E} et \mathcal{A} et l'équation (III.2) assure que l'équation (III.4) définit bien une densité de probabilité.

Nous nommerons par la suite les vecteurs aléatoires dont la probabilité jointe peut s'exprimer ainsi vecteur aléatoire à représentation matricielle. Il est important de remarquer que dans l'expression précédente le produit $\prod_{k=1}^n R_d(x_k) = R_d(x_1) \dots R_d(x_n)$ est un produit orienté ; en effet le produit matriciel n'est pas commutatif contrairement à la multiplication scalaire. Cette non-commutativité du produit matriciel est essentielle parce qu'elle brise la symétrie par permutation de la probabilité jointe, créant par contre-coup une notion de temps pour le vecteur aléatoire \underline{X} . On peut également vérifier que les conditions de positivité exigées sur les matrices \mathcal{A} , \mathcal{E} et \mathcal{P} sont des conditions suffisantes pour que la fonction multivariée définie par l'Eq. (III.4) soit bien une densité de probabilité (i.e. positive et d'intégrale égale à 1). Toutefois, il ne s'agit pas de conditions nécessaires. En effet, la positivité des coefficients des matrices n'est pas une notion stable par changement de base. Or, pour un couple (\mathcal{A}, R_d) donné, il est facile de construire un second couple (\mathcal{A}', R'_d) correspondant à la même probabilité jointe par changement de base B . Il suffit de définir

$$\begin{aligned} \mathcal{A}' &= B^{-1T} \mathcal{A} B^T, \\ R'_d(x) &= B R_d(x) B^{-1}. \end{aligned} \quad (\text{III.5})$$

Il est alors facile de vérifier que $\mathbb{P}_{\underline{X}'} = \mathbb{P}_{\underline{X}}$. Cependant, les nouvelles matrices \mathcal{A}' , \mathcal{E}' et \mathcal{P}' n'ont alors aucune raison de vérifier les conditions de positivité imposées aux matrices originelles.

1.2 Marginale, moments et dépendance

À partir de la définition (III.4), il est possible d'exploiter la structure en produit de la probabilité jointe pour calculer analytiquement nombre de propriétés statistiques utilisant la commutativité de l'intégration et du produit matriciel.

Notamment, les distributions univariées s'expriment simplement comme

$$\begin{aligned} \mathbb{P}_k(X_k = x_k) &= \frac{\int \mathcal{L}(\prod_{i=1}^n R_d(x_i)) \prod_{i \neq k} dx_i}{\mathcal{L}(\mathcal{E}^n)} \\ &= \frac{\mathcal{L}(\int R_d(x_1) dx_1 \dots R_d(x_k) \dots R_d(x_n) dx_n)}{\mathcal{L}(\mathcal{E}^n)} \\ &= \frac{\mathcal{L}(\mathcal{E}^{k-1} R_d(x_k) \mathcal{E}^{n-k})}{\mathcal{L}(\mathcal{E}^n)}. \end{aligned} \quad (\text{III.6})$$

Ces équations impliquent que la probabilité marginale s'exprime forcément comme une moyenne pondérée des distributions $\mathcal{P}_{i,j}$

$$\mathbb{P}_k(X_k = x) = \sum_{i,j} c_{i,j,k} \mathcal{P}_{i,j}(x). \quad (\text{III.7})$$

Les coefficients $c_{i,j,k}$ dépendent des matrices \mathcal{A} et \mathcal{E} et vérifient $\forall k, \sum c_{i,j,k} = 1$. Afin de simplifier l'étude des moments, définissons dès à présent les matrices des moments $M(q)$

$$M(q)_{i,j} = \mathcal{E}_{i,j} \int_{\mathbb{R}} x^q \mathcal{P}_{i,j}(x) dx. \quad (\text{III.8})$$

Avec ces définitions, l'expression des moments au point k est simplement

$$\mathbb{E} [X_k^q] = \frac{\mathcal{L}(\mathcal{E}^{k-1} M(q) \mathcal{E}^{n-k})}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.9})$$

Il est également possible d'obtenir une expression analytique de la probabilité jointe à p variables

$$\mathbb{P}(X_{k_1} = x_{k_1}, \dots, X_{k_p} = x_{k_p}) = \frac{\mathcal{L}(\mathcal{E}^{k_1-1} (\prod_{r=1}^{p-1} R_d(x_{k_r}) \mathcal{E}^{k_{r+1}-k_r-1}) R_d(x_{k_p}) \mathcal{E}^{n-k_p})}{\mathcal{L}(\mathcal{E}^n)}, \quad (\text{III.10})$$

ce qui mène immédiatement à l'expression des moments à p points (avec q_r l'ordre associé au point k_r)

$$\mathbb{E} \left[\prod_{r=1}^p X_{k_r}^{q_r} \right] = \frac{\mathcal{L}(\mathcal{E}^{k_1-1} (\prod_{r=1}^{p-1} M(q_r) \mathcal{E}^{k_{r+1}-k_r-1}) M(q_p) \mathcal{E}^{n-k_p})}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.11})$$

L'Eq. (III.11) est un résultat clef puisqu'il montre que les moments du vecteur \underline{X} ne dépendent que des matrices $M(q)$ et du couple $(\mathcal{A}, \mathcal{E})$. Les moments de \underline{X} peuvent donc être directement déduits à partir des moments des $\mathcal{P}_{i,j}$, de la matrice de structure \mathcal{E} et de la forme linéaire \mathcal{L} . Cette sobriété ouvre des perspectives intéressantes sur lequel nous reviendrons en section 2.

1.3 Condition de stationnarité

Le vecteur \underline{X} ne forme pas en général une série stationnaire. Les équations (III.6) et (III.9) montrent clairement que ce caractère non stationnaire émerge de la non-commutativité du produit matriciel. Cependant, sous certaines conditions, il est possible d'obtenir un vecteur stationnaire. Une condition suffisante particulièrement intéressante est d'imposer la commutativité de la matrice \mathcal{A}^T et \mathcal{E} .

$$[\mathcal{A}^T, \mathcal{E}] \equiv \mathcal{A}^T \mathcal{E} - \mathcal{E} \mathcal{A}^T = 0. \quad (\text{III.12})$$

Cette condition entraîne que

$$\forall x, \mathcal{L}(\mathcal{E} R_d(x)) = \mathcal{L}(R_d(x) \mathcal{E}). \quad (\text{III.13})$$

Cette propriété permet de simplifier l'expression de la distribution marginale en

$$\mathbb{P}_S(X = x) = \frac{\mathcal{L}(R_d(x) \mathcal{E}^{n-1})}{\mathcal{L}(\mathcal{E}^n)} = \sum_{i,j} c_{i,j} \mathcal{P}_{i,j}(x). \quad (\text{III.14})$$

Cette expression ne dépend plus de k , toutes les composantes de X partagent donc la même distribution marginale. De manière similaire, l'expression de la probabilité jointe et des moments à p points devient

$$\mathbb{P}(X_{k_1} = x_{k_1}, \dots, X_{k_p} = x_{k_p}) = \frac{\mathcal{L} \left(\left(\prod_{r=1}^{p-1} R_d(x_{k_r}) \mathcal{E}^{k_{r+1}-k_r-1} \right) R_d(x_{k_p}) \mathcal{E}^{n-(k_p-k_1)-1} \right)}{\mathcal{L}(\mathcal{E}^n)}, \quad (\text{III.15})$$

$$\mathbb{E} \left[\prod_{r=1}^p X_{k_r}^{q_r} \right] = \frac{\mathcal{L} \left(\left(\prod_{r=1}^{p-1} M(q_r) \mathcal{E}^{k_{r+1}-k_r-1} \right) M(q_p) \mathcal{E}^{n-(k_p-k_1)-1} \right)}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.16})$$

Ces relations indiquent clairement que le vecteur \underline{X} peut être considéré comme une série temporelle stationnaire : toutes les statistiques à p points ne dépendent plus que des différences de temps $k_{r+1} - k_r$.

Les seules matrices \mathcal{A} vérifiant l'Eq. (III.12), ce quelle que soit la matrice \mathcal{E} , sont les matrices multiples de la matrice identité. Dans ce cas, \mathcal{L} est exactement un multiple (scalaire) de la trace. Cependant, l'opérateur trace induit une structure de dépendance circulaire

$$l \geq 2k, \quad \mathbb{E}[X_k X_l] = \mathbb{E}[X_k X_{n+2k-l}], \quad (\text{III.17})$$

ce qui peut être indésirable pour certaines applications. Afin d'éviter cet écueil tout en respectant la condition de stationnarité (III.12), il suffit de choisir des paires spécifiques de matrice $(\mathcal{A}, \mathcal{E})$. Un exemple particulièrement intéressant apparaît lorsqu'on considère une matrice \mathcal{A} définie par

$$\mathcal{A}_{i,j} = \frac{1}{d}. \quad (\text{III.18})$$

Dans ce cas l'ensemble des matrices commutant avec \mathcal{A} n'est autre que l'ensemble des matrices doublement stochastiques (et leurs multiples) :

$$\forall i, j, \quad \sum_k \mathcal{E}_{k,j} = \sum_k \mathcal{E}_{i,k} \in \mathbb{R}. \quad (\text{III.19})$$

Avec cette paire $(\mathcal{A}, \mathcal{E})$, il est facile de vérifier que pour toute matrice M

$$\forall k \in \mathbb{N}, \quad \mathcal{L}(M\mathcal{E}^k) = \mathcal{L}(M), \quad (\text{III.20})$$

ce qui permet de simplifier fortement les équations (III.14) à (III.16). Les distributions marginales et partielles ne dépendent plus de la taille du signal n . De plus, la distribution marginale ne dépend plus de \mathcal{E} :

$$\mathbb{P}_k(X_k = x_k) = \mathcal{L}(R_d(x_k)), \quad (\text{III.21})$$

$$\mathbb{P}(X_{k_1} = x_{k_1}, \dots, X_{k_p} = x_{k_p}) = \mathcal{L} \left(\left(\prod_{r=1}^{p-1} R_d(x_{k_r}) \mathcal{E}^{k_{r+1}-k_r-1} \right) R_d(x_{k_p}) \right), \quad (\text{III.22})$$

$$\mathbb{E} \left[\prod_{r=1}^p X_{k_r}^{q_r} \right] = \mathcal{L} \left(\left(\prod_{r=1}^{p-1} M(q_r) \mathcal{E}^{k_{r+1}-k_r-1} \right) M(q_p) \right). \quad (\text{III.23})$$

Ce choix de \mathcal{A} et de \mathcal{E} est donc particulièrement intéressant pour l'élaboration de séries temporelles stationnaires avec des moments prescrits a priori. Nous en reparlerons en section 7.

2 Structure de dépendance

Les équations (III.10) et (III.11) indiquent que la matrice \mathcal{E} joue un rôle primordial dans la structure de dépendance du vecteur \underline{X} tandis que les matrices des moments $M(q)$ ne contribuent à l'expression des moments qu'en tant que constantes, certes matricielles. Plus précisément, la structure de dépendance de \underline{X} va dépendre finement du comportement des puissances \mathcal{E}^k de cette matrice de structure \mathcal{E} . Suivant le caractère diagonalisable ou non de la matrice \mathcal{E} , cette structure de dépendance sera radicalement différente. Une étude précise des formes potentielles de \mathcal{E}^k et leurs implications sur la corrélation s'impose donc.

2.1 Matrice de structure diagonalisable

Commençons donc par le cas le plus simple et le plus familier et supposons que \mathcal{E} est une matrice diagonalisable de valeurs propres distinctes $\lambda_1, \dots, \lambda_r$ (avec $r \leq d$). Il est alors possible de décomposer \mathcal{E} en r matrices E_1, \dots, E_r telles que :

$$\mathcal{E}^k = \sum_{i=1}^r \lambda_i^k E_i, \quad (\text{III.24})$$

ce qui permet en utilisant la linéarité de \mathcal{L} de réécrire l'Eq. (III.6)

$$\mathbb{P}_k(X_k = x) = \frac{\sum_{i,j \in \{1, \dots, r\}^2} \lambda_i^{k-1} \lambda_j^{n-k} \mathcal{L}(E_i R_d(x) E_j)}{\sum_{i=1}^r \lambda_i^n \mathcal{L}(E_i)}, \quad (\text{III.25})$$

faisant apparaître explicitement la dépendance en k de la distribution marginale du k -ème coefficient. Cette forme fait d'ailleurs écho avec l'Eq. (III.7) étant donné que le terme $\mathcal{L}(E_i R_d(x) E_j)$ correspond à une combinaison linéaire des $\mathcal{P}_{i,j}$. De même, pour les statistiques à deux points (i.e. la corrélation), l'équation générale (III.11) se simplifie en :

$$\mathbb{E}[X_k X_l] = \frac{\sum_{i,j,m \in \{1, \dots, r\}^3} \lambda_i^{k-1} \lambda_j^{n-l} \lambda_m^{|l-k|-1} \mathcal{L}(E_i M(1) E_m M(1) E_j)}{\sum_{i=1}^r \lambda_i^n \mathcal{L}(E_i)}. \quad (\text{III.26})$$

Dans le cas stationnaire, cette expression peut être simplifiée davantage et devient

$$\mathbb{E}[X_k X_l] = \frac{\sum_{j,m} \lambda_j^{n-2} \left(\frac{\lambda_m}{\lambda_j} \right)^{|l-k|-1} \mathcal{L}(M(1) E_m M(1) E_j)}{\sum_{i=1}^r \lambda_i^n \mathcal{L}(E_i)}. \quad (\text{III.27})$$

Cette dernière équation dévoile une structure en somme pondérée d'exponentielles $\exp(-(|l-k|-1)(\ln \lambda_j - \ln \lambda_m))$ pour la fonction de covariance. Si nous nommons r_M le nombre de modules distincts parmi les valeurs propres λ_k , le nombre d'échelles

de temps $\tau_{j,m} = (\ln |\lambda_j| - \ln |\lambda_m|)^{-1}$ distinctes est au plus $r_M(r_M - 1)/2 = \binom{r_M}{2}$. La dimension d de la matrice \mathcal{E} impose donc une limite sur la complexité potentielle de la fonction de corrélation.

Ce comportement en somme d'exponentielles décroissantes de la corrélation est toujours valide dans le cas non stationnaire même si les équations ne sont pas aussi claires. Cependant, si une des valeurs propres domine en module les autres valeurs propres, il est possible de faire émerger un comportement asymptotique simple même dans le cas général. Supposons donc que $|\lambda_1| > |\lambda_k|$ et que $\mathcal{L}(E_1) \neq 0$. Le terme de normalisation est alors asymptotiquement dominé par

$$\mathcal{L}(\mathcal{E}^n) \sim \lambda_1^n \mathcal{L}(E_1), \quad n \rightarrow +\infty. \quad (\text{III.28})$$

En injectant cette forme asymptotique dans l'Eq. (III.6), on obtient une forme simplifiée de la distribution marginale

$$\mathbb{P}_k(X_k = x) = \sum_{i,j} \lambda_i^{k-1} \lambda_j^{n-k} \frac{\mathcal{L}(E_i R_d(x) E_j)}{\mathcal{L}(\mathcal{E}^n)} \sim \frac{\mathcal{L}(E_1 R_d(x) E_1)}{\lambda_1 \mathcal{L}(E_1)}. \quad (\text{III.29})$$

De même, Eq. (III.26) devient

$$\mathbb{E}[X_k X_l] \sim \sum_i \left(\frac{\lambda_i}{\lambda_1} \right)^{|l-k|-1} \frac{\mathcal{L}(E_1 M(1) E_i M(1) E_1)}{\mathcal{L}(E_1) \lambda_1^2}. \quad (\text{III.30})$$

Ces équations montrent que dans le cas d'une matrice \mathcal{E} diagonalisable avec une valeur propre dominante, chaque composante du vecteur \underline{X} partage asymptotiquement la même distribution marginale. De plus, la fonction d'autocovariance aux points k et l s'écrit comme une somme d'exponentielles ne dépendant plus que de la distance $|k - l|$. Dans la limite $n \rightarrow +\infty$, le vecteur \underline{X} est donc asymptotiquement quasi-stationnaire.

2.2 Matrice de structure non diagonalisable

Dans le cas non diagonalisable, le panorama est radicalement différent. Les corrélations exponentielles s'effacent pour laisser la place à des structures de corrélations bien plus exotiques. Pour aborder en douceur ces cas exotiques, contentons-nous tout d'abord d'une matrice non diagonalisable classique $\mathcal{E} = \text{Id} + H$, où Id est la matrice identité et H est une matrice nilpotente d'ordre $p + 1$ (i.e. $H^{p+1} = 0$ et $H^p \neq 0$). Les puissances de \mathcal{E} sont alors

$$\mathcal{E}^k = \sum_{j=0}^p \binom{k}{j} H^j, \quad (\text{III.31})$$

ce qui permet de réécrire l'Eq. (III.6) en

$$\mathbb{P}_k(X_k = x) = \frac{\sum_{i,j \leq p} \binom{k-1}{i} \binom{n-k}{j} \mathcal{L}(H^i R_d(x) H^j)}{\sum_{i=1}^p \binom{n}{i} \mathcal{L}(H^i)}. \quad (\text{III.32})$$

De façon similaire, la fonction de covariance devient :

$$\mathbb{E}[X_k X_l] = \frac{\sum_{i,j,m \leq p} \binom{k-1}{i} \binom{|l-k|-1}{m} \binom{n-l}{j} \mathcal{L}(H^i M(1) H^m M(1) H^j)}{\sum_{i=1}^p \binom{n}{i} \mathcal{L}(H^i)}. \quad (\text{III.33})$$

Afin de mieux comprendre ces équations, il est intéressant d'étudier leurs comportements asymptotiques. Dans la limite $n \rightarrow +\infty$, on sait que

$$\binom{n}{p} \sim \frac{n^p}{p!}, \quad n \rightarrow +\infty. \quad (\text{III.34})$$

Si $\mathcal{L}(H^p) \neq 0$, cela implique que

$$\begin{aligned} \mathbb{P}_k(X_k = x) &= \sum_{i,j \leq p} \binom{k-1}{i} \binom{n-k}{j} \frac{\mathcal{L}(H^i R_d(x) H^j)}{\mathcal{L}(\mathcal{E}^n)} \\ &\sim \sum_{i+j=p} \binom{p}{i} \left(\frac{k}{n}\right)^i \left(1 - \frac{k}{n}\right)^j \frac{\mathcal{L}(H^j R_d(x) H^i)}{\mathcal{L}(H^p)}. \end{aligned} \quad (\text{III.35})$$

Comparé avec l'Eq. (III.25), ce résultat montre que la distribution marginale de X_k est un mélange de distributions dont les poids dépendent de la position relative k/n plutôt que de la position absolue k comme dans le cas \mathcal{E} diagonalisable.

La fonction de covariance diffère fortement, elle aussi, du cas diagonalisable

$$\begin{aligned} \mathbb{E}[X_k X_l] &\approx \sum_{i+j+m=p} \\ &\left[\frac{p!}{i!m!j!} \left(\frac{k}{n}\right)^i \left(\frac{|l-k|}{n}\right)^j \left(1 - \frac{l}{n}\right)^m \frac{\mathcal{L}(H^i M(1) H^j M(1) H^m)}{\mathcal{L}(H^p)} \right]. \end{aligned} \quad (\text{III.36})$$

La présence des termes algébriques $\left(\frac{|l-k|}{n}\right)^j$ indique l'existence de corrélations à longue portée qui s'étendent sur l'ensemble du vecteur \underline{X} .

À partir de ce choix particulier de matrice non diagonalisable \mathcal{E} , il est possible d'esquisser le comportement du cas général. Pour une matrice \mathcal{E} quelconque, la décomposition de Jordan de \mathcal{E} en blocs diagonaux permet de se ramener à une combinaison des deux cas précédents. Et on peut voir que dans le cas le plus général, la fonction de covariance peut être décomposée comme une somme de produits d'exponentielles décroissantes et de fonctions algébriques.

Chapitre 7

Application pratique à la synthèse de signaux

1 Motivation

Le traitement du signal moderne requiert très souvent de générer numériquement des données synthétiques de façon rapide et efficace tout en garantissant certaines de leurs propriétés statistiques (distributions marginales, covariance, spectre, distribution multivariée, . . .) prescrites a priori et, ce, avec un degré de fiabilité contrôlé. C'est par exemple le cas lorsqu'on souhaite mesurer et vérifier les performances de nouvelles analyses statistiques. Le calcul théorique de ces performances peut, dans certains cas, s'avérer trop complexe pour être mené à bien, et ce particulièrement lorsqu'on souhaite utiliser ces outils sur des données réelles, dont les propriétés ne sont pas parfaitement connues. Une alternative dans ces situations est alors de recourir à des simulations de Monte-Carlo. Les performances sont alors mesurées en moyennant les performances obtenues sur des réalisations indépendantes de données synthétiques conçues pour être aussi proches que possible des données réelles. Un autre exemple est fourni par les méthodes Bayésiennes d'estimation, qui nécessitent généralement l'usage de méthodes de Monte Carlo par chaînes de Markov (MCMC) [15] ou des schémas de résolution basés sur des méthodes variationnelles (cf. e.g., [106, 21]) pour générer numériquement des copies indépendantes d'hyperparamètres avec des distributions (potentiellement complexes) dérivées du formalisme Bayésien.

Dans le cas de synthèse numérique de séries temporelles gaussiennes, avec une covariance imposée, la méthode de synthèse par *Circulant Embedded Matrix* (cf. [42, 39, 35]) est considérée de nos jours comme l'état de l'art et est largement utilisée. Cette méthode a été étendue pour la synthèse de séries temporelles gaussiennes multivariées, avec des auto-covariances et des covariances croisées imposées, notamment dans [109, 32] (mais aussi [65] comme variation).

La synthèse de séries temporelles non gaussiennes s'est révélée être bien plus complexe. La prescription de la distribution jointe complète s'avère difficile à réaliser tandis que la covariance et les distributions marginales ne suffisent plus à définir complètement le processus. De nombreuses approches ont été proposées pour résoudre ces problèmes (e.g., [58, 66] et les références qui y sont présentes pour une revue du sujet). D'autres stratégies reposent sur des techniques par substituts (*surrogates*) : en partant de données réelles (expérimentales) originales issues du problème

analysé, des copies de substitution sont obtenues en redistribuant aléatoirement un des attributs des données (par exemple la phase de sa transformée de Fourier) tout en maintenant les autres fixés (l'amplitude de sa transformée de Fourier) (cf. e.g., [103]). Récemment, une méthode par *transport optimal* a été proposée, tentant de modifier itérativement la distribution jointe des vecteurs aléatoires pour atteindre une cible donnée a priori [24]. Le cadre théorique des simulations par chaînes de Markov offre une autre alternative et une vaste classe de solutions, se focalisant sur les propriétés de la dynamique locale sans emphase explicite sur la distribution jointe du processus. Les procédures par chaînes de Markov sont aussi largement utilisées pour la synthèse de réalisations indépendantes de vecteurs aléatoires avec des propriétés imposées (cf. e.g., [33] pour une revue).

Dans la suite, nous proposons d'utiliser nos modèles de variables aléatoires à représentation matricielle pour mettre au point une méthode alternative de synthèse de signaux. L'avantage d'utiliser des variables aléatoires à représentation matricielle réside dans le fait que nous connaissons exactement la densité de probabilité jointe. De plus, comme nous l'avons détaillé dans la chapitre 6, il est relativement aisé de déduire les moments et autres statistiques à k points à partir de ces représentations matricielles. Il est donc tentant d'utiliser cette classe de variables aléatoires pour construire des signaux synthétiques aux propriétés très contrôlées, notamment afin de pouvoir tester la robustesse de certains estimateurs vis-à-vis des statistiques d'ordres élevés.

Dans la section 2, nous commencerons par réinterpréter le modèle théorique par représentation matricielle comme un cas particulier de modèle à chaîne de Markov cachée, ce qui fait émerger une procédure de synthèse numérique rapide et efficace; nous explicitons ensuite comment atteindre des distributions marginales et autres propriétés statistiques ciblées. Enfin, la section 4 présente deux ensembles différents d'exemples pédagogiques ayant pour but d'illustrer l'efficacité et la versatilité de notre méthode. En premier vient un jeu de séries temporelles dont les distributions marginales, fonction de covariances et statistiques d'ordre 4 ont été imposées a priori. En second, un jeu de vecteurs aléatoires, avec des marginales différentes pour chaque composante, est synthétisé et nous montrons comment il est possible de régler la corrélation et la statistique d'ordre 4 au sein des composantes des vecteurs sans altérer la distribution marginale desdits vecteurs.

2 Modèle par chaîne de Markov cachée

Si la définition de \underline{X} à partir de sa distribution jointe $\mathbb{P}(\underline{X})$ (i.e. l'Eq. (III.4)) est élégante et pratique pour le calcul des propriétés statistiques, elle n'offre malheureusement pas de moyen clair pour construire et synthétiser efficacement les variables aléatoires associées. Nous allons montrer que ce modèle peut être reformulé comme un modèle à chaîne de Markov cachée afin de pallier ce défaut, gagnant par là-même un outil très intéressant pour la compréhension des propriétés statistiques de \underline{X} . La première étape de cette reformulation consiste malheureusement à développer le produit matriciel apparaissant dans l'Eq. (III.4). Étendre la formule classique $(ABC)_{i,j} = \sum_{k,l} a_{i,k} b_{k,l} c_{l,j}$ au produit de n matrices, nous donne

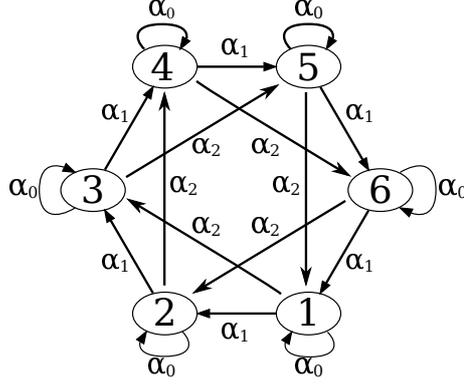


FIGURE 7.1 – **Graphe de transition.** Exemple pour $d = 6$ et $\mathcal{E} = \alpha_0 \text{Id} + \alpha_1 J_d + \alpha_2 J_d^2$ ($\alpha_0 + \alpha_1 + \alpha_2 = 1$) où J_d est une matrice de saut circulaire (cf. Eq. (III.53))

$$\mathbb{P}(x_1, \dots, x_n) = \sum_{\underline{\Gamma}} \kappa(\underline{\Gamma}) \prod_{k=1}^n \mathcal{P}(x_k)_{\Gamma_{k-1}, \Gamma_k}, \quad (\text{III.37})$$

avec $\underline{\Gamma} \equiv \{\Gamma_0, \dots, \Gamma_k, \dots, \Gamma_n\} \in [1, \dots, d]^{n+1}$ et

$$\kappa(\underline{\Gamma}) = \frac{\mathcal{A}_{\Gamma_0, \Gamma_n}}{\mathcal{L}(\mathcal{E}^n)} \prod_{k=1}^n \mathcal{E}_{\Gamma_{k-1}, \Gamma_k}, \quad (\text{III.38})$$

où $\sum_{\underline{\Gamma}} \kappa(\underline{\Gamma}) = 1$. Par conséquent, $\kappa(\underline{\Gamma})$ et $\mathbb{P}(\underline{X})$ peuvent être interprétés respectivement comme la fonction de probabilité de $\underline{\Gamma}$ et comme un mélange pondéré par $\kappa(\underline{\Gamma})$ de lois dont la densité est définie par le produit $\prod_{k=1}^n \mathcal{P}_{\Gamma_{k-1}, \Gamma_k}(x_k)$.

Pour la seconde étape, appelons $\Gamma_{\geq t}$ l'ensemble des chaînes commençant à l'indice t et s'arrêtant à l'indice $n-1$: $\Gamma_{\geq t} \equiv (\Gamma_t, \dots, \Gamma_{n-1})$. Pour une paire donnée (Γ_0, Γ_n) , l'Eq. (III.38) implique que :

$$\begin{aligned} \mathbb{P}(\Gamma_k = j | \Gamma_0 = \gamma_0, \dots, \Gamma_{k-1} = \gamma_{k-1}) &= \\ &= \frac{\mathcal{E}_{\gamma_0, \gamma_1} \dots \mathcal{E}_{\gamma_{k-1}, j} \sum_{\Gamma_{\geq k+1}} \mathcal{E}_{j, \Gamma_{k+1}} \dots \mathcal{E}_{\Gamma_{n-1}, \Gamma_n}}{\mathcal{E}_{\gamma_0, \gamma_1} \dots \mathcal{E}_{\gamma_{k-2}, \gamma_{k-1}} \sum_{\Gamma_{\geq k}} \mathcal{E}_{\gamma_{k-1}, \Gamma_k} \dots \mathcal{E}_{\Gamma_{n-1}, \Gamma_n}} \\ &= \mathcal{E}_{\gamma_{k-1}, j} \frac{\binom{\mathcal{E}^{n-k}}{j, \Gamma_n}}{\binom{\mathcal{E}^{n-k+1}}{\gamma_{k-1}, \Gamma_n}} = \mathbb{P}(\Gamma_k = j | \Gamma_{k-1} = \gamma_{k-1}). \end{aligned} \quad (\text{III.39})$$

Ainsi $\underline{\Gamma}$ est une chaîne de Markov inhomogène à d états avec une matrice de transition à l'étape k définie par :

$$\mathbb{P}(\Gamma_k = j | \Gamma_{k-1} = i) = \mathcal{E}_{i,j} \frac{\binom{\mathcal{E}^{n-k}}{j, \Gamma_n}}{\binom{\mathcal{E}^{n-k+1}}{i, \Gamma_n}}. \quad (\text{III.40})$$

En d'autres termes, la matrice \mathcal{E} peut être interprétée comme la base de la matrice de transition de la chaîne de Markov sous-jacente $\underline{\Gamma}$, comme l'illustre la Fig. 7.1.

De plus, l'Eq. (III.38) nous permet de calculer explicitement la distribution de l'état initial (Γ_0, Γ_n) de la chaîne :

$$\mathbb{P}(\Gamma_0 = i, \Gamma_n = j) = \frac{\mathcal{A}_{i,j}(\mathcal{E}^n)_{i,j}}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.41})$$

Alliant ces résultats, il est possible de synthétiser numériquement le vecteur \underline{X} en utilisant un modèle de chaîne de Markov cachée avec quatre variables cachées : l'état courant Γ_k , l'état précédent Γ_{k-1} et la paire formée par les états initial et final (Γ_0, Γ_n) . En combinant les Eqs. (III.40) et (III.41), l'algorithme de synthèse peut être esquissé en 2 étapes

Étape 1 : Initialisation :

Utiliser l'Eq. (III.41) pour générer les états Γ_0 et Γ_n .

Étape 2 : Itération pour $k = 1, \dots, n$:

Étape 2.1 : Choisir aléatoirement un état Γ_k , selon les probabilités de transition données par l'Eq. (III.40) ;

Étape 2.2 : Générer X_k selon $\mathcal{P}_{\Gamma_{k-1}, \Gamma_k}$.

Si $\mathcal{A}_{i,j} = 1/d$ (cf. l'Eq. (III.18)) et \mathcal{E} est doublement stochastique (cf. l'Eq. (III.19)), il peut être montré qu'il n'y a nul besoin de fixer a priori l'état final Γ_n .

Si on réécrit dans ces conditions, l'équation (III.40), on obtient

$$\mathbb{P}(\Gamma_k = j | \Gamma_{k-1} = i) = \mathcal{E}_{i,j} \frac{\sum_{f=1}^d (\mathcal{E}^{n-k})_{j,f}}{\sum_{f=1}^d (\mathcal{E}^{n-k+1})_{i,f}}. \quad (\text{III.42})$$

Or, si la matrice \mathcal{E} est doublement stochastique, \mathcal{E}^{n-k} et \mathcal{E}^{n-k+1} le sont également et par conséquent

$$\mathbb{P}(\Gamma_k = j | \Gamma_{k-1} = i) = \mathcal{E}_{i,j}. \quad (\text{III.43})$$

La chaîne de Markov devient alors homogène et les Eqs. (III.40) et (III.41) peuvent être réécrites en :

$$\mathbb{P}(\Gamma_k = j | \Gamma_{k-1} = i) = \mathcal{E}_{i,j}, \quad (\text{III.44})$$

$$\mathbb{P}(\Gamma_0 = i) = \frac{1}{d}. \quad (\text{III.45})$$

De telles simplifications impliquent que \mathcal{E} est exactement la matrice de transition de la chaîne de Markov Γ . De plus, l'état initial de cette chaîne de Markov Γ_0 est distribué uniformément.

En outre, il est intéressant d'évaluer le coût algorithmique de cette procédure. Dans le cas général où \mathcal{E} n'est pas inversible, la complexité algorithmique est dominée par le coût associé au calcul des puissances de matrices \mathcal{E}^k , ce qui mène à une complexité totale en $\mathcal{O}(d^3 n \ln n)$ (avec un algorithme de multiplication matricielle en $\mathcal{O}(d^3)$ et un algorithme de calcul de puissance en $\mathcal{O}(\ln n)$). Si \mathcal{E} est inversible, le coût global est ramené à $\mathcal{O}(d^3 n)$ (puisque une seule multiplication matricielle est requise à chaque étape). Avec le choix de matrices \mathcal{E} doublement stochastiques et $\mathcal{A}_{i,j} = 1/d$, le coût algorithmique n'est plus que $\mathcal{O}(n \ln d)$. Par conséquent, notre procédure de synthèse s'adapte efficacement non seulement pour des signaux de grande taille (n grand) mais aussi pour des structures de dépendances complexes (grande dimension d). Il est d'ailleurs potentiellement aisé d'optimiser plus avant l'algorithme pour des cas spécifiques, par exemple pour des matrices de structures \mathcal{E} creuses.

3 Conception pratique

Dans la section précédente, nous avons mis en évidence un algorithme de synthèse pour les vecteurs aléatoires dont la probabilité jointe s'exprime sous la forme matricielle décrite par l'Eq. (III.4). Cependant s'il est intéressant théoriquement de prescrire exactement la densité de probabilité jointe, il s'agit d'une contrainte bien trop forte et artificielle pour des applications pratiques. Il est plus naturel de décomposer cette contrainte en deux sous-contraintes plus simples et plus amènes en fixant a priori les lois marginales du vecteur \underline{X} en premier lieu et en second lieu la structure de dépendance (corrélation et autres statistiques d'ordres plus élevés). Les sections 1 et 2 montrent qu'il ne peut exister au plus $\binom{r_M}{2}$ échelles de temps distinctes au sein des fonctions de corrélations (avec r_M le nombre de valeurs propres de module distincts). Le choix de la dimension d limite donc la complexité de la structure de dépendance. La forme précise de cette structure de dépendance est déterminée par le choix conjoint de \mathcal{E} et \mathcal{A} (i.e., stationnarité, covariance). La section 4 explicite ce lien dans un cas particulier intéressant et propose une méthode pour choisir \mathcal{E} et \mathcal{A} en fonction de la structure de dépendance souhaitée.

Pour la suite, nous nous intéressons à la question du choix de la fonction matricielle $\mathcal{P}(x)$ pour \mathcal{E} et \mathcal{A} fixés. Par souci de simplicité, nous discuterons tout d'abord du cas stationnaire avant de considérer les extensions possibles pour les cas non-stationnaires. Nous avons donc deux contraintes à satisfaire : une contrainte sur la distribution marginale du vecteur \underline{X} mais aussi une contrainte sur la fonction de dépendance. La contrainte sur la dépendance peut être décomposée en deux sous-parties. Les matrices \mathcal{E} et \mathcal{A} déterminent la forme générale de la dépendance, c'est-à-dire le nombre de longueurs de corrélations potentielles distinctes et les périodes associées. Les détails spécifiques (les coefficients) dépendent des matrices de moments $M(q)$ et par conséquent de la matrice de probabilité \mathcal{P} . Il est vital de noter que cela implique que les contraintes sur les fonctions de covariances et la distribution marginale ne sont pas indépendantes : le choix de \mathcal{P} fixe simultanément la distribution marginale (comme indiqué par l'Eq. (III.7)) et les coefficients gouvernant les amplitudes relatives des structures de dépendance, à travers les matrices $M(q)$ (cf. l'Eq. (III.11)). De plus, la distribution marginale impose des contraintes non triviales sur l'ensemble des matrices $M(q)$ accessibles. Par exemple, l'Eq. (III.14) implique que :

$$M(1)_{i,j} < \mathcal{E}_{i,j} \mathbb{E}[X|X > x], \quad x \equiv F_S^{-1}(1 - c_{i,j}), \quad (\text{III.46})$$

avec X généré selon \mathbb{P}_S (comme défini par l'Eq. (III.14)), et F_S la fonction cumulative associée à la distribution marginale \mathbb{P}_S . La difficulté réside donc dans la mise au point d'une approche numérique qui permette d'obtenir facilement des lois à marginales et matrice des moments fixées de façon conjointe. Afin d'essayer de séparer ces deux contraintes, nous proposons de construire la fonction matricielle $\mathcal{P}(x)$ à partir d'une procédure de normalisation d'une matrice de fonctions réelles ("strictement") positives $g_{i,j}(x)$

$$c_{i,j} \mathcal{P}_{i,j}(x) = \frac{g_{i,j}(x)}{\sum_{l,m} g_{l,m}(x)} \mathbb{P}_S(x). \quad (\text{III.47})$$

Avec cette paramétrisation la contrainte (III.14) est automatiquement vérifiée. Cependant, afin d'assurer que tous les $\mathcal{P}_{i,j}$ soient bien des distributions, les $g_{i,j}$ doivent

satisfaire l'équation suivante :

$$\int \frac{g_{i,j}(x)}{\sum_{l,m} g_{l,m}(x)} \mathbb{P}_S(x) dx = c_{i,j}. \quad (\text{III.48})$$

En factorisant $g_{i,j}(x)$ en $g_{i,j}(x) = \mu_{i,j} h_{i,j}(x)$, il est possible d'utiliser le degré de liberté fourni par $\underline{\mu} \equiv \{\mu_{i,j}\}$ pour résoudre l'équation précédente pour un \underline{h} donné :

$$\int \frac{\mu[\underline{h}]_{i,j} h_{i,j}(x)}{\sum_{l,m} \mu[\underline{h}]_{l,m} h_{l,m}(x)} \mathbb{P}_S(x) dx = c_{i,j}. \quad (\text{III.49})$$

En injectant la solution $\mu[\underline{h}]$ de l'Eq. (III.49) dans l'Eq. (III.47) on obtient :

$$\mathcal{P}[\underline{h}]_{i,j}(x) = \frac{\mu[\underline{h}]_{i,j} h_{i,j}(x)}{c_{i,j} \sum_{l,m} \mu[\underline{h}]_{l,m} h_{l,m}(x)} \mathbb{P}_S(x). \quad (\text{III.50})$$

La dernière étape consiste alors à trouver un h tel que

$$M(q)_{i,j} = \mathcal{E}_{i,j} \int x^q \mathcal{P}[\underline{h}]_{i,j}(x) dx \quad (\text{III.51})$$

où les matrices $M(q)$ sont les matrices des moments fixées a priori. Cette paramétrisation de $\mathcal{P}(x)$ par la matrice de fonctions \underline{h} permet essentiellement de réinterpréter l'équation (III.8) comme une équation non-linéaire en \underline{h} . Afin de résoudre l'Eq. (III.51), nous limitons en pratique les fonctions h à une famille de noyaux paramétrés $K_p : h_{i,j}(x) = K_{p_{i,j}}(x)$. Un choix naturel pour les paramètres du noyau serait d'utiliser des paramètres vectoriels avec autant de degré de liberté que le nombre d'ordres de moments fixés a priori. Par exemple, lorsque la marginale souhaitée est une loi normale et que $M(1)$ et $M(2)$ sont fixées a priori, il est naturel (mais non obligatoire) de sélectionner $K_{m,\sigma} = \mathcal{N}_{m,\sigma}$. D'autres exemples seront donnés en section 4. Une fois ces bases théoriques posées, les équations (III.49) à (III.51) peuvent être résolues numériquement, étape par étape. Des exemples de résultats obtenues par cette méthode seront détaillés en section 4 et illustrés par la Fig. 7.4.

Pour résumer, on peut dérouler la construction d'un vecteur aléatoire \underline{X}_n , avec une marginale et une structure de dépendance fixée, en quatre étapes

1. Choisir d , \mathcal{E} et \mathcal{A} en accord avec la structure de dépendance souhaitée.
2. Sélectionner les matrices des moments $M(q)$ pour obtenir les bons coefficients de dépendance.
3. Définir une famille de noyaux K_p .
4. Résoudre les Eqs. (III.49) à (III.51) afin de construire la matrice de distributions \mathcal{P} en accord avec la distribution marginale \mathbb{P}_k .

Une implémentation en MATLAB de cette procédure a été réalisée. Les sources sont disponibles sur demande. Le potentiel pratique de cette méthode est exploré plus avant dans la section 4 où des exemples simples (et pédagogiques) sont présentés et détaillés.

Dans le cas général non-stationnaire, l'algorithme décrit plus haut ne peut être utilisé à cause du trop grand nombre de contraintes qui apparaissent. Cependant, si on restreint le cas non-stationnaire au cas des matrices \mathcal{E} diagonales supérieures par bloc décrit en section 4.2 (qui est suffisamment riche pour décrire tous les cas

utilisables en pratique), il est alors aisé d'étendre l'algorithme en l'appliquant de façon indépendante pour chacune des composantes du vecteur (non-stationnaire) \underline{X} . Comme illustré dans la section 4.2, l'alliance de cette structure de matrice en blocs et de l'algorithme décrit précédemment donne naissance à un outil général et efficace pour la synthèse numérique de vecteurs aléatoires.

4 Illustrations

Dans cette section, nous allons tenter d'illustrer tout le potentiel de la construction générale décrite plus haut dans deux contextes spécifiques. Tout d'abord, nous revisiterons le cas des vecteurs stationnaires \underline{X} , c'est-à-dire le problème de la synthèse de séries temporelles stationnaires. Ensuite, nous illustrerons la flexibilité avec une série de trois exemples trivariés partageant les distributions marginales et les mêmes corrélations.

4.1 Série temporelle stationnaire

Matrice de structure circulaire

Pour façonner des séries temporelles stationnaires, nous sélectionnons la matrice $\mathcal{A}_{i,j} = (1/d)$ associée avec une matrice \mathcal{E} doublement stochastique :

$$\mathcal{E} = \sum_{k=0}^{d-1} \alpha_k J_d^k, \quad \sum_{k=0}^{d-1} \alpha_k \equiv 1, \quad (\text{III.52})$$

où $J_d \in M_d(\mathbb{R})$ est définie par :

$$J_d = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}, \quad (\text{III.53})$$

et $J_d^0 = I_d$ est la matrice identité. Il est aisé de vérifier que $[\mathcal{A}^T, J_d] = 0$. Par suite, $[\mathcal{A}^T, \mathcal{E}] = 0$ et \underline{X}_n est stationnaire. De plus, les valeurs propres de J_d sont bien connues, il s'agit des racines d -ème de l'unité. Si on appelle $\omega = \exp(2i\pi/d)$, les valeurs propres λ_i de \mathcal{E} s'expriment simplement comme :

$$\lambda_i = \sum_{k=0}^{d-1} \alpha_k \omega^{ik}, \quad i = 1, \dots, d. \quad (\text{III.54})$$

Les vecteurs propres de J_d ont eux aussi une expression simple :

$$B_{i,j} = \omega^{ij}. \quad (\text{III.55})$$

Avec ces définitions, il est utile d'opérer un changement de base pour se placer dans la base des vecteurs propres de \mathcal{E} :

$$\begin{aligned}\tilde{\mathcal{A}} &= B\mathcal{A}B^{-1} = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & & \dots \\ 0 & \vdots & 1 \end{pmatrix} \\ \tilde{\mathcal{E}} &= B^{-1}\mathcal{E}B = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{pmatrix} \\ \tilde{M}(q) &= B^{-1}M(q)B.\end{aligned}\tag{III.56}$$

Ces calculs mènent à une expression intéressante des dépendances à deux points :

$$\mathbb{E}[X_k^{q_1} X_l^{q_2}] = \sum_i \lambda_i^{k-l} \tilde{M}(q_1)_{d,i} \tilde{M}(q_2)_{i,d}.\tag{III.57}$$

Pour simplifier les notations, définissons le vecteur colonne $(C_{\tilde{M}(q)})_k = \sum_i \tilde{M}(q)_{i,k}$, le vecteur ligne $(L_{\tilde{M}(q)})_k = \sum_j \tilde{M}(q)_{k,j}$ et la transformée de Fourier discrète (normalisée) d'un vecteur v :

$$\mathcal{F}(v)_k = \frac{1}{d} \sum_{l=1}^d v_l e^{\frac{2i\pi kl}{d}} = \frac{1}{d} \sum_l v_l \omega^{kl}.\tag{III.58}$$

Avec ces nouvelles notations, l'Eq. (III.57) devient

$$\begin{aligned}\mathbb{E}[X_0^q X_t^q] - \mathbb{E}[X_0^q] \mathbb{E}[X_t^q] &= \\ \sum_{k=1}^{\lfloor d/2 \rfloor} m_k \Re \left\{ \mathcal{F}(R_{\tilde{M}(q)})_k \overline{\mathcal{F}(C_{\tilde{M}(q)})_k} e^{-\frac{t-1}{\tau_k}} \lambda_k^{t-1} \right\},\end{aligned}\tag{III.59}$$

où $\lfloor z \rfloor$ représente la partie entière de z , $m_k = 1$ si $2k = d$ et $m_k = 2$ sinon, et $\Re \{ \}$ dénote la partie réelle d'un complexe. Dans ce cas spécifique, il n'y a qu'au plus $\lfloor d/2 \rfloor$ échelles de temps distinctes dans le signal. Comme exemple pédagogique, choisir $\mathcal{E} = \alpha_0 I_d + \alpha_1 J_d$ (où $\alpha_0 + \alpha_1 = 1$) nous permet d'analyser plus en profondeur la forme de la fonction de covariance. Les valeurs propres de \mathcal{E} s'écrivent $\lambda_k = \alpha_0 + \alpha_1 e^{\frac{2i\pi k}{d}}$ ce qui peut se mettre sous la forme $\lambda_k = e^{-\frac{1}{\tau_k}} e^{\pm i \frac{2\pi}{T_k}}$, $k = 1, \dots, \lfloor d/2 \rfloor$. En combinant ces résultats avec l'Eq. (III.59), on montre que $\tau_k = -1/\ln|\lambda_k|$ et $T_k = 2\pi/\arg(\lambda_k)$ sont les échelles de temps et périodes caractéristiques de la structure de dépendance ; ces échelles et périodes dépendent donc du choix conjoint de d et α_0 .

On peut aller plus loin et obtenir un développement asymptotique de τ_k : $\tau_k \approx_{\alpha_0 \alpha_1 \rightarrow 0} [\alpha_0 \alpha_1 (1 - \cos \frac{2\pi k}{d})]^{-1}$. De plus, accroître d augmente simultanément le nombre d'échelles de temps distinctes et le rapport entre la plus petite et plus grande de ces échelles caractéristiques. Plus précisément, on a asymptotiquement (pour $\alpha_0 \alpha_1 \rightarrow 0$)

$$\frac{\lambda_1}{\lambda_d} \sim \frac{(d/2\pi)^2}{2}.\tag{III.60}$$

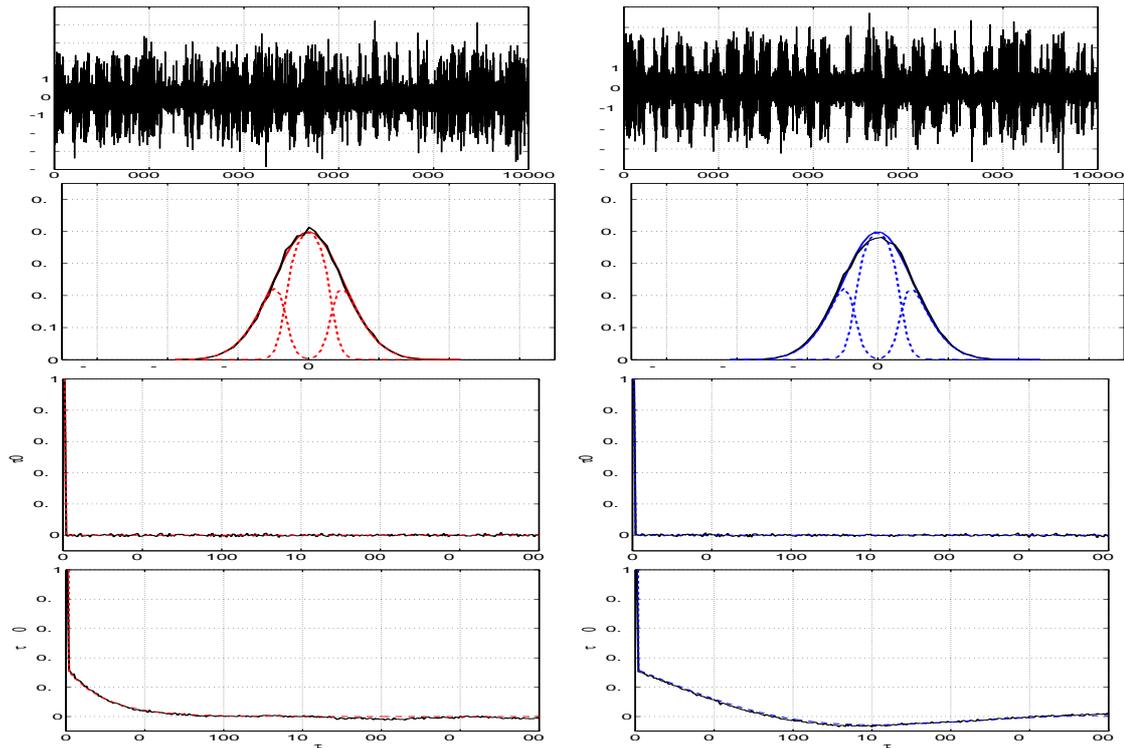


FIGURE 7.2 – **Synthèse numérique.** Deux séries temporelles différentes avec la même distribution marginale (une gaussienne de moyenne nulle et de variance $\sigma = 1$), la même fonction de covariance (choisie pour être une fonction δ de Dirac) mais des fonctions de covariance de leur carré différentes, et donc une distribution jointe différente ($\alpha_0 = 0.98$). À gauche \underline{X} , à droite \underline{Y} . Première ligne, une réalisation de la série temporelle. Ensuite, de haut en bas, distribution marginale estimée (ligne noire pleine) et théorique (lignes en tiret colorées), fonction de corrélation et fonction de corrélation des carrés.

Exemple

Pour illustrer le potentiel de notre construction théorique couplée avec notre méthode de synthèse, nous proposons un exemple pédagogique. Tout d'abord, nous souhaitons construire une série stationnaire de marginale gaussienne $P(x) = \mathcal{N}_{0,1}$, avec une fonction d'autocovariance qui consiste en une fonction de Dirac δ (i.e., aucune corrélation et $M(1) \equiv 0$), mais avec une structure de dépendance fixée par la statistique d'ordre 4, c'est-à-dire par le choix de $M(2)$. Par définition, X ne peut être un processus conjointement gaussien (sa distribution jointe diffère en effet d'une loi gaussienne multivariée). En second lieu, nous utilisons les mêmes d , \mathcal{A} et \mathcal{E} , pour construire une seconde série temporelle Y de taille n , indépendante de X et partageant la même distribution marginale et la même fonction d'autocovariance. Cependant, les fonctions d'autocovariance des carrés de X et Y –et par conséquent leur distribution jointe– diffèrent (via une matrice $M(2)$ différente).

Pour ces exemples, nous choisissons $d = 6$ pour obtenir trois échelles de temps distinctes et $\mathcal{E} = \alpha_0 I_d + \alpha_1 J_d$ ($\alpha_0 + \alpha_1 = 1$) dans un souci de simplicité. Avec ces choix, les trois échelles de temps disponibles sont :

$$\tau_1 \underset{\alpha_1 \rightarrow 0}{\approx} \frac{2}{\alpha_1}, \tau_2 \underset{\alpha_1 \rightarrow 0}{\approx} \frac{2}{3\alpha_1}, \tau_3 \underset{\alpha_1 \rightarrow 0}{\approx} \frac{1}{2\alpha_1}. \quad (\text{III.61})$$

Les deux séries X et Y diffèrent seulement par l'autocovariance de leur carrés ($\mathbb{E}[X_0^2 X_t^2] - \mathbb{E}[X_0^2] \mathbb{E}[X_t^2]$). Cela implique que leur matrices $M(2)$ doivent différer. Une simplification intéressante à ce point est de considérer les matrices $M(2)$ de la forme $M(2) = \alpha_0 D + \alpha_1 D J_d$ où D est une matrice diagonale. Dans ce cas, l'Eq. (III.59) se simplifie plus avant en :

$$\begin{aligned} \mathbb{E}[X_0^q X_t^q] - \mathbb{E}[X_0^q] \mathbb{E}[X_t^q] = \\ \sum_{k=1}^{\lfloor d/2 \rfloor} m_k |\mathcal{F}(\text{diag}(D))_k|^2 \Re \left\{ (\alpha_0 + \alpha_1 \omega^k) e^{-\frac{t-1}{\tau_k}} e^{2\pi i \frac{t-1}{T_k}} \right\}, \end{aligned} \quad (\text{III.62})$$

où $\text{diag}(D)_i = D_{i,i}$ est le vecteur diagonal de la matrice D . Par suite, choisir pour X ,

$$D_{i,i} = \begin{cases} \sigma_1 & \text{i pair} \\ \sigma_2 & \text{sinon} \end{cases}, \quad \sigma_1 + \sigma_2 = 1 \quad (\text{III.63})$$

entraîne que seulement τ_3 , c'est-à-dire la plus courte des échelles de temps, apparaît dans l'auto-covariance de X^2 :

$$\mathbb{E}[X_0^2 X_t^2] - \mathbb{E}[X_0^2] \mathbb{E}[X_t^2] = \frac{(\sigma_1 - \sigma_2)^2}{4} (\alpha_0 - \alpha_1)^t. \quad (\text{III.64})$$

Inversement, imposer pour Y

$$D_{i,i} = \begin{cases} \sigma_1 & i \leq 3 \\ \sigma_2 & i > 3 \end{cases}, \quad \sigma_1 + \sigma_2 = 1 \quad (\text{III.65})$$

implique que τ_1 et τ_3 contribuent simultanément à l'autocovariance de Y^2 :

$$\begin{aligned} \mathbb{E}[Y_0^2 Y_t^2] - \mathbb{E}[Y_0^2] \mathbb{E}[Y_t^2] = \\ \frac{(\sigma_1 - \sigma_2)^2}{36} \left[(\alpha_0 - \alpha_1)^t + 4 \Re \left\{ (2 - \alpha_1 - \sqrt{3}i\alpha_1) \lambda_1^{t-1} \right\} \right]. \end{aligned} \quad (\text{III.66})$$

Cependant, $\tau_1 \approx 4\tau_3$ est prépondérant pour les grands t et constitue donc le terme dominant.

Avec ces choix pédagogiques et documentés de $M(2)$ pour X et Y , nous pouvons utiliser les procédures numériques décrites en section 3, avec un noyau gaussien $K_{m,\sigma}(x) = \exp(-(x - m)^2/(2\sigma^2))$, pour construire la matrice de distribution $\mathcal{P}_{i,j}$. La méthode par chaîne de Markov cachée introduite en section 2 est ensuite utilisée pour synthétiser les signaux X et Y . L'analyse des signaux synthétisés est reportée en Fig. 7.2. Cette figure présente pour X (à gauche) et Y (à droite) un exemple de réalisation de la série temporelle, une comparaison des distributions marginales estimées et souhaitées, des fonctions de covariance, et des fonctions de covariance du carré de la série (de haut en bas). En résumé, la Fig. 7.2 montre clairement que X et Y partage la même distribution marginale et la même covariance mais avec des distributions différentes (comme nous le souhaitions).

En utilisant le même algorithme de synthèse, il est possible de construire bien d'autres exemples de séries temporelles partageant la même distribution marginale, la même corrélation mais différant en termes de distributions jointes. Cela ouvre d'intéressantes perspectives pour tester la dégradation des performances d'estimateurs classiques face à de la corrélation d'ordre supérieur.

4.2 Vecteur aléatoire

Conception multivariée :

Considérons maintenant la question de la synthèse de vecteurs aléatoires dont les composantes sont inter-dépendantes et possèdent des distributions marginales différentes. Comme nous l'avons vu dans la section 2, pour une dimension d fixée, lorsque n augmente, \underline{X} a tendance à devenir stationnaire. Une méthode pour circonvenir cet inconvénient est d'augmenter la taille (d) de la matrice R_d simultanément avec la taille du vecteur \underline{X}_n . Une possibilité est alors de choisir $d = (n + 1)d^*$, cela couplé avec une structure triangulaire (supérieure) par bloc afin de ne pas accroître plus que nécessaire le nombre de degrés de liberté :

$$R_d(x) = \begin{pmatrix} 0_{d^*} & R_{d^*}^{(1)}(x) & & 0_{d^*} \\ & 0_{d^*} & \ddots & \\ & & 0_{d^*} & R_{d^*}^{(n)}(x) \\ 0_{d^*} & & & 0_{d^*} \end{pmatrix}, \quad (\text{III.67})$$

où 0_{d^*} est un bloc de zéros de taille $d^* \times d^*$ et $R_{d^*}^{(k)}$ est une famille de $d^* \times d^*$ matrices, définie comme pour l'Eq. (III.1) :

$$R_{d^*}^{(k)}(x) = \mathcal{E}^{(*k)} \mathcal{P}^{(*k)}(x), \quad \int_{-\infty}^{+\infty} \mathcal{P}_{i,j}^{(*k)}(x) dx = 1.$$

Pour compléter ce tableau, nous lui adjoignons la matrice \mathcal{A} suivante

$$\mathcal{A} = \begin{pmatrix} 0_{d^*} & \dots & 0_{d^*} \\ \vdots & & \vdots \\ \mathcal{A}^* & \dots & 0_{d^*} \end{pmatrix}, \quad (\text{III.68})$$

où \mathcal{A}^* est une matrice de coefficients positifs de taille $d^* \times d^*$. De plus, si on définit $\mathcal{L}^*(M) = \text{tr}(\mathcal{A}^{*T}M)$ (en tant que forme linéaire sur les matrices de dimension $d^* \times d^*$), l'Eq. (III.4) devient :

$$\mathbb{P}(x_1, \dots, x_n) = \frac{\mathcal{L}^*\left(\prod_{k=1}^n R_{d^*}^{(k)}(x_k)\right)}{\mathcal{L}^*\left(\prod_{k=1}^n \mathcal{E}^{(*k)}\right)}. \quad (\text{III.69})$$

En se limitant au cas $\mathcal{E}^{(*k)} = \mathcal{E}^*$, la distribution jointe se simplifie en :

$$\mathbb{P}(x_1, \dots, x_n) = \frac{\mathcal{L}^*\left(\prod_{k=1}^n R_{d^*}^{(k)}(x_k)\right)}{\mathcal{L}^*\left((\mathcal{E}^*)^n\right)}. \quad (\text{III.70})$$

La distribution jointe décrite par l'Eq. (III.70) consiste en une légère variation sur le thème de l'Eq. (III.4), où l'immuable matrice de représentation $R_d(x)$ a été remplacée par une matrice de représentation variable $R_{d^*}^{(k)}(x)$. Les expressions générales (III.67) et (III.68) montrent que cette variation n'est rien de plus qu'une notation pratique, mais qui, au final, est équivalent au cas développé précédemment. Tous les résultats développés dans les sections 1 à 7 s'appliquent donc pleinement.

Cela signifie notamment qu'il est aisé de retrouver les expressions analytiques de la distribution marginale, des distributions jointes partielles (à p points) et des moments :

$$\mathbb{P}_k(X_k = x_k) = \frac{1}{\mathcal{L}((\mathcal{E}^*)^n)} \mathcal{L}\left((\mathcal{E}^*)^{k_1-1} R_{d^*}^{(k)}(x_k) (\mathcal{E}^*)^{n-k}\right), \quad (\text{III.71})$$

$$\begin{aligned} \mathbb{P}(X_{k_1} = x_{k_1}, \dots, X_{k_p} = x_{k_p}) &= \frac{1}{\mathcal{L}((\mathcal{E}^*)^n)} \\ &\mathcal{L}\left((\mathcal{E}^*)^{k_1-1} \left(\prod_{r=1}^{p-1} R_{d^*}^{(k_r)}(x_{k_r}) (\mathcal{E}^*)^{k_{r+1}-k_r-1}\right) R_{d^*}^{(k_p)}(x_{k_p}) (\mathcal{E}^*)^{n-k_p}\right), \end{aligned} \quad (\text{III.72})$$

$$\begin{aligned} \mathbb{E}\left[\prod_{r=1}^p X_{k_r}^{q_r}\right] &= \frac{1}{\mathcal{L}((\mathcal{E}^*)^n)} \\ &\mathcal{L}\left((\mathcal{E}^*)^{k_1-1} \left(\prod_{r=1}^{p-1} M^{(*k_r)}(q_r) (\mathcal{E}^*)^{k_{r+1}-k_r-1}\right) M^{(*k_p)}(q_p) (\mathcal{E}^*)^{n-k_p}\right), \end{aligned} \quad (\text{III.73})$$

avec $M^{(*k)}(q) = \int_{-\infty}^{+\infty} x^q R_{d^*}^{(k)}(x) dx$. Avec cette structure très particulière de la matrice R_d , chaque composante du vecteur aléatoire \underline{X} est relativement indépendante des autres. Notamment, les distributions univariées de X_k ne dépendent que de $\mathcal{P}^{(*k)}$. De plus, les formes spécifiques de la matrice \mathcal{A} et R_d imposent que la chaîne de Markov cachée Γ commence dans le premier bloc diagonal supérieur (de taille d^*) et finit dans le dernier bloc diagonal supérieur. En d'autres termes, à chaque pas k , la chaîne de Markov cachée se déplace du $k^{\text{ème}}$ bloc au $(k+1)^{\text{ème}}$ en suivant une transition i, j qui correspond à une loi de probabilité appartenant au bloc $R_{d^*}^{(k)}$.

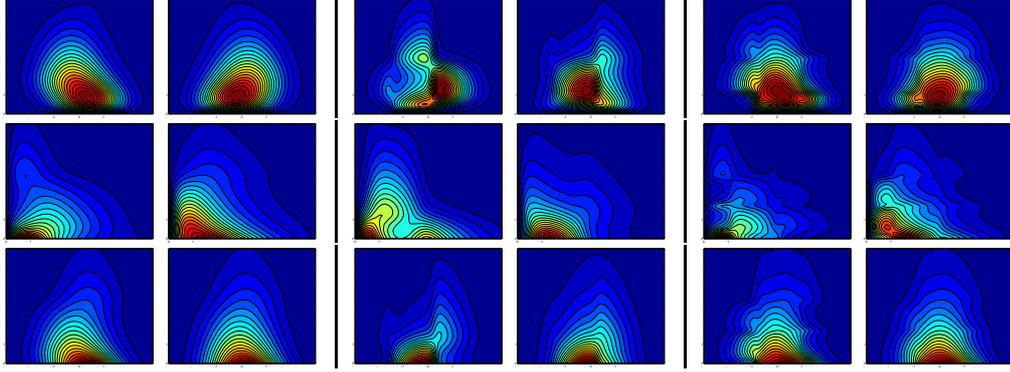


FIGURE 7.3 – Distributions bivariées de \underline{X}_n (à gauche), \underline{Y}_n (au milieu) et \underline{Z}_n (à droite), pour les pairs de composantes 1-2 (en haut), 2-3 (milieu) et 1-3 (en bas). Pour chacun des vecteurs aléatoires : à gauche : $p = 0.1$. À droite : $p = 0.8$.

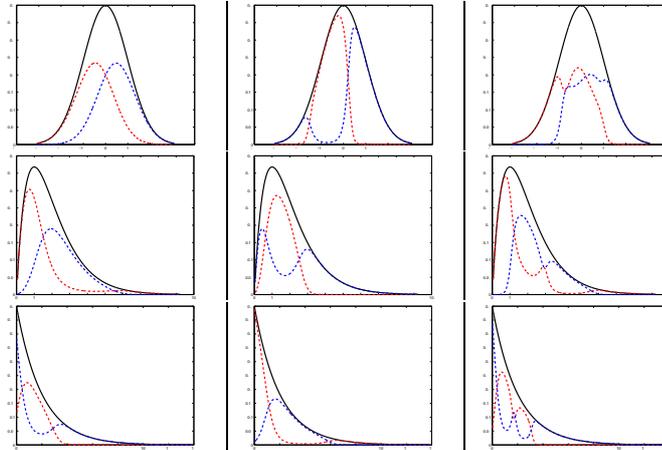


FIGURE 7.4 – Distributions marginales \mathbb{P}_k (lignes pleines) pour \underline{X}_n (à gauche), \underline{Y}_n (au milieu) et \underline{Z}_n (à droite) et les $c_{i,j,k} \mathcal{P}_{i,j}$ correspondant (lignes en pointillés), pour les composants X_1, Y_1, Z_1 (en haut), X_2, Y_2, Z_2 (au milieu) et X_3, Y_3, Z_3 (en bas).

Exemple trivarié

Nous allons maintenant expliciter la construction de deux vecteurs aléatoires trivariés différents $\underline{X} = (X_1, X_2, X_3)$ et $\underline{Y} = (Y_1, Y_2, Y_3)$. Leurs marginales seront fixées a priori à une distribution gaussienne $\mathcal{N}_{0,1}$ pour X_1 et Y_1 , une distribution Γ , avec un paramètre de forme $\alpha = 2$ et un paramètre d'échelle $\beta = 1$ pour X_2 et Y_2 , et une autre distribution Γ avec $\alpha = 1$ et $\beta = 2$ pour X_3 et Y_3 . Pour illustrer le potentiel de notre méthode, \underline{X} et \underline{Y} auront également les mêmes corrélations, mais des distributions jointes différentes.

Tout d'abord, nous fixons la dimension $d^* = 2$. Par simplicité, la matrice de structure sera $\mathcal{E}^* = \alpha_0 I_d + \alpha_1 J_d$ avec de plus $\mathcal{A}_{i,j}^* = (1/d^*)$. Ensuite, pour contrôler la corrélation, la matrice des moments $M^{(*k)}(1)$ est fixée, pour les deux vecteurs, à ($k = 1, 2, 3$) :

$$M^{(*k)}(1) = \mathcal{E}^* \circ \begin{pmatrix} m_{k,1} & m_{k,1} \\ m_{k,2} & m_{k,2} \end{pmatrix}, \quad (\text{III.74})$$

avec la contrainte que :

$$m_{k,1} + m_{k,2} = 2\mathbb{E}[X_k].$$

Si on définit $\Delta_k = m_{k,1} - m_{k,2}$, la covariance s'écrit

$$\text{Cov}[X_1, X_2] = \text{Cov}[Y_1, Y_2] = (1 - 2\alpha_0)\Delta_1\Delta_2,$$

$$\text{Cov}[X_2, X_3] = \text{Cov}[Y_2, Y_3] = (1 - 2\alpha_0)\Delta_2\Delta_3,$$

$$\text{Cov}[X_1, X_3] = \text{Cov}[Y_1, Y_3] = (1 - 2\alpha_0)^2\Delta_1\Delta_3.$$

Ainsi, la covariance de deux composantes successifs dépend linéairement de α_0 , et, si $\Delta_k\Delta_l > 0$, cette covariance est maximum pour $\alpha_0 = 1$, nulle à $\alpha_0 = 0.5$ et minimum pour $\alpha_0 = 0$.

On peut alors forcer les distributions jointes trivariées à être différentes en fixant les matrices des moments $M^{(**k)}(q)$ d'ordre $q = 2$ à des valeurs différentes. Celles de \underline{X} sont fixées à :

$$\begin{aligned} M^{(*1)}(2) &= \mathcal{E}^* \circ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad M^{(*3)}(2) = \mathcal{E}^* \circ \begin{pmatrix} 4.5 & 4.5 \\ 11.5 & 11.5 \end{pmatrix}, \\ M^{(*2)}(2) &= \mathcal{E}^* \circ \begin{pmatrix} 4.5 & 4.5 \\ 7.5 & 7.5 \end{pmatrix}; \end{aligned} \tag{III.75}$$

tandis que pour \underline{Y} , nous choisissons :

$$\begin{aligned} M^{(*1)}(2) &= \mathcal{E}^* \circ \begin{pmatrix} 0.5 & 0.5 \\ 1.5 & 1.5 \end{pmatrix}, \quad M^{(*3)}(2) = \mathcal{E}^* \circ \begin{pmatrix} 8 & 8 \\ 8 & 8 \end{pmatrix}. \\ M^{(*2)}(2) &= \mathcal{E}^* \circ \begin{pmatrix} 2.75 & 2.75 \\ 9.25 & 9.25 \end{pmatrix}. \end{aligned} \tag{III.76}$$

La construction précise des matrices $\mathcal{P}^{(**k)}$ suit alors la procédure développée en section 3, avec un noyau gaussien $K_{m,\sigma}(x) = \exp(-(x-m)^2/(2\sigma^2))$. Les matrices de probabilité $\mathcal{P}^{(**k)}$ résultant de cette procédure n'ont que deux composantes distinctes. La Fig. 7.4 montre de quelle manière les probabilités (pondérées) $c_{i,j,k}\mathcal{P}_{i,j}^{(**k)}$ sont combinées afin d'obtenir les distributions \mathbb{P}_k pour \underline{X}_n (à gauche) et \underline{Y}_n (au milieu).

De plus, en changeant le noyau K utilisé pour découper $\mathcal{P}_{i,j}^{(**k)}$, il est possible de varier aisément la probabilité jointe tout en satisfaisant le même jeu de contraintes. Le choix du noyau K introduit donc une versatilité intéressante dans la procédure de construction des matrices $\mathcal{P}^{(**k)}$. Par exemple, K pourrait être choisi pour correspondre à la famille des distributions Γ , voire à l'union des familles de densité gaussiennes et Γ . Afin d'illustrer cette versatilité, construisons donc un troisième vecteur aléatoire trivarié \underline{Z}_n , partageant les mêmes distributions marginales et covariances que \underline{X}_n et \underline{Y}_n (\underline{Z}_n possède en fait exactement les mêmes matrices de moments $M^{(**k)}(1)$ et $M^{(**k)}(2)$ que \underline{X}_n), mais obtenu avec un noyau non classique $K_{m,\sigma}(x) = (0.1 + ((x-m)/\sigma)^2) \exp(-((x-m)/\sigma)^2)$.

Les distributions bivariées partielles (à cause des difficultés inhérentes à la représentation d'une distribution trivariée sur du papier essentiellement bidimensionnel) sont représentées sur la Fig. 7.3 pour les paires (X_1, X_2) (ligne du haut), (X_2, X_3) (ligne du milieu) et (X_1, X_3) (ligne du bas) pour \underline{X}_n (bloc de gauche), \underline{Y}_n (bloc du milieu) et \underline{Z}_n (bloc de droite), dans le cas d'une covariance négative (colonne

de gauche dans chaque bloc) et positive (colonne de droite). Ces figures montrent clairement que, quelles que soit les paires considérées, les distributions bivariées (et a fortiori les distributions jointes trivariées) diffèrent pour les trois vecteurs, bien que leurs composantes aient les mêmes distributions univariées et corrélations. En complément, la Fig. 7.4 montre que les distributions marginales ciblées \mathbb{P}_k sont obtenues par un mélange des densités $c_{i,j,k}\mathcal{P}_{i,j}^{(*k)}$ générées par l'algorithme décrit en section 3. Il est à noter que dans le contexte particulier des exemples proposés, ce mélange fut limité (pour un souci de simplicité) pour n'impliquer que deux densités $\mathcal{P}^{(*k)}$ différentes pour chaque composante alors que $d^* \times d^* = 4$ distributions distinctes auraient pu être utilisées dans le cas général. De façon intéressante, une comparaison pour chacune des composantes entre les trois vecteurs (i.e., pour une ligne donnée) illustre le fait que la même distribution marginale \mathbb{P}_k est reconstituée en utilisant différents mélanges de densités $c_{i,j,k}\mathcal{P}_{i,j}^{(*k)}$ pour chacun des trois vecteurs, ce qui constitue une des signatures potentielles du fait que les distributions jointes de \underline{X}_n , \underline{Y}_n et \underline{Z}_n diffèrent.

Nos exemples trivariés furent choisis ici en tant que compromis, suffisamment simples pour être pédagogiques (Il est encore possible de lire facilement trois distributions jointe partielles de paires de composantes) mais aussi suffisamment complexes pour illustrer le potentiel et la richesse de notre méthode. Il est assez facile de construire d'autres exemples multivariés.

4.3 Exemple en dimension infinie

Il est possible de généraliser la procédure de synthèse décrite ci-dessus pour générer des variables aléatoires dont la représentation matricielle est de dimension infinie. En dimension infinie, la définition de la classe des opérateurs considérée devient bien plus délicate. Néanmoins, on peut commencer par se limiter aux matrices \mathcal{E} infinies possédant un nombre fini de coefficients non nuls pour chaque colonne. On peut, par exemple, considérer la matrice $\mathcal{E}_{k \in \mathbb{Z}, l \in \mathbb{Z}}$ définie par

$$\mathcal{E}_{k,l} = \delta_{l+1,k} + \delta_{l-1,k}. \tag{III.77}$$

Cette matrice \mathcal{E} modélise un mouvement brownien sur l'espace \mathbb{Z} des états. La figure 7.6 présente une présentation schématique de la structure de la matrice \mathcal{E} et du mouvement brownien associé. Cette matrice est doublement stochastique, il est alors intéressant de définir \mathcal{L} par

$$\mathcal{L}(M) = \sum_{k \in \mathbb{Z}} M_{k,0}. \tag{III.78}$$

Avec cette définition de \mathcal{L} et \mathcal{E} la chaîne de Markov résultante est homogène. Il ne reste ensuite plus qu'à choisir la matrice des distributions \mathcal{P} , par exemple des lois normales, avec

$$\mathcal{P}_{k,l} = \mathcal{N}_{\mu=l, \sigma=5}. \tag{III.79}$$

La Fig. (7.5) donne un exemple d'une réalisation d'un tel processus.

On peut cependant construire des matrices \mathcal{E} plus exotiques. On peut par exemple simuler un déplacement le long d'un arbre binaire complet de profondeur infinie.

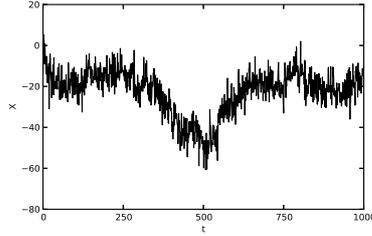


FIGURE 7.5 – Exemple d’une réalisation de X dans le cas où \mathcal{E} correspond à une matrice de transition d’un mouvement brownien sur \mathbb{Z} .

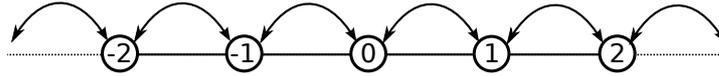


FIGURE 7.6 – Esquisse de la structure de la matrice \mathcal{E} sur \mathbb{Z} . Chaque ligne k de la matrice correspond à une position sur l’axe \mathbb{Z} . Les coefficients non nuls de la matrice correspondent à des sauts soit vers la gauche soit vers la droite.

Appelons $\mathbb{B} = \{(p, k), 0 \leq k < 2^p, p \in \mathbb{N}\}$ l’ensemble des couples d’entier correspondant à une position dans un arbre binaire. On peut ensuite définir une matrice $\mathcal{E}_{(p,k) \in \mathbb{B}, (q,l) \in \mathbb{B}}$ par

$$\begin{cases} \mathcal{E}_{(p,k),(0,0)} = \frac{1}{2} \left(\delta_{(1,0),(p,k)} + \delta_{(1,1),(p,k)} \right), \\ \mathcal{E}_{(p,k),(q,l)} = \frac{1}{4} \left(\delta_{(q+1,2l),(p,k)} + \delta_{(q+1,2l+1),(p,k)} + 2\delta_{(q-1, \lfloor \frac{l}{2} \rfloor), (p,k)} \right). \end{cases} \quad (\text{III.80})$$

La Fig. 7.8 présente une vue simplifiée de la structure de cette matrice \mathcal{E} . On peut ensuite allier cette matrice \mathcal{E} avec

$$\mathcal{L}(M) = \sum_{t \in \mathbb{B}} M_{t,(0,0)}, \quad (\text{III.81})$$

et

$$\mathcal{P}_{(p,k),(q,l)} = \mathcal{N}_{\mu=k 2^{-p}, \sigma=2^{-p}}. \quad (\text{III.82})$$

Une illustration d’un tel processus est visible sur la Fig. (7.7).

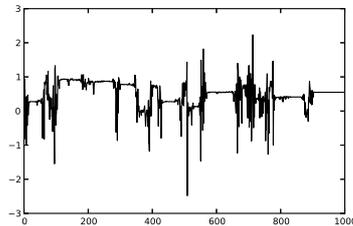


FIGURE 7.7 – Exemple d’une réalisation de X dans le cas où \mathcal{E} correspond à une matrice de transition d’un mouvement brownien sur un arbre binaire complet.

D’autres exemples sont bien sûr possibles, les deux cas présentés plus haut ne constituent qu’un bref aperçu des possibilités des variables aléatoires à représentation matricielle infinie. Ce sujet n’a cependant pas pu être abordé en profondeur au sein de cette thèse, par manque de temps.

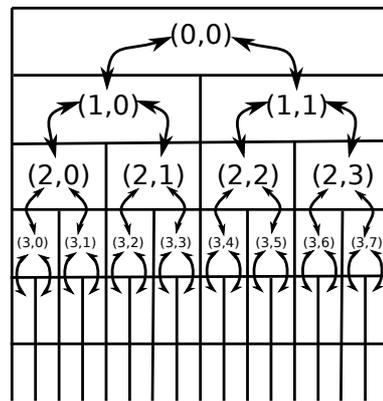


FIGURE 7.8 – Esquisse de la structure de la matrice \mathcal{E} sur l'arbre binaire \mathbb{B} . Chaque ligne (p, k) de la matrice correspond à une position dans l'arbre binaire \mathbb{B} . Les coefficients non nuls de la matrice correspondent à des sauts soit vers le nœud père (s'il existe) soit vers un des deux nœuds fils .

Chapitre 8

Sommes et extrêmes revisités

1 Motivation

Il est possible de synthétiser des variables aléatoires dont la probabilité jointe possède une représentation matricielle en utilisant les modèles à chaînes de Markov cachées proposés au chapitre 7. Nous allons montrer dans ce chapitre que ces modèles de chaînes de Markov cachées peuvent aussi être utilisés pour caractériser le comportement de la somme ou du maximum de ces variables aléatoires. Cela nous permettra également de réinterpréter les différences fondamentales entre le cas \mathcal{E} diagonalisable et \mathcal{E} non diagonalisable comme une conséquence de la structure de graphe induite par \mathcal{E} . Cette structure de graphe est également étroitement liée à l'apparition de distributions limites non standards pour la somme et le maximum dans certaines situations. Afin de fermer le cercle, nous concluons sur l'apparition d'une rupture potentielle de la loi des grands nombres pour les matrices \mathcal{E} non diagonalisables.

2 Équations matricielles

Nous nous intéressons au comportement de la somme

$$\mathfrak{S}(X) = \sum_{k=1}^n X_k, \quad (\text{III.83})$$

ou du maximum

$$M(X) = \max_{k=1..n} \{X_k\} \quad (\text{III.84})$$

des composantes d'un vecteur aléatoire X de taille n à représentation matricielle. Un des avantages des représentations matricielles est la facilité avec laquelle il est possible d'exprimer les sommes et maxima. En effet, si on considère le maximum sur une réalisation du processus matriciel, on a

$$\begin{aligned} \mathbb{P}(X_1, \dots, X_n < x) &= \frac{1}{\mathcal{L}(\mathcal{E}^n)} \mathcal{L} \left(\prod_{i=1}^n \int_{-\infty}^x R_d(x_i) dx_i \right), \\ &= \frac{1}{\mathcal{L}(\mathcal{E}^n)} \mathcal{L} \left(\left(\int_{-\infty}^x R_d(t) dt \right)^n \right). \end{aligned} \quad (\text{III.85})$$

Si on définit la fonction cumulative matricielle \mathcal{F} par

$$\mathcal{F}(x) = \int_{-\infty}^x R_d(t) dt, \quad (\text{III.86})$$

on obtient alors

$$\mathbb{P}(X_1 < x, \dots, X_n < x) = \frac{\mathcal{L}(\mathcal{F}(x)^n)}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.87})$$

On retrouve donc un résultat très similaire au cas i.i.d. (cf. Chapitre I), si ce n'est que la fonction cumulative a été remplacée par une fonction cumulative matricielle \mathcal{F} , qu'un terme de normalisation $\mathcal{L}(\mathcal{E}^n)$ est apparu, complété par la projection \mathcal{L} . On peut remarquer que dans le cas scalaire, ces termes disparaissent et on retrouve bien les relations classiques.

On a un résultat très similaire pour la somme de variables aléatoires, si on s'intéresse à la fonction caractéristique de la somme \mathfrak{S} :

$$\begin{aligned} \phi_{\mathfrak{S}}(z) &= \mathbb{E} \left[e^{-iz \sum_{i=1}^n X_i} \right], \\ &= \frac{1}{\mathcal{L}(\mathcal{E}^n)} \mathcal{L} \left(\prod_{i=1}^n \int_{-\infty}^{+\infty} e^{-izx_i} R_d(x_i) dx_i \right). \end{aligned} \quad (\text{III.88})$$

Par suite, si on définit la fonction caractéristique matricielle

$$\Phi(z) = \int_{-\infty}^{+\infty} e^{-izx} R_d(x) dx, \quad (\text{III.89})$$

on peut simplifier l'équation précédente en

$$\phi_{\mathfrak{S}}(z) = \frac{\mathcal{L}(\Phi(z)^n)}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.90})$$

On retrouve exactement le même type de relation que dans le cas du maximum. Maxima et sommes continuent donc de partager une structure mathématique commune dans le cas des processus à représentation matricielle. Il serait donc possible d'utiliser ces représentations matricielles pour caractériser l'ensemble des lois limites possibles pour \mathfrak{S} et M . Toutefois, certaines difficultés apparaissent rapidement. La définition des lois stables n'est pas triviale en présence d'une taille explicite du signal. De plus, la définition des distributions limites elles-mêmes peut poser problème. En effet, on peut définir deux formes différentes de stabilité pour ces distributions. Par exemple, pour le maximum, on peut d'abord, considérer une condition de stabilité sur \mathcal{F} (ou Φ dans le cas de la somme) :

$$\mathcal{F}(a_n x + b_n) = \frac{\mathcal{F}(x)^n}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.91})$$

Cette équation ressemble fortement à une version matricielle de l'équation fonctionnelle classique qui caractérise les lois max-stables dans le cas i.i.d.,

$$F(a_n x + b_n) = F(x)^n. \quad (\text{III.92})$$

Cette formulation est néanmoins problématique puisque \mathcal{F} n'a pas vraiment de sens physique (ou statistique) ; seule $\mathcal{L}(\mathcal{F})$ en a un. Une formulation plus exacte pour les distributions limites serait alors

$$\mathcal{L}(\mathcal{F}(a_n x + b_n)) = \frac{\mathcal{L}(\mathcal{F}(x)^n)}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.93})$$

Il semble difficile de résoudre exhaustivement cette équation fonctionnelle dans le cas général. Alternativement, nous avons vu dans le chapitre 7 qu'il est possible de réinterpréter les processus à représentations matricielle comme des modèles à chaîne de Markov cachée. Cette interprétation n'est finalement qu'une manière différente d'aborder les produits de matrices précédents. Elle présente l'avantage qu'il est possible de réutiliser en grande partie les résultats déjà connus sur les propriétés de convergence des chaînes de Markov.

3 Modèle à chaîne de Markov cachée

Dans la reformulation sous forme de modèle à chaîne de Markov cachée, nous savons que

$$\mathbb{P}(x_1, \dots, x_n) = \sum_{\Gamma} \kappa(\Gamma) \prod_{i=1}^n \mathcal{P}_{\Gamma_i, \Gamma_{i+1}}(x_i). \quad (\text{III.94})$$

La probabilité $\kappa(\Gamma)$ de la chaîne Γ s'écrit

$$\kappa(\Gamma) = \frac{A_{\Gamma_1, \Gamma_{n+1}}}{\mathcal{L}(\mathcal{E}^n)} \prod_{i=1}^n \mathcal{E}_{\Gamma_i, \Gamma_{i+1}}. \quad (\text{III.95})$$

En utilisant cette formulation, il est possible de séparer la synthèse du vecteur aléatoire X en deux étapes distinctes. Tout d'abord, la chaîne de Markov Γ est générée. Ensuite, la réalisation de Γ détermine quelles sont les lois aléatoires (indépendantes) utilisées pour générer X . Notre approche est fondée sur une simple remarque : pour une chaîne de Markov donnée Γ , il est très simple d'établir la fonction cumulative de M ou la fonction caractéristique de \mathfrak{S} puisque les composantes de X sont indépendantes à Γ fixée, i.e.,

$$\mathbb{P}(x_1, \dots, x_n | \Gamma) = \prod_{i=1}^n \mathcal{P}_{\Gamma_i, \Gamma_{i+1}}(x_i). \quad (\text{III.96})$$

Dans le cas du maximum, on a donc

$$F_M^\Gamma(x) = \prod_{i=1}^n F_{\mathcal{P}_{\Gamma_i, \Gamma_{i+1}}}(x). \quad (\text{III.97})$$

Si on considère $N_{i,j}$ le nombre de fois que la chaîne Γ passe par la transition $i \rightarrow j$

$$N_{i,j} = \text{card}\{k, \Gamma_k = i \text{ et } \Gamma_{k+1} = j\}, \quad (\text{III.98})$$

on peut réécrire l'équation (III.97) sous la forme

$$F_M^\Gamma(x) = \prod_{1 \leq i, j \leq d} F_{\mathcal{P}_{i,j}}(x)^{N_{i,j}}. \quad (\text{III.99})$$

La loi suivie par le maximum ne dépend donc que de $\mathcal{P}_{i,j}$ et $N_{i,j}$ pour une chaîne Γ fixée. De plus, une fois que l'on connaît F_M^Γ , obtenir la cumulative dans le cas général revient juste à intégrer sur toutes les chaînes possibles Γ

$$F_M(x) = \sum_{\Gamma} \kappa(\Gamma) F_M^\Gamma(x). \quad (\text{III.100})$$

Si on s'intéresse à la somme \mathfrak{S} , on retrouve le même type de résultat pour la fonction caractéristique le long d'une chaîne

$$\phi_{\mathfrak{S}}^{\Gamma}(x) = \prod_{1 \leq i, j \leq d} \phi_{\mathcal{P}_{i,j}}(x)^{N_{i,j}}, \quad (\text{III.101})$$

et la fonction caractéristique globale

$$\phi_{\mathfrak{S}}(x) = \sum_{\Gamma} \kappa(\Gamma) \phi_M^{\Gamma}(x). \quad (\text{III.102})$$

Que ce soit dans le cas du maximum et de la somme, il est donc inutile de connaître précisément le comportement de Γ , il suffit de connaître le comportement de $N_{i,j}$. On peut voir ce phénomène comme une conséquence de la commutativité et de l'associativité de la somme ou du maximum. Puisque la somme et le maximum sont invariants par permutation des termes, il n'est pas nécessaire de savoir dans quel ordre (Γ) on effectue la somme (ou prend le maximum) des composantes de X et on peut se contenter de connaître la nature globale (résumée par $N_{i,j}$) des termes sommés (ou maximisés).

La question est donc de savoir s'il est possible de prédire le comportement de $N_{i,j}$ pour les chaînes de Markov que nous considérons. Les chaînes Γ ne sont cependant pas des chaînes homogènes, ce qui rend le problème plus complexe et intéressant. En exploitant la riche littérature sur les chaînes de Markov et les matrices de voisinage, il est cependant possible de caractériser précisément le comportement de la somme et du maximum. Afin de considérer simultanément le cas de la somme et du maximum, nous appellerons \mathfrak{F} la fonction naturelle associée, i.e., la cumulative dans le cas du maximum et la fonction caractéristique pour la somme et nous noterons X^{\oplus} pour indiquer soit la somme soit le maximum. Avec cette notation, les équations (III.101) et (III.99) peuvent être fusionnées en

$$\mathfrak{F}(X^{\oplus})^{\Gamma}(x) = \prod_{1 \leq i, j \leq d} \mathfrak{F}_{\mathcal{P}_{i,j}}(x)^{N_{i,j}}. \quad (\text{III.103})$$

Cette formulation permet d'étudier simultanément le comportement de la somme et du maximum en utilisant leur structure algébrique commune. À partir de là, il suffit d'arriver à décrire le comportement de $N_{i,j}$. Comme on peut s'y attendre, le cas général est complexe. Il est cependant possible de réduire le cas général à un système équivalent plus simple. Pour ce faire, il est tout d'abord nécessaire d'analyser les différents types de comportements possibles de la chaîne Γ . On peut distinguer deux cas différents. Le premier cas correspond à une chaîne irréductible dans lequel le système va être mélangé rapidement. La corrélation du système est donc forcément à courte portée. Il s'agit du cas le plus simple que nous étudierons dans la section 4. Le deuxième cas correspond aux chaînes réductibles pour lesquelles il existe des transitions "irréversibles". Nous reviendrons plus en détail sur la définition de ces transitions irréversibles. Cependant, dans notre étude, nous commencerons par étudier des chaînes réductibles simples, c'est-à-dire des chaînes Γ dont la matrice \mathcal{E} peut être décomposée en $\mathcal{E} = \text{Id} + H$ où H est une matrice nilpotente. Pour ces chaînes simples, toutes les transitions non diagonales (i.e. la partie nilpotente de \mathcal{E}) sont irréversibles. Pour donner une idée du type de système considéré, nous étudierons en premier lieu le cas des chaînes réductibles simples à deux états dans la section

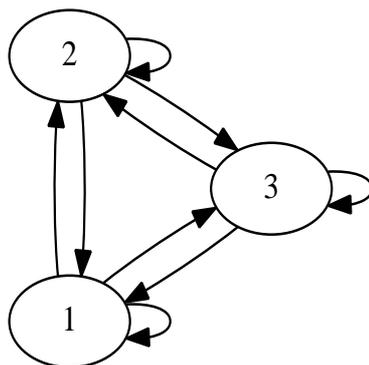


FIGURE 8.1 – Exemple de graphe de matrice \mathcal{E} irréductible. L'arête $i \rightarrow j$ n'est présente que si $\mathcal{E}_{i,j} \neq 0$.

5 avant d'aborder le cas des chaînes réductibles simples à d états dans la section 6. Enfin, nous concluerons en montrant dans la section 7 qu'il est possible de décrire le cas général en le réduisant à un modèle simplifié de chaînes réductibles simples en utilisant notre connaissance des chaînes irréductibles.

4 Chaînes irréductibles

4.1 Chaînes homogènes

Le cas le plus simple correspond au cas où les chaînes Γ sont des chaînes de Markov homogènes de matrice de transition \mathcal{E} irréductible. On dit qu'une matrice \mathcal{E} est une matrice irréductible si elle est fortement connectée, i.e.,

$$\forall i, j, \exists n, \quad (\mathcal{E}^n)_{i,j} > 0. \quad (\text{III.104})$$

Une manière pratique de visualiser cette notion de matrice irréductible est d'observer le graphe orienté $G(E)$ associé à la matrice \mathcal{E} : chaque nœud de ce graphe correspond à une ligne (ou une colonne) de la matrice \mathcal{E} . On ajoute ensuite une arête $i \rightarrow j$ si et seulement si $\mathcal{E}_{i,j} \neq 0$. La matrice \mathcal{E} est alors irréductible si et seulement si, pour tout couple de nœuds i, j du graphe $G(E)$, il existe un chemin de i vers j et un chemin de j vers i . Un exemple de matrice \mathcal{E} irréductible est donné par la figure 8.1. Si la matrice \mathcal{E} est irréductible, le théorème de Perron-Frobenius [94] implique qu'il existe un ensemble de r (r divisant d) valeurs propres dominantes λ_k de module ν telles que

$$\lambda_k = \nu \omega_r^{k-1} \quad (\text{III.105})$$

où $\omega_r = e^{i\frac{2\pi}{r}}$. Grâce à l'existence de ces valeurs propres dominantes (et la nature homogène de la chaîne), on peut montrer que les chaînes Γ vont rapidement mélanger le système, ce qui permet d'obtenir le théorème ergodique suivant [94] : il existe des constantes $\alpha_{i,j} \in \mathbb{R}^+$ telles que, pour presque toutes les chaînes Γ ,

$$\lim_{n \rightarrow +\infty} \frac{N_{i,j}}{n} = \alpha_{i,j}. \quad (\text{III.106})$$

Le taux de transition $N_{i,j}$ est donc prévisible et nous pouvons donc réécrire l'Eq. (III.103) sous la forme

$$\mathfrak{F}_{X^\oplus}^\Gamma(x) = \left(\prod_{1 \leq i,j \leq d} \mathfrak{F}_{\mathcal{P}_{i,j}}(x)^{\alpha_{i,j}} \right)^n. \quad (\text{III.107})$$

On peut donc exprimer \mathfrak{F} comme un produit pondéré de fonctions naturelles élevées à la puissance n . Cette forme suggère de construire une variable aléatoire synthétique i.i.d. W telle que

$$\mathfrak{F}_W(x) = \prod_{i,j} \mathfrak{F}_{\mathcal{P}_{i,j}}(x)^{\alpha_{i,j}}. \quad (\text{III.108})$$

Si on considère le maximum ou la somme W^\oplus de n copies i.i.d. de W , la fonction naturelle de W^\oplus est alors

$$\mathfrak{F}_{W^\oplus}(x) = \prod_{i,j} \mathfrak{F}_{\mathcal{P}_{i,j}}(x)^{\alpha_{i,j}n} = \mathfrak{F}_{X^\oplus}^\Gamma(x). \quad (\text{III.109})$$

Les fonctions naturelles de W^\oplus et X^\oplus coïncident donc. Grâce à l'isomorphisme entre fonction naturelle et variables aléatoires, cela signifie que W^\oplus et X^\oplus sont de mêmes lois. Le comportement du maximum (respectivement de la somme) de X est donc identique à celui du maximum (respectivement la somme) de n variables i.i.d. W_k . Par conséquent, pour décrire le comportement asymptotique de X^\oplus , il suffit d'appliquer les résultats classiques sur la somme ou le maximum de variables i.i.d. à la variable synthétique W : les lois limites de la somme et du maximum X sont donc forcément des lois limites standards.

4.2 Chaînes non homogènes

Il faut cependant noter que dans le cas général les chaînes de Markov ne sont pas homogènes, puisque les probabilités de transition pour une chaîne d'état final f sont de la forme

$$\mathbb{P}(\Gamma_{k+1} = j | \Gamma_k = i) = \mathcal{E}_{i,j} \frac{(\mathcal{E}^{n-k})_{j,f}}{(\mathcal{E}^{n-k+1})_{i,f}}. \quad (\text{III.110})$$

Cependant, si \mathcal{E} est irréductible et si on se trouve suffisamment loin de l'état final, cette chaîne de Markov inhomogène peut être approchée par une chaîne de Markov homogène. Si la matrice \mathcal{E} est apériodique et si on note $\mu^{(k)}$ le vecteur propre associé à la valeur propre λ_k de \mathcal{E} , avec pour $i \neq 1$, $|\lambda_1| > |\lambda_i|$ on a

$$\mathbb{P}(\Gamma_{k+1} = j | \Gamma_k = i) \approx \mathcal{E}_{i,j} \frac{\mu_j^{(1)}}{\lambda_1 \mu_i^{(1)}}. \quad (\text{III.111})$$

Loin de l'état final, la chaîne de Markov Γ se comporte donc comme une chaîne de Markov homogène. On retrouve le même type de comportement sur une chaîne périodique. En premier lieu, si la chaîne est périodique on peut associer un ordre $o : \mathbb{N} \rightarrow \mathbb{Z}/r\mathbb{Z}$ à chaque état tel que

$$\mu_i^{(k)} = \mu_i^{(1)} \omega_r^{ko(i)}, \quad (\text{III.112})$$

ce qui permet de montrer que si $n - k + o(i) \equiv o(f)[r]$

$$\mathbb{P}(\Gamma_{k+1} = j | \Gamma_k = i) \approx \mathcal{E}_{i,j} \frac{\mu_j^{(1)}}{\lambda_1 \mu_i^{(1)}}. \quad (\text{III.113})$$

On peut donc bien approcher la chaîne Γ par une chaîne de Markov homogène même dans le cas périodique lorsqu'on est suffisamment éloigné de l'état final. De plus, la convergence de la matrice de transition vers la matrice homogène se fait de manière exponentielle en $n - k$, le temps de convergence est donc trop court pour perturber le comportement ergodique de la chaîne dans la limite $n \rightarrow +\infty$.

Les matrices \mathcal{E} irréductibles correspondent donc au cas facile et peu intéressant lorsqu'on étudie le comportement de la somme et du maximum. Or, dans notre étude des propriétés statistiques des processus à représentation matricielle au chapitre 6, nous avons mis en évidence que ces propriétés statistiques diffèrent fortement selon que \mathcal{E} est diagonalisable ou non. On peut donc se demander quels sont les liens entre la classe des matrices \mathcal{E} diagonalisables et celle des matrices \mathcal{E} irréductibles. On peut commencer par remarquer que les matrices diagonalisables ne sont pas nécessairement irréductibles. Par exemple, les matrices \mathcal{E} diagonales sont diagonalisables mais pas irréductibles. Réciproquement, une matrice irréductible peut ne pas être diagonalisable. Cependant, il faut noter que dans ce cas le comportement de Γ sera dominée par la présence d'un ensemble de valeurs propres dominantes d'espace propre de dimension 1, et asymptotiquement la structure de corrélation de \underline{X} sera semblable à celle d'une matrice diagonalisable.

5 Chaîne réductible simple à deux états

Les cas intéressants apparaissent donc si on considère les matrices \mathcal{E} non irréductibles. Pour faciliter la lecture, nous appellerons par la suite "matrices réductibles" de telles matrices et "chaînes réductibles", les chaînes dont la matrice \mathcal{E} est réductible. Le cas général est cependant assez complexe, nous allons l'aborder étape par étape. Dans un premier temps, nous nous restreindrons à l'étude des chaînes réductibles "simples" dont la matrice de structure \mathcal{E} peut s'écrire sous la forme

$$\mathcal{E} = \text{Id} + H, \quad (\text{III.114})$$

où H est une matrice nilpotente d'ordre $p + 1 \in \mathbb{N}$:

$$p + 1 = \min\{p, H^p = 0, p \in \mathbb{N}\}. \quad (\text{III.115})$$

Dans cette section, nous nous focaliserons sur le cas particulier des chaînes réductibles simples à deux états. Ce modèle simple permet déjà de mettre en lumière l'existence de lois limites non standards dans le cas du maximum et de la somme, mais aussi l'existence d'une rupture de la loi des grands nombres. Il nous permettra également d'introduire les notions nécessaires pour traiter le cas moins spécifique des chaînes réductibles simples à d états.

Nous commençons donc par l'exemple le plus simple possible d'une matrice \mathcal{E} réductible

$$\mathcal{E} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \quad (\text{III.116})$$

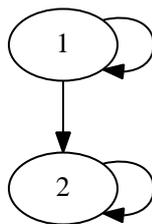


FIGURE 8.2 – Structure de la matrice \mathcal{E} . Seules les arêtes non nulles sont représentés.

La matrice \mathcal{E} n'est pas irréductible car il n'existe pas de chemin de l'état 2 vers l'état 1 comme le montre la Fig. 8.2. Les puissances de \mathcal{E} sont

$$\mathcal{E}^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}. \quad (\text{III.117})$$

On définit \mathcal{L} par

$$\mathcal{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (\text{III.118})$$

Avec ces définitions, seules les chaînes Γ qui commencent dans l'état 1, restent un certain temps dans cet état puis basculent dans l'état 2, sont possibles.

En d'autres termes, toutes les chaînes de probabilité non nulle sont de la forme

$$\Gamma_k^{(l)} = \begin{cases} 1 & k < l, \\ 2 & k \geq l. \end{cases} \quad (\text{III.119})$$

De plus, si on utilise l'Eq. (III.95) pour calculer les probabilités des chaînes $\Gamma^{(l)}$, on obtient

$$1 \leq l \leq n, \quad \mathbb{P}(\Gamma = \Gamma^{(l)}) = \frac{1}{n}. \quad (\text{III.120})$$

Toutes les chaînes $\Gamma^{(l)}$ sont équiprobables, ce qui implique que

$$\begin{aligned} p(N_{1,1} = k) &= p(N_{2,2} = k) = \frac{1}{n}, \\ p(N_{1,2} = k) &= \delta_{k,1}. \end{aligned} \quad (\text{III.121})$$

Les $N_{i,i}$ suivent donc une loi uniforme, avec toutefois la contrainte que $N_{1,1} + N_{2,2} = n - 1$. Contrairement au cas irréductible, $\frac{N_{i,i}}{n}$ ne converge pas presque sûrement vers une constante déterministe. Ici la chaîne de Markov ne mélange pas les états initiaux, elle les sépare en deux blocs indépendants dont la taille fluctue fortement. La difficulté par rapport à la section 4 consiste donc à normaliser simultanément ces deux blocs et ce quel que soit l . À partir de ce point, il est nécessaire de scinder en deux l'étude de la somme et du maximum, car leur comportement vis-à-vis de la normalisation diffère fortement.

5.1 Maximum

Pour ce faire, si on renomme Y_k les variables aléatoires i.i.d. du premier bloc (i.e., $1 \leq k < l, Y_k \equiv X_k$) et Z_k les variables i.i.d. du second bloc (i.e., $1 \leq k \leq n-l, Z_k \equiv X_{l+k}$) on peut réécrire le maximum $M(X|\Gamma)$ comme

$$M(X|\Gamma) = \max\left\{\max_{k=1\dots l-1}\{Y_k\}, \max_{k=1\dots(n-l)}\{Z_k\}, X_l\right\} \quad (\text{III.122})$$

Si on normalise $M(X|\Gamma^{(l)})$ avec la suite (a_n, b_n) , on peut négliger la variable de transition X_l et obtenir

$$\frac{M(X|\Gamma^{(l)}) - b_n}{a_n} = \max\left\{\frac{\max_{k=1\dots l-1}\{Y_k\} - b_n}{a_n}, \frac{\max_{k=1\dots(n-l)}\{Z_k\} - b_n}{a_n}, \frac{X_l - b_n}{a_n}\right\}. \quad (\text{III.123})$$

On obtient donc exactement la même normalisation sur les deux blocs. Sauf cas particuliers, sur lesquels nous reviendrons plus tard, la normalisation devrait également permettre de négliger la variable de transition X_l – notamment lorsque Y et X ne sont pas majorés. Supposons pour l'instant qu'une telle normalisation (a_n, b_n) existe, si on note $x_n = a_n x + b_n$, on peut alors calculer $F_{\Gamma^{(l)}}^{\text{lim}}$ la distribution cumulative associée à la chaîne $\Gamma^{(l)}$. Nous avons

$$\begin{aligned} F_{\alpha,n}(x) &= F_Y(x_n)^{\alpha n} F_Z(x_n)^{(1-\alpha)n} \\ &\sim \exp\left(n\left[\alpha \overline{F}_Y(x_n) + (1-\alpha)\overline{F}_Z(x_n)\right]\right). \end{aligned} \quad (\text{III.124})$$

Pour obtenir la convergence en distribution de $F_{\alpha,n}(x)$, il est nécessaire que (a_n, b_n) vérifie

$$\exists f, \quad \lim_n n F_Y(x_n) = f(x), \quad (\text{III.125})$$

$$\exists g, \quad \lim_n n F_Z(x_n) = g(x). \quad (\text{III.126})$$

On peut d'ores et déjà noter que ces équations correspondent aux conditions standards d'existence d'une distribution limite pour le maximum des Y et celui des Z . Les lois limites de cumulative $F_Y^{\text{lim}}(x) = \exp(-f(x))$ et $F_Z^{\text{lim}} = \exp(-g(x))$ sont donc soit des lois limites standards soit une loi limite dégénérée, c'est-à-dire soit une masse ponctuelle en 0, soit une "masse ponctuelle" en $-\infty$.

Si on injecte ensuite ces limites (III.125) et (III.126) dans l'équation (III.124), on obtient la distribution limite F_α pour le chemin $\Gamma^{(l)}$

$$F_{\alpha,n}(x_n) \xrightarrow[n \rightarrow +\infty]{} F_\alpha^{\text{lim}}(x) \equiv e^{\alpha f(x) + (1-\alpha)g(x)}. \quad (\text{III.127})$$

On peut ensuite intégrer sur l'ensemble des valeurs de α pour obtenir la fonction cumulative du maximum dans le cas général

$$F^{\text{lim}}(x) = \int_0^1 e^{\alpha f(x) + (1-\alpha)g(x)} d\alpha, \quad (\text{III.128})$$

ce qui conduit à

$$F^{\text{lim}}(x) = \frac{e^{f(x)} - e^{g(x)}}{f(x) - g(x)} = \frac{f+g}{2} + \frac{f^2 + fg + g^2}{6} + \sum_{k=3}^{+\infty} \frac{\sum_{l=0}^k f^l g^{k-l}}{(k+1)!}. \quad (\text{III.129})$$

On obtient donc une loi limite qui est un mélange continu des deux lois limites standards (mais potentiellement dégénérées) F_Y^{lim} et F_Z^{lim} .

5.2 Lois limites pour le maximum

On peut ensuite caractériser l'ensemble des lois limites et les domaines d'attractions correspondant. Premièrement, les conditions (III.125) et (III.126) ne peuvent être vérifiées que si Y et Z appartiennent à un domaine d'attraction standard.

Appelons ζ_Y et ζ_Z les paramètres de ces domaines d'attraction. On peut supposer sans perte de généralité que $\zeta_Y \geq \zeta_Z$. La difficulté est alors de trouver une suite de normalisation (a_n, b_n) qui soit compatible avec X et Z simultanément. Pour ce faire, considérons les suites de normalisation classiques (a_n, b_n) et (c_n, d_n) associées respectivement à Y et Z , i.e.

$$\begin{aligned} \frac{M_Y(n) - b_n}{a_n} &\xrightarrow{\mathcal{D}} \mathcal{F}_{\zeta_Y}, \\ \frac{M_Z(n) - d_n}{c_n} &\xrightarrow{\mathcal{D}} \mathcal{F}_{\zeta_Z}. \end{aligned} \quad (\text{III.130})$$

Trois cas se présentent alors : les suites de normalisation peuvent être compatibles entre elles, une des suites peut dominer l'autre ou il est également possible qu'aucune des deux ne domine l'autre sans qu'elles soient compatibles. Elles peuvent par exemple se dominer alternativement. Dans ce cas, il n'existe pas de loi limite correspondante.

Compatibilité :

Les suites (a_n, b_n) et (c_n, d_n) sont compatibles entre elles si

$$\begin{cases} \lim_{n \rightarrow +\infty} \frac{a_n}{c_n} = \gamma \in \mathbb{R}^{+*}, \\ \lim_{n \rightarrow +\infty} \frac{b_n - d_n}{a_n} = \delta \in \mathbb{R}. \end{cases} \quad (\text{III.131})$$

Si deux suites de normalisation sont compatibles, on a pour une suite de variables aléatoires X_n

$$\frac{X_n - a_n}{b_n} \xrightarrow{\mathcal{D}} \mathcal{L} \Leftrightarrow \frac{X_n - c_n}{d_n} \xrightarrow{\mathcal{D}} \gamma(\mathcal{L} + \delta). \quad (\text{III.132})$$

En d'autres termes, si deux suites de normalisation sont compatibles, elles mènent à la même distribution limite à une dilatation et une translation près. Les suites de normalisation (a_n, b_n) ne peuvent être compatibles entre elles que si $\zeta_Y = \zeta_Z$. Cependant cette condition n'est pas suffisante. L'équation (III.131) implique que les suites (a_n, b_n) et (c_n, d_n) ne sont compatibles entre elles que si

$$\begin{cases} \lim_{x \rightarrow +\infty} \frac{\overline{F}_Y}{\overline{F}_Z}(x) = c \in \mathbb{R} & \text{si } \zeta > 0. \\ y_f = x_f \text{ et } \overline{F}_Y^{-1}(x) \underset{t \rightarrow y_f}{\sim} \overline{F}_Z^{-1}(t) & \text{si } \zeta = 0. \\ y_f = x_f \text{ et } \lim_{x \rightarrow y_f} \frac{\overline{F}_Y}{\overline{F}_Z}(x) = c \in \mathbb{R} & \text{si } \zeta < 0. \end{cases} \quad (\text{III.133})$$

Dans ce cas de figure, la distribution limite sera donc un mélange continu d'une loi limite standard et d'une version translatée ou dilatée de celle-ci. Plus précisément, si $\zeta > 0$, on peut choisir comme suite de normalisation $(a_n = \overline{F}_Y^{-1}(1/n), b_n = 0)$; la loi limite de Z est donc forcément une dilatation de la loi limite de Y . On obtient un résultat identique si on considère le cas $\zeta < 0$. En effet, pour $\zeta < 0$, on peut choisir comme suite de normalisation $(a_n = y_f - \overline{F}_Y^{-1}(1/n), b_n = y_f)$, d'où il résulte que

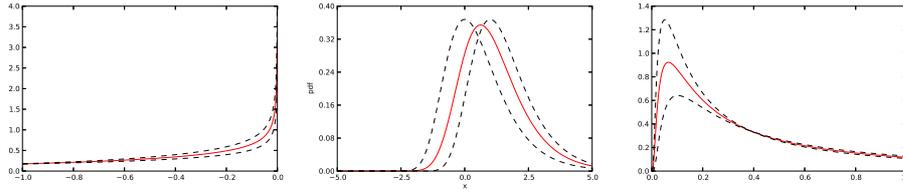


FIGURE 8.3 –

Illustration des lois limites non standards (en trait plein rouge) pour $\zeta = -1.5$ à gauche, $\zeta = 0$ au centre et $\zeta = 2$ à droite. Les lois non standards résultent d'un mélange continu d'une loi standard et d'une version dilatée (pour les lois de type Fréchet ou Weibull) ou translatée (pour les lois de type Gumbel) de celle-ci. Ces lois sources sont représentées en lignes tiretées noires sur la figure.

deux suites de normalisation compatibles ne peuvent différer asymptotiquement que par une dilatation. Au contraire, dans le cas Gumbel, la condition (III.133) implique qu'asymptotiquement pour deux suites de normalisation (a_n, b_n) et (c_n, d_n) , $a_n \sim c_n$. Les lois limites ne peuvent donc différer que par une translation.

Pour résumer dans le cas $\zeta \neq 0$, $M(X)$ tend vers une distribution limite qui est un mélange continu d'une loi limite classique \mathcal{F}_ζ et de sa version dilatée $\gamma\mathcal{F}_\zeta$. Pour $\zeta = 0$, la distribution limite de $M(X)$ est un mélange continu d'une loi Gumbel et d'une loi Gumbel translatée. Une illustration de ces lois limites est présentée sur la Fig. 8.3.

Domination :

Inversement, si on a $\zeta_Y > \zeta_Z$, et que X et Z partagent le même point de terminaison, la variable Y va dominer Z . Plus précisément, on aura

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{c_n}{a_n} &= 0, \\ \lim_{n \rightarrow +\infty} \frac{b_n - d_n}{d_n} &= 0. \end{aligned} \quad (\text{III.134})$$

Dans le cas $\zeta_Y > 0$ et $\zeta_Z > 0$, cela implique

$$\frac{Z_n - b_n}{a_n} \xrightarrow{\mathcal{D}} \delta_0 \quad (\text{III.135})$$

où δ_0 est une masse ponctuelle en 0. En d'autres termes, la suite (a_n, b_n) replie sur un Dirac les suites de variables aléatoires qui sont normalisables par (c_n, d_n) . Dans le cas général, la cumulative de Z va converger vers la cumulative dégénérée

$$\lim_{n \rightarrow +\infty} F_{(Z-b_n)/a_n}(x) = 1. \quad (\text{III.136})$$

Si Y domine X , on va donc dans ce cas obtenir une loi limite de la forme

$$F^{\text{lim}}(x) = \frac{e^{f(x)}}{f(x)}. \quad (\text{III.137})$$

Il s'agit donc finalement d'un mélange continu entre une loi limite standard et la loi dégénérée.

On peut remarquer que si Y et Z ont des points de terminaison finis y_f et z_f , Y domine Z implique que $y_f \geq z_f$. Il faut faire cependant attention dans ce cas. Si la loi de la variable de transition X_l possède un point de terminaison $x_f > y_f$ et $x_f > z_f$, $M(X)$ va alors converger en probabilité vers

$$M(X) \xrightarrow{P} \max\{X_l, y_f, z_f\}. \quad (\text{III.138})$$

Il s'agit d'un cas pathologique dans lequel la variable aléatoire de transition X_l joue un rôle prépondérant dans la convergence en distribution du maximum, et ce malgré son unique apparition.

5.3 Normalisation de la somme

On peut poursuivre le même raisonnement dans le cas de somme. Pour commencer, on peut réécrire \mathfrak{S} en séparant les blocs de variables Y et Z

$$\mathfrak{S}(X) = \sum_{i=1}^{\alpha n} Y_i + X_l + \sum_{i=1}^{(1-\alpha)n} Z_i. \quad (\text{III.139})$$

Le comportement de la somme vis-à-vis de la normalisation diffère cependant du maximum. En effet, si on normalise \mathfrak{S} par $(\mathfrak{S} - b_n)/a_n$, on obtient

$$\frac{\mathfrak{S}(X) - b_n}{a_n} = \frac{\sum_{i=1}^{\alpha n} Y_i - \beta_n^Y}{a_n} + \frac{X_l - \beta_n^{X_l}}{a_n} + \frac{\sum_{i=1}^{(1-\alpha)n} Z_i - \beta_n^Z}{a_n} \quad (\text{III.140})$$

avec $\beta^Y + \beta^{X_l} + \beta^Z = b_n$. Deux points contrastent fortement avec la normalisation du maximum. Premièrement, la variable esseulée X_l sera forcément négligeable par rapport à la somme de n termes lorsque $n \rightarrow +\infty$. Deuxièmement, il est possible de séparer la constante de normalisation b_n en trois sous-facteurs de normalisation β^Y , β^{X_l} et β^Z . Le choix de ces facteurs de normalisation demeure cependant assez simple. Supposons en effet, que Y et Z soient de moyennes finies. Dans ce cas, si on souhaite normaliser Y et X en utilisant les théorèmes de convergence vers les lois stables, il est nécessaire que

$$\begin{aligned} \beta_n^Y &= \alpha n \mathbb{E}[Y]. \\ \beta_n^Z &= (1 - \alpha)n \mathbb{E}[Z]. \end{aligned} \quad (\text{III.141})$$

Or, on doit avoir $\beta_n^Y + \beta_n^Z = b_n + o(n)$ et $\partial_\alpha b_n = 0$. Cela implique donc que

$$\mathbb{E}[Y] = \mathbb{E}[Z]. \quad (\text{III.142})$$

On ne peut donc utiliser les théorèmes de convergence vers lois stables pour des lois de moyenne finie que si $\mathbb{E}[Y] = \mathbb{E}[Z]$ et dans ce cas, il n'existe qu'un seul choix acceptable pour β^Y et β^Z . Au contraire, si une des lois ne possède pas de moyenne finie, il est possible de la normaliser en utilisant une suite de la forme $(a_n = n^{1/\zeta} L(n), b_n = 0)$ avec $\zeta < 1$ et L une fonction lentement variable. Il est donc inutile de tenir compte des facteurs de normalisation additifs β qui seront effacés par la normalisation en a_n .

5.4 Lois limites pour la somme

Lois limites pour des moyennes concordantes ou infinies

Si toutes les lois sont de moyennes finies et concordantes ou si au moins une des lois est de moyenne infinie, le comportement de la limite va dépendre du comportement relatif de la queue des distributions Y et Z . Appelons (a_n, b_n) et (c_n, d_n) les suites de normalisation classiques de la somme associées aux variables Y et Z . On peut déjà remarquer que Y ne peut être compatible avec Z (cf. équation (III.131)) que si les deux variables possèdent la même moyenne, ou si au moins une des deux variables est de moyenne infinie.

Compatibilité :

Si (a_n, b_n) et (c_n, d_n) sont compatibles (i.e. $\mathbb{E}[Y] = \mathbb{E}[Z]$ et $\lim_n a_n/c_n = \gamma$), il est alors possible de normaliser simultanément Y et Z en utilisant la suite classique de normalisation de Y . On peut remarquer que si Y et Z sont de variances finies, a_n et c_n sont forcément compatibles. Inversement, si a_n et c_n sont compatibles Y et Z appartiennent au même domaine d'attraction ζ (la réciproque est, elle, fautive). De plus, on sait que la suite de normalisation de Y sera de la forme $a_n = L(n)n^{1/\zeta}$, $b_n = \mathbb{E}[X]n$ avec $0 < \zeta \leq 2$ le paramètre de la loi stable limite. En utilisant le fait que dans le cas symétrique normalisé la fonction caractéristique de la loi stable \mathcal{S}_ζ (cf. définition (I.9)) est $\phi_\zeta(z) = \exp(-|z|^\zeta)$, on peut alors exprimer la fonction caractéristique limite de X par

$$\phi_\alpha^{(X)}(z) = (\phi_\zeta(z))^\alpha (\phi_\zeta(\gamma z))^{1-\alpha}. \quad (\text{III.143})$$

Fonctions caractéristiques :

On peut ensuite intégrer sur l'ensemble des valeurs de α pour obtenir la fonction caractéristique de $\mathfrak{S}(X)$ normalisée dans le cas général

$$\phi_l^{(X)}(x) = \int_0^1 (\phi_\zeta(z))^\alpha (\phi_\zeta(\gamma z))^{1-\alpha} d\alpha. \quad (\text{III.144})$$

Ce qui conduit à

$$\phi_l(z) = \frac{\phi_\zeta(z) - \phi_\zeta(\gamma z)}{|\ln \phi_\zeta(z)| - |\ln \phi_\zeta(\gamma z)|}. \quad (\text{III.145})$$

La fonction caractéristique de la loi limite est donc un mélange continu de la fonction caractéristique classique et d'une version dilatée de celle-ci. On peut mettre en miroir ce résultat avec l'équation (III.129) où la fonction cumulative limite du maximum est un mélange continu de la fonction cumulative limite classique et d'une version dilatée ($\zeta \neq 0$) ou translatée ($\zeta = 0$) de celle-ci. Les comportements de la somme et du maximum sont donc encore une fois similaires sans être identiques.

Domination :

Dans le cas de la somme, les points de terminaison ne jouent pas de rôle particulier, il est donc impossible d'effacer des variables aléatoires comme dans le cas du maximum.

Un second sous-cas apparaît si la suite (a_n, b_n) domine la suite (c_n, d_n) (ou l'inverse). Dans ce cas, afin de normaliser la somme des Y , il est nécessaire de réduire la somme Z à un Dirac. On obtient donc une loi limite qui est un mélange continu de ce Dirac et de la fonction limite classique

$$\phi_l(z) = \frac{\phi_\zeta(z) - 1}{|\ln \phi_\zeta(z)|}. \quad (\text{III.146})$$

5.5 Rupture de la loi des grands nombres

Que se passe-t-il dans le cas des lois de moyennes finies mais discordantes? Si $\mathbb{E}[X] \neq \mathbb{E}[Y]$, nous avons vu qu'il est impossible d'utiliser les théorèmes de convergence vers les lois stables pour normaliser $\mathfrak{S}(X)$. Face à cette limitation, il est naturel de s'en remettre à la loi des grands nombres. La loi des grands nombres implique que pour une chaîne $\Gamma^{(l)}$,

$$\lim_{n \rightarrow +\infty} \frac{\mathfrak{S}(\Gamma^{(l)})}{n} \stackrel{\text{p.s.}}{=} \alpha \mathbb{E}[Y] + (1 - \alpha) \mathbb{E}[Z]. \quad (\text{III.147})$$

La loi des grands nombre implique donc que la moyenne de X à Γ fixé dépend de α qui est distribué uniformément sur $[0, 1]$. La moyenne de X converge donc en distribution vers une distribution non dégénérée.

Plus précisément, on peut traduire l'équation (III.147) en terme de convergence en distribution par

$$\frac{\mathfrak{S}(\Gamma^{(l)})}{n} \xrightarrow{\mathcal{D}} \delta(x - \alpha \mathbb{E}[Y] + (1 - \alpha) \mathbb{E}[Z]). \quad (\text{III.148})$$

En intégrant cette distribution limite sur α , on obtient une distribution limite générale

$$p(x) = \frac{1}{|\mathbb{E}[Y] - \mathbb{E}[Z]|} \begin{cases} 1 & \text{si } x_{\min} < x < x_{\max}, \\ 0 & \text{sinon.} \end{cases} \quad (\text{III.149})$$

où $x_{\min} = \min\{\mathbb{E}[Y], \mathbb{E}[Z]\}$ et $x_{\max} = \max\{\mathbb{E}[Y], \mathbb{E}[Z]\}$. La moyenne de X suit donc une distribution uniforme sur l'intervalle $[x_{\min}, x_{\max}]$. La loi des grands nombres est donc fortement affaiblie dans ce contexte des chaînes réductibles puisqu'on passe d'une convergence presque sûre de la moyenne vers une quantité déterministe à une convergence en distribution vers une distribution uniforme. Cette instabilité de la moyenne de X peut être vue comme une conséquence à la fois de la structure de corrélation de X et de sa nature non stationnaire. Le fait que toutes les chaînes $\Gamma^{(l)}$ soient équiprobables signifie qu'il est impossible de prévoir de manière quasi-déterministe la moyenne de X parce que celle-ci peut osciller librement entre la moyenne de Y et celle de Z .

6 Chaînes réductibles simples à d états

6.1 Définitions

La prochaine étape est d'étendre nos résultats à l'ensemble des chaînes réductibles simples, c'est-à-dire l'ensemble des chaînes dont la matrice de structure \mathcal{E} peut être écrite sous la forme

$$\mathcal{E} = \text{Id} + H, \quad (\text{III.150})$$

où H est une matrice nilpotente d'ordre $p + 1 \in \mathbb{N}$:

$$p + 1 = \min\{p, H^p = 0, p \in \mathbb{N}\}. \quad (\text{III.151})$$

Cette décomposition de la matrice \mathcal{E} nous renseigne principalement sur ses propriétés algébriques. Afin d'obtenir une idée plus concrète de la forme de la matrice \mathcal{E} ,

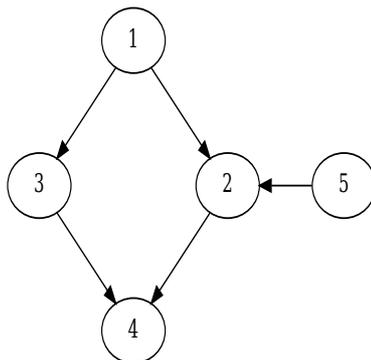


FIGURE 8.4 – Exemple de matrice \mathcal{E} de la forme $\text{Id} + H$. Seuls les liens non nuls sont représentés.

nous allons montrer que le graphe orienté $G(H)$ des coefficients non-nuls de H est nécessairement acyclique. En effet, si \mathcal{E} est une matrice non négative ($\forall i, j, \mathcal{E}_{i,j} \geq 0$) la matrice H doit vérifier

$$\begin{aligned} \forall i, \quad H_{i,i} &\geq -1, \\ \forall i, j, \quad i \neq j, \quad H_{i,j} &\geq 0. \end{aligned} \tag{III.152}$$

De plus, puisque $H^p = 0$, la seule valeur propre de H est 0. Une conséquence directe est alors que $\text{tr}(H^2) = 0$, ce qui implique

$$\sum_i H_{i,i}^2 + \sum_{i,j,i \neq j} H_{i,j} H_{j,i} = 0. \tag{III.153}$$

D'après l'équation (III.152), tous les termes de gauche de l'équation (III.153) sont positifs. Par conséquent, les coefficients diagonaux $H_{i,i}$ sont tous nuls et, pour tout $i \neq j$, soit $H_{i,j} = 0$ soit $H_{j,i} = 0$: la matrice H ne contient donc pas d'aller-retour (cycle d'ordre 2) entre deux états. La matrice H est donc une matrice nilpotente à coefficients positifs, et il est bien connu que dans ce cas le graphe orienté associé $G(H)$ est acyclique¹.

En d'autres termes, le graphe $G(H)$ est composé d'un ensemble de transitions allant toujours dans le même sens, sans retour en arrière possible. La figure 8.4 donne un exemple de tel graphe. L'existence d'un sens de parcours du graphe –conséquence du caractère acyclique– implique également l'existence d'une permutation des noms des indices telle que $H_{i,j}$ soit triangulaire supérieure (i.e. $H_{i,j} \neq 0 \Rightarrow j > i$, les arêtes du graphe relient toujours un nœud i à un nœud d'indice supérieur). Il est donc possible de considérer uniquement les matrices H triangulaires supérieures sans perte de généralité.

6.2 Séparation des transitions

Considérons donc un chemin Γ de la chaîne de Markov cachée. La décomposition de la matrice de structure en $\mathcal{E} = \text{Id} + H$ signifie que Γ peut suivre deux types de

1. Simplement parce qu'en présence d'un cycle de longueur k , on devrait avoir pour tout p entier, $H^{kp} \geq 0$

transition : une transition de répétition utilisant la partie Id de \mathcal{E} et une transition irréversible utilisant la matrice nilpotente H . Il est donc possible de séparer Γ en un schéma de répétition R et un chemin irréversible \mathcal{T} de longueur r en posant

$$\Gamma = \left(\underbrace{\mathcal{T}_1, \dots, \mathcal{T}_1}_{R_1 \text{ fois}}, \dots, \underbrace{\mathcal{T}_k, \dots, \mathcal{T}_k}_{R_k \text{ fois}}, \dots, \underbrace{\mathcal{T}_r, \dots, \mathcal{T}_r}_{R_r \text{ fois}} \right). \quad (\text{III.154})$$

Il faut cependant que r et R_k vérifient

$$r \leq p + 1, \quad (\text{III.155})$$

$$\sum_{k=1}^r R_k = n + 1. \quad (\text{III.156})$$

L'intérêt de cette décomposition réside dans le fait que la probabilité $\kappa(\Gamma)$ de la chaîne Γ ne dépend que de \mathcal{T} :

$$\begin{aligned} \kappa(\Gamma) &= \frac{1}{Z} A_{\Gamma_1, \Gamma_{n+1}} \prod_{i=1}^n \mathcal{E}_{\Gamma_i, \Gamma_{i+1}} \\ &= \frac{A_{\mathcal{T}_1, \mathcal{T}_r}}{Z} \prod_{i=1}^{r-1} \mathcal{E}_{\mathcal{T}_i, \mathcal{T}_{i+1}} = \kappa(\mathcal{T}). \end{aligned} \quad (\text{III.157})$$

Toutes les chaînes Γ qui partagent le même chemin irréversible \mathcal{T} sont donc équiprobables. De plus, le nombre $N_R(\mathcal{T})$ de chaînes de même chemin irréversible \mathcal{T} est simplement

$$N_R(\mathcal{T}) = \binom{n}{r}. \quad (\text{III.158})$$

La probabilité d'observer un chemin irréversible \mathcal{T} est alors

$$p(\mathcal{T}) = N_R(\mathcal{T}) \kappa(\mathcal{T}). \quad (\text{III.159})$$

L'Eq. (III.159) montre que cette probabilité $p(\mathcal{T})$ ne dépend que de \mathcal{T} et $\binom{n}{r}$ (avec $r \leq p + 1 \leq d$). Par conséquent, dans la limite $n \rightarrow +\infty$, la statistique de Γ est dominée par la classe des chemins irréversibles de longueur maximale r_{\max} . Si nous supposons que $\mathcal{L}(H^p) > 0$, cela signifie que $r_{\max} = p + 1$. Dans cette limite asymptotique, si nous appelons $\mathcal{C}_{\mathcal{T}}$ la classe des chaînes Γ de chemin irréversible \mathcal{T} et $|\mathcal{T}| \equiv r$ la longueur du chemin irréversible, nous avons donc

$$p(\Gamma \in \mathcal{C}_{\mathcal{T}}) = \delta_{|\mathcal{T}|, p+1} \frac{\kappa(\mathcal{T})}{Z'}, \quad (\text{III.160})$$

$$\text{où } Z' = \sum_{|\mathcal{T}|=p+1} p(\mathcal{T}). \quad (\text{III.161})$$

Combinés avec l'Eq. (III.94), ces résultats signifient qu'il suffit d'étudier tout d'abord les distributions limites pour chacune des classes de chemins irréversibles avant de les recombinaison pour obtenir la distribution limite générale. La figure 8.5 montre l'exemple de la séparation du graphe de la figure 8.4 en trois chemins irréversibles de longueur maximale.

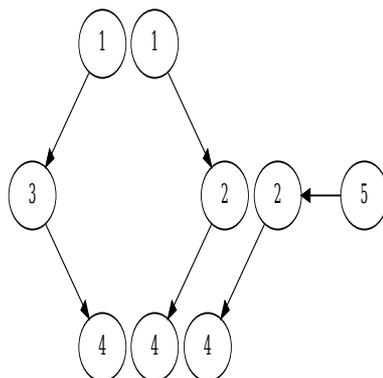


FIGURE 8.5 – Séparation du graphe de la figure 8.4 en ses trois chemins irréversibles de longueur maximale.

6.3 Distribution limite le long d'un chemin irréversible

Considérons donc le comportement asymptotique de la fonction naturelle \mathfrak{F} sous l'hypothèse que $\Gamma \in \mathcal{C}_{\mathcal{T}}$ pour un chemin irréversible \mathcal{T} de longueur maximale. Nous savons qu'à l'intérieur de la classe tous les schémas de répétition R_k sont équiprobables. De plus, pour un schéma R_k donné, la fonction naturelle \mathfrak{F}_R associée est

$$\mathfrak{F}_R(x) = \prod_{k=1}^r \mathfrak{F}_{T_k, T_k}^{R_k}(x). \quad (\text{III.162})$$

Si nous appelons $\underline{\alpha} = \underline{R}/n$ le taux de répétition avec $\underline{R} = \{R_k\}$, nous avons

$$\mathfrak{F}_{\underline{\alpha}}(x) = \prod_{k=1}^r \mathfrak{F}_{T_k, T_k}^{n\alpha_k}(x). \quad (\text{III.163})$$

En d'autres termes, comme dans la section 5, il est possible de séparer la somme (respectivement le maximum) en blocs de sommes (respectivement de maxima) de variables i.i.d.. Nous avons cependant p blocs à la place des deux blocs de la section 5. Afin de normaliser ces p blocs, il est donc nécessaire de trouver la plus *grande* suite de normalisation parmi les suites de normalisation classiques des variables \mathcal{P}_{T_k, T_k} , c'est-à-dire une suite de normalisation qui permet à au moins une distribution de converger vers une loi limite classique tout en assurant la convergence en distribution de l'ensemble des autres variables vers une distribution potentiellement dégénérée. Il ne faut pas oublier non plus que cette suite de normalisation doit être valide quel que soit \mathcal{T} . Si on considère

$$\{Y_k, Y_k \stackrel{\mathcal{D}}{=} \mathcal{P}_{k,k} \text{ et } k \in \{\mathcal{T}_i, |\mathcal{T}| = p + 1\}\} \quad (\text{III.164})$$

l'ensemble des variables aléatoires qui apparaissent comme des répétitions le long d'un chemin irréversible de longueur maximale, il est donc nécessaire de considérer la suite de normalisation extrême sur l'ensemble $\{Y_k\}$. Fort heureusement, il est possible d'adapter les idées développées dans la section 5 pour donner corps à cette notion de plus grande suite de normalisation. Pour ce faire, nous pouvons définir deux ordres partiels $>_M$ et $>_{\mathfrak{S}}$ en utilisant les notions de domination de suite de

normalisation établie par l'équation (III.134). Plus précisément, nous noterons $Y >_M Z$ (respectivement $Y >_{\mathfrak{S}} Z$) si la suite de normalisation classique du maximum (respectivement de la somme) associée à Y domine celle associée à Z . Il s'agit d'un ordre partiel qui nous permet de définir les variables extrémales $\{Z_k\}$ de l'ensemble $\{Y_k\}$:

$$\{Z_k\} = \{Y_k\} \setminus \{Y_k, \exists l, Y_l > Y_k\}. \quad (\text{III.165})$$

Par définition de l'ensemble extrémal $\{Z_k\}$, aucune des variables Z_k n'est dominée par une autre variable Z_l . Il existe donc deux possibilités distinctes : soit les $\{Z_k\}$ sont compatibles entre elles, et il est alors possible de normaliser simultanément tous les Y_k , soit elles ne sont pas compatibles (par exemple, il est possible que deux variables Z se dominant alternativement) et dans ce cas, il ne peut y avoir de convergence en distribution de $M(X)$ ou de $\mathfrak{S}(X)$ en utilisant les théorèmes de convergence vers les lois stables (ou max-stables). Cependant dans le cas de la somme, si toutes les lois Z sont de moyennes finies, il est possible d'utiliser la loi des grands nombres pour obtenir une convergence en distribution.

Compatibilité :

Si les Z_k sont bien compatibles entre elles, il suffit de choisir une suite de normalisation parmi les suites de normalisation classiques Z_k . Pour chacune des variables Y_k , il existe alors une fonction naturelle \mathfrak{F}_k telle que la fonction naturelle limite de X le long de \mathcal{T} soit

$$\mathfrak{F}_{\mathcal{T},\alpha}(x) = \prod_{k=1}^r \mathfrak{F}_{\mathcal{T}_k}^{\alpha_k}(x). \quad (\text{III.166})$$

Si Y est une des variables extrémales de $\{Z\}$, \mathfrak{F}_k sera une des fonctions naturelles (cumulative ou fonction caractéristique) limites classiques, potentiellement dilatée ou translatée. Sinon, la normalisation aura réduit \mathfrak{F}_k à une fonction naturelle dégénérée.

Pour un schéma de répétition donné $\underline{\alpha}$, la loi limite du maximum ou de la somme est donc caractérisée par sa fonction naturelle $\mathfrak{F}_{\mathcal{T},\underline{\alpha}}$. Afin d'obtenir la distribution limite globale au sein de la classe $\mathcal{C}_{\mathcal{T}}$, il suffit alors d'intégrer sur l'ensemble des valeurs possibles de $\underline{\alpha}$

$$\mathfrak{F}_{\mathcal{T}}(x) = \int_{\sum_{k=1}^r \alpha_k=1} F_{\alpha}^{\text{lim}}(x) \prod_{k=1}^r d\alpha_k. \quad (\text{III.167})$$

Pour calculer cette intégrale, nous utilisons le changement de variable

$$\beta_i = \sum_{k \leq i} \alpha_k, \quad (\text{III.168})$$

et obtenons

$$\begin{aligned} \mathfrak{F}_{\mathcal{T}}(x) &= \int_{0=\beta_0 < \beta_1 < \dots < \beta_r=1} \prod_{i=1}^r e^{-(\beta_i - \beta_{i-1}) |\ln \mathfrak{F}_{\mathcal{T}_i}|} d\beta \\ &= \int_{0 < \beta_1 < \dots < \beta_{r-1} < 1} \prod_{i=1}^r e^{\beta_i (|\ln \mathfrak{F}_{\mathcal{T}_{i+1}}| - |\ln \mathfrak{F}_{\mathcal{T}_i}|)} \prod_{i=1}^{r-1} d\beta_i. \end{aligned} \quad (\text{III.169})$$

Quelques calculs par récurrence permettent de simplifier cette expression en

$$\mathfrak{F}_{\mathcal{T}} = \sum_{k=1}^r \frac{\mathfrak{F}_{\mathcal{T}_k}(x)}{\prod_{j \neq k} (|\ln \mathfrak{F}_{\mathcal{T}_j}(x)| - |\ln \mathfrak{F}_{\mathcal{T}_k}(x)|)}. \quad (\text{III.170})$$

La distribution limite de X le long du chemin irréversible \mathcal{T} est donc un mélange continu de lois limites classiques (potentiellement dilatées dans le cas de la somme ou du maximum pour $\zeta \neq 0$ ou translatées pour le maximum et $\zeta = 0$) et de lois dégénérées.

La distribution limite de $M(X)$ ou $\mathfrak{S}(X)$ est alors un mélange discret de ces distributions continûment mélangées et prend la forme

$$\mathfrak{f}_l = \sum_{|\mathcal{T}|=p+1} p(\mathcal{T}) \mathfrak{f}_{\mathcal{T}}. \quad (\text{III.171})$$

On peut remarquer que cette somme porte sur un nombre fini de termes puisque le nombre maximum de chemins irréversibles distincts de longueur $p + 1$ dans un graphe comportant d nœuds est $\binom{d}{p+1}$.

Rupture de la loi des grands nombres :

Dans le cas de la somme \mathfrak{S} , si toutes les variables Z sont de moyennes finies, mais différentes, les variables Z seront nécessairement incompatibles. Cependant dans ce cas, en utilisant la loi des grands nombres, il est facile de montrer que

$$\frac{\mathfrak{S}(X|\mathcal{T}, \alpha)}{n} \xrightarrow{\mathcal{D}} \delta \left(x - \sum_k \alpha_k \mathbb{E}[Y_{\mathcal{T}_k}] \right). \quad (\text{III.172})$$

Il faut noter que dans ce cas la distinction entre variables extrémales Z_k et variables non extrémales n'est plus significative. Si on intègre ensuite sur α pour obtenir la distribution de $\mathfrak{S}(X|\mathcal{T})$, on obtient la densité de probabilité

$$p_{\mathcal{T}}(x) = \int_{\sum \alpha_k=1} \delta \left(x - \sum_k \alpha_k \mathbb{E}[Y_{\mathcal{T}_k}] \right) \prod_k d\alpha_k. \quad (\text{III.173})$$

Cette densité de probabilité est donc un polynôme par morceaux dont le support correspond à l'enveloppe convexe des espérances $\{\mathbb{E}[Y_{\mathcal{T}_k}]\}$. Si on somme alors sur les chemins irréversibles possibles de longueur maximale, on obtient

$$p_l(x) = \sum_{|\mathcal{T}|=p+1} p(\mathcal{T}) p_{\mathcal{T}}(x). \quad (\text{III.174})$$

Comme dans le cas précédent, il s'agit d'un polynôme par morceaux dont le support est l'enveloppe convexe des espérances $\mathbb{E}[Y_k]$.

7 Cas général

Nous allons maintenant décrire le comportement général de \mathfrak{S} et M pour des processus X à représentation matricielle en utilisant l'intuition que nous avons obtenue en section 4 et 6. La section 4 nous a appris que si la chaîne de Markov cachée est suffisamment mélangeante, le maximum et la somme ne sont pas sensibles à la structure de corrélation, et les distributions limites sont préservées. Dans la section 6, nous avons pu montrer que dans le cas des matrices de structure de la forme $\mathcal{E} = \text{Id} + H$ avec H nilpotente, le comportement global de la chaîne est hautement instable à cause de la présence de transitions irréversibles et cette instabilité fait

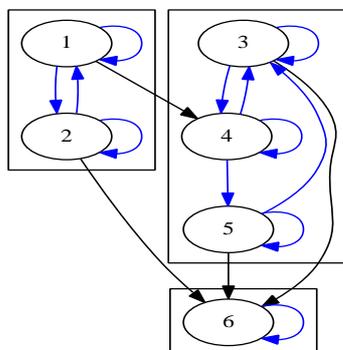


FIGURE 8.6 – Exemple d'un cas générique, les nœuds appartenant à une même composante fortement connexe sont encadrés. Les transitions réversibles sont en bleues, les irréversibles en noir.

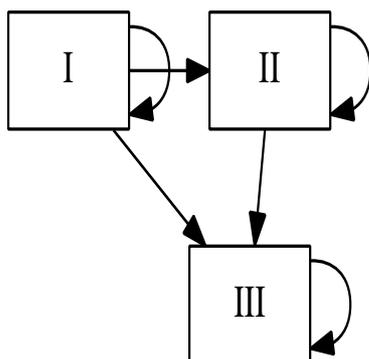


FIGURE 8.7 – Réduction du graphe de la figure 8.6 en un graphe irréversible simple. La structure interne de chaque composante connexe est oubliée et on ne conserve que les arêtes reliant les différentes composantes.

apparaître de nouvelles lois limites qui correspondent à des mélanges continus entre des lois limites classiques de même paramètre. Nous allons montrer qu'il est en fait possible dans le cas général de se ramener au cas des chaînes réductibles simples.

Pour ce faire, nous allons utiliser la décomposition en blocs irréductibles qui est un des corollaires du théorème de Perron-Frobenius. Considérons notre matrice de structure \mathcal{E} et le graphe orienté $G(\mathcal{E})$ associé. Nous dirons que i est connecté à j , ce que nous noterons $i \xrightarrow{c} j$, s'il existe un chemin orienté de i à j dans $G(\mathcal{E})$, i.e.,

$$(i \xrightarrow{c} j) \Leftrightarrow (\exists k, (\mathcal{E}^k)_{i,j} > 0). \quad (\text{III.175})$$

Nous souhaitons identifier dans \mathcal{E} les transitions irréversibles qui apparaissaient dans la section 4. Pour ce faire, il est plus simple de commencer par caractériser les transitions réversibles. Nous dirons qu'il existe une transition réversible entre i et j , noté $i \xleftrightarrow{c} j$, si

$$(i \xleftrightarrow{c} j) \Leftrightarrow (i \xrightarrow{c} j \text{ et } j \xrightarrow{c} i) \quad (\text{III.176})$$

La relation binaire \xleftrightarrow{c} est une relation d'équivalence (transitive, réflexive et symétrique). Les classes \mathcal{C}_k associées à cette relation d'équivalence sont appelées les composantes fortement connexes du graphe $G(\mathcal{E})$, il s'agit des classes d'états au sein desquelles les transitions sont toujours réversibles. Les transitions irréversibles sont alors les transitions entre des états qui appartiennent à des classes différentes. Cela signifie que la relation de connexion $i \xrightarrow{c} j$ induit une relation d'ordre sur l'ensemble des classes. Si on définit

$$\mathcal{C}_i \xrightarrow{c} \mathcal{C}_j \Leftrightarrow \exists i \in \mathcal{C}_i, j \in \mathcal{C}_j, \quad i \xrightarrow{c} j, \quad (\text{III.177})$$

alors \xrightarrow{c} est une relation d'ordre. La transitivité est immédiate, la réflexivité également. Le caractère antisymétrique résulte du fait que si on a $\mathcal{C}_i \xrightarrow{c} \mathcal{C}_j$ et $\mathcal{C}_j \xrightarrow{c} \mathcal{C}_i$ alors \mathcal{C}_i et \mathcal{C}_j sont fortement connectées, ce qui implique que $\mathcal{C}_i = \mathcal{C}_j$. La Fig. 8.6 donne un exemple de cette séparation d'un graphe réductible en ces composantes fortement connexes.

On peut alors utiliser ces informations sur la structure de $G(\mathcal{E})$ pour simplifier la forme de \mathcal{E} . En effet, si on applique une permutation telle que les états appartenant à une même classe soient adjacents et que les classes soient ordonnées suivant \xrightarrow{c} , on obtient nécessairement une matrice triangulaire supérieure par blocs. De plus, chacun des blocs diagonaux correspondant à la classe \mathcal{C}_k est irréductible. Si on suppose de plus que ces blocs sont également apériodiques², ils possèdent alors une valeur propre dominante ρ_k associée à un vecteur propre $\mu^{(k)}$ (choisi positif et tel que $\|\mu^{(k)}\|_2 = 1$). L'existence de cette valeur propre dominante implique qu'à l'intérieur des blocs, les chaînes Γ vont rapidement converger vers la mesure invariante. Il est donc possible d'oublier les détails du comportement de la chaîne à l'intérieur des composantes fortement connexes tout en conservant la même distribution limite.

Nous proposons dans la suite une méthode permettant de construire formellement une variable réduite où ces détails ont été effacés tout en conservant la même distribution limite de la somme ou du maximum. Pour commencer, nous décomposons le chemin Γ en une suite de r sous-chaînes $\Gamma^{(k)}$, de longueurs l_k et dont tous les

2. Le cas périodique se comporte essentiellement comme le cas apériodique dans le cadre de notre étude de la somme ou du maximum.

états appartiennent à même classe $\mathcal{C}_{\mathbf{g}_k}$, entrecoupés par des transitions irréversibles $\Delta_{2k} \rightarrow \Delta_{2k+1}$ de telle sorte que

$$\begin{aligned} \Gamma_1^{(k)} &= \Delta_{2k-1}, & \Gamma_{l_k}^{(k)} &= \Delta_{2k}, \\ \Gamma &= \Gamma^{(1)} \Delta_2 \Delta_3 \Gamma^{(2)} \dots \Gamma^{(r-1)} \Delta_{2r-2} \Delta_{2r-1} \Gamma^{(r)}. \end{aligned} \quad (\text{III.178})$$

On peut remarquer que pour \underline{l} et $\underline{\Delta}$ fixés, les sous-chaînes $\Gamma^{(k)}$ sont elles-mêmes des chaînes de Markov issues d'une représentation matricielle dont la matrice \mathcal{E} est le bloc diagonal de la classe $\mathcal{C}_{\mathbf{g}_k}$ et la matrice \mathcal{A} est $\mathcal{A}_{i,j}^{(k)} = \delta_{i,\Delta_{2k-1}} \delta_{j,\Delta_{2k}}$ – ce qui assure que les états initial et final de la chaîne sont bien Δ_{2k-1} et Δ_{2k} . Les sous-chaînes $\Gamma^{(k)}$ vont donc converger rapidement vers leur mesure invariante $\mu^{(k)}$, ce qui entraîne que X^\oplus ne dépend que de \underline{l} et $\underline{\mathbf{g}} \equiv \{\mathbf{g}_k\}$. Plus précisément, si on définit W_k comme le mélange naturel au sein de la classe \mathcal{C}_i , i.e.

$$\begin{aligned} \mathfrak{F}_{W_k}(z) &= \prod_{i,j \in \mathcal{C}_k} \mathfrak{F}_{\mathcal{P}_{i,j}}(z)^{s_{i,j}}, \\ \text{où } s_{i,j} &= \frac{\mathcal{E}_{i,j} \mu_j^{(k)}}{\rho^k \|\mu^{(k)}\|_1}. \end{aligned} \quad (\text{III.179})$$

la fonction naturelle de X^\oplus sera

$$\mathfrak{F}^{\underline{l}, \underline{\mathbf{g}}}(x) = \prod_{k=1}^r \mathfrak{F}_{W_{\mathbf{g}_k}}(z)^{l_k}. \quad (\text{III.180})$$

Pour caractériser le comportement de la somme ou du maximum, il suffit donc d'analyser le comportement de \underline{l} et $\underline{\mathbf{g}}$. Pour ce faire, il est plus facile de commencer par estimer la densité de probabilité des couples $(\underline{l}, \underline{\Delta})$. Si on appelle $\mathcal{E}^{(k)}$ le bloc diagonal de la classe \mathcal{C}_k , on peut noter qu'il est possible d'exprimer simplement les puissances de $\mathcal{E}^{(k)}$ à partir de $\mu^{(k)}$:

$$\left(\mathcal{E}^{(k)}\right)_{i,j}^m \approx \rho_k^m \mu_i^{(k)} \mu_j^{(k)}. \quad (\text{III.181})$$

Il est alors possible d'exprimer la distribution de $(\underline{l}, \underline{\Delta})$ en utilisant ces grandeurs matricielles, on obtient

$$p(\underline{l}, \underline{\Delta}) \underset{n \rightarrow +\infty}{\sim} \prod_k \rho_k^{l_k} \mu_{\Delta_1}^{(\mathbf{g}_1)} \mathcal{A}_{\Delta_1, \Delta_{2r}} \mu_{\Delta_{2r}}^{(\mathbf{g}_{2r})} \frac{\prod_{j=1}^r \mu_{\Delta_{2j}}^{(\mathbf{g}_j)} \mathcal{E}_{\Delta_{2j}, \Delta_{2j+1}} \mu_{\Delta_{2j+1}}^{(\mathbf{g}_{j+1})}}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.182})$$

Cette forme de $p(\underline{l}, \underline{\Delta})$ implique que seules les composantes fortement connexes de valeur propre ρ_k maximale contribuent de manière significative à la somme ou au maximum dans la limite $n \rightarrow +\infty$. Les classes de valeur propre ρ_k non maximale peuvent toutefois jouer le rôle d'étape transitoire entre des classes de valeur propre maximale. Cependant, pour simplifier le raisonnement attendant, nous nous restreindrons dans la suite au cas où toutes les valeurs propres ρ_k sont maximales ($\forall k, \rho_k = \rho$). On peut alors reconnaître dans l'expression de $p(\underline{l}, \mathcal{T})$ la probabilité d'une chaîne de Markov dérivée d'une représentation matricielle réductible simple, où les termes diagonaux correspondent à ρ et les termes non diagonaux correspondent aux $\mathcal{E}_{\Delta_i, \Delta_j}$. On peut

cependant simplifier davantage le problème en sommant l'équation (III.182) sur l'ensemble des transitions irréversibles Δ compatibles avec la suite de classe $\mathcal{C}_{\mathfrak{g}_k}$. Pour cela, on définit

$$\tilde{p}(\underline{l}, \underline{\mathfrak{g}}) = \sum_{\substack{\underline{\Delta} \\ \forall k, \Delta_{2k} \in \mathcal{C}_{\mathfrak{g}_k}}} p(\underline{l}, \underline{\Delta}). \quad (\text{III.183})$$

En injectant l'équation (III.182) dans l'équation (III.183), on obtient alors

$$p(\underline{l}, \underline{\mathfrak{g}}) = \rho^{n-r+1} \sum_{\substack{\underline{\Delta} \\ \forall k, \Delta_{2k} \in \mathcal{C}_{\mathfrak{g}_k}}} \mu_{\Delta_1}^{(\mathfrak{g}_1)} \mathcal{A}_{\Delta_1, \Delta_{2r}} \mu_{\Delta_{2r}}^{(\mathfrak{g}_{2r})} \frac{\prod_{j=1}^r \mu_{\Delta_{2j}}^{(\mathfrak{g}_j)} \mathcal{E}_{\Delta_{2j}, \Delta_{2j+1}} \mu_{\Delta_{2j+1}}^{(\mathfrak{g}_{j+1})}}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.184})$$

L'équation (III.184) présente une certaine ressemblance avec l'expression (III.37) de la densité de probabilité de la chaîne Γ . Pour accentuer cette ressemblance, on peut définir

$$\begin{aligned} \tilde{\mathcal{E}}_{i,i} &= \rho, \\ i \neq j, \quad \tilde{\mathcal{E}}_{i,j} &= \sum_{k \in \mathcal{C}_i, l \in \mathcal{C}_j} \mu_k^{(i)} \mathcal{E}_{k,l} \mu_l^{(j)}, \end{aligned} \quad (\text{III.185})$$

et

$$\tilde{A}_{i,j} = \sum_{k \in \mathcal{C}_i, l \in \mathcal{C}_j} \mu_k^{(i)} A_{k,l} \mu_l^{(j)}. \quad (\text{III.186})$$

On peut alors réécrire l'équation (III.184) sous la forme

$$\tilde{p}(\underline{l}, \underline{\mathfrak{g}}) = \left(\prod_k \tilde{\mathcal{E}}_{k,k}^{l_k} \right) \tilde{A}_{\mathfrak{g}_1, \mathfrak{g}_r} \frac{\prod_{j=1}^{r-1} \tilde{\mathcal{E}}_{\mathfrak{g}_j, \mathfrak{g}_{j+1}}}{\mathcal{L}(\mathcal{E}^n)}. \quad (\text{III.187})$$

À partir de ce point, il est possible de réinterpréter $\tilde{p}(\underline{l}, \underline{\mathfrak{g}})$ comme la densité de probabilité d'une chaîne réductible simple $\tilde{\Gamma}$: les longueurs \underline{l} des sous-chaînes correspondent alors au nombre de répétitions dans la chaîne $\tilde{\Gamma}$ de l'état $\mathcal{C}_{\mathfrak{g}_k}$ et la chaîne des classes $\underline{\mathfrak{g}}$ n'est autre que la chaîne des transitions irréversibles (cf. section 5). Si on associe cette chaîne $\tilde{\Gamma}$ avec la matrice de probabilité

$$\begin{aligned} \tilde{\mathcal{P}}_{i,i}(w) &= p(W_i = w), \\ i \neq j, \quad \tilde{\mathcal{P}}_{i,j} &= p(W_i = w), \end{aligned} \quad (\text{III.188})$$

on obtient alors un processus réduit Y dont la distribution de la somme ou du maximum va converger asymptotique vers la distribution de la somme ou du maximum de X . On peut noter que grâce à la nature transitoire des transitions irréversibles, le choix des fonctions $\tilde{\mathcal{P}}_{i,j}$ non diagonales importe peu (sauf en présence de points de terminaisons distincts dans le cas du maximum).

Les états de $\tilde{\Gamma}$ vont correspondre aux classes d'états fortement connexes de Γ . De même, les coefficients $\tilde{\mathcal{E}}_{i,j}$ traduisent les probabilités de transitions entre les classes \mathcal{C}_i et \mathcal{C}_j , une fois que les sous-chaînes $\Gamma^{(i)}$ et $\Gamma^{(j)}$ ont atteint leur état d'équilibre. Le processus Y correspond donc à une version simplifiée du processus X dans laquelle on ne s'intéresse plus qu'au comportement de X à l'échelle des classes d'états fortement connexes plutôt qu'à l'échelle des états. La Fig. 8.7 montre un exemple de ce processus de réduction de graphe. Dans cette figure, le graphe initial de la Fig. 8.6 a été épuré pour ne représenter que les transitions à l'échelle des classes. La matrice de

structure réduite $\tilde{\mathcal{E}}$ est alors une version simplifiée de \mathcal{E} qui appartient à la classe des matrices réductibles simples. Cela implique notamment que le comportement local de Y diffère de celui de X , la structure de corrélation ayant été simplifiée par la réduction ainsi opérée. Cependant le comportement global est conservé, et on peut donc étudier le comportement de la somme ou du maximum de X en étudiant le comportement de la somme ou maximum de Y . Quelle que soit la forme de la matrice \mathcal{E} , il est donc possible de se ramener au cas des matrices \mathcal{E} réductibles simples par cette opération de réduction et les résultats de convergence de la somme et du maximum obtenus dans la section 6 s'appliquent donc dans le cas général.

8 Conclusion

Les sommes et maxima de variables aléatoires à représentation matricielle présentent donc de nombreuses similitudes avec les sommes de variables aléatoires i.i.d.. L'évolution de la fonction cumulative du maximum ou de la fonction caractéristique de la somme en fonction de la taille n du signal est ainsi décrite par des variantes matricielles des équations i.i.d.. Il s'est avéré cependant plus facile d'étudier les sommes et les maxima en utilisant le modèle de chaîne de Markov cachée développé au chapitre 7.

La commutativité et l'associativité de la somme et du maximum impliquent que ceux-ci ne sont pas sensibles à la structure locale de la dépendance des variables aléatoires. Seul le comportement global des variables à représentation matricielle joue donc un rôle significatif dans la convergence vers des distributions limites. Il est donc peu surprenant que dans les cas des variables aléatoires à représentation matricielle dont la corrélation est à courte portée (i.e., dont la matrice \mathcal{E} est irréductible), on retrouve exactement les lois limites classiques.

Cependant, en présence de corrélation à l'échelle du système, un comportement riche apparaît. En effet, pour les représentations matricielles de matrice \mathcal{E} réductible, la présence de transitions irréversibles fait apparaître des fluctuations dont la taille dépend linéairement du système. Ces fluctuations sont suffisamment importantes pour déstabiliser la convergence vers les lois limites standards et font apparaître des lois limites qui sont des mélanges discrets de mélanges continus de lois standards. Plus précisément, les fonctions naturelles (cumulative ou fonction caractéristique) limites prennent la forme d'un mélange discret

$$\mathfrak{F}_l = \sum_{|\mathcal{T}|=p+1} p(\mathcal{T}) \mathfrak{F}_{\mathcal{T}}, \quad (\text{III.189})$$

où $\mathfrak{F}_{\mathcal{T}}$ correspond à la loi limite le long du chemin irréversible \mathcal{T} et $p(\mathcal{T})$ correspond à la probabilité d'observer ce chemin irréversible.

Dans le cas du maximum, les fonctions naturelles sont les fonctions cumulatives. Les fonctions cumulatives limites $F_{\mathcal{T}}(\equiv \mathfrak{F}_{\mathcal{T}})$ sont caractérisées par le paramètre ζ correspondant au domaine d'attraction classique associé, et les $p + 1$ paramètres de translation ou dilatation $\underline{\mu} \equiv \{\mu_i\}$. L'ensemble des fonctions cumulatives limites

$F_{\zeta, \underline{\mu}}$ est donc

$$\zeta < 0, \quad F_{\zeta, \underline{\mu}}(x) = \sum_{k=1}^{p+1} \frac{|x|^{p/\zeta} e^{-(\mu_k |x|)^{-1/\zeta}}}{\prod_{j \neq k} (\mu_j^{-1/\zeta} - \mu_k^{-1/\zeta})}, \quad x < 0, \quad (\text{III.190})$$

$$\zeta = 0, \quad F_{\zeta, \underline{\mu}}(x) = \sum_{k=1}^{p+1} \frac{\exp(px - e^{-x+\mu_k})}{\prod_{j \neq k} (e^{\mu_j} - e^{\mu_k})}, \quad x \in \mathbb{R}, \quad (\text{III.191})$$

$$\zeta > 0, \quad F_{\underline{\mu}}(x) = \sum_{k=1}^{p+1} \frac{x^{p/\zeta} e^{-(\mu_k x)^{-1/\zeta}}}{\prod_{j \neq k} (\mu_j^{-1/\zeta} - \mu_k^{-1/\zeta})}, \quad x > 0. \quad (\text{III.192})$$

Dans le cas de la somme, deux situations coexistent. Si les moyennes des variables aléatoires apparaissant le long de \mathcal{T} diffèrent, la moyenne S du système va fluctuer en fonction du temps passé dans chacun des états. Or dans le cas des chaînes réductibles, ces temps de piégeage sont distribués uniformément et croissent proportionnellement avec la taille du système. En présence de ces fluctuations à l'échelle du signal, la loi des grands nombres est mise en défaut : la moyenne \mathfrak{S}/n ne converge plus presque sûrement vers l'espérance $\mathbb{E}[X]$. À la place, on observe une convergence en distribution de \mathfrak{S}/n vers des lois limites polynomiales par morceaux $\mathcal{L}_{\underline{\mu}}$

$$\frac{\mathfrak{S}}{n} \xrightarrow{\mathcal{D}} \mathcal{L}_{\underline{\mu}}. \quad (\text{III.193})$$

Ces lois polynomiales par morceaux sont déterminées par les moyennes $\mu_k = \mathbb{E}[Y_{\mathcal{T}_k}]$ des variables Y rencontrées le long du chemin \mathcal{T} . Plus explicitement, la densité de probabilité $p_{\underline{\mu}}$ de ces lois $\mathcal{L}_{\underline{\mu}}$ s'écrit

$$p_{\underline{\mu}}(x) = \int_{\sum \alpha_k = 1} \delta\left(x - \sum_k \alpha_k \mu_k\right) \prod_k^{p+1} d\alpha_k. \quad (\text{III.194})$$

Si les moyennes sont concordantes, la loi des grands nombres est vérifiée, et il est possible d'utiliser les résultats de convergence classiques des sommes de variables i.i.d. pour dériver les fonctions naturelles limites non classiques.

Dans le cas de la somme, les fonctions naturelles sont les fonctions caractéristiques. Les fonctions caractéristiques limites de ces sommes $\phi_{\mathcal{T}}(\equiv \mathfrak{F}_{\mathcal{T}})$ sont donc déterminées par le paramètre ζ décrivant le type de lois stables et les paramètres $\underline{\sigma} = \{\sigma_k\}$ de dilatation associés aux différentes variables. Par exemple, dans le cas gaussien, on a

$$\phi_{\underline{\sigma}}(z) = \sum_{k=1}^{p+1} \frac{e^{-\frac{\sigma_k^2 z^2}{2}}}{x^{2p} \prod_{j \neq k} (\sigma_j^2 - \sigma_k^2)}. \quad (\text{III.195})$$

Excepté dans le cas $p = 1$, la densité de probabilité associée à ces fonctions caractéristiques ne semble pas avoir de forme analytique simple. Dans les cas non gaussiens, on pourrait utiliser l'équation générale (III.170) pour obtenir une forme explicite des fonctions caractéristiques limites $\phi_{\zeta, \underline{\sigma}}$. Malheureusement, cette forme explicite est nettement moins lisible que l'expression générale, nous nous abstenons donc de la reproduire ici.

Il est remarquable de constater que ces résultats de convergence s'appliquent à toute variable aléatoire X à représentation matricielle positive de dimension finie. Il

est en effet toujours possible d'utiliser les résultats de la section 7 pour se ramener au cas des variables dont la matrice \mathcal{E} est réductible simple que nous avons étudiées en détail dans la section 6.

Conclusion générale

Dans ces ultimes lignes, notre périple à travers cette thèse touche à sa fin. Nous nous demandions naguère, 161 pages plus tôt, ce que nous pouvions apprendre sur la somme et le maximum de variables aléatoires en alliant les perspectives du traitement du signal et de la physique statistique. Dans le cas des variables aléatoires i.i.d., nous avons déjà contemplé les résultats exhaustifs rassemblés dans la littérature. Mais l'éclat de ces résultats ne peut que rehausser les ombres qui demeurent au-delà des variables i.i.d. : que pouvons-nous apprendre sur les sommes et les maxima lorsque les variables sont couplées à la taille du système, ou en présence de corrélation ? La physique statistique des systèmes désordonnés ou l'analyse multifractale évoquaient d'ores et déjà des situations riches et intrigantes où les résultats de convergence classiques, voire la loi des grands nombres, sont mis en défaut. Force était de constater que ces ruptures de convergences, âpres, surprenantes, intrigantes, offraient un signe clair d'un comportement physiquement intéressant. Nous partîmes donc en quête de ces ruptures de convergence en nous appuyant sur trois axes d'exploration : l'importance des événements rares, le couplage entre les variables aléatoires individuelles et la taille du système, et la corrélation.

Quelles ruptures de convergence avons-nous observées ?

Maximum de variables aléatoires élevées à une puissance dépendant de la taille du signal/système :

Dans l'étude du comportement du maximum de variables aléatoires élevées à une puissance q_n dépendant de la taille du système, nous avons vu apparaître des lois limites non standards, directement reliées aux distributions limites standards. Ces distributions limites non standards n'apparaissent cependant que lorsque la suite des puissances q_n croît plus rapidement qu'une suite de puissance caractéristique q_n^* . Cette suite de puissance caractéristique définit donc un seuil entre les transformations asymptotiquement négligeables ou significatives. Malheureusement, le sens physique de cette transition nous échappe encore, même si on peut conjecturer qu'elle trahit l'existence d'une synergie entre le couplage avec la taille du système et l'importance des événements rares. Une étude plus approfondie de cette interaction pourrait nous permettre de mieux comprendre le comportement à taille finie de la convergence du maximum vers les lois limites. Le lien assez direct entre la famille des lois limites non-standards et des lois limites standards laisse entrevoir la possibilité qu'il existe d'autres familles de distributions limites non standards liées à des transformations différentes de l'élévation à une puissance.

Estimateur des moments et transition vitreuse :

Nous avons retrouvé cette notion de puissance caractéristique q^* dans notre étude de l'estimation des moments pour les variables aléatoires de lois *irrégulières* ou de processus multifractals. Dans ce contexte, l'existence d'une rupture de convergence pour l'estimateur des moments lorsque l'ordre des moments est couplé à la taille du système se manifeste par un effet de linéarisation des moments. Passé un ordre critique q^* , le logarithme de l'estimateur empirique des moments se comporte comme une droite et s'écarte des moments théoriques. Cet effet était déjà connu au sein de la littérature avec par exemple [6, 68]. En utilisant une approche physique fondée sur

l'étude des modèles REM, nous avons vu que cet effet de linéarisation est analogue à une transition vitreuse dans un modèle REM.

En poursuivant cette analogie, nous nous sommes concentrés sur la question de l'évaluation quantitative et expérimentale de l'ordre critique des moments q^* au-delà duquel l'estimateur des moments ne converge plus vers les moments théoriques. Pour la classe des lois log-quasi-exponentielles étirées, nous avons établi des modèles théoriques du comportement asymptotique de cet ordre critique q^* . Il est intéressant de noter que pour cette classe de variables aléatoires, cet ordre critique des moments évolue comme une puissance strictement inférieure à 1 du logarithme de la taille du signal. Il peut donc être extrêmement coûteux d'évaluer de manière fiable des moments d'ordre élevé dans ces situations. Cet état de fait rend notre construction d'un estimateur de l'ordre critique d'autant plus intéressante pour de potentielles applications pratiques puisqu'il offre non seulement une estimation de cet ordre critique q^* pour un nombre d'échantillons n donné mais aussi une estimation grossière de la variation de cet ordre critique q^* en fonction de n . L'étude du comportement de l'estimateur des moments le long d'une courbe (n, q) offre ainsi un point de vue intéressant sur le comportement du système à taille finie en permettant de raffiner les comportements asymptotiques étudiés.

Une des conclusions les plus frappantes de cette étude du comportement de l'estimateur de moments en présence de corrélation apparaît dans le cadre de l'analyse multifractale. En effet, nous avons pu montrer que pour les cascades de Poisson composées, l'interaction entre les corrélations à longue portée et le couplage avec l'échelle d'analyse du système entraîne l'apparition d'un ordre critique q^* qui ne dépend plus que des propriétés statistiques du processus multifractal : aussi fine que soit l'échelle d'analyse considérée, il est impossible d'obtenir des informations sur les moments théoriques au-delà de cet ordre critique. Il est d'ailleurs intéressant de constater que, malgré la présence de corrélations à longue portée, les rapports de participation de ces processus multifractals suivent un comportement similaire à celui prédit pour le modèle REM, modèle sans corrélation, une fois la notion de nombre effectif d'échantillons indépendants introduite.

Variables aléatoires à représentation matricielle :

En nous intéressant aux variables aléatoires à représentation matricielle, nous avons mis en évidence le fait que la seule présence de corrélation à l'échelle du signal suffit à mettre en défaut la loi des grands nombres ou les théorèmes de convergence classiques. Nous avons montré que si la matrice de structure \mathcal{E} de cette représentation matricielle est réductible, la dynamique du signal est dominée par l'existence de transitions irréversibles au sein de la chaîne de Markov cachée. La prédominance des transitions irréversibles entraîne l'apparition de lois limites non standards qui sont des mixtures d'interpolations de lois limites standards. Il est intéressant de constater que ces variables aléatoires sont suffisamment simples pour nous permettre de caractériser en détail les lois limites de leurs sommes et maxima et suffisamment complexes pour mettre en défaut la loi des grands nombres.

Au cours de cette thèse, nous avons également pu mettre en avant une nature duale intrigante de ces variables aléatoires à représentation matricielle : si leurs propriétés statistiques sont plus facilement décrites à partir de leur représentation matricielle, il est bien plus simple pour synthétiser ces variables de les exprimer sous

la forme de modèles à chaîne de Markov cachée particuliers. De manière similaire, si on s'intéresse au comportement de leurs sommes ou maxima, on retrouve cette approche duale puisque deux approches distinctes sont possibles. On peut soit caractériser les lois limites en utilisant directement la représentation matricielle des probabilités jointes et en résolvant les équations fonctionnelles matricielles associées, soit utiliser l'interprétation sous forme de modèle à chaîne de Markov cachée pour caractériser le comportement global de cette chaîne. Nous avons opté dans cette thèse pour la seconde solution qui nous a semblé offrir une résolution plus aisée, mais ce choix est probablement fort subjectif, et la généralisation du raisonnement que nous avons adopté au cas des représentations matricielles de dimension infinie semble non triviale.

Ces représentations de dimension infinie représente un défi intéressant : il est relativement aisé d'étendre les définitions que nous avons adoptées pour englober des variables aléatoires dont la représentation est de dimension infinie. Cependant dans ce cas, la caractérisation de leurs propriétés statistiques et l'étude détaillée des lois limites du maximum ou de la somme deviennent bien plus ardues. Parmi les autres généralisations potentiellement intéressantes de ces lois à représentation matricielle, on peut noter que la restriction des matrices R_d à des matrices positives, si elle permet de s'assurer que l'on obtient uniquement des probabilités jointes valides, n'est pas strictement nécessaire. Lever cette restriction ouvre des possibilités intéressantes mais introduit également des difficultés considérables dans la reformulation sous forme de modèle à chaîne de Markov cachée. De manière indépendante, on peut également constater que par essence, le produit de matrices est une opération essentiellement unidimensionnelle, ce qui peut limiter les possibilités d'utilisation des variables aléatoires à représentation matricielle pour des signaux non unidimensionnel. Pour pallier cet inconvénient, on pourrait songer à utiliser des variables aléatoires à représentation tensorielle qui remplacerait le produit matriciel, opérateur binaire, par un opérateur ternaire, voire n-aire, plus adapté pour décrire des structures de dimension quelconque.

Perspectives

Alors que cette thèse approche inéluctablement de sa fin, il est difficile de résister à la tentation de combiner les approches précédentes. Les variables aléatoires à représentation matricielle n'ouvrent-elles pas des perspectives intrigantes pour l'étude du comportement du maximum ou de la somme de variables aléatoires corrélées dont le comportement individuel dépend de la taille du système ? Qu'advient-t-il de nos résultats sur les distributions limites de maximum de variables aléatoires élevées à une puissance dépendant de la taille du signal lorsque les dites variables aléatoires ne sont plus des variables i.i.d. mais des variables à représentation matricielle ? De manière similaire, on pourrait appliquer notre étude de l'ordre critique des moments au cas des variables à représentation matricielle. Dans cette perspective, il est intéressant de noter que notre étude de l'effet de la corrélation sur l'estimateur de l'ordre critique des moments repose essentiellement sur la notion de nombre effectif d'échantillons indépendants. Cependant les tests que nous avons effectués sur des variables corrélées reposent sur une interprétation très empirique de ce nombre effectif d'échantillons. Or pour les variables aléatoires à représentations matricielles,

il est possible de définir et calculer un nombre effectif d'échantillons indépendants théorique. Est-ce que ces deux notions coïncident ? Sont-elles comparables ?

Ces questions attendent encore fébrilement leur réponse, et peut être que leur futur y pourvoira ; mais cette thèse est arrivée là où *ex quo nata est, ducetur*, et s'achève ici.

Appendices

Annexe A

Fonctions lentement variables

La classe des fonctions lentement variables introduite par Karamata [69] permet de caractériser efficacement un sous-ensemble de fonctions qui varient plus lentement que n'importe quelle fonction de puissance. Nous nous proposons dans la suite de cette annexe de faire un résumé rapide des propriétés de ces fonctions lentement variables. Formellement, une fonction L est une fonction lentement variable si, et seulement si,

$$\forall t \in \mathbb{R}, \quad \lim_{x \rightarrow +\infty} \frac{L(tx)}{L(x)} = 1. \quad (\text{A.1})$$

Un exemple classique de ces fonctions lentement variables sont les fonctions constantes, les fonctions logarithmes mais aussi les fonctions de la forme

$$f(x) = e^{(\ln x)^\alpha (\ln \ln x)^\beta \dots} \text{ avec } \alpha < 1. \quad (\text{A.2})$$

Deux propriétés rendent ces fonctions lentement variables particulièrement attractives. Premièrement, la classe des fonctions lentement variables est stable par multiplication. Deuxièmement, si une fonction L est lentement variable alors

$$\forall \alpha > 0, \quad \lim_{x \rightarrow +\infty} x^\alpha L(x) = +\infty, \quad (\text{A.3})$$

$$\forall \alpha > 0, \quad \lim_{x \rightarrow +\infty} x^{-\alpha} L(x) = 0. \quad (\text{A.4})$$

On peut aller plus loin et montrer [19] que toute fonction lentement variable L peut s'écrire sous la forme

$$L(x) = \exp \left(c(x) \int_{t_0}^x \frac{\epsilon(t)}{t} dt \right), \quad (\text{A.5})$$

où les fonctions $c(x)$ et $\epsilon(x)$ vérifient

$$\lim_{x \rightarrow +\infty} c(x) \in \mathbb{R}, \quad (\text{A.6})$$

$$\lim_{x \rightarrow +\infty} \epsilon(x) = 0. \quad (\text{A.7})$$

Une des limites de ces fonctions lentement variables tient au fait que la dérivée L' d'une fonction lentement variable L peut osciller rapidement, ce qui peut être problématique pour les applications pratiques. Une solution potentielle à ce problème consiste à considérer les fonctions lentement variables normalisées, pour lesquelles la

fonction $c(x)$ de l'équation (A.5) est une constante. C'est-à-dire qu'une fonction L est lentement variable normalisée si, et seulement si, il existe $c \in \mathbb{R}$ et une fonction $\epsilon : \mathbb{R} \rightarrow \mathbb{R}$ telles que

$$L(x) = \exp \left(c \int_{t_0}^x \frac{\epsilon(t)}{t} dt \right), \quad (\text{A.8})$$

avec $\lim_{x \rightarrow +\infty} \epsilon(x) = 0$. Dans ces conditions, le comportement de L' par rapport à L devient bien plus prévisible et on peut montrer que la condition (A.8) est équivalente à la définition suivante : une fonction L est une fonction lentement variable normalisée [19] si et seulement si

$$\lim_{x \rightarrow +\infty} \frac{xL'(x)}{L(x)} = 0. \quad (\text{A.9})$$

Les fonctions lentement variables normalisées permettent donc d'exiger des conditions plus strictes sur le comportement de L' . On peut aller plus loin et considérer les fonctions lentement variables k fois normalisées qui vérifient

$$\forall j \leq k, \quad \lim_{x \rightarrow +\infty} \frac{\partial^j \ln L(x)}{(\partial \ln x)^j} = 0. \quad (\text{A.10})$$

Chaque degré de normalisation k supplémentaire ajoute alors une contrainte sur le comportement de la k -ème dérivée $L^{(k)}$. On peut également définir une classe des fonctions lentement variables régulières qui vérifient

$$\forall k \in \mathbb{N}, \quad \lim_{x \rightarrow +\infty} \frac{\partial^k \ln L(x)}{(\partial \ln x)^k} = 0. \quad (\text{A.11})$$

Ces restrictions supplémentaires pourraient sembler draconiennes, on peut cependant montrer [19] que pour toute fonction lentement variable L , il existe une fonction lentement variable régulière L_r telle que

$$L(x) \underset{x \rightarrow +\infty}{\sim} L_r(x). \quad (\text{A.12})$$

Un autre résultat important que nous utilisons abondamment dans le cadre de cette thèse est l'existence d'inverse asymptotique pour les fonctions régulièrement variables. On appelle fonction régulièrement variable d'indice α [19] les fonctions f de la forme

$$f(x) = x^\alpha L(x). \quad (\text{A.13})$$

Ces fonctions régulièrement variables se comportent de manière essentiellement semblable à des fonctions de puissance. Par exemple, si f et g sont régulièrement variables d'indice α et β alors fg est régulièrement variable d'indice $\alpha + \beta$. De même $1/f$ est régulièrement variable d'indice $-\alpha$. De plus, toute fonction régulièrement variable d'indice α admet une fonction réciproque asymptotique g régulièrement variable d'indice $1/\alpha$ [19] telle que

$$f(g(x)) \underset{x \rightarrow +\infty}{\sim} g(f(x)) \underset{x \rightarrow +\infty}{\sim} x. \quad (\text{A.14})$$

Ces résultats permettent de manipuler la classe des fonctions régulièrement variables comme une extension assez naturelle de la classe des lois de puissances partageant de nombreuses propriétés communes.

Annexe B

REM (*Random Energy Model*)

Proposé par Derrida [40], le REM (*Random Energy Model*) consiste en un des modèles les plus simples de la physique des systèmes désordonnés, mais il capture néanmoins une grande partie de la phénoménologie des systèmes désordonnés plus réalistes. Il peut être représenté comme un modèle de spins, ce bien que les spins ne jouent aucun rôle essentiel dans la description du modèle – si ce n’est pour définir le nombre de configurations. Nous considérons un système de N spins, ce qui correspond à un nombre de configurations de $n = 2^N$. À chaque configuration j est associée une énergie aléatoire E_j qui est tirée de manière i.i.d. avec une probabilité $P(E)$:

$$P(E) = \frac{1}{\sqrt{N\pi J^2}} \exp\left(-\frac{E^2}{NJ^2}\right). \quad (\text{B.1})$$

Les énergies E_j sont des variables aléatoires i.i.d.. Nous notons $\rho(E)dE$ le nombre de configurations dont l’énergie se trouve dans l’intervalle $[E, E + dE]$, de telle sorte que $\rho(E)$ est la densité de configurations d’énergie E . La densité $\rho(E)$ est également une variable aléatoire, mais ses fluctuations sont faibles si $\rho(E)$ est grand, plus précisément $\rho(E) \approx \mathbb{E}[\rho(E)]$. Par définition, $\mathbb{E}[\rho(E)] = nP(E)$, ce qui mène à

$$\begin{aligned} \mathbb{E}[\rho(E)] &= \exp\left(N \ln 2 - \frac{E^2}{NJ^2}\right) \\ &= \exp\left[\ln n \left(1 - \frac{\varepsilon^2}{J^2 \ln 2}\right)\right] \end{aligned} \quad (\text{B.2})$$

où la densité d’énergie $\varepsilon = E/N$ a été introduite. On voit que si $1 - \varepsilon^2/(J^2 \ln 2) > 0$, ce qui correspond à $|\varepsilon| < \varepsilon^\dagger = J\sqrt{\ln 2}$, $\mathbb{E}[\rho(E)]$ est exponentiellement grand en fonction de N , de telle sorte qu’il y a un grand nombre de configurations de densité d’énergie ε , et l’hypothèse $\rho(E) \approx \mathbb{E}[\rho(E)]$ est justifiée. Au contraire, si $|\varepsilon| > \varepsilon^\dagger$, $\mathbb{E}[\rho(E)]$ est très faible pour de grands n . Cela signifie que pour la plupart des réalisations, il n’existe aucune configuration de densité d’énergie $|\varepsilon| > \varepsilon^\dagger$. La valeur non nulle, mais faible de $\mathbb{E}[\rho(E)]$ provient essentiellement de la contribution à la moyenne de quelques rares échantillons, qui incluent des configurations dotées de niveaux d’énergie exceptionnellement bas (ou haut).

Nous pouvons ensuite évaluer la fonction de partition du REM, définie comme

$$Z = \sum_{k=1}^{2^N} e^{-E_k/T}. \quad (\text{B.3})$$

Cette fonction de partition est une variable aléatoire, dont la valeur typique peut être évaluée de la manière suivante :

$$Z \approx Z_{\text{typ}} = \int_{-\varepsilon^\dagger}^{\varepsilon^\dagger} d\varepsilon \mathbb{E} [\tilde{\rho}(\varepsilon)] e^{-N\varepsilon/T}, \quad (\text{B.4})$$

avec la notation $\tilde{\rho}(\varepsilon) = N\rho(N\varepsilon)$. Dans l'équation plus haut, nous avons remplacé $\tilde{\rho}(\varepsilon)$ par $\mathbb{E} [\tilde{\rho}(\varepsilon)]$ pour $|\varepsilon| < \varepsilon^\dagger$, et par 0 pour $|\varepsilon| > \varepsilon^\dagger$. Nous pouvons alors écrire

$$Z_{\text{typ}} = \int_{-\varepsilon^\dagger}^{\varepsilon^\dagger} d\varepsilon e^{-(\ln n)g(\varepsilon)} \quad (\text{B.5})$$

avec

$$g(\varepsilon) = \frac{\varepsilon^2}{J^2 \ln 2} + \frac{\varepsilon}{T \ln 2} - 1. \quad (\text{B.6})$$

Dans la limite des grands n , nous pouvons évaluer Z_{typ} à travers une méthode du col, plus précisément

$$Z_{\text{typ}} \approx e^{-(\ln n)g_{\min}(\varepsilon^\dagger)} \quad (\text{B.7})$$

où $g_{\min}(\varepsilon^\dagger)$ est la valeur minimale de $g(\varepsilon)$ sur l'intervalle $[-\varepsilon^\dagger, \varepsilon^\dagger]$. Considérons tout d'abord la valeur ε_m qui minimise $g(\varepsilon)$ sur l'ensemble de la ligne réelle. En prenant la dérivée de $g(\varepsilon)$, on a

$$g'(\varepsilon) = \frac{2\varepsilon}{J^2 \ln 2} + \frac{1}{T \ln 2}. \quad (\text{B.8})$$

À partir de $g'(\varepsilon) = 0$, nous obtenons

$$\varepsilon_m = -\frac{J^2}{2T}. \quad (\text{B.9})$$

Comme $g(\varepsilon)$ est une parabole, elle décroît pour $\varepsilon < \varepsilon_m$ et augmente pour $\varepsilon > \varepsilon_m$. Si $\varepsilon_m > -\varepsilon^\dagger$, alors $g_{\min}(\varepsilon^\dagger) = g(\varepsilon_m)$, par suite

$$Z_{\text{typ}} \approx e^{-Ng(\varepsilon_m)}. \quad (\text{B.10})$$

La condition $\varepsilon_m > -\varepsilon^\dagger$ se traduit en $T > T_g$, où la température de transition vitreuse T_g est définie par

$$T_g = \frac{J}{2\sqrt{\ln 2}}. \quad (\text{B.11})$$

Pour $\varepsilon_m < -\varepsilon^\dagger$, ce qui équivaut à $T < T_g$, $g(\varepsilon)$ est une fonction croissante de ε sur l'ensemble de l'intervalle $[-\varepsilon^\dagger, \varepsilon^\dagger]$, par suite $g_{\min}(\varepsilon^\dagger) = g(-\varepsilon^\dagger)$, et

$$Z_{\text{typ}} \approx e^{-Ng(-\varepsilon^\dagger)}. \quad (\text{B.12})$$

À partir de ces estimations de Z_{typ} , on peut calculer l'énergie libre $F = -T \ln Z_{\text{typ}}$, et l'entropie $S = -\partial F / \partial T$. Pour $T > T_g$, on obtient

$$F = -N \left(T \ln 2 + \frac{J^2}{4T} \right), \quad (\text{B.13})$$

ce qui mène pour l'entropie à

$$S = N \left(\ln 2 - \frac{J^2}{4T^2} \right). \quad (\text{B.14})$$

Pour $T < T_g$, nous avons

$$F = TNg(-\varepsilon^\dagger) = -NJ\sqrt{\ln 2}. \quad (\text{B.15})$$

L'énergie libre ne dépend donc pas de la température dans cette région, par conséquent l'entropie correspondante s'annule :

$$S = 0, \quad T < T_g. \quad (\text{B.16})$$

On peut aussi vérifier que l'entropie donnée par l'Eq. (B.14) pour $T > T_g$ s'annule continûment pour $T \rightarrow T_g$. Ainsi la température T_g correspond à une température de transition vitreuse, où l'entropie s'annule lorsqu'on diminue la température et demeure nulle pour toute température inférieure à T_g . En fait, pour rendre cet énoncé plus précis, seule la densité d'entropie S/N tend vers zéro pour $T < T_g$, dans la limite N tendant vers l'infini. Le calcul des termes correctifs suivants pour l'entropie montre que S est indépendant de N , mais non nulle, pour $T < T_g$. L'entropie est donc intensive dans cette région de température, ce qui signifie que seul un nombre fini de configurations, parmi les 2^N configurations possibles, est effectivement occupé : le système est gelé dans ses états de plus basses énergies.

Annexe C

Convergence du maximum transformé vers les distributions limites non standards.

Les résultats de convergence obtenus dans la section 4 sont dépendants de la condition donnée par l'Eq. (I.116). Cette condition est équivalente à $F((a_n x + b_n)^{1/q_n})^n - F(\alpha_n \nu(x) + \beta_n)^n \rightarrow 0$. Cette relation est automatiquement valide si g_W est uniformément continue. Cependant, une telle condition est superflue. Définissons

$$x_n = \alpha_n \nu(x) + \beta_n, \quad (\text{C.1})$$

$$\epsilon_n = \frac{(a_n x + b_n)^{1/q_n} - x_n}{\alpha_n}. \quad (\text{C.2})$$

Il est alors possible de montrer que pour les quatre cas présentés en section 4, $\lim_{n \rightarrow \infty} \epsilon_n = 0$. Pour les q_n divergeant dans le cas $q_n/q_n^* \rightarrow 0$, l'Eq. (I.120) mène à

$$\epsilon_n \underset{n \rightarrow +\infty}{\sim} \left(\frac{1}{q_n} - 1 \right) \left(\frac{q_n}{q_n^*} \right). \quad (\text{C.3})$$

Dans le cas $q_n \sim \lambda q_n^*$, nous avons

$$\epsilon_n = \ln(1 + \lambda x) \left(\frac{1}{\lambda} - \frac{q_n}{q_n^*} \right) + \mathcal{O} \left(\frac{q_n}{(q_n^*)^2} \right). \quad (\text{C.4})$$

De même, pour les q_n convergeant vers 0, l'Eq. (I.125) implique

$$\epsilon_n = \left(q_n^* - \frac{1}{\zeta} \right) (e^{\zeta x} - 1) + \mathcal{O}(q_n). \quad (\text{C.5})$$

Puisque $q_n \rightarrow 0$ dans ce cas et que $q_n^* \rightarrow 1/\zeta$, ϵ_n est bien un terme asymptotiquement nul. Et enfin, dans le cas des puissances avec une limite finie et $q_n \sim \lambda q_n^*$, on a

$$\epsilon_n \underset{n \rightarrow +\infty}{\approx} \frac{1}{\zeta} \left((1 + \lambda x)^{1/q_n} - (1 + \lambda x)^{\zeta/\lambda} \right) + \frac{1}{\zeta} - q_n^*. \quad (\text{C.6})$$

Combiné avec $q_n^* \rightarrow 1/\zeta$, cela confirme bien que $\lim_{n \rightarrow \infty} \epsilon_n = 0$.

De plus, par construction $\lim_{n \rightarrow \infty} n\overline{F}(x_n) \in \mathbb{R}$. Il est possible de montrer que $F((a_n x + b_n)^{1/q_n})^n - F(\alpha_n \nu(x) + \beta_n)^n \rightarrow 0$ est équivalent à

$$\lim_{n \rightarrow \infty} \frac{\overline{F}(x_n + \alpha_n \epsilon_n)}{\overline{F}(x_n)} = 1. \quad (\text{C.7})$$

Si l'Eq. (C.7) est vérifiée, alors la convergence en distribution de M_n^U est assurée. Dans les sections 1, 2 et 3, nous vérifions pas à pas que cette condition est satisfaite pour les tous les domaines classiques. Par conséquent, les transitions décrites plus haut sont toujours valides.

1 Domaine Gumbel

Le domaine Gumbel est le plus difficile à caractériser. Afin de prouver la convergence dans le cas général nous emploierons temporairement les facteurs de normalisation classiques

$$\beta_n = \overline{F}^{-1}(1/n), \quad (\text{C.8})$$

$$\alpha_n = E(\beta_n), \quad (\text{C.9})$$

où la fonction $E(x)$ est définie par

$$E(x) = \frac{1}{\overline{F}} \int_x^{+\infty} \overline{F}(t) dt, \quad (\text{C.10})$$

et vérifie :

$$\begin{aligned} \lim_{x \rightarrow +\infty} E'(x) &= 0, \\ \forall r > 0, \quad \lim_{x \rightarrow +\infty} \frac{\overline{F}(x + rE(x))}{\overline{F}(x)} &= e^{-r}. \end{aligned} \quad (\text{C.11})$$

Une des propriétés intéressantes de E est que pour tout réel positif r

$$\frac{E(x + rE(x))}{E(x)} \xrightarrow{x \rightarrow +\infty} 1. \quad (\text{C.12})$$

La combinaison des Eqs. (C.11), (C.12) et (C.12) mène à

$$\begin{aligned} E(x_n) &= E(\alpha_n x + \beta_n) \\ &= E(\beta_n + xE(\beta_n)) \sim E(\beta_n), \end{aligned} \quad (\text{C.13})$$

de telle sorte que $E(x_n) \sim \alpha_n$. En conséquence, $\alpha_n \epsilon_n / E(x_n) \rightarrow 0$, ce qui en conjonction avec l'Eq. (C.11) conduit à

$$\begin{aligned} \overline{F}(x_n + \alpha_n \epsilon_n) &= \overline{F}\left(x_n + \frac{\alpha_n \epsilon_n}{x_n E(x_n)} E(x_n) x_n\right) \\ &\sim \exp\left(-\frac{\alpha_n \epsilon_n}{x_n E(x_n)}\right) \overline{F}(x_n) \\ &\sim \overline{F}(x_n). \end{aligned} \quad (\text{C.14})$$

Ainsi l'Eq. (C.7) est vérifiée pour le domaine Gumbel.

2 Domaine Weibull

Comme énoncé dans l'Eq. (I.93), nous avons pour le domaine Weibull

$$\bar{F}(x) = L\left((x_F - x)^{-1}\right) (x_F - x)^{-1/\zeta}. \quad (\text{C.15})$$

De plus, la définition des facteurs de normalisation mène à

$$\frac{x_F - \beta_n}{\alpha_n} \underset{n \rightarrow +\infty}{\sim} \frac{(x_F - \beta_n)\bar{F}(\beta_n)}{\bar{F}'(\beta_n)}. \quad (\text{C.16})$$

Par suite, sous l'hypothèse que $L'/L \rightarrow 0$ nous avons

$$\lim_{n \rightarrow \infty} \frac{x_F - \beta_n}{\alpha_n} = \zeta. \quad (\text{C.17})$$

Il est donc possible de montrer que

$$x_F - (x_n + \alpha_n \epsilon_n) \sim \alpha_n x \left(1 + \frac{\zeta}{x}\right). \quad (\text{C.18})$$

Les propriétés des fonctions variant lentement induisent que, pour $n \rightarrow \infty$

$$\begin{aligned} \bar{F}(x_n + \alpha_n \epsilon_n) &= L\left(\frac{1}{x_F - x_n - \alpha_n \epsilon_n}\right) (x_F - x_n - \alpha_n \epsilon_n)^\zeta \\ &\sim L\left(\frac{1}{\alpha_n x}\right) (\alpha_n x (1 + \zeta/x))^{-\frac{1}{\zeta}} \\ &\sim \bar{F}(x_n). \end{aligned} \quad (\text{C.19})$$

Une fois encore, l'Eq. (C.7) est vérifiée, cette fois pour le domaine Weibull.

3 Domaine Fréchet

Au sein du domaine Fréchet, la preuve est presque immédiate en utilisant $\alpha_n \epsilon_n / x_n \rightarrow 0$ qui est une conséquence directe de $\lim_{n \rightarrow \infty} \epsilon = 0$ et $\lim_{n \rightarrow \infty} q_n^* > 0$. Nous avons à partir de l'Eq. (I.92)

$$\begin{aligned} \bar{F}(x_n + \alpha_n \epsilon_n) &= (x_n + \alpha_n \epsilon_n)^{-1/\zeta} L(x_n + \alpha_n \epsilon_n) \\ &\sim x_n^{-1/\zeta} L(x_n) = \bar{F}(x_n). \end{aligned} \quad (\text{C.20})$$

Le résultat est immédiat et l'Eq. (C.7) est également vérifiée pour le domaine Fréchet.

Annexe D

Évaluation par la méthode du col de $\mathbb{E}[X^q]$

La méthode du col est une technique classique permettant d'évaluer le comportement asymptotique d'intégrales de la forme $\int_{-\infty}^{+\infty} g(x)e^{tf(x)}dx$ pour les grands t [102, 36]. Plus précisément, si x^* est le seul maximum f , l'évaluation par la méthode du col s'écrit :

$$\int_{-\infty}^{+\infty} g(x)e^{tf(x)}dx \underset{t \rightarrow +\infty}{\sim} g(x^*)e^{tf(x^*)}, \quad \text{avec } f'(x^*) = 0. \quad (\text{D.1})$$

Pour le présent problème, l'Eq. (II.76) mène à

$$\mathbb{E}[X^q] = \int_{-\infty}^{+\infty} e^{qy - \phi(y)} dy, \quad (\text{D.2})$$

où $\phi(y) = -\ln \mathbb{P}_Y(y)$ est la densité logarithmique de probabilité de Y . Une tentative d'identification naïve avec l'intégrale de la méthode du col mènerait dans ce cas à $t \equiv q$, $f(x) \equiv x$ et $g(x) \equiv \exp(-\phi(x))$. Puisque la fonction f n'admet pas de maximum, une application directe de la méthode du col est inefficace. Cependant, il est possible de dériver une variante de la méthode du col en utilisant le fait que ϕ est essentiellement une fonction de puissance.

La première étape consiste à localiser le maximum y_m de l'intégrande dans l'Eq. (D.2) :

$$\phi'(y_m) = q. \quad (\text{D.3})$$

Le changement de variable $y = \nu y_m$ nous permet de réécrire l'Eq. (D.2) en :

$$\mathbb{E}[X^q] = y_m \int_{-\infty}^{+\infty} e^{qy_m \nu - \phi(\nu y_m)} d\nu. \quad (\text{D.4})$$

Si ϕ était exactement une fonction puissance $\phi(y) = y^\rho/\rho$, nous aurions $y_m = q^{1/(\rho-1)}$ et l'équation précédente deviendrait

$$\mathbb{E}[X^q] = y_m \int_{-\infty}^{+\infty} e^{q \frac{\rho}{\rho-1} (\nu - \nu^\rho/\rho)} d\nu. \quad (\text{D.5})$$

Dans cette situation, il serait possible d'appliquer directement la méthode du col avec $t \equiv q^{\rho/(\rho-1)}$ et $f(x) \equiv x - x^\rho$, obtenant ainsi l'Eq. (II.77). Cependant, dans le cas général, $\phi(y)$ n'est pas exactement une fonction de puissance et une approche plus

indirecte est nécessaire. Les conditions (II.20) et (II.21) assurent que $\phi(y) \underset{y \rightarrow +\infty}{\sim} h(y)$ et $\phi'(y) \underset{y \rightarrow +\infty}{\sim} h'(y)$, ce qui implique que

$$\lim_{y \rightarrow +\infty} \frac{y\phi'(y)}{\phi(y)} = \rho, \quad (\text{D.6})$$

$$\lim_{y \rightarrow +\infty} \frac{\phi(\nu y)}{\phi(y)} = \nu^\rho. \quad (\text{D.7})$$

La combinaison des Eqs. (D.6), (D.7) et (D.4) donne

$$qy_m\nu - \phi(\nu y_m) = \phi(y_m)(\rho\nu - \nu^\rho + \varepsilon(y_m, \nu)), \quad (\text{D.8})$$

avec $\forall \nu, \lim_{y_m \rightarrow +\infty} \varepsilon(y_m, \nu) = 0$; et donc

$$\mathbb{E} [X^q] = y_m \int_{-\infty}^{+\infty} e^{\phi(y_m)(\rho\nu - \nu^\rho + \varepsilon(y_m, \nu))} d\nu. \quad (\text{D.9})$$

Comme précédemment, l'intégrale précédente peut être vue comme une intégrale de type méthode du col avec $t \equiv \phi(y_m)$ et $f(x) \equiv \rho x - x^\rho + \varepsilon(y_m, \nu)$. Cependant, avec ces définitions, la fonction f dépend de t à travers le terme correctif $\varepsilon(y_m, \nu)$. La prochaine étape est de vérifier que ce terme correctif n'altère pas le comportement asymptotique de l'intégrale.

Pour ce faire, nous commençons par réécrire l'équation ci-dessus en

$$\mathbb{E} [X^q] = y_m e^{qy_m - \phi(y_m)} \int_{-\infty}^{+\infty} e^{\phi(y_m)(1 - \rho + \rho\nu - \nu^\rho + \varepsilon_2(y_m, \nu))} d\nu. \quad (\text{D.10})$$

avec $\varepsilon_2(y_m, \nu) = \varepsilon(y_m, \nu) - \varepsilon(y_m, 1)$. Pour $\nu \neq 1$, $1 - \rho + \rho\nu - \nu^\rho + \varepsilon_2(y_m, \nu) < 0$ donc seul le voisinage de $\nu = 1$ contribue significativement à l'intégrale lorsque $y \rightarrow +\infty$. Dans le voisinage de $\nu = 1$, il est plus simple de repartir de l' Eq. (D.4) et de développer à l'ordre 2 $q\nu y_m - \phi(\nu y_m)$ pour $\nu = 1 + h$ et $h \rightarrow 0$. On a alors

$$\begin{aligned} q(1+h)y_m - \phi((1+h)y_m) &= qy_m - \phi(y_m) + hy_m(q - \phi'(y_m)) - \frac{h^2}{2}y_m^2\phi''(y_m) + o(h^2) \\ &= qy_m - \phi(y_m) - \frac{h^2}{2}y_m^2\phi''(y_m) + o(h^2) \end{aligned} \quad (\text{D.11})$$

En injectant ce développement dans l'intégrale (D.4), on obtient l'intégrale gaussienne suivante

$$\int_{-\infty}^{+\infty} e^{-y_m^2\phi''(y_m)\frac{(\nu-1)^2}{2}} d\nu = \sqrt{\frac{2\pi}{y_m^2\phi''(y_m)}}, \quad (\text{D.12})$$

Or ce qui redonne bien l'équation (II.77) :

$$\mathbb{E} [X^q] \sim e^{qy_m - \phi(y_m)} \sqrt{\frac{2\pi}{\phi''(y_m)}}. \quad (\text{D.13})$$

Annexe E

Moments tronqués et estimateur des moments

Dans cette annexe, nous démontrerons la convergence presque sûre de $f(n, q) = \ln S(n, q) / \ln n$ vers $\lim_n \ln T(n, q) / \ln n$, lorsque $n \rightarrow +\infty$, pour les variables aléatoires de classe A. De manière similaire, nous montrerons que $\ln S(n, q) / |\ln a|$ converge vers $\lim_a \ln T(a, q) / |\ln a|$, lorsque $a \rightarrow 0$, pour les variables de classe B.

L'essence de la démonstration est fort similaire dans les deux cas. Dans un premier temps, nous montrerons que presque sûrement tous les échantillons y appartiennent à l'intervalle $] -\infty, y^\dagger]$. Nous étudions ensuite plus en détails le comportement de S en sectionnant plus finement cet intervalle. Cependant, les contraintes sont légèrement différentes dans les deux cas, nous développerons donc les preuves séparément malgré leur forte similarité.

1 Classe A

Le point le plus fondamental est de rendre précis le rôle clef joué par y_τ^\dagger en tant que frontière. Par définition,

$$\mathbb{P}(Y_1, \dots, Y_n > y^\dagger) = 1 - e^{-\tau(n)} \underset{\tau(vn) \rightarrow 0}{\sim} \tau(n). \quad (\text{E.1})$$

Le lemme de Borel-Cantelli [18] nous dit que si une séquence d'événements aléatoires A_n satisfait $\sum_n P(A_n) < +\infty$, alors les événements A_n ne se produisent qu'un nombre fini de fois. Par suite, un choix naturel d'événements serait $A_n = \{\exists i < n, Y_i > y_{\tau(n)}^\dagger\}$. Si $\sum_n \tau(n) < +\infty$, alors pour un n suffisamment grand, tous les Y_i sont presque sûrement plus petits que $y_{\tau(n)}^\dagger$. Cependant, pour préserver la propriété $y_{\tau(n)}^\dagger \rightarrow y_1^\dagger$ pour $n \rightarrow +\infty$, il est également nécessaire que $\ln \tau \ll \ln n$, i.e. :

$$\lim_{x \rightarrow +\infty} \frac{\ln \tau}{\ln n} = 0. \quad (\text{E.2})$$

Malheureusement, ces deux conditions sont incompatibles.

Pour résoudre ce dilemme, considérons donc une sous-suite $n(r)$ d'échantillons (par exemple $n(r) = 2^r$) et les événements $A_r = \{\exists i < n(r), Y_i > y_{\tau(n(r))}^\dagger\}$. La condition du lemme de Borel-Cantelli devient alors :

$$\sum_r \tau(n(r)) < +\infty. \quad (\text{E.3})$$

On peut d'ores et déjà remarquer que cette condition, pour $n(r) = 2^r$, est compatible avec $\ln \tau \ll \ln n$ si on choisit $\tau = 1/\ln n$. De plus, si la valeur de $\ln S/\ln n$ aux points $n(r)$ est connue, le comportement de S peut être inféré avec suffisamment de précision pour tous les points. Considérons deux points $k < l$, nous avons :

$$f(k, q(k)) < f(l, q(l)) \frac{\ln l}{\ln k} + 1 - \frac{\ln l}{\ln k}. \quad (\text{E.4})$$

Pour $a = n(r) < k < b = n(r+1)$, l'inégalité précédente devient :

$$f(a, q(a)) \frac{\ln a}{\ln b} + 1 - \frac{\ln a}{\ln b} < f(k, q(k)) < f(b, q(b)) \frac{\ln b}{\ln a} + 1 - \frac{\ln b}{\ln a}. \quad (\text{E.5})$$

Si $\ln(n(r+1))/\ln(r) \rightarrow 1$, l'inégalité plus haut décrit précisément f , dans la limite $n \rightarrow +\infty$. Ceci est équivalent à définir $n(r)$ comme

$$n(r) = \exp(e^{\nu(r)}), \text{ avec } \lim_{r \rightarrow +\infty} \nu(r+1) - \nu(r) = 0. \quad (\text{E.6})$$

Une solution simple est de choisir (presque) arbitrairement $\nu(r) = \sqrt{r}$ et $\tau(r) = \frac{1}{r^2}$, alors $n(r) = \exp(e^{\sqrt{r}})$ et

$$\tau(n) = \left(\frac{1}{\ln \ln n} \right)^4. \quad (\text{E.7})$$

Par conséquent, si on observe uniquement les points $n(r)$, pour r suffisamment grand, presque sûrement, tous les Y_i sont plus petits que y_τ^\dagger .

Nous noterons χ_I la fonction caractéristique de l'ensemble I ,

$$\chi_I(x) = \begin{cases} 1 & \text{si } x \in I, \\ 0 & \text{sinon.} \end{cases} \quad (\text{E.8})$$

On peut écrire

$$S(n, q) \stackrel{\text{p.s.}}{=} \frac{1}{n} \sum_i e^{qY_i} \chi_{(-\infty, y_\tau^\dagger]}(Y_i). \quad (\text{E.9})$$

On peut évaluer plus précisément cette somme en scindant l'intervalle restant $(-\infty, y_\tau^\dagger]$ en $2R(n) + 1$ sous-intervalles avec

$$a_r = k \frac{y_\tau^\dagger}{R(n)}, \quad (\text{E.10})$$

$$I_k = (a_r, a_{k+1}], k \in -R, \dots, R-1, \quad (\text{E.11})$$

$$I_{-R(n)-1} = I_{-\infty} = (-\infty, a_{-R}]. \quad (\text{E.12})$$

Pour des raisons qui seront plus explicites par la suite, nous choisissons :

$$R(n) = \lfloor \frac{\ln n}{4 \ln \ln n} \rfloor, \quad (\text{E.13})$$

où $\lfloor x \rfloor$ représente la partie entière de x . Comme $y_\tau^\dagger = \mathcal{L}(\ln n)(\ln n)^{1/\rho}$, la longueur des sous-intervalles tend vers 0. $S(n, q)$ peut donc être réécrit

$$S(n, q) \stackrel{\text{p.s.}}{=} \sum_{i=1}^n \sum_{k=-R(n)-1}^{R(n)-1} e^{qY_i} \chi_{I_k}(Y_i). \quad (\text{E.14})$$

Il est possible de compter le nombre de points dans chaque intervalle :

$$n_r = \sum_{i=1}^n \chi_{I_k}(Y_i). \quad (\text{E.15})$$

Sous certaines hypothèses, n_r ne devrait pas être très différent de

$$\mathbb{E}[n_r] = n\mathbb{E}[\chi_{I_k}]. \quad (\text{E.16})$$

Plus précisément, l'inégalité de Markov [18] nous donne presque sûrement une borne supérieure :

$$\mathbb{P}(\exists k, n_r > \alpha(n)\mathbb{E}[n_r]) < \frac{2R(n) + 1}{\alpha(n)}. \quad (\text{E.17})$$

Afin d'utiliser le lemme de Borell-Cantelli, il faut que $\sum_r \frac{2R(n(r))+1}{\alpha(n(r))} < +\infty$. Plus α croît lentement, plus grande est la quantité d'information que l'on peut obtenir du lemme de Borel-Cantelli. Un choix valide ici nécessite $\ln \alpha \ll \ln n$. À cause de la parcimonie de $n(r)$,

$$\alpha(n) = R(n) \ln n \quad (\text{E.18})$$

est un choix compatible, qui mène à :

$$S(n, q) \stackrel{\text{p.s.}}{<} \sum_{k=-R(n)-1}^{R(n)-1} \alpha(n)\mathbb{E}[\chi_{I_k}] e^{qa_{k+1}}. \quad (\text{E.19})$$

De plus, si on note ϕ la densité logarithmique (comme défini en Annexe D),

$$\mathbb{E}[\chi_{I_k}] < 2 \frac{y^\dagger}{R} e^{-\phi(a_r)}. \quad (\text{E.20})$$

La contribution majeure à la somme provient de l'intervalle I_m contenant $y_m(q)$ défini par

$$\phi'(y_m) = q. \quad (\text{E.21})$$

Cela mène à une borne supérieure assez grossière,

$$S(n, q) \stackrel{\text{p.s.}}{<} \alpha(n) \left(y_r^\dagger e^{qa_{m+1} - \phi(a_m)} + \mathbb{E}[\chi_{I_{-\infty}}] e^{-qy^\dagger} \right). \quad (\text{E.22})$$

La plupart des termes de cette équation sont logarithmiquement négligeables par rapport à $\ln n$. Par suite,

$$\lim_n \frac{\ln S(n, q)}{\ln n} \stackrel{\text{p.s.}}{<} \lim_n \frac{qa_{m+1}(n) - \phi(a_m(n))}{\ln n}. \quad (\text{E.23})$$

La borne inférieure est plus difficile à atteindre et la méthode proposée ici utilise l'inégalité de Bienaymé-Tchebitchev et ne s'applique donc que dans le cas de variables i.i.d. .

$$\mathbb{P} \left(N_m > \frac{\mathbb{E}[N_m]}{2} \right) < \frac{4}{n\mathbb{E}[\chi_{I_m}]}. \quad (\text{E.24})$$

Or, on peut minorer $n\mathbb{E}[\chi_{I_m}]$ par

$$\begin{aligned}
 n\mathbb{E}[\chi_{I_m}] &> n\mathbb{E}[\chi_{I_{R-1}}] \\
 &> n \int_{y^\dagger(1-\frac{1}{R})}^{y^\dagger(1-\frac{1}{2R})} e^{-\phi(y)} dy. \\
 &> n \frac{y^\dagger}{2R} e^{-\phi(y^\dagger(1-\frac{1}{2R}))} \\
 &= \frac{y^\dagger}{2R} (\ln n)^{2\rho(1+\epsilon(y))} \\
 &\sim \frac{\ln \ln n}{4} (\ln n)^{(2\rho-1+1/\rho)(1+\epsilon(y))}.
 \end{aligned} \tag{E.25}$$

On veut de plus que

$$\sum_r \mathbb{P}\left(N_m > \frac{\mathbb{E}[N]_m}{2}\right) < +\infty. \tag{E.26}$$

Cette contrainte (E.26) alliée à l'équation (E.25) impose la condition la plus restrictive sur R , elle est en grande partie responsable du choix de $R(n)$ dans l'Eq. (E.13). Avec ce choix de $R(n)$, on peut utiliser le lemme de Borel-Cantelli pour obtenir

$$\frac{1}{2} e^{qa_m - \phi(a_{m+1})} \stackrel{\text{p.s.}}{<} S(n, q). \tag{E.27}$$

On obtient alors une borne inférieure sur la limite de f :

$$\lim_n \frac{qa_m - \phi(a_{m+1})}{\ln n} \stackrel{\text{p.s.}}{<} \lim_n f(n, q). \tag{E.28}$$

La dernière étape est de vérifier que les deux bornes (E.23) et (E.28) convergent vers la même valeur lorsque $n \rightarrow +\infty$. Il faut d'abord vérifier que

$$\lim_{n \rightarrow +\infty} \frac{q(a_m - a_{m+1})}{\ln n} = 0. \tag{E.29}$$

Cela revient à vérifier que

$$\lim_{n \rightarrow +\infty} \frac{q(n)y^\dagger}{R \ln n} = \lim_{n \rightarrow +\infty} \frac{q(n)y^\dagger \ln \ln n}{(\ln n)^2} = 0. \tag{E.30}$$

Cette convergence est obtenue au prix d'une condition sur $q(n)$

$$\exists \epsilon > 0, q \ll (\ln n)^{2-1/\rho-\epsilon}. \tag{E.31}$$

On veut ensuite s'assurer que

$$\lim_{n \rightarrow +\infty} \frac{\phi(a_m) - \phi(a_{m+1})}{\ln n} = 0. \tag{E.32}$$

Pour ce faire on peut remarquer que

$$\phi(a_m) - \phi(a_{m+1}) \sim \frac{y^\dagger}{R} \phi'(a_m) < \frac{y^\dagger}{R} \phi'(y^\dagger) \tag{E.33}$$

et

$$\frac{\phi'(y^\dagger)y^\dagger}{R \ln n} \sim \frac{4 \ln \ln n}{\ln n}. \quad (\text{E.34})$$

Combinées, les équations (E.33) et (E.34) impliquent que la condition E.32 est vérifiée. Les bornes supérieure (E.23) et inférieure (E.28) convergent vers la même valeur et on obtient donc

$$\lim_n \frac{\ln S(n, q(n))}{\ln n} \stackrel{\text{p.s}}{=} \lim_n \frac{qy_m(n) - \phi(y_m(n))}{\ln n}. \quad (\text{E.35})$$

Le terme de droite de l'Eq. (E.35) correspond au résultat par l'évaluation des moments tronqués par la méthode du col spécifique de l'annexe D. Par conséquent, l'Eq. (II.116) :

$$\lim_{n \rightarrow +\infty} \frac{\ln S(n, q(n))}{|\ln n|} \stackrel{\text{p.s}}{=} \lim_{n \rightarrow +\infty} \frac{\ln T(n, q(n))}{|\ln n|} \quad (\text{E.36})$$

est bien démontrée.

2 classe B

Dans cette section, nous nous proposons d'adapter la preuve précédente au cas de processus multifractals de la classe B. Nous souhaitons démontrer que

$$\lim_{r \rightarrow +\infty} \frac{\ln S_{n_r}(a_r, q)}{|\ln a_r|} \stackrel{\text{p.s}}{=} \lim_{a \rightarrow 0} \frac{\ln M(a, q)}{|\ln a|}. \quad (\text{E.37})$$

Une première différence avec la classe A est l'existence d'un choix naturel pour la sous-suite n_r . En effet, si on s'intéresse aux accroissements sur des échelles discrètes, il est assez naturel de considérer que $n_r = 2^r$ et $a_r = 2^{-r}L$.

Une deuxième différence réside dans le fait que pour la classe B, il est plus aisé de raisonner en utilisant la variable h plutôt que y . Pour commencer, nous raffinons la définition de h^\dagger en choisissant

$$\tau_r = \frac{1}{r^2} \quad (\text{E.38})$$

dans l'équation (II.95), à la place d'avoir un τ constant. Cette altération n'affecte pas h_0^\dagger à cause de la propriété

$$\lim_{r \rightarrow +\infty} \frac{\ln \tau_r}{|\ln a_r|} = 0. \quad (\text{E.39})$$

Avec ce choix de τ_r , nous avons

$$\sum_r P(\exists i < n_r, h_{a_r}(i) < h_r^\dagger) < +\infty. \quad (\text{E.40})$$

En utilisant le lemme de Borel-Cantelli et en choisissant comme événements $A_r = \{\exists i < n_r, h_{a_r}(i) < h_r^\dagger\}$, nous pouvons donc affirmer que pour r suffisamment grand, tous les h_a sont presque sûrement supérieurs à h_r^\dagger . Nous notons χ_I la fonction caractéristique de l'ensemble I ,

$$\chi_I(x) = \begin{cases} 1 & \text{si } x \in I, \\ 0 & \text{sinon.} \end{cases} \quad (\text{E.41})$$

Puisque presque sûrement tous les h_a sont supérieurs à h_r^\dagger , nous pouvons écrire

$$S_{n_r}(a_r, q) \stackrel{\text{p.s.}}{=} \frac{1}{n_r} \sum_{i=1}^{n_r} a^{qh_i} \chi_{[h_r^\dagger, +\infty)}(h_i). \quad (\text{E.42})$$

Afin de raffiner ce résultat, nous découpons l'intervalle $[h^\dagger, +\infty)$ en un ensemble de sous-intervalles. Tout d'abord, nous définissons un point de séparation c_∞ par

$$\phi'(c_\infty) = 0. \quad (\text{E.43})$$

Le sous-intervalle correspondant est :

$$I_\infty = (c_\infty, +\infty). \quad (\text{E.44})$$

L'intervalle restant $[h_r^\dagger, c_\infty]$ possède une longueur finie l_r

$$l_r = c_\infty - h_r^\dagger. \quad (\text{E.45})$$

Nous divisons cet intervalle en $\lfloor 2 \ln n_r \rfloor$ sous-intervalles (où $\lfloor x \rfloor$ est la partie entière de x), avec

$$c_p = h^\dagger + p \frac{l_r}{1 + \lfloor \ln n_r \rfloor}, \quad (\text{E.46})$$

$$I_p = [c_p, c_{p+1}], \quad |p| < 1 + \lfloor \ln n_r \rfloor. \quad (\text{E.47})$$

Nous appelons m_p la densité de points à l'intérieur de l'intervalle I_p

$$m_p = \frac{1}{n_r} \sum_{i=1}^{n_r} \chi_{I_p}(h(i)). \quad (\text{E.48})$$

Afin d'obtenir une borne supérieure sur m_p , on peut utiliser l'inégalité de Markov :

$$\mathbb{P}(\exists j, m_j > (\ln n_r)^3) < \frac{1}{(\ln n_r)^2}. \quad (\text{E.49})$$

On peut alors vérifier que

$$\sum_r \mathbb{P}(\exists j, m_j > (\ln n_r)^3 \mathbb{E}[\chi_{I_p}]) = \sum_r \frac{1}{(r \ln 2)^2} < +\infty. \quad (\text{E.50})$$

On peut donc utiliser le lemme de Borel-Cantelli, pour obtenir une borne supérieure presque-sûre de m_p :

$$m_p \stackrel{\text{p.s.}}{<} (\ln n_r)^3 \mathbb{E}[\chi_{I_p}]. \quad (\text{E.51})$$

La borne inférieure est cependant plus problématique, si on utilise l'inégalité de Bienaymé-Tchebitchev, on obtient

$$\mathbb{P}\left(m_j > \frac{\mathbb{E}[\chi_{I_p}]}{2}\right) < \frac{1}{n_r^2} \text{Var}\left[\sum_{i=1}^{n_r} \chi_{I_p}(H_i)\right]. \quad (\text{E.52})$$

Pour que ce résultat soit utilisable, il faut donc que la somme des corrélations des χ_{I_p} croisse moins rapidement que $n_r/(\ln n_r)^2$. Faisons donc l'hypothèse réaliste que

tel est le cas. On peut alors utiliser le lemme de Borel-Cantelli pour obtenir une borne inférieure presque-sûre

$$\frac{\mathbb{E}[\chi_{I_p}]}{2} \stackrel{\text{p.s.}}{<} m_p. \quad (\text{E.53})$$

En combinant les bornes supérieures et inférieures, on peut encadrer presque sûrement m_p par

$$\frac{\mathbb{E}[\chi_{I_p}]}{2} \stackrel{\text{p.s.}}{<} m_p \stackrel{\text{p.s.}}{<} (\ln n_r)^3 \mathbb{E}[\chi_{I_p}]. \quad (\text{E.54})$$

De plus, la longueur des intervalles I_p tend vers 0, cela implique que :

$$\mathbb{E}[\chi_{I_p}] \underset{r \rightarrow +\infty}{\sim} \frac{l_r}{\ln n_r} a^{\phi(c_p)}. \quad (\text{E.55})$$

Par suite, nous avons un $\varepsilon_1 \in (0, 1)$ tel que pour r suffisamment grand,

$$S_{n_r}(a_r, q) \stackrel{\text{p.s.}}{<} \sum_p (\ln n_r)^2 l_r (1 + \varepsilon_1) a^{q c_p + \phi(c_p)} + m_\infty a^{c_\infty}. \quad (\text{E.56})$$

Nous sommes principalement intéressés par la contribution maximale, qui provient de c_m :

$$(1 - \varepsilon_2) l_r \frac{a^{q c_m + \phi(c_m)}}{2 \ln r} \stackrel{\text{p.s.}}{<} S_{n_r}(a_r, q) \stackrel{\text{p.s.}}{<} (1 + \varepsilon_3) (\ln n_r)^3 l_r a^{q c_m + \phi(c_m)}, \quad (\text{E.57})$$

avec $\varepsilon_2, \varepsilon_3 \in (0, 1)$. Ces deux inégalités peuvent être réécrites en termes logarithmiques,

$$\begin{aligned} & -(q c_m + \phi(c_m)) + \frac{\ln(1 - \varepsilon_2) + \ln l_r - \ln 2 - \ln \ln n_r}{|\ln a|} \\ & \stackrel{\text{p.s.}}{<} \frac{\ln S_{n_r}(a_r, q)}{|\ln a_r|} \stackrel{\text{p.s.}}{<} \\ & -(q c_m + \phi(c_m)) + \frac{3 \ln \ln n_r + \ln(1 + \varepsilon_3) + \ln l_r}{|\ln a|}. \end{aligned} \quad (\text{E.58})$$

La longueur totale l_r est bornée parce que les h_r^\dagger et c_∞ le sont. Lorsque $r \rightarrow +\infty$, la longueur de l'intervalle I_p tend vers 0, ce qui signifie que

$$c_m \rightarrow \max(h_m, h^\dagger). \quad (\text{E.59})$$

En prenant la limite $r \rightarrow +\infty$ dans l'Eq. (E.58), on obtient donc enfin

$$\lim_{r \rightarrow +\infty} \frac{\ln S_{n_r}(a_r, q)}{|\ln a_r|} \stackrel{\text{p.s.}}{=} -q \max(h_m, h^\dagger) - \phi(\max(h_m, h^\dagger)). \quad (\text{E.60})$$

Le membre de droite de l'équation précédente correspond exactement à l'évaluation des moments tronqués par la méthode du col obtenue dans les Eqs. (II.119) et (II.120), ce qui démontre la validité de l'Eq. (E.37). Un résultat similaire pour les cascades multiplicatives dyadiques (cf [96]) suggère que la convergence presque sûre tient dans le cas général, et ce sans avoir besoin d'hypothèse pour obtenir la borne inférieure.

Annexe F

Estimation de θ : résultats théoriques

1 Construction

Afin de démontrer la consistance et la normalité asymptotique de Ω_k , il est plus simple de considérer les variables (G_1, \dots, G_n) de la loi $\Lambda^{(k)}$. Si nous définissons ensuite

$$\Delta_i = i(G_i - G_{i+1}), \quad (\text{F.1})$$

on peut voir que

$$p_{\Lambda^{(k)}}(\delta_1, \dots, \delta_{k-1}, g_k) = \exp\left(-e^{-g_k} - kg_k - \sum_{i < k} i \delta_i\right). \quad (\text{F.2})$$

En d'autres termes, Δ_i et G_k sont indépendants. Des calculs directs mènent à :

$$\mathbb{E}[\Delta_i] = 1, \quad (\text{F.3})$$

$$\text{Var}[\Delta_i] = 1. \quad (\text{F.4})$$

De la même manière, avec γ la constante Euler-Mascheroni :

$$\mathbb{E}[G_k] = \gamma - \zeta(1; k-1), \quad (\text{F.5})$$

$$\text{Var}[G_k] = \frac{\pi^2}{6} - \zeta(2; k-1). \quad (\text{F.6})$$

où $\zeta(s; n)$ est une fonction ζ de Riemman incomplète :

$$\zeta(s; n) = \sum_{j=1}^n \frac{1}{j^s}. \quad (\text{F.7})$$

Si nous essayons d'appliquer Ω_k aux variables G_i , nous pouvons transcrire

$$\Omega_k = \sum_i \alpha_i G_i \quad (\text{F.8})$$

en

$$\Omega_k = G_k + \sum_{i=1}^{k-1} \beta_i \Delta_i, \quad (\text{F.9})$$

avec $\beta_i = \frac{1}{i} \sum_{j \leq i} \alpha_j$. Nous pouvons donc calculer l'erreur quadratique avec

$$\mathbb{E} [\Omega_k^2] = \mathbb{E} [\Omega_k]^2 + \text{Var} [\Omega_k]. \quad (\text{F.10})$$

ce qui nous amène à

$$\mathbb{E} [\Omega_k^2] = \frac{\pi^2}{6} - \zeta(2, k-1) + \sum_{i=1}^{k-1} \beta_i^2 + \left(\gamma - \zeta(1; k-1) + \sum_1^{k-1} \beta_i \right)^2. \quad (\text{F.11})$$

Il est possible de minimiser $\mathbb{E} [\Omega_k^2]$ en fonction des β_i directement avec un multiplicateur de Lagrange λ afin de s'assurer que $\mathbb{E} [\Omega_k] = 0$:

$$\partial_{\beta_i} (\mathbb{E} [\Omega_k^2] - \lambda \mathbb{E} [\Omega_k]) = 2(\gamma - \zeta(1; k-1)) - \lambda + 2 \sum_{j=1}^{k-1} \beta_j + 2\beta_i. \quad (\text{F.12})$$

Une solution évidente est alors $\beta_i = \beta = \frac{\zeta(1; k-1) - \gamma}{k-1}$ et par conséquent :

$$\begin{cases} \forall i, i \neq k, & \alpha_i = \frac{\gamma - \sum_{l=1}^k \frac{1}{l}}{k-1} \\ & \alpha_k = 1 - (k-1)\alpha_1. \end{cases} \quad (\text{F.13})$$

2 Consistance

Nous pouvons maintenant vérifier que $\lim_k \mathbb{E} [\Omega_k^2] = 0$ pour les variables G_k

$$\mathbb{E} [\Omega_k^2] = \frac{\pi^2}{6} - \zeta(2, k-1) + \frac{(\zeta(1; k-1) - \gamma)^2}{k-1}. \quad (\text{F.14})$$

Le résultat vient immédiatement à partir de $\zeta(2) = \frac{\pi^2}{6}$. Utilisant la convergence en distribution de U_k , nous obtenons alors

$$\mathbb{E} [\Omega_k^2(U)] \xrightarrow[k]{} 0. \quad (\text{F.15})$$

3 Normalité asymptotique

On peut écrire Ω_k comme

$$\Omega_k = g_k + \beta \sum_{i=1}^{k-1} \Delta_i. \quad (\text{F.16})$$

À partir de l'Eq. (F.2), il est facile de vérifier que les variables Δ_i sont i.i.d., donc,

$$\frac{\Omega_k - g_k - \zeta(1; k-1) + \gamma}{\zeta(1; k-1) - \gamma} \xrightarrow{\mathcal{D}} \mathcal{N}_{0,1}. \quad (\text{F.17})$$

De plus, les relations

$$\mathbb{E} [g_k - \zeta(1; k-1) + \gamma] = 0, \quad (\text{F.18})$$

$$\text{Var} [g_k - \zeta(1; k-1) + \gamma] = \frac{\pi^2}{6} - \zeta(2; k-1), \quad (\text{F.19})$$

impliquent que

$$g_k - \zeta(1; k-1) + \gamma \xrightarrow{P} 0. \quad (\text{F.20})$$

Il est alors possible d'éliminer ce terme dans l'équation précédente :

$$\frac{\Omega_k}{\zeta(1; k-1) - \gamma} \xrightarrow{D} \mathcal{N}_{0,1}. \quad (\text{F.21})$$

La combinaison de ces résultats avec la convergence en distribution des U_k démontre qu'il existe une fonction $k(n)$ telle que Ω_k est asymptotiquement normal.

Liste de publications

- [1] F. Angeletti, E. Bertin, and P. Abry. Critical moment definition and estimation, for finite size observation of log-exponential-power law random variables. *Signal Processing*, 92 :2848 – 2865, 2012.
- [2] F. Angeletti, E. Bertin, and P. Abry. Matrix Products for the Synthesis of Stationary Time Series with a priori Prescribed Joint Distributions. In *Proceeding of the IEEE Int. Conf. on Acoust. Speech and Sig. Proc. (ICASSP)*, pages 3897 – 3900, 2012.
- [3] F. Angeletti, E. Bertin, and P. Abry. Renormalization flow for extreme value statistics of random variables raised to a varying power. *J.Phys A*, 45 :115004, 2012.
- [4] F. Angeletti, M. Mézard, E. Bertin, and P. Abry. Linearization effect in multifractal analysis : Insights from Random Energy Model analysis. *Physica D*, 240 :1245–1253, 2011.

Bibliographie

- [5] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch. Multiscale nature of network traffic. *IEEE Signal Proc. Mag.*, 19(3) :28–46, 2002.
- [6] P. Abry, V. Pipiras, and H. Wendt. Extreme Values, Heavy Tails and Linearization Effect : A Contribution to Empirical Multifractal Analysis. In *21st GRETSI Symposium on Signal and Image Processing*, Troyes, France, 2007.
- [7] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki. Valence bond ground states in isotropic quantum antiferromagnets. *Comm. Math. Phys*, 115(3) :477–528, 1988.
- [8] N. I Akhiezer. *The classical moment problem : And some related questions in analysis*. Oliver & Boyd, 1965.
- [9] A. Arneodo, E. Bacry, and J.-F. Muzy. The thermodynamics of fractals revisited with wavelets. *Physica A*, 213(1-2) :232–275, 1995.
- [10] G. Ben Arous, L.V. Bogachev, and S.A. Molchanov. Limit theorems for sums of random exponentials. *Probab. Theory Relat. Fields*, 132 :579–612, 2005.
- [11] E. Bacry, A. Gloter, M. Hoffmann, and J.-F. Muzy. Multifractal analysis in a mixed asymptotic framework. *Annals of Applied Probabilit*, 20(5) :1729–1760, 2010.
- [12] E. Bacry and J.F. Muzy. Log-infinitely divisible multifractal processes. *Commun. Math. Phys.*, 236 :449–475, 2003.
- [13] J. Barral and B. Mandelbrot. Multiplicative Products of Cylindrical Pulses. *Probab. Theory Relat. Fields*, 124 :409–430, 2002.
- [14] U. Basu and P. K. Mohanty. Active–absorbing-state phase transition beyond directed percolation : A class of exactly solvable models. *Phys. Rev. E*, 79 :041143, 2009.
- [15] B. A. Berg. *Markov chain Monte Carlo simulations and their statistical analysis*. World Scientific, 2004.
- [16] J. Bernoulli. *Ars Conjectandi*. 1713.
- [17] E. Bertin and G. Györgyi. Renormalization group theory for finite-size scaling in extreme statistics. *J. Stat. Mech.*, page P08022, 2010.
- [18] P. Billingsley. *Probability and Measure*. Wiley, 1995.
- [19] N. H. Bingham, C.M Goldie, and J.L. Teugels. *Regular variation*. Cambridge University Press, 1989.
- [20] N. H. Bingham and J. L. Teugels. Conditions implying domains of attraction. pages 23–34, Braşov, 1981. Sixth Conf. Probab. Th.

- [21] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [22] R. A. Blythe and M. R. Evans. Nonequilibrium steady states of matrix-product form : a solver's guide. *Journal of Physics A : Mathematical and Theoretical*, 40(46) :R333–R441, 2007.
- [23] L.V. Bogachev. Extreme value theory for random exponentials. *CRM Proceedings and Lecture Notes*, 42 :41, 2007.
- [24] P. Borgnat, P. Abry, and P. Flandrin. Using surrogates and optimal transport for synthesis of stationary multivariate series with prescribed covariance function and non-Gaussian joint-distribution. In *Proceeding of the IEEE Int. Conf. on Acoust. Speech and Sig. Proc. (ICASSP)*, Kyoto, 2012.
- [25] J.-P. Bouchaud. Weak ergodicity breaking and aging in disordered systems. *J. Phys. I (France)*, 2, 1992.
- [26] J.-P. Bouchaud and G. Biroli. On the Adam-Gibbs-Kirkpatrick-Thirumalai-Wolynes scenario for the viscosity increase in glasses. *J. Chem. Phys.*, 121 :7347, 2004.
- [27] J.-P. Bouchaud and M. Mézard. Universality classes for extreme value statistics. *J. Phys. A : Math. Gen.*, 30 :7997, 1997.
- [28] I. Calvo, J. C. Cuchí, J. G. Esteve, and F. Falceto. Extreme value distributions and Renormalization Group. *ArXiv :1109.5841*, 2011.
- [29] G. Cardano. *Liber de ludo aleae*. 1663.
- [30] D. Carpentier and P. Le Doussal. Glass transition of a particle in a random potential, front selection in nonlinear renormalization group, and entropic phenomena in Liouville and sinh-Gordon models. *Phys. Rev. E*, 63 :026110, 2001.
- [31] P. Chainais, R. Riedi, and P. Abry. On non scale invariant infinitely divisible cascades. *IEEE Transactions on Information Theory*, 51(3), 2005.
- [32] G. Chan and A.T.A. Wood. Simulation of stationary Gaussian vector fields. *Statistics and Computing*, 9(4) :265–268, 1999.
- [33] M.-H. Chen, Q.-M. Shao, and J. G. Ibrahim. *Monte Carlo methods in Bayesian computation*. Springer Series in Statistics. Springer-Verlag, New York, 2000.
- [34] M. Clusel and E. Bertin. Global fluctuations in physical systems : a subtle interplay between sum and extreme value statistics. *Int. J. Mod. Phys. B*, 22 :3311, 2008.
- [35] R.B. Davies and D.S. Harte. Tests for Hurst Effect. *Biometrika*, 74(4) :95–101, 1987.
- [36] N. G. De Bruijn. *Asymptotic Methods in Analysis*. Dover Publications, 1958.
- [37] P.-S. De Laplace. *Théorie analytique des probabilités*. 1812.
- [38] A. De Moivre. *Miscellanea Analytica*. 1730.
- [39] A. Dembo, C. L. Mallows, and L. A. Shepp. Embedding Nonnegative Definite Toeplitz Matrices in Nonnegative Definite Circulant Matrices, with Applications to Covariance Estimation. *IEEE Trans. Inform. Theory*, 35 :1206–1212, 1989.
- [40] B. Derrida. Random-Energy Model : an exactly solvable model of disordered systems. *Phys. Rev. B*, 24 :2613, 1981.

- [41] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier. Exact solution of a 1D asymmetric exclusion model using a matrix formulation. *J Phys. A*, 26 :1493–1517, 1993.
- [42] C. R. Dietrich and G.N. Newsam. Fast and Exact Simulation of Stationary Gaussian Processes through Circulant Embedding of the Covariance Matrix. *SIAM J. Sci. Comput.*, 18(4) :1088–1107, 1997.
- [43] B. Diu, C. Guthmann, D. Lederer, and B. Roulet. *Elements de Physique statistique*. Hermann, 1996.
- [44] R. S. Ellis. Large deviations for a general class of random vectors. *Ann. Probab.*, 12(1) :1–12, 1984.
- [45] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for insurance and finance*. Springer, 1997.
- [46] M. Evans, P. Ferrari, and K. Mallick. Matrix Representation of the Stationary Measure for the Multispecies TASEP. *Journal of Statistical Physics*, 135 :217–239, 2009.
- [47] M. Fannes, B. Nachtergaele, and R. F. Werner. Finitely correlated states on quantum spin chains. *Comm. Math. Phys.*, 144(3) :443–490, 1992.
- [48] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Procs. Cambridge Philos. Soc.*, 24 :180, 1928.
- [49] U. Frisch. *Turbulence, the Legacy of A.N. Kolmogorov*. Cambridge University Press, 1995.
- [50] U. Frisch and G. Parisi. Fully developed turbulence and intermittency. *Proc. Int. Summer School on Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, pages 84–88, 1985.
- [51] Y. V. Fyodorov and J.-P. Bouchaud. Freezing and extreme-value statistics in a random energy model with logarithmically correlated potential. *J. Phys. A : Math. Theor.*, 41 :372001, 2008.
- [52] Y. V. Fyodorov, P. Le Doussal, and A. Rosso. Statistical mechanics of logarithmic REM : duality, freezing and extreme value statistics of $1/f$ noises generated by Gaussian free fields. *J. Stat. Mech. : Theor. Exp.*, 2009 :P10005, 2009.
- [53] J. Galambos. *The Asymptotic Theory of Extreme Value Statistics*. John Wiley & Sons, 1978.
- [54] J. Gärtner. On large deviations from an invariant measure. *Teor. Veroyatnost. i Primenen.*, 22(1) :27–42, 1977.
- [55] B. V. Gnedenko. Sur la distribution limite du terme maximum d’une série aléatoire. *Ann. Math.*, 44 :423, 1943.
- [56] B.V. Gnedenko and A.N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley, 1968.
- [57] B. Grigelionis. On the extreme-value theory for stationary diffusions under power normalization. *Lithuanian Mathematical Journal*, 44 :36, 2004.
- [58] M. Grigoriu. *Applied Non-Gaussian Processes*. Prentice-Hall, 1995.

- [59] E. J. Gumbel. *Statistics of Extremes*. Dover Publications, 1958.
- [60] G. Györgyi, P. C. W. Holdsworth, B. Portelli, and Z. Rácz. Statistics of extremal intensities for Gaussian interfaces. *Phys. Rev. E*, 68 :056116, 2003.
- [61] G. Györgyi, N. R. Moloney, K. Ozogány, and Z. Rácz. Finite size scaling in extreme statistics. *Phys. Rev. Lett.*, 100 :210601, 2008.
- [62] G. Györgyi, N. R. Moloney, K. Ozogány, Z. Rácz, and M. Droz. Renormalization group theory for finite-size scaling in extreme statistics. *Phys. Rev. E*, 81 :041135, 2010.
- [63] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, 1980.
- [64] H. Helgason, V. Pipiras, and P. Abry. Synthesis of Multivariate Stationary Series with Prescribed Marginal Distributions and Covariance using Circulant Matrix Embedding. *Preprint*, 2010.
- [65] H. Helgason, V. Pipiras, and P. Abry. Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding. *Signal Processing*, 91(5) :1123–1133, 2011.
- [66] H. Helgason, V. Pipiras, and P. Abry. Synthesis of multivariate stationary series with prescribed marginal distributions and covariance using circulant embedding. *Signal Processing*, 91(8) :1741–1758, 2011.
- [67] S. Jaffard. Wavelet techniques in multifractal analysis. In *Fractal Geometry and Applications : A Jubilee of Benoît Mandelbrot, M. Lapidus et M. van Frankenhuysen Eds., Proceedings of Symposia in Pure Mathematics*, volume 72(2), pages 91–152. AMS, 2004.
- [68] A. Kagan and S. Nagaev. How many moments can be estimated from a large sample? *Statistics & Probability Letters*, 55(1) :99–105, 2001.
- [69] J. Karamata. Sur une mode de croissance régulière fonctions. *Math. (Cluj)*, 4 :38–53, 1930.
- [70] R. W. Katz, M. B. Parlange, and P. Naveau. Statistics of extremes in hydrology. *Adv. Water Resour.*, 25 :1287, 2002.
- [71] T. R. Kirkpatrick, D. Thirumalai, and P. G. Wolynes. Scaling concepts for the dynamics of viscous liquids near an ideal glassy state. *Phys. Rev. A*, 40 :1045, 1989.
- [72] K. Klauck and A. Schadschneider. On the ubiquity of matrix-product states in one-dimensional stochastic processes with boundary interactions. *Physica A : Statistical Mechanics and its Applications*, 271(1–2) :102 – 117, 1999.
- [73] S. A. Ladoucette and J. L. Teugels. Asymptotics for Ratios with Applications to Reinsurance. *Methodology and Computing in Applied Probability*, 9 :225–242, 2007.
- [74] B. Lashermes, P. Abry, and P. Chainais. New insight in the estimation of scaling exponents. *Int. J. on Wavelets, Multiresolution and Information Processing*, 2(4), 2004.
- [75] P. Le Doussal and C. Monthus. Exact solutions for the statistics of extrema of some random 1D landscapes, Application to the equilibrium and the dynamics of the toy model. *Physica A*, 317 :140, 2003.

- [76] F. Longin. From VaR to stress testing : the extreme value approach. *J. Bank. Finance*, 24 :1097, 2000.
- [77] E. Lukacs. *Characteristic Functions*. Griffin, 1987.
- [78] S. N. Majumdar and A. Comtet. Exact Maximal Height Distribution of Fluctuating Interfaces. *Phys. Rev. Lett.*, 92 :225501, 2004.
- [79] S. N. Majumdar and P. L. Krapvisky. Extremal Paths on a Random Cayley Tree. *Phys. Rev. E*, 62 :7735, 2000.
- [80] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, 1998.
- [81] K. Mallick and S. Sandow. Finite dimensional representations of the quadratic algebra : Applications to the exclusion process. *J. Phys. A : Math. Gen.*, 30 :4513, 1997.
- [82] B. B. Mandelbrot. Intermittent turbulence in self-similar cascades : divergence of high moments and dimension of the carrier. *J. Fluid Mech.*, 62 :331–358, 1974.
- [83] B. B. Mandelbrot. Negative Fractal Dimensions and Multifractals. *Physica A*, 163 :306–315, 1990.
- [84] B. B. Mandelbrot. A multifractal walk down Wall Street. *Sci. Am.*, 280(2) :70–73, 1999.
- [85] B. B. Mandelbrot. Multifractal power law distributions : negative and Critical Dimensions and other anomalies, explained by a simple example. *J. Stat. Phys.*, 110(3-6) :739–774, 2003.
- [86] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. A. Virasoro. Replica symmetry breaking and the nature of the spin glass phase. *J. de Physique*, 45 :843, 1984.
- [87] M. Mézard, G. Parisi, and M. Virasoro. *Spin Glass Theory and Beyond*. World Scientific Publishing Company, 1987.
- [88] N. R. Mohan and S. Ravi. Max domains of attraction of univariate and multivariate p-max stable laws. *Theory Probab. Appl.*, 37 :632, 1993.
- [89] G. M. Molchan. Scaling exponents and multifractal dimensions for independent random cascades. *Commun. Math. Phys.*, 179(3) :681–702, 1996.
- [90] G. M. Molchan. Turbulent cascades : Limitations and a statistical test of the lognormal hypothesis. *Phys. Fluids*, 9(8) :2387–2396, 1997.
- [91] J.-F. Muzy, E. Bacry, R. Baile, and P. Poggi. Uncovering latent singularities from multifractal scaling laws in mixed asymptotic regime. Application to turbulence. *EPL*, 82(6) :60007, 2008.
- [92] J.-F. Muzy, E. Bacry, and A. Kozhemyak. Extreme values and fat tails of multifractal fluctuations. *Phys. Rev. E*, 73 :066114, 2006.
- [93] C. Ngô and H. Ngô. *Physique statistique : Introduction*. Dunod, 2008.
- [94] J. R. Norris. *Markov Chains*. Cambridge University Press, 1998.
- [95] E. Olivieri and P. Picco. On the Existence of Thermodynamics for the Random Energy Model. *Commun. Math. Phys.*, 96 :125–144, 1984.

- [96] M. Ossiander and E. C. Waymire. Statistical estimation for multiplicative cascades. *Ann. Stat.*, 28(6) :1533–1560, 2000.
- [97] G. Paladin and A. Vulpiani. Anomalous scaling laws in multifractal objects. *Physics Reports*, 156(4) :147 – 225, 1987.
- [98] E. Pantcheva. Limit theorems for extreme order statistics under nonlinear normalization. *Lecture Notes in Math.*, 1155 :284, 1985.
- [99] K. A. Penson and K. Górska. Exact and Explicit Probability Densities for One-Sided Lévy Stable Distributions. *Phys. Rev. Lett.*, 105 :210604, 2010.
- [100] S. Ravi and A. S. Praveena. A Note on Tail Behaviour of Distributions in the Max Domain of Attraction of the Frechet/ Weibull Law under Power Normalization. *ProbStat Forum*, 3 :1, 2010.
- [101] R. H. Riedi. Multifractal Processes. In P. Doukhan, G. Oppenheim, and M.S. Taqqu, editors, *Theory and applications of long range dependence*, pages 625–717. Birkhäuser, 2003.
- [102] H. Royden. *Real Analysis*. Macmillan, 1968.
- [103] T. Schreiber and A. Schmitz. Surrogate time series. *Physica D*, 142(3-4) :346–382, 2000.
- [104] D. Sherrington and S. Kirkpatrick. Solvable Model of a Spin Glass. *Phys. Rev. Lett.*, 35 :1792–1796, 1975.
- [105] H. Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 478(1-3) :1–69, 2009.
- [106] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos. The Variational Approximation for Bayesian Inference. *IEEE Signal Processing Magazine*, 25(6) :131–146, 2008.
- [107] B. Védel, H. Wendt, P. Abry, and S. Jaffard. On the impact of the number of vanishing moments on the dependence structures of compound Poisson motion and fractional Brownian motion in multifractal time. In *Dependence in Probability and Statistics*, P. Doukhan, G. Lang, G. Teyssière, D. Surgailis (Eds.),. Springer, 2010.
- [108] H. Wendt, P. Abry, and S. Jaffard. Bootstrap for Empirical Multifractal Analysis. *IEEE Signal Processing Mag.*, 24(4) :38–48, 2007.
- [109] A. T. A. Wood and G. Chan. Simulation of stationary Gaussian Processes in $[0, 1]^d$. *J. of Comput. and Graph. Stat.*, 3(4) :409–432, 1994.

Résumé : Cette thèse s'est développée à l'interface entre physique statistique et traitement statistique du signal, afin d'allier les perspectives de ces deux disciplines sur les problèmes de sommes et maxima de variables aléatoires. Nous avons exploré trois axes d'études qui mènent à s'éloigner des conditions classiques (i.i.d.) : l'importance des événements rares, le couplage avec la taille du système, et la corrélation. Combinés, ces trois axes mènent à des situations dans lesquelles les théorèmes de convergence classiques sont mis en défaut. Pour mieux comprendre l'effet du couplage avec la taille du système, nous avons étudié le comportement de la somme et du maximum de variables aléatoires indépendantes élevées à une puissance dépendante de la taille du signal. Dans le cas du maximum, nous avons mis en évidence l'apparition de lois limites non standards. Dans le cas de la somme, nous nous sommes intéressés au lien entre effet de linéarisation et transition vitreuse en physique statistique. Grâce à ce lien, nous avons pu définir une notion d'ordre critique des moments, montrant que, pour un processus multifractal, celui-ci ne dépend pas de la résolution du signal. Parallèlement, nous avons construit et étudié, théoriquement et numériquement, les performances d'un estimateur de cet ordre critique pour une classe de variables aléatoires indépendantes. Pour mieux cerner l'effet de la corrélation sur le maximum et la somme de variables aléatoires, nous nous sommes inspirés de la physique statistique pour construire une classe de variable aléatoires dont la probabilité jointe peut s'écrire comme un produit de matrices. Après une étude détaillée de ses propriétés statistiques, qui a montré la présence potentielle de corrélation à longue portée, nous avons proposé pour ces variables une méthode de synthèse en réussissant à reformuler le problème en termes de modèles à chaîne de Markov cachée. Enfin, nous concluons sur une analyse en profondeur du comportement limite de leur somme et de leur maximum.

Abstract : This thesis has grown at the interface between statistical physics and signal processing, combining the perspectives of both disciplines to study the issues of sums and maxima of random variables. Three main axes, venturing beyond the classical (i.i.d) conditions, have been explored : The importance of rare events, the coupling between the behavior of individual random variable and the size of the system, and correlation. Together, these three axes have led us to situations where classical convergence theorems are no longer valid. To improve our understanding of the impact of the coupling with the system size, we have studied the behavior of the sum and the maximum of independent random variables raised to a power depending of the size of the signal. In the case of the maximum, we have brought to light non standard limit laws. In the case of the sum, we have studied the link between linearisation effect and glass transition in statistical physics. Following this link, we have defined a critical moment order such that for a multifractal process, this critical order does not depend on the signal resolution. Similarly, a critical moment estimator has been designed and studied theoretically and numerically for a class of independent random variables. To gain some intuition on the impact of correlation on the maximum or sum of random variables, following insights from statistical physics, we have constructed a class of random variables where the joint distribution probability can be expressed as a matrix product. After a detailed study of its statistical properties, showing that these variables can exhibit long range correlations, we have managed to recast this model into the framework of Hidden Markov Chain models, enabling us to design a synthesis procedure. Finally, we conclude by an in-depth study of the limit behavior of the sum and maximum of these random variables.
