



HAL
open science

Lois stables et processus ponctuels : liens et estimation des paramètres

Shuyan Liu

► **To cite this version:**

Shuyan Liu. Lois stables et processus ponctuels : liens et estimation des paramètres. Mathématiques [math]. Université des Sciences et Technologie de Lille - Lille I, 2009. Français. NNT : . tel-00463817

HAL Id: tel-00463817

<https://theses.hal.science/tel-00463817v1>

Submitted on 15 Mar 2010

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.

N° d'ordre : 40183



Université des Sciences et Technologies de Lille
Ecole Doctorale régionale Sciences Pour l'Ingénieur
Lille Nord-de-France

THÈSE

pour obtenir

LE GRADE DE DOCTEUR DE L'UNIVERSITÉ
DISCIPLINE : MATHÉMATIQUES APPLIQUÉES
SPÉCIALITÉ : PROBABILITÉS

présentée par

Shuyan LIU

LOIS STABLES ET PROCESSUS PONCTUELS : LIENS ET ESTIMATION DES PARAMÈTRES

soutenue le 10 décembre 2009 devant le jury composé de

Président :	Philippe SOULIER	Université Paris 10
Directeur de Thèse :	Youri DAVYDOV	Université Lille 1
Rapporteurs :	Jean-Marc BARDET	Université Paris 1
	Bernard GAREL	Institut National Polytechnique de Toulouse
Examineurs :	Johan SEGERS	Institut de statistique, UCL
	Radu STOICA	Université Lille 1
	Alain VIENNE	Université Lille 1

Résumé

L'objectif de cette thèse est d'étendre une méthode d'estimation des paramètres d'une loi stable dans \mathbb{R}^d aux lois à queue régulière dans un cône arbitraire. La méthode d'échantillonnage par paquets est modifiée afin d'optimiser la vitesse de convergence des estimateurs. Un nouvel estimateur de la masse totale de mesure spectrale est proposé. Des résultats sur les statistiques d'ordre des lois à queue régulière dans un cône et la loi des grands nombres pour le schéma triangulaire sont établis. La consistance et la normalité asymptotique des estimateurs sont démontrées. La performance des estimateurs est étudiée par simulation. On compare ces estimateurs avec quelques estimateurs connus. Les tableaux de performance sont fournis. La méthode de noyau est utilisée pour estimer la densité d'une mesure spectrale absolument continue d'une loi à queue régulière. On prouve la consistance de l'estimateur dans notre cas particulier. Pour augmenter le nombre de points utilisés dans l'échantillon, on propose une méthode d'estimation utilisant les permutations aléatoires de l'échantillon. La variation régulière a la propriété d'être préservée par plusieurs opérations et transformations. On considère trois sortes de transformations. Des conditions suffisantes pour cette préservation sont proposées et quelques contre-exemples sont présentés.

Les modèles de lois stables et de lois à queue lourde sont très utilisés dans plusieurs domaines d'application. On considère deux jeux de données réelles : les cours des 30 valeurs de l'indice DJIA et les perturbations planétaires des comètes du nuage de Oort. En appliquant la méthode d'estimation présentée on obtient des descriptions statistiques de ces données.

Abstract

The objective of this thesis is to extend an estimation method of parameters of a stable distribution in \mathbb{R}^d to the regularly varying tail distributions in an arbitrary cone. The sampling method of regrouping is modified to optimize the rate of convergence of estimators. A new estimator of total mass of the spectral measure is proposed. Some results about order statistics of regularly varying tail laws in a cone and the strong law of large numbers on the triangular schema are established. The consistency and the asymptotic normality of estimators are proved. The performance of proposed estimators is studied by simulation. We compare these estimators with some known estimators. The performance tables are provided. The kernel density estimation is used to estimate the density of an absolutely continuous spectral measure of a regularly varying tail law. We prove the consistency of the estimator in our particular case. To increase the number of points used in the sample, an estimation method using the random permutations of sample is proposed. The property of regular variation can be preserved by several operations and transformations. We consider three kinds of transformations. The sufficient conditions for this preservation are proposed and some counter-examples are presented.

The models of stable distributions and heavy tailed distributions are widely used in several application areas. We consider two sets of real data : the prices of 30 stocks of the DJIA index and the planetary perturbations of comets of the Oort cloud. By applying the estimation method presented previously we obtain some statistical descriptions of these data.

Remerciements

Je remercie tout à bord Youri DAVYDOV qui avant de diriger ma thèse, avait encadré mon mémoire de DEA, pour tout son dynamisme, ses compétences mathématiques et sa patience qui m'ont permis d'accomplir cette étude. Je suis très heureuse de le connaître dans ma vie. Pendant ces années, il était toujours disponible pour m'aider et me guider dans mon travail. Je suis profondément reconnaissante pour son soutien continu et ses nombreux conseils durant cette thèse.

Ensuite je remercie Radu STOICA pour m'avoir donné la possibilité de mener à bien une étude sur les applications de la méthode développée théoriquement. Son enthousiasme et son encouragement m'ont chaque fois permis de rebondir dans les moments difficiles. Je le remercie vivement pour son encadrement et les nombreuses discussions que nous avons eues.

Je remercie Jean-Marc BARDET pour avoir accepté d'être rapporteur de ce travail. Les conseils qu'il a apporté m'ont été très précieux pour structurer et pour améliorer la qualité des différentes sections.

Je remercie Bernard GAREL pour avoir accepté d'être rapporteur de ce travail. Ses remarques constructifs m'ont permis d'affiner ma rédaction et de clarifier certains points du manuscrit.

Je remercie Philippe SOULIER pour avoir accepté d'examiner ce travail avec diligence et pour l'honneur qu'il m'a fait de présider le jury de cette thèse.

Je remercie Alain VIENNE pour son soutien et pour avoir accepté de participer à ce jury.

Je remercie Johan SEGERS pour l'intérêt qu'il porte à ce travail et sa participation à ce jury.

Le travail sur les perturbations planétaires des comètes a été réalisé avec la précieuse aide des astronomes de l'observatoire de Lille. J'adresse mes remerciements à Marc FOUCHARD et Giovanni B. VALSECCHI pour leur collaboration.

Je remercie l'ensemble des membres du laboratoire Paul Painlevé, en particulier ceux de l'équipe de probabilité et statistiques. Je remercie tout spécialement Marie-Claude VIANO pour ses aides chaleureuses à chaque fois lorsque j'en avais besoin.

Je tient à remercier l'ensemble de mes camarades doctorants : Alexis, Anne, Bénédicte, Benoît, Chang, Eric, Julien, Léon, Martin, Oscar, Qidi, Raphaël, Vincent et Youcef, sans oublier ceux du début Siham et Frédéric. Merci également à mes autres amis qui m'ont encouragé à finir ce travail : Fabrice, Guillaume, Jie, Lily, Min, Qi et Qing.

J'ai la chance d'avoir passé quelques années ici où la culture est si différente que la nôtre. Merci enfin à mes parents pour leur confiance en moi depuis le début de cette expérience extraordinaire qui fait grandir.

Table des matières

Liste des notations	v
Introduction	1
1 Lois à queue régulière et lois stables dans un cône convexe	7
1.1 Préliminaires	7
1.1.1 Cônes convexes	7
1.1.2 Variation régulière	9
1.1.3 Processus ponctuels dans un cône	11
1.1.4 Lois $St\alpha\mathcal{S}$, série de LePage et domaines d'attraction	13
1.2 Transformations des lois à queue régulière	20
1.2.1 Transformations sphériques	20
1.2.2 Transformations radiales	22
1.2.3 Projections	25
1.2.4 Preuves	28
1.3 Simulation des vecteurs aléatoires à queue régulière	33
1.3.1 Lois α -stables	33
1.3.2 Lois appartenant au domaine d'attraction d'une loi $St\alpha\mathcal{S}$	40
1.3.3 Lois max-stables	43
2 Estimation des paramètres des lois à queue régulière	51
2.1 Estimation de l'indice de queue et de la mesure spectrale	51
2.1.1 Résultats auxiliaires	52
2.1.2 Définition et consistance des estimateurs	55
2.1.3 Normalité asymptotique	61
2.2 Estimation de la densité de mesure spectrale	72
2.2.1 Convergence de la densité de θ_{mi}	74
2.2.2 Estimation de la densité de σ par la méthode de noyau	77
2.3 Estimation basée sur l'échantillon permuté	85
2.4 Performance des estimateurs	90
2.4.1 Résultats d'estimation	90
2.4.2 Influence du regroupement d'échantillon	96
2.4.3 Illustration de la convergence des estimateurs	96
2.4.4 Comparaison par simulation avec d'autres méthodes	101
3 Applications	109

3.1	Quelques outils	109
3.2	Étude des cours des 30 valeurs de l'indice DJIA	111
3.2.1	Données unidimensionnelles	112
3.2.2	Données tri-variées	116
3.2.3	Amplitude des données	116
3.3	Étude des perturbations planétaires des comètes du nuage de Oort	120
3.3.1	Introduction	120
3.3.2	Données	120
3.3.3	Résultats	121
Conclusions et perspectives		131
A Bilan sur les estimateurs		133
A.1	Liste des variables et calcul des estimateurs	133
A.2	Consistance	134
A.3	Normalité asymptotique	134
B Programmes MATLAB		137
B.1	Programme pour estimer les paramètres : mvrfit.m	137
B.2	Programmes pour simuler les vecteurs aléatoires	140
B.2.1	stabsym1d.m	140
B.2.2	stabiso2d.m	141
B.2.3	maxstabuni.m	142
B.2.4	maxstab2p.m	143

Table des figures

1.1	Simulations des processus ponctuels poissonniens et empiriques de loi max-stable bivariee.	19
1.2	Influence des paramètres sur la forme de la loi stable.	35
1.3	Log-log graphe de $\mathbf{P}\{X > x\}$ où $X \sim \mathcal{S}_1(\alpha, (0, 1), 0)$	37
1.4	Convergence de la série de LePage.	40
1.5	Simulations des lois α -stables isotropiques bivariées.	41
1.6	Simulations des lois $\mathcal{S}_2(\alpha, \sigma_i, 0), i = 1, 2, 3$	42
1.7	Simulations des lois appartenant au $\text{Dom}(\alpha, \sigma_i), i = 1, 2, 3, \alpha = 0.75, 1.5$ par transformations sphériques (1.3.12).	43
1.8	Simulations des lois appartenant au $\text{Dom}(\alpha, \sigma_i), i = 1, 2, 3, \alpha = 0.75, 1.5$ par transformations radiales (1.3.13).	44
1.9	Simulations des lois $\mathcal{MS}_2(0.75, \sigma)$, dépendance de σ	48
1.10	Simulations des lois $\mathcal{MS}_2(\alpha, \sigma)$ où σ est définie par (1.3.17), dépendance de α	49

2.1	Estimation des densités de mesures spectrales des lois $\mathcal{S}_2(0.75, \sigma, 0)$	89
2.2	F.d.r. empiriques des mesures spectrales des lois $\mathcal{S}_2(\alpha, \sigma)$	93
2.3	F.d.r. empiriques des mesures spectrales des lois appartenant au domaine d'attraction d'une loi $\mathcal{St}\alpha\mathcal{S}$, simulées par transformation (1.3.12).	94
2.4	F.d.r. empiriques des mesures spectrales des lois appartenant au domaine d'attraction d'une loi $\mathcal{St}\alpha\mathcal{S}$, simulées par transformation (1.3.13).	94
2.5	F.d.r. empiriques des mesures spectrales des lois $\mathcal{MS}_2(0.75, \sigma)$	95
2.6	Influence du regroupement d'échantillon.	97
2.7	Résultats d'estimation pour des échantillons de taille variée.	99
2.8	Convergence des estimateurs $\hat{\alpha}_N, \widehat{\sigma(S)}_N$ et $\hat{\sigma}_N$	100
2.9	Comparaison des f.d.r. empiriques de mesure spectrale des lois $\mathcal{S}_2(\alpha, \sigma, 0)$ estimées par trois méthodes, où σ est uniforme ou définie par (1.3.9).	104
2.10	Comparaison des f.d.r. empiriques de mesure spectrale des lois $\mathcal{S}_2(\alpha, \sigma, 0)$ estimées par trois méthodes, où σ est définie par (1.3.10) et (1.3.11).	105
2.11	Comparaison des convergences des estimateurs différents.	106
3.1	Les paramètres α estimés pour les données de 30 composantes de l'indice DJIA.	113
3.2	Les paramètres β estimés pour les données de 30 composantes de l'indice DJIA.	114
3.3	Comparaison deux par deux des données de AIG, Citigroup et DIS.	117
3.4	Résultats d'estimation de la mesure spectrale des données 3-dimensionnelles de AIG, Citigroup et DIS.	118
3.5	a. Ellipse du mouvement képlérien. b. Éléments d'orbite.	121
3.6	Différences des quantiles empiriques pour les perturbations autour des grandes planètes.	123
3.7	Analyse de la symétrie en utilisant la différence des quantiles.	124
3.8	p -valeurs obtenues pour le test de normalité des quantiles empiriques $\hat{z}_{0.95}$ autour de l'orbite de Jupiter.	125
3.9	Validation de l'analyse basée sur l'indicateur de différence $\hat{z}_{0.99} - \hat{n}_{0.99}$ calculé autour de l'orbite de Jupiter.	126
3.10	Résultats d'estimation de l'indice de queue α sur les marques des perturbations.	127
3.11	Résultats d'estimation du paramètre d'échelle γ sur les marques des perturbations.	127
3.12	Résultats d'estimation du paramètre de symétrie β sur les marques des perturbations.	128
3.13	Résultats d'estimation du paramètre de position δ sur les marques des perturbations.	128
3.14	Les p -valeurs calculées autour de l'orbite de Jupiter pour le test des quantiles empiriques $\hat{z}_{0.99}$ des lois à queue régulière.	129
3.15	Les p -valeurs du test statistique de χ^2 pour les perturbations avec $\alpha < 2$	129
3.16	Les p -valeurs du test statistique de χ^2 pour les perturbations avec $\alpha \geq 2$	130

Liste des tableaux

2.1	Paramètres estimés des lois α -stables unidimensionnelles.	91
2.2	Paramètres estimés des lois α -stables bivariées.	91
2.3	Paramètres estimés des lois appartenant au domaine d'attraction d'une loi $St\alpha\mathcal{S}$	91
2.4	Paramètres estimés des lois max-stables bivariées d'indice $\alpha = 0.75$	92
2.5	Paramètres estimés des lois max-stables bivariées d'indice $\alpha = 1.5, 3$	92
2.6	Tableau de performance des estimateurs des paramètres des lois α -stables bivariées.	107
3.1	Résultats d'estimation et des tests pour les 30 composantes de l'indice DJIA.	115
3.2	Résultats d'estimation des paramètres des lois à queue régulière pour les données de trois composantes de l'indice DJIA : AIG, C et DIS.	117
3.3	Les p -valeurs obtenues du test des quantiles pour l'amplitude des données.	119

Liste des notations

$v.a.$	vecteur aléatoire
$e.a.$	élément aléatoire
\mathcal{B}	tribu borélienne
$Sta\mathcal{S}$	strictement α -stable
$\underline{\underline{\mathcal{L}}}$	égalité en loi
\Rightarrow	convergence faible
\xrightarrow{P}	convergence en probabilité
R_α	ensemble des fonctions à variation régulière d'indice α à l'infini
$R_\alpha(0+)$	ensemble des fonctions à variation régulière d'indice α à l'origine
$\text{Dom}(\alpha, \sigma)$	domaine d'attraction d'une loi α -stable de mesure spectrale σ
$\text{VR}(\alpha, \sigma)$	ensemble des éléments aléatoires à queue régulière d'indice α et de mesure spectrale σ
Γ_k	somme de k variables aléatoires i.i.d. de loi exponentielle standard
$\Lambda_{\alpha, \sigma}$	mesure d'intensité d'un processus ponctuel poissonnien
$\pi_{\alpha, \sigma}$	processus ponctuel poissonnien de mesure d'intensité $\Lambda_{\alpha, \sigma}$
β_n	processus ponctuel empirique
$f.d.r.$	fonction de répartition

Introduction

La variation régulière des mesures boréliennes est une propriété qu'on rencontre dans de nombreux domaines de la théorie des probabilités et ses applications. Un vecteur aléatoire (v.a.) X à valeurs dans \mathbb{R}^d a une loi à *queue régulière* d'exposant caractéristique $\alpha > 0$ si il existe une mesure finie σ sur la sphère unité S^{d-1} telle que $\forall B \in \mathcal{B}(S^{d-1})$ et $\sigma(\partial B) = 0$,

$$\lim_{x \rightarrow \infty} \frac{x^\alpha}{L(x)} \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| > x \right\} = \sigma(B), \quad (1)$$

où L est une fonction à variation lente, i.e. $\frac{L(\lambda x)}{L(x)} \rightarrow 1$ quand $x \rightarrow \infty$, $\forall \lambda > 0$. Cette condition apparaît souvent dans les études sur le comportement asymptotique des sommes partielles d'éléments i.i.d. (le théorème central limite général, voir [Rva62]), des maxima partiels et des processus ponctuels (la théorie des valeurs extrêmes, voir [Res87]).

La propriété de variation régulière peut être définie dans un espace plus général où l'addition des éléments et la multiplication par des nombres réels positifs sont bien définies, i.e. dans un espace qui a la structure d'un cône. Pour les mesures dans un espace métrique général, la notion appropriée de convergence et les résultats de base (théorème de portemanteau, théorème de convergence d'une fonction d'un élément aléatoire et caractérisation de la compacité relative) sont présentés dans [HL06].

La propriété de variation régulière est liée étroitement avec la caractérisation des domaines d'attraction des lois stables d'indice de stabilité α et de mesure spectrale σ . Un v.a. X à valeurs dans un espace de Banach a une loi *strictement α -stable* ($Sta\mathcal{S}$) si pour tous $a, b > 0$

$$a^{1/\alpha} X_1 + b^{1/\alpha} X_2 \stackrel{\mathcal{L}}{=} (a+b)^{1/\alpha} X, \quad (2)$$

où X_1, X_2 sont des copies indépendantes de X , et $\stackrel{\mathcal{L}}{=}$ représente l'égalité en loi. Cette définition a un sens dans tous les cônes convexes. Quel que soit le choix du cône, l'étude de la loi stable dans ce cône ramène toujours à la propriété de variation régulière. Dans \mathbb{R}^d la convergence (1) avec $\alpha \in (0, 2)$ est nécessaire et suffisante pour que X appartienne au domaine d'attraction d'une loi α -stable de mesure spectrale σ , ce que nous noterons " $X \in \text{Dom}(\alpha, \sigma)$ ". Dans un cône général muni d'une norme sous-invariante cette convergence avec $\alpha \in (0, 1)$ est suffisante pour que $X \in \text{Dom}(\alpha, \sigma)$ (voir Th. 4.7 [DMZ08]).

Les lois stables ont été proposées comme modèle pour de nombreux types de systèmes. Dans \mathbb{R}^d sauf les lois normales pour lesquelles $\alpha = 2$, les lois stables sont leptokurtiques et à queue lourde. Elles donnent souvent un très bon ajustement par exemple pour certains bruits de télécommunication [SK74] [NS95] et pour les rendements des actifs financiers [AB88] [MR93]. Bien qu'il existe d'autres lois à queue lourde - la loi t , la loi log-normale

ou la loi de Paréto - il y a au moins une bonne raison d'utiliser le modèle de lois stables : le théorème central limite généralisé dit que la seule limite possible non triviale des sommes normalisées des v.a. i.i.d. est une loi stable. De nombreuses quantités observées sont l'effet cumulé de plusieurs petits facteurs comme les cours de la bourse ou le bruit de communication, de sorte que le modèle de lois stables peut être utilisé.

Une des difficultés techniques dans l'étude des lois stables est que, sauf pour quelques cas particuliers, il n'y a pas de formule explicite pour la densité. D'ailleurs une loi stable multivariée est définie par l'indice de stabilité $\alpha \in (0, 2]$ et la mesure spectrale σ qui est une mesure finie sur la sphère unité S^{d-1} . Le problème d'estimation de la mesure spectrale est de nature non-paramétrique. Plusieurs méthodes d'estimation des paramètres des lois stables sont disponibles dans la littérature : méthode du maximum de vraisemblance [DuM71] [DuM73], méthode des moments [Zol86], méthode des quantiles [FR68] [FR71] [McC86], méthode de la fonction caractéristique empirique [Pre72b] [NPM01], etc.. Néanmoins, ces méthodes ont toutes le même inconvénient : elles sont valides si les données constituent un échantillon de loi stable. Si les données proviennent d'une loi différente, l'inférence sur l'indice de queue peut être fortement trompeuse. Dans la pratique, par précaution, on estime d'abord l'indice α en utilisant les méthodes d'estimation des lois à queue lourde, par exemple [Hil75]. Si la valeur estimée de α se trouve entre 0 et 2, les données seraient considérées comme un échantillon de loi stable, et les méthodes d'estimation présentées auparavant pourraient être utilisées pour estimer les autres paramètres. Remarquons que les lois α -stables de mesure spectrale σ sont des lois à queue régulière d'indice caractéristique α et de même mesure spectrale σ . Sous l'hypothèse que les données proviennent d'une loi à queue régulière, une fois que les paramètres sont estimés, les résultats contiennent en même temps l'indice de queue d'une loi stable ou d'une autre loi à queue lourde, et la mesure spectrale correspondante. Dans cette circonstance, l'estimation des lois à queue régulière devient intéressante.

On sait que la variation régulière multivariée est équivalente à la convergence des processus ponctuels empiriques vers un processus ponctuel poissonnien qui possède des propriétés très intéressantes et dont la mesure d'intensité a une forme particulière. Ce résultat est généralisé pour les processus ponctuels dans un cône [DMZ08]. Davydov et Paulauskas (voir [DP99] et [DPR00]) ont proposé une méthode pour estimer l'indice α et la mesure spectrale σ d'une loi stable dans \mathbb{R}^d en utilisant le lien entre les lois stables et les processus ponctuels. L'avantage principal de cette méthode est la simplicité d'usage de ces estimateurs et l'indépendance de leur expression par rapport à la fonction caractéristique. Le but de ce travail est de généraliser cette méthode aux lois à queue régulière dans un cône arbitraire. En ajoutant certaines conditions supplémentaires, nous prouvons la normalité asymptotique des estimateurs. La performance des estimateurs est étudiée par simulation.

Une autre partie de notre travail consiste à étudier quelques transformations des lois à queue régulière. La variation régulière a la propriété importante d'être préservée par plusieurs opérations et transformations. Ces propriétés des transformations sont utiles pour les simulations de vecteurs appartenant au domaine d'attraction d'une loi stable ayant une mesure spectrale donnée. Nous considérons trois sortes de transformations. Des conditions suffisantes pour cette préservation sont proposées et quelques contre-exemples sont présentés.

Les modèles de lois stables et de lois à queue lourde sont très utilisés dans plusieurs domaines d'application. Les méthodes d'estimation classiques fonctionnent souvent sous l'hypothèse de stabilité ou de queue lourde. La méthode développée dans cette thèse est valable sous l'hypothèse plus générale de variation régulière. Les résultats d'estimation nous donnent en même temps l'indice de queue qui distingue la loi stable des autres lois à queue lourde, et aussi la mesure spectrale correspondante. Nous appliquons dans le dernier chapitre cette méthode sur deux jeux de données réelles : les cours des 30 valeurs de l'indice DJIA et les perturbations planétaires des comètes du nuage de Oort.

Plan de la thèse

Chapitre 1

Nous commençons par un résumé des définitions et théorèmes de base concernant les cônes convexes et les semigroupes. Nous donnons dans la sous-section 1.1.2 plusieurs définitions équivalentes de la variation régulière d'une mesure dans un cône. L'équivalence entre la variation régulière multivariée et la convergence des processus ponctuels empiriques vers un processus ponctuel poissonnien est généralisée dans un cône [DMZ08]. On obtient le théorème central limite pour une suite d'éléments aléatoires i.i.d. dans un cône : la somme normalisée d'éléments aléatoires converge faiblement vers une représentation de type de LePage de la loi stable correspondante. Nous discutons ces résultats dans les sous-sections 1.1.3 et 1.1.4. Un exemple de loi max-stable multivariée par simulation des processus ponctuels correspondants est donné à la fin de la section 1.1.

Dans la section 1.2 nous étudions la préservation de régularité par des transformations. Les transformations sphériques, radiales et les projections sont considérées respectivement dans les sous-sections 1.2.1, 1.2.2 et 1.2.3. Des conditions suffisantes pour cette préservation sont proposées et quelques contre-exemples sont présentés.

La section 1.3 contient des exemples de données simulées qu'on utilise pour tester la performance des estimateurs. Vu la difficulté de simulation des v.a. α -stables en haute dimension surtout dans le cas où la mesure spectrale est absolument continue, nous simulons les v.a. $Sta\mathcal{S}$ en utilisant la série de LePage. C'est une méthode de simulation directe, mais la convergence de la série est lente si la mesure spectrale n'est pas symétrique. Ce phénomène est montré dans la sous-section 1.3.1 par simulation de lois stables unidimensionnelles. Des v.a. appartenant au domaine d'attraction d'une loi $Sta\mathcal{S}$ sont simulés dans la sous-section 1.3.2 par la transformation d'une loi $Sta\mathcal{S}$ de mesure spectrale uniforme, car, pour cette loi le problème de simulation a été parfaitement résolu. Des données suivant une loi max-stable sont aussi simulées en utilisant la série de LePage dans la sous-section 1.3.3.

Chapitre 2

Dans la section 2.1 nous généralisons la méthode d'estimation des paramètres des lois stables dans \mathbb{R}^d ([DP99] et [DPR00]) aux lois à queue régulière dans un cône arbitraire. La méthode d'échantillonnage par paquets présentée dans [DP99] et [DPR00] est modifiée

afin d'optimiser la vitesse de convergence des estimateurs. Un nouvel estimateur de la masse totale de la mesure spectrale est proposé sous la condition que la loi appartienne au domaine d'attraction normal d'une loi $St\alpha S$. D'abord, des résultats auxiliaires sur les statistiques d'ordre des lois à queue régulière dans un cône, ainsi que la loi forte des grands nombres pour un schéma triangulaire sont établis dans la sous-section 2.1.1. Ensuite la consistance des estimateurs est présentée dans la sous-section 2.1.2. En ajoutant des conditions asymptotiques du second ordre sur la loi marginale, nous prouvons la normalité asymptotique des estimateurs dans la sous-section 2.1.3.

L'estimateur de la mesure spectrale d'une loi à queue régulière proposé auparavant nous permet d'obtenir la fonction de répartition de la mesure spectrale. Pour estimer la densité d'une mesure spectrale absolument continue, l'estimation par la méthode du noyau de la densité d'un v.a. sur la sphère peut être utilisée. La difficulté dans notre cas est que la densité à estimer est la limite des densités marginales. La convergence des densités marginales est démontrée dans la sous-section 2.2.1. La consistance de l'estimateur de la densité est prouvée dans la sous-section 2.2.2.

D'après les études de performance précédentes, on constate que notre méthode est fiable asymptotiquement, mais qu'elle se détériore pour les petites tailles d'échantillon. Pour augmenter le nombre de points utilisés dans l'échantillon on propose une méthode d'estimation utilisant les permutations aléatoires d'échantillon dans la section 2.3.

La performance des estimateurs proposés est étudiée par simulation dans la section 2.4. Les données simulées dans la section 1.3 sont utilisées. Des résultats d'estimation en utilisant l'échantillonnage par paquets optimisé sont présentés dans la sous-section 2.4.1. Ensuite, en changeant la façon de former les paquets nous montrons l'influence du regroupement dans la sous-section 2.4.2. La vitesse de convergence des estimateurs est illustrée dans la sous-section 2.4.3. Quelques méthodes connues pour estimer les paramètres des lois α -stables sont présentées dans la sous-section 2.4.4. Nous comparons ces méthodes avec notre méthode sur des échantillons de taille différente et de différentes valeurs des paramètres. Les tableaux de performance sont fournis à la fin.

Chapitre 3

Ce chapitre commence par présenter une famille de lois à queue régulière qui est alternative aux lois stables et leurs statistiques d'ordre. Ces deux outils seront très utilisés pour nos problèmes suivants.

Dans la section 3.2 un jeu de données des cours des 30 valeurs de l'indice Dow Jones Industrial Average (DJIA) est considéré. Cet ensemble de données a été traité par Nolan dans [Nol05] pour tester si il est approximativement sous-gaussien et ensuite estimer ses paramètres. Dans la sous-section 3.2.1 nous effectuons la procédure d'estimation des paramètres des lois à queue régulière sur chaque composante. Nous comparons les résultats obtenus avec ceux obtenus par la méthode du maximum de vraisemblance pour les lois stables. En prenant quelques combinaisons des composantes nous formons des ensembles de données multivariées et nous étudions la structure de leur mesure spectrale. Les résultats sont montrés et discutés dans la sous-section 3.2.2. La distribution de la norme des données est étudiée à la fin.

Dans la section 3.3 des perturbations planétaires des comètes du nuage de Oort sont

considérées. Nous donnons dans cette section une description statistique de ces perturbations. La sous-section 3.3.2 présente la compilation des données. Le modèle de lois à queue régulière présenté dans le chapitre 1 s'adapte particulièrement à notre situation, car il prend en compte en même temps l'état d'engraissement de la queue et l'asymétrie de la loi. Les lois stables et une famille alternative de lois à queue régulière d'indice de queue $\alpha \geq 2$ sont proposées pour modéliser les données. Les résultats obtenus sont montrés et interprétés dans la sous-section 3.3.3.

Chapitre 1

Lois à queue régulière et lois stables dans un cône convexe

Dans ce chapitre nous présentons d'abord des résultats connus sur la variation régulière d'une mesure, les lois $\text{St}\alpha\mathcal{S}$ et les processus ponctuels dans un cône abstrait. Ce sont des études nécessaires pour généraliser une méthode d'estimation des paramètres d'une loi stable dans \mathbb{R}^d aux lois à queue régulière dans un cône. Ensuite nous considérons trois sortes de transformations par lesquelles la propriété de régularité est préservée. Cette préservation est utile pour simuler des données appartenant au domaine d'attraction d'une loi $\text{St}\alpha\mathcal{S}$. A la fin quelques exemples de simulation des v.a. de loi à queue régulière sont donnés.

1.1 Préliminaires

1.1.1 Cônes convexes

Nous résumons des définitions et théorèmes de base concernant les cônes convexes et les semigroupes. Pour la présentation plus détaillée nous renvoyons le lecteur à l'article [\[DMZ08\]](#).

Définition 1.1.1. *Un semigroupe abélien topologique est un espace topologique \mathbb{K} muni de l'opération binaire commutative et associative. Il est supposé que \mathbb{K} possède l'élément neutre \mathbf{e} satisfaisant $x + \mathbf{e} = x$ pour tout $x \in \mathbb{K}$.*

Définition 1.1.2. *Un cône convexe \mathbb{K} est un semigroupe abélien topologique, supposé complet et séparable, avec une opération de multiplication par des scalaires positifs, $(x, a) \rightarrow ax$, continue pour $x \in \mathbb{K}$ et $a > 0$ tel que les conditions suivantes sont remplies :*

- 1) $a(x + y) = ax + ay$, $a > 0$, $x, y \in \mathbb{K}$
- 2) $a(bx) = (ab)x$, $a, b > 0$, $x \in \mathbb{K}$
- 3) $1x = x$, $x \in \mathbb{K}$
- 4) $a\mathbf{e} = \mathbf{e}$, $a > 0$, \mathbf{e} est l'élément neutre de \mathbb{K} .

\mathbb{K} s'appelle cône pointé si il y a un élément unique $\mathbf{0}$ dit l'origine tel que $ax \rightarrow \mathbf{0}$ quand $a \downarrow 0$ pour tout $x \in \mathbb{K} \setminus \{\mathbf{e}\}$.

Remarquons qu'ici la condition de distributivité suivante n'est pas nécessaire

$$(a + b)x = ax + bx, \quad a, b > 0, \quad x \in \mathbb{K}. \quad (1.1.1)$$

On appelle (1.1.1) la *seconde loi de distributivité*. Cette condition n'est pas satisfaite pour le cône des sous-ensembles compacts de l'espace de Banach avec l'addition de Minkowski, c'est-à-dire

$$A + B := \{a + b : a \in A, b \in B\},$$

pour $A, B \in \mathcal{K}(\mathbb{B})$ où $\mathcal{K}(\mathbb{B})$ est la famille des sous-ensembles non-vides compacts de l'espace de Banach séparable \mathbb{B} . Cette condition n'est pas satisfaite non plus pour le cône $\mathbb{R}_+ = [0, \infty)$ avec l'opération d'addition définie par le maximum ou le minimum.

Le lemme suivant montre que l'élément neutre ne coïncide pas nécessairement avec l'origine.

Lemme 1.1.1. *Soit \mathbb{K} un cône pointé.*

- 1) *Si la seconde loi de distributivité a lieu, alors $\mathbf{e} = \mathbf{0}$.*
- 2) *Si il existe l'élément $x \neq \mathbf{e}$ qui possède l'inverse $(-x)$, i.e. $x + (-x) = \mathbf{e}$, alors $\mathbf{e} = \mathbf{0}$.*

Définition 1.1.3. *Un élément $z \in \mathbb{K}$ est α -stable avec $\alpha \neq 0$, si*

$$a^{1/\alpha}z + b^{1/\alpha}z = (a + b)^{1/\alpha}z$$

pour tout $a, b > 0$.

Dans la suite on note $\mathbb{K}(\alpha)$ l'ensemble des éléments α -stables de \mathbb{K} . Il est clair que $\mathbf{e}, \mathbf{0} \in \mathbb{K}(\alpha)$ pour tout $\alpha \neq 0$.

Définition 1.1.4. *Le cône \mathbb{K} s'appelle un cône normé si \mathbb{K} est métrisable par une distance d qui est homogène à l'origine, i.e. $d(ax, \mathbf{0}) = ad(x, \mathbf{0})$ pour tout $a > 0$ et $x \in \mathbb{K}$. La valeur $\|x\| = d(x, \mathbf{0})$ s'appelle la norme de x .*

Dans les sections suivantes on suppose que \mathbb{K} est un cône normé. Il est clair que $\|x\| = 0$ si et seulement si $x = \mathbf{0}$. De plus si $\mathbf{e} \neq \mathbf{0}$, alors la propriété 4) de la définition 1.1.2. implique que $\|\mathbf{e}\| = d(\mathbf{e}, \mathbf{0}) = \infty$. Il est donc essentiel d'admettre que d peut prendre la valeur infinie. Par exemple, si \mathbb{K} est le cône $\overline{\mathbb{R}}_+ = [0, \infty]$ où l'opération d'addition est définie par le minimum, noté $(\overline{\mathbb{R}}_+, \wedge)$, alors la distance euclidienne d'un élément non-zéro $x \in \mathbb{R}_+$ à ∞ (étant l'élément neutre) est infinie.

La boule ouverte de rayon r centrée au point $\mathbf{0}$ est notée par

$$B_r = \{x \in \mathbb{K} \mid \|x\| < r\}.$$

L'intérieur de son complément est noté par

$$B^r = \{x \mid \|x\| > r\}.$$

L'ensemble

$$S = \{x \mid \|x\| = 1\}$$

s'appelle la *sphère unité*. Remarquons que S est complète par rapport à la topologie induite par la distance d sur \mathbb{K} . L'existence d'une origine implique que $\|x\| < \infty$ pour tout $x \in \mathbb{K} \setminus \{\mathbf{e}\}$, ainsi \mathbb{K} admet la *décomposition polaire*. Cette décomposition est réalisée par la bijection $x \leftrightarrow (\|x\|, x/\|x\|)$ entre

$$\mathbb{K}' = \mathbb{K} \setminus \{\mathbf{0}, \mathbf{e}\}$$

et $(0, \infty) \times S$.

En plus de l'homogénéité de la distance, on a parfois besoin de la propriété suivante

$$d(x+h, x) \leq d(h, \mathbf{0}) = \|h\|, \quad x, h \in \mathbb{K}. \quad (1.1.2)$$

Alors la distance (ou la norme) dans \mathbb{K} est dite *sous-invariante*. Si \mathbb{K} est un groupe, alors une distance *invariante* (donc aussi sous-invariante) existe toujours, c'est-à-dire (1.1.2) a lieu avec l'égalité. Cette contrainte n'est pas triviale, par exemple le cône $(\overline{\mathbb{R}_+}, \wedge)$ avec la distance euclidienne ne vérifie pas cette propriété.

Lemme 1.1.2. *Si \mathbb{K} a une norme sous-invariante, alors $\mathbf{0} = \mathbf{e}$, et pour tout $\alpha \in (0, 1)$, \mathbf{e} est l'élément unique de norme finie qui appartient à $\mathbb{K}(\alpha)$.*

Les exemples typiques de cônes qui possèdent ces propriétés sont les espaces de Banach ou les cônes convexes dans l'espace de Banach; la famille des sous-ensembles compacts (ou compacts convexes) d'un espace de Banach avec l'addition de Minkowski [DPR00] [GH85] [GHZ82]; la famille des ensembles compacts dans \mathbb{R}^d avec l'opération d'addition définie par l'union [Mol05]; la famille des mesures finies avec l'addition conventionnelle et la multiplication scalaire [DVJ03] [Rac91]. Un autre exemple typique est l'espace $\mathbb{R}_+ = [0, \infty)$ avec l'opération d'addition définie par le maximum, i.e. $x + y = x \vee y = \max(x, y)$. Plus d'informations sur ces exemples et d'autres peuvent être trouvées dans [DMZ08].

1.1.2 Variation régulière

La notion de variation régulière est un des concepts de base qui apparaît de façon naturelle dans des différents contextes de la théorie des probabilités et ses applications. On donne ici la définition de la régularité pour l'élément aléatoire dans un cône \mathbb{K} . Rappelons d'abord la définition d'une fonction à variation régulière. On dit que L est *une fonction à variation régulière d'indice α à l'infini (respectivement à l'origine)* et on note $L \in R_\alpha$ (respectivement $L \in R_\alpha(0+)$) si

$$\frac{L(\lambda x)}{L(x)} \rightarrow x^\alpha, \quad \text{quand } x \rightarrow \infty \text{ (} x \rightarrow 0 \text{) pour tout } \lambda > 0.$$

En particulier, si $\alpha = 0$ la fonction L est dite *fonction à variation lente*.

Définition 1.1.5. La loi de l'élément aléatoire (e.a.) X dans \mathbb{K}' est dite à queue à variation régulière s'il existe une mesure finie σ sur la sphère unité S , un nombre $\alpha > 0$ (resp. $\alpha < 0$) et une fonction L à variation lente tels que

$$\lim_{x \rightarrow \infty} \frac{x^\alpha}{L(x)} \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| > x \right\} = \sigma(B), \quad (1.1.3)$$

$$\left(\text{resp. } \lim_{x \rightarrow 0^+} \frac{x^\alpha}{L(x)} \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| < x \right\} = \sigma(B), \right) \quad (1.1.4)$$

pour tout $B \in \mathcal{B}(S)$ avec $\sigma(\partial B) = 0$; ici $\|\cdot\|$ est la norme correspondante dans \mathbb{K} .

On appelle σ la mesure spectrale de X , et α s'appelle exposant de variation régulière. Le fait que X admet une queue à variation régulière d'indice α et de mesure spectrale σ sera dans la suite noté par l'écriture " $X \in \text{VR}(\alpha, \sigma)$ ". Pour simplifier on dit aussi X a une loi à queue régulière. La condition que X est à valeurs dans \mathbb{K}' assure que $0 < \|X\| < \infty$ presque sûrement. Remarquons qu'il est facile de déduire (1.1.4) de (1.1.3) ou en sens inverse en utilisant la transformation d'inverse $X \mapsto X/\|X\|^2$. Dans la suite on ne considère que le cas où $\alpha > 0$.

Il est clair que sans perdre de généralité on peut considérer σ comme une mesure normalisée, c'est-à-dire $\sigma(S) = 1$. En prenant $B = S$ dans (1.1.3), on déduit immédiatement que

$$\frac{x^\alpha}{L(x)} \mathbf{P}\{\|X\| > x\} \rightarrow 1, \quad x \rightarrow \infty, \quad (1.1.5)$$

c'est-à-dire que $\|X\|$ est une variable aléatoire positive de loi à queue régulière.

Les relations (1.1.3) et (1.1.5) donnent

$$\mathbf{P} \left\{ \frac{X}{\|X\|} \in B \mid \|X\| > x \right\} \rightarrow \sigma(B), \quad x \rightarrow \infty,$$

pour tout $B \in \mathcal{B}(S)$ avec $\sigma(\partial B) = 0$, ce qui signifie que la loi conditionnelle de $\frac{X}{\|X\|}$ sachant $\{\|X\| > x\}$ converge faiblement vers σ .

Il existe différentes caractérisations de la propriété $X \in \text{VR}(\alpha, \sigma)$ (voir, par exemple, Mikosch [Mik03]). On n'en donne ici que deux.

- 1) L'e.a. $X \in \text{VR}(\alpha, \sigma)$ si et seulement si il existe une fonction \tilde{L} à variation lente telle que pour tout $r > 0$ et $B \in \mathcal{B}(S)$ avec $\sigma(\partial B) = 0$

$$\lim_{n \rightarrow \infty} n \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| > r b_n \right\} = \sigma(B) r^{-\alpha}, \quad (1.1.6)$$

où $b_n = n^{1/\alpha} \tilde{L}(n)$.

- 2) Pour formuler le deuxième critère on passe aux coordonnées polaires et on identifie \mathbb{K}' avec le produit $(0, \infty) \times S$. On introduit les mesures Q_n et Q sur $\mathcal{B}((0, \infty) \times S)$

$$Q_n((r, \infty) \times B) = n \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| > r b_n \right\}, \quad (1.1.7)$$

$$Q = m_\alpha \times \sigma, \quad (1.1.8)$$

où $m_\alpha(dr) = \alpha r^{-\alpha-1} dr$. Soit \mathbb{C} la famille des fonctions continues bornées $f : \mathbb{K} \rightarrow \mathbb{R}$ pour lesquelles il existe un nombre $r > 0$ tel que $f(x) = 0$ pour tout $x \in B_r$.

Alors $X \in \text{VR}(\alpha, \sigma)$ si et seulement si

$$\int f dQ_n \rightarrow \int f dQ, \quad n \rightarrow \infty \quad (1.1.9)$$

pour toute $f \in \mathbb{C}$.

Remarquons que la convergence (1.1.9) est équivalente à la convergence vague pour \mathbb{K} localement compact.

On renvoie le lecteur aux monographies de Feller [Fel71], de Araujo & Giné [AG80], de Resnick [Res87] et de Bingham et al. [BGT87] pour la présentation exhaustive de la notion de variation régulière. Elle est liée étroitement avec la caractérisation des domaines d'attraction des lois stables multidimensionnelles (voir Araujo et al. [AG80], Samorodnitsky et Taqqu [ST94] et Davydov et al. [DMZ08]). On trouve plusieurs informations sur les propriétés et les applications dans les articles récents de Mikosch [Mik03], Basrak et al. [BDM02] et Jacobsen et al. [JMRS07].

1.1.3 Processus ponctuels dans un cône

Considérons un cône normé \mathbb{K} muni de la σ -algèbre borélienne $\mathcal{B}(\mathbb{K})$. Une *mesure de comptage* est une mesure qui peut être représentée par

$$m = \delta_{x_1} + \delta_{x_2} + \dots = \sum_i \delta_{x_i}$$

où δ_x représente la mesure de Dirac au point $x \in \mathbb{K}$ et $\{x_1, x_2, \dots\}$ est un ensemble de points dénombrable au maximum. Soit \mathcal{M}_0 (respectivement \mathcal{M}) la famille des mesures de comptage m sur $\mathcal{B}(\mathbb{K})$ telles que $m(B^r) < \infty$ (respectivement $m(B_r) < \infty$) pour tout $r > 0$. Pour couvrir ces deux cas avec la même notation, on note A_r au lieu de B^r (respectivement B_r) dans le cas où l'on considère les mesures dans \mathcal{M}_0 (respectivement \mathcal{M}). Alors nous avons toujours $m(A_r) < \infty$, c'est-à-dire seulement un nombre fini de points x_i du support de m se situent dans A_r .

Un *processus ponctuel* μ est une application mesurable d'un espace de probabilité à \mathcal{M}_0 (ou \mathcal{M}) muni de la σ -algèbre engendrée par les ensembles des mesures $m \in \mathcal{M}_0$ (ou $m \in \mathcal{M}$) telle que $m(B) = n$ pour les ensembles boréliens $B \subset \mathbb{K}$ et $n \geq 0$.

Soit Λ une mesure sur \mathbb{K} qui est finie sur A_r pour tout $r > 0$. Un processus ponctuel π est un *processus poissonnien* de mesure d'intensité Λ , si pour toute famille d'ensembles boréliens disjoints F_1, \dots, F_n , les variables aléatoires $\pi(F_1), \dots, \pi(F_n)$ sont indépendantes mutuellement et distribuées suivant les lois de Poisson d'espérance $\Lambda(F_1), \dots, \Lambda(F_n)$ respectivement.

Définition 1.1.6. (*Processus ponctuel poissonnien stable* $\pi_{\alpha, \sigma}$) Soit $\Lambda_{\alpha, \sigma}$ une mesure sur \mathbb{K}' définie par $m_\alpha \times \sigma$ où σ est une mesure finie sur $\mathcal{B}(S)$ et m_α est une mesure sur

$(0, \infty)$ telle que pour $\alpha \neq 0$ on a

$$\begin{cases} m_\alpha((r, \infty)) = r^{-\alpha} & \text{si } \alpha > 0, \\ m_\alpha((0, r)) = r^{-\alpha} & \text{si } \alpha < 0, \end{cases} \quad r > 0. \quad (1.1.10)$$

Le processus ponctuel poissonnien stable est un processus poissonnien de mesure d'intensité $\Lambda_{\alpha, \sigma}$, noté $\pi_{\alpha, \sigma}$.

La mesure σ s'appelle la *mesure spectrale*. L'importance du processus $\pi_{\alpha, \sigma}$ vient du fait qu'il est *stable* par rapport à l'addition appliquée aux mesures de comptage correspondantes (Th. 3.2 [DMZ08]). En effet, soient $\pi'_{\alpha, \sigma}$ et $\pi''_{\alpha, \sigma}$ des copies indépendantes de $\pi_{\alpha, \sigma}$, alors pour tous $a, b > 0$,

$$D_{a^{1/\alpha}}\pi'_{\alpha, \sigma} + D_{b^{1/\alpha}}\pi''_{\alpha, \sigma} \stackrel{\mathcal{L}}{=} D_{(a+b)^{1/\alpha}}\pi_{\alpha, \sigma} \quad (1.1.11)$$

où D_a est un automorphisme continu défini par $D_a : x \mapsto ax$. Ici $D_a\pi$ représente la mesure image de π , i.e. $(D_a\pi)(A) = \pi(D_a^{-1}A) = \pi(D_{a^{-1}}A)$ pour tout ensemble borélien A . En fait l'égalité (1.1.11) est équivalente à

$$a^{1/\alpha}\text{supp}(\pi'_{\alpha, \sigma}) \cup b^{1/\alpha}\text{supp}(\pi''_{\alpha, \sigma}) \stackrel{\mathcal{L}}{=} (a+b)^{1/\alpha}\text{supp}(\pi_{\alpha, \sigma}).$$

Cela signifie que $\text{supp}(\pi_{\alpha, \sigma})$ est un ensemble fermé aléatoire union-stable, i.e. un e.a. $\mathcal{St}\alpha\mathcal{S}$ dans le cône des ensembles fermés avec l'opération d'addition définie par l'union.

Pour tout $\alpha \neq 0$, $\pi_{\alpha, \sigma}$ a la représentation suivante (Th. 3.3 [DMZ08]) :

$$\pi_{\alpha, \sigma} \stackrel{\mathcal{L}}{=} \sum_{k=1}^{\infty} \delta_{c\Gamma_k^{-1/\alpha}\epsilon_k} \quad (1.1.12)$$

où $\{\lambda_k, k \geq 1\}$ et $\{\epsilon_k, k \geq 1\}$ sont deux suites indépendantes de variables aléatoires i.i.d. et $\Gamma_k = \lambda_1 + \dots + \lambda_k, k \geq 1$. Les variables aléatoires λ_k ont la loi exponentielle standard, les e.a. ϵ_k sont de même loi $\tilde{\sigma}(\cdot) = \frac{\sigma(\cdot)}{\sigma(S)^{1/\alpha}}$ et $c = \sigma(S)^{1/\alpha}$.

Soit $\{\xi_k, k \geq 1\}$ une suite d'e.a. i.i.d. à valeurs dans \mathbb{K} . Un *processus ponctuel empirique* est défini par

$$\beta_n = \sum_{k=1}^n \delta_{\xi_k/b_n}, \quad n \geq 1, \quad (1.1.13)$$

où $\{b_n, n \geq 1\}$ est une suite de constantes de normalisation telle que $b_n \rightarrow \infty$. On va montrer plus tard (voir (2.1.11)) que pour avoir de bonnes propriétés, les constantes b_n doivent avoir la forme suivante

$$b_n = n^{1/\alpha}L(n), \quad n \geq 1, \quad (1.1.14)$$

avec $\alpha \neq 0$ et $L(n)$ une fonction à variation lente à l'infini.

Le processus poissonnien $\pi_{\alpha, \sigma}$ est la limite faible du processus empirique β_n si la loi des ξ_k a la queue à variation régulière [Res87]. On donne ici ce résultat sous la forme de [DMZ08].

Théorème 1.1.3. Soient ξ, ξ_1, ξ_2, \dots des e.a. i.i.d. à valeurs dans \mathbb{K}' . Alors $\beta_n \Rightarrow \pi_{\alpha, \sigma}$ quand $n \rightarrow \infty$ pour $\alpha > 0$ si et seulement si $\xi \in \text{VR}(\alpha, \sigma)$, c'est-à-dire s'il existe une mesure finie σ sur $\mathcal{B}(S)$ telle que la convergence (1.1.6) a lieu pour tout $r > 0$ et $B \in \mathcal{B}(S)$ avec $\sigma(\partial B) = 0$.

Remarque 1. La condition (1.1.6) est typique pour les théorèmes limites de sommes d'éléments aléatoires, voir [Res87] page 154 et [AG80] page 167. Il existe un résultat similaire pour $\alpha < 0$ (Cor. 4.4 [DMZ08]).

Ce résultat qui s'appelle parfois *la transformation de Poisson* [Res06] montre que la variation régulière multivariée a une équivalence probabiliste exacte en terme de convergence des mesures empiriques vers une mesure aléatoire poissonnienne limite.

1.1.4 Lois $\text{St}\alpha\mathcal{S}$, série de LePage et domaines d'attraction

Un e.a. X dans \mathbb{K} a une loi *strictement α -stable* ($\text{St}\alpha\mathcal{S}$) s'il vérifie la condition (2) de l'introduction avec les opérations d'addition et multiplication définies sur \mathbb{K} . Le théorème ci-dessous donne une famille riche de lois $\text{St}\alpha\mathcal{S}$ par leur décomposition en série.

Théorème 1.1.4. (Th. 3.6 [DMZ08]) Soient $\{\lambda_k, k \geq 1\}$ et $\{\epsilon_k, k \geq 1\}$ deux suites indépendantes de variables aléatoires i.i.d.. Les variables aléatoires λ_k ont la loi exponentielle standard, les e.a. ϵ_k sont de même loi $\tilde{\sigma}(\cdot)$ qui est une mesure finie normalisée sur S . Notons $\Gamma_k = \lambda_1 + \dots + \lambda_k$, $k \geq 1$. Si la valeur principale de l'intégrale $\int x \pi_{\alpha, \tilde{\sigma}}(dx)$ est finie, i.e. $\int_{A_r} x \pi_{\alpha, \tilde{\sigma}}(dx) < \infty$, avec probabilité 1, alors pour tout $z \in \mathbb{K}(\alpha)$ et $c \geq 0$, la série

$$\xi_{\alpha, \sigma} = z + c \sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \epsilon_k \quad (1.1.15)$$

converge presque sûrement (p.s.) et $\xi_{\alpha, \sigma}$ admet une loi $\text{St}\alpha\mathcal{S}$ sur \mathbb{K} . Le paramètre α s'appelle l'exposant caractéristique et la mesure $\sigma(\cdot) = c^\alpha \tilde{\sigma}(\cdot)$ s'appelle la mesure spectrale.

Si la norme de \mathbb{K} est sous-invariante, alors la somme infinie (1.1.15) converge absolument p.s. pour tout $\alpha \in (0, 1)$.

Définition 1.1.7. Un e.a. X à valeurs dans \mathbb{K} appartient au domaine d'attraction d'un e.a. $\text{St}\alpha\mathcal{S}$, noté $\xi_{\alpha, \sigma}$, si pour $\{X_n, n \geq 1\}$, une suite de copies indépendantes de X , on a

$$b_n^{-1}(X_1 + \dots + X_n) \Rightarrow \xi_{\alpha, \sigma}, \quad (1.1.16)$$

où $\{b_n, n \geq 1\}$ est une suite de constantes de normalisation positives et \Rightarrow signifie la convergence faible des e.a. à valeurs dans \mathbb{K} .

Le fait que l'e.a. X appartient au domaine d'attraction de $\xi_{\alpha, \sigma}$ sera noté par l'écriture " $X \in \text{Dom}(\alpha, \sigma)$ ". Le théorème suivant montre qu'avec une condition supplémentaire sur \mathbb{K} la condition $X \in \text{VR}(\alpha, \sigma)$ implique $X \in \text{Dom}(\alpha, \sigma)$ pour $0 < \alpha < 1$.

Théorème 1.1.5. (Th. 4.7 [DMZ08]) Supposons que \mathbb{K} a la norme sous-invariante. Si $X \in \mathbb{K}'$ satisfait la condition (1.1.6) avec $\alpha \in (0, 1)$, alors X appartient au domaine

d'attraction d'un e.a. $St\alpha S$, noté $\xi_{\alpha,\sigma}$, qui admet la représentation de type de LePage suivante

$$\xi_{\alpha,\sigma} \stackrel{\mathcal{L}}{=} c \sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \epsilon_k, \quad (1.1.17)$$

où Γ_k , ϵ_k et c sont définis comme dans le théorème 1.1.4..

Idée de la démonstration : Il suffit de montrer la convergence du genre de (1.1.16) vers la somme $c \sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \epsilon_k$. La somme normalisée $b_n^{-1}(X_1 + \dots + X_n)$ peut être interprétée comme l'intégrale $\int x \beta_n(dx)$, et la somme $c \sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \epsilon_k$ peut être écrite aussi comme $\int x \pi_{\alpha,\sigma}(dx)$. Donc pour prouver $X \in \text{Dom}(\alpha, \sigma)$ il suffit de montrer la convergence faible de la fonctionnelle $f(\beta_n)$ vers $f(\pi_{\alpha,\sigma})$ où $f(\mu) = \int x \mu(dx)$. Pour cela on a besoin de la convergence $\beta_n \Rightarrow \pi_{\alpha,\sigma}$ et

$$\limsup_n \mathbf{P} \left\{ \left\| \int_{B_r} x \beta_n(dx) \right\| \geq \varepsilon \right\} \rightarrow 0 \text{ quand } r \rightarrow 0 \quad (1.1.18)$$

pour tout $\varepsilon > 0$ (voir [DMZ08] Lemme 4.5).

D'après la convergence de la somme des points ordonnés du support de β_n vers celle de $\pi_{\alpha,\sigma}$ on peut montrer

$$\left\| \int_{B_r} x \beta_n(dx) \right\| \leq \int_{B_r} \|x\| \beta_n(dx) \rightarrow \int_{B_r} \|x\| \pi_{\alpha,\sigma}(dx) \quad (1.1.19)$$

quand $n \rightarrow \infty$ avec probabilité 1. Puisque la norme de \mathbb{K} est sous-invariante l'intégrale $\int \|x\| \pi_{\alpha,\sigma}(dx)$ converge presque sûrement pour tout $\alpha \in (0, 1)$. La convergence (1.1.18) a lieu d'après (1.1.19) et le lemme de Fatou. \square

Le résultat précédent confirme que la condition suffisante pour qu'un v.a. dans \mathbb{R}^d appartienne au $\text{Dom}(\alpha, \sigma)$ est aussi valable dans le cône convexe.

On résume les relations entre les processus ponctuels et la convergence des sommes des e.a. par le diagramme suivant.

$$\begin{array}{ccc} \forall B \in \mathcal{B}(S), \sigma(\partial B) = 0, & & \\ n\mathbf{P} \left\{ \frac{\xi}{\|\xi\|} \in B, \|\xi\| > r b_n \right\} \rightarrow \sigma(B) r^{-\alpha} & & \\ & \begin{array}{c} \downarrow \uparrow \\ \Rightarrow \end{array} & \\ \beta_n = \sum_{k=1}^n \delta_{\xi_k/b_n} & \Rightarrow & \Pi_{\alpha,\sigma} \stackrel{\mathcal{L}}{=} \sum_{k=1}^{\infty} \delta_{\Gamma_k^{-1/\alpha} \epsilon_k c} \\ \downarrow f(\mu) = \int x \mu(dx) & & \downarrow f(\mu) = \int x \mu(dx) \\ f(\beta_n) = b_n^{-1}(\xi_1 + \dots + \xi_n) & \Rightarrow & f(\pi_{\alpha,\sigma}) = c \sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \epsilon_k \end{array}$$

L'existence de la mesure spectrale d'une loi $\mathcal{St}\alpha\mathcal{S}$ est connue dans un espace de Banach séparable. Afin de prouver l'existence de la représentation de LePage dans un semigroupe plus général, la technique de l'analyse harmonique des semigroupes est utilisée [DMZ08], et les e.a. $\mathcal{St}\alpha\mathcal{S}$ sont caractérisés. On donne ici un exemple de loi $\mathcal{St}\alpha\mathcal{S}$ dans $\mathbb{K} = (\mathbb{R}_+^d, \vee)$.

Exemple 1. *Notation :* Les vecteurs dans $\mathbb{R}_+^d = [0, \infty)^d$ sont notés par $\mathbf{x} = (x_1, \dots, x_d)$. Les relations et opérations sont définies par composante, c'est-à-dire pour $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^d$

$$\begin{aligned} \mathbf{x} < \mathbf{y} & \text{ signifie } x_i < y_i, i = 1, \dots, d, \\ \mathbf{x} \leq \mathbf{y} & \text{ signifie } x_i \leq y_i, i = 1, \dots, d. \end{aligned}$$

L'addition et la multiplication par des scalaires sont définies par

$$\begin{aligned} \mathbf{x} + \mathbf{y} & := \mathbf{x} \vee \mathbf{y} = (x_1 \vee y_1, \dots, x_d \vee y_d), \\ a\mathbf{x} & := (ax_1, \dots, ax_d), \quad a > 0. \end{aligned}$$

Les rectangles sont notés par

$$\begin{aligned} (\mathbf{a}, \mathbf{b}) & = \{(x_1, \dots, x_d) \mid a_1 < x_1 < b_1, \dots, a_d < x_d < b_d\}, \\ [\mathbf{a}, \mathbf{b}] & = \{(x_1, \dots, x_d) \mid a_1 \leq x_1 \leq b_1, \dots, a_d \leq x_d \leq b_d\}. \end{aligned}$$

Prenons la L_∞ -norme, i.e. $\|\mathbf{x}\| = \max(x_1, \dots, x_d)$. La sphère unité dans cette norme $\{\mathbf{x} \mid \|\mathbf{x}\| = 1\}$ est notée par S_{\vee}^{d-1} .

a) Prenons d'abord la mesure spectrale définie par

$$\sigma(\cdot) = \sum_{i=1}^d p_i \delta_{e_i}(\cdot),$$

où $e_i = (e_i^{(1)}, \dots, e_i^{(d)})$, $e_i^{(j)} = \begin{cases} 1, & i = j \\ 0, & \text{sinon} \end{cases}$ et $\sum_{i=1}^d p_i = 1$, $p_i \geq 0$, $i = 1, \dots, d$.

La mesure σ est concentrée sur les points où les axes et la sphère unité se croisent. La série de LePage dans ce cas converge absolument presque sûrement pour tout $\alpha > 0$. Elle produit un v.a. max-stable. Définissons un v.a. par

$$\xi_{\alpha, \sigma} \stackrel{\mathcal{L}}{=} \bigvee_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \epsilon_k \tag{1.1.20}$$

où Γ_k est la somme des k variables aléatoires i.i.d. exponentielles standards, ϵ_k , $k = 1, 2, \dots$, sont les v.a. i.i.d. de loi σ et les deux suites $\{\Gamma_k\}$ et $\{\epsilon_k\}$ sont indépendantes. En notant

$$\epsilon_k = (\epsilon_k^{(1)}, \dots, \epsilon_k^{(d)})$$

on a

$$\xi_{\alpha, \sigma} \stackrel{\mathcal{L}}{=} \left(\bigvee_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \epsilon_k^{(1)}, \dots, \bigvee_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \epsilon_k^{(d)} \right) = (\Gamma_{\tau_1}^{-1/\alpha}, \dots, \Gamma_{\tau_d}^{-1/\alpha}),$$

où $\tau_i = \min\{k \mid \epsilon_k^{(i)} = 1\}$, $i = 1, \dots, d$. Pour chaque i , $\{\epsilon_k^{(i)}\}$ est une suite de variables aléatoires i.i.d. à valeurs dans $\{0, 1\}$ de la loi $\sigma^{(i)}(\cdot) = p_i \mathbb{1}_{\{1\}}(\cdot)$. En notant $\xi_{\alpha, \sigma} = (\xi_{\alpha, \sigma}^{(1)}, \dots, \xi_{\alpha, \sigma}^{(d)})$ considérons les lois marginales

$$\begin{aligned}
\mathbf{P}\{\xi_{\alpha, \sigma}^{(i)} \leq x_i\} &= \mathbf{P}\{\Gamma_{\tau_i}^{-1/\alpha} \leq x_i\} \\
&= \sum_{k=1}^{\infty} \mathbf{P}\{\Gamma_k^{-1/\alpha} \leq x_i\} \mathbf{P}\{\tau_i = k\} \\
&= \sum_{k=1}^{\infty} \int_{x_i^{-\alpha}}^{\infty} \frac{t^{k-1} e^{-t}}{\Gamma(k)} dt (1-p_i)^{k-1} p_i \\
&= \int_{x_i^{-\alpha}}^{\infty} e^{-t} p_i e^{t(1-p_i)} dt \\
&= e^{-p_i x_i^{-\alpha}}.
\end{aligned} \tag{1.1.21}$$

D'un autre côté définissons le processus ponctuel poissonnien

$$\pi_{\alpha, \sigma}(\cdot) = \sum_{k=1}^{\infty} \delta_{\Gamma_k^{-1/\alpha} \epsilon_k}(\cdot). \tag{1.1.22}$$

La mesure d'intensité est $m_{\alpha} \times \sigma$ où m_{α} est la mesure sur \mathbb{R}_+^1 telle que $m_{\alpha}((r, \infty)) = r^{-\alpha}$ pour tout $r > 0$. On en déduit la fonction de répartition de $\xi_{\alpha, \sigma}$, pour $\mathbf{x} > 0$

$$\begin{aligned}
\mathbf{P}\{\xi_{\alpha, \sigma} \in [0, \mathbf{x}]\} &= \mathbf{P}\{\pi_{\alpha, \sigma}([0, \mathbf{x}]^{\mathbb{G}}) = 0\} \\
&= \exp(-m_{\alpha} \times \sigma([0, \mathbf{x}]^{\mathbb{G}})) \\
&= \exp\left(-\sum_{i=1}^d p_i x_i^{-\alpha}\right).
\end{aligned}$$

La formule (1.1.21) peut être aussi déduite en utilisant le processus ponctuel associé

$$\pi_{\alpha, \sigma^{(i)}} = \sum_{k=1}^{\infty} \delta_{\Gamma_k^{-1/\alpha} \epsilon_k^{(i)}}, \quad i = 1, \dots, d. \tag{1.1.23}$$

Par la théorie de processus d'amincissement (voir par exemple [Res92]), les processus $\pi_{\alpha, \sigma^{(i)}}$ sont indépendants et poissonniens de mesure d'intensité $m_{\alpha} \times \sigma^{(i)}$ respectivement. Alors les probabilités marginales de $\xi_{\alpha, \sigma}$ sont

$$\mathbf{P}\{\xi_{\alpha, \sigma}^{(i)} \leq x_i\} = \mathbf{P}\{\pi_{\alpha, \sigma^{(i)}}((x_i, \infty)) = 0\} = \exp(-p_i x_i^{-\alpha}).$$

En conclusion, le v.a. max-stable $\xi_{\alpha, \sigma}$ a ses composantes indépendantes si la mesure spectrale est concentrée sur l'intersection des axes et la sphère unité. C'est un résultat connu qu'on peut trouver par exemple dans [Res87] (Cor. 5.25). La colonne de gauche de la Figure 1.1 présente les courbes de niveau de la densité de loi max-stable bivariable et les simulations des processus ponctuels β_n et $\pi_{\alpha, \sigma}$ avec $\alpha = 0.75$ et σ concentrée sur les points $e_i, i = 1, 2$.

b) Cette fois-ci la mesure spectrale σ est uniforme sur $S_{\sqrt{V}}^{d-1}$.

Le v.a. $\xi_{\alpha,\sigma}$ et les processus ponctuels $\pi_{\alpha,\sigma}$ et $\pi_{\alpha,\sigma^{(i)}}$ sont définis comme avant par (1.1.20) (1.1.22) et (1.1.23). On montre d'abord que dans ce cas la loi des $\epsilon_k^{(i)}$, $i = 1, \dots, d$ est

$$\sigma^{(i)}(\cdot) = \frac{1}{d} \mathbb{I}_{\{1\}}(\cdot) + \mu(\cdot) \mathbb{I}_{\mathcal{B}([0,1])}(\cdot), \quad (1.1.24)$$

où μ est une mesure sur $[0, 1)$ absolument continue par rapport à la mesure de Lebesgue λ de densité $\frac{d-1}{d}$, i.e. $d\mu/d\lambda = \frac{d-1}{d}$.

Notons

$$S_i = \{\mathbf{x} \mid x_i = 1, \mathbf{x} \in S_{\sqrt{V}}^{d-1}\}, \quad i \in I = \{1, \dots, d\}.$$

En fait pour chaque i l'ensemble S_i est une face de la sphère unité. La mesure de Lebesgue sur \mathbb{R}^{d-1} est notée λ^{d-1} . Alors on a

$$S_{\sqrt{V}}^{d-1} = \bigcup_{i=1}^d S_i, \quad \lambda^{d-1}(S_i) = \lambda^{d-1}(S_j), \quad i \neq j,$$

et pour l'ensemble $A \subset I$ et $A \neq \emptyset$

$$\lambda^{d-1}\left(\bigcap_{i \in A} S_i\right) = \lambda^{d-1}(\{\mathbf{x} \mid x_i = 1, i \in A, \mathbf{x} \in S_{\sqrt{V}}^{d-1}\}) = 0.$$

Donc

$$\lambda^{d-1}(S_{\sqrt{V}}^{d-1}) = \sum_{i=1}^d \lambda^{d-1}(S_i).$$

On en déduit

$$\mathbf{P}\{\epsilon_k^{(i)} = 1\} = \mathbf{P}\{\epsilon_k \in S_i\} = \frac{\lambda^{d-1}(S_i)}{\lambda^{d-1}(S_{\sqrt{V}}^{d-1})} = \frac{1}{d}. \quad (1.1.25)$$

Puisque ϵ_k est un v.a. uniforme sur $S_{\sqrt{V}}^{d-1}$, pour $i \neq j$ la loi $\mathbf{P}\{\epsilon_k^{(i)} \in \cdot \mid \epsilon_k \in S_j\}$ est la mesure de Lebesgue sur l'intervalle $[0, 1]$. Donc pour l'ensemble $B \in \mathcal{B}([0, 1])$

$$\begin{aligned} \mathbf{P}\{\epsilon_k^{(i)} \in B\} &= \sum_{j \in I \setminus \{i\}} \mathbf{P}\{\epsilon_k^{(i)} \in B \mid \epsilon_k \in S_j\} \mathbf{P}\{\epsilon_k \in S_j\} \\ &= \frac{d-1}{d} \lambda(B). \end{aligned} \quad (1.1.26)$$

Les égalités (1.1.25) et (1.1.26) impliquent (1.1.24).

Considérons les probabilités marginales de $\xi_{\alpha,\sigma}$

$$\begin{aligned} \mathbf{P}\{\xi_{\alpha,\sigma}^{(i)} \leq x_i\} &= \mathbf{P}\{\pi_{\alpha,\sigma^{(i)}}((x_i, \infty)) = 0\} \\ &= \exp(-m_{\alpha} \times \sigma^{(i)}((x_i, \infty))) \\ &= \exp\left(-\frac{\alpha + d}{(\alpha + 1)d} x_i^{-\alpha}\right). \end{aligned}$$

En effet en notant η la variable aléatoire de loi m_α , on a

$$\begin{aligned}
m_\alpha \times \sigma^{(i)}((x_i, \infty)) &= \mathbf{P}\{\eta \epsilon_k^{(i)} > x_i\} \\
&= \mathbf{P}\{\eta > x_i, \epsilon_k^{(i)} = 1\} + \mathbf{P}\{\eta \epsilon_k^{(i)} > x_i, \epsilon_k^{(i)} \in [0, 1)\} \\
&= \frac{1}{d}(x_i)^{-\alpha} + \int_{x_i}^{\infty} \int_x^{\infty} \alpha t^{-\alpha-1} \frac{d-1}{d} / t dt dx \\
&= \frac{\alpha + d}{(\alpha + 1)d} x_i^{-\alpha}.
\end{aligned}$$

De manière analogue on déduit la fonction de répartition de $\xi_{\alpha, \sigma}$, pour $\mathbf{x} > 0$

$$F_{\xi_{\alpha, \sigma}}(\mathbf{x}) = \mathbf{P}\{\xi_{\alpha, \sigma} \in [0, \mathbf{x}]\} = \mathbf{P}\{\pi_{\alpha, \sigma}([0, \mathbf{x}]^{\mathfrak{L}}) = 0\} = \exp\{-\Lambda_{\alpha, \sigma}([0, \mathbf{x}]^{\mathfrak{L}})\},$$

où $\Lambda_{\alpha, \sigma}$ est la mesure d'intensité d'un processus poissonnien. Dans le cas où $d = 2$ l'expression de la loi précédente est

$$F_{\xi_{\alpha, \sigma}}(x_1, x_2) = \exp\left(-\frac{1}{2(\alpha + 1)}(\alpha x_{(1)} x_{(2)}^{-\alpha-1} + (\alpha + 2)x_{(1)}^{-\alpha})\right),$$

où $x_{(1)} = x_1 \wedge x_2$ et $x_{(2)} = x_1 \vee x_2$. La colonne de droite de la Figure 1.1 présente les courbes de niveau de la densité de loi max-stable bivariable et les simulations des processus ponctuels β_n et $\pi_{\alpha, \sigma}$ avec $\alpha = 0.75$ et σ uniforme.

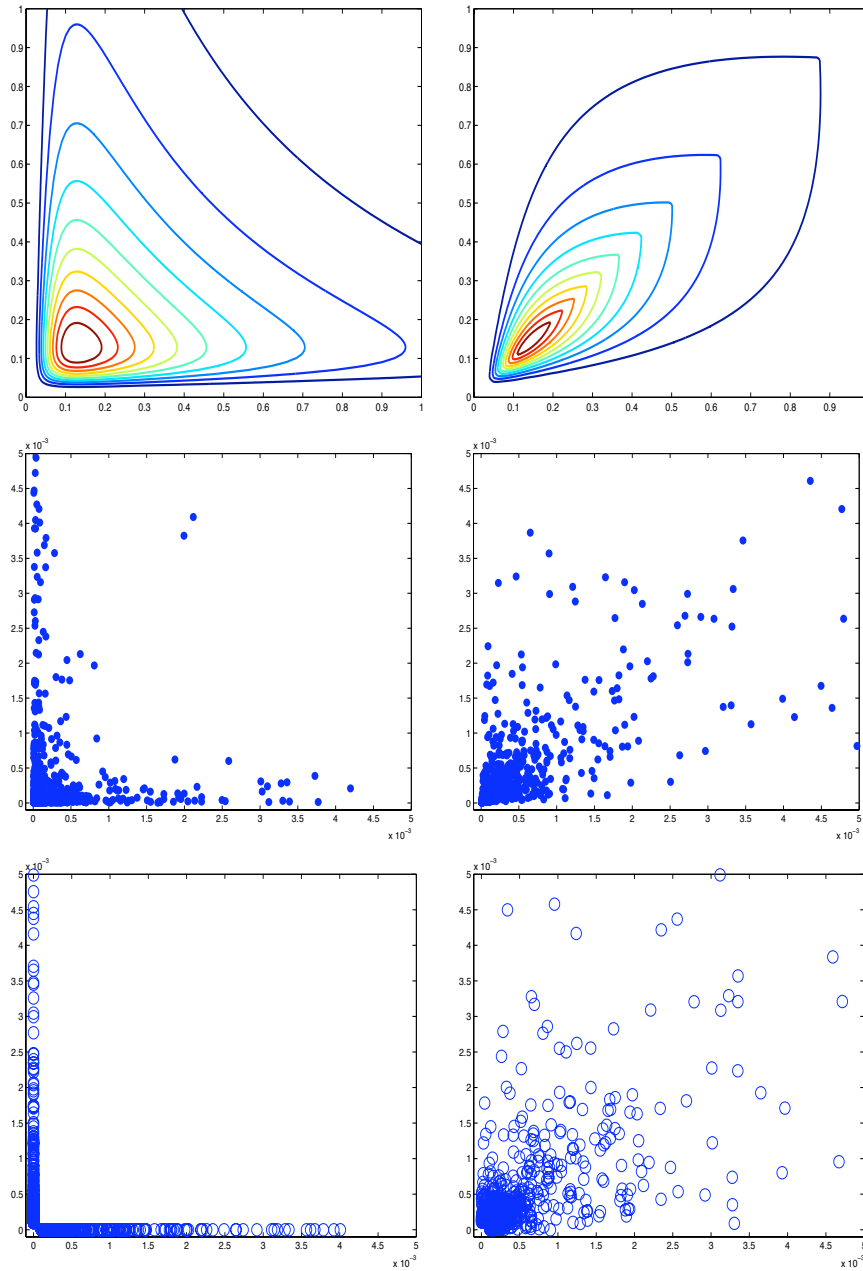


FIG. 1.1: Courbes de niveau des densités max-stables bivariées (en haut) et les simulations du processus ponctuel empirique β_n (au milieu) et du processus ponctuel poissonnien $\pi_{\alpha, \sigma}$ (en bas), $\alpha = 0.75$, la taille d'échantillon $N = 1000$. Colonne de gauche : σ est concentrée sur deux points $(1, 0)$ et $(0, 1)$ de probabilités $p = q = 0.5$. Colonne de droite : σ est uniforme sur la sphère unité S^1_V .

1.2 Transformations des lois à queue régulière

La variation régulière a la propriété importante d'être préservée sous plusieurs opérations et transformations qu'on utilise souvent en pratique. Une large collection de résultats de ce type est présentée dans le survey de Jessen et Mikosch [JM06]. Le but de notre travail dans cette section est de compléter les investigations dans cette direction. On considère ici trois sortes de transformations. Premièrement, on étudie le passage de l'e.a. initial X à l'e.a. $Y = \|X\|f\left(\frac{X}{\|X\|}\right)$ où f est une application de S dans S . Dans le deuxième cas on transforme la partie radiale de X , plus exactement on s'intéresse à la variation régulière de l'e.a. $Y = Xf\left(\frac{X}{\|X\|}\right)$, où la fonction f cette fois-ci est une application de S dans \mathbb{R}_+ . Supposons que X est un v.a. dans \mathbb{R}^d . Nous considérons dans le troisième cas les projections du v.a. X dans l'espace \mathbb{R}^k , $k < d$. On propose des conditions suffisantes et on donne des exemples qui montrent que ces conditions ne pourront être affaiblies sensiblement. En conclusion remarquons que les propriétés des transformations présentées ici seront utiles pour les simulations des vecteurs appartenant au domaine d'attraction d'une loi stable ayant une mesure spectrale donnée. Dans ce contexte on peut mentionner les articles de Chambers et al. [CMS76] et de Modarres et al. [MN94].

1.2.1 Transformations sphériques

Soit X un e.a. dans \mathbb{K}' ayant une queue à variation régulière et $\left(\|X\|, \frac{X}{\|X\|}\right)$ sa décomposition polaire. On va d'abord s'intéresser aux transformations qui ne changent que la partie sphérique de X . Plus exactement on prend une application mesurable $f : S \rightarrow S$ et on définit le nouveau vecteur $Y = \|X\|f\left(\frac{X}{\|X\|}\right)$. Il est clair qu'en coordonnées polaires $Y = \left(\|X\|, f\left(\frac{X}{\|X\|}\right)\right)$. Sous quelles conditions Y reste-t-il encore un vecteur aléatoire à queue régulière? Quelle est la mesure spectrale de ce nouveau vecteur? Le théorème suivant répond à ces questions.

Théorème 1.2.1. *Soit X un e.a. dans \mathbb{K}' et $X \in VR(\alpha, \sigma)$. Soit f une application σ -p.p. continue sur S à valeurs dans S , et μ la mesure image définie par $\mu = \sigma f^{-1}$. Alors l'e.a. transformé $Y = \left(\|X\|, f\left(\frac{X}{\|X\|}\right)\right)$ a la queue à variation régulière de même exposant que X et de mesure spectrale μ , c'est-à-dire $Y \in VR(\alpha, \mu)$.*

Démonstration: Prenons $B \in \mathcal{B}(S)$ tel que $\mu(\partial B) = 0$. Alors $\forall r > 0$,

$$\begin{aligned} \frac{x^\alpha}{L(x)} \mathbf{P} \left\{ \frac{Y}{\|Y\|} \in B, \|Y\| > x \right\} &= \frac{x^\alpha}{L(x)} \mathbf{P} \left\{ f\left(\frac{X}{\|X\|}\right) \in B, \|X\| > x \right\} \\ &= \frac{x^\alpha}{L(x)} \mathbf{P} \left\{ \frac{X}{\|X\|} \in f^{-1}(B), \|X\| > x \right\}. \end{aligned}$$

Pour que le dernier terme converge vers $\mu(B)r^{-\alpha}$, il suffit d'assurer grâce à (1.1.5.) que $\sigma(\partial f^{-1}(B)) = 0$. Notons D l'ensemble des discontinuités de f , alors $\sigma(D) = 0$. On va montrer que

$$\partial f^{-1}(B) \subset f^{-1}(\partial B) \cup D. \quad (1.2.1)$$

Si $x \in \partial f^{-1}(B) \setminus D$ alors f est continue en x . Comme $x \in \partial f^{-1}(B)$, il existe deux suites $x_n \in f^{-1}(B)$ et $y_n \in f^{-1}(B)^c$, $n = 1, 2, \dots$, telles que $x_n \rightarrow x$ et $y_n \rightarrow x$. On en déduit

$$f(x_n) \rightarrow f(x), f(x_n) \in B$$

et

$$f(y_n) \rightarrow f(x), f(y_n) \notin B.$$

Cela implique $f(x) \in \partial B$, ainsi $x \in f^{-1}(\partial B)$ d'où (1.2.1). Puisque $\mu(\partial B) = \sigma f^{-1}(\partial B) = 0$, on en déduit

$$\sigma(\partial f^{-1}(B)) \leq \sigma f^{-1}(\partial B) + \sigma(D) = 0$$

qui complète la démonstration. \square

Corollaire 1.2.2. Soit X un vecteur aléatoire dans \mathbb{R}^2 et $X \in VR(\alpha, \sigma)$. On identifie S^1 avec l'intervalle $[0, 2\pi)$ et on suppose que la mesure spectrale de X est uniforme, c'est-à-dire $d\sigma/d\theta = 1/2\pi$, $\theta \in [0, 2\pi)$. Soit μ une mesure de probabilité sur S^1 avec la fonction de répartition $F(x) = \mu([0, x])$, $x \in [0, 2\pi)$. Si Y est un vecteur aléatoire défini par $Y = \left(\|X\|, F^{-1}\left(\frac{X}{2\pi\|X\|}\right) \right)$ où F^{-1} est la fonction de quantile correspondant à F , alors $Y \in VR(\alpha, \mu)$.

Remarque 2. Ce corollaire montre avec évidence l'utilité des applications du théorème 1.2.1. au problème de simulation.

Remarque 3. La condition de continuité de f σ -p.p. est importante. L'exemple suivant montre que le résultat du théorème 1.2.1. n'est plus vrai si l'on omet cette condition.

Exemple 2. Soient F_1 et F_2 deux lois définies sur \mathbb{R}_+ telles que

- 1) la loi F_i n'a pas de queue à variation régulière, $i = 1, 2$,
- 2) la loi $F = \frac{1}{2}(F_1 + F_2)$ a la queue à variation régulière d'exposant α et de constantes de normalisation b_n .

Soient $\{1/n\}$ et $\{-1/n\}$ deux suites des points sur la sphère unité $S^1 = (-\pi, \pi]$, notées $\{x_n^+\}$ et $\{x_n^-\}$. Soit l_n^\pm la demi-droite sortant de 0 et passant par le point x_n^\pm . Définissons deux suites des segments $\{\Delta_n^+\}$ et $\{\Delta_n^-\}$ par $\Delta_n^\pm = l_n^\pm \cap ((-\pi, \pi] \times [n, n+1))$. Soit \mathbf{P} la loi sur $(-\pi, \pi] \times (\mathbb{R}_+ \setminus \{0\})$ définie de la façon suivante : son support est la réunion de tous les intervalles Δ_i^\pm , $i = 1, 2, \dots$, et la restriction de \mathbf{P} sur $\bigcup_{i=1}^\infty \Delta_i^+$ (respectivement sur $\bigcup_{i=1}^\infty \Delta_i^-$) coïncide avec $1/2F_1$ transférée sur l_n^+ (respectivement avec $1/2F_2$ transférée sur l_n^-). Soit X un v.a. de loi \mathbf{P} . On vérifie facilement que \mathbf{P} est la loi ayant la queue régulière avec b_n comme les constantes de normalisation dont la mesure spectrale est $\sigma = \delta_{\{0\}}$.

Si l'on définit $f : (-\pi, \pi] \rightarrow (-\pi, \pi]$ par

$$f(x) = \begin{cases} \frac{\pi}{2} & x > 0, \\ 0 & x = 0, \\ -\frac{\pi}{2} & x < 0, \end{cases}$$

on obtient que $\sigma f^{-1} = \sigma$ tandis que $Y = (\|X\|, f(\frac{X}{\|X\|}))$ n'a pas de queue régulière. Réellement, par exemple, on a pour $\varepsilon > 0$

$$\begin{aligned} n\mathbf{P} \left\{ \frac{Y}{\|Y\|} \in \left(\frac{\pi}{2} - \varepsilon, \frac{\pi}{2} + \varepsilon \right), \|Y\| > rb_n \right\} &= n\mathbf{P} \left\{ \frac{X}{\|X\|} \in (0, \pi], \|X\| > rb_n \right\} \\ &= n(1 - F_1(rb_n)) \end{aligned}$$

qui ne converge nulle part par le choix de F_1 .

1.2.2 Transformations radiales

On considère maintenant les transformations ne modifiant que la partie radiale de l'élément initial. Etant donné une fonction $h : S \rightarrow \mathbb{R}_+$ on définit le nouvel e.a. $Y = Xh\left(\frac{X}{\|X\|}\right) = \left(\|X\|h\left(\frac{X}{\|X\|}\right), \frac{X}{\|X\|}\right)$. Le résultat suivant donne des conditions sous lesquelles la propriété de régularité de queue est préservée.

Théorème 1.2.3. *Soit X un e.a. dans \mathbb{K}' et $X \in VR(\alpha, \sigma)$. Soit h une fonction σ -p.p. continue et bornée sur S à valeurs dans \mathbb{R}_+ , μ une mesure finie sur S de densité $h(x)^\alpha$ par rapport à σ . Alors l'e.a. $Y = \left(\|X\|h\left(\frac{X}{\|X\|}\right), \frac{X}{\|X\|}\right) \in VR(\alpha, \mu)$.*

La démonstration est reportée dans la section suivante.

Les deux contre-exemples ci-dessous montrent que la condition que h est σ -p.p. continue et bornée est réellement importante pour préserver la régularité. L'exemple 3 présente une fonction h σ -p.p. continue mais non-bornée pour laquelle le résultat du théorème 1.2.3. n'a pas lieu, tandis que la fonction h de l'exemple 4 sera bornée mais non σ -p.p. continue.

Exemple 3. *Définition de X .* Soit τ une mesure discrète sur S^1 définie par

$$\tau(\{b_k\}) = q_k = \frac{1}{k(k+1)}, \quad b_k = \pi - \frac{\pi}{2^{k-1}}, \quad k = 1, 2, \dots$$

Il est clair que $\sum_{k=1}^{\infty} q_k = 1$ et $b_k \in [0, \pi)$. Notons L_k la demi-droite sortant de 0 et passant par le point b_k , i.e. $L_k = \{cb_k, c > 0\}$. Soit Q_k une mesure sur L_k dont la fonction de répartition est définie par

$$F_k(x) = \begin{cases} 0 & 0 < x < 1 \\ 1 - k^{-\nu}x^{-\alpha} & x \geq 1 \end{cases}$$

où $\nu > 0$, $k = 1, 2, \dots$. Supposons que X soit un v.a. dans \mathbb{R}^2 de loi \mathbf{P} définie par :

$$\mathbf{P}(A) = \sum_{k=1}^{\infty} q_k Q_k(A \cap L_k), \quad A \in \mathcal{B}(\mathbb{R}^2).$$

Définissons la mesure σ sur S^1 par

$$\sigma(B) = \sum_{\{k|b_k \in B\}} q_k k^{-\nu}, \quad B \in \mathcal{B}(S^1). \quad (1.2.2)$$

Cette mesure est bien définie car $\sum_{k=1}^{\infty} q_k k^{-\nu} < 1$. Maintenant pour tous $B \in \mathcal{B}(S^1)$ avec $\sigma(\partial B) = 0$ et $\forall r > 1$ on a

$$\begin{aligned}
r^\alpha \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| > r \right\} &= \sum_{\{k|b_k \in B\}} r^\alpha \mathbf{P} \left\{ \frac{X}{\|X\|} = b_k, \|X\| > r \right\} \\
&= r^\alpha \sum_{\{k|b_k \in B\}} q_k (1 - F_k(r)) \\
&= r^\alpha r^{-\alpha} \sum_{\{k|b_k \in B\}} q_k k^{-\nu} \\
&= \sigma(B).
\end{aligned} \tag{1.2.3}$$

Cela signifie que X a la loi à queue régulière d'exposant α et de mesure spectrale σ .

On passe à la construction de notre fonction h . Prenons les intervalles I_k , $k = 1, 2, \dots$ sur $S^1 = [0, 2\pi)$

$$I_1 = \left(\frac{7\pi}{4}, 2\pi \right) \cup \left[0, \frac{\pi}{4} \right), \quad I_k = \left(b_k - \frac{\pi}{2^{k+1}}, b_k + \frac{\pi}{2^{k+1}} \right), \quad k \geq 2.$$

Puisque la distance entre b_k et b_{k+1} est $\frac{\pi}{2^k}$, les intervalles I_1, I_2, \dots sont disjoints et $b_k \in I_k$ pour chaque k . Notre fonction h est définie par

$$h(x) = \sum_{k=1}^{\infty} k^\beta \mathbf{1}_{I_k}(x), \tag{1.2.4}$$

où β est tel que $\frac{1}{\alpha} < \beta < \frac{1+\nu}{\alpha}$. Évidemment h est σ -p.p. continue et non-bornée. Si $Y = Xh\left(\frac{X}{\|X\|}\right)$, alors pour $\forall r > 1$

$$\begin{aligned}
r^\alpha \mathbf{P} \left\{ \frac{Y}{\|Y\|} \in S^1, \|Y\| > r \right\} &= r^\alpha \mathbf{P} \left\{ \|X\| h\left(\frac{X}{\|X\|}\right) > r \right\} \\
&= r^\alpha \sum_{k=1}^{\infty} \mathbf{P} \left\{ \frac{X}{\|X\|} = b_k, \|X\| > \frac{r}{k^\beta} \right\} \\
&\geq r^\alpha \sum_{\{k|\frac{r}{k^\beta} < 1\}} \frac{1}{k(k+1)} \\
&\geq \frac{r^\alpha}{r^{1/\beta} + 1},
\end{aligned}$$

d'où suit la convergence vers l'infini quand $r \rightarrow \infty$ du terme de droite, ce qui n'aurait pas lieu si le théorème 1.2.3. était applicable.

Exemple 4. *Définition de X .* On considère la fonction $g(x)$ sur \mathbb{R}_+

$$g(x) = \sum_{k=0}^{\infty} \frac{1}{2^k} \mathbf{1}_{(k, k+1]}(x).$$

Notons son graphe par D_g :

$$D_g = \{(x, y) | y = g(x), x > 0\}.$$

Soit Q une mesure sur \mathbb{R}_+ dont la fonction de répartition est définie par

$$F_Q(x) = 1 - G(x) = 1 - 1 \wedge x^{-\alpha}.$$

L'application π de \mathbb{R}_+ à \mathbb{R}^2 définie par

$$\pi : r \mapsto (r, g(r))$$

transforme Q en mesure image $Q\pi^{-1}$ qui sera concentrée sur D_g .

Supposons que X est un v.a. dans \mathbb{R}^2 de loi \mathbf{P} suivante :

$$\mathbf{P}(A) = \frac{1}{2}Q\pi^{-1}(A \cap D_g) + \frac{1}{2}Q(A \cap E), \quad A \in \mathcal{B}(\mathbb{R}^2),$$

où $E = \{(x, 0) | x > 0\}$. On vérifie que X satisfait la condition (1.1.6). Prenons d'abord $B = [a, 2\pi)$, $a \in (0, 2\pi)$. Notons $k_a = \min\{k | \frac{1}{2^k} \leq a\}$, alors

$$((k_a, \infty) \times B) \cap D_g = \emptyset.$$

Donc pour tout $r > k_a$ on a

$$r^\alpha \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| > r \right\} = 0. \quad (1.2.5)$$

Ensuite, si $B = [0, a)$, pour $\forall r > k_a$

$$\begin{aligned} r^\alpha \mathbf{P} \{(r, \infty) \times B\} &= r^\alpha \mathbf{P}\{(r, \infty) \times (0, a) \cup ((r, \infty) \times \{0\})\} \\ &= r^\alpha \left(\frac{1}{2}Q\pi^{-1}(((r, \infty) \times (0, a)) \cap D_g) + \frac{1}{2}Q((r, \infty)) \right) \\ &= r^\alpha G(r) \\ &= 1. \end{aligned} \quad (1.2.6)$$

Les relations (1.2.5), (1.2.6) donnent (1.1.3) avec $\sigma = \delta_{\{0\}}$.

Définition de h . On pose $h(x) = \mathbb{1}_{(0, 2\pi)}(x)$. Alors l'ensemble des discontinuités de h sera $\{0\}$ et puisque $\sigma(\partial\{0\}) = \sigma(\{0\}) = 1$, la fonction h n'est pas σ -p.p. continue.

Par les arguments analogues aux précédents on trouve que le vecteur $Y = Xh\left(\frac{X}{\|X\|}\right)$ satisfait la condition de régularité avec la mesure spectrale $\mu = \frac{1}{2}\delta_{\{0\}}$. Par conséquent $d\mu/d\sigma \neq h$.

Si les variables aléatoires $\frac{X}{\|X\|}$ et $\|X\|$ sont indépendantes il y a une condition moins forte sur h telle que la régularité soit préservée sous la transformation.

Théorème 1.2.4. *Soit X un e.a. dans \mathbb{K}' et $X \in VR(\alpha, \sigma)$, h une fonction définie sur S à valeurs dans \mathbb{R}_+ telle que $\int_S h^{\alpha+\varepsilon} d\sigma < \infty$ pour un $\varepsilon > 0$. Si les variables $\frac{X}{\|X\|}$ et $\|X\|$ sont indépendantes, alors l'élément transformé $Y = Xh\left(\frac{X}{\|X\|}\right) = \left(\|X\|h\left(\frac{X}{\|X\|}\right), \frac{X}{\|X\|}\right) \in VR(\alpha, \mu)$ où μ a la densité $d\mu/d\sigma = h^\alpha$.*

Remarque 4. Dans l'espace \mathbb{R}^d en représentant Y sous la forme $Y = \|X\|Z$, où $Z = \frac{X}{\|X\|}h\left(\frac{X}{\|X\|}\right)$, on remarque que le résultat suit directement du Th. 4.15 [RW91]. En connections de ce résultat on peut mentionner [Bre65] (pour $d = 1$) et [DE00], Lemme 3.9.

La démonstration est reportée dans la sous-section 1.2.4.

En randomisant la fonction h , on déduit immédiatement du théorème 1.2.4. le corollaire suivant.

Corollaire 1.2.5. *Supposons que $X \in VR(\alpha, \sigma)$. Soit $\{Z(\theta), \theta \in S\}$ un processus stochastique indépendant de X dont les trajectoires sont presque sûrement positives et σ -p.p. continues. Si pour un $\varepsilon > 0$*

$$\int_{S^{d-1}} \mathbf{E}(Z^{\alpha+\varepsilon}(\theta))\sigma(d\theta) < \infty$$

alors le vecteur $Y = XZ\left(\frac{X}{\|X\|}\right) \in VR(\alpha, \mu)$ où μ a la densité $d\mu/d\sigma = \mathbf{E}(Z(\theta))^\alpha$.

Remarque 5. En vu du théorème 1.2.4. on pourrait penser que la condition suivante et moins restrictive

$$\exists \delta > 0 \text{ tel que } \int_{S^1} h^{\alpha+\delta} d\sigma < \infty$$

sera suffisante pour préserver la régularité de queue. Notre exemple montre que ce n'est pas le cas. Réellement, si δ est suffisamment petit,

$$\int_{S^1} h^{\alpha+\delta}(\theta)\sigma(\theta) = \sum_{k=1}^{\infty} k^{(\alpha+\delta)\beta-\nu} q_k \leq \sum_{k=1}^{\infty} k^{(\alpha+\delta)\beta-\nu-2} < \infty.$$

1.2.3 Projections

Dans cette sous-section on discute deux genres de projections d'un v.a. à queue régulière dans \mathbb{R}^d . Soit $\mathbf{x} = (x_1, \dots, x_d)$ un vecteur dans \mathbb{R}^d . La norme de \mathbf{x} est définie par

$$\|\mathbf{x}\| = \sqrt{x_1^2 + \dots + x_d^2}.$$

Pour un nombre entier k , $1 \leq k < d$, notons

$$\mathbf{x}_k = (x_1, \dots, x_k)$$

et

$$\|\mathbf{x}_k\| = \sqrt{x_1^2 + \dots + x_k^2}.$$

La première projection est définie par

$$\begin{aligned} \Lambda_k : \mathbb{R}^d \setminus \{\mathbf{x} \mid x_1 = \dots = x_k = 0\} &\rightarrow \mathbb{R}^k, \\ \mathbf{x} &\mapsto \frac{\mathbf{x}_k}{\|\mathbf{x}_k\|} \|\mathbf{x}\|. \end{aligned} \quad (1.2.7)$$

En notant

$$\begin{aligned} \Upsilon_k : S^{d-1} \setminus \{\mathbf{x} \mid x_1 = \dots = x_k = 0\} &\rightarrow S^{k-1}, \\ \mathbf{x} &\mapsto \frac{\mathbf{x}_k}{\|\mathbf{x}_k\|}, \end{aligned} \quad (1.2.8)$$

on a la proposition suivante.

Proposition 1.2.1. *Soit X un v.a. à valeurs dans \mathbb{R}^d et $X \in VR(\alpha, \sigma)$. Alors pour un nombre entier k , $1 \leq k < d$, on a $\Lambda_k(X) \in VR(\alpha, \sigma_k)$ où $\sigma_k = \Upsilon_k(\sigma)$.*

Démonstration: Prenons l'ensemble $B \in \mathcal{B}(S^{k-1})$ et $\sigma_k(\partial B) = 0$. Par la continuité de Υ_k , on peut prouver que

$$\partial\Upsilon_k^{-1}(B) \subset \Upsilon_k^{-1}(\partial B). \quad (1.2.9)$$

En effet $\forall \mathbf{x} \in \partial\Upsilon_k^{-1}(B)$, on peut trouver deux suites $\{\mathbf{x}_n\} \in \Upsilon_k^{-1}(B)$ et $\{\mathbf{y}_n\} \notin \Upsilon_k^{-1}(B)$ telles que

$$\mathbf{x}_n \xrightarrow[n \rightarrow \infty]{} \mathbf{x} \quad \text{et} \quad \mathbf{y}_n \xrightarrow[n \rightarrow \infty]{} \mathbf{x}.$$

Puisque Υ_k est continue sur S^{d-1} , on en déduit que

$$\Upsilon_k(\mathbf{x}_n) \xrightarrow[n \rightarrow \infty]{} \Upsilon_k(\mathbf{x}) \quad \text{et} \quad \Upsilon_k(\mathbf{y}_n) \xrightarrow[n \rightarrow \infty]{} \Upsilon_k(\mathbf{x}).$$

Ainsi $\Upsilon_k(\mathbf{x}) \in \partial B$ d'où vient (1.2.9). On déduit de (1.2.9) l'inégalité suivante

$$\sigma(\partial\Upsilon_k^{-1}(B)) \leq \sigma(\Upsilon_k^{-1}(\partial B)) = \sigma_k(\partial B) = 0. \quad (1.2.10)$$

Il est clair que

$$\mathbf{P}\{\Lambda_k(X) \in (x, \infty) \times B\} = \mathbf{P}\{X \in (x, \infty) \times \Upsilon_k^{-1}(B)\}. \quad (1.2.11)$$

Par la régularité de X et (1.2.10) on a

$$\lim_{x \rightarrow \infty} \frac{x^\alpha}{L(x)} \mathbf{P}\{X \in (x, \infty) \times \Upsilon_k^{-1}(B)\} = \sigma(\Upsilon_k^{-1}(B)). \quad (1.2.12)$$

En considérant (1.2.11) et (1.2.12) et la définition de σ_k on obtient

$$\lim_{x \rightarrow \infty} \frac{x^\alpha}{L(x)} \mathbf{P}\{\Lambda_k(X) \in (x, \infty) \times B\} = \sigma_k(B).$$

□

Considérons maintenant le cas où la mesure spectrale σ est uniforme. C'est-à-dire si \mathbf{e} est un v.a. à valeurs dans S^{d-1} de loi σ , alors quelle que soit la rotation vectorielle d -dimensionnelle, notée U_d , on a

$$U_d(\mathbf{e}) \stackrel{\mathcal{L}}{=} \mathbf{e}.$$

D'après la proposition 1.2.1. la loi du v.a. $\Lambda_k(X)$, $1 \leq k < d$ est la loi à queue régulière d'indice α et de mesure spectrale $\sigma_k = \Upsilon_k(\sigma)$. Pour la rotation k -dimensionnelle U_k , définissons la rotation d -dimensionnelle $U_{d,k}$ par

$$U_{d,k}(\mathbf{x}) = (U_k((x_1, \dots, x_k)), x_{k+1}, \dots, x_d). \quad (1.2.13)$$

Ainsi pour tout U_k , on a

$$\begin{aligned}
 U_k(\Upsilon_k(\mathbf{e})) &= U_k\left(\frac{e_1}{\|\mathbf{e}_k\|}, \dots, \frac{e_k}{\|\mathbf{e}_k\|}\right) \\
 &= \Upsilon_k(U_k((e_1, \dots, e_k)), e_{k+1}, \dots, e_d) \\
 &= \Upsilon_k(U_{d,k}(\mathbf{e})) \\
 &\stackrel{\mathcal{L}}{=} \Upsilon_k(\mathbf{e}).
 \end{aligned}$$

Donc $\Upsilon_k(\mathbf{e})$ est un v.a. à valeurs dans S^{k-1} de loi uniforme, c'est-à-dire σ_k est uniforme. On en déduit la proposition suivante.

Proposition 1.2.2. *Soit X un v.a. à valeurs dans \mathbb{R}^d et $X \in VR(\alpha, \sigma)$ où la mesure spectrale σ est uniforme sur S^{d-1} , alors pour un nombre entier $1 \leq k < d$, on a $\Lambda_k(X) \in VR(\alpha, \sigma_k)$ où σ_k est uniforme.*

La deuxième projection est définie par

$$\begin{aligned}
 T_k : \mathbb{R}^d \setminus \{\mathbf{x} \mid x_1 = \dots = x_k = 0\} &\rightarrow \mathbb{R}^k, \\
 \mathbf{x} &\mapsto \mathbf{x}_k.
 \end{aligned} \tag{1.2.14}$$

Proposition 1.2.3. *Soit X un v.a. à valeurs dans \mathbb{R}^d et $X \in VR(\alpha, \sigma)$. Alors pour un nombre entier k , $1 \leq k < d$, la loi du v.a. $T_k(X)$ a la queue à variation régulière d'indice α .*

Démonstration: C'est un résultat direct du lemme 4.6 de [JM06] appliqué à la matrice

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 \end{pmatrix}_{k \times d}.$$

□

Remarquons que cette proposition ne donne pas la forme de la mesure spectrale du v.a. $T_k(X)$. Soit X un v.a. à valeurs dans \mathbb{R}^d et $X \in VR(\alpha, \sigma)$. Considérons le cas où σ est uniforme. Supposons que $T_k(X) \in VR(\alpha, \sigma^*)$ où σ^* dépend de α et σ . Notons U_k la rotation sur l'espace \mathbb{R}^k . Il est clair que

$$U_k(T_k(X)) \in VR(\alpha, U_k(\sigma^*)). \tag{1.2.15}$$

D'ailleurs puisque σ est uniforme, quelle que soit la rotation d -dimensionnelle, notée U_d , on a

$$U_d(X) \in VR(\alpha, \sigma). \tag{1.2.16}$$

Ainsi pour tout U_k , on a

$$\begin{aligned} U_k(T_k(X)) &= U_k(X^{(1)}, \dots, X^{(k)}) \\ &= T_k(U_k((X^{(1)}, \dots, X^{(k)})), X^{(k+1)}, \dots, X^{(d)}) \\ &= T_k(U_{d,k}(X)), \end{aligned} \tag{1.2.17}$$

où $U_{d,k}$ est défini par (1.2.13). D'après (1.2.16) et (1.2.17) on a $U_k(T_k(X)) \in \text{VR}(\alpha, \sigma^*)$. En considérant (1.2.15) et l'unicité de la mesure spectrale, on obtient $U_k(\sigma^*) \stackrel{\mathcal{L}}{=} \sigma^*$, $\forall U_k$, c'est-à-dire σ^* est uniforme.

Proposition 1.2.4. *Soit X un v.a. à valeurs dans \mathbb{R}^d et $X \in \text{VR}(\alpha, \sigma)$ avec la mesure spectrale σ uniforme sur S^{d-1} , alors pour le nombre entier $1 \leq k < d$, on a $T_k(X) \in \text{VR}(\alpha, \sigma^*)$ où σ^* est uniforme.*

1.2.4 Preuves

Démonstration du théorème 1.2.3. Rappelons que Q_n, Q sont les mesures associées avec X et définies par (1.1.7) et (1.1.8). Soit \tilde{Q}_n, \tilde{Q} les mesures associées avec le vecteur $Y = Xh(\frac{X}{\|X\|})$ et définies de même façon, c'est-à-dire

$$\tilde{Q}_n((r, \infty) \times B) = n\mathbf{P} \left\{ \frac{Y}{\|Y\|} \in B, \|Y\| > rb_n \right\}, \quad \tilde{Q} = m_\alpha \times \mu, \tag{1.2.18}$$

où $B \in \mathcal{B}(S)$, $\sigma(\partial B) = 0$, et $r > 0$, b_n est le même que dans (1.1.7). Définissons la famille des fonctions $f : \mathbb{K} \rightarrow \mathbb{R}$

$$\mathbb{C} = \{f \mid f \text{ est continue bornée et } f(x) = 0 \text{ pour tous } x \in (r, \infty) \times S, r > 0\}.$$

Pour démontrer le théorème, il nous suffit d'établir la convergence

$$\int f d\tilde{Q}_n \rightarrow \int f d\tilde{Q}, \quad n \rightarrow \infty, \tag{1.2.19}$$

pour toute $f \in \mathbb{C}$. Définissons l'application φ par

$$\begin{aligned} \varphi : \mathbb{R}_+ \times S &\rightarrow \mathbb{R}_+ \times S, \\ (\rho, \theta) &\mapsto (\rho h(\theta), \theta). \end{aligned}$$

On remarque que $Y = \varphi(X)$, et

$$\begin{aligned} Q_n \varphi^{-1}((r, \infty) \times B) &= n\mathbf{P} \left\{ \frac{Y}{\|Y\|} \in B, \|Y\| > rb_n \right\}, \\ Q \varphi^{-1}((r, \infty) \times B) &= Q\{(\rho, \theta) \mid \theta \in B, \rho h(\theta) \in (r, \infty)\} \\ &= \int_B \int \mathbb{1}_{(r, \infty)}(\rho h(\theta)) m_\alpha(d\rho) \sigma(d\theta) \\ &= \int_B \sigma(d\theta) \alpha \int \mathbb{1}_{(r, \infty)}(\rho h(\theta)) \rho^{-\alpha-1} d\rho \\ &= \mu(B) r^{-\alpha}. \end{aligned}$$

Par conséquent,

$$\int f d\tilde{Q}_n = \int (f \circ \varphi) dQ_n,$$

$$\int f d\tilde{Q} = \int (f \circ \varphi) dQ.$$

La fonction $f \circ \varphi$ est bornée et Q -p.p. continue grâce à Q -p.p. continuité de φ . De plus, $f \circ \varphi \in \mathbb{C}$ puisque h est supposée bornée. En effet si $h \leq M$ et $\text{supp}(f) \subset (r, \infty) \times S$, alors $\text{supp}(f \circ \varphi) \subset (r/M, \infty) \times S$. Car X a la queue régulière d'exposant α et de mesure spectrale σ , grâce à la convergence équivalente (1.1.9), on a pour toute $f \in \mathbb{C}$

$$\int (f \circ \varphi) dQ_n \rightarrow \int (f \circ \varphi) dQ$$

ce qui donne (1.2.19). □

Démonstration du théorème 1.2.4. En représentant Y sous la forme $Y = \|X\|Z$, où $Z = \frac{X}{\|X\|} h(\frac{X}{\|X\|})$, on remarque que le résultat découle directement du Th. 4.15 [RW91]. En connection de ce résultat on peut mentionner [Bre65] (pour $d = 1$) et [DE00], Lemme 3.9. Ici on donne une autre démonstration.

Par la définition de variation régulière nous avons la condition (1.1.3) pour tous $r > 0$ et $B \in \mathcal{B}(S)$ avec $\sigma(\partial B) = 0$. Sans perdre la généralité on suppose que $\sigma(S) = 1$. Puisque $\frac{X}{\|X\|}$ et $\|X\|$ sont indépendantes, en posant $B = S$ on déduit de (1.1.3) que la loi de $\frac{X}{\|X\|}$ est σ et que

$$\lim_{r \rightarrow \infty} \frac{r^\alpha}{L(r)} \mathbf{P}\{\|X\| > r\} = 1. \quad (1.2.20)$$

Nous allons montrer que pour tous $B \in \mathcal{B}(S)$ avec $\mu(\partial B) = 0$

$$\lim_{r \rightarrow \infty} \frac{r^\alpha}{L(r)} \mathbf{P}\left\{\frac{Y}{\|Y\|} \in B, \|Y\| > r\right\} = \int_B h^\alpha d\sigma = \mu(B). \quad (1.2.21)$$

On a besoin d'un résultat bien connu sur les fonctions à variation lente (voir [DE00] lemme 3.11).

Lemme 1.2.6. *Soit L une fonction à variation lente, alors il existe une constante $C > 0$ telle que pour tout $\delta \in (0, 1]$ et une constante T_δ l'inégalité suivante a lieu :*

$$\frac{1}{C} \psi^{-\delta}(x) \leq \frac{L(Tx)}{L(T)} \leq C \psi^\delta(x), \quad T > T_\delta, \quad x \geq \frac{T_\delta}{T},$$

ici $\psi(x) = \max\{x, \frac{1}{x}\}$.

Soit $\delta_1 \in (0, \min\{\delta, \alpha\})$, T_{δ_1} la constante du lemme 1.2.6. correspondant à δ_1 . Alors

$$\begin{aligned} p^{(1)}(r) &= \frac{r^\alpha}{L(r)} \mathbf{P} \left\{ \frac{Y}{\|Y\|} \in B, \|Y\| > r, h\left(\frac{X}{\|X\|}\right) > \frac{r}{T_{\delta_1}} \right\} \\ &= \frac{r^\alpha}{L(r)} \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| h\left(\frac{X}{\|X\|}\right) > r, h\left(\frac{X}{\|X\|}\right) > \frac{r}{T_{\delta_1}} \right\} \\ &\leq \frac{r^\alpha}{L(r)} \frac{\mathbf{E} |h(\frac{X}{\|X\|})|^{\alpha+\delta}}{r^{\alpha+\delta}} T_{\delta_1}^{\alpha+\delta} \\ &= \frac{1}{L(r)r^\delta} \int_{S^{d-1}} h(\theta)^{\alpha+\delta} \sigma(d\theta) T_{\delta_1}^{\alpha+\delta}. \end{aligned}$$

Par conséquent,

$$\lim_{r \rightarrow \infty} p^{(1)}(r) = 0. \quad (1.2.22)$$

Maintenant on a

$$\begin{aligned} p^{(2)}(r) &= \frac{r^\alpha}{L(r)} \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| h\left(\frac{X}{\|X\|}\right) > r, h\left(\frac{X}{\|X\|}\right) \leq \frac{r}{T_{\delta_1}} \right\} \\ &= \frac{r^\alpha}{L(r)} \int_B \mathbb{1}_{(0, \frac{r}{T_{\delta_1}})}(h(\theta)) \mathbf{P} \left\{ \|X\| > \frac{r}{h(\theta)} \right\} \sigma(d\theta) \\ &= \int_B \frac{L(\frac{r}{h(\theta)})}{L(r)} \mathbb{1}_{(0, \frac{r}{T_{\delta_1}})}(h(\theta)) h(\theta)^\alpha \Psi(r, \theta) \sigma(d\theta), \end{aligned}$$

où $\Psi(r, \theta) = (\frac{r}{h(\theta)})^\alpha (L(\frac{r}{h(\theta)}))^{-1} \mathbf{P}\{\|X\| > \frac{r}{h(\theta)}\}$. Grâce à (1.2.20) pour tout θ , $\Psi(r, \theta) \rightarrow 1$ quand $r \rightarrow \infty$. De plus cette fonction est bornée. Le lemme 1.2.6. nous montre que

$$\left| \frac{L(\frac{r}{h(\theta)})}{L(r)} \right| \leq C \max\{h(\theta)^{\delta_1}, h(\theta)^{-\delta_1}\}$$

si $r > T_{\delta_1}$ et $h(\theta) \leq \frac{r}{T_{\delta_1}}$. On a donc le droit d'appliquer le théorème de convergence dominée qui donne

$$\lim_{r \rightarrow \infty} p^{(2)}(r) = \int_B h(\theta)^\alpha \sigma(d\theta) = \mu(B).$$

□

Dans le théorème 1.2.3. la famille des fonctions définies sur S à valeurs dans \mathbb{R}_+ , σ -p.p continues est considérée. On va présenter une famille des fonctions équivalente. Cette équivalence nous donne une autre démonstration du théorème 1.2.3.. De manière plus générale, supposons que S est un espace métrique, \mathcal{S} est la σ -algèbre engendrée par les ensembles ouverts de S . Soit σ une mesure finie sur (S, \mathcal{S}) . Notons \mathfrak{L} la famille des fonctions définies sur S positives et σ -p.p. continues, i.e.

$$\mathfrak{L} = \{f \mid f \text{ est positive et } \sigma\text{-p.p. continue}\}.$$

Définition 1.2.1. La fonction f est σ -approximable si pour tout $\varepsilon > 0$ il existe deux fonctions en escaliers f_ε^- et f_ε^+ définies par $f_\varepsilon^\pm = \sum_{i=1}^{\infty} r_i^\pm \mathbb{1}_{I_i}$ où $r_i^\pm \geq 0$ et $\{I_i\}, i = 1, \dots$ est une partition de S telle que

- 1) $\sigma(\partial I_i) = 0, i = 1, \dots;$
- 2) $0 \leq f_\varepsilon^-(x) \leq f(x) \leq f_\varepsilon^+(x), \forall x \in S;$
- 3) $f_\varepsilon^+(x) - f_\varepsilon^-(x) < \varepsilon, \forall x \in S.$

Proposition 1.2.5. *Soit \mathfrak{U} la famille des fonctions positives et σ -approximables, i.e. $\mathfrak{U} = \{f \mid f \text{ est positive et } \sigma\text{-approximable}\}$, alors $\mathfrak{L} = \mathfrak{U}$.*

Remarque 6. Le théorème 1.2.3. est une conséquence directe de cette proposition. Supposons que les conditions du théorème 1.2.3. sont remplies, alors pour tout $\varepsilon > 0$ il existe deux fonctions en escaliers h_ε^- et h_ε^+ qui satisfont les conditions 1), 2) et 3) de la proposition 1.2.5. telles que

$$\begin{aligned} \int_B h_\varepsilon^-(\theta)^\alpha \sigma(d\theta) r^{-\alpha} &= \sum_{i=1}^{\infty} \sigma(B \cap I_i) (r_i^-)^\alpha r^{-\alpha} \leq \tilde{Q}_n((r, \infty) \times B) \\ &\leq \sum_{i=1}^{\infty} \sigma(B \cap I_i) (r_i^+)^\alpha r^{-\alpha} \\ &= \int_B h_\varepsilon^+(\theta)^\alpha \sigma(d\theta) r^{-\alpha}, \end{aligned}$$

où $B \in \mathcal{B}(S)$, $\sigma(\partial B) = 0$ et $r > 0$, $\tilde{Q}_n((r, \infty) \times B)$ est défini par (1.2.18).

Démonstration: On montre d'abord $\mathfrak{L} \subset \mathfrak{U}$. Les points a dans \mathbb{R}_+ tels que $\sigma(f^{-1}(a)) \neq 0$ sont au maximum dénombrables. Parce qu'avec les valeurs différentes de a on a des points disjoints $f^{-1}(a)$ dans S , comme σ est finie sur S ces points qui ont la mesure positive sont au maximum dénombrables. Notons $D = \{a \mid a \in \mathbb{R}_+, \sigma(f^{-1}(a)) \neq 0\}$.

On construit maintenant les fonctions en escaliers. Pour tout $\varepsilon > 0$ on définit une partition de demi-droite \mathbb{R}_+ par deux étapes : premièrement on prend une suite de constantes croissantes $\{a_0, a_1, \dots\}$ comme suit

$$a_0 = 0, a_i = a_{i-1} + \varepsilon/2, \quad i = 1, \dots$$

Deuxièmement si il y a un a_i défini comme avant qui appartient à D , on remplace a_i par a'_i tel que $a'_i \in (a_i - \varepsilon/4, a_i + \varepsilon/4) \setminus D$, note encore par a_i . Après cette procédure la distance entre a_i et a_{i+1} est toujours inférieure à ε . De plus $\sigma(f^{-1}(a_i)) = 0, i = 0, \dots$. Notons $I_{\varepsilon, i} = f^{-1}([a_{i-1}, a_i]), i = 1, \dots$. Construisons deux fonctions en escaliers suivantes :

$$f_\varepsilon^- = \sum_{i=1}^{\infty} a_{i-1} \mathbb{1}_{I_{\varepsilon, i}}, \quad f_\varepsilon^+ = \sum_{i=1}^{\infty} a_i \mathbb{1}_{I_{\varepsilon, i}}.$$

Notons T l'ensemble des discontinuités de f , alors $\sigma(T) = 0$. On va montrer que

$$\partial f^{-1}([a_{i-1}, a_i]) \subset f^{-1}(\partial[a_{i-1}, a_i]) \cup T, \quad i = 1, \dots$$

Si $x \in \partial f^{-1}([a_{i-1}, a_i]) \setminus T$ alors f est continue en x . Comme $x \in \partial f^{-1}([a_{i-1}, a_i])$, il existe deux suites $x_n \in f^{-1}([a_{i-1}, a_i])$ et $y_n \in f^{-1}([a_{i-1}, a_i])^c, n = 1, 2, \dots$, telles que $x_n \rightarrow x$ et $y_n \rightarrow x$. On en déduit

$$f(x_n) \rightarrow f(x), \quad f(x_n) \in [a_{i-1}, a_i]$$

et

$$f(y_n) \rightarrow f(x), \quad f(y_n) \notin [a_{i-1}, a_i].$$

Cela implique $f(x) \in \partial[a_{i-1}, a_i]$, ainsi $x \in f^{-1}(\partial[a_{i-1}, a_i])$. Par conséquent,

$$\sigma(\partial I_i) = \sigma(\partial f^{-1}([a_{i-1}, a_i])) \leq \sigma(f^{-1}(a_{i-1})) + \sigma(f^{-1}(a_i)) + \sigma(T) = 0.$$

Il est clair que $0 \leq f_\varepsilon^-(x) \leq f(x) \leq f_\varepsilon^+(x)$, pour tout $x \in S$. De plus

$$f_\varepsilon^+(x) - f_\varepsilon^-(x) = \sum_{i=1}^{\infty} (a_i - a_{i-1}) \mathbb{I}_{I_{\varepsilon,i}}(x) < \varepsilon, \quad \forall x \in S.$$

Maintenant on vérifie que $\mathfrak{U} \subset \mathfrak{L}$. Si $f \in \mathfrak{U}$, alors il existe une suite des fonctions en escaliers f_n définie par $f_n = \sum_{i=1}^{\infty} r_{n,i} \mathbb{I}_{I_{n,i}}$ telle que

$$\sup_{x \in S} |f(x) - f_n(x)| \xrightarrow{n \rightarrow \infty} 0,$$

et

$$\sigma\left(\bigcup_{n,i=1}^{\infty} \partial I_{n,i}\right) = 0.$$

Prenons un point $x_0 \notin \bigcup_{n,i=1}^{\infty} \partial I_{n,i}$. Pour $\forall \varepsilon > 0$ il existe n_0 tel que

$$|f(x) - f_{n_0}(x)| < \varepsilon, \quad \forall x \in S.$$

Comme $x_0 \notin \bigcup_{i=1}^{\infty} \partial I_{n_0,i}$ et $x_0 \in \bigcup_{i=1}^{\infty} I_{n_0,i} = S$, il existe i_0 unique tel que $x_0 \in I_{n_0,i_0}^\circ$. Donc il existe un voisinage de x_0 noté V_{x_0} tel que $V_{x_0} \subset I_{n_0,i_0}$. Alors pour $\forall x \in V_{x_0}$, on a

$$|f(x_0) - f(x)| \leq |f(x_0) - f_{n_0}(x_0)| + |f_{n_0}(x_0) - f_{n_0}(x)| + |f_{n_0}(x) - f(x)| < 2\varepsilon.$$

Par conséquent, f est continue sur x_0 . □

1.3 Simulation des vecteurs aléatoires à queue régulière

Dans cette section nous présentons quelques exemples de lois à queue régulière et des résultats de simulation. Les lois α -stables et les lois max-stables sont considérées. Ces deux familles de lois apparaissent souvent dans les domaines d'application des événements rares et en particulier des valeurs extrêmes. Pour les lois stables, des spécifications différentes ont amené des paramétrisations différentes dans les littératures. Nous commençons dans la première sous-section par préciser la paramétrisation et rappeler des propriétés principales de lois stables. Ensuite nous simulons des v.a. de loi stable en utilisant la représentation de LePage. En utilisant la propriété de préservation de régularité par des transformations, nous pouvons aussi simuler des v.a. appartenant au domaine d'attraction d'une loi stable. Pour les lois max-stables, nous choisissons quelques exemples particuliers tels que l'expression de la fonction de répartition peut être déduite.

1.3.1 Lois α -stables

Paramétrisation et propriétés de base

Il existe plusieurs définitions équivalentes de lois stables ainsi que plusieurs paramétrisations. Nous présentons ici trois définitions équivalentes.

Définition 1.3.1. *Un v.a. X dans \mathbb{R}^d a une loi stable si pour tout $a, b \in \mathbb{R}_+$ il existe $c \in \mathbb{R}_+$ et un vecteur $D \in \mathbb{R}^d$ tels que*

$$aX_1 + bX_2 \stackrel{\mathcal{L}}{=} cX + D, \quad (1.3.1)$$

où X_1 et X_2 sont des copies indépendantes de X .

Le nombre c dans (1.3.1) vérifie $c^\alpha = a^\alpha + b^\alpha$ pour un nombre $0 < \alpha \leq 2$ (voir [Fel71] section VI.1 pour la démonstration). La loi de X est dite α -stable. Si $D = 0$ on dit que X est *strictement α -stable*, et la condition (1.3.1) devient (2) dans l'introduction. Remarquons que si X est de loi gaussienne $\mathcal{N}(\mu, \sigma^2)$, alors les termes de gauche dans (1.3.1) sont respectivement $\mathcal{N}(a\mu, (a\sigma)^2)$ et $\mathcal{N}(b\mu, (b\sigma)^2)$, et le terme de droite est $\mathcal{N}(c\mu + D, (c\sigma)^2)$. Si on prend $c^2 = a^2 + b^2$ et $D = (a + b - c)\mu$, l'égalité (1.3.1) est établie, c'est-à-dire que les lois normales sont 2-stables.

La difficulté technique dans l'étude de lois α -stables est que, sauf dans les cas particuliers ($\alpha = 0.5, 1$ et 2), il n'y a pas de forme explicite pour la densité. Les seules informations utilisables pour les v.a. α -stables sont leur fonction caractéristique.

Définition 1.3.2. *La fonction caractéristique d'un v.a. α -stable ($0 < \alpha < 2$) dans \mathbb{R}^d s'exprime par l'expression suivante :*

$$\phi_{\alpha, \sigma}(t) = \exp \left(-\frac{1}{C_\alpha} \int_{S^{d-1}} \psi_\alpha(\langle t, s \rangle) \sigma(ds) + i\langle \delta, t \rangle \right), \quad (1.3.2)$$

où $S^{d-1} = \{x \in \mathbb{R}^d \mid \|x\| = 1\}$, σ est une mesure finie sur la sphère unité S^{d-1} , δ est un vecteur dans \mathbb{R}^d , $\langle t, s \rangle$ représente le produit scalaire,

$$C_\alpha = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha) \cos(\pi\alpha/2)}, & \alpha \neq 1, \\ \frac{2}{\pi}, & \alpha = 1, \end{cases} \quad (1.3.3)$$

et

$$\psi_\alpha(x) = \begin{cases} |x|^\alpha \left(1 - i \operatorname{sign}(x) \tan \frac{\pi\alpha}{2} \right), & \alpha \neq 1, \\ |x| \left(1 + i \frac{\pi}{2} \operatorname{sign}(x) \ln |x| \right), & \alpha = 1. \end{cases}$$

Cette définition montre que la loi stable dans \mathbb{R}^d est spécifiée par un nombre α entre 0 et 2, *indice de stabilité*, une mesure finie σ sur S^{d-1} dite *mesure spectrale* et un vecteur δ dans \mathbb{R}^d . L'expression (1.3.2) est déduite de la fonction caractéristique d'une loi infiniment divisible (voir par exemple XVII [Fel71] et [AG80]). Pourtant la densité de la loi stable existe, on peut obtenir une formule intégrale par la transformée de Fourier inverse. Il existe des méthodes numériques pour obtenir la densité et la fonction de répartition approximatives d'une loi stable [Nol97].

Dans le cas unidimensionnel la sphère unité ne contient que deux points, i.e. $S^0 = \{-1, 1\}$. La mesure spectrale σ se réduit à deux valeurs $\sigma(-1)$ et $\sigma(1)$. La fonction caractéristique peut être écrite comme suit

$$\phi_{\alpha,\sigma}(t) = \begin{cases} \exp \left(-\frac{\gamma}{C_\alpha} |t|^\alpha \left(1 - i\beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2} \right) + i\delta t \right), & \alpha \neq 1, \\ \exp \left(-\frac{\gamma}{C_\alpha} |t| \left(1 + i\beta \frac{\pi}{2} \operatorname{sign}(t) \ln |t| \right) + i\delta t \right), & \alpha = 1, \end{cases}$$

où $\gamma = \sigma(1) + \sigma(-1)$ et $\beta = (\sigma(1) - \sigma(-1))/\gamma$. La loi stable unidimensionnelle est déterminée par quatre paramètres : $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\gamma > 0$ et $\delta \in \mathbb{R}^1$. Le paramètre β contrôle l'asymétrie. La loi est dite *biaisée totalement vers la droite* si $\beta = 1$ et *biaisée totalement vers la gauche* si $\beta = -1$. Si $\beta = 0$ la densité est symétrique par rapport à δ . Les paramètres γ et δ sont les paramètres d'échelle et de position.

Les paramétrisations qu'on rencontre le plus souvent en pratique sont celles de Samorodnitsky et Taqqu [ST94] notée (S) et de Zolotarev [Zol86] notée (Z). La différence entre notre paramétrisation et la paramétrisation (S) concerne la masse totale de la mesure spectrale σ . Notons σ_s le paramètre d'échelle et la mesure spectrale dans la paramétrisation (S). Dans le cas unidimensionnel

$$\gamma = C_\alpha \sigma_s^\alpha,$$

et dans le cas multidimensionnel

$$\sigma(\cdot) = C_\alpha \sigma_s(\cdot),$$

où C_α est définie par (1.3.3).

Nous utilisons la notation $\mathcal{S}_d(\alpha, \sigma, \delta)$ pour la loi α -stable d -dimensionnelle. Si $d \geq 2$, σ représente une mesure finie sur S^{d-1} . Si $d = 1$, la mesure spectrale σ est remplacée par un couple (β, γ) , c'est-à-dire on utilise la notation $\mathcal{S}_1(\alpha, (\beta, \gamma), \delta)$. Le fait que X a la loi $\mathcal{S}_d(\alpha, \sigma, \delta)$ est noté par l'écriture " $X \sim \mathcal{S}_d(\alpha, \sigma, \delta)$ ". La Figure 1.2 présente l'influence des paramètres sur la forme de la densité d'une loi stable.

Nous présentons maintenant quelques propriétés des lois α -stables.

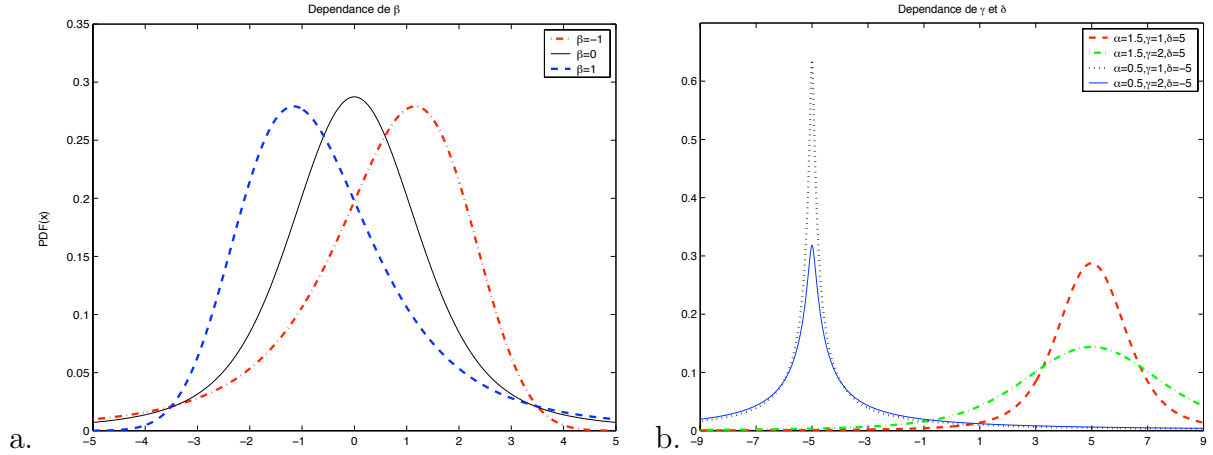


FIG. 1.2: Influence des paramètres sur la forme de la loi stable. Les densités de la loi $\mathcal{S}_1(1.5, (\beta, 1), 0)$ $\beta = -1, 0, 1$ (a) et de la loi $\mathcal{S}_1(\alpha, (0, \gamma), \delta)$ (b).

Propriété 1. Si $X \sim \mathcal{S}_d(\alpha, \sigma, \delta)$ avec $0 < \alpha < 2$ alors

$$\begin{aligned} \mathbf{E}\|X\|^p &< \infty \quad \text{pour tout } 0 < p < \alpha, \\ \mathbf{E}\|X\|^p &= \infty \quad \text{pour tout } p \geq \alpha. \end{aligned}$$

Propriété 2. Si $1 < \alpha \leq 2$, le paramètre de position δ est égal à l'espérance.

Propriété 3. Le v.a. X est $\mathcal{St}\alpha\mathcal{S}$ dans \mathbb{R}^d avec $0 < \alpha \leq 2$ si et seulement si

- (a) $\alpha \neq 1, \delta = 0$,
- (b) $\alpha = 1, \int_{S^{d-1}} s^{(k)} \sigma(ds) = 0$, pour $k = 1, 2, \dots, d$.

Propriété 4. Le v.a. α -stable est *symétrique* si et seulement si $\delta = 0$ et σ est une mesure symétrique sur S^{d-1} , i.e. $\sigma(A) = \sigma(-A)$ pour tout ensemble borélien A dans S^{d-1} .

La troisième définition montre que les lois α -stables sont la seule limite possible non-triviale des sommes normalisées des v.a. i.i.d..

Définition 1.3.3. Un v.a. X dans \mathbb{R}^d a une loi stable s'il possède un domaine d'attraction, i.e. s'il existe une suite de v.a. i.i.d. Y_1, Y_2, \dots dans \mathbb{R}^d , une suite de nombres positifs $\{b_n\}$ et une suite de vecteurs réels $\{a_n\}$ telles que

$$\frac{Y_1 + Y_2 + \dots + Y_n}{b_n} + a_n \Rightarrow X, \quad (1.3.4)$$

où \Rightarrow représente la convergence faible.

La loi de Y_i est dite *appartenant au domaine d'attraction* d'une loi α -stable de mesure spectrale σ , notée par " $Y_i \in \text{Dom}(\alpha, \sigma)$ ". Si $b_n = n^{1/\alpha}$ la loi de Y_i est dite *appartenant au domaine d'attraction normal* d'une loi α -stable. Un v.a. stable appartient à son propre domaine d'attraction normal. Si X est une variable aléatoire gaussienne et si les Y_i sont

des variables aléatoires i.i.d. de variance finie, alors (1.3.4) est le théorème central limite ordinaire. En général, $b_n = n^{1/\alpha}L(n)$ où $L(x)$ est une fonction à variation lente. La condition de variation régulière de la queue (1.1.6) est nécessaire pour que $X \in \text{Dom}(\alpha, \sigma)$ avec $0 < \alpha < 2$.

Théorème 1.3.1. (Cor. 6.20 [AG80]) Soit X un v.a. dans l'espace \mathbb{R}^d et $X \in \text{Dom}(\alpha, \sigma)$, $0 < \alpha < 2$, avec les nombres de normalisation b_n . Alors pour tout $r > 0$ et $B \in \mathcal{B}(S^{d-1})$ avec $\sigma(\partial B) = 0$ on a (1.1.6).

Remarque 7. La condition (1.1.6) est aussi suffisante pour que X appartienne au domaine d'attraction d'une loi strictement stable (Th. 7.11 dans [AG80] page 167).

Remarque 8. Ce théorème montre que dans \mathbb{R}^d la famille de lois appartenant au domaine d'attraction d'une loi α -stable avec $\alpha < 2$ est un sous-ensemble des lois à queue régulière.

Définissons pour $(r, s) \in \mathbb{R}_+ \times S^{d-1} = \mathbb{R}^d \setminus \{0\}$ la mesure μ par

$$\mu\{0\} = 0, \quad d\mu(r, s) = d\sigma(s)dr/r^{1+\alpha}.$$

La mesure μ est la mesure de Lévy associée à la loi stable. Si X est un v.a. α -stable non-gaussien, alors $X \in \text{Dom}(\alpha, \sigma)$ avec $b_n = n^{1/\alpha}$. On en déduit la condition de régularité

$$\lim_{r \rightarrow \infty} r^\alpha \mathbf{P} \left\{ \frac{X}{\|X\|} \in B, \|X\| > r \right\} = \sigma(B), \quad (1.3.5)$$

pour tout $r > 0$ et $B \in \mathcal{B}(S)$ avec $\sigma(\partial B) = 0$.

Il y a plusieurs paramétrisations pour les lois stables. Beaucoup de confusions ont été provoquées par ces différentes paramétrisations. La variété des paramétrisations vient de l'évolution historique, ainsi que des nombreux problèmes qui ont été analysés à l'aide des formes spécifiques des lois stables. Il y a de bonnes raisons pour utiliser les différentes paramétrisations en fonction des situations : le travail numérique, l'étude des propriétés algébriques ou l'étude des propriétés analytiques, etc.. L'avantage de notre paramétrisation est que la mesure spectrale d'une loi stable définie par sa fonction caractéristique (1.3.2) coïncide avec celle qui apparaît dans la propriété de régularité (1.1.6). Dans le cas unidimensionnel la condition (1.3.5) devient :

$$\begin{cases} \lim_{x \rightarrow \infty} x^\alpha \mathbf{P}\{X > x\} & = \sigma(1), \\ \lim_{x \rightarrow \infty} x^\alpha \mathbf{P}\{X < -x\} & = \sigma(-1). \end{cases} \quad (1.3.6)$$

Les relations des paramètres sont

$$\sigma(1) = \frac{(1 + \beta)}{2}\gamma, \quad \sigma(-1) = \frac{(1 - \beta)}{2}\gamma.$$

Les preuves de (1.3.5) et (1.3.6) peuvent être trouvées dans [ST94], page 16 et 197. La Figure 1.3 présente le log-log graphe de la queue de lois stables.

Les v.a. sous-gaussiens stables jouent un rôle important dans la famille des lois stables. La proposition suivante donne la définition et des propriétés de la loi sous-gaussienne stable.

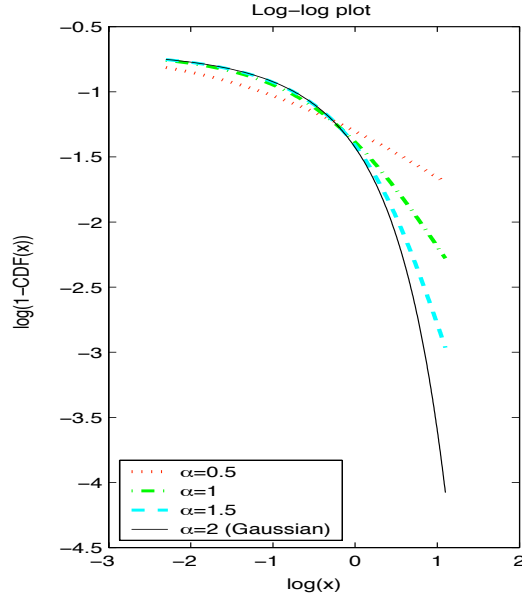


FIG. 1.3: Log-log graphe de $\mathbf{P}\{X > x\}$ où $X \sim \mathcal{S}_1(\alpha, (0, 1), 0)$, $\alpha = 0.5, 1, 1.5$ et 2 .

Proposition 1.3.1. ([ST94] Prop. 2.5.2) Soit X un v.a. dans \mathbb{R}^d , $G = (G_1, \dots, G_d)$ un vecteur gaussien de loi $\mathcal{N}(0, R)$ où $R = (R_{i,j})_{d \times d} = (\mathbf{E}G_i G_j)_{d \times d}$ est la matrice de covariances de G , alors les assertions suivantes sont équivalentes :

- (a) X est sous-gaussien α -stable de vecteur gaussien sous-jacent G , i.e. $X \stackrel{\mathcal{L}}{=} A^{1/2}G$ où $A \sim \mathcal{S}_1(\alpha/2, (1, C_\alpha(\cos \frac{\pi\alpha}{4})^2), 0)$, A et G sont indépendants.
- (b) La fonction caractéristique de X est

$$\mathbf{E} \exp(i\langle u, X \rangle) = \exp \left(- \left| \frac{1}{2} u R u^T \right|^{\alpha/2} \right).$$

- (c) X est de contour elliptique autour de l'origine.

Le caractère principal est que toutes les lois sous-gaussiennes stables d -dimensionnelles sont caractérisées par l'indice α et une matrice $d \times d$ symétrique définie positive. Il y a $1 + d(d+1)/2$ paramètres au total. Le problème d'estimation est donc essentiellement différent du cas général. Au lieu d'estimer une mesure finie sur la sphère S^{d-1} , il suffit d'estimer une matrice $d \times d$. Par ailleurs le problème de simulation se réduit à simuler une variable aléatoire stable unidimensionnelle et un vecteur aléatoire gaussien. Chambers et al. [CMS76] ont proposé une méthode pour simuler les variables aléatoires stables unidimensionnelles générales. En utilisant cette méthode pour engendrer la variable stable positive A et la méthode standard pour engendrer le vecteur gaussien G , tous les vecteurs sous-gaussiens stables peuvent être simulés. Les v.a. appartenant au domaine d'attraction d'une loi α -stable de mesure spectrale donnée peuvent être simulés par transformation d'une loi α -stable de mesure spectrale uniforme qui caractérise un cas particulier de la famille des lois sous-gaussiennes stables.

Proposition 1.3.2. ([ST94] Prop. 2.5.5) Soit X un v.a. α -stable symétrique dans \mathbb{R}^d , alors les assertions suivantes sont équivalentes :

- (a) X est sous-gaussien α -stable avec le vecteur gaussien sous-jacent G ayant les composants i.i.d. $\mathcal{N}(0, \sigma^2)$, i.e. $G = (G_1, \dots, G_d)$, $G_i \sim \mathcal{N}(0, \sigma^2)$, les variables aléatoires G_i sont indépendantes, $i = 1, \dots, d$.
- (b) La fonction caractéristique de X est

$$\mathbf{E} \exp(i\langle u, X \rangle) = \exp(-2^{-\alpha/2} \sigma^\alpha \|u\|^\alpha),$$

i.e. elle ne dépend que de la norme du vecteur $u = (u_1, \dots, u_d)$.

- (c) La mesure spectrale de X est uniforme.

La loi StaS de mesure spectrale uniforme s'appelle α -stable isotropique.

Proposition 1.3.3. Soit X un v.a. de loi α -stable isotropique, alors les variables aléatoires $\|X\|$ et $\frac{X}{\|X\|}$ sont indépendantes.

Démonstration: D'après la proposition 1.3.2. on a

$$X \stackrel{\mathcal{L}}{=} A^{1/2} G = (A^{1/2} G_1, \dots, A^{1/2} G_d),$$

où A est une variable aléatoire $\alpha/2$ -stable totalement biaisée vers droite et G_i sont i.i.d. de loi $\mathcal{N}(0, \sigma^2)$. Ainsi

$$\|X\| \stackrel{\mathcal{L}}{=} A^{1/2} \|G\| \quad \text{et} \quad \frac{X}{\|X\|} \stackrel{\mathcal{L}}{=} \left(\frac{G_1}{\|G_1\|}, \dots, \frac{G_d}{\|G_d\|} \right).$$

Il suffit de montrer que $\|G\|$ et $\frac{G}{\|G\|}$ sont indépendants. La densité de G est

$$f(x) = \frac{1}{(2\pi)^{d/2} \sigma^d} \exp\left(-\frac{1}{2} \left(\frac{\|x\|}{\sigma}\right)^2\right).$$

C'est-à-dire en coordonnées polaires la densité de G sur le point x ne dépend que de la norme $\|x\|$, donc elle est uniforme par rapport aux coordonnées angulaires, les variables aléatoires $\|G\|$ et $\frac{G}{\|G\|}$ sont indépendantes. \square

Données α -stables simulées

Il existe des méthodes connues pour simuler les v.a. stables. Chambers et al. [CMS76] ont proposé une méthode pour simuler les variables aléatoires stables unidimensionnelles arbitraires. Plus récemment une technique précise a été construite dans [Wer96] en utilisant une transformation non-linéaire d'une paire des variables aléatoires indépendantes de loi uniforme et de loi exponentielle. Une méthode pour simuler les vecteurs aléatoires stables basée sur la discrétisation de la mesure spectrale a été présentée par Modarres et Nolan [MN94]. Ici on simule les données stables en utilisant la représentation de série de LePage d'une loi stable.

Soient $\{\lambda_k, k \geq 1\}$ et $\{\epsilon_k, k \geq 1\}$ deux suites indépendantes de variables aléatoires i.i.d.. Les variables aléatoires λ_k sont de loi exponentielle standard, les v.a. ϵ_k sont de loi $\tilde{\sigma}(\cdot)$ qui est une mesure finie normalisée sur S^{d-1} . Notons $\Gamma_k = \lambda_1 + \dots + \lambda_k$, $k \geq 1$, c une constante positive. D'après le Théorème 1.1.4., si la série

$$\xi_{\alpha, \sigma} \stackrel{\mathcal{L}}{=} c \sum_{k=1}^{\infty} \Gamma_k^{-1/\alpha} \epsilon_k \quad (1.3.7)$$

converge p.s. alors $\xi_{\alpha, \sigma}$ a la loi \mathcal{StaS} sur \mathbb{R}^d de mesure spectrale $\sigma(\cdot) = c^\alpha \tilde{\sigma}(\cdot)$. Remarquons que si $\alpha < 1$ ou si ϵ_k a une loi symétrique, alors la série (1.3.7) converge p.s. d'après le théorème des trois séries appliqué conditionnellement étant donnée la suite $\{\Gamma_k\}$. Ce résultat peut être trouvé dans [LWZ81], théorème 3.

Théorème 1.3.2. ([LWZ81], Th. 3) *La série définie par (1.3.7) avec $\alpha < 1$ ou σ symétrique a la loi \mathcal{StaS} .*

Donc on peut simuler un v.a. \mathcal{StaS} noté ξ , satisfaisant la condition du théorème précédent en utilisant la somme partielle de la série de LePage, i.e.

$$\sum_{i=1}^k c \Gamma_i^{-1/\alpha} \epsilon_i \quad (1.3.8)$$

où Γ_i , ϵ_i et c sont définis comme avant. Cette somme donne une bonne approximation si la loi est symétrique. Dans le cas non-symétrique, l'erreur est grande pour une valeur de k petite. Deux exemples unidimensionnels $\mathcal{S}_1(0.75, (0.5, 1), 0)$ (non symétrique) et $\mathcal{S}_1(0.75, (0, 1), 0)$ (symétrique) sont considérés. On simule les variables aléatoires en utilisant la somme (1.3.8) avec différentes valeurs de k . Quand le nombre de termes k augmente, on constate que la densité empirique approche beaucoup plus rapidement vers la densité réelle dans le deuxième exemple que dans le premier exemple, voir la Figure 1.4.

Dans la suite les v.a. 2-dimensionnels de lois 0.75 et 1.5-stables isotropiques sont simulés. La taille d'échantillon $N = 10^7$ et le nombre de termes $k = 10$. On calcule la densité réelle en utilisant la fonction Matlab "mvstablepdf" dans le package appelé "STABLE" obtenu sur demande au site <http://www.RobustAnalysis.com/>. Les courbes de niveau des densités réelles et empiriques sont présentées dans la Figure 1.5. On simule ensuite les v.a. \mathcal{StaS} symétriques de mesures spectrales σ_1 , σ_2 et σ_3 dont les densités sont définies par

$$f_1(\theta) = \begin{cases} \frac{1}{\pi}, & \theta \in [0, \frac{\pi}{2}] \cup [\pi, \frac{3\pi}{2}), \\ 0, & \text{sinon,} \end{cases} \quad (1.3.9)$$

$$f_2(\theta) = \begin{cases} \frac{1}{2} \cos(2\theta), & \theta \in [-\frac{\pi}{4}, \frac{\pi}{4}] \cup [\frac{3\pi}{4}, \frac{5\pi}{4}), \\ 0, & \text{sinon,} \end{cases} \quad (1.3.10)$$

$$f_3(\theta) = \frac{1}{4} |\cos(2\theta)|, \quad \theta \in [0, 2\pi). \quad (1.3.11)$$

Le package de programmes qu'on utilise ne permet pas de calculer la densité d'une loi stable de mesure spectrale absolument continue. Nous présentons ici seulement les

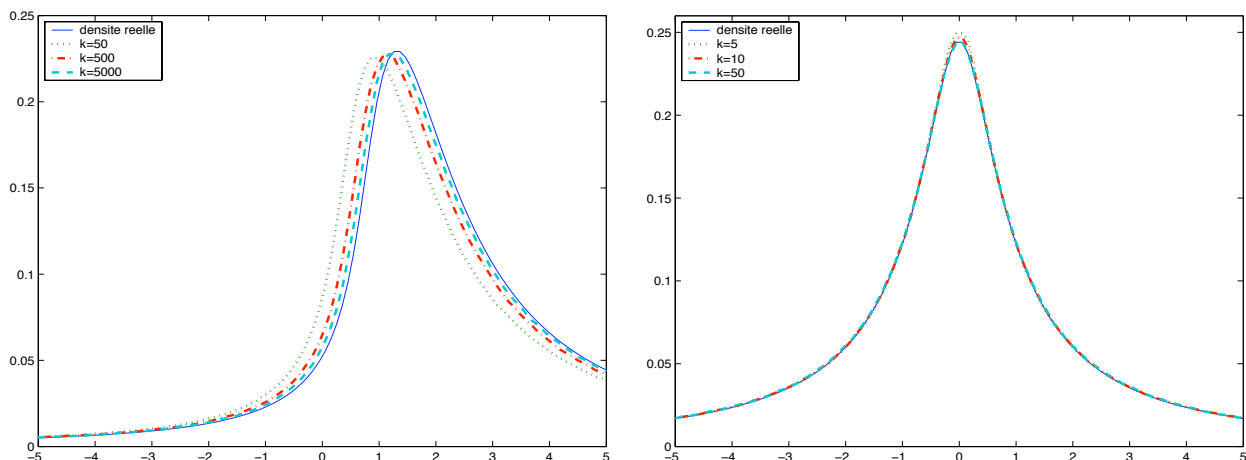


FIG. 1.4: Densités des lois $\mathcal{S}_1(0.75, (0.5, 1), 0)$ (gauche) et $\mathcal{S}_1(0.75, (0, 1), 0)$ (droite) simulées par la somme partielle de série de LePage avec différentes valeurs de k , la taille d'échantillon $N = 10^7$.

courbes de niveau des densités empiriques dans la Figure 1.6. Les lois $\mathcal{S}_2(0.75, \sigma_i, 0)$ et $\mathcal{S}_2(1.5, \sigma_i, 0)$, $i = 1, 2, 3$ sont considérées.

Un des usages de ces données est de tester la robustesse des procédures statistiques multivariées : engendrer aléatoirement l'ensemble de données et évaluer les statistiques par la procédure considérée. Un autre usage possible est de calculer la probabilité $\mathbf{P}\{X \in A\}$ pour l'ensemble $A \subset \mathbb{R}^d$ où X est le v.a. qu'on simule. En général le calcul numérique de ces probabilités est difficile, voir par exemple [NR95] et [Nol97]. On peut les estimer par la méthode standard de Monte Carlo une fois que nous savons générer les vecteurs de loi prescrite. Mais le calcul de simulation devient coûteux si on prend le nombre de termes k grand. Donc cette méthode n'est applicable que dans le cas où la mesure spectrale possède une symétrie radiale. Le troisième usage de ces vecteurs est la simulation dans les domaines d'application, par exemple l'analyse du portefeuille stable [Pre72a].

1.3.2 Lois appartenant au domaine d'attraction d'une loi $\mathcal{St}\alpha\mathcal{S}$

Pour simuler les v.a. appartenant au domaine d'attraction d'une loi $\mathcal{St}\alpha\mathcal{S}$, on applique les transformations présentées dans la section 1.2 aux v.a. α -stables isotropiques. Soit X un v.a. dans \mathbb{R}^2 et $X \in \text{VR}(\alpha, \sigma)$ de mesure spectrale σ uniforme. On identifie la sphère unité S^1 avec l'intervalle $[0, 2\pi)$. Alors $d\sigma/d\theta = 1/2\pi$, $\theta \in [0, 2\pi)$. Soit μ une mesure de probabilité sur S^1 . Sa fonction de répartition (f.d.r.) est notée par $F(x)$, $x \in [0, 2\pi)$. Si Y est un v.a. défini par

$$Y = \|X\|F^{-1}\left(\frac{X}{2\pi\|X\|}\right) = \left(\|X\|, F^{-1}\left(\frac{X}{2\pi\|X\|}\right)\right), \quad (1.3.12)$$

où F^{-1} est la fonction de quantile correspondante de F , alors d'après le corollaire 1.2.2. on a $Y \in \text{VR}(\alpha, \mu)$. On appelle cette transformation "sphérique".

Supposons que la mesure μ est absolument continue par rapport à la mesure de Lebesgue de densité f p.p. continue et bornée. D'après le théorème 1.2.3., le v.a. défini

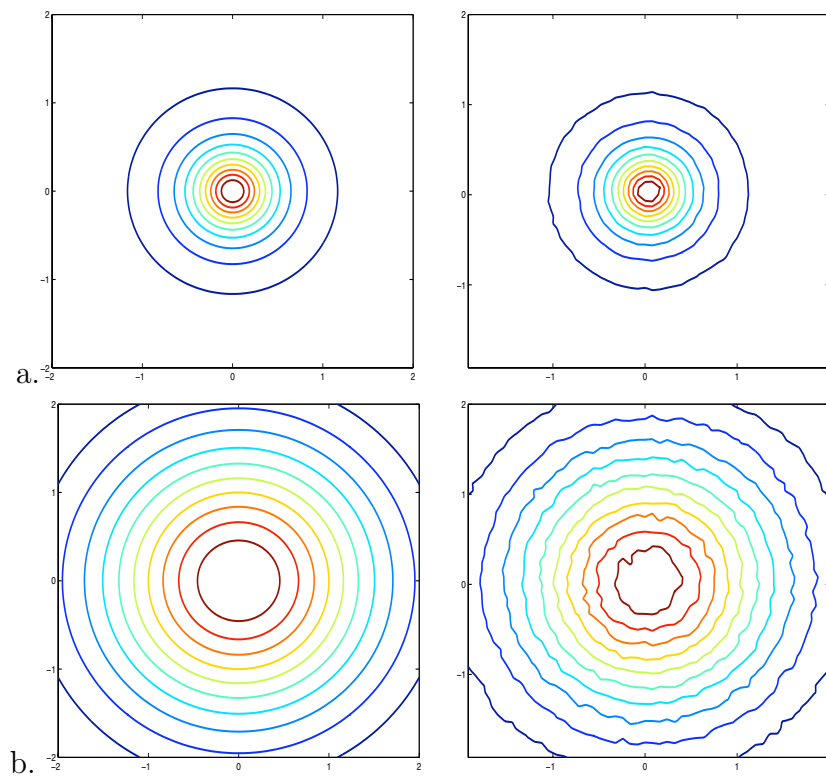


FIG. 1.5: Courbes de niveau des densités réelles (gauche) et empiriques (droite) de lois α -stables isotropiques bivariées, $\alpha = 0.75$ (a), 1.5 (b), $k = 10$, $N = 10^7$.

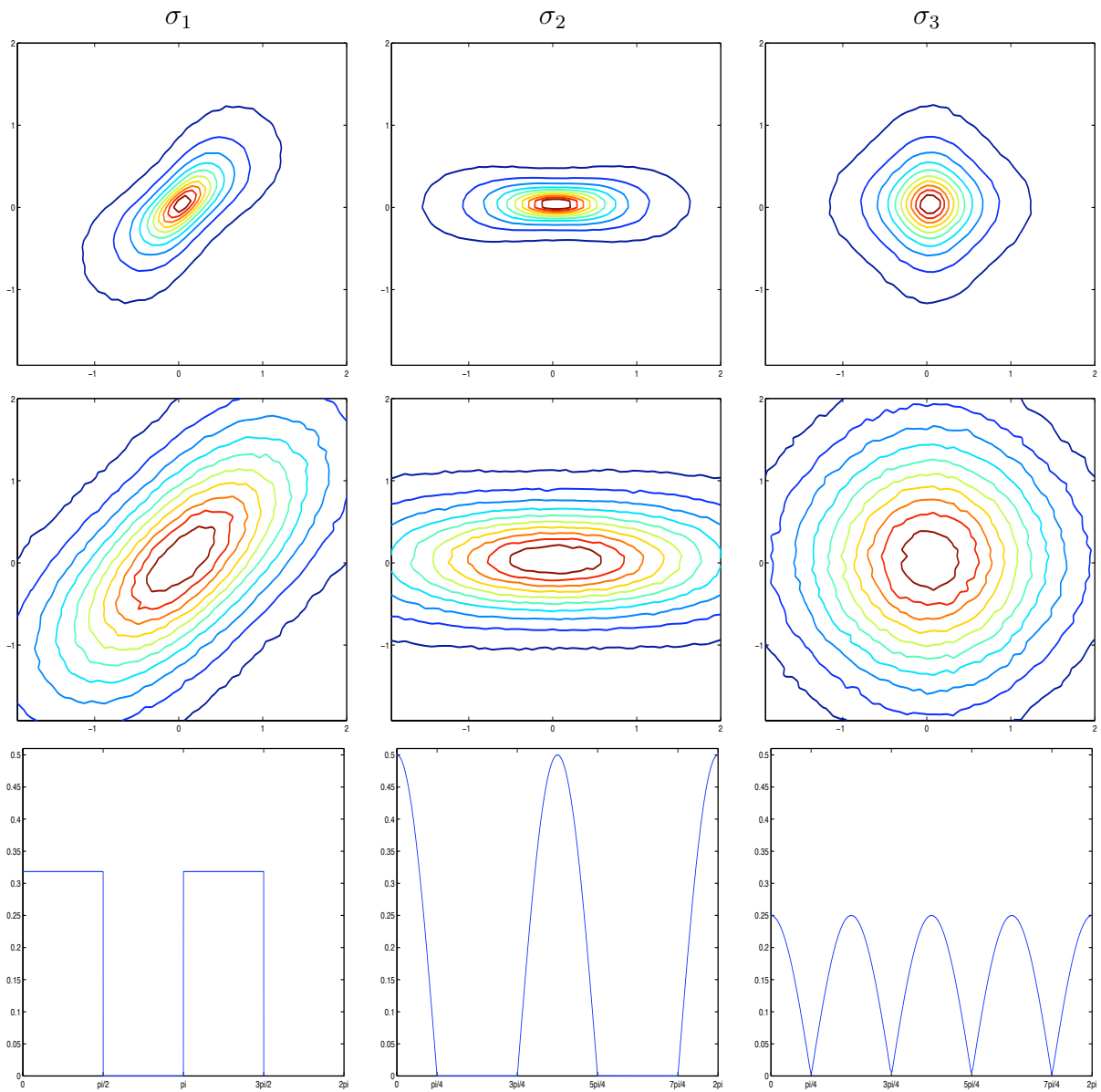


FIG. 1.6: Courbes de niveau des densités empiriques des lois $\mathcal{S}_2(\alpha, \sigma_i, 0)$, $i = 1, 2, 3$, où les densités de σ_i sont définies par (1.3.9), (1.3.10) et (1.3.11), $\alpha = 0.75$ (en haut), 1.5 (au milieu) et les graphes des densités de la mesure spectrale correspondante (en bas), $k = 10$, $N = 10^7$.

par

$$Z = X \left(2\pi f \left(\frac{X}{\|X\|} \right) \right)^{1/\alpha} = \left(\|X\| \left(2\pi f \left(\frac{X}{\|X\|} \right) \right)^{1/\alpha}, \frac{X}{\|X\|} \right). \quad (1.3.13)$$

a une queue à variation régulière de mesure spectrale μ , i.e. $Z \in \text{VR}(\alpha, \mu)$. On appelle cette transformation “*radiale*”. Puisque dans \mathbb{R}^d la condition de variation régulière d’un v.a. est nécessaire et suffisante pour que ce v.a. appartienne au domaine d’attraction d’une loi $\text{Sta}\mathcal{S}$, on a $Y \in \text{Dom}(\alpha, \mu)$ et $Z \in \text{Dom}(\alpha, \mu)$.

Prenons les mesures spectrales $\sigma_i, i = 1, 2, 3$ dont les densités f_i sont définies par (1.3.9) (1.3.10) et (1.3.11) comme μ . On transforme les données simulées de loi 0.75 et 1.5-stable isotropique par (1.3.12) et (1.3.13). Les résultats sont présentés dans les Figures 1.7 et 1.8.

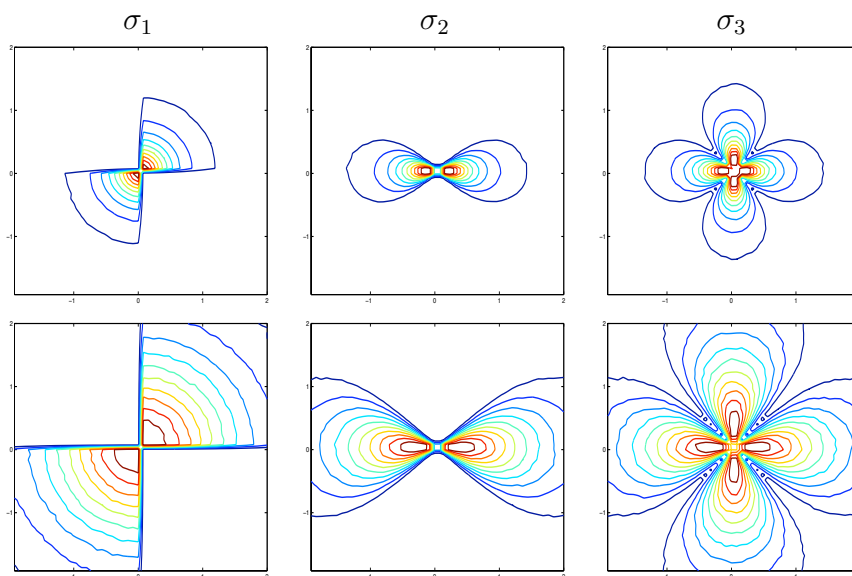


FIG. 1.7: Courbes de niveau des densités empiriques des lois appartenant au $\text{Dom}(\alpha, \sigma_i), i = 1, 2, 3$, $\alpha = 0.75$ (en haut), 1.5 (en bas), simulées par transformations sphériques (1.3.12) d’une loi α -stable isotropique, $N = 10^7$.

1.3.3 Lois max-stables

On considère maintenant les lois stables dans l’espace (\mathbb{R}_+^d, \vee) . Notons S_{\vee}^{d-1} la sphère unité de l’espace (\mathbb{R}_+^d, \vee) , i.e. $S_{\vee}^{d-1} = \{\mathbf{x} \mid \|\mathbf{x}\| = 1\}$ où $\|\cdot\|$ est la norme L_{∞} . La définition de loi max-stable est déduite de la définition de loi $\text{Sta}\mathcal{S}$ générale (2) en remplaçant l’addition par le maximum. Un v.a. $X = (X^{(1)}, \dots, X^{(d)})$ admet une loi *strictement max-stable* si pour tout $a, b > 0$

$$\left(a^{1/\alpha} X_1^{(1)} \vee b^{1/\alpha} X_2^{(1)}, \dots, a^{1/\alpha} X_1^{(d)} \vee b^{1/\alpha} X_2^{(d)} \right) \stackrel{\mathcal{L}}{=} (a + b)^{1/\alpha} (X^{(1)}, \dots, X^{(d)})$$

où X_1, X_2 sont des copies indépendantes de X .

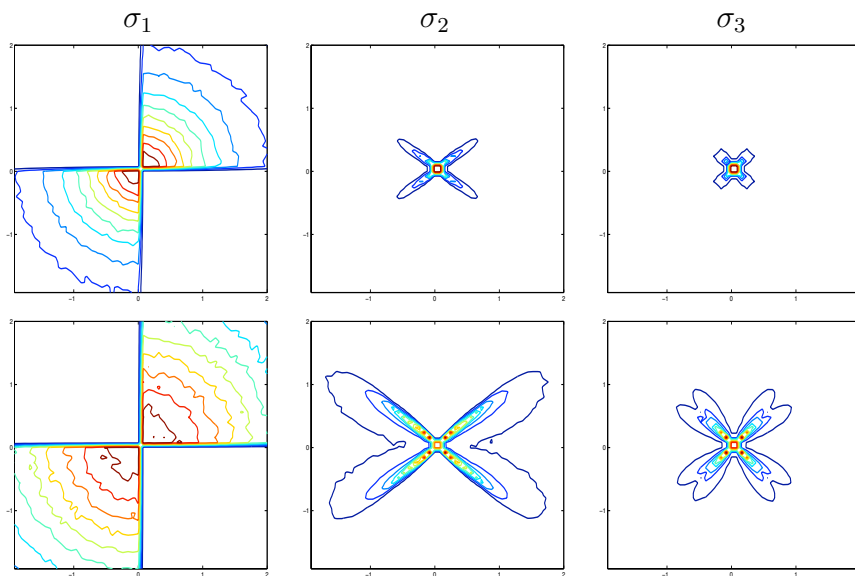


FIG. 1.8: Courbes de niveau des densités empiriques des lois appartenant au $\text{Dom}(\alpha, \sigma_i)$, $i = 1, 2, 3$, $\alpha = 0.75$ (en haut), 1.5 (en bas), simulées par transformations radiales (1.3.13) d'une loi α -stable isotropique, $N = 10^7$.

La série suivante converge absolument p.s. pour tout $\alpha > 0$

$$\xi_{\alpha, \sigma} \stackrel{\mathcal{L}}{=} \bigvee_{i=1}^{\infty} c \Gamma_i^{-1/\alpha} \epsilon_i \quad (1.3.14)$$

où $c > 0$, $\Gamma_i = \sum_{j=1}^i \lambda_j$, $\{\lambda_j\}$ est une suite de variables aléatoires i.i.d. de loi exponentielle standard, ϵ_i est une suite de v.a. i.i.d. dans S_V^{d-1} de loi $\tilde{\sigma}$, λ_j et ϵ_i sont indépendants. Le résultat direct du Théorème 7.15 dans [DMZ08] montre que tous les v.a. strictement max-stables sans composante gaussienne admettent la représentation de LePage (1.3.14). Donc la série (1.3.14) produit un v.a. strictement max-stable déterminé par l'indice α et la mesure finie $\sigma = c^\alpha \tilde{\sigma}(\cdot)$ sur S_V^{d-1} .

Définition 1.3.4. *Un v.a. X à valeurs dans \mathbb{R}_+^d appartient au domaine d'attraction maximum d'un v.a. strictement max-stable, noté ξ , si pour une suite $\{X_n, n \geq 1\}$ de copies indépendantes de X on a*

$$b_n^{-1} \left(\bigvee_{i=1}^n X_i^{(1)}, \dots, \bigvee_{i=1}^n X_i^{(d)} \right) \Rightarrow \xi, \quad (1.3.15)$$

où $\{b_n, n \geq 1\}$ est une suite de constantes de normalisation positives.

On peut montrer que si $\{X_i, i \geq 1\}$ est une suite de v.a. i.i.d. telle que la suite $\left\{ b_n^{-1} \max_{i \in \{1, \dots, n\}} X_i, n \geq 1 \right\}$ converge en loi pour une suite appropriée $b_n > 0$ vers une limite non triviale, c'est-à-dire vers un v.a. non constant, alors la limite est une loi max-

stable. Dans la suite on utilise l'écriture " $X \sim \mathcal{MS}_d(\alpha, \sigma)$ " pour indiquer que X a une loi strictement max-stable d -dimensionnelle d'indice α et de mesure spectrale σ .

On simule les données de loi $\mathcal{MS}_d(\alpha, \sigma)$ en utilisant la somme partielle de la série (1.3.14), i.e.

$$\hat{\xi}_k \stackrel{\mathcal{L}}{=} \bigvee_{i=1}^k \Gamma_i^{-1/\alpha} \epsilon_i = \left(\max_{1 \leq i \leq k} \Gamma_i^{-1/\alpha} \epsilon_i^{(1)}, \dots, \max_{1 \leq i \leq k} \Gamma_i^{-1/\alpha} \epsilon_i^{(d)} \right)$$

où $\epsilon_i = (\epsilon_i^{(1)}, \dots, \epsilon_i^{(d)}) \sim \sigma$ sont v.a. i.i.d. sur la sphère S_V^{d-1} , Γ_i sont définis comme dans (1.3.14). On discute d'abord l'erreur de simulation. Pour chaque ϵ_i une de ses coordonnées $\epsilon_i^{(j)}$ est égale à 1. Supposons que

$$\mathbf{P}\{\epsilon_i^{(j)} = 1\} = p_j > 0, \quad j = 1, \dots, d.$$

Définissons le moment d'arrêt $\tau^{(j)} = \min\{n \mid \epsilon_n^{(j)} = 1\}$. La loi de $\tau^{(j)}$ est géométrique

$$\mathbf{P}\{\tau^{(j)} = n\} = \mathbf{P}\{\epsilon_1^{(j)} < 1, \dots, \epsilon_{n-1}^{(j)} < 1, \epsilon_n^{(j)} = 1\} = (1 - p_j)^{n-1} p_j.$$

La représentation de LePage (1.3.14) peut être écrite comme suit

$$\xi_{\alpha, \sigma} \stackrel{\mathcal{L}}{=} \left(\max_{1 \leq i \leq \tau^{(1)}} \Gamma_i^{-1/\alpha} \epsilon_i^{(1)}, \dots, \max_{1 \leq i \leq \tau^{(d)}} \Gamma_i^{-1/\alpha} \epsilon_i^{(d)} \right).$$

Donc

$$\mathbf{P}\{\hat{\xi}_k \neq \xi_{\alpha, \sigma}\} = \mathbf{P}\{\exists j \text{ tel que } \tau^{(j)} > k\} \leq \sum_{j=1}^d \mathbf{P}\{\tau^{(j)} > k\} = \sum_{j=1}^d (1 - p_j)^k.$$

Cette inégalité montre que la probabilité que $\hat{\xi}_k$ ne soit pas max-stable peut être suffisamment petite pour la valeur de k acceptable.

Dans l'espace (\mathbb{R}_+, \vee) la f.d.r. de loi $\mathcal{MS}_1(\alpha, \sigma)$ est $\exp(-x^{-\alpha} \sigma(1)^\alpha)$.

Dans la suite on considère deux exemples de lois $\mathcal{MS}_2(\alpha, \sigma)$.

Exemple 5. La mesure spectrale est uniforme.

Cet exemple est étudié dans l'Exemple 1 b) de la section 1.1. Les courbes de niveau de la densité empirique sont présentées dans la Figure 1.9 b).

On discute ici la régularité de la queue de cette loi. Soit $\xi_{\alpha, \sigma}$ un v.a. max-stable dans \mathbb{R}^2 défini par (1.3.14) de mesure spectrale σ uniforme. Utilisons les notations de l'Exemple 1 de la section 1.1. Le processus ponctuel poissonnien $\pi_{\alpha, \sigma}$ est défini par (1.1.22). Sa mesure d'intensité est $\sigma \times m$ où m est la mesure sur \mathbb{R}_+^1 telle que $m((r, \infty)) = r^{-\alpha}$ pour tout $r > 0$. En prenant l'ensemble $B = [(1, 0), (1, t)]$ ou $B = [(0, 1), (t, 1)]$, $0 \leq t \leq 1$ on a $\forall r > 0$

$$\begin{aligned} I_1 = \mathbf{P}\{\pi_{\alpha, \sigma}(B \times (r, \infty)) \neq 0, \pi_{\alpha, \sigma}(B^{\mathbb{G}} \times (tr, \infty)) = 0\} &\leq \mathbf{P}\{\xi_{\alpha, \sigma} \in B \times (r, \infty)\} \\ &\leq \mathbf{P}\{\pi_{\alpha, \sigma}(B \times (r, \infty)) \neq 0\} = I_2, \end{aligned}$$

où

$$I_1 = (1 - \exp(-\sigma(B)r^{-\alpha})) \exp(-\sigma(B^{\mathbb{G}})(tr)^{-\alpha}) \text{ et } I_2 = 1 - \exp(-\sigma(B)r^{-\alpha}).$$

On en déduit l'équivalence

$$\mathbf{P}\{\xi_{\alpha,\sigma} \in B \times (r, \infty)\} \sim \sigma(B)r^{-\alpha} \text{ quand } r \rightarrow \infty. \quad (1.3.16)$$

Donc la loi de $\xi_{\alpha,\sigma}$ vérifie la condition de variation régulière.

Exemple 6. La mesure spectrale est concentrée sur deux points.

Notons

$$\sigma(\cdot) = p\delta_{e_1}(\cdot) + q\delta_{e_2}(\cdot) \quad (1.3.17)$$

où $e_1 = (1, a)$, $e_2 = (b, 1)$, $0 \leq a, b \leq 1$ et $p + q = 1$. Il est facile de voir que le v.a. a ses valeurs dans la région entre deux rayons sortant du point zéro vers les points e_1 et e_2 . On en déduit la f.d.r.

$$\mathbf{P}\{\xi \in [0, \mathbf{x}]\} = \begin{cases} \exp(-px_1^{-\alpha} - qx_2^{-\alpha}), & a < x_2/x_1 < 1/b, \\ \exp(-(pa^\alpha + q)x_2^{-\alpha}), & x_2/x_1 < a, \\ \exp(-(p + qb^\alpha)x_1^{-\alpha}), & x_2/x_1 > 1/b. \end{cases}$$

En notant le moment d'arrêt $\tau = \min\{i \mid \epsilon_i = e_2\}$, considérons les probabilités que le v.a. ait ses valeurs dans l'ensemble $\{e_1 \times (0, \infty)\}$

$$\begin{aligned} \mathbf{P}\{\xi_{\alpha,\sigma} \in e_1 \times (0, r)\} &= \sum_{i=2}^{\infty} \mathbf{P}\{\epsilon_1 = e_1, \tau = i, \Gamma_i^{-1/\alpha} < a\Gamma_1^{-1/\alpha}, \Gamma_1^{-1/\alpha} < r\} \\ &= \sum_{i=2}^{\infty} p^{i-1}q \mathbf{P}\left\{\frac{\lambda_2 + \dots + \lambda_i}{\lambda_1} > a^{-\alpha} - 1, \lambda_1 > r^{-\alpha}\right\} \\ &= \left(p - \frac{pq(1 - a^\alpha)}{q + pa^\alpha}\right) \exp(-(p + qa^{-\alpha})r^{-\alpha}). \end{aligned}$$

De manière analogue on a

$$\mathbf{P}\{\xi_{\alpha,\sigma} \in e_2 \times (0, r)\} = \left(q - \frac{pq(1 - b^\alpha)}{p + qb^\alpha}\right) \exp(-(q + pb^{-\alpha})r^{-\alpha}).$$

Cela signifie que plus a et b sont proches de 1, plus il est possible que ξ tombe sur les deux lignes du bord. Si $a = b = 0$ alors ξ a ses valeurs strictement à l'intérieur d'un quart de plan \mathbb{R}^2 et ses composantes sont indépendantes. Cela coïncide avec l'Exemple 1 a) dans la section 1.1. Les Figures 1.9 et 1.10 présentent l'influence des paramètres σ et α pour la densité de $\mathcal{MS}_2(\alpha, \sigma)$ avec σ définie par (1.3.17).

Étudions la régularité de cette loi. Définissons deux v.a.

$$\epsilon^{(1)} \sim p\delta_{e_1}, \quad \epsilon^{(2)} \sim q\delta_{e_2}.$$

Notons $\epsilon_1^{(i)}, \epsilon_2^{(i)}, \dots, i = 1, 2$ des copies indépendantes de $\epsilon^{(i)}$. D'après la théorie des processus d'amincissement (voir par exemple [Res92]), les processus π_1 et π_2 définis par

$$\pi_i = \sum_{j=1}^{\infty} \delta_{\Gamma_j^{-1/\alpha} \epsilon_j^{(i)}}, \quad i = 1, 2$$

sont indépendants et poissonniens. Les mesures d'intensités sont respectivement pm et qm . Prenons l'ensemble

$$B = [(1, s), (1, t)], \quad 0 \leq s < a < t \leq 1.$$

Il est facile de voir que $\forall r > 0$

$$\begin{aligned} I_1 = \mathbf{P}\{\pi_1(e_1 \times (r, \infty)) \neq 0, \pi_2(e_2 \times (tr, \infty)) = 0\} &\leq \mathbf{P}\{\xi_{\alpha, \sigma} \in B \times (r, \infty)\} \\ &\leq \mathbf{P}\{\pi_1(e_1 \times (r, \infty)) \neq 0\} = I_2, \end{aligned}$$

où

$$I_1 = (1 - \exp(-pr^{-\alpha})) \exp(-q(tr)^{-\alpha}) \text{ et } I_2 = 1 - \exp(-pr^{-\alpha}).$$

Ces deux termes sont équivalents à $pr^{-\alpha}$ quand $r \rightarrow \infty$, ainsi on obtient

$$\mathbf{P}\{\xi_{\alpha, \sigma} \in B \times (r, \infty)\} \sim pr^{-\alpha} \text{ quand } r \rightarrow \infty.$$

De même manière, pour $B = [(s, 1), (t, 1)], 0 \leq s < b < t \leq 1$ on a $\mathbf{P}\{\xi_{\alpha, \sigma} \in B \times (r, \infty)\} \sim qr^{-\alpha}$ quand $r \rightarrow \infty$.

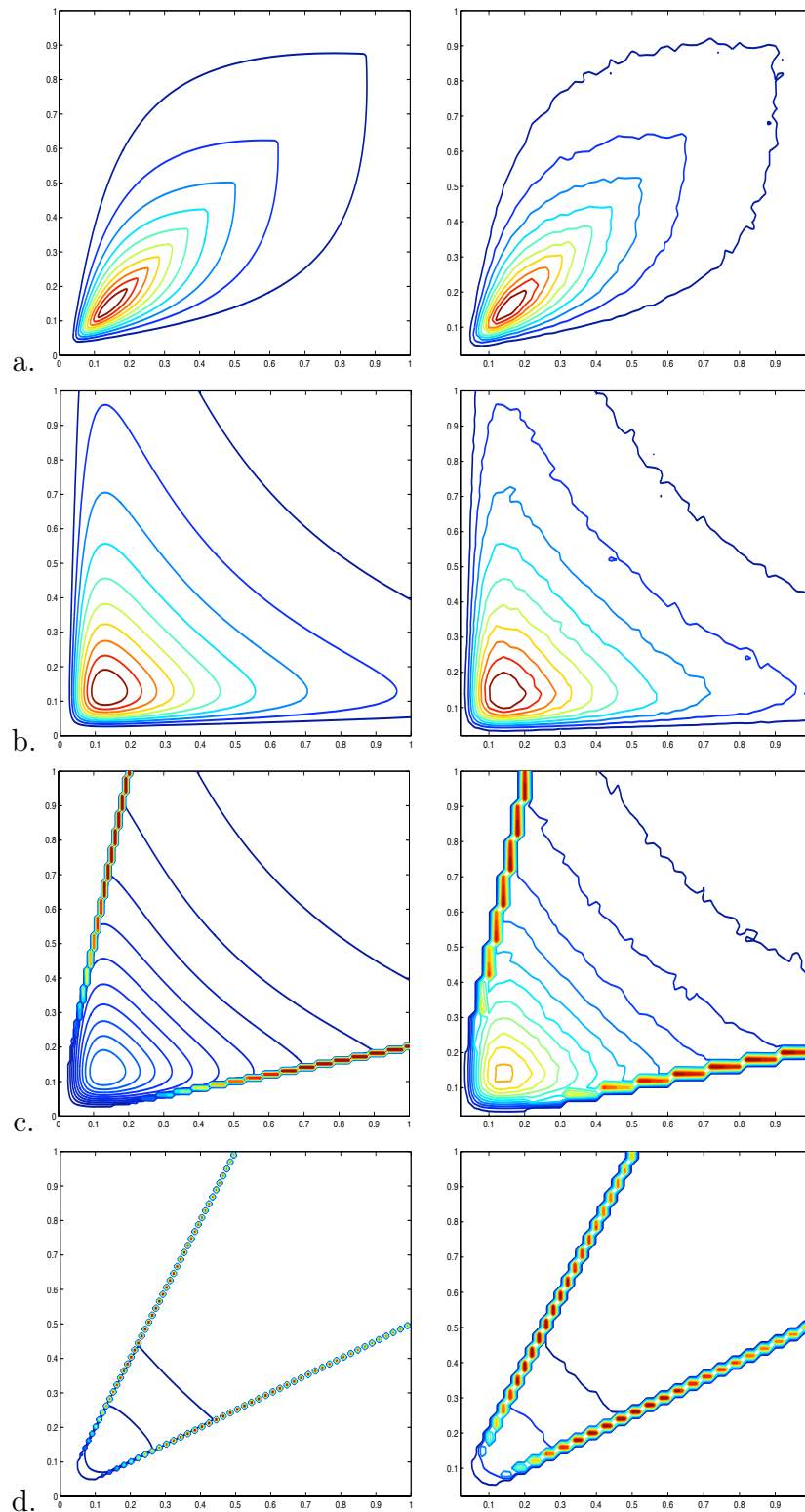


FIG. 1.9: Courbes de niveau des densités réelles (gauche) et empiriques (droite) de la loi $\mathcal{MS}_2(0.75, \sigma)$, σ est uniforme (a), σ est définie par (1.3.17) avec $p = q = 0.5$, $a = b = 0$ (b), $a = b = 0.2$ (c), $a = b = 0.5$ (d), $N = 10^7$.

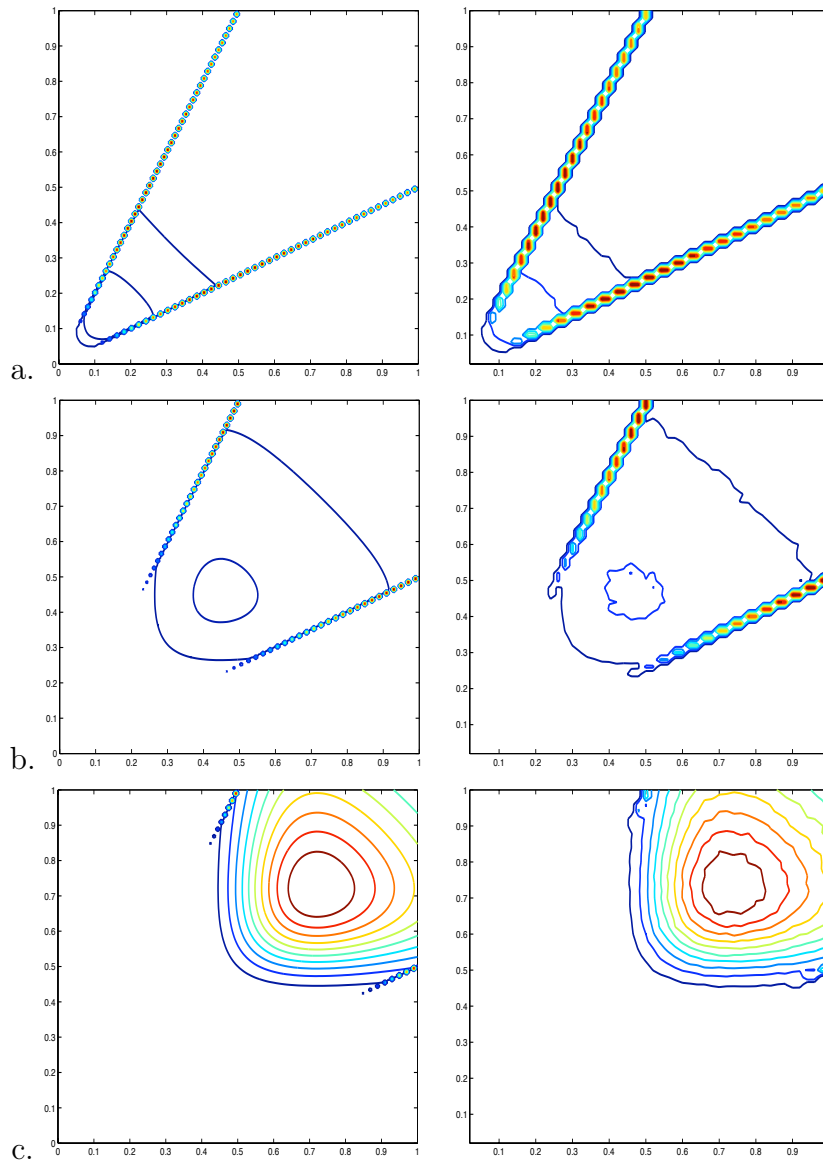


FIG. 1.10: Courbes de niveau des densités réelles (gauche) et empiriques (droite) de la loi $\mathcal{MS}_2(\alpha, \sigma)$ où σ est définie par (1.3.17) avec $p = q = 0.5$, $a = b = 0.5$, $\alpha = 0.75$ (a), $\alpha = 1.5$ (b), $\alpha = 3$ (c), $N = 10^7$.

Chapitre 2

Estimation des paramètres des lois à queue régulière

Nous présentons dans ce chapitre une méthode d'estimation des paramètres des lois à queue régulière dans un cône arbitraire. Nous prouvons d'abord la convergence des statistiques d'ordre de lois à queue régulière dans un cône. Ce résultat nous permet de généraliser une méthode d'estimation des paramètres de lois stables dans \mathbb{R}^d aux lois à queue régulière dans un cône. Un nouvel estimateur de la masse totale de la mesure spectrale est proposé en supposant que la loi marginale appartienne au domaine d'attraction normal d'une loi stable. Pour prouver la normalité asymptotique d'estimation, deux conditions du second ordre sont présentées. La vitesse de convergence des estimateurs dépend de l'exposant du second ordre et de la façon de former l'échantillon en paquets.

Si la mesure spectrale σ d'une loi à queue régulière est absolument continue par rapport à la mesure de Lebesgue, nous cherchons à estimer la densité de σ . La méthode d'estimation de noyau pour les données directionnelles est utilisée.

La méthode d'estimation présentée se détériore pour les petites tailles d'échantillon. Pour résoudre ce problème, nous proposons une méthode de type bootstrap pour augmenter la taille de l'échantillon utilisé.

La performance des estimateurs proposés est étudiée à l'aide de simulation. Quelques méthodes connues d'estimation des paramètres des lois stables sont comparées avec notre méthode sur des échantillons de différentes tailles et de différentes valeurs de paramètres. Les tableaux de performance sont fournis à la fin.

2.1 Estimation de l'indice de queue et de la mesure spectrale

Soit $\xi_1, \xi_2, \dots, \xi_N$ des e.a. i.i.d. dans \mathbb{K} , d'une loi ayant une queue régulière de paramètre inconnu α et de mesure spectrale inconnue σ . Notre but est d'estimer α et σ à partir de cet échantillon. Une méthode basée sur les statistiques d'ordre de loi stable dans \mathbb{R}^d a été introduite dans [DP99] (voir aussi [DPR00]). On généralise cette méthode aux lois à queue régulière dans un cône convexe. La sous-section 2.1.1 contient quelques résultats auxiliaires concernant les statistiques d'ordre d'une loi à queue régulière et la loi forte des grands

nombres pour un schéma triangulaire. La sous-section 2.1.2 présente les estimateurs de l'exposant caractéristique α et de la mesure spectrale σ . Ces estimateurs sont consistants. Sous certaines conditions supplémentaires, nous prouvons la normalité asymptotique des estimateurs dans la sous-section 2.1.3. Comme cette méthode est souvent mentionnée dans la suite, nous résumons les résultats principaux dans l'annexe A.

2.1.1 Résultats auxiliaires

Soit X un e.a. dans \mathbb{K}' qui satisfait la condition de variation régulière (1.1.6). Notons $Y = \|X\|$ et $G(x) = \mathbf{P}\{Y > x\}$. On a donc

$$nG(b_n x) \rightarrow \sigma(S)x^{-\alpha}, \text{ quand } n \rightarrow \infty, \text{ pour tout } x > 0. \quad (2.1.1)$$

Pour x positif fixé, choisissons n le plus petit nombre entier tel que $b_{n+1} > x$. Alors $b_n \leq x < b_{n+1}$ et pour la fonction non-croissante G on a

$$\frac{G(\lambda b_{n+1})}{G(b_n)} \leq \frac{G(\lambda x)}{G(x)} \leq \frac{G(\lambda b_n)}{G(b_{n+1})}, \text{ pour tout } \lambda > 0.$$

Selon (2.1.1) $nG(b_n) \rightarrow \sigma(S)$, on a donc

$$\frac{G(\lambda x)}{G(x)} \rightarrow \lambda^{-\alpha}, \text{ quand } x \rightarrow \infty \text{ pour tout } \lambda > 0.$$

On en déduit que $G \in R_{-\alpha}$, cela nous permet d'écrire l'expression asymptotique

$$G(x) \sim x^{-\alpha}L(x), \quad (2.1.2)$$

où $L(x)$ est une fonction à variation lente.

Rappelons un résultat connu sur l'inverse asymptotique d'une fonction à variation régulière.

Théorème 2.1.1. (*[BGT87] Th. 1.5.12*) Soit $f \in R_\alpha$ avec $\alpha > 0$, alors $\exists g(x) \in R_{1/\alpha}$ telle que la relation suivante a lieu

$$f(g(x)) \sim g(f(x)) \sim x \text{ quand } x \rightarrow \infty. \quad (2.1.3)$$

Ici g (l'inverse asymptotique de f) est déterminé uniquement par rapport à l'équivalence asymptotique, et une version de g est

$$f^-(x) = \inf\{y : f(y) \leq x\}.$$

Notons

$$f(x) = \frac{1}{G(x)} \sim x^\alpha \frac{1}{L(x)}, \quad (2.1.4)$$

alors $f(x) \in R_\alpha$ avec $\alpha > 0$. En appliquant le théorème précédent, on obtient l'inverse $g \in R_{1/\alpha}$ de la forme suivante :

$$g(x) = x^{1/\alpha}L^\sharp(x)$$

où la fonction à variation lente L^\sharp vérifie les relations suivantes

$$L(x)^{-1/\alpha} L^\sharp(f(x)) \rightarrow 1, \quad (2.1.5)$$

et

$$L(g(x))^{-1} L^\sharp(x)^\alpha \rightarrow 1, \quad x \rightarrow \infty. \quad (2.1.6)$$

D'après (2.1.3) et (2.1.4) on obtient

$$G\left(g\left(\frac{1}{x}\right)\right) \sim x, \quad x \rightarrow 0. \quad (2.1.7)$$

Définissons l'inverse généralisée

$$G^{-1}(x) := \inf\{y : G(y) < x\},$$

on va montrer que

$$G(G^{-1}(x)) \sim x, \quad x \rightarrow 0. \quad (2.1.8)$$

Pour cela choisissons $\lambda > 1$, $A > 1$, $\delta \in (0, \infty)$, alors d'après le théorème de Potter (Th. 1.5.6 [BGT87] page 25) il existe u_0 telle que

$$A^{-1}\lambda^{-\alpha-\delta}G(v) \leq G(u) \leq A\lambda^{\alpha+\delta}G(v), \quad \forall v \in [\lambda^{-1}u, \lambda u], \quad u \geq u_0.$$

Prenons x assez petit tel que $G^{-1}(x) \geq u_0$, alors par définition de G^{-1} il existe $y \in [\lambda^{-1}G^{-1}(x), G^{-1}(x)]$ tel que $G(y) \geq x$, et il existe $y' \in [G^{-1}(x), \lambda G^{-1}(x)]$ tel que $G(y') < x$. En prenant $G^{-1}(x)$ comme u , y et y' comme v , on obtient

$$A^{-1}\lambda^{-\alpha-\delta}G(y) \leq G(G^{-1}(x)) \leq A\lambda^{\alpha+\delta}G(y').$$

Donc les lim sup et lim inf de $G(G^{-1}(x))/x$ se trouvent entre $A\lambda^{\alpha+\delta}$ et $A^{-1}\lambda^{-\alpha-\delta}$ quand $x \rightarrow \infty$. En prenant $A, \lambda \downarrow 1$, on obtient $G(G^{-1}(x))/x \rightarrow 1$.

Alors les relations (2.1.7) et (2.1.8) nous donnent immédiatement

$$G^{-1}(x) \sim g(1/x), \quad x \rightarrow 0.$$

On en déduit donc l'expression équivalente de l'inverse de $G(x)$:

$$G^{-1}(x) \sim x^{-1/\alpha} L^\sharp(1/x) \in R_{-1/\alpha}(0+). \quad (2.1.9)$$

Lemme 2.1.2. *Soit X un e.a. dans \mathbb{K}' , $G(x) = \mathbf{P}\{\|X\| > x\}$. Si $G(x)$ satisfait la condition (2.1.1), alors pour chaque $i = 1, 2, \dots$,*

$$b_n^{-1} G^{-1}\left(\frac{\Gamma_i}{\Gamma_{n+1}}\right) \xrightarrow[n \rightarrow \infty]{} \sigma(S)^{1/\alpha} \Gamma_i^{-1/\alpha} \text{ avec probabilité } 1, \quad (2.1.10)$$

où Γ_i est la somme de i variables aléatoires i.i.d. de loi exponentielle standard.

Démonstration: La condition (2.1.1) implique

$$G(x_n) \sim \frac{\sigma(S)}{n} \left(\frac{b_n}{x_n} \right)^\alpha, \quad n \rightarrow \infty$$

où $x_n = b_n x$, $x_n \rightarrow \infty$, quand $n \rightarrow \infty$. En remplaçant l'expression gauche de la formule précédente par (2.1.2), on obtient une expression équivalente de b_n en terme de $L(x)$:

$$b_n \sim \left(\frac{nL(x_n)}{\sigma(S)} \right)^{1/\alpha}, \quad n \rightarrow \infty. \quad (2.1.11)$$

En considérant (2.1.9), on a une expression équivalente avec probabilité 1 pour chaque i fixé :

$$G^{-1} \left(\frac{\Gamma_i}{\Gamma_{n+1}} \right) \sim \left(\frac{\Gamma_i}{\Gamma_{n+1}} \right)^{-1/\alpha} L^\# \left(\frac{\Gamma_{n+1}}{\Gamma_i} \right), \quad n \rightarrow \infty \quad (2.1.12)$$

où $L^\#$ vérifie (2.1.5) qui signifie ici

$$L(x_n)^{-1/\alpha} L^\# \left(\frac{\Gamma_{n+1}}{\Gamma_i} \right) \rightarrow 1, \quad n \rightarrow \infty. \quad (2.1.13)$$

De (2.1.11)-(2.1.13) on déduit qu'avec probabilité 1

$$b_n^{-1} G^{-1} \left(\frac{\Gamma_i}{\Gamma_{n+1}} \right) \sim \sigma(S)^{1/\alpha} \left(\frac{\Gamma_{n+1}}{n\Gamma_i} \right)^{1/\alpha} L(x_n)^{-1/\alpha} L^\# \left(\frac{\Gamma_{n+1}}{\Gamma_i} \right) \sim \left(\frac{\sigma(S)}{\Gamma_i} \right)^{1/\alpha},$$

quand $n \rightarrow \infty$, puisque $\Gamma_{n+1}/n \xrightarrow[n \rightarrow \infty]{p.s.} 1$; ce qui complète la démonstration. \square

Soient Y_1, Y_2, \dots des variables aléatoires i.i.d. suivant la loi $1-G$. Les variables aléatoires $Y_{n,1}, Y_{n,2}, \dots, Y_{n,n}$, $Y_{n,1} \geq Y_{n,2} \geq \dots \geq Y_{n,n}$, représentent les statistiques d'ordre de Y_1, Y_2, \dots, Y_n . Il est bien connu que (voir, par exemple, [Bre68] Section 13.6)

$$(Y_{n,1}, \dots, Y_{n,n}) \stackrel{\mathcal{L}}{=} \left(G^{-1} \left(\frac{\Gamma_1}{\Gamma_{n+1}} \right), \dots, G^{-1} \left(\frac{\Gamma_n}{\Gamma_{n+1}} \right) \right). \quad (2.1.14)$$

Lemme 2.1.3. *Le vecteur $b_n^{-1}(Y_{n,1}, \dots, Y_{n,n}, 0, 0, \dots)$ converge faiblement dans \mathbb{R}^∞ vers $\sigma(S)^{1/\alpha}(\Gamma_1^{-1/\alpha}, \Gamma_2^{-1/\alpha}, \dots)$.*

Démonstration: C'est un résultat qui découle directement de (2.1.10) et (2.1.14). \square

Proposition 2.1.1. *Soient $\{X_{m,i}, 1 \leq i \leq n\}$ des variables aléatoires réelles i.i.d. pour chaque m . Supposons que les indices n et m vérifient les relations suivantes*

$$n \sim N^r, \quad m \sim N^{1-r} \quad \text{quand } N \rightarrow \infty \quad (2.1.15)$$

où $0 < r < 1$ est une constante et $N \in \mathbb{N}$. S'il existe un nombre réel $k > \frac{2}{r}$ et une constante $M > 0$ tels que $\mathbf{E}|X_{m,1}|^k \leq M < \infty$, alors

$$\frac{1}{n} \sum_{i=1}^n X_{m,i} - \mathbf{E}X_{m,1} \xrightarrow[N \rightarrow \infty]{p.s.} 0. \quad (2.1.16)$$

Remarque 9. Ceci est une variante de la loi forte des grands nombres pour un schéma triangulaire. En fait la convergence (2.1.16) reste vraie si l'on remplace la condition $\mathbf{E}|X_{m,1}|^k \leq M$, $k > \frac{2}{r}$ par une hypothèse moins restrictive $\sum_{m=1}^{\infty} m^{-\frac{kr}{2(1-r)}} \mathbf{E}|X_{m,1}|^k < \infty$.

Démonstration: Notons $Y_{m,i} = X_{m,i} - \mathbf{E}X_{m,1}$, alors les variables aléatoires $\{Y_{m,i}, 1 \leq i \leq n\}$ sont i.i.d. et centrées. On a

$$\mathbf{E}|Y_{m,1}|^k \leq \mathbf{E}(|X_{m,1}| + |\mathbf{E}X_{m,1}|)^k \leq \mathbf{E}(2^{k-1}(|X_{m,1}|^k + |\mathbf{E}X_{m,1}|^k)) \leq 2^k \mathbf{E}|X_{m,1}|^k.$$

Il est connu (voir [Ros70]) que pour $k \geq 2$ on a

$$\mathbf{E} \left| \sum_{i=1}^n Y_{m,i} \right|^k \leq c(k) n^{k/2} \mathbf{E}|Y_{m,1}|^k,$$

où $c(k)$ est une constante positive qui ne dépend que de k . On déduit de la condition (2.1.15) qu'il existe une constante $C > 0$ telle que $n \geq CN^r$. Ainsi pour tout $\varepsilon > 0$ on a

$$\mathbf{P} \left\{ \left| \frac{1}{n} \sum_{i=1}^n Y_{m,i} \right| > \varepsilon \right\} \leq \frac{\mathbf{E} \left| \sum_{i=1}^n Y_{m,i} \right|^k}{n^k \varepsilon^k} \leq \frac{2^k c(k) \mathbf{E}|X_{m,1}|^k}{C^{\frac{k}{2}} N^{\frac{kr}{2}} \varepsilon^k} = \frac{c_0}{N^{\frac{kr}{2}} \varepsilon^k},$$

où $c_0 = 2^k c(k) M / C^{\frac{k}{2}}$. Puisque $\frac{kr}{2} > 1$, on peut trouver un nombre positif suffisamment petit ε' tel que $\frac{kr}{2} - \varepsilon' > 1$. En prenant $\varepsilon = \varepsilon_N = N^{-\frac{\varepsilon'}{k}}$ et en appliquant le lemme de Borel-Cantelli, on obtient qu'avec probabilité 1 et pour N suffisamment grand

$$\left| \frac{1}{n} \sum_{i=1}^n X_{m,i} - \mathbf{E}X_{m,1} \right| \leq N^{-\frac{\varepsilon'}{k}},$$

d'où vient la convergence (2.1.16). □

2.1.2 Définition et consistance des estimateurs

On divise l'échantillon $\xi_1, \xi_2, \dots, \xi_N$ en n groupes disjoints $G_{m,1}, \dots, G_{m,n}$, chacun contient m éléments. En pratique on choisit d'abord m et on pose $n = [N/m]$ ($[a]$ représente la partie entière d'un nombre a positif). Quand N tend vers l'infini on a $nm = [N/m]m \sim N$. De plus on demande $n, m \rightarrow \infty$ quand $N \rightarrow \infty$.

On estime d'abord le paramètre α . Posons

$$M_{m,i}^{(1)} = \max\{\|\xi\| \mid \xi \in G_{m,i}\}, \quad i = 1, \dots, n, \quad (2.1.17)$$

c'est-à-dire $M_{m,i}^{(1)}$ est la plus grande norme dans le groupe $G_{m,i}$. Notons $\xi_{m,i}$ tel que

$$\|\xi_{m,i}\| = M_{m,i}^{(1)}, \quad (2.1.18)$$

et

$$M_{m,i}^{(2)} = \max\{\|\xi\| \mid \xi \in G_{m,i} \setminus \{\xi_{m,i}\}\}, \quad i = 1, \dots, n, \quad (2.1.19)$$

$M_{m,i}^{(2)}$ est alors la deuxième grande norme dans le même groupe.

D'après le lemme 2.1.3., il est clair que pour chaque i

$$\left(\frac{M_{m,i}^{(1)}}{b_m}, \frac{M_{m,i}^{(2)}}{b_m} \right) \Rightarrow c(\Gamma_1^{-1/\alpha}, \Gamma_2^{-1/\alpha}) \quad \text{quand } m \rightarrow \infty, \quad (2.1.20)$$

où b_m est défini par (1.1.14) et $c = \sigma(S)^{1/\alpha}$. Ici $\Gamma_i = \sum_{j=1}^i \lambda_j$, et $\lambda_1, \lambda_2, \dots$ sont des variables aléatoires i.i.d. suivant la loi exponentielle standard, i.e. $\mathbf{E}(\lambda_i) = 1$. En posant

$$\varkappa_{m,i} = \frac{M_{m,i}^{(2)}}{M_{m,i}^{(1)}}, \quad S_n = \sum_{i=1}^n \varkappa_{m,i} \quad (2.1.21)$$

on construit un estimateur du paramètre α comme suit,

$$\hat{\alpha}_N = \frac{S_n}{n - S_n}. \quad (2.1.22)$$

On va prouver que cet estimateur est consistant, c'est-à-dire

$$\hat{\alpha}_N \xrightarrow[N \rightarrow \infty]{p.s.} \alpha. \quad (2.1.23)$$

Théorème 2.1.4. Soient ξ_1, \dots, ξ_N des e.a. i.i.d. à valeurs dans \mathbb{K}' tels que pour tout $x > 0$ on a

$$n\mathbf{P}(\|\xi_1\| > b_n x) \sim \sigma(S)x^{-\alpha} \quad \text{quand } n \rightarrow \infty. \quad (2.1.24)$$

Si S_n est défini par (2.1.21) avec $n \sim N^r$, $0 < r < 1$, alors

$$\frac{1}{n} S_n \xrightarrow[N \rightarrow \infty]{p.s.} \frac{\alpha}{1 + \alpha}. \quad (2.1.25)$$

Remarque 10. Cela signifie que la quantité $\frac{S_n}{n - S_n}$ donne un estimateur consistant du paramètre α .

Démonstration: D'après (2.1.20), on obtient pour tout $i, 1 \leq i \leq n$,

$$\varkappa_{m,i} \Rightarrow \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{1/\alpha} \quad \text{quand } m \rightarrow \infty.$$

Comme $0 \leq \varkappa_{m,i} \leq 1$, on a pour tout k entier

$$\mathbf{E} \varkappa_{m,i}^k \xrightarrow[m \rightarrow \infty]{} \mathbf{E} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{k/\alpha}. \quad (2.1.26)$$

Un calcul simple nous donne la formule suivante

$$\mathbf{E} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{1/\alpha} = \frac{\alpha}{\alpha + 1}. \quad (2.1.27)$$

Les variables aléatoires $\varkappa_{m,1}, \dots, \varkappa_{m,n}$ sont i.i.d. et $0 \leq \varkappa_{m,i} \leq 1$. En utilisant la proposition 2.1.1., on obtient que

$$\frac{1}{n} S_n - \mathbf{E} \varkappa_{m,1} \xrightarrow[N \rightarrow \infty]{p.s.} 0. \quad (2.1.28)$$

Ceci avec (2.1.26) et (2.1.27) impliquent (2.1.25), ce qui complète la démonstration du théorème. \square

Maintenant on considère l'estimation de la masse totale $\sigma(S)$. Remarquons qu'en prenant $B = S$ dans (1.1.3) on a (1.1.5) qui implique que la norme de ξ est une variable aléatoire positive ayant une queue à variation régulière d'indice α . La mesure spectrale de cette variable aléatoire, notée σ^* , est une mesure concentrée sur un seul point 1 et $\sigma^*(1) = \sigma(S)$. Donc le problème d'estimation de la mesure spectrale de ξ se réduit à l'estimation du paramètre d'une loi à queue régulière unidimensionnelle. Par ailleurs d'après le lemme 2.1.3., on a $b_n^{-1} \|\xi\|_{N1} \Rightarrow \sigma(S)^{1/\alpha} \Gamma_1^{-1/\alpha}$ où $\|\xi\|_{N1}$ représente la plus grande norme parmi ξ_1, \dots, ξ_N . Donc la valeur de $\sigma(S)$ dépend de la suite de constantes de normalisation b_n . L'estimation de b_n est discutée dans [RS97] et [Res06]. Si l'échantillon est issu d'une loi stable dans \mathbb{R}^d , prenons v un vecteur dans \mathbb{R}^d , alors la projection $\langle \xi, v \rangle$ est de loi stable dans \mathbb{R}^1 . Dans ce cas la masse totale $\sigma(S)$ est en fonction du vecteur v et le paramètre d'échelle de la loi de variable $\langle \xi, v \rangle$ qui peut être estimé par la méthode des moments.

Ici on propose un estimateur de $\sigma(S)$ sous l'hypothèse que l'e.a. ξ satisfait la condition (1.1.6) avec $b_n = n^{1/\alpha}$. Cela signifie que ξ appartient au *domaine d'attraction normal* d'une loi stable. Un exemple typique est la loi $\mathcal{St}\alpha\mathcal{S}$. Notons

$$q_{m,i} = \frac{M_{m,i}^{(1)}}{m^{1/\alpha}} \quad (2.1.29)$$

où $M_{m,i}^{(1)}$ est défini par (2.1.17). L'estimateur qu'on propose est défini par

$$\widehat{\sigma(S)}_N = \left(\frac{1}{n\Gamma(1 - \frac{t}{\alpha})} \sum_{i=1}^n q_{m,i}^t \right)^{\frac{\alpha}{t}}, \quad t > 0. \quad (2.1.30)$$

Lemme 2.1.5. *Si l'e.a. ξ satisfait la condition (2.1.24) avec $b_n = n^{1/\alpha}$, alors pour un nombre réel $0 < t < \alpha$ il existe une constante $C > 0$ telle que pour tous m et i*

$$\mathbf{E} q_{m,i}^t \leq C < \infty. \quad (2.1.31)$$

Démonstration: On déduit de la condition (2.1.24) et $b_n = n^{1/\alpha}$ l'équivalence suivante

$$\mathbf{P}\{\|\xi\| > x\} \sim \sigma(S)x^{-\alpha} \quad \text{quand } x \rightarrow \infty.$$

Alors il existe $\delta > 0$ et une constante $M > 0$ telle que pour $x > \delta$ on a

$$\begin{aligned} \mathbf{P}\{q_{m,i}^t \geq x\} &= \mathbf{P}\{M_{m,i}^{(1)} \geq m^{1/\alpha} x^{1/t}\} \\ &= 1 - (\mathbf{P}\{\|\xi\| \leq m^{1/\alpha} x^{1/t}\})^m \\ &\leq m \mathbf{P}\{\|\xi\| > m^{1/\alpha} x^{1/t}\} \\ &\leq M x^{-\alpha/t}. \end{aligned}$$

Ainsi en prenant $x_0 > \delta$ et $C = x_0 + M x_0^{1-\alpha/t} \frac{\alpha}{\alpha-t}$ on a

$$\begin{aligned} \mathbf{E}q_{m,i}^t &= \int_{\{x < x_0\} \cup \{x \geq x_0\}} x d\mathbf{P}_{q_{m,i}^t}(x) \\ &\leq x_0 + x_0 \mathbf{P}\{q_{m,i}^t \geq x_0\} + \int_{x_0}^{\infty} \mathbf{P}\{q_{m,i}^t \geq x\} dx \\ &\leq x_0 + M x_0^{1-\alpha/t} + \int_{x_0}^{\infty} M x^{-\alpha/t} dx \\ &= C. \end{aligned}$$

□

Théorème 2.1.6. *Supposons que l'e.a. ξ satisfait la condition (2.1.24) avec $b_n = n^{1/\alpha}$, l'estimateur $\widehat{\sigma(S)}_N$ est définie par (2.1.30) avec $n \sim N^r$, $0 < r < 1$. Alors pour $0 < t < \frac{\alpha r}{2}$ on a*

$$\widehat{\sigma(S)}_N - \sigma(S) \xrightarrow[N \rightarrow \infty]{p.s.} 0. \quad (2.1.32)$$

Démonstration: Il suffit de montrer la convergence suivante

$$\frac{1}{n} \sum_{i=1}^n q_{m,i}^t - \Gamma\left(1 - \frac{t}{\alpha}\right) \sigma(S)^{\frac{t}{\alpha}} \xrightarrow[N \rightarrow \infty]{p.s.} 0. \quad (2.1.33)$$

En considérant (2.1.20) on a pour chaque $i = 1, \dots, n$

$$q_{m,i}^t \Rightarrow \frac{\sigma(S)^{t/\alpha}}{\Gamma_1^{t/\alpha}} \text{ quand } m \rightarrow \infty.$$

En prenant $t = \frac{\alpha r \delta}{2}$ avec $0 < \delta < 1$ et $k = \frac{2\delta'}{r}$ avec $1 < \delta' < \frac{1}{\delta}$, on a $2t < tk < \alpha$, ainsi il existe une constante C telle que $\mathbf{E}q_{m,i}^{tk} \leq C < \infty$ d'après le lemme 2.1.5.. Puisque $k > \frac{2}{r}$, d'après la proposition 2.1.1. on obtient la convergence (2.1.33). □

Dans la suite nous supposons que la mesure spectrale σ est normalisée, c'est-à-dire $\sigma(S) = 1$. Notons

$$\theta_{m,i} = \frac{\xi_{m,i}}{\|\xi_{m,i}\|}, \quad i = 1, \dots, n, \quad (2.1.34)$$

où $\xi_{m,i}$ est défini par (2.1.18). Les e.a. $\theta_{m,1}, \dots, \theta_{m,n}$ sont i.i.d. à valeurs dans S .

Lemme 2.1.7. Soient ξ_1, \dots, ξ_N des e.a. i.i.d. à valeurs dans \mathbb{K}' tels que la condition de régularité (1.1.6) est satisfaite. Si $\theta_{m,i}$ est défini par (2.1.34), alors

$$\theta_{m,i} \Rightarrow \sigma \text{ quand } m \rightarrow \infty \quad (2.1.35)$$

pour chaque i .

Démonstration: En prenant un ensemble borélien B sur la sphère unité S tel que $\sigma(\partial B) = 0$, on a

$$\begin{aligned} \mathbf{P}\{\theta_{m,i} \in B\} &= \mathbf{P}\{\xi_{m,1}/\|\xi_{m,1}\| \in B\} \\ &= \sum_{k=1}^m \mathbf{P}\{\xi_{m,1}/\|\xi_{m,1}\| \in B, \xi_{m,1} = \xi_k\} \\ &= m\mathbf{P}\{\xi_m/\|\xi_m\| \in B, \xi_{m,1} = \xi_m\} \\ &= m\mathbf{P}\{\xi_m/\|\xi_m\| \in B, \|\xi_m\| \geq \|\xi_k\|, \forall k = 1, \dots, m-1\} \\ &= m\mathbf{P}\{\xi_m/\|\xi_m\| \in B, \|\xi_m\| \geq b_m\tau_{m-1}\} \\ &= \int m\mathbf{P}\{\xi_m/\|\xi_m\| \in B, \|\xi_m\| \geq b_mx\} \mathbf{P}_{\tau_{m-1}}(dx). \end{aligned}$$

où $\mathbf{P}_{\tau_{m-1}}$ est la loi de $\tau_{m-1} = \max_{1 \leq k \leq m-1} (\|\xi_k\| b_m^{-1})$.

D'après (1.1.6) et le lemme 2.1.3. le dernier terme converge vers

$$\int \sigma(B)x^{-\alpha} \mathbf{P}_{\Gamma_1^{-1/\alpha}}(dx) = \sigma(B) \mathbf{E}(\Gamma_1^{-1/\alpha})^{-\alpha} = \sigma(B).$$

□

Ainsi pour chaque ensemble borélien B sur la sphère unité S tel que $\sigma(\partial B) = 0$, on a

$$\mathbb{I}_B(\theta_{m,i}) \Rightarrow \mathbb{I}_B(\eta) \text{ quand } m \rightarrow \infty,$$

où η est un e.a. de loi σ . Cela implique

$$\mathbf{E}\mathbb{I}_B(\theta_{m,i}) \xrightarrow{m \rightarrow \infty} \sigma(B). \quad (2.1.36)$$

S'il existe une constante $r > 0$ telle que $n \sim N^r$, en appliquant la proposition 2.1.1. au schéma triangulaire des variables aléatoires $\mathbb{I}_B(\theta_{m,1}), \dots, \mathbb{I}_B(\theta_{m,n})$, on obtient que pour chaque $B \in \mathcal{B}(S)$, $\sigma(\partial B) = 0$ fixé,

$$\frac{1}{n} \sum_{i=1}^n \mathbb{I}_B(\theta_{m,i}) - \mathbf{E}\mathbb{I}_B(\theta_{m,1}) \xrightarrow[N \rightarrow \infty]{p.s.} 0. \quad (2.1.37)$$

En considérant (2.1.36) et (2.1.37) on obtient le résultat suivant,

Théorème 2.1.8. Soient ξ_1, \dots, ξ_N des éléments aléatoires i.i.d. à valeurs dans \mathbb{K}' tels que la condition de régularité (1.1.6) est satisfaite. Si $\theta_{m,i}$ est défini par (2.1.34) avec $n \sim N^r$, $0 < r < 1$, alors la loi empirique basée sur l'échantillon $\theta_{m,1}, \dots, \theta_{m,n}$ est un estimateur consistant de la mesure spectrale σ , c'est-à-dire $\forall B \in \mathcal{B}(S)$ tel que $\sigma(\partial B) = 0$,

$$\hat{\sigma}_N(B) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_B(\theta_{m,i}) \xrightarrow[N \rightarrow \infty]{p.s.} \sigma(B). \quad (2.1.38)$$

C'est un résultat sur les ensembles fixés. D'après la proposition suivante on a une convergence plus forte.

Proposition 2.1.2. *Soit (S, \mathcal{S}) un espace métrique séparable et complet. Soit $\{\sigma_n\}$ une suite des mesures de probabilité aléatoires sur S . Si σ est une mesure de probabilité sur (S, \mathcal{S}) telle que pour chaque ensemble $B \in \mathcal{B}(S)$ et $\sigma(\partial B) = 0$ on a la convergence $\sigma_n(B) \xrightarrow{p.s.} \sigma(B)$, alors*

$$\sigma_n \xrightarrow{p.s.} \sigma \text{ quand } n \rightarrow \infty.$$

Démonstration: Notons la famille des ensembles de σ -continuité

$$\mathcal{D}_\sigma = \{B \mid B \in \mathcal{B}(S), \sigma(\partial B) = 0\}.$$

Comme S est séparable, il existe un ensemble dénombrable et dense dans S noté W :

$$W = \{x_1, x_2, \dots\}, \quad x_i \in S, \quad i = 1, 2, \dots$$

Notons les voisinages des boules pour les points dans W

$$V(x_i, r) = \{x \mid x \in S, \|x - x_i\| < r\}.$$

Puisque pour chaque $x_i \in W$ les frontières $\partial\{V(x_i, r)\} \subset \{x \mid \|x - x_i\| = r\}$ sont disjointes pour les r différents, ainsi au maximum des ensembles dénombrables d'entre eux peuvent avoir σ -mesure positive. Par conséquent, il existe une suite de nombres positifs $r_k^i \downarrow 0$ quand $k \rightarrow \infty$ pour chaque x_i telle que

$$\mathbb{L}_i = \{V(x_i, r_k^i), k = 1, 2, \dots\} \subset \mathcal{D}_\sigma.$$

On obtient alors une famille dénombrable $\mathbb{L} = \bigcup_{x_i \in W} \mathbb{L}_i$. Il est clair que pour chaque $x_i \in W$, la famille \mathbb{L}_i est une base locale au point x_i pour la topologie \mathcal{S} . Puisque W est dense dans S , \mathbb{L} est une base de \mathcal{S} . La tribu engendrée par \mathbb{L} , notée $\sigma(\mathbb{L})$, est égale à la tribu borélienne $\mathcal{B}(S)$.

Maintenant on élargit \mathbb{L} en ajoutant des intersections finies des membres de \mathbb{L} :

$$\mathcal{L} = \mathbb{L} \cup \left\{ \bigcap_{i \in I} V_i \mid V_i \in \mathbb{L}, I \subset \mathbb{N}, \text{card}(I) < \infty \right\}.$$

Il est clair que \mathcal{L} est encore dénombrable et $\sigma(\mathcal{L}) = \mathcal{B}(S)$, de plus $\mathcal{L} \subset \mathcal{D}_\sigma$. Puisque $\sigma_n(B) \xrightarrow{p.s.} \sigma(B)$ pour tout $B \in \mathcal{B}(S)$ et $\sigma(\partial B) = 0$, donc $\forall V \in \mathcal{L}, \exists \Lambda_V \subset \Omega$ et $\mathbf{P}(\Lambda_V) = 0$, tel que $\forall \omega \in \Lambda_V^c$ on a

$$\sigma_n(\omega, V) \rightarrow \sigma(V). \quad (2.1.39)$$

Si on note $\Lambda = \bigcup_{V \in \mathcal{L}} \Lambda_V$, alors $\mathbf{P}(\Lambda) = 0$. De plus $\forall \omega \in \Lambda^c$ on a toujours la convergence (2.1.39) pour tout $V \in \mathcal{L}$. La famille \mathcal{L} est fermée sous l'opération d'intersection finie. Grâce au théorème 2.2 dans [Bil68] (page 14) on obtient $\sigma_n \Rightarrow \sigma, \forall \omega \in \Lambda^c$, ce qui implique $\sigma_n \xrightarrow{p.s.} \sigma$. \square

Corollaire 2.1.9. *Soit $\hat{\sigma}_N$ défini comme dans le théorème 2.1.8., alors*

$$\hat{\sigma}_N \xrightarrow{p.s.} \sigma \text{ quand } N \rightarrow \infty.$$

Démonstration: C'est un résultat direct du théorème 2.1.8. et de la proposition précédente. \square

2.1.3 Normalité asymptotique

Notre but est de choisir n et m d'une manière telle qu'avec une normalisation appropriée les lois des estimateurs $\hat{\alpha}_N$, $\widehat{\sigma(S)}_N$ et $\hat{\sigma}_N$ soient asymptotiquement normales. Cela nous donne l'idée sur la vitesse optimale de convergence des estimateurs.

Supposons que ξ appartient au domaine d'attraction normale d'une loi $\mathcal{St}\alpha\mathcal{S}$, c'est-à-dire la condition de régularité (1.1.6) est satisfaite avec $b_n = n^{1/\alpha}$. Dans ce cas on en déduit

$$\lim_{x \rightarrow \infty} x^\alpha G(x) = \sigma(S),$$

où $G(x) = \mathbf{P}\{\|\xi\| > x\}$. Ici on suppose une condition plus forte appelée *la relation asymptotique du second ordre*

$$G(x) = c_1 x^{-\alpha} + c_2 x^{-\rho} + o(x^{-\rho}), \quad x \rightarrow \infty, \quad (2.1.40)$$

avec $0 < \alpha < \rho \leq \infty$, où $c_1 = \sigma(S)$. Cette relation est connue dans \mathbb{R}^d . Fristedt [Fri72] a donné un développement asymptotique pour la loi de la norme d'un v.a. $\mathcal{St}\alpha\mathcal{S}$ dans \mathbb{R}^d

$$G(x) = c_1 x^{-\alpha} + c_2 x^{-2\alpha} + O(x^{-3\alpha}), \quad \text{quand } x \rightarrow \infty, \quad (2.1.41)$$

c'est-à-dire $\rho = 2\alpha$. Si $\rho = \infty$, $G(x)$ est la loi de Paréto. Mais dans le cas de cône, la situation est beaucoup plus compliquée, cette relation n'est plus évidente.

Théorème 2.1.10. *Supposons que l'e.a. ξ satisfait la condition (2.1.40). Soit S_n défini par (2.1.21). Si on choisit*

$$n = N^{2\zeta/(1+2\zeta)-\varepsilon}, \quad m = N^{1/(1+2\zeta)+\varepsilon}, \quad (2.1.42)$$

où $\zeta = (\rho - \alpha)/\alpha$ et ε est une constante strictement positive et suffisamment petite, alors

$$\frac{\sqrt{n} \left(\frac{1}{n} S_n - \frac{\alpha}{\alpha + 1} \right)}{\left(\frac{1}{n} \sum_{i=1}^n \chi_{m,i}^2 - \left(\frac{1}{n} S_n \right)^2 \right)^{1/2}} \Rightarrow \mathcal{N}(0, 1). \quad (2.1.43)$$

Remarque 11.

- 1) Comme $\hat{\alpha}_N$ est construit en utilisant les normes d'échantillon, la propriété de $\hat{\alpha}_N$ reste la même comme celle dans \mathbb{R}^d . Donc la démonstration de ce théorème ressemble à celle du Théorème 1 dans [Pau03].

2) Si l'expression de gauche de (2.1.43) était remplacée tout simplement par

$$\sqrt{n} \left(\frac{1}{n} S_n - \frac{\alpha}{\alpha + 1} \right),$$

la loi normale limite aurait la variance

$$\sigma^2 = \alpha((\alpha + 1)^2(\alpha + 2))^{-1}.$$

C'est-à-dire que la loi limite dépendrait du paramètre inconnu α . Cette forme n'est pas pratique pour les applications. C'est pourquoi on considère la somme de type auto-normalisé au lieu de la forme habituelle.

3) Dans le cas où l'échantillon vient d'une loi stable dans \mathbb{R}^d , le résultat de Fristedt (2.1.41) implique $\rho = 2\alpha$. Cela nous permet de choisir $n = N^{2/3-\varepsilon}$, $m = N^{1/3+\varepsilon}$, alors $\mathbf{E} \left(\frac{1}{n} S_n - \frac{\alpha}{\alpha + 1} \right) = o(n^{-1/2}) = o(N^{-1/3+\varepsilon/2})$. Donc la vitesse de convergence d'estimateur dans \mathcal{L}_1 s'approche de $N^{1/3}$.

Démonstration: D'après la démonstration du théorème 2.1.4., notons

$$a_m = \mathbf{E} \mathcal{X}_{m,1} = \frac{\alpha}{\alpha + 1} + r_m$$

où $r_m \rightarrow 0$ quand $m \rightarrow \infty$. Puisque

$$\sqrt{n} \left(\frac{1}{n} S_n - \frac{\alpha}{\alpha + 1} \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathcal{X}_{m,i} - a_m) + \sqrt{n} r_m,$$

il est facile de voir que (2.1.43) aura lieu si on a les trois relations suivantes :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathcal{X}_{m,i} - a_m) \Rightarrow \mathcal{N}(0, \sigma^2), \quad (2.1.44)$$

$$\sqrt{n} r_m \rightarrow 0, \quad (2.1.45)$$

$$\frac{1}{n} \sum_{i=1}^n \mathcal{X}_{m,i}^2 - \left(\frac{1}{n} S_n \right)^2 \xrightarrow[N \rightarrow \infty]{P} \sigma^2. \quad (2.1.46)$$

Puisque $0 < \mathcal{X}_{m,i} \leq 1$, les variables aléatoires $\{\mathcal{X}_{m,i} - a_m, 1 \leq i \leq n\}$ sont i.i.d. et uniformément bornées. Supposons que $|\mathcal{X}_{m,i} - a_m| \leq M < \infty$. La variance limite est

$$\sigma_m^2 := \mathbf{E}(\mathcal{X}_{m,i} - a_m)^2 \rightarrow \sigma^2 = \alpha((\alpha + 1)^2(\alpha + 2))^{-1}. \quad (2.1.47)$$

Cette relation est une conséquence de (2.1.26) où la valeur limite est obtenue par l'évaluation de $\mathbf{E}(\lambda_1/(\lambda_1 + \lambda_2))^{2/\alpha}$. Considérons la condition de Lindeberg : pour tout $\varepsilon > 0$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n \sigma_m^2} \sum_{i=1}^n \int_{\{|\mathcal{X}_{m,i} - a_m| > \varepsilon \sqrt{n} \sigma_m\}} (\mathcal{X}_{m,i} - a_m)^2 d\mathbf{P} \\ & \leq \lim_{n \rightarrow \infty} \frac{M^2}{\sigma_m^2} \mathbf{P}\{|\mathcal{X}_{m,1} - a_m| > \varepsilon \sqrt{n} \sigma_m\} \\ & = 0. \end{aligned} \quad (2.1.48)$$

La relation (2.1.44) découle du théorème central limite appliqué au schéma triangulaire $\{\varkappa_{m,i} - a_m, 1 \leq i \leq n\}$.

Comme la suite des variables aléatoires $\{\varkappa_{m,i}, 1 \leq i \leq n\}$ est uniformément bornée, le moment $\mathbf{E}\varkappa_{m,i}^{2k}$ existe pour tout $k > 0$, et pour tous m et i . En appliquant la proposition 2.1.1. au schéma triangulaire $\{\varkappa_{m,i}^2, 1 \leq i \leq n\}$, on a

$$\frac{1}{n} \sum_{i=1}^n \varkappa_{m,i}^2 - \mathbf{E}\varkappa_{m,i}^2 \xrightarrow[N \rightarrow \infty]{p.s.} 0.$$

D'après (2.1.28) on a

$$\left(\frac{1}{n} S_n\right)^2 - a_m^2 \xrightarrow[N \rightarrow \infty]{p.s.} 0.$$

Donc

$$\frac{1}{n} \sum_{i=1}^n \varkappa_{m,i}^2 - \left(\frac{1}{n} S_n\right)^2 - (\mathbf{E}\varkappa_{m,i}^2 - a_m^2) \xrightarrow[N \rightarrow \infty]{p.s.} 0,$$

en considérant (2.1.47) on obtient (2.1.46).

Maintenant (2.1.44), (2.1.47) et le Théorème 3.2 de [Bil68] impliquent la convergence des lois jointes :

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (\varkappa_{m,i} - a_m), \frac{1}{n} \sum_{i=1}^n (\varkappa_{m,i} - a_m)^2\right) \Rightarrow (N(0, \sigma^2), \sigma^2).$$

Ainsi il reste à vérifier la convergence (2.1.45). Sous la condition (2.1.40) on a $|r_m| \leq Cm^{-\zeta}$ où $\zeta = (\rho - \alpha)/\alpha$ et C dépend de c_1, c_2, α et ρ (voir [Pau03], Lemma). Donc en prenant

$$n = N^{2\zeta/(1+2\zeta)-\varepsilon}, m = N^{1/(1+2\zeta)+\varepsilon}$$

où ε est une constante suffisamment petite, on obtient $\sqrt{nr_m} \rightarrow 0$. \square

En utilisant la Delta-method on obtient un théorème central limite ayant la forme plus commode pour les applications.

Corollaire 2.1.11. *Supposons que les hypothèses dans le théorème 2.1.10. sont toutes remplies, alors*

$$\frac{\sqrt{n} \left(\frac{S_n}{n - S_n} - \alpha\right)}{(\alpha + 1)^2 \left(\frac{1}{n} \sum_{i=1}^n \varkappa_{m,i}^2 - \left(\frac{1}{n} S_n\right)^2\right)^{1/2}} \Rightarrow \mathcal{N}(0, 1). \quad (2.1.49)$$

Remarque 12. En considérant (2.1.46) et (2.1.47) on obtient que la variance asymptotique de l'estimateur $\hat{\alpha}_N$ est $\frac{\alpha(\alpha+1)^2}{\alpha+2}$. Cela signifie que les intervalles de confiance augmentent avec α .

Démonstration: Prenons la fonction $f(x) = \frac{x}{1-x}$, alors

$$\frac{S_n}{n - S_n} - \alpha = f\left(\frac{S_n}{n}\right) - f\left(\frac{\alpha}{\alpha+1}\right) = f'(\tau_n) \left(\frac{S_n}{n} - \frac{\alpha}{\alpha+1}\right) \quad (2.1.50)$$

où τ_n est un réel positif compris entre $\frac{S_n}{n}$ et $\frac{\alpha}{\alpha+1}$. La convergence (2.1.25) implique

$$\tau_n \xrightarrow[N \rightarrow \infty]{p.s.} \frac{\alpha}{\alpha+1}.$$

Puisque $f'(x) = \frac{1}{(1-x)^2}$ est continue en $\frac{\alpha}{\alpha+1}$, on a

$$f'(\tau_n) \xrightarrow[N \rightarrow \infty]{p.s.} f'\left(\frac{\alpha}{\alpha+1}\right). \quad (2.1.51)$$

Les convergences (2.1.43) et (2.1.51) et l'égalité (2.1.50) implique (2.1.49) \square

Dans la suite on va discuter la normalité asymptotique de l'estimateur de la masse totale défini par (2.1.30).

Théorème 2.1.12. *Supposons que l'e.a. ξ satisfait la condition (2.1.40) avec $\rho > \alpha + 1$, la quantité $q_{m,i}$ est définie par (2.1.29) avec $n \sim N^r$, $0 < r < 1$. Choisissons*

$$\begin{aligned} r &= \frac{3\alpha - 4(\rho - 1) + \sqrt{16(\rho - 1)^2 - 8\alpha(\rho - 1) - 7\alpha^2}}{2\alpha} - \varepsilon, \quad \text{si } \rho \leq \frac{11}{8}\alpha + 1, \\ r &= \frac{1}{2} - \varepsilon, \quad \text{si } \rho > \frac{11}{8}\alpha + 1, \end{aligned} \quad (2.1.52)$$

où ε est une constante suffisamment petite. Alors dans deux cas

$$\begin{aligned} \text{a) } 0 < t < \frac{\alpha r}{4} \wedge 1 & \quad \text{si } \alpha + 1 < \rho \leq \frac{11}{8}\alpha + 1 \quad \text{ou} \quad \rho \geq \frac{3}{2}\alpha + 1, \\ \text{b) } 0 < t < \frac{3\alpha + 2 - 2\rho}{2} \wedge 1 & \quad \text{si } \frac{11}{8}\alpha + 1 < \rho < \frac{3}{2}\alpha + 1, \end{aligned} \quad (2.1.53)$$

on a

$$\frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n q_{m,i}^t - \Gamma\left(1 - \frac{t}{\alpha}\right) \sigma(S)^{t/\alpha} \right)}{\left(\frac{1}{n} \sum_{i=1}^n q_{m,i}^{2t} - \left(\frac{1}{n} \sum_{i=1}^n q_{m,i}^t \right)^2 \right)^{1/2}} \Rightarrow \mathcal{N}(0, 1). \quad (2.1.54)$$

Remarque 13. Dans le cas où l'échantillon est issu d'une loi stable dans \mathbb{R}^d , on a $\rho = 2\alpha$ d'après le résultat de Fristedt (2.1.41). Les conditions du théorème sont satisfaites si $\alpha > 1$. En prenant $n = N^r$, on a $\mathbf{E} \left(\frac{1}{n} \sum_{i=1}^n q_{m,i}^t - \Gamma\left(1 - \frac{t}{\alpha}\right) \sigma(S)^{t/\alpha} \right) = o(n^{-1/2}) = o(N^{-r/2})$. Donc la vitesse de convergence d'estimateur dans \mathcal{L}_1 peut atteindre à $N^{1/4-\varepsilon/2}$ si $\alpha > 8/5$.

Démonstration: Notons $X_{m,i} = q_{m,i}^t$, $\mu_m = \mathbf{E}q_{m,i}^t$, $\sigma_m^2 = \text{Var}(X_{m,i})$ et $r_m = \mu_m - \Gamma(1 - t/\alpha)\sigma(S)^{t/\alpha}$. Ainsi la convergence (2.1.54) aura lieu si on a les trois relations suivantes :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_{m,i} - \mu_m) \Rightarrow \mathcal{N}(0, \sigma^2) \quad (2.1.55)$$

$$\frac{1}{n} \sum_{i=1}^n X_{m,i}^2 - \left(\frac{1}{n} \sum_{i=1}^n X_{m,i} \right)^2 \xrightarrow[N \rightarrow \infty]{P} \sigma^2. \quad (2.1.56)$$

et

$$\sqrt{nr_m} \rightarrow 0 \quad (2.1.57)$$

D'après le lemme 2.1.5., les moments $\{\mathbf{E}X_{m,i}^4\}$ sont uniformément bornés. On a donc la convergence suivante

$$\sigma_m^2 := \mathbf{E}X_{m,i}^2 - (\mathbf{E}X_{m,i})^2 \rightarrow \sigma^2 = \sigma(S)^{2t/\alpha} (\Gamma(1 - 2t/\alpha) - (\Gamma(1 - t/\alpha))^2), \quad (2.1.58)$$

et la condition de Lindeberg, c'est-à-dire pour tout $\varepsilon > 0$

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n\sigma_m^2} \sum_{i=1}^n \int_{\{|X_{m,i} - \mu_m| > \varepsilon \sqrt{n}\sigma_m\}} (X_{m,i} - \mu_m)^2 d\mathbf{P} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sigma_m^2} \int_{\{|X_{m,1} - \mu_m| > \varepsilon \sqrt{n}\sigma_m\}} (X_{m,1} - \mu_m)^2 d\mathbf{P} \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{\sigma_m^2} (\mathbf{E}(X_{m,1} - \mu_m)^4)^{1/2} (\mathbf{P}\{|X_{m,1} - \mu_m| > \varepsilon \sqrt{n}\sigma_m\})^{1/2} \\ &= 0. \end{aligned}$$

La convergence (2.1.55) découle du théorème central limite appliqué au schéma triangulaire $\{X_{m,i}, 1 \leq i \leq n\}$. D'après la proposition 2.1.1. et le lemme 2.1.5., pour obtenir les convergences suivantes

$$\frac{1}{n} \sum_{i=1}^n X_{m,i}^2 - \mathbf{E}X_{m,i}^2 \xrightarrow[N \rightarrow \infty]{p.s.} 0 \quad (2.1.59)$$

et

$$\frac{1}{n} \sum_{i=1}^n X_{m,i} - \mu_m \xrightarrow[N \rightarrow \infty]{p.s.} 0, \quad (2.1.60)$$

il suffit de prendre $0 < t < \frac{\alpha r}{4}$. On en déduit

$$\frac{1}{n} \sum_{i=1}^n X_{m,i}^2 - \left(\frac{1}{n} \sum_{i=1}^n X_{m,i} \right)^2 - (\mathbf{E}X_{m,i}^2 - \mu_m^2) \xrightarrow[N \rightarrow \infty]{p.s.} 0,$$

en considérant (2.1.58) on obtient (2.1.56). Maintenant il reste à vérifier (2.1.57).

Lemme 2.1.13. *Si ξ satisfait la condition (2.1.40) avec $\rho > \alpha + 1$ et $0 < t < 1$, alors*

$$|r_m| \leq Cm^{-\zeta} \quad (2.1.61)$$

où $\zeta = \min(\frac{1}{2}, \frac{t+\rho-\alpha-1}{\alpha})$ et C ne dépend que de c_1, c_2, α, ρ et t .

Démonstration: D'après la relation (2.1.40) si x est suffisamment grand on a

$$G(x) = c_1 x^{-\alpha} + c_2 x^{-\rho} + o(x^{-\rho}),$$

où $c_1 = \sigma(S)$. Il est possible d'écrire la fonction réciproque quand x est assez petit,

$$G^{-1}(x) = \sigma(S)^{1/\alpha} x^{-1/\alpha} + b x^s + O(x^{s+(\rho-\alpha)/\alpha}) \quad (2.1.62)$$

avec $b = \alpha^{-1} c_2 \sigma(S)^{(1-\rho)/\alpha}$ et $s = (\rho - \alpha - 1)/\alpha$. On en déduit que pour $\delta > 0$ petit et $0 < x < \delta$

$$G^{-1}(x) - \sigma(S)^{1/\alpha} x^{-1/\alpha} = x^s (b + O(x^{(\rho-\alpha)/\alpha})).$$

Choisissons δ tel que $|O(x^{(\rho-\alpha)/\alpha})| \leq |b|$ (cela nous donne $\delta = o(|b|^{\frac{\alpha}{\rho-\alpha}})$), alors on peut écrire

$$|G^{-1}(x) - \sigma(S)^{1/\alpha} x^{-1/\alpha}| \leq 2|b|x^s. \quad (2.1.63)$$

Notons

$$R_+^{m+1} = \{\bar{x} = (x_1, \dots, x_{m+1}) : x_i \geq 0, i = 1, \dots, m+1\},$$

$$\Sigma_m = x_1 + \dots + x_m,$$

$$A = \left\{ \bar{x} \in R_+^{m+1} : \frac{x_1}{\Sigma_{m+1}} \geq \delta \right\}, A^c = R_+^{m+1} \setminus A,$$

où δ est choisi pour que (2.1.63) ait lieu. D'après sa définition et la relation (2.1.14) la variable aléatoire $q_{m,i} = M_{m,i}^{(1)}/m^{1/\alpha}$ est de loi

$$G^{-1} \left(\frac{\Gamma_1}{\Gamma_{m+1}} \right) / m^{1/\alpha}$$

alors

$$|r_m| = \left| \mathbf{E} \left(G^{-1} \left(\frac{\Gamma_1}{\Gamma_{m+1}} \right) \right)^t / m^{t/\alpha} - \mathbf{E} \left(\frac{\sigma(S)}{\Gamma_1} \right)^{t/\alpha} \right| \leq I_1 + I_2 + I_3 + I_4, \quad (2.1.64)$$

où

$$I_1 = \int_A m^{-t/\alpha} \left| \left(G^{-1} \left(\frac{x_1}{\Sigma_{m+1}} \right) \right)^t \right| \exp(-\Sigma_{m+1}) d\bar{x},$$

$$I_2 = \int_{A^c} m^{-t/\alpha} \left| \left(G^{-1} \left(\frac{x_1}{\Sigma_{m+1}} \right) \right)^t - \left(\frac{x_1}{\Sigma_{m+1}} \right)^{-t/\alpha} \sigma(S)^{t/\alpha} \right| \exp(-\Sigma_{m+1}) d\bar{x},$$

$$I_3 = \int_{A^c} m^{-t/\alpha} \sigma(S)^{t/\alpha} x_1^{-t/\alpha} \left| \Sigma_{m+1}^{t/\alpha} - m^{t/\alpha} \right| \exp(-\Sigma_{m+1}) d\bar{x},$$

et

$$I_4 = \int_A \sigma(S)^{t/\alpha} x_1^{-t/\alpha} \exp(-\Sigma_{m+1}) d\bar{x}.$$

Puisque dans A on a $G^{-1}(x_1/\Sigma_{m+1}) \leq G^{-1}(\delta)$, ainsi pour tout $\tau_1 > 0$ (qui sera choisi plus tard)

$$I_1 \leq m^{-t/\alpha} (G^{-1}(\delta))^t \mathbf{P} \left\{ \frac{\Gamma_1}{\Gamma_{m+1}} \geq \delta \right\} \leq m^{-t/\alpha} C_1 \mathbf{E} \left(\frac{\Gamma_1}{\Gamma_{m+1}} \right)^{\tau_1} \quad (2.1.65)$$

où $C_1 = (G^{-1}(\delta))^t \delta^{-\tau_1}$. De même manière, on a pour tout $\tau_2 > 0$,

$$I_4 \leq \sigma(S)^{t/\alpha} (\mathbf{E} \Gamma_1^{-2t/\alpha})^{1/2} \left(\mathbf{P} \left\{ \frac{\Gamma_1}{\Gamma_{m+1}} \geq \delta \right\} \right)^{1/2} \leq C_2 \mathbf{E} \left(\frac{\Gamma_1}{\Gamma_{m+1}} \right)^{\tau_2} \quad (2.1.66)$$

où $C_2 = \sigma(S)^{t/\alpha} (\Gamma(1 - 2t/\alpha))^{1/2} \delta^{-\tau_2}$. La fonction $f(x) = x^t, 0 < t < 1$ est Lipschitz continue sur $[0, \delta]$. En considérant (2.1.63) on a

$$\begin{aligned} I_2 &\leq \int_{A^c} m^{-t/\alpha} C_3 \left| G^{-1} \left(\frac{x_1}{\Sigma_{m+1}} \right) - \left(\frac{x_1}{\Sigma_{m+1}} \right)^{-1/\alpha} \sigma(S)^{1/\alpha} \right| \exp(-\Sigma_{m+1}) d\bar{x} \\ &\leq 2|b| C_3 m^{-t/\alpha} \int_{A^c} \left(\frac{x_1}{\Sigma_{m+1}} \right)^s \exp(-\Sigma_{m+1}) d\bar{x} \\ &\leq 2|b| C_3 m^{-t/\alpha} \mathbf{E} \left(\frac{\Gamma_1}{\Gamma_{m+1}} \right)^s \end{aligned} \quad (2.1.67)$$

où $C_3 = t(\min(G^{-1}(\delta), \delta^{-1/\alpha} \sigma(S)^{1/\alpha}))^{t-1}$. Il est connu (voir par exemple [Bre68]) que le v.a. m -dimensionnel $\left(\frac{\Gamma_1}{\Gamma_{m+1}}, \dots, \frac{\Gamma_m}{\Gamma_{m+1}} \right)$ a la même densité que le vecteur des statistiques d'ordre des variables aléatoires uniformes sur $[0, 1)$. En particulier la variable aléatoire $\frac{\Gamma_1}{\Gamma_{m+1}}$ a la densité suivante (on suppose que $m \geq 2$)

$$g(x) = \begin{cases} m(1-x)^{m-1}, & \text{si } 0 \leq x \leq 1, \\ 0, & \text{sinon.} \end{cases}$$

Donc

$$\mathbf{E} \left(\frac{\Gamma_1}{\Gamma_{m+1}} \right)^s = m \int_0^1 x^s (1-x)^{m-1} dx = m B(s+1, m) \leq 4\Gamma(s+1) m^{-s}. \quad (2.1.68)$$

Par l'inégalité de Cauchy Schwarz on a

$$I_3 \leq m^{-t/\alpha} \sigma(S)^{t/\alpha} (\mathbf{E} \Gamma_1^{-2t/\alpha})^{1/2} (\mathbf{E} (\Gamma_m^{t/\alpha} - m^{t/\alpha})^2)^{1/2}. \quad (2.1.69)$$

Il reste à évaluer

$$\mathbf{E} (\Gamma_m^{t/\alpha} - m^{t/\alpha})^2 = \frac{\Gamma(m+1+2t/\alpha)}{\Gamma(m+1)} - 2m^{t/\alpha} \frac{\Gamma(m+1+t/\alpha)}{\Gamma(m+1)} + m^{2t/\alpha}.$$

En utilisant l'approximation de la fonction gamma

$$\Gamma(x) = \sqrt{2\pi} x^{x-1/2} e^{-x} \left(1 + \frac{1}{12x} + \frac{1}{288x^2} + o\left(\frac{1}{x^2}\right) \right), \quad x \rightarrow \infty,$$

on obtient

$$\mathbf{E} (\Gamma_m^{t/\alpha} - m^{t/\alpha})^2 = A_m (m+1+2t/\alpha)^{2t/\alpha} H_1 - 2B_m m^{t/\alpha} (m+1+t/\alpha)^{t/\alpha} H_2 + m^{2t/\alpha}, \quad (2.1.70)$$

où

$$\begin{aligned} A_m &= \left(1 + \frac{2t/\alpha}{m+1}\right)^{m+1} \left(1 + \frac{2t/\alpha}{m+1}\right)^{-1/2} e^{-2t/\alpha}, \\ B_m &= \left(1 + \frac{t/\alpha}{m+1}\right)^{m+1} \left(1 + \frac{t/\alpha}{m+1}\right)^{-1/2} e^{-t/\alpha}, \\ H_1 &= 1 - \frac{t}{6\alpha m^2} + o\left(\frac{1}{m^2}\right) \quad \text{et} \quad H_2 = 1 - \frac{t}{12\alpha m^2} + o\left(\frac{1}{m^2}\right). \end{aligned} \quad (2.1.71)$$

Par un calcul simple on obtient les développements suivants

$$A_m = 1 - \left(\frac{t}{\alpha} + \frac{2t^2}{\alpha^2}\right) \frac{1}{m+1} + o\left(\frac{1}{m}\right), \quad (2.1.72)$$

$$B_m = 1 - \left(\frac{t}{2\alpha} + \frac{t^2}{2\alpha^2}\right) \frac{1}{m+1} + o\left(\frac{1}{m}\right). \quad (2.1.73)$$

En considérant (2.1.70) (2.1.71) (2.1.72) et (2.1.73) on a

$$\mathbf{E}(\Gamma_m^{t/\alpha} - m^{t/\alpha})^2 = \frac{t^2}{\alpha^2} m^{2t/\alpha-1} + o(m^{2t/\alpha-1}).$$

Ainsi on déduit de (2.1.69) qu'il existe une constante positive C_4 telle que

$$I_3 \leq C_4 m^{-1/2}. \quad (2.1.74)$$

En considérant les estimations (2.1.65)-(2.1.74) et en choisissant $\tau_1 = (\rho - \alpha - 1)/\alpha$ et $\tau_2 = (t + \rho - \alpha - 1)/\alpha$ dans (2.1.65) et (2.1.66), on obtient l'estimation (2.1.61) avec $C = C(\alpha, \rho, c_1, c_2, t)$. Le lemme est prouvé. \square

Ayant l'estimation de r_m , et en prenant $n \sim N^r$ avec $r < \frac{2\zeta}{1+2\zeta}$ on obtient $\sqrt{nr}r_m \rightarrow 0$. D'après un calcul simple on obtient les relations (2.1.52) et (2.1.53). \square

Maintenant nous considérons les propriétés asymptotiques de l'estimateur $\hat{\sigma}_N$ défini dans le Théorème 2.1.8.. On est toujours sous l'hypothèse que ξ appartient au domaine d'attraction normale d'une loi $\mathcal{St}\alpha\mathcal{S}$. On présente maintenant une relation plus forte : $\forall B \in \mathcal{B}(S)$ tel que $\sigma(\partial B) = 0$,

$$\mathbf{P} \left\{ \frac{\xi}{\|\xi\|} \in B, \|\xi\| > x \right\} = \sigma(B)x^{-\alpha} + Cx^{-\rho} + o(x^{-\rho}) \text{ quand } x \rightarrow \infty, \quad (2.1.75)$$

où $\rho > \alpha > 0$. Remarquons que cette condition implique les conditions (1.1.6) et (2.1.40).

Notons $\sigma(B) = b$, $\mathbb{I}_B(\theta_{m,i}) = \eta_{m,i}$, $i = 1, 2, \dots, n$. Alors

$$\begin{aligned} Z_n &= \hat{\sigma}_N(B) - \sigma(B) = \frac{1}{n} \sum_{i=1}^n (\eta_{m,i} - \sigma(B)), \\ \sqrt{n}Z_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\eta_{m,i} - \sigma(B)) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\eta_{m,i} - \mathbf{E}\eta_{m,1}) + \sqrt{n}(\mathbf{E}\eta_{m,1} - \sigma(B)), \end{aligned}$$

$$U_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\eta_{m,i} - \mathbf{E}\eta_{m,1}), \quad r_m = \mathbf{E}\eta_{m,1} - \sigma(B).$$

On pose

$$T_N = \frac{\sqrt{n}Z_n}{\left(\frac{1}{n} \sum_{i=1}^n \eta_{m,i}^2 - \left(\frac{1}{n} \sum_{i=1}^n \eta_{m,i} \right)^2 \right)^{1/2}},$$

alors le résultat de propriété asymptotique de $\hat{\sigma}_N$ peut être écrit comme suit :

Théorème 2.1.14. *Si ξ satisfait la condition (2.1.75) et*

$$n = N^{2\zeta/(1+2\zeta)-\varepsilon}, \quad m = N^{1/(1+2\zeta)+\varepsilon},$$

où $\zeta = \min(\frac{\rho-\alpha}{\alpha}, 1)$ et ε est une constante strictement positive et arbitrairement petite, alors

$$T_N \Rightarrow \mathcal{N}(0, 1). \quad (2.1.76)$$

Remarque 14.

- 1) Si on écrit tout simplement $\sqrt{n}Z_n$ dans le côté gauche de (2.1.76), la variance de la loi limite devient $\sigma(B)(1 - \sigma(B))$ qui dépend d'un paramètre inconnu $\sigma(B)$.
- 2) Rappelons qu'il y a une expansion asymptotique (2.1.41) pour la distribution strictement α -stable dans \mathbb{R}^d , c'est-à-dire $\rho = 2\alpha$ [Fri72]. Dans ce cas on prend $n = N^{2/3-\varepsilon}$, $m = N^{1/3+\varepsilon}$, donc la vitesse de convergence de $\hat{\sigma}_N(B)$ dans \mathcal{L}_1 est proche de $N^{1/3}$.

Démonstration: Notons $c^2 = b(1 - b)$. Pour prouver (2.1.76) il suffit de montrer les relations suivantes :

$$U_n \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(0, c^2), \quad (2.1.77)$$

$$\sqrt{n}r_m \rightarrow 0, \quad (2.1.78)$$

$$\frac{1}{n} \sum_{i=1}^n \eta_{m,i}^2 - \left(\frac{1}{n} \sum_{i=1}^n \eta_{m,i} \right)^2 \xrightarrow[N \rightarrow \infty]{P} c^2. \quad (2.1.79)$$

Puisque $0 < \mathbb{I}_B(\theta_{m,i}) \leq 1$, les moments $\mathbf{E}|\mathbb{I}_B(\theta_{m,i}) - \mathbf{E}\mathbb{I}_B(\theta_{m,i})|^k$ sont uniformément bornés pour tout $k > 0$. De manière similaire à (2.1.48) on a la condition de Lindeberg pour le schéma triangulaire $\{\mathbb{I}_B(\theta_{m,i}) - \mathbf{E}\mathbb{I}_B(\theta_{m,i})\}$ dont la variance est

$$c_m^2 := \mathbf{E}(\eta_{m,1} - \mathbf{E}\eta_{m,1})^2 = \mathbf{E}\eta_{m,1}^2 - (\mathbf{E}\eta_{m,1})^2 = b + r_m - (b + r_m)^2 \rightarrow c^2,$$

si on a (2.1.78). La relation (2.1.77) découle du théorème central limite appliqué au schéma triangulaire $\{\mathbb{I}_B(\theta_{m,i}) - \mathbf{E}\mathbb{I}_B(\theta_{m,i})\}$.

Il est facile de voir qu'en appliquant le lemme 2.1.1. on a (2.1.79). Afin de prouver le théorème, il reste à établir la relation (2.1.78). En fait, on a $|r_m| \leq C \max(m^{-1}, m^{-(\rho-\alpha)/\alpha})$ d'après le lemme suivant. En prenant une constante positive arbitrairement petite ε et

$$n = N^{\frac{2\zeta}{1+2\zeta}-\varepsilon}, \quad m = N^{\frac{1}{1+2\zeta}+\varepsilon},$$

on obtient $\sqrt{nr_m} \rightarrow 0$, le théorème est prouvé. \square

Lemme 2.1.15. *Si la condition (2.1.75) est satisfaite, alors*

$$|r_m| \leq C \max(m^{-1}, m^{-(\rho-\alpha)/\alpha}). \quad (2.1.80)$$

Démonstration: Pour prouver (2.1.80), il suffit de montrer que

$$\mathbf{P}\{\theta_{m,i} \in B\} = \sigma(B) + R_m, \quad (2.1.81)$$

avec le résiduel $R_m = O(\max(m^{-1}, m^{-(\rho-\alpha)/\alpha}))$. Notons

$$G_m(x) = \mathbf{P} \left\{ \max_{1 \leq i \leq m-1} \|\xi_i\| \leq x \right\}.$$

En utilisant la définition de $\theta_{m,i}$, il n'est pas difficile de voir que

$$\mathbf{P}\{\theta_{m,i} \in B\} = m \int_0^\infty \mathbf{P} \left\{ \frac{\xi_1}{\|\xi_1\|} \in B, \|\xi_1\| > r \right\} G_m(dr).$$

Notons $\tilde{G}_m(x) = G_m(xm^{1/\alpha})$. L'hypothèse (2.1.75) implique (rappelons que $\sigma(S) = 1$) pour s assez grand,

$$\mathbf{P}\{\|\xi\| > s\} = s^{-\alpha} + Cs^{-\rho} + o(s^{-\rho}). \quad (2.1.82)$$

Ainsi, il est facile de voir la relation

$$\lim_{m \rightarrow \infty} \tilde{G}_m(x) = G_0(x) = \begin{cases} \exp(-x^{-\alpha}), & x > 0, \\ 0, & x \leq 0. \end{cases}$$

En utilisant (2.1.75) et le fait que $\int_0^\infty y^{-\alpha} dG_0(y) = 1$, on a (2.1.81) avec $R_m = \sum_{i=1}^4 R_{m,i}$, où

$$\begin{aligned} R_{m,1} &= m \int_0^s \mathbf{P} \left\{ \frac{\xi_1}{\|\xi_1\|} \in B, \|\xi_1\| > r \right\} dG_m(r), \\ R_{m,2} &= -\sigma(B) \int_0^{s'} y^{-\alpha} dG_0(y), \\ R_{m,3} &= \sigma(B) \int_{s'}^\infty y^{-\alpha} d(\tilde{G}_m(y) - G_0(y)), \\ R_{m,4} &= Cm^{-(\rho-\alpha)/\alpha} \int_{s'}^\infty y^{-\rho} d\tilde{G}_m(y). \end{aligned}$$

Ici $s' = sm^{-1/\alpha}$ et nous allons choisir s plus tard. Il est facile de voir que

$$R_{m,1} \leq m(1 - \mathbf{P}\{\|\xi_1\| > s\})^m = m(1 - hm^{-1})^m \leq me^{-\frac{1}{2}h},$$

où $h = h(m, s) = m\mathbf{P}\{\|\xi_1\| > s\} \geq \frac{1}{2}ms^{-\alpha}$. Nous avons utilisé (2.1.82) pour la dernière inégalité. Alors, si on choisit

$$s = \left(\frac{m}{K \ln m}\right)^{1/\alpha},$$

avec K suffisamment grand, on obtient

$$R_{m,1} = o(m^{-1}). \quad (2.1.83)$$

Un calcul simple nous donne

$$R_{m,2} = o(m^{-1}). \quad (2.1.84)$$

Le terme résiduel principal est $R_{m,3}$. Pour estimer ce terme on doit estimer d'abord la différence $\tilde{G}_m(y) - G_0(y)$. Un développement assez simple de la fonction logarithmique donne une estimation suffisante pour nous.

Lemme 2.1.16. *Soient $\xi_i, i \geq 1$ les e.a. i.i.d. satisfaisant la condition (2.1.82). Alors pour $y > cm^{-1/\alpha}$*

$$|\tilde{G}_m(y) - G_0(y)| \leq C(\alpha, \rho) \exp(-y^{-\alpha})(m^{-(\rho-\alpha)/\alpha}y^{-\rho} + m^{-1}y^{-2\alpha}) \quad (2.1.85)$$

et

$$\sup_y |\tilde{G}_m(y) - G_0(y)| = C(\alpha, \rho) \max(m^{-1}, m^{-(\rho-\alpha)/\alpha}). \quad (2.1.86)$$

Démonstration: Pour prouver (2.1.85) on écrit

$$\ln \tilde{G}_m(y) = m \ln(1 - \mathbf{P}\{\|\xi_1\| > m^{1/\alpha}y\}),$$

puis on utilise (2.1.82) et l'expansion de la fonction logarithmique. Après un calcul assez simple, on obtient (2.1.85). Puisque pour $0 < y < cm^{-1/\alpha}$, les termes $\tilde{G}_m(y)$ et $G_0(y)$ ont de l'ordre $o(m^{-1})$, la relation (2.1.86) vient de (2.1.85). Le lemme est prouvé. \square

Maintenant on peut estimer le terme $R_{m,3}$. En intégrant par partie on obtient

$$|R_{m,3}| = \sigma(B)(R_{m,3}^{(1)} + R_{m,3}^{(2)}), \quad (2.1.87)$$

où

$$R_{m,3}^{(1)} = s'^{-\alpha} |\tilde{G}_m(s') - G_0(s')|,$$

$$R_{m,3}^{(2)} = \alpha \int_{s'}^{\infty} |\tilde{G}_m(y) - G_0(y)| y^{-\alpha-1} dy.$$

Comme $s' = (K \ln m)^{-1/\alpha} > Cm^{-1/\alpha}$, on peut utiliser (2.1.85) pour estimer les deux quantités $R_{m,3}^{(i)}, i = 1, 2$. Par quelques calculs simples, on a

$$R_{m,3}^{(1)} = o(m^{-1}),$$

$$R_{m,3}^{(2)} \leq C(\alpha, \rho) \max(m^{-1}, m^{-(\rho-\alpha)/\alpha}).$$

De façon similaire on estime $R_{m,4}$:

$$R_{m,4} = C m^{-(\rho-\alpha)/\alpha} \int_{s'}^{\infty} y^{-\rho} d\tilde{G}_m(y) = C m^{-(\rho-\alpha)/\alpha} (R_{m,4}^{(1)} + R_{m,4}^{(2)}), \quad (2.1.88)$$

où

$$R_{m,4}^{(1)} = \int_{s'}^{\infty} y^{-\rho} dG_0(y),$$

$$R_{m,4}^{(2)} = \int_s^{\infty} y^{-\rho} d(\tilde{G}_m(y) - G_0(y)).$$

Il est facile de voir que

$$R_{m,4}^{(1)} \leq C(\alpha, \rho) \quad (2.1.89)$$

et $R_{m,4}^{(2)}$ peut être estimé de façon similaire à $R_{m,3}$:

$$R_{m,4}^{(2)} \leq C(\alpha, \rho) \max(m^{-1}, m^{-(\rho-\alpha)/\alpha}). \quad (2.1.90)$$

En collectionnant les estimations (2.1.83), (2.1.84), et (2.1.87)-(2.1.90) on obtient (2.1.80). \square

Un résumé des résultats principaux des estimateurs $\hat{\alpha}_N \widehat{\sigma(S)}_N$ et $\hat{\sigma}_N$ peut être trouvé dans l'annexe A.

2.2 Estimation de la densité de mesure spectrale

Dans cette section nous supposons que la mesure spectrale σ est normalisée, i.e. $\sigma(S) = 1$, et absolument continue par rapport à la mesure de Lebesgue λ avec la densité p . On considère le problème d'estimation de la densité de la mesure spectrale à partir de l'échantillon ξ_1, \dots, ξ_N .

Rappelons d'abord les propriétés d'estimateur $\hat{\sigma}_n$ présentées dans la sous-section 2.1.2 (voir aussi l'annexe A) :

(Lemme 2.1.7., Corollaire 2.1.9.) Si ξ satisfait la condition de régularité alors pour chaque i on a

$$\theta_{mi} \Rightarrow \sigma, \quad m \rightarrow \infty \quad (2.2.1)$$

et

$$\hat{\sigma}_n \xrightarrow{p.s.} \sigma, \quad n \rightarrow \infty. \quad (2.2.2)$$

L'estimateur naturel de p est

$$\hat{p}_{k,n}(x) = \sum_{i=1}^k \frac{\hat{\sigma}_n(A_{k,i})}{\lambda(A_{k,i})} \mathbb{I}_{A_{k,i}}(x) \quad (2.2.3)$$

où $\mathcal{A}_k = \{A_{k,i}, i = 1, \dots, k\}$ est une partition de la sphère unité telle que $\sigma(\partial A_{k,i}) = 0$ pour tout i et le diamètre maximal de $A_{k,i}$, noté d_k , tend vers 0 quand $k \rightarrow \infty$.

Lemme 2.2.1. Soit p la densité de loi σ , $\hat{p}_{k,n}$ l'estimateur de p défini par (2.2.3). Alors pour x un point de continuité de p on a

$$\lim_{k \rightarrow \infty} \lim_{n \rightarrow \infty} |\hat{p}_{k,n}(x) - p(x)| = 0, \quad p.s.. \quad (2.2.4)$$

Démonstration: Notons $A_{k,i(x)}$ l'ensemble dans la partition \mathcal{A}_k et $i(x)$ l'indice i tel que $x \in A_{k,i(x)}$. On a

$$\begin{aligned} |\hat{p}_{k,n}(x) - p(x)| &= \frac{1}{\lambda(A_{k,i(x)})} |\hat{\sigma}_n(A_{k,i(x)}) - \lambda(A_{k,i(x)})p(x)| \\ &\leq \frac{1}{\lambda(A_{k,i(x)})} |\hat{\sigma}_n(A_{k,i(x)}) - \sigma(A_{k,i(x)})| \\ &\quad + \left| \frac{\sigma(A_{k,i(x)})}{\lambda(A_{k,i(x)})} - p(x) \right| \\ &= I_1 + I_2. \end{aligned} \quad (2.2.5)$$

D'après (2.2.2) pour k fixé, quel que soit $\varepsilon > 0$, $\exists n(k)$ tel que pour tout $n > n(k)$

$$I_1 = \frac{1}{\lambda(A_{k,i(x)})} |\hat{\sigma}_n(A_{k,i(x)}) - \sigma(A_{k,i(x)})| < \varepsilon, \quad p.s.. \quad (2.2.6)$$

Puisque x est un point de continuité de p , si k est suffisamment grand on a

$$I_2 < \varepsilon. \quad (2.2.7)$$

Les inégalités (2.2.5), (2.2.6) et (2.2.7) impliquent (2.2.4). \square

Remarque 15. L'ordre de deux limites dans la convergence (2.2.4) est important. Considérons pour n fixé

$$\hat{p}_{k,n}(x) = \frac{\hat{\sigma}_n(A_{k,i(x)})}{\lambda(A_{k,i(x)})} = \frac{\sum_{j=1}^n \mathbb{1}_{A_{k,i(x)}} \theta_{mj}}{n\lambda(A_{k,i(x)})}.$$

La variable aléatoire $X = \sum_{j=1}^n \mathbb{1}_{A_{k,i(x)}} \theta_{mj}$ est de loi binomiale de paramètres n et $\sigma_m(A_{k,i(x)})$ où σ_m représente la loi de θ_{mi} . Quel que soit $\varepsilon > 0$ on peut trouver une suite des partitions \mathcal{A}_k telle que pour k suffisamment grand $n\lambda(A_{k,i(x)})\varepsilon < 1$ et $\sigma_m(A_{k,i(x)}^{\mathbf{G}}) > (1 - k^{-1-\varepsilon})^{1/n}$. Ainsi

$$\mathbf{P}\{\hat{p}_{k,n}(x) > \varepsilon\} = \mathbf{P}\{X > n\lambda(A_{k,i(x)})\varepsilon\} = 1 - (1 - \sigma_m(A_{k,i(x)}))^n \leq k^{-1-\varepsilon}.$$

On en déduit pour n fixé

$$\lim_{k \rightarrow \infty} \hat{p}_{k,n}(x) = 0, \quad p.s., \quad (2.2.8)$$

donc

$$\lim_{n \rightarrow \infty} \lim_{k \rightarrow \infty} \hat{p}_{k,n}(x) = 0, \quad p.s..$$

La convergence (2.2.8) signifie que pour le nombre des points du support de $\hat{\sigma}_n$ fixé, la densité estimée tend vers zéro presque partout quand on augmente le nombre des ensembles de la partition. En effet cet estimateur est l'histogramme des points du support de $\hat{\sigma}_n$.

Avant de présenter l'estimation de la densité de σ par la méthode de noyau, on discute d'abord les propriétés de la densité des v.a. i.i.d. θ_{mi} .

2.2.1 Convergence de la densité de θ_{mi}

Proposition 2.2.1. *Soient ξ_1, \dots, ξ_N des v.a. i.i.d. dans \mathbb{R}^d . Si la loi des ξ_i est absolument continue par rapport à la mesure de Lebesgue λ , alors la loi des v.a. i.i.d. θ_{mi} est absolument continue par rapport à λ .*

Démonstration: Soient ξ_1, \dots, ξ_m des v.a. i.i.d. dans le premier paquet G_{m1} de la partition de l'échantillon. Notons $\mathbb{R}^{d \times m} = \mathbb{R}^d \times \dots \times \mathbb{R}^d$, le produit de m fois l'espace \mathbb{R}^d , Π_m l'ensemble des permutations sur $\{1, \dots, m\}$. Le v.a. θ_{m1} peut être exprimé par

$$\theta_{m1} = \frac{\xi_{\pi(1)}}{\|\xi_{\pi(1)}\|}, \quad (2.2.9)$$

où π est l'élément dans Π_m tel que $\|\xi_{\pi(1)}\| \geq \|\xi_{\pi(2)}\| \geq \dots \geq \|\xi_{\pi(m)}\|$.

L'élément dans $\mathbb{R}^{d \times m}$ est noté par $\mathbf{x} = (x_1, \dots, x_m)$ avec $x_i \in \mathbb{R}^d$. Définissons les ensembles

$$E = \{\mathbf{x} \mid \|x_1\| > \|x_2\| > \dots > \|x_m\|, \mathbf{x} \in \mathbb{R}^{d \times m}\},$$

$$E_\pi = \{\mathbf{x} \mid \|x_{\pi(1)}\| > \|x_{\pi(2)}\| > \dots > \|x_{\pi(m)}\|, \mathbf{x} \in \mathbb{R}^{d \times m}\}, \quad \pi \in \Pi_m.$$

L'ensemble E_π est un cône, c'est-à-dire $\forall a > 0, \forall \mathbf{x} \in E_\pi$, on a $a\mathbf{x} \in E_\pi$. Il est clair que $\text{Card}(\Pi_m) = m!$ et

$$\mathbb{R}^{d \times m} = \left(\bigcup_{\pi \in \Pi_m} E_\pi \right) \cup \Delta$$

où Δ est un ensemble de mesure de Lebesgue nulle. Définissons les applications

$$\begin{aligned} f_\pi : E_\pi &\rightarrow E \\ (x_1, \dots, x_m) &\mapsto (x_{\pi(1)}, \dots, x_{\pi(m)}), \\ g : \mathbb{R}^{d \times m} &\rightarrow \mathbb{R}^d \\ (x_1, \dots, x_m) &\mapsto x_1, \end{aligned}$$

et

$$\begin{aligned} h : \mathbb{R}^d &\rightarrow S^{d-1} \\ x &\mapsto \frac{x}{\|x\|}. \end{aligned}$$

Par l'indépendance, le m -uplet (ξ_1, \dots, ξ_m) a la densité p^m par rapport à λ^m dans l'espace $\mathbb{R}^{d \times m}$. Donc $(\xi_1, \dots, \xi_m) \in \bigcup_{\pi \in \Pi_m} E_\pi$, λ^m -p.p.. D'après la définition de θ_{m1} (2.2.9) on a

$$\theta_{m1} = h \circ g \circ f_\pi(\xi_1, \dots, \xi_m), \quad \lambda^m - \text{p.p.} \quad (2.2.10)$$

Si $\mathbf{x}, \mathbf{y} \in E_\pi$ et $a\mathbf{x} + b\mathbf{y} \in E_\pi$ avec a, b les nombres réels, alors $f_\pi(a\mathbf{x} + b\mathbf{y}) = af_\pi(\mathbf{x}) + bf_\pi(\mathbf{y})$. Cela signifie que l'application f_π est linéaire dans E_π . On en déduit que la loi de $f_\pi(\xi_1, \dots, \xi_m)$ est absolument continue par rapport à λ de densité notée q_m . Ainsi la densité du v.a. $g \circ f_\pi(\xi_1, \dots, \xi_m)$ est la densité marginale de q_m , notée q .

Notons $\mathbf{x} = (\theta, r)$ la décomposition polaire d'un vecteur dans \mathbb{R}^d . Soit X le v.a. dans \mathbb{R}^d de densité q . En notant $Y = \frac{X}{\|X\|}$, on a pour toute fonction f positive définie sur S^{d-1}

$$\begin{aligned} \mathbf{E}f(Y) &= \int_{\mathbb{R}^d} f\left(\frac{x}{\|x\|}\right) q(x) dx \\ &= \int_{S^{d-1}} \int_0^\infty f(\theta) r^{d-1} q(\theta, r) dr d\theta \\ &= \int_{S^{d-1}} f(\theta) \int_0^\infty r^{d-1} q(\theta, r) dr d\theta. \end{aligned} \quad (2.2.11)$$

Les égalités (2.2.10), (2.2.11) et la linéarité de f_π impliquent l'existence de la densité de θ_{m1} . \square

Notons la loi de θ_{m1} par σ_m , la densité de θ_{m1} par p_m . On va montrer que sous la condition **(FVR)** (voir (2.2.14)) on a $p_m \xrightarrow{\mathcal{L}_1} p$ quand $m \rightarrow \infty$.

On associe aux v.a. ξ_1, \dots, ξ_m le processus empirique

$$\beta_m = \sum_{i=1}^m \delta_{\xi_i/b_m},$$

où $\{b_m\}$ est une suite de constantes de normalisation telle que $b_m \rightarrow \infty$. Rappelons que la condition de variation régulière (1.1.6) est équivalente à la convergence

$$\beta_m \Rightarrow \pi_{\alpha, \sigma}, \quad m \rightarrow \infty,$$

où $\pi_{\alpha, \sigma}$ est le processus ponctuel poissonnien de mesure de contrôle

$$m_\alpha = \mu_\alpha \times \sigma,$$

où μ_α est une mesure σ -finie sur $(0, \infty)$ de densité $d\mu_\alpha(x)/d\lambda(x) = \alpha x^{-1-\alpha}$. Notons F la loi de ξ_i . En notant

$$\mu_m(A) = mF(b_m A), \quad A \in \mathcal{B}(\mathbb{R}^d \setminus \{0\}), \quad (2.2.12)$$

la condition (1.1.6) peut être écrite sous la forme suivante

$$\mu_m((r, \infty) \times B) \rightarrow m_\alpha((r, \infty) \times B), \quad m \rightarrow \infty, \quad (2.2.13)$$

pour tout $B \in \mathcal{B}(S^{d-1})$ avec $\sigma(\partial B) = 0$ et $\forall r > 0$.

Soit μ une mesure définie sur \mathbb{R}^d , A l'ensemble borélien dans \mathbb{R}^d . La notation $\mu|_A$ représente la restriction de la mesure μ sur A . On présente une condition appelée *la condition forte de variation régulière (FVR)* :

Pour la mesure μ_m définie par (2.2.12), la convergence en variation a lieu sur $B(0, \varepsilon)^c$, si $\forall \varepsilon > 0$

$$\mu_m|_{B(0, \varepsilon)^c} \xrightarrow{\text{var}} m_\alpha|_{B(0, \varepsilon)^c}, \quad (2.2.14)$$

où $B(0, \varepsilon)$ est la boule ouverte de centre 0 et de rayon ε dans \mathbb{R}^d .

Remarque 16. La condition (2.2.14) entraîne la condition (2.2.13). En effet (2.2.14) implique une convergence plus faible

$$mF(b_m A) \rightarrow m_\alpha(A), \quad A \in \mathcal{B}(B(0, \varepsilon))^{\mathbb{G}}.$$

Remarque 17. Il est bien connu, voir [Rei93], que la condition (2.2.14) est équivalente à la convergence forte des lois des restrictions des processus β_m : pour chaque $D \subset B(0, r)^{\mathbb{G}}$

$$\mathcal{L}_{\beta_m|D} \xrightarrow{var} \mathcal{L}_{\Pi_\alpha|D}. \quad (2.2.15)$$

Théorème 2.2.2. Soit ξ_1, \dots, ξ_m des v.a. i.i.d. dans \mathbb{R}^d de loi F . Si F vérifie la condition (FVR), alors

$$p_m \xrightarrow{\mathcal{L}_1} p, \quad m \rightarrow \infty, \quad (2.2.16)$$

où p est la densité de la mesure spectrale σ , p_m est la densité de θ_{m1} .

Démonstration: Notons η un v.a. dans S^{d-1} de loi σ . Pour montrer (2.2.16) il suffit de montrer

$$\|\mathcal{L}_{\theta_{m1}} - \mathcal{L}_\eta\|_{var} \xrightarrow{n \rightarrow \infty} 0. \quad (2.2.17)$$

Rappelons que le processus ponctuel poissonnien $\pi_{\alpha, \sigma}$ peut être représenté en loi comme suit :

$$\pi_{\alpha, \sigma} \stackrel{\mathcal{L}}{=} \sum_{i=1}^{\infty} \delta_{\Gamma_i^{-1/\alpha} \epsilon_i},$$

où ϵ_i sont des v.a. i.i.d. sur S^{d-1} de loi σ et $\Gamma_i = \sum_{j=1}^i \lambda_j$ où λ_j sont des variables i.i.d. de loi exponentielle standard, les suites de v.a. $\{\epsilon_i\}$ et $\{\lambda_j\}$ sont indépendantes.

Notons Y_1 et Z_{m1} les éléments dont la norme est la plus grande dans la configuration de $\pi_{\alpha, \sigma}$ et de β_m , i.e.

$$Y_1 = \Gamma_1^{-1/\alpha} \epsilon_1, \quad Z_{m1} = \xi_{m1}/b_m.$$

Pour tout $\varepsilon > 0$ notons les ensembles

$$D_\varepsilon = \{\Gamma_1^{-1/\alpha} \epsilon_1, \Gamma_2^{-1/\alpha} \epsilon_2, \dots\} \cap B(0, \varepsilon)^{\mathbb{G}},$$

$$D_{\varepsilon, m} = \{\xi_{m1}/b_m, \dots, \xi_{mm}/b_m\} \cap B(0, \varepsilon)^{\mathbb{G}}.$$

Les éléments dont la norme est la plus grande dans les restrictions des processus $\pi_{\alpha, \sigma}$ et β_m sont notés respectivement par

$$Y_1^\varepsilon = \left\{ Y \mid \|Y\| = \max_{Y_i \in D_\varepsilon} (\|Y_i\|), Y \in D_\varepsilon \right\},$$

$$Z_{m1}^\varepsilon = \left\{ Z \mid \|Z\| = \max_{Z_i \in D_{\varepsilon, m}} (\|Z_i\|), Z \in D_{\varepsilon, m} \right\}.$$

Ainsi

$$\|\mathcal{L}_{Y_1^\varepsilon} - \mathcal{L}_{Y_1}\|_{var} \leq 2\mathbf{P}\{Y_1 \neq Y_1^\varepsilon\} = 2\mathbf{P}\{\Gamma_1^{-1/\alpha} \leq \varepsilon\} = 2\exp(-\varepsilon^{-\alpha}). \quad (2.2.18)$$

D'après le lemme 2.1.3. on a

$$\|\mathcal{L}_{Z_{m1}^\varepsilon} - \mathcal{L}_{Z_{m1}}\|_{var} \leq 2\mathbf{P}\{\|Z_{m1}\| \leq \varepsilon\} \xrightarrow{m \rightarrow \infty} 2\mathbf{P}\{\|Y_1\| \leq \varepsilon\} = 2\exp(-\varepsilon^{-\alpha}). \quad (2.2.19)$$

La condition de **(FVR)** donne la convergence

$$\|\mathcal{L}_{Z_{m1}^\varepsilon} - \mathcal{L}_{Y_1^\varepsilon}\|_{var} \xrightarrow{m \rightarrow \infty} 0. \quad (2.2.20)$$

En considérant les inégalités (2.2.18)-(2.2.20) on a

$$\|\mathcal{L}_{Y_1} - \mathcal{L}_{Z_{m1}}\|_{var} \leq \|\mathcal{L}_{Y_1^\varepsilon} - \mathcal{L}_{Y_1}\|_{var} + \|\mathcal{L}_{Z_{m1}^\varepsilon} - \mathcal{L}_{Z_{m1}}\|_{var} + \|\mathcal{L}_{Z_{m1}^\varepsilon} - \mathcal{L}_{Y_1^\varepsilon}\|_{var},$$

d'où

$$\|\mathcal{L}_{Y_1} - \mathcal{L}_{Z_{m1}}\|_{var} \xrightarrow{n \rightarrow \infty} 0,$$

ce qui implique (2.2.17). \square

2.2.2 Estimation de la densité de σ par la méthode de noyau

Il y a beaucoup d'études concernant l'estimation non-paramétrique de la fonction de densité de probabilité d'un v.a. directionnel par la méthode de noyau. On considère l'estimateur de noyau proposé dans [BRRZ88],

$$\hat{p}_n(x; y_1, \dots, y_n) = \frac{C(h_n)}{nh_n^{d-1}} \sum_{i=1}^n K\left(\frac{1 - x^T y_i}{h_n^2}\right), \quad x, y_i \in S^{d-1}, \quad (2.2.21)$$

où $h_n > 0$, $C(h_n) > 0$, $K(\cdot)$ est une fonction non-négative définie sur $[0, \infty)$ telle que

$$\frac{h_n^{d-1}}{C(h_n)} = \int_{S^{d-1}} K\left(\frac{1 - x^T y}{h_n^2}\right) \lambda(dy). \quad (2.2.22)$$

Ici cette intégrale est indépendante de x . En utilisant le résultat (2.2.2) dans [Wat83] page 11, l'intégrale (2.2.22) peut être écrite comme

$$(C(h_n))^{-1} = \frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^{\frac{2}{h_n^2}} K(v) v^{\frac{d-3}{2}} (2 - vh_n^2)^{\frac{d-3}{2}} dv.$$

Si $h_n \xrightarrow{n \rightarrow \infty} 0$ et

$$(C1) \quad 0 < \int_0^\infty K(v) v^{\frac{d-3}{2}} dv < \infty$$

alors d'après le théorème de convergence dominée

$$\lim_{n \rightarrow \infty} (C(h_n))^{-1} = \frac{(2\pi)^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^\infty K(v) v^{\frac{d-3}{2}} dv. \quad (2.2.23)$$

Quelques exemples de densité de noyau sont

$$\begin{aligned} K(v) &= e^{-v} && \text{(loi de Von Mises),} \\ K(v) &= \begin{cases} 1 & \text{si } v < 1 \\ 0 & \text{sinon} \end{cases} && \text{(loi uniforme).} \end{aligned}$$

Soit η_1, \dots, η_n des v.a. i.i.d. de loi σ . Notons l'estimateur défini par (2.2.21) avec $y_i = \eta_i$ par

$$\hat{p}_n(x; \eta). \quad (2.2.24)$$

Alors $\hat{p}_n(x; \eta)$ est un estimateur de la densité de σ . Nous résumons des résultats connus sur l'estimateur $\hat{p}_n(x; \eta)$ présentés dans [BRRZ88]. Nous commençons par énoncer des conditions sur $K(\cdot)$ et $\{h_n\}$,

- (C2) K est bornée sur $[0, \infty)$,
- (C3) $\lim_{v \rightarrow \infty} K(v)v^{\frac{d-1}{2}} = 0$,
- (C4) $\lim_{n \rightarrow \infty} h_n = 0$,
- (C5) $nh_n^{d-1} / \log n \xrightarrow{n \rightarrow \infty} \infty$,
- (C6) $nh_n^{d-1} \xrightarrow{n \rightarrow \infty} \infty$.

Les propriétés de $\hat{p}_n(x; \eta)$ sont les suivantes :

- (Lemma 5. [BRRZ88]) Si les conditions (C1)-(C4) sont remplies alors pour x un point de continuité de p

$$|\mathbf{E}\hat{p}_n(x; \eta) - p(x)| \xrightarrow{n \rightarrow \infty} 0. \quad (2.2.25)$$

- (Theorem 1. [BRRZ88]) Si les conditions (C1)-(C5) sont remplies alors pour x un point de continuité de p

$$\hat{p}_n(x; \eta) \xrightarrow[n \rightarrow \infty]{p.s.} p(x). \quad (2.2.26)$$

- (Theorem 2. [BRRZ88]) Si les conditions (C1), (C4) et (C5) sont remplies, p est continue sur S^{d-1} et K est intégrable au sens de Riemann sur tout intervalle fini dans $[0, \infty)$ avec

$$\int_0^\infty \sup\{K(u) \mid |\sqrt{u} - \sqrt{v}| < 1\} v^{\frac{d-3}{2}} dv < \infty, \quad (2.2.27)$$

alors

$$\sup_x |\hat{p}_n(x; \eta) - p(x)| \xrightarrow[n \rightarrow \infty]{p.s.} 0. \quad (2.2.28)$$

- (Theorem 3. [BRRZ88]) Si les conditions (C2), (C4) et (C6) sont remplies alors

$$\int_{S^{d-1}} |\hat{p}_n(x; \eta) - p(x)| \lambda(dx) \xrightarrow[n \rightarrow \infty]{p.s.} 0. \quad (2.2.29)$$

La difficulté dans notre cas est que l'échantillon $\theta_{m1}, \dots, \theta_{mn}$ sur lequel on estime la densité de σ n'est pas de loi σ . Les convergences de la loi et de la densité de θ_{mi} sont données

par (2.2.1) et (2.2.16). Dans cette sous-section on discute les propriétés de l'estimation de la densité de σ à partir des v.a. θ_{mi} , $i = 1, \dots, n$.

Notons

$$\hat{p}_n(x; \theta), \quad (2.2.30)$$

l'estimateur de p défini par (2.2.21) avec $y_i = \theta_{mi}$. Pour simplifier la notation, on utilise θ_{ni} au lieu de θ_{mi} car m dépend de n . Pour montrer les consistances de l'estimateur on a besoin de quelques lemmes.

Lemme 2.2.3. *Soient ξ_1, \dots, ξ_n des variables aléatoires indépendantes telles que $\mathbf{E}(\xi_i) = 0$ et $\text{Var}(\xi_i) = \sigma_i^2$, $i = 1, \dots, n$. Si il existe une constante positive b telle que $\mathbf{P}\{|\xi_i| \leq b\} = 1$, $i = 1, \dots, n$, alors quel que soit $\varepsilon > 0$ et pour tout n , on a*

$$\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^n \xi_i\right| \geq \varepsilon\right\} \leq 2 \exp\left(-\frac{n\varepsilon^2}{2\sigma^2 + b\varepsilon}\right), \quad (2.2.31)$$

où $\sigma^2 = n^{-1}(\sigma_1^2 + \dots + \sigma_n^2)$.

Pour la démonstration voir [Hoe63]. Le lemme suivant est une généralisation du Lemma 3. dans [Dev83].

Lemme 2.2.4. *(L'inégalité de loi multinomiale). Soit (X_1, \dots, X_k) un v.a. de loi multinomiale $(n, p_{n1}, \dots, p_{nk})$. Pour tout $\varepsilon \in (0, 1)$ et pour tout k tel que $k/n \leq \varepsilon^2/20$, on a*

$$\mathbf{P}\left\{\sum_{i=1}^k |X_i - \mathbf{E}X_i| > n\varepsilon\right\} \leq 3 \exp\left(-\frac{n\varepsilon^2}{25}\right). \quad (2.2.32)$$

Démonstration: D'après le Lemme 3 dans [Dev83], si (X_1, \dots, X_k) est de loi multinomiale (n, p_1, \dots, p_k) , on a (2.2.32). Remarquons que le terme de droite de l'inégalité (2.2.32) ne dépend pas de p_i . Donc ce résultat peut être généralisé directement dans notre cas. \square

Soient ξ_1, \dots, ξ_N des v.a. i.i.d. suivant la loi F à queue régulière de mesure spectrale σ . Supposons que σ est absolument continue par rapport à la mesure de Lebesgue avec la densité p .

Lemme 2.2.5. *Si la loi F vérifie la condition (FVR) et que les conditions (C1)-(C4) sont remplies, alors pour x un point de continuité de p on a*

$$|\mathbf{E}\hat{p}_n(x; \theta) - p(x)| \xrightarrow[n \rightarrow \infty]{} 0, \quad (2.2.33)$$

où $\hat{p}_n(x; \theta)$ est l'estimateur de p défini par (2.2.30).

Si p est continue sur S^{d-1} , alors

$$\sup_x |\mathbf{E}\hat{p}_n(x; \theta) - p(x)| \xrightarrow[n \rightarrow \infty]{} 0. \quad (2.2.34)$$

Démonstration: En notant la loi de θ_{ni} par σ_n , et

$$K_n(x) = \frac{C(h_n)}{h_n^{d-1}} K(x),$$

on a

$$\begin{aligned} |\mathbf{E}\hat{p}_n(x; \theta) - p(x)| &\leq |\mathbf{E}\hat{p}_n(x; \theta) - \mathbf{E}\hat{p}_n(x; \eta)| + |\mathbf{E}\hat{p}_n(x; \eta) - p(x)| \\ &= R_1 + R_2. \end{aligned}$$

D'après (2.2.25), R_2 converge vers 0 quand $n \rightarrow \infty$. On ne considère que le terme R_1 .

D'après la condition **(C2)** $\exists M < \infty$ tel que $K(x) < M, \forall x \in [0, \infty)$. Pour chaque n définissons une suite de constantes croissante $\{a_i, i = 0, \dots, n\}$ par

$$a_0 = 0, a_{i+1} = a_i + \frac{M}{n}, i = 0, \dots, n-1.$$

En notant $I_{n,i} = K^{-1}\left(\frac{h_n^{d-1}}{C(h_n)}[a_i, a_{i+1}]\right), i = 0, \dots, n-1$ on construit la fonction en escaliers

$$K_n^*(x) = \sum_{i=0}^{n-1} a_i \mathbb{I}_{I_{n,i}}(x).$$

Alors on a $0 \leq K_n(x) - K_n^*(x) < M/n$. Considérons

$$\begin{aligned} R_1 &= \left| \int_{S^{d-1}} K_n \left(\frac{1-x^T y}{h_n^2} \right) \sigma_n(dy) - \int_{S^{d-1}} K_n \left(\frac{1-x^T y}{h_n^2} \right) \sigma(dy) \right| \\ &\leq \left| \int_{S^{d-1}} K_n \left(\frac{1-x^T y}{h_n^2} \right) - K_n^* \left(\frac{1-x^T y}{h_n^2} \right) \sigma_n(dy) \right| \\ &\quad + \left| \int_{S^{d-1}} K_n^* \left(\frac{1-x^T y}{h_n^2} \right) \sigma_n(dy) - \int_{S^{d-1}} K_n^* \left(\frac{1-x^T y}{h_n^2} \right) \sigma(dy) \right| \\ &\quad + \left| \int_{S^{d-1}} K_n^* \left(\frac{1-x^T y}{h_n^2} \right) - K_n \left(\frac{1-x^T y}{h_n^2} \right) \sigma(dy) \right| \\ &= R_{11} + R_{12} + R_{13}. \end{aligned}$$

Il est clair que $R_{11} < M/n$ et $R_{13} < M/n$. Notons

$$\Delta_{n,i}(x) = \left\{ y \mid \frac{1-x^T y}{h_n^2} \in I_{n,i}, y \in S^{d-1} \right\}.$$

On a

$$\begin{aligned} R_{12} &= \left| \sum_{i=0}^{n-1} a_i (\sigma_n(\Delta_{n,i}(x)) - \sigma(\Delta_{n,i}(x))) \right| \\ &\leq M \sum_{i=0}^{n-1} |\sigma_n(\Delta_{n,i}(x)) - \sigma(\Delta_{n,i}(x))| \\ &\leq M \|\sigma_n - \sigma\|_{var}. \end{aligned}$$

D'après le théorème 2.2.2., $\|\sigma_n - \sigma\|_{var} \rightarrow 0$, ce qui complète la démonstration de (2.2.33). La convergence (2.2.34) peut être prouvée de manière similaire si p est continue. \square

On considère d'abord la consistance forte de $\hat{p}_n(x; \theta)$ pour les points fixés.

Théorème 2.2.6. *Si la loi F vérifie la condition (FVR) et que les conditions (C1)-(C5) sont remplies, alors pour x un point de continuité de p*

$$\hat{p}_n(x; \theta) \xrightarrow[n \rightarrow \infty]{p.s.} p(x). \quad (2.2.35)$$

Démonstration: D'après le lemme 2.2.5. on a

$$\lim_{n \rightarrow \infty} \mathbf{E}\hat{p}_n(x; \theta) = p(x).$$

Alors pour avoir la convergence (2.2.35) il suffit de montrer que

$$|\hat{p}_n(x; \theta) - \mathbf{E}\hat{p}_n(x; \theta)| \xrightarrow[n \rightarrow \infty]{p.s.} 0. \quad (2.2.36)$$

D'après la condition (C2) $\exists M < \infty$ tel que $K(x) < M, \forall x \in [0, \infty)$. La loi de θ_{ni} est notée par σ_n . Notons

$$\xi_i = \frac{C(h_n)}{h_n^{d-1}} \left(K \left(\frac{1 - x^T \theta_{ni}}{h_n^2} \right) - \mathbf{E}K \left(\frac{1 - x^T \theta_{ni}}{h_n^2} \right) \right)$$

alors les variables aléatoires ξ_1, \dots, ξ_n sont i.i.d. et

$$\mathbf{E}\xi_i = 0, \quad |\xi_i| \leq 2h_n^{1-d}C(h_n)M,$$

$$\begin{aligned} \mathbf{E}\xi_i^2 &\leq \frac{C^2(h_n)}{h_n^{2(d-1)}} \int_{S^{d-1}} K^2 \left(\frac{1 - x^T y}{h_n^2} \right) \sigma_n(dy) \\ &\leq \frac{MC^2(h_n)}{h_n^{2(d-1)}} \int_{S^{d-1}} K \left(\frac{1 - x^T y}{h_n^2} \right) \sigma_n(dy). \end{aligned}$$

D'après (2.2.23) et le lemme 2.2.5., on peut trouver les constantes $a > 0$ et $a(x) > 0$ telles que

$$|\xi_i| \leq ah_n^{1-d}, \quad \mathbf{E}\xi_i^2 \leq a(x)h_n^{1-d}.$$

D'après le lemme 2.2.3.

$$\begin{aligned} \mathbf{P}\{|\hat{p}_n(x; \theta) - \mathbf{E}\hat{p}_n(x; \theta)| \geq \varepsilon\} &= \mathbf{P}\left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| \geq \varepsilon \right\} \\ &\leq 2 \exp \left(- \frac{n\varepsilon^2}{2a(x)h_n^{1-d} + ah_n^{1-d}\varepsilon} \right) \\ &= 2 \exp \left(- \frac{nh_n^{d-1}\varepsilon^2}{2a(x) + a\varepsilon} \right). \end{aligned}$$

D'après la condition (C5)

$$\sum_{n=1}^{\infty} \mathbf{P}\{|\hat{p}_n(x; \theta) - \mathbf{E}\hat{p}_n(x; \theta)| \geq \varepsilon\} < \infty,$$

ceci implique (2.2.36). \square

Ensuite nous établissons la consistance forte de $\hat{p}_n(x, \theta)$ dans la norme \mathcal{L}_1 sous certaines conditions.

Théorème 2.2.7. *Supposons que les conditions (C1), (C4) et (C6) sont remplies. Supposons aussi que les densités p_n de θ_{ni} sont équicontinues, c'est-à-dire*

$$\sup_n \omega_{p_n}(\delta) \rightarrow 0, \quad \delta \rightarrow 0, \quad (2.2.37)$$

$$\text{où } \omega_{p_n}(\delta) = \sup_{|x-y| \leq \delta} |p_n(x) - p_n(y)|.$$

Alors pour $\varepsilon > 0$ donné, il existe une constante $c > 0$ telle que

$$\mathbf{P}\left\{\int_{S^{d-1}} |\hat{p}_n(x; \theta) - p_n(x)| \lambda(dx) \geq \varepsilon\right\} \leq e^{-cn}. \quad (2.2.38)$$

Démonstration: Notons

$$K_n(x) = \frac{C(h_n)}{h_n^{d-1}} K(x),$$

alors

$$\hat{p}_n(x; \theta) = \frac{1}{n} \sum_{i=1}^n K_n\left(\frac{1 - x^T \theta_{ni}}{h_n^2}\right)$$

et

$$\int_{S^{d-1}} K_n\left(\frac{1 - x^T y}{h_n^2}\right) \lambda(dy) = 1. \quad (2.2.39)$$

Considérons

$$\begin{aligned} \int_{S^{d-1}} |\hat{p}_n(x; \theta) - p_n(x)| \lambda(dx) &\leq \int_{S^{d-1}} |\mathbf{E}\hat{p}_n(x; \theta) - p_n(x)| \lambda(dx) \\ &\quad + \int_{S^{d-1}} |\hat{p}_n(x; \theta) - \mathbf{E}\hat{p}_n(x; \theta)| \lambda(dx) \\ &= I_1 + I_2. \end{aligned} \quad (2.2.40)$$

D'après (2.2.39)

$$I_1 \leq \int_{S^{d-1}} \lambda(dx) \int_{S^{d-1}} K_n\left(\frac{1 - x^T y}{h_n^2}\right) |p_n(y) - p_n(x)| \lambda(dy). \quad (2.2.41)$$

Notons $M_n = \sup\{p_n(x), x \in S^{d-1}\}$ et $\Delta_n(x) = \{y \in S^{d-1} \mid 1 - x^T y > \rho h_n^2\}$. Prenons ρ suffisamment grand tel que

$$\begin{aligned}
& \int_{S^{d-1}} \lambda(dx) \int_{\Delta_n(x)} K_n \left(\frac{1 - x^T y}{h_n^2} \right) |p_n(y) - p_n(x)| \lambda(dy) \\
& \leq M_n \int_{S^{d-1}} \lambda(dx) \int_{\Delta_n(x)} K_n \left(\frac{1 - x^T y}{h_n^2} \right) \lambda(dy) \\
& \leq M_n \frac{(2\pi)^{\frac{d-1}{2}} h_n^{d-1}}{\Gamma(\frac{d-1}{2})} \int_{S^{d-1}} \lambda(dx) \int_{\rho}^{\infty} K_n(v) v^{\frac{d-3}{2}} \lambda(dy) \\
& < \frac{\varepsilon}{4}.
\end{aligned} \tag{2.2.42}$$

D'après la condition d'équicontinuité de p_n (2.2.37) et la condition (2.2.39), on a pour n suffisamment grand

$$\begin{aligned}
& \int_{S^{d-1}} \lambda(dx) \int_{S^{d-1} \setminus \Delta_n(x)} K_n \left(\frac{1 - x^T y}{h_n^2} \right) |p_n(y) - p_n(x)| \lambda(dy) \\
& \leq \frac{\varepsilon}{4\lambda(S^{d-1})} \int_{S^{d-1}} \lambda(dx) \int_{S^{d-1} \setminus \Delta_n(x)} K_n \left(\frac{1 - x^T y}{h_n^2} \right) \lambda(dy) \\
& < \frac{\varepsilon}{4}.
\end{aligned} \tag{2.2.43}$$

D'après (2.2.41)-(2.2.43) pour n suffisamment grand

$$I_1 < \frac{\varepsilon}{2}. \tag{2.2.44}$$

Prenons la fonction en escaliers

$$K^*(x) = \sum_{j=1}^k \alpha_j \mathbb{1}_{A_j}(x) \geq 0,$$

où A_1, \dots, A_k sont les ensembles disjoints dans $[0, \infty)$, telle que

$$\frac{C(h_n)(2\pi)^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_0^{\infty} |K(v) - K^*(v)| v^{\frac{d-3}{2}} dv < \frac{\varepsilon}{6}$$

et posons

$$\hat{p}_n^*(x; \theta) = \frac{C(h_n)}{nh_n^{d-1}} \sum_{i=1}^n K^* \left(\frac{1 - x^T \theta_{ni}}{h_n^2} \right).$$

D'après (2.2.23) on a

$$\begin{aligned}
& \int_{S^{d-1}} |\hat{p}_n(x; \theta) - \hat{p}_n^*(x; \theta)| \lambda(dx) \\
& \leq \frac{C(h_n)}{nh_n^{d-1}} \sum_{i=1}^n \int_{S^{d-1}} \left| K \left(\frac{1 - x^T \theta_{ni}}{h_n^2} \right) - K^* \left(\frac{1 - x^T \theta_{ni}}{h_n^2} \right) \right| \lambda(dx) \\
& < \frac{\varepsilon}{6}
\end{aligned} \tag{2.2.45}$$

et

$$\int_{S^{d-1}} |\mathbf{E}\hat{p}_n(x; \theta) - \mathbf{E}\hat{p}_n^*(x; \theta)| \lambda(dx) < \frac{\varepsilon}{6}. \quad (2.2.46)$$

D'après (2.2.40) et (2.2.44)-(2.2.46) pour avoir (2.2.38) il suffit de prouver que pour tout $\varepsilon_1 > 0$ il existe une constante positive c telle que

$$\mathbf{P} \left\{ \int_{S^{d-1}} |\hat{p}_n^*(x; \theta) - \mathbf{E}\hat{p}_n^*(x; \theta)| \lambda(dx) \geq \varepsilon_1 \right\} \leq e^{-cn}. \quad (2.2.47)$$

Sans perdre la généralité on peut prendre ici $K^*(x) = \mathbb{I}_{[a,b]}(x)$.

Pour $x = (x_1, \dots, x_d)^T \in S^{d-1}$ on peut représenter x en coordonnées polaires par

$$\begin{aligned} x_1 &= \cos \phi_1 \\ x_2 &= \sin \phi_1 \cos \phi_2 \\ &\dots \\ x_{d-1} &= \sin \phi_1 \cdots \sin \phi_{d-2} \cos \phi_{d-1} \\ x_d &= \sin \phi_1 \cdots \sin \phi_{d-1} \end{aligned} \quad (2.2.48)$$

avec $0 \leq \phi_i \leq \pi, i = 1, \dots, d-2$ et $0 \leq \phi_{d-1} \leq 2\pi$. Cette représentation est unique dans l'ensemble $H \subset S^{d-1}$ avec $\lambda(H) = 1$. Pour $L > 0$, prenons

$$J_{i_j}^{(j)} = \{x = x(\phi_1, \dots, \phi_{d-1}) \in H \mid L^{-1}h_n(i_j - 1) \leq \phi_j < L^{-1}h_n i_j\},$$

$$i_j = 1, \dots, u - 1 = [h_n^{-1}L\pi], j = 1, \dots, d - 2,$$

$$i_{d-1} = 1, \dots, v - 1 = [h_n^{-1}2L\pi],$$

$$J_u^{(j)} = \{x = x(\phi_1, \dots, \phi_{d-1}) \in H \mid (u - 1)L^{-1}h_n \leq \phi_j < \pi\},$$

$$j = 1, \dots, d - 2,$$

$$J_v^{(d-1)} = \{x = x(\phi_1, \dots, \phi_{d-1}) \in H \mid (v - 1)L^{-1}h_n \leq \phi_{d-1} < 2\pi\}$$

et

$$J_{i_1, \dots, i_{d-1}} = \bigcap_{j=1}^{d-1} J_{i_j}^{(j)}, \quad i_1, \dots, i_{d-2} = 1, \dots, u; i_{d-1} = 1, \dots, v.$$

Tous les ensembles $J_{i_1, \dots, i_{d-1}}$ constituent une partition Ψ de H .

En notant la loi des v.a. $\theta_{n1}, \dots, \theta_{nm}$ par σ_n , la loi empirique de $\theta_{n1}, \dots, \theta_{nm}$ par σ_{nn} , les ensembles

$$A^*(x) = \left\{ y \in H \mid \frac{1 - x^T y}{h_n^2} \in [a, b] \right\}, \quad x \in H,$$

on a

$$\int_{S^{d-1}} |\hat{p}_n^*(x; \theta) - \mathbf{E}\hat{p}_n^*(x; \theta)| \lambda(dx) = \frac{C(h_n)}{h_n^{d-1}} \int_{S^{d-1}} |\sigma_{nn}(A^*(x)) - \sigma_n(A^*(x))| \lambda(dx).$$

Bai *et al.* dans [BRRZ88] ont montré que

$$\begin{aligned} & \mathbf{P} \left\{ \frac{C(h_n)}{h_n^{d-1}} \int_{S^{d-1}} |\sigma_{nn}(A^*(x)) - \sigma_n(A^*(x))| \lambda(dx) \geq \varepsilon_1 \right\} \\ & \leq \mathbf{P} \left\{ \sum_{J \in \Psi} |\sigma_{nn}(J) - \sigma_n(J)| \geq \frac{\varepsilon_1}{3c} \right\} \end{aligned}$$

où c est une constante positive. En appliquant le lemme 2.2.4. sur le terme de droite dans l'inégalité précédente, on obtient (2.2.47). \square

Corollaire 2.2.8. *Si la loi F vérifie la condition (FVR) et que les conditions (C1), (C4) et (C6) sont remplies, les densités de θ_{ni} , notées p_n , satisfont la condition d'équicontinuité (2.2.37), alors*

$$\int_{S^{d-1}} |\hat{p}_n(x; \theta) - p(x)| \lambda(dx) \xrightarrow[n \rightarrow \infty]{p.s.} 0. \quad (2.2.49)$$

Démonstration: C'est une conséquence immédiate des théorèmes 2.2.2. et 2.2.7.. \square

2.3 Estimation basée sur l'échantillon permuté

D'après les résultats de normalité asymptotique des estimateurs présentés dans la sous-section 2.1.3, la valeur optimale de r est $2/3$. C'est-à-dire qu'étant donné N données issues d'une loi à queue régulière, nous en utilisons donc $2 \times \lceil N^{2/3} \rceil$ entre eux pour estimer les paramètres. D'après des expériences sur des données simulées, nous avons besoin d'au moins 5000 données pour avoir un résultat d'estimation fiable. Puisqu'en pratique nous n'avons pas toujours autant de données, les résultats d'estimation ne sont pas satisfaisants. Pour résoudre ce problème nous proposons une méthode de type de bootstrap pour augmenter le nombre des points utilisés dans l'échantillon.

Soit $\Sigma_N = \{\pi_1, \dots, \pi_{N!}\}$ l'ensemble des permutations de $\{1, \dots, N\}$, Π l'élément aléatoire uniformément réparti sur Σ_N , c'est-à-dire

$$\mathbf{P}\{\Pi = \pi_i\} = \frac{1}{N!}, \quad 1 \leq i \leq N!$$

Pour chaque π_i la permutation de l'échantillon $\xi_{\pi_i(1)}, \dots, \xi_{\pi_i(N)}$ est aussi un échantillon i.i.d. issu de même loi que ξ_1, \dots, ξ_N . Nous présentons des estimateurs des paramètres α et σ basés sur la permutation de l'échantillon.

Notons

$$u_{n,\pi_i} = \frac{S_{n,\pi_i}}{n} \quad \text{et} \quad \alpha_{n,\pi_i} = \frac{u_{n,\pi_i}}{1 - u_{n,\pi_i}}, \quad 1 \leq i \leq N!,$$

où S_{n,π_i} est défini pour la permutation π_i de l'échantillon, noté $\xi_{\pi_i(1)}, \dots, \xi_{\pi_i(N)}$ (voir l'annexe A pour la définition de S_n). Pour n fixé les variables aléatoires u_{n,π_i} sont de même loi. C'est-à-dire

$$u_{n,\pi_i} \stackrel{\mathcal{L}}{=} u_{n,\pi_j}, \quad i \neq j.$$

D'après le théorème 2.1.4., pour chaque $\pi_i \in \Sigma_N$, on a

$$u_{n,\pi_i} \xrightarrow[n \rightarrow \infty]{p.s.} \frac{\alpha}{1 + \alpha}, \quad (2.3.1)$$

d'où

$$\alpha_{n,\pi_i} \xrightarrow[n \rightarrow \infty]{p.s.} \alpha. \quad (2.3.2)$$

Puisque $0 \leq u_{n,\pi_i} \leq 1$, $\forall n, i$, d'après le théorème de convergence dominée, on a pour chaque $\pi_i \in \Sigma_N$

$$\mathbf{E} \left| u_{n,\pi_i} - \frac{\alpha}{1 + \alpha} \right| \rightarrow 0, \quad n \rightarrow \infty. \quad (2.3.3)$$

Ensuite notons

$$u_{n,\Pi} = \frac{S_{n,\Pi}}{n}, \quad \alpha_{n,\Pi} = \frac{u_{n,\Pi}}{1 - u_{n,\Pi}},$$

où $S_{n,\Pi}$ est défini sur la permutation aléatoire Π de l'échantillon, noté $\xi_{\Pi(1)}, \dots, \xi_{\Pi(N)}$. La loi de la variable aléatoire $u_{n,\Pi}$ est la loi marginale de la loi jointe du vecteur aléatoire $(u_{n,\Pi}, \Pi)$. Ainsi pour tout ensemble borélien $B \subset [0, 1]$ on a

$$\mathbf{P}\{u_{n,\Pi} \in B\} = \sum_{i=1}^N \mathbf{P}\{u_{n,\Pi} \in B \mid \Pi = \pi_i\} \mathbf{P}\{\Pi = \pi_i\} = \mathbf{P}\{u_{n,\pi_1} \in B\} \quad (2.3.4)$$

d'où

$$u_{n,\Pi} \stackrel{\mathcal{L}}{=} u_{n,\pi_1} \quad \text{et} \quad \alpha_{n,\Pi} \stackrel{\mathcal{L}}{=} \alpha_{n,\pi_1}. \quad (2.3.5)$$

Ainsi d'après les convergences (2.3.1), (2.3.2) et (2.3.3) on a

$$u_{n,\Pi} \xrightarrow[n \rightarrow \infty]{P} \frac{\alpha}{1 + \alpha}, \quad (2.3.6)$$

$$\alpha_{n,\Pi} \xrightarrow[n \rightarrow \infty]{P} \alpha, \quad (2.3.7)$$

où $\xrightarrow[n \rightarrow \infty]{P}$ représente la convergence en probabilité, et

$$\mathbf{E} \left| u_{n,\Pi} - \frac{\alpha}{1 + \alpha} \right| \rightarrow 0, \quad n \rightarrow \infty. \quad (2.3.8)$$

Soient Π_1, \dots, Π_k une suite des copies indépendantes de Π . En notant

$$u_{n,k} = \frac{1}{k} \sum_{i=1}^k u_{n,\Pi_i},$$

définissons deux estimateurs de α

$$\hat{\alpha}_{n,k}^{(1)} = \frac{u_{n,k}}{1 - u_{n,k}} \quad \text{et} \quad \hat{\alpha}_{n,k}^{(2)} = \frac{1}{k} \sum_{i=1}^k \alpha_{n,\Pi_i}. \quad (2.3.9)$$

D'après (2.3.5) et les convergences (2.3.6)-(2.3.8) on a pour k fixé

$$u_{n,k} \xrightarrow[n \rightarrow \infty]{P} \frac{\alpha}{1 + \alpha},$$

$$u_{n,k} \xrightarrow[n \rightarrow \infty]{\mathcal{L}_1} \frac{\alpha}{1 + \alpha},$$

$$\hat{\alpha}_{n,k}^{(1)} \xrightarrow[n \rightarrow \infty]{P} \alpha,$$

et

$$\hat{\alpha}_{n,k}^{(2)} \xrightarrow[n \rightarrow \infty]{P} \alpha.$$

Pour l'estimateur de la masse totale de la mesure spectrale $\sigma(S)$, prenons

$$\widehat{\sigma(S)}_{n,k} = \left(\frac{1}{n\Gamma(1 - \frac{t}{\alpha})} \sum_{i=1}^n \omega_{k,i} \right)^{\frac{\alpha}{t}} \quad \text{avec} \quad \omega_{k,i} = \frac{1}{k} \sum_{j=1}^k (q_{mi}^{\Pi_j})^t, \quad (2.3.10)$$

où $q_{mi}^{\Pi_j}$ est défini sur la permutation de l'échantillon $\xi_{\Pi_j(1)}, \dots, \xi_{\Pi_j(N)}$ (voir l'annexe A pour la définition de q_{mi}). De manière analogue à (2.3.4) on peut déduire que

$$q_{mi}^{\Pi} \stackrel{\mathcal{L}}{=} q_{mi}^{\pi_1}, \quad (2.3.11)$$

où Π est un élément aléatoire uniformément réparti sur Σ_N et π_1 est un élément dans Σ_N . Supposons que le v.a. ξ vérifie la condition (1.1.6) avec $b_n = n^{1/\alpha}$. D'après le théorème 2.1.6. (voir aussi l'annexe A), on a pour $0 < t < \frac{\alpha r}{2}$

$$\frac{1}{n} \sum_{i=1}^n (q_{mi}^{\pi_1})^t - \Gamma\left(1 - \frac{t}{\alpha}\right) \sigma(S)^{\frac{t}{\alpha}} \xrightarrow[N \rightarrow \infty]{p.s.} 0.$$

En considérant (2.3.11) on obtient pour k fixé

$$\frac{1}{n} \sum_{i=1}^n \omega_{k,i} \xrightarrow[n \rightarrow \infty]{P} \Gamma\left(1 - \frac{t}{\alpha}\right) \sigma(S)^{\frac{t}{\alpha}},$$

d'où

$$\widehat{\sigma(S)}_{n,k} \xrightarrow[n \rightarrow \infty]{P} \sigma(S).$$

Maintenant considérons l'estimateur de la mesure spectrale σ . Notons $\theta_{m1}^{\Pi_j}, \dots, \theta_{mn}^{\Pi_j}$ les v.a. sur la sphère définis pour la permutation aléatoire Π_j de l'échantillon, $\xi_{\Pi_j(1)}, \dots, \xi_{\Pi_j(N)}$ (voir l'annexe A pour la définition de θ_{mi}). Les v.a. $\{\theta_{mi}^{\Pi_j}, i = 1, \dots, n; j = 1, \dots, k\}$ sont de même loi. De plus pour tout ensemble borélien B sur la sphère unité S tel que $\sigma(\partial B) = 0$ on a

$$\mathbf{P}\{\theta_{m1}^{\Pi_1} \in B\} = \sum_{j=1}^{N!} \mathbf{P}\{\theta_{m1}^{\pi_j} \in B\} \mathbf{P}\{\Pi_1 = \pi_j\} = \mathbf{P}\{\theta_{m1}^{\pi_1} \in B\}.$$

C'est-à-dire

$$\theta_{mi}^{\Pi_j} \stackrel{\mathcal{L}}{=} \theta_{m1}^{\pi_1}, \quad 1 \leq i \leq n, 1 \leq j \leq k.$$

D'après le théorème 2.1.7. on a pour chaque i et j

$$\theta_{mi}^{\Pi_j} \Rightarrow \sigma, \quad m \rightarrow \infty. \quad (2.3.12)$$

Pour chaque j en appliquant la proposition 2.1.1. sur les variables aléatoires $\{\mathbb{I}_B(\theta_{mi}^{\Pi_j}), i = 1, \dots, n\}$, on obtient pour $\forall B \in \mathcal{B}(S)$ tel que $\sigma(\partial B) = 0$,

$$\frac{1}{n} \sum_{i=1}^n \mathbb{I}_B(\theta_{mi}^{\Pi_j}) \xrightarrow[n \rightarrow \infty]{P} \sigma(B). \quad (2.3.13)$$

Définissons l'estimateur de σ par

$$\hat{\sigma}_{n,k} = \frac{1}{nk} \sum_{i=1}^n \sum_{j=1}^k \delta_{\theta_{mi}^{\Pi_j}}. \quad (2.3.14)$$

La convergence (2.3.13) implique que pour k fixé et $\forall B \in \mathcal{B}(S)$ tel que $\sigma(\partial B) = 0$,

$$\hat{\sigma}_{n,k}(B) \xrightarrow[n \rightarrow \infty]{P} \sigma(B).$$

Supposons que σ est absolument continue avec la densité p . Définissons l'estimateur de p par

$$\hat{p}_n(x; \theta_{m1}^{\Pi_j}, \dots, \theta_{mn}^{\Pi_j}, j = 1, \dots, k) = \frac{C(h_n)}{nh_n^{d-1}} \sum_{j=1}^k \sum_{i=1}^n K \left(\frac{1 - x^T \theta_{mi}^{\Pi_j}}{h_n^2} \right). \quad (2.3.15)$$

C'est-à-dire qu'on utilise $n \times k$ points pour estimer la densité de σ . Pour simplifier, on utilise la notation $\hat{p}_{n,k}(x; \theta)$ pour l'estimateur de p défini par (2.3.15). Les v.a. $\{\theta_{m1}^{\Pi_j}, \dots, \theta_{mn}^{\Pi_j}\}$ sont i.i.d. pour chaque j . Si les conditions du théorème 2.2.6. sont remplies, alors pour x un point de continuité de p on a

$$\frac{C(h_n)}{nh_n^{d-1}} \sum_{i=1}^n K \left(\frac{1 - x^T \theta_{mi}^{\Pi_j}}{h_n^2} \right) \xrightarrow[n \rightarrow \infty]{P} p(x),$$

d'où

$$\hat{p}_{n,k}(x; \theta) \xrightarrow[n \rightarrow \infty]{P} p(x).$$

On illustre l'amélioration de l'estimation de la densité de σ basée sur l'échantillon permuté dans la Figure 2.1. Les lois $\mathcal{S}_2(0.75, \sigma_i, 0)$, $i = 1, 2, 3$ avec σ_i définis par (1.3.9), (1.3.10) et (1.3.11) sont considérées. La taille d'échantillon $N = 10^5$. Les résultats d'estimation par la méthode de noyau basée sur l'échantillon initial sont présentés dans la colonne de gauche. Ensuite on permute l'échantillon 100 fois aléatoirement, i.e. $k = 100$. L'estimateur défini par (2.3.15) est utilisé. Les résultats obtenus sont présentés dans la colonne de droite. On constate que les densités estimées par (2.3.15) sont plus proches des densités théoriques.

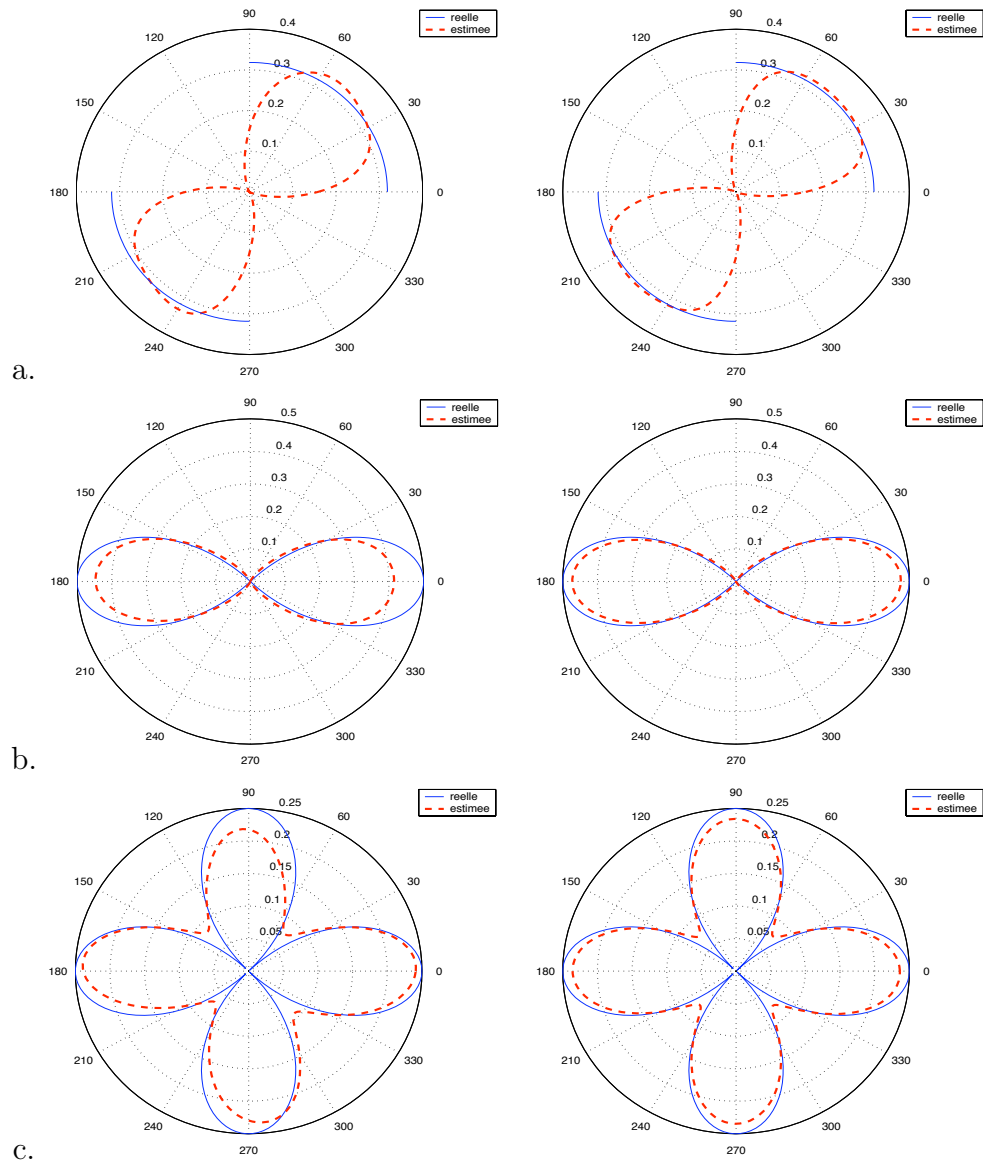


FIG. 2.1: Densités empiriques de mesures spectrales des lois $\mathcal{S}_2(0.75, \sigma)$ estimées par la méthode de noyau. Les densités de σ sont définies par (1.3.9) (a), (1.3.10) (b) et (1.3.11) (c), $N = 10^5$. Estimation sur l'échantillon initial (gauche), permuté (droite), $k = 100$.

2.4 Performance des estimateurs

Afin de vérifier les propriétés des estimateurs présentés auparavant, nous avons écrit le programme MATLAB “mvrfit.mat”. Ce programme permet d’estimer l’indice caractéristique α et la mesure spectrale σ d’une loi à queue régulière dans l’espace \mathbb{R}^d , $d \leq 30$. La performance des estimateurs est étudiée en utilisant des données simulées dans la section 1.3. Des résultats d’estimation sur les échantillons de taille fixe sont présentés dans la sous-section 2.4.1. Nous montrons l’influence du regroupement d’échantillon dans la sous-section 2.4.2. La vitesse de convergence des estimateurs est illustrée dans la sous-section 2.4.3. Nous présentons brièvement des méthodes connues pour estimer les paramètres des lois α -stables dans la sous-section 2.4.4. Elles sont comparées avec notre méthodes en utilisant des données simulées.

2.4.1 Résultats d’estimation

Nous résumons l’algorithme et des propriétés principales des estimateurs dans l’annexe A. Le programme est présenté dans l’annexe B. Les données simulées dans la section 1.3 sont utilisées pour évaluer les statistiques. Les mesures spectrales qui apparaissent dans cette section sont normalisées, c’est-à-dire $\sigma(S) = 1$.

Les résultats d’estimation des paramètres des lois α -stables sont présentés dans les Tableaux 2.1 et 2.2. Les f.d.r. empiriques de la mesure spectrale sont présentées dans la Figure 2.2. Les résultats d’estimation des paramètres des lois appartenant au domaine d’attraction d’une loi $\mathcal{S}\mathcal{t}\alpha\mathcal{S}$ sont présentés dans le Tableau 2.3. Les f.d.r. empiriques de la mesure spectrale correspondante sont présentées dans les Figures 2.3 et 2.4.

Les résultats d’estimation des paramètres des lois max-stables sont présentés dans les Tableaux 2.4 et 2.5. Pour dessiner la f.d.r. de mesure spectrale d’une loi $\mathcal{M}\mathcal{S}_2(\alpha, \sigma)$, nous identifions la sphère unité S_{\vee}^1 avec l’intervalle $[-1, 1]$. Soit $\theta = (\theta^{(1)}, \theta^{(2)})$ le point dans S_{\vee}^1 . Définissons la bijection

$$\begin{aligned} \Pi : S_{\vee}^1 &\rightarrow [-1, 1], \\ \theta &\mapsto \text{sign}(\theta^{(1)} - \theta^{(2)})(1 - \theta^{(1)} \wedge \theta^{(2)}). \end{aligned} \quad (2.4.1)$$

Sous cette bijection la loi uniforme sur S_{\vee}^1 devient la loi uniforme sur l’intervalle $[-1, 1]$. La mesure concentrée sur deux points $(1, a)$ et $(b, 1)$, $0 \leq a, b \leq 1$, devient la mesure concentrée sur les points $(1 - a, 0)$ et $(b - 1, 0)$. Les f.d.r. empiriques de la mesure spectrale estimée sont présentées dans la Figure 2.5.

Dans les tableaux suivants on observe que le paramètre $\sigma(S)$ est mal estimé à chaque fois que le paramètre α correspondant n’est pas bien estimé, par exemple dans le cas $\mathcal{S}_2(1.5, \sigma, 0)$ où σ est uniforme. Ce phénomène montre la dépendance entre l’estimateur de $\sigma(S)$ et celui de α . On observe aussi que les longueurs des intervalles de confiance de α sont plus grandes pour $\alpha = 1.5$ que pour $\alpha = 0.75$.

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$
$\mathcal{S}_1(0.5, (0.5, 1), 0)$	0.46 (+0.10/-0.09)	0.57 (± 0.13)	0.73 (+0.16/-0.14)
$\mathcal{S}_1(1.5, (-0.5, 1), 0)$	1.57 (+0.30/-0.25)	-0.30 (± 0.15)	1.45 (+0.30/-0.25)

TAB. 2.1: Paramètres estimés des lois α -stables unidimensionnelles. Les valeurs dans les parenthèses sont les intervalles de confiance 95%, $N = 5000$.

σ	$\mathcal{S}_2(0.75, \sigma, 0)$		$\mathcal{S}_2(1.5, \sigma, 0)$	
	$\hat{\alpha}$	$\widehat{\sigma(S^1)}$	$\hat{\alpha}$	$\widehat{\sigma(S^1)}$
uniforme	0.71 (+0.07/-0.06)	0.82 (+0.09/-0.08)	1.35 (+0.11/-0.10)	0.66 (+0.07/-0.07)
f_1	0.79 (+0.08/-0.07)	1.24 (+0.14/-0.13)	1.46 (+0.13/-0.11)	1.00 (+0.11/-0.10)
f_2	0.69 (+0.07/-0.06)	0.76 (+0.09/-0.08)	1.49 (+0.13/-0.12)	0.99 (+0.11/-0.10)
f_3	0.74 (+0.07/-0.07)	0.99 (+0.13/-0.11)	1.46 (+0.13/-0.12)	1.06 (+0.13/-0.12)

TAB. 2.2: Paramètres estimés des lois α -stables bivariées où les densités de mesure spectrale $f_i, i = 1, 2, 3$ sont définies par (1.3.9), (1.3.10) et (1.3.11). Les mesures spectrales estimées correspondantes sont présentées dans la Figure 2.2. Les valeurs dans les parenthèses sont les intervalles de confiance 95%, $N = 5 \times 10^4$.

σ	Dom(0.75, σ)		Dom(1.5, σ)	
	$\hat{\alpha}$	$\widehat{\sigma(S^1)}$	$\hat{\alpha}$	$\widehat{\sigma(S^1)}$
f_1	0.73 (+0.07/-0.06)	0.87 (+0.10/-0.09)	1.57 (+0.14/-0.12)	1.26 (+0.14/-0.13)
Y f_2	0.79 (+0.07/-0.07)	1.27 (+0.14/-0.13)	1.50 (+0.13/-0.12)	1.13 (+0.13/-0.12)
f_3	0.70 (+0.07/-0.06)	0.78 (+0.09/-0.08)	1.43 (+0.12/-0.11)	0.85 (+0.09/-0.08)
f_1	0.76 (+0.07/-0.07)	2.11 (+0.24/-0.22)	1.59 (+0.14/-0.13)	2.87 (+0.32/-0.29)
Z f_2	0.77 (+0.07/-0.07)	2.15 (+0.24/-0.22)	1.54 (+0.13/-0.12)	2.48 (+0.27/-0.25)
f_3	0.70 (+0.07/-0.06)	0.72 (+0.08/-0.07)	1.54 (+0.14/-0.12)	1.14 (+0.12/-0.11)

TAB. 2.3: Paramètres estimés des lois appartenant au domaine d'attraction d'une loi $St\alpha S$ simulées par les transformations d'une loi stable isotropique. Les transformations Y et Z sont définies par (1.3.12) et (1.3.13). Les densités de mesure spectrale $f_i, i = 1, 2, 3$ sont définies par (1.3.9), (1.3.10) et (1.3.11). Les mesures spectrales estimées correspondantes sont présentées dans les Figures 2.3 et 2.4. Les valeurs dans les parenthèses sont les intervalles de confiance 95%, $N = 5 \times 10^4$.

σ	$\mathcal{MS}_2(0.75, \sigma)$	
	$\hat{\alpha}$	$\widehat{\sigma(S_V^1)}$
uniforme	0.78 (+0.08/-0.07)	1.19 (+0.14/-0.12)
$\frac{1}{2}(\delta_{(1,0)} + \delta_{(0,1)})$	0.71 (+0.07/-0.06)	0.81 (+0.09/-0.08)
$\frac{1}{2}(\delta_{(1,0.2)} + \delta_{(0.2,1)})$	0.74 (+0.07/-0.07)	0.92 (+0.10/-0.09)
$\frac{1}{2}(\delta_{(1,0.5)} + \delta_{(0.5,1)})$	0.73 (+0.07/-0.07)	0.90 (+0.10/-0.09)

TAB. 2.4: Paramètres estimés des lois max-stables bivariées d'indice $\alpha = 0.75$. Les mesures spectrales estimées correspondantes sont présentées dans la Figure 2.5. Les valeurs dans les parenthèses sont les intervalles de confiance 95%, $N = 5 \times 10^4$.

	$\hat{\alpha}$	$\widehat{\sigma(S_V^1)}$
$\mathcal{MS}_2(1.5, \sigma)$	1.53 (+0.14/-0.12)	0.98 (+0.11/-0.10)
$\mathcal{MS}_2(3, \sigma)$	2.91 (+0.24/-0.22)	0.85 (+0.09/-0.08)

TAB. 2.5: Paramètres estimés des lois max-stables bivariées d'indice $\alpha = 1.5, 3$, la mesure spectrale est définie par $\sigma = \frac{1}{2}(\delta_{(1,0.5)} + \delta_{(0.5,1)})$. Les valeurs dans les parenthèses sont les intervalles de confiance 95%, $N = 5 \times 10^4$.

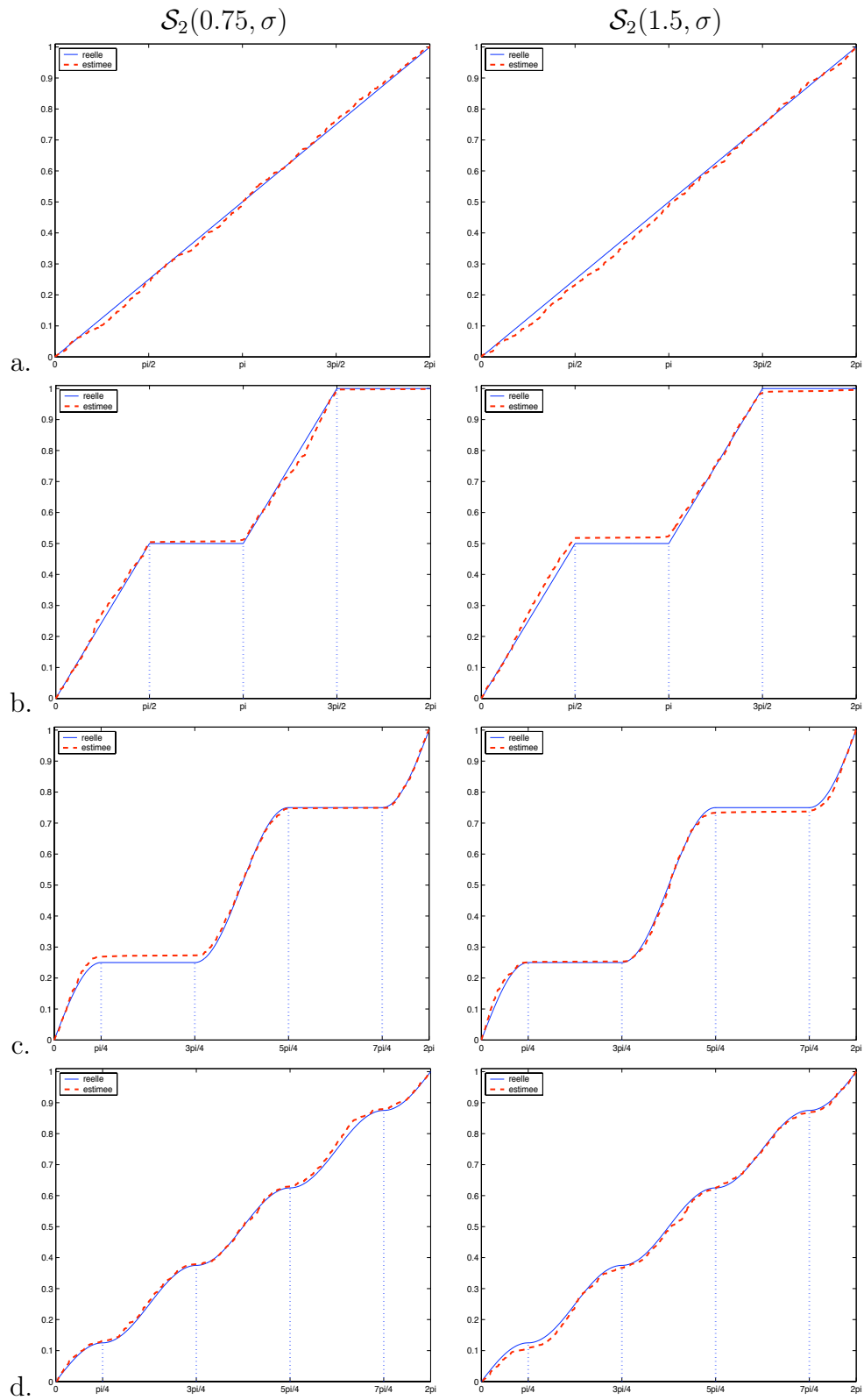


FIG. 2.2: F.d.r. empiriques des mesures spectrales des lois $\mathcal{S}_2(0.75, \sigma)$ (gauche) et des lois $\mathcal{S}_2(1.5, \sigma)$ (droite), σ est uniforme (a), les densités de σ sont définies par (1.3.9) (b), (1.3.10) (c) et (1.3.11) (d), $N = 5 \times 10^4$.

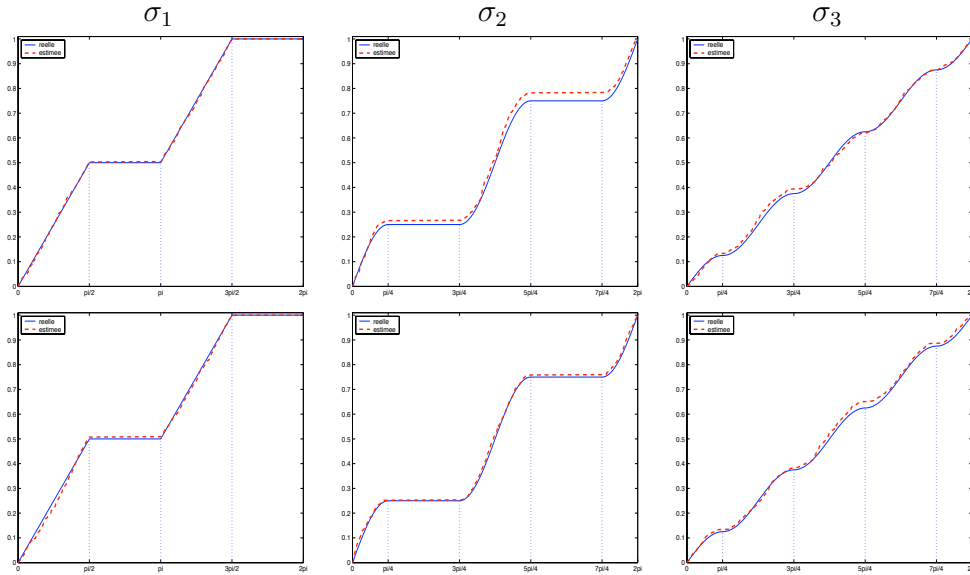


FIG. 2.3: F.d.r. empiriques des mesures spectrales des lois appartenant au domaine d'attraction de la loi $\mathcal{S}_2(0.75, \sigma_i)$ (en haut), et de la loi $\mathcal{S}_2(1.5, \sigma)$ (en bas), simulées par transformation (1.3.12), les densités de σ_i , $i = 1, 2, 3$, sont définies par (1.3.9), (1.3.10) et (1.3.11), $N = 5 \times 10^4$.

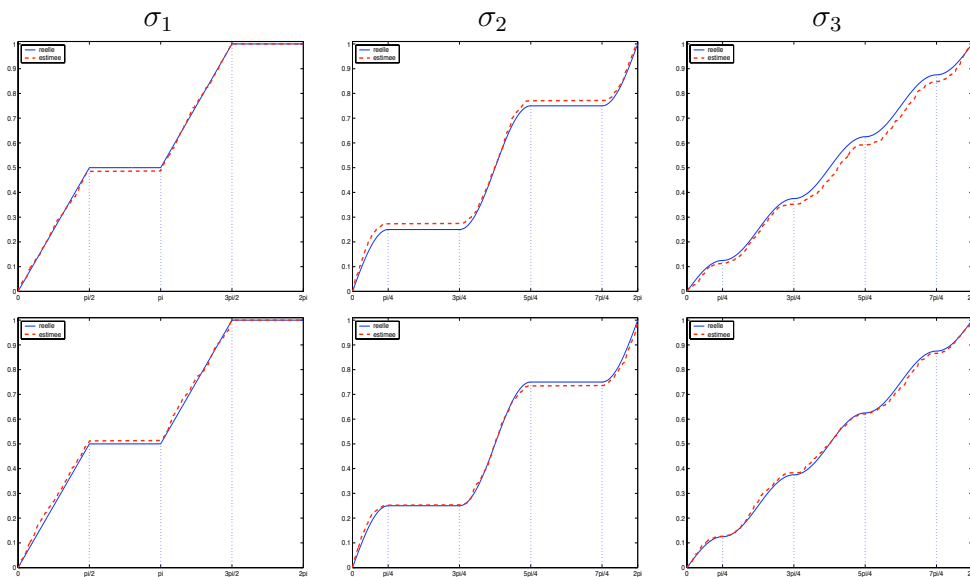


FIG. 2.4: F.d.r. empiriques des mesures spectrales des lois appartenant au domaine d'attraction de la loi $\mathcal{S}_2(0.75, \sigma)$ (en haut), et de la loi $\mathcal{S}_2(1.5, \sigma)$ (en bas), simulées par transformation (1.3.13), les densités de σ_i , $i = 1, 2, 3$, sont définies par (1.3.9), (1.3.10) et (1.3.11), $N = 5 \times 10^4$.

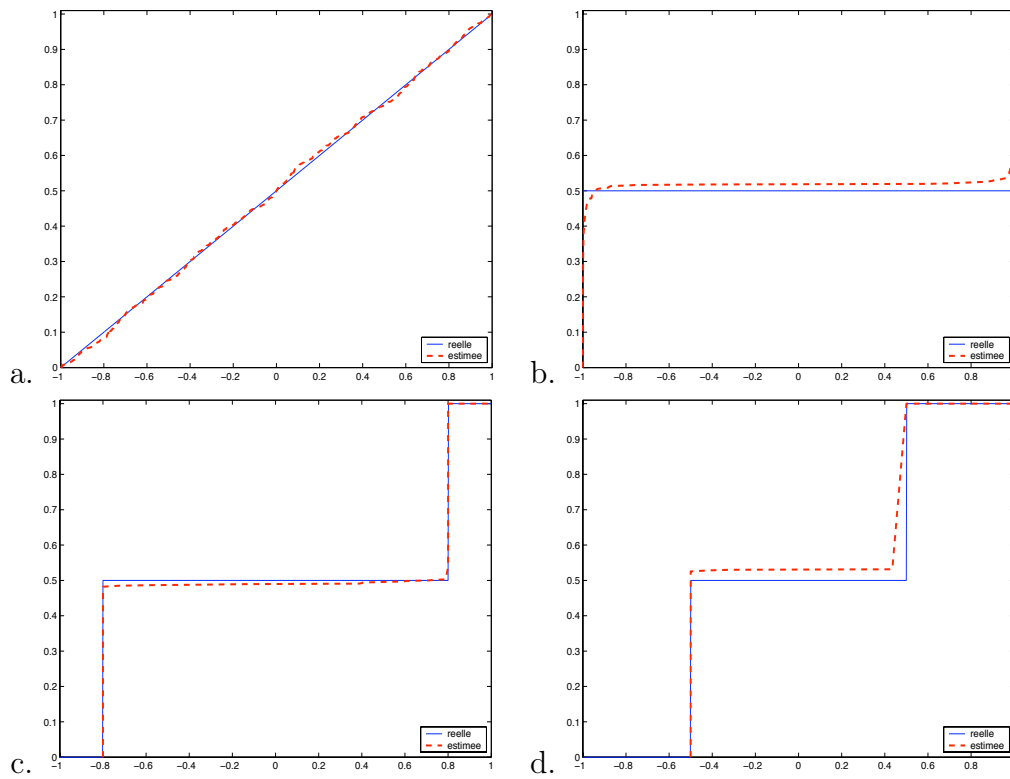


FIG. 2.5: F.d.r. empiriques des mesures spectrales des lois $\mathcal{MS}_2(0.75, \sigma)$, où σ est uniforme (a), σ est définie par (1.3.17) avec $p = q = 0.5$ et $a = b = 0$ (b), $a = b = 0.2$ (c), $a = b = 0.5$ (d), $N = 5 \times 10^4$.

2.4.2 Influence du regroupement d'échantillon

Les propriétés asymptotiques des estimateurs dépendent de la façon de former les paquets. On va étudier l'influence du regroupement d'échantillon par simulation. Cela nous permet non seulement de constater l'effet du regroupement sur les résultats d'estimation, mais aussi d'avoir une idée de choix pour n quand la taille d'échantillon est petite.

Pour le v.a. $\mathcal{St}\alpha\mathcal{S}$ dans \mathbb{R}^d on a une relation asymptotique du second ordre avec $\rho = 2\alpha$ (voir (2.1.41)). Dans ce cas d'après les théorèmes 2.1.10. et 2.1.14. (voir aussi l'annexe A) si on choisit

$$n = N^{2/3-\varepsilon}, \quad m = N^{1/3+\varepsilon},$$

où ε est une constante strictement positive et arbitrairement petite, alors on a une convergence du type théorème central limite pour les estimateurs $\hat{\alpha}_N$ et $\hat{\sigma}_N$. C'est-à-dire la vitesse de convergence des estimateurs dans \mathcal{L}_1 peut atteindre $\sqrt{n} \sim N^{1/3-\varepsilon/2}$. Les propriétés asymptotiques de $\widehat{\sigma(S)}_N$ sont plus compliquées car cet estimateur dépend de l'indice α .

On estime les paramètres de N -échantillon issu d'une loi stable unidimensionnelle en changeant l'indice de regroupement r . Les estimateurs $\hat{\alpha}_N$, $\widehat{\sigma(S)}_N$ et $\hat{\beta}_N$ sont définis par (2.1.22), (2.1.30) et (2.1.38) (voir aussi l'annexe A). Les procédures d'estimation sont répétées 50 fois sur des échantillons indépendants. La moyenne des valeurs obtenues est prise comme la valeur estimée. Notons $s = 1 - r$. Les graphes des points $(s, \hat{\alpha}_N)$ et $(s, \hat{\beta}_N)$ sont présentés dans la Figure 2.6. Il semble que la valeur optimale de s soit autour de 0.4 qui est plus grande que la valeur théorique $1/3$. Cela peut être expliqué, pour une valeur petite de N , le rôle de constante ε devient important.

2.4.3 Illustration de la convergence des estimateurs

Pour étudier la convergence de l'estimateur de la mesure spectrale, nous avons besoin de choisir une métrique appropriée pour mesurer la distance entre deux mesures. Nous considérons ici les lois à queue régulière dans l'espace \mathbb{R}^2 , la mesure spectrale est donc sur la sphère S^1 . Soient θ_1 et θ_2 deux variables aléatoires à valeurs dans S^1 . Notons F_1 et F_2 les f.d.r. des θ_1 et θ_2 . Il est connu que la métrique de Lévy définie par

$$L(\theta_1, \theta_2) := \inf\{\varepsilon > 0 \mid F_1(x - \varepsilon) - \varepsilon \leq F_2(x) \leq F_1(x + \varepsilon) + \varepsilon\}.$$

est une métrique de la convergence faible, c'est-à-dire si $L(\theta_n, \theta) \rightarrow 0$ quand $n \rightarrow \infty$, alors $\theta_n \Rightarrow \theta$. Le calcul de cette métrique n'est pas facile. Comme la sphère unité est un espace compact, il est préférable d'utiliser une autre distance plus simple pour mesurer la distance entre les mesures. Rappelons que la distance de \mathcal{L}_1 entre deux fonctions F_1 et F_2 est définie par

$$\|F_1 - F_2\|_1 = \int_{S^1} |F_1 - F_2| \lambda(dx),$$

où λ est la mesure de Lebesgue sur \mathbb{R}^1 . Soient $\{F_n\}$ une suite de f.d.r., F une f.d.r. dans S^1 . La convergence $F_n \xrightarrow{\mathcal{L}_1} F$ a lieu si et seulement si $\|F_n - F\|_1 \rightarrow 0$.

Proposition 2.4.1. *Soient σ_n, σ des lois de probabilité sur la sphère unité S^1 avec les*

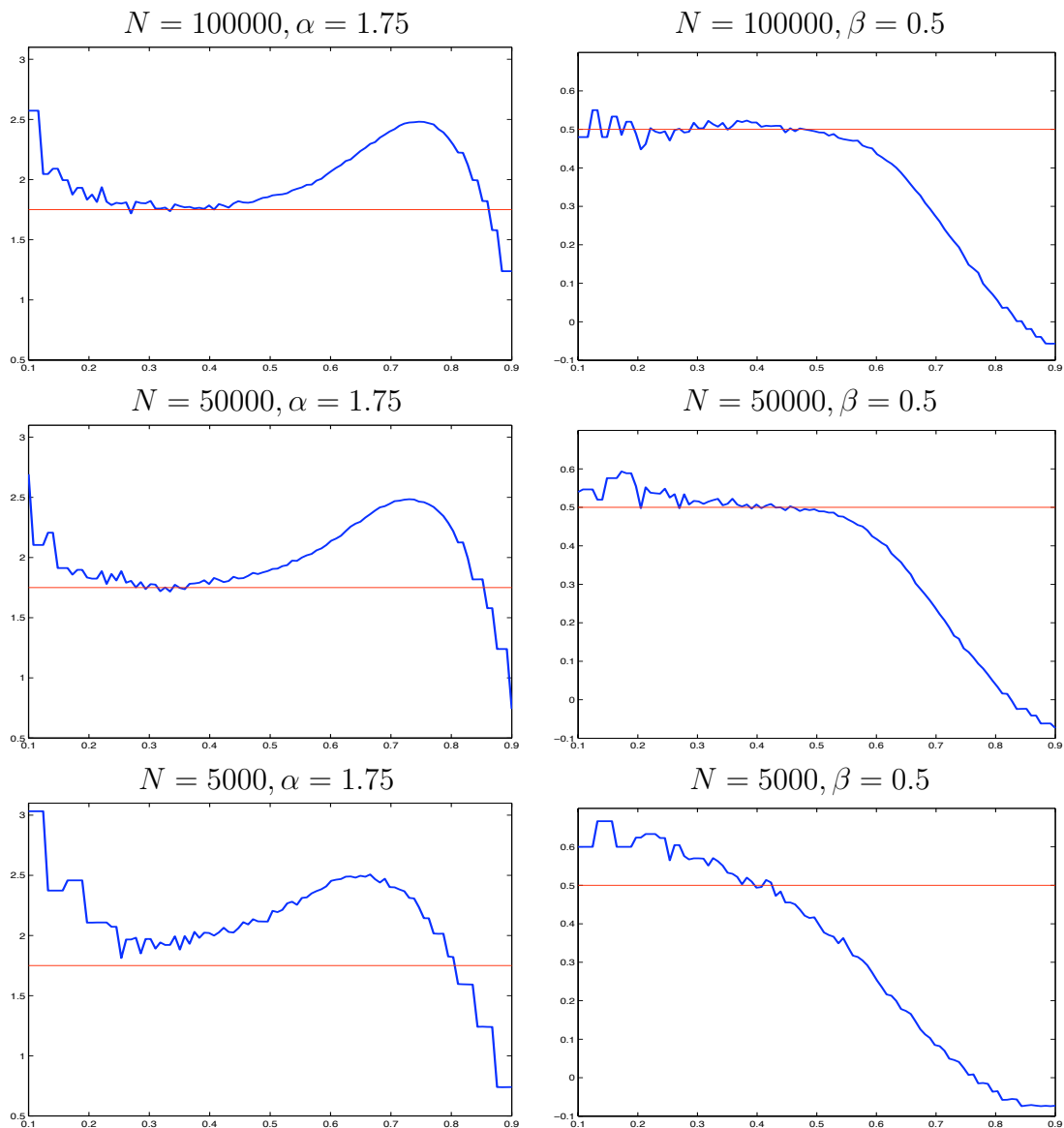


FIG. 2.6: Résultats d'estimation des paramètres α (gauche) et β (droite) de loi $\mathcal{S}_1(1.75, (0.5, 1), 0)$ en prenant les différentes valeurs de l'indice r . L'axe x représente $s = 1 - r$. La ligne au milieu représente la valeur réelle.

f.d.r. F_n et F . On a la convergence faible $\sigma_n \Rightarrow \sigma, n \rightarrow \infty$ si et seulement si $F_n \xrightarrow{\mathcal{L}_1} F, n \rightarrow \infty$.

Démonstration: La nécessité est triviale. Supposons que $F_n \xrightarrow{\mathcal{L}_1} F$ quand $n \rightarrow \infty$. Puisque la sphère est compacte, la famille $\{\sigma_n\}$ est tendue ainsi relativement compacte. C'est-à-dire quelle que soit une suite $\{\sigma_{n'}\} \subset \{\sigma_n\}$, il existe une sous-suite $\{\sigma_{n''}\} \subset \{\sigma_{n'}\}$ telle que $\sigma_{n''} \Rightarrow \mu$ où μ est une mesure définie sur S^{d-1} avec la f.d.r. notée G . Notons $F_{n''}$ la f.d.r. de $\sigma_{n''}$. Ainsi on a $F_{n''}(x) \rightarrow G(x)$ sur tous les points de continuité de G . On déduit $F_n \xrightarrow{\mathcal{L}_1} G$ du fait que la f.d.r. est *càdlàg* (continue à droite avec limite à gauche). Par l'unicité de la limite dans \mathcal{L}_1 , F et G coïncident. Par conséquent $\sigma = \mu$, d'où vient la convergence $\sigma_n \Rightarrow \sigma$. \square

Nous étudions la convergence des estimateurs à l'aide d'une simulation de la loi $\mathcal{S}_2(0.75, \sigma, 0)$ où σ a la densité définie par (1.3.11). La procédure d'estimation des paramètres est effectuée pour des échantillons de taille N variée de 500 à 5×10^6 . La Figure 2.7 présente le graphe des paramètres α et $\sigma(S)$ estimés en fonction de la taille d'échantillon. La dépendance entre le paramètre de masse totale et le paramètre α peut être observée en comparant les deux premiers diagrammes de cette figure. Le dernier diagramme présente les distances de \mathcal{L}_1 entre les f.d.r. empiriques et réelles de mesure spectrale en fonction de la taille d'échantillon. On constate que la convergence de $\hat{\sigma}(\cdot)$ est assez lente.

Ensuite la procédure d'estimation sur $k = 100$ échantillons indépendants issus de la même loi qu'avant est effectuée. La taille d'échantillon varie de 500 à 10^5 . On obtient donc les valeurs estimées des paramètres pour chaque échantillon, notées $\hat{\alpha}_{N,i}$, $\widehat{\sigma(S)}_{N,i}$ et $\hat{\sigma}_{N,i}$, $i = 1, \dots, k$. Pour chaque taille on calcule les moyennes des distances

$$\bar{d}_{\alpha,N} = \frac{1}{k} \sum_{i=1}^k |\hat{\alpha}_{N,i} - \alpha| \quad (2.4.2)$$

$$\bar{d}_{\sigma(S),N} = \frac{1}{k} \sum_{i=1}^k |\widehat{\sigma(S)}_{N,i} - \sigma(S)| \quad (2.4.3)$$

$$\bar{d}_{\sigma,N} = \frac{1}{k} \sum_{i=1}^k \int_{S^1} |F_{\sigma} - F_{\hat{\sigma}_{N,i}}| d\theta. \quad (2.4.4)$$

Les graphes des points $(N, \bar{d}_{\alpha,N})$, $(N, \bar{d}_{\sigma(S),N})$ et $(N, \bar{d}_{\sigma,N})$ sont présentés dans la Figure 2.8. La vitesse de convergence de $\widehat{\sigma(S)}_N$ est beaucoup plus lente que celles de deux autres estimateurs. Cela peut être observé par l'échelle des valeurs de distance.

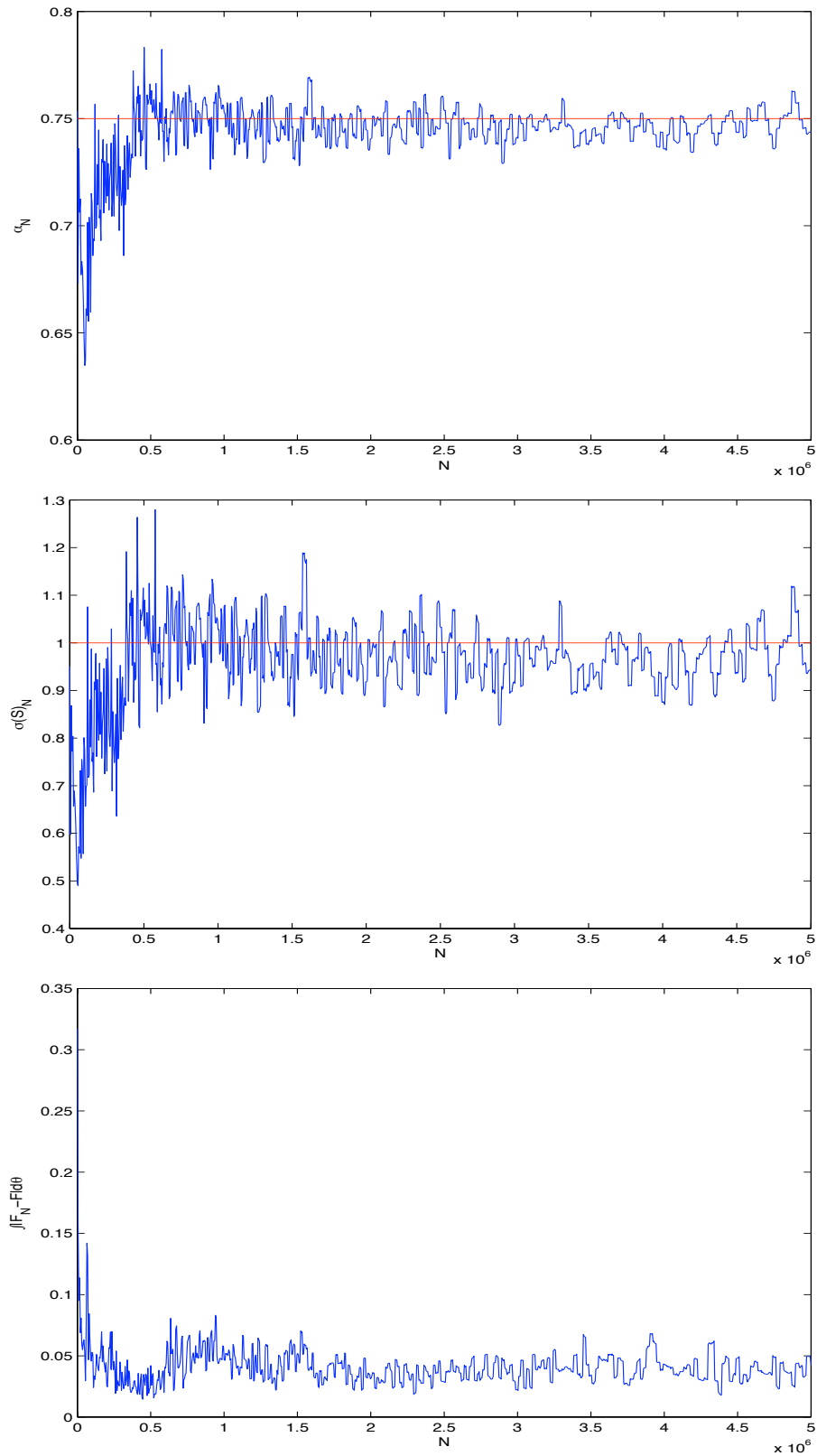


FIG. 2.7: Graphes des points $(N, \hat{\alpha}_N)$ (en haut), $(N, \widehat{\sigma(S)}_N)$ (au milieu) et $(N, \|F_\sigma - F_{\hat{\sigma}_N}\|_1)$ (en bas). Les données simulées sont de loi $\mathcal{S}_2(0.75, \sigma, 0)$ où σ a la densité définie par (1.3.11). La ligne horizontale représente la valeur réelle.

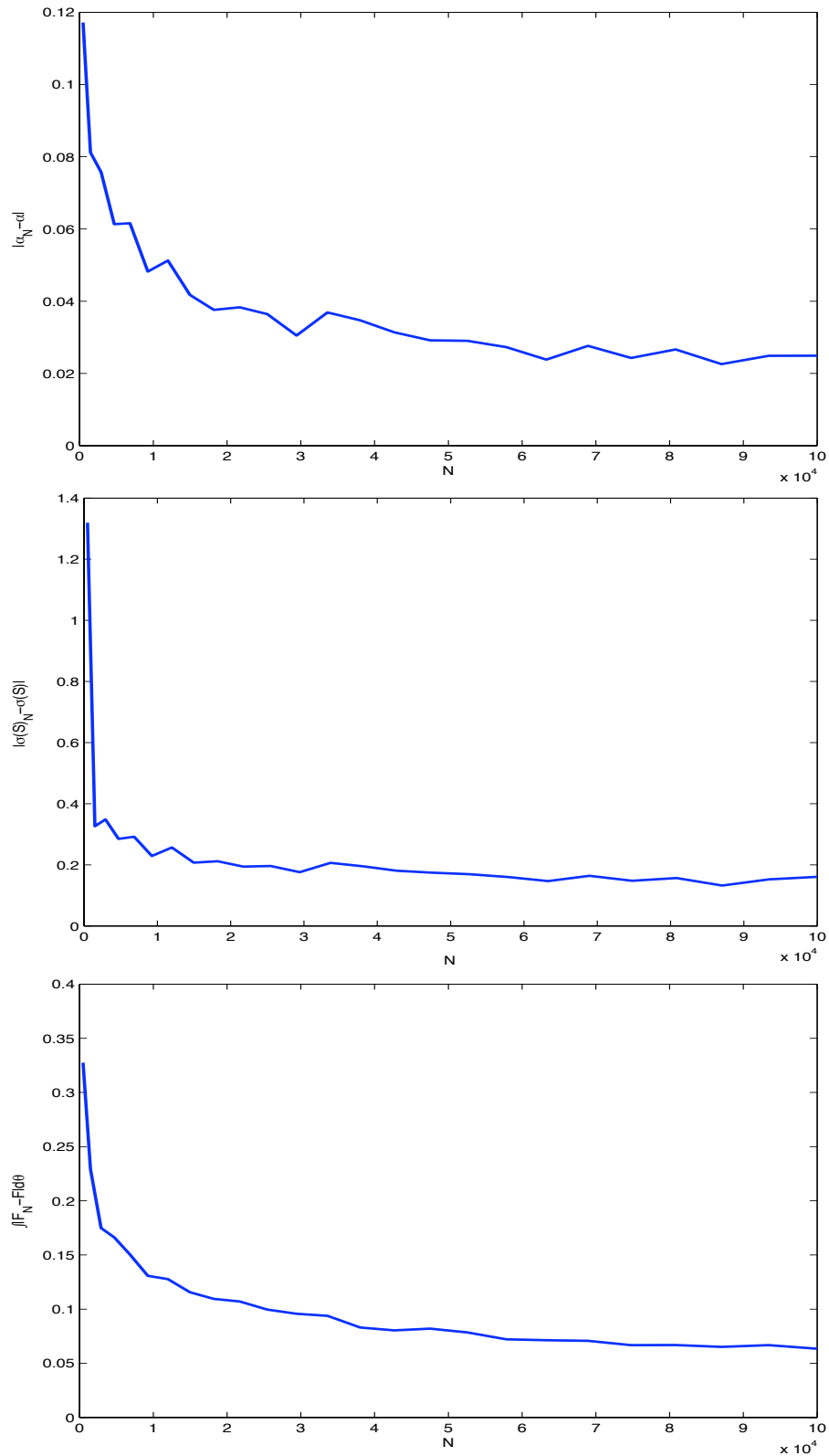


FIG. 2.8: Graphes des points $(N, \bar{d}_{\alpha, N})$ (en haut), $(N, \bar{d}_{\sigma(S), N})$ (au milieu), $(N, \bar{d}_{\sigma, N})$ (en bas). Les données simulées sont de loi $\mathcal{S}_2(0.75, \sigma, 0)$ où σ a la densité définie par (1.3.11).

2.4.4 Comparaison par simulation avec d'autres méthodes

Ces dernières, on s'est beaucoup intéressé aux lois α -stables multivariées dans une variété de problèmes pratiques. La mesure spectrale d'une loi stable porte les informations sur la structure de dépendance du vecteur. Un vecteur α -stable a ses composantes indépendantes si et seulement si sa mesure spectrale est discrète et concentrée sur les intersections des axes et la sphère. Nous présentons ici brièvement deux méthodes d'estimation de la mesure spectrale d'une loi stable proposées par Nolan, Panorska et McCulloch [NPM01]. Ces méthodes sont basées sur la discrétisation de la mesure spectrale. Il existe d'autres méthodes d'estimation, voir par exemple [RX93] [CR95] et [Nag01].

Méthode de la fonction caractéristique empirique

L'idée principale consiste à minimiser la distance entre la fonction caractéristique empirique et la fonction caractéristique réelle par une norme appropriée. L'idée de la méthode vient de la correspondance bijective entre la f.d.r. et sa transformée de Fourier. La méthode générale peut être décrite de façon suivante.

Soit X_1, \dots, X_n un n -échantillon issu d'une loi d -dimensionnelle de fonction caractéristique

$$\Phi(\mathbf{t}) = \mathbf{E} \exp(i\langle X, \mathbf{t} \rangle) = \exp(-I(\mathbf{t})), \quad (2.4.5)$$

où $X = (X^{(1)}, \dots, X^{(d)})$ est un vecteur aléatoire stable, $\mathbf{t} = (t^{(1)}, \dots, t^{(d)})$, et la fonction d'exposant I est donnée par

$$I(\mathbf{t}) = \int_{S^{d-1}} \psi_\alpha(\langle \mathbf{t}, \theta \rangle) \sigma(d\theta)$$

où

$$\psi_\alpha(x) = \begin{cases} |x|^\alpha \left(1 - i \operatorname{sign}(x) \tan \frac{\pi\alpha}{2}\right), & \alpha \neq 1, \\ |x| \left(1 + i \frac{\pi}{2} \operatorname{sign}(x) \ln |x|\right), & \alpha = 1. \end{cases}$$

Soit $\hat{\Phi}_n(\mathbf{t})$ la fonction caractéristique empirique et $\hat{I}_n(\mathbf{t})$ la fonction d'exposant empirique, i.e.

$$\hat{\Phi}_n(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \exp(i\langle X_i, \mathbf{t} \rangle), \quad \hat{I}_n(\mathbf{t}) = -\ln \hat{\Phi}_n(\mathbf{t}).$$

Étant donné une grille $\mathbf{t}_1, \dots, \mathbf{t}_k \in S^{d-1}$, $\hat{I}_n = (\hat{I}_n(\mathbf{t}_1), \dots, \hat{I}_n(\mathbf{t}_k))^T$ est l'estimateur de $I(\cdot)$. Considérons l'approximation discrète de σ ,

$$\sigma^* = \sum_{j=1}^k s_j \delta_{\theta_j}$$

où $s_j = \sigma(A_j)$, $j = 1, \dots, k$, est le poids, et θ_j est un point dans A_j . En général les A_j sont les éléments d'une partition de la sphère unité S^{d-1} de "centre" θ_j . Dans ce cas

$I(\mathbf{t}) = \sum_{j=1}^k \psi_\alpha(\langle \mathbf{t}, \theta_j \rangle) s_j$. Définissons la matrice $k \times k$

$$\Psi = \Psi(\mathbf{t}_1, \dots, \mathbf{t}_k, \theta_1, \dots, \theta_k) = \begin{pmatrix} \psi_\alpha(\langle \mathbf{t}_1, \theta_1 \rangle) & \dots & \psi_\alpha(\langle \mathbf{t}_1, \theta_k \rangle) \\ \dots & \dots & \dots \\ \psi_\alpha(\langle \mathbf{t}_k, \theta_1 \rangle) & \dots & \psi_\alpha(\langle \mathbf{t}_k, \theta_k \rangle) \end{pmatrix},$$

alors $(s_1, \dots, s_k)^T = \Psi^{-1} \hat{I}_n$ est une approximation des masses des éléments contenant $\theta_j, j = 1, \dots, k$. Pour plus de détails voir [NPM01].

Méthode de projection

Cette méthode est une généralisation de la méthode de McCulloch [McC00]. Elle est basée sur les projections unidimensionnelles. Soit X un v.a. α -stable de fonction caractéristique définie par (2.4.5), alors pour tout $\mathbf{t} \in S^{d-1}$, la projection unidimensionnelle $\langle \mathbf{t}, X \rangle$ est une variable aléatoire α -stable de fonction caractéristique $\Phi_{\mathbf{t}}(u) = \mathbf{E} \exp(iu \langle \mathbf{t}, X \rangle)$. Ses paramètres d'échelle et de symétrie sont donnés par

$$\sigma_{\mathbf{t}} = \int_{S^{d-1}} |\langle \mathbf{t}, \theta \rangle|^\alpha \sigma(d\theta),$$

$$\beta_{\mathbf{t}} = \frac{\int_{S^{d-1}} |\langle \mathbf{t}, \theta \rangle|^\alpha \text{sign}(\langle \mathbf{t}, \theta \rangle) \sigma(d\theta)}{\int_{S^{d-1}} |\langle \mathbf{t}, \theta \rangle|^\alpha \sigma(d\theta)}.$$

Considérons l'échantillon X_1, \dots, X_n . Fixons la grille $\mathbf{t}_1, \dots, \mathbf{t}_k \in S^{d-1}$ et estimons les paramètres de la loi de $\langle \mathbf{t}, X \rangle$ pour chaque direction \mathbf{t} par les méthodes unidimensionnelles. On obtient un estimateur de $\hat{I}_n = (\hat{I}_n(\mathbf{t}_1), \dots, \hat{I}_n(\mathbf{t}_k))^T$. Ensuite la procédure est la même que celle de la méthode de la fonction caractéristique empirique.

Performance des estimateurs

Les méthodes d'estimation de mesure spectrale présentées auparavant, méthode de la fonction caractéristique empirique (f.c.e.) et méthode de projection (proj.), sont comparées à l'aide d'une étude de simulation avec notre estimateur basé sur les propriétés des processus ponctuels poissonniens (p.p.p.). Les données α -stables simulées dans la sous-section 1.3.1 sont utilisées. L'estimation par les méthodes de la f.c.e. et de proj. est effectuée en utilisant la fonction MATLAB "mvstablefit" dans le package "STABLE" obtenu sur demande au site <http://www.RobustAnalysis.com/>.

Les estimateurs de mesure spectrale de f.c.e. et de proj. dépendent du paramètre α . En pratique ce paramètre est calculé par des méthodes d'estimation unidimensionnelles. Plus précisément, en appliquant une certaine méthode unidimensionnelle, on estime d'abord les paramètres $\alpha_i, i = 1, \dots, d$, pour chaque dimension des données. Ensuite la moyenne $\frac{1}{d} \sum_{i=1}^d \alpha_i$ est prise comme paramètre α . Pour cette raison la comparaison des méthodes d'estimation de lois stables multivariées est composée de deux parties. Premièrement nous comparons trois méthodes d'estimation de mesure spectrale en utilisant les mêmes valeurs estimées de α . Remarquons que notre estimateur de mesure spectrale ne dépend pas de l'indice α . Deuxièmement nous comparons les paramètres α et $\sigma(S)$ estimés sur des données stables bivariées en utilisant notre méthode et deux autres méthodes unidimensionnelles.

Il n'est pas facile de comparer ces estimateurs de mesure spectrale, car dans les méthodes de la f.c.e. et de proj., il s'agit de choisir une grille de la sphère unité sur laquelle on estime les masses, tandis que le nombre de points du support de mesure spectrale estimée par la méthode de p.p.p. est fixé par rapport à la taille d'échantillon. Ici on prend simplement le nombre de points de grille pour les méthodes de la f.c.e. et de proj. égal au nombre des paquets pour la méthode de p.p.p..

D'abord l'estimation de mesure spectrale sur des échantillons issus de huit lois stables bivariées de taille $N = 3 \times 10^4$ est effectuée. Les méthodes de la f.c.e. et de proj. sont appliquées en utilisant les mêmes valeurs estimées de α . La méthode unidimensionnelle utilisée pour estimer le paramètre α est la méthode de la f.c.e.. Les f.d.r. empiriques de mesure spectrale obtenues sont présentées dans les Figures 2.9 et 2.10.

Ensuite nous étudions la convergence des estimateurs des paramètres des lois stables bivariées. Nous prenons des échantillons issus de deux lois stables bivariées de taille variée de 500 à 3.5×10^4 . Les lois $\mathcal{S}_2(0.75, \sigma_3, 0)$ et $\mathcal{S}_2(1.5, \sigma_1, 0)$ où σ_3 et σ_1 ont les densités définies par (1.3.11) et (1.3.9) sont considérées. La procédure d'estimation précitée est répétée sur $k = 100$ échantillons indépendants pour chaque taille. Nous calculons les moyennes des valeurs estimées $\bar{\alpha}_N$ et $\bar{\sigma}(S)_N$ ainsi que les moyennes des distances des mesures spectrales $\bar{d}_{\sigma,N}$ définies par (2.4.4). Les graphes des points $(N, \bar{\alpha}_N)$, $(N, \bar{\sigma}(S)_N)$ et $(N, \bar{d}_{\sigma,N})$ sont présentés dans la Figure 2.11. Dans les diagrammes concernant les résultats d'estimation des paramètres α et $\sigma(S)$, aux deux premières lignes de la Figure 2.11, nous observons que la méthode de la f.c.e. sous-estime toujours les paramètres. Les résultats obtenus par cette méthode ne dépendent presque pas de la taille d'échantillon. Dans le cas $\alpha > 1$ un biais important est observé. La méthode de p.p.p. sous-estime les paramètres dans le cas $\alpha < 1$ et surestime les paramètres dans le cas $\alpha > 1$. Ces graphes montrent que les convergences des résultats obtenus par la méthode de p.p.p. ont une vitesse plus lente que celles obtenues par la méthode de la f.c.e.. Dans les diagrammes concernant les résultats d'estimation des mesures spectrales, à la dernière ligne de la Figure 2.11, nous observons que la vitesse de convergence de l'estimateur de p.p.p. est entre celles de l'estimateur de f.c.e. et de proj..

A la fin nous comparons des résultats d'estimation des paramètres α et $\sigma(S)$ des lois stables en utilisant la méthode de p.p.p., la méthode de la f.c.e. et la méthode du maximum de vraisemblance (m.v.). Les échantillons issus de huit lois stables bivariées de taille $N = 7000$ sont utilisés. La procédure d'estimation est répétée $k = 100$ fois sur des échantillons indépendants de même loi. La moyenne, le minimum, le maximum et la moyenne d'erreur quadratique, i.e.

$$\text{MEQ}_\theta = \frac{1}{k} \sum_{i=1}^k (\hat{\theta}_i - \theta)^2,$$

de $\hat{\alpha}$ et $\widehat{\sigma(S)}$ sont calculés. Les résultats obtenus sont présentés dans le Tableau 2.6. La dernière colonne de ce tableau présente les moyennes du temps de calcul correspondantes. Nous observons toujours les sous-estimations des estimateurs de f.c.e. et de m.v.. Les moyennes d'erreur quadratique de notre méthode sont beaucoup plus élevées que celles des autres méthodes. Néanmoins nous avons un avantage évident par rapport au temps de calcul.

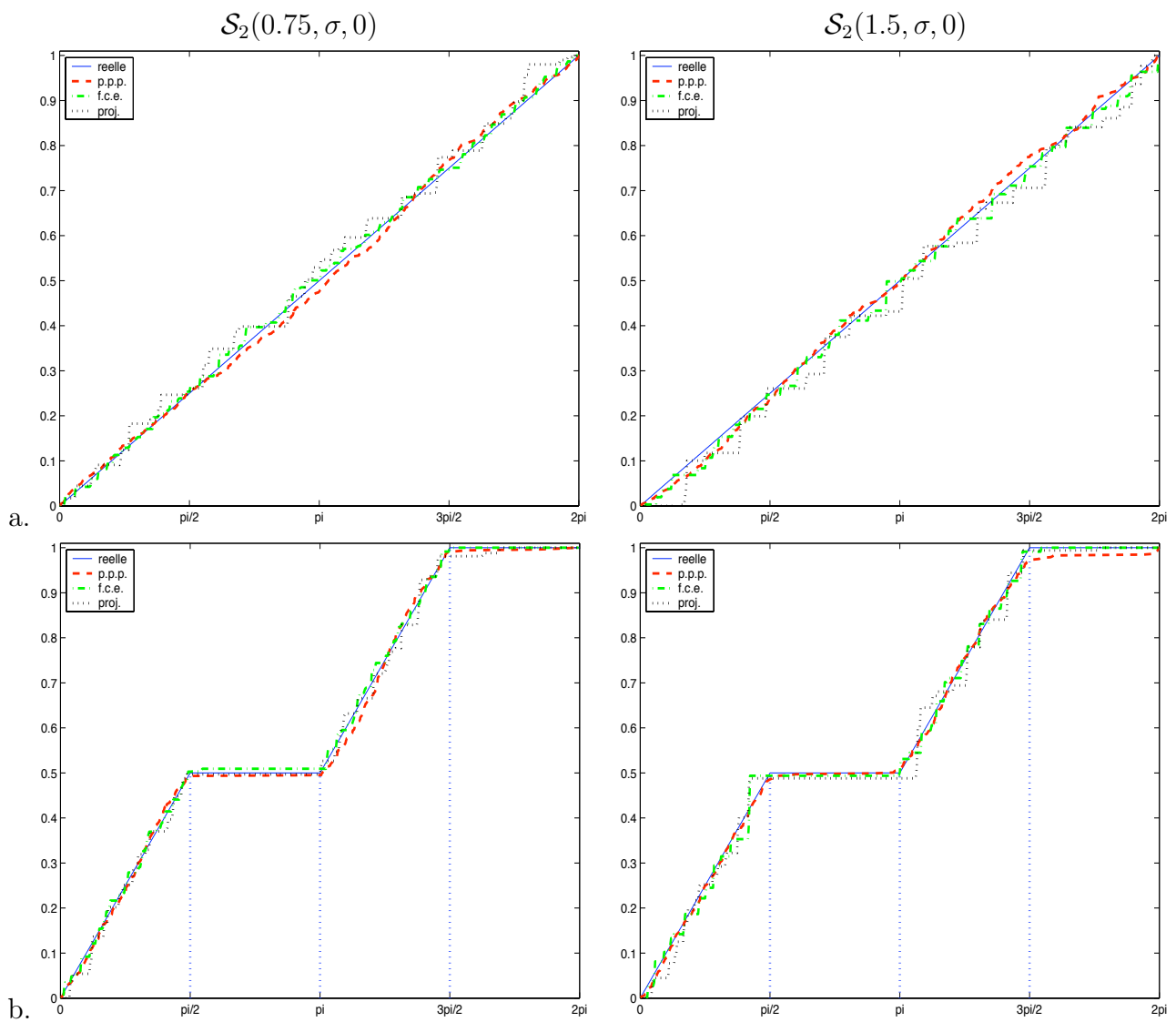


FIG. 2.9: F.d.r. empiriques de mesure spectrale des lois $\mathcal{S}_2(0.75, \sigma, 0)$ (gauche) et $\mathcal{S}_2(1.5, \sigma)$ (droite), σ est uniforme (a), la densités de σ est définie par (1.3.9) (b), $N = 3 \times 10^4$.

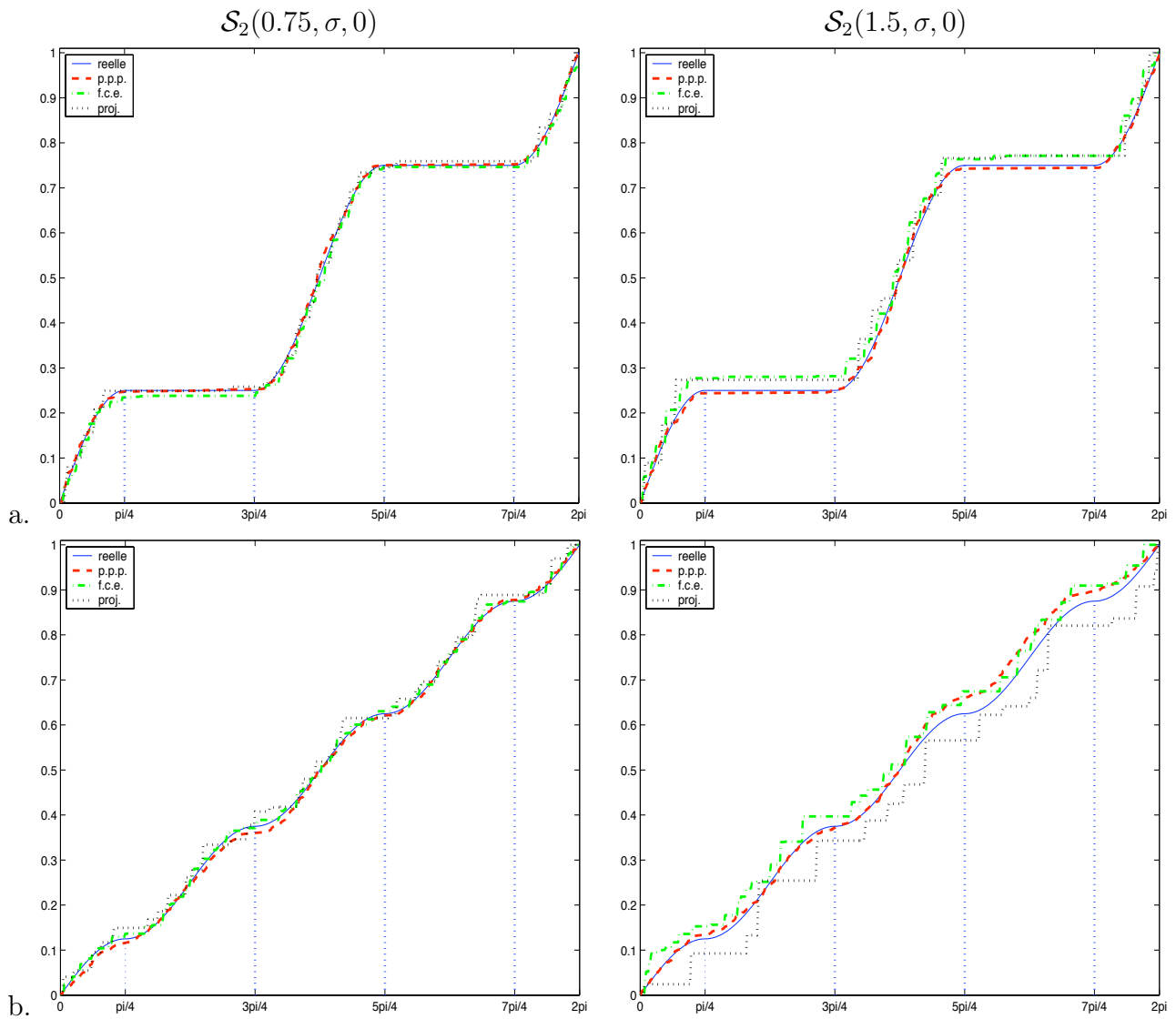


FIG. 2.10: F.d.r. empiriques de mesure spectrale des lois $\mathcal{S}_2(0.75, \sigma, 0)$ (gauche) et $\mathcal{S}_2(1.5, \sigma)$ (droite), la densité de σ est définie par (1.3.10) (a) et (1.3.11) (b), $N = 3 \times 10^4$.

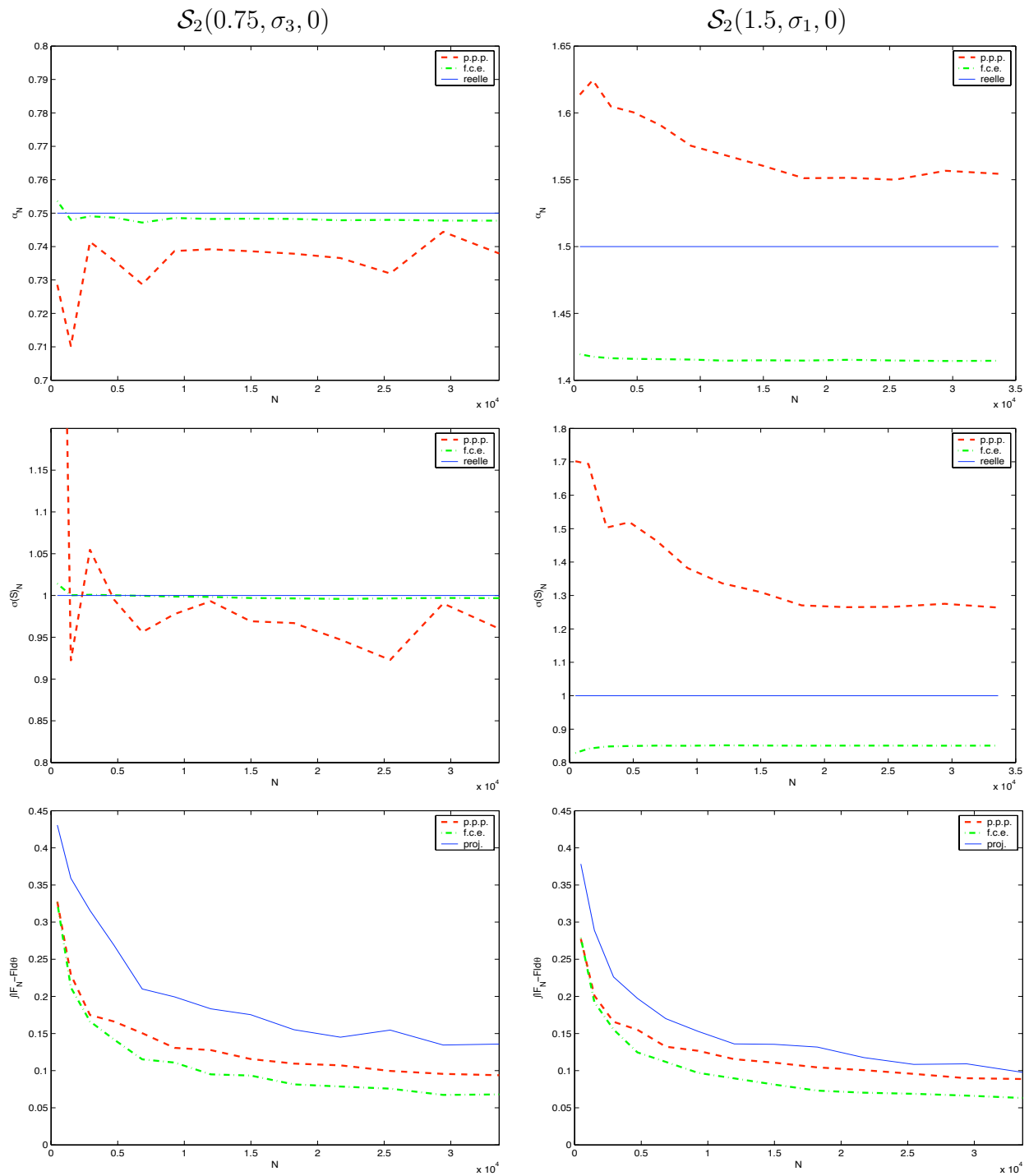


FIG. 2.11: Comparaison des convergences des estimateurs différents. Les échantillons viennent des lois $\mathcal{S}_2(0.75, \sigma_3, 0)$ et $\mathcal{S}_2(1.5, \sigma_1, 0)$, où σ_3 et σ_1 ont les densités définies par (1.3.11) et (1.3.9). Graphes des distances entre les paramètres estimés et réels en fonction de la taille d'échantillon, i.e. $(N, \bar{d}_{\alpha, N})$ (en haut), $(N, \bar{d}_{\sigma(S), N})$ (au milieu), $(N, \bar{d}_{\sigma, N})$ (en bas).

méthode	$\bar{\alpha}$	α_{min}	α_{max}	MEQ_{α}	$\overline{\sigma(S)}$	$\sigma(S)_{min}$	$\sigma(S)_{max}$	MEQ_{σ}	temps de calcul (en seconde)
$\mathcal{S}_2(0.75, \sigma, 0)$ isotropique									
p.p.p.	0.731	0.573	0.925	4.28e-003	0.963	0.417	2.309	1.167e-001	1.870e-002
f.c.e.	0.746	0.712	0.788	1.52e-004	0.999	0.944	1.043	2.774e-004	1.550e-001
m.v.	0.748	0.730	0.769	7.21e-005	0.997	0.963	1.025	1.616e-004	3.207e+001
$\mathcal{S}_2(1.5, \sigma, 0)$ isotropique									
p.p.p.	1.587	1.341	1.881	1.18e-002	1.419	0.817	3.777	1.852e-001	3.812e-002
f.c.e.	1.416	1.366	1.459	3.42e-004	0.850	0.819	0.883	1.951e-004	1.727e-001
m.v.	1.397	1.362	1.439	2.07e-004	0.847	0.819	0.877	1.701e-004	2.981e+001
$\mathcal{S}_2(0.75, \sigma_1, 0)$									
p.p.p.	0.740	0.598	0.899	3.62e-003	0.976	0.469	2.027	8.830e-002	1.653e-002
f.c.e.	0.748	0.712	0.789	2.02e-004	0.998	0.952	1.037	3.273e-004	1.628e-001
m.v.	0.749	0.729	0.774	7.40e-005	0.995	0.963	1.031	1.995e-004	3.500e+001
$\mathcal{S}_2(1.5, \sigma_1, 0)$									
p.p.p.	1.571	1.337	1.962	1.73e-002	1.386	0.733	3.728	2.664e-001	5.699e-002
f.c.e.	1.416	1.385	1.454	2.36e-004	0.851	0.819	0.898	2.388e-004	2.836e-001
m.v.	1.398	1.366	1.424	1.46e-004	0.851	0.815	0.889	2.121e-004	3.307e+001
$\mathcal{S}_2(0.75, \sigma_2, 0)$									
p.p.p.	0.745	0.565	0.972	4.28e-003	1.035	0.421	2.947	1.475e-001	1.613e-002
f.c.e.	0.748	0.722	0.776	1.67e-004	0.998	0.967	1.048	2.862e-004	1.660e-001
m.v.	0.748	0.725	0.777	8.18e-005	0.996	0.970	1.031	1.868e-004	3.519e+001
$\mathcal{S}_2(1.5, \sigma_2, 0)$									
p.p.p.	1.573	1.236	1.993	1.42e-002	1.398	0.569	3.098	1.872e-001	3.256e-002
f.c.e.	1.412	1.360	1.459	4.42e-004	0.851	0.811	0.901	3.880e-004	2.092e-001
m.v.	1.395	1.360	1.423	2.10e-004	0.854	0.816	0.903	3.190e-004	3.390e+001
$\mathcal{S}_2(0.75, \sigma_3, 0)$									
p.p.p.	0.733	0.581	0.877	4.18e-003	0.967	0.437	1.907	1.086e-001	1.840e-002
f.c.e.	0.747	0.710	0.791	1.45e-004	1.000	0.933	1.041	2.705e-004	1.558e-001
m.v.	0.748	0.728	0.771	6.61e-005	0.997	0.954	1.027	1.785e-004	3.190e+001
$\mathcal{S}_2(1.5, \sigma_3, 0)$									
p.p.p.	1.582	1.280	1.908	1.41e-002	1.429	0.643	2.983	1.991e-001	3.157e-002
f.c.e.	1.406	1.357	1.460	3.19e-004	0.853	0.805	0.898	3.024e-004	1.699e-001
m.v.	1.393	1.352	1.434	2.34e-004	0.849	0.805	0.887	2.675e-004	2.887e+001

TAB. 2.6: Tableau de performance des estimateurs des paramètres des lois α -stables bivariées, $N = 7000, r = 0.6, k = 100$.

Chapitre 3

Applications

Dans ce chapitre, les lois à queue régulière sont utilisées pour étudier deux jeux de données. La première section présente l'étude d'un jeu de données financières avec 30 composantes. Nous comparons les résultats d'estimation de notre méthode avec ceux obtenus par la méthode du maximum de vraisemblance pour les lois stables [Nol05]. En choisissant quelques composantes représentatives nous formons des jeux de données multivariées, et nous étudions la structure de leur mesure spectrale. Le comportement de la queue de l'amplitude des données est étudié à la fin. Cette étude sert également à mettre en évidence deux perspectives importantes liées à ce travail.

Dans la deuxième section, nous travaillons sur un jeu de données provenant de l'astronomie. Ces données représentent les perturbations induites par les grandes planètes à la dynamique des comètes. Nous commençons par une analyse exploratoire basée sur les statistiques d'ordre. Les résultats empiriques nous suggèrent des lois à queue lourde. Les lois stables et une famille alternative de lois à queue régulière d'indice $\alpha \geq 2$ sont proposées pour la modélisation. Des tests statistiques sont également effectués pour vérifier les choix de cette modélisation.

3.1 Quelques outils

Avant de commencer la présentation des applications, nous présentons deux outils très utiles pour nos problèmes : une famille de lois à queue régulière et la statistique d'ordre. La famille de lois à queue régulière que nous présentons sera utilisée comme alternative aux lois stables. La statistique d'ordre nous permettra de mettre en place des tests pour les résultats de nos estimations ou de construire quelques outils d'analyse exploratoire.

Rappelons la définition de la régularité d'une loi unidimensionnelle qui peut être déduite de la définition 1.1.5.. La variable aléatoire X dans \mathbb{R}^1 a une loi à queue régulière si il existe les constantes $\alpha > 0$, $p, q \geq 0$ telles que

$$\begin{cases} \lim_{x \rightarrow \infty} \frac{x^\alpha}{L(x)} \mathbf{P}\{X > x\} & = p, \\ \lim_{x \rightarrow \infty} \frac{x^\alpha}{L(x)} \mathbf{P}\{X < -x\} & = q, \end{cases} \quad (3.1.1)$$

où $L(x)$ est une fonction à variation lente. Ce modèle décrit non seulement la vitesse de

décroissance de la queue, mais aussi la symétrie de la loi. La flexibilité de cette famille de lois s'adapte particulièrement à notre cas. Les paramètres qu'on va estimer sont

$$\alpha, \quad \beta = \frac{p-q}{p+q} \quad \text{et} \quad \gamma = p+q.$$

Le paramètre α contrôle le kurtosis, β contrôle la symétrie, γ contrôle l'échelle. La méthode d'estimation présentée dans le premier chapitre est mise en place. L'algorithme et le programme sont présentés dans les annexes A et B. Une famille importante de lois à queue régulière est celle de lois α -stables, présentée dans la sous-section 1.3.1. La condition de régularité (1.3.6) est un cas particulier de la condition (3.1.1) avec $L(x) = 1$. Pour $\alpha \geq 2$, nous proposons une famille de lois dont la densité est définie par

$$f(z) = \frac{C_{\kappa,\alpha}}{1 + \kappa|z - \omega|^{\alpha+1}}, \quad (3.1.2)$$

où $C_{\kappa,\alpha}$ est une constante de normalisation, κ est le paramètre d'échelle, ω est le paramètre de position et α est l'indice de queue. Cette famille de lois a une queue régulière d'indice α , i.e.

$$\mathbf{P}\{|X| > x\} \sim Cx^{-\alpha}, \quad \text{quand } x \rightarrow \infty.$$

Cette équivalence est un cas particulier de la condition (3.1.1) avec $p = q = C$, $L(x) = 1$. Comme ces lois sont symétriques par rapport à ω , elles peuvent être utilisées pour modéliser les données issues d'une loi à queue régulière dont le paramètre de symétrie β est proche de 0. Les moments d'ordre un et deux existent si $\alpha \geq 2$. Soit X une variable aléatoire à densité définie par (3.1.2). En posant

$$C_1 = \int_0^\infty \frac{1}{1+x^{\alpha+1}} dx, \quad C_2 = \int_0^\infty \frac{x^2}{1+x^{\alpha+1}} dx,$$

des relations entre les paramètres et les moments de X sont les suivantes

$$C_{\kappa,\alpha} = \frac{\kappa^{\frac{1}{\alpha+1}}}{2C_1}, \quad \mathbf{E}X = \omega, \quad \text{Var}X = \frac{C_2}{C_1 \kappa^{\frac{2}{\alpha+1}}}.$$

L'estimation des paramètres de ces lois est effectuée en plusieurs étapes. Tout d'abord, l'indice de queue α est estimé par l'algorithme précédent. Ensuite le paramètre de position ω est estimé par la moyenne empirique d'échantillon. A la fin, la constante de normalisation $C_{\kappa,\alpha}$ et le paramètre d'échelle κ sont estimés en utilisant la méthode des moments.

Soit Z_1, \dots, Z_n une suite de variables aléatoires i.i.d. et $F(z) = \mathbf{P}\{Z \leq z\}$, $z \in \mathbb{R}$, la fonction de répartition correspondante. Considérons aussi Σ_n , l'ensemble des permutations sur $\{1, \dots, n\}$. La statistique d'ordre de l'échantillon (Z_1, \dots, Z_n) est représentée par les éléments de cet échantillon arrangés en ordre croissant et on les note par $(Z_{(1,n)}, \dots, Z_{(n,n)})$. Ainsi $Z_{(1,n)} \leq \dots \leq Z_{(n,n)}$ et il existe une permutation aléatoire $\sigma_n \in \Sigma_n$ telle que

$$(Z_{(1,n)}, \dots, Z_{(n,n)}) = (Z_{\sigma_n(1)}, \dots, Z_{\sigma_n(n)}). \quad (3.1.3)$$

Dans la suite, quelques résultats classiques de la littérature sont présentés [Dav81, DJ06]. Si F est continue, alors $Z_{(1,n)} < \dots < Z_{(n,n)}$ presque sûrement et la permutation

σ_n de la définition (3.1.3) est unique. Si Z_1 a la densité f , alors la densité de la loi commune des statistiques d'ordre est donnée par

$$n! \mathbb{I}\{z_1 < \dots < z_n\} f(z_1) \dots f(z_n).$$

La statistique d'ordre permet de déterminer la plupart des caractéristiques de la loi d'où est issu l'échantillon. En particulier pour définir la fonction de répartition empirique, il suffit d'ordonner l'échantillon par valeurs croissantes. Pour un nombre réel $q \in (0, 1)$, la statistique $Z_{([nq]+1, n)}$ (où $[nq]$ désigne la partie entière de nq) s'appelle *le quantile empirique d'ordre q* de l'échantillon. Nous présentons deux résultats importants concernant le quantile empirique.

Supposons que F est continue et qu'il existe une solution unique z_q pour l'équation $F(z) = q$ avec $q \in (0, 1)$. La solution z_q est appelée le q -quantile de F . Ainsi soit $\{k(n), n \geq 1\}$ une suite des entiers telle que $1 \leq k(n) \leq n$ et $\lim_{n \rightarrow \infty} \frac{k(n)}{n} = q$. Alors la suite des quantiles $\{Z_{(k(n), n)}, n \geq 1\}$ converge presque sûrement vers z_q . Ce premier résultat montre que le quantile empirique est un estimateur du quantile théorique.

Si la densité de Z_1 , notée f , est continue et strictement positive, alors $\{Z_{([nq]+1, n)}\}$ converge en loi vers z_q :

$$\sqrt{n}(Z_{([nq]+1, n)} - z_q) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{q(1-q)}{f(z_q)^2}\right). \quad (3.1.4)$$

Ce deuxième résultat permet de calculer des intervalles de confiance et construire des tests d'hypothèses. D'un point de vue pratique, une analyse basée sur les quantiles empiriques permet d'examiner la queue, la symétrie et la structure générale spatiale de la loi des données. D'un point de vue théorique, les mathématiques sur lesquelles cet outil est basé permettent une analyse des données rigoureuse.

3.2 Première application : étude des cours des 30 valeurs de l'indice DJIA

Les cours ajustés de clôture quotidiens pour les 30 valeurs de l'indice Dow Jones Industrial Average (DJIA) ont été enregistrés entre le 3 janvier 2000 et le 31 décembre 2004. Ils sont transformés séparément pour chaque titre par $y_t = \ln(x_{t+1}/x_t)$, étant donné $N = 1256$ valeurs de cours par titre. Nous obtenons un ensemble de données 30-dimensionnelles de taille 1255. Cet ensemble de données a déjà été traité dans [Nol05], l'auteur proposant des lois sous-gaussiennes stables pour la modélisation. Ici nous utilisons des lois à queue régulière pour modéliser ce jeu de données.

Nous commençons par un ajustement de loi à queue régulière unidimensionnelle sur chaque dimension des données en utilisant la méthode d'estimation présentée dans la section 2.1 (voir aussi l'annexe A). Si les paramètres caractéristiques α sont significativement différents entre eux, alors ces données ne sont pas issues d'une loi à queue régulière multivariée. Si les paramètres d'asymétrie β ne sont pas tous proches de 0, alors la loi n'est pas symétrique et la mesure spectrale du vecteur aléatoire ne devrait pas être uniforme.

Ensuite nous prenons trois composantes de l'indice DJIA pour étudier le comportement de la mesure spectrale. La mesure spectrale contient des informations essentielles sur le vecteur, en particulier sur la structure de dépendance entre les titres individuels qui composent l'indice DJIA. Supposons que X est un vecteur aléatoire d -dimensionnel et $X \in \text{VR}(\alpha, \sigma)$ avec σ uniforme. D'après la proposition 1.2.4. toute projection k -dimensionnelle de X , $1 \leq k < d$, forme un vecteur aléatoire de loi à queue régulière de mesure spectrale uniforme. Donc on peut étudier l'uniformité de la mesure spectrale des données multivariées en examinant les lois bivariées. Si on ne peut pas le faire pour toutes les projections, on peut vérifier quelques paires visuellement. C'est-à-dire que nous estimons d'abord les mesures spectrales des données de petite dimension formées par les composantes choisies. Ensuite nous comparons ces mesures spectrales estimées avec la mesure uniforme sur la sphère unité en regardant les nuages ou les histogrammes des points du support de mesure correspondante.

A la fin de cette section nous considérons la loi d'amplitude du vecteur aléatoire, i.e. la loi de la norme des données. Soit X un v.a. de loi à queue régulière. D'après (1.1.5) la norme $\|X\|$ a une loi à queue régulière de même indice caractéristique. Nous utilisons le paramètre α estimé du vecteur 30-dimensionnel comme indice de la queue de la loi de $\|X\|$. La loi de type Paréto est utilisée pour ajuster la distribution des normes des données.

3.2.1 Données unidimensionnelles

Nous effectuons la procédure d'estimation des paramètres des lois à queue régulière unidimensionnelles sur chaque composante de l'indice DJIA pour obtenir

$$\hat{\Theta}_i = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i), \quad i = 1, \dots, 30.$$

Puisque les données sont considérées comme des échantillons de variables aléatoires i.i.d., la permutation des données est aussi i.i.d.. On répète la procédure d'estimation $k = 1000$ fois sur les données permutées aléatoirement pour obtenir

$$\hat{\Theta}_{ij} = (\hat{\alpha}_{ij}, \hat{\beta}_{ij}, \hat{\sigma}_{ij}), \quad i = 1, \dots, 30, \quad j = 1, \dots, k.$$

Alors

$$\bar{\Theta}_i = (\bar{\alpha}_i, \bar{\beta}_i, \bar{\sigma}_i) = \frac{1}{k} \sum_{j=1}^k \hat{\Theta}_{ij}$$

donnent une estimation des paramètres α , β et σ . Les propriétés de ces estimateurs sont présentées dans la section 2.3.

Les résultats d'estimation des paramètres pour chaque composante de l'indice DJIA sont présentés dans le Tableau 3.1. La Figure 3.1 présente les valeurs obtenues de $\bar{\alpha}_i$ en comparant avec les résultats d'estimation de α par la méthode du maximum de vraisemblance (m.v.) sous l'hypothèse que les données viennent d'une loi α -stable. Notons ces valeurs par $\alpha_{i,stab}$, $i = 1, \dots, 30$. Les valeurs de $\bar{\alpha}_i$ sont toutes plus grandes que $\alpha_{i,stab}$. Il y a une corrélation entre ces deux résultats. Le manque de données pourrait être la cause de ce phénomène. Nous avons observé dans la sous-section 2.4.4 que la méthode

proposée surestime le paramètre α , tandis que la méthode du m.v. sous-estime α dans le cas où $\alpha > 1$. Malgré ces différences les résultats indiquent que les lois des données pour chaque composante ont des indices de queue voisins l'un de l'autre. Dans ce cas l'ensemble de données 30-dimensionnelles peut être considéré comme un échantillon de loi à queue régulière multivariée. Les résultats d'estimation des paramètres β sont présentés dans la Figure 3.2. La plupart des valeurs sont autour de 0. Cependant il existe des points qui s'éloignent de 0. Cela indique que les lois de chaque composante ne sont pas toutes symétriques.

Pour vérifier ces résultats, les tests statistiques des quantiles $q_{0.99}$ et de Kolmogorov-Smirnov (KS) ont été construits. Le test sur les quantiles vérifie si les quantiles empiriques $\hat{z}_{0.99}$ de chaque composante viennent d'une loi de type Paréto ou d'une loi stable avec les paramètres correspondants. En même temps le test KS est déterminé plus par le corps de la distribution que par la queue. Les hypothèses nulles sont respectivement que les données sont issues d'une loi dont la densité est définie par (3.1.2) ou d'une loi stable. En conclusion, dans la plupart des cas, les deux approches sont validées par les tests statistiques mis en place. Nous sommes devant la première perspective mentionnée au tout début du chapitre : quel est le meilleur choix du modèle pour un jeu de données ? A priori, pour les données présentées, le maximum de vraisemblance pour les lois stables a des "meilleurs" résultats aux tests que notre méthode. Les différences entre ces deux méthodes pourraient être expliquées par le fait que notre modèle ne prend pas en compte l'asymétrie. Néanmoins, ce que nous proposons est un modèle ayant une densité de probabilité simple avec des résultats aux tests statistiques au moins comparables.

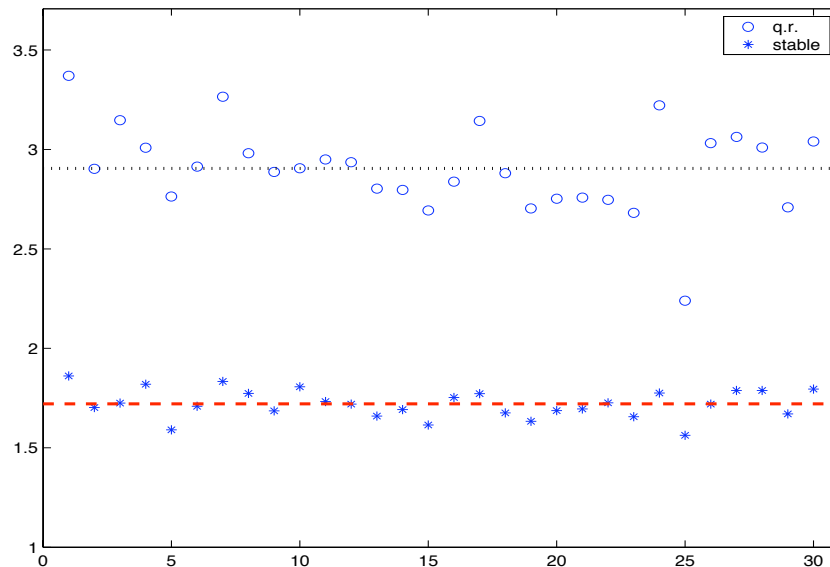


FIG. 3.1: Les paramètres α estimés pour les données de 30 composantes de l'indice DJIA. L'ordre est donné dans le tableau 3.1. Les lignes droites horizontales représentent les moyennes des α estimées de modèle de loi à queue régulière (q.r.) (en pointillés) et de modèle de loi stable utilisant la méthode d'estimation du maximum de vraisemblance (en tirets).

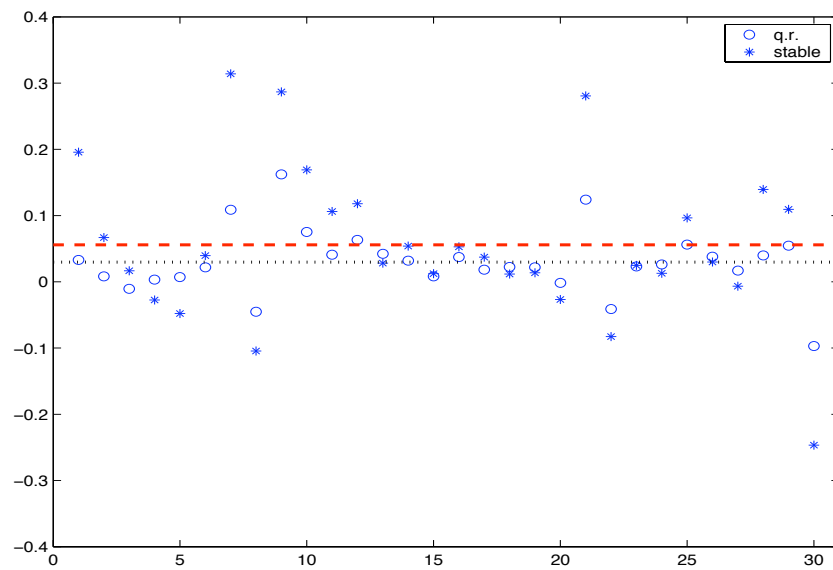


FIG. 3.2: Les paramètres β estimés pour les données de 30 composantes de l'indice DJIA. L'ordre est donné dans le tableau 3.1. Les lignes droites horizontales représentent les moyennes des β estimées de modèle de loi à queue régulière (q.r.) (en pointillés) et de modèle de loi stable utilisant la méthode d'estimation du maximum de vraisemblance (en tirets).

	code	$\bar{\alpha}$	α_{stab}	$\bar{\beta}$	β_{stab}	$\bar{\sigma}$	σ_{stab}	$\bar{\delta}$	δ_{stab}	$q_{0.99}$	KS		
1	AA	3.37	1.86	0.03	0.20	2.62e-006	6.11e-005	-1.25e-004	3.14e-005	0.60	0.74	0.02	0.41
2	AIG	2.90	1.70	0.02	0.07	5.54e-006	1.34e-004	-3.10e-005	4.98e-005	0.84	0.68	0.05	0.28
3	AXP	3.15	1.72	0.01	0.02	4.14e-006	1.52e-004	9.04e-005	2.56e-004	0.18	0.29	0.02	0.17
4	BA	3.01	1.82	0.00	-0.03	5.32e-006	6.91e-005	2.66e-004	5.06e-004	0.27	0.84	0.57	0.75
5	BAC	2.76	1.59	0.01	-0.05	8.37e-006	2.45e-004	6.85e-004	5.07e-004	0.77	0.46	0.04	0.53
6	C	2.91	1.71	0.03	0.04	6.87e-006	1.50e-004	2.89e-004	3.93e-004	0.44	0.48	0.39	0.61
7	CAT	3.26	1.83	0.11	0.31	2.17e-006	5.70e-005	6.65e-004	8.68e-004	0.86	0.78	0.32	0.74
8	CVX	2.98	1.77	-0.05	-0.10	2.02e-006	4.92e-005	3.17e-004	2.21e-004	0.24	0.74	0.31	0.94
9	DD	2.89	1.69	0.15	0.29	5.53e-006	1.41e-004	-9.61e-005	3.07e-004	0.64	0.27	0.00	0.65
10	DIS	2.91	1.81	0.09	0.17	9.43e-006	8.95e-005	-1.90e-005	2.01e-004	0.28	0.76	0.32	0.81
11	GE	2.95	1.73	0.05	0.11	5.93e-006	1.23e-004	-1.68e-004	-5.59e-005	0.80	0.85	0.02	0.49
12	GM	2.94	1.72	0.06	0.12	6.84e-006	1.45e-004	-3.29e-004	-5.25e-005	0.54	0.47	0.20	0.97
13	HD	2.80	1.66	0.05	0.03	1.49e-005	2.58e-004	-3.12e-004	-1.08e-004	0.99	0.68	0.07	0.45
14	HPQ	2.80	1.69	0.03	0.05	2.59e-005	3.02e-004	-5.66e-004	-4.08e-004	0.60	0.50	0.11	0.42
15	IBM	2.69	1.61	0.02	0.01	1.37e-005	2.52e-004	-1.04e-004	-1.67e-004	0.81	0.50	0.01	0.16
16	INTC	2.84	1.75	0.03	0.05	2.77e-005	2.40e-004	-4.81e-004	-3.04e-005	0.47	0.63	0.28	0.96
17	JNJ	3.14	1.77	0.02	0.04	1.29e-006	5.43e-005	3.17e-004	4.31e-004	0.08	0.37	0.02	0.16
18	JPM	2.88	1.68	0.04	0.01	1.18e-005	2.34e-004	-2.54e-005	-2.79e-004	0.91	0.75	0.05	0.06
19	KO	2.70	1.63	0.02	0.01	7.35e-006	1.58e-004	-1.75e-004	-1.40e-004	0.47	0.40	0.08	0.61
20	MCD	2.75	1.69	-0.00	-0.03	8.55e-006	1.42e-004	-1.19e-004	-9.57e-005	0.25	0.43	0.06	0.46
21	MMM	2.76	1.70	0.14	0.28	5.70e-006	1.03e-004	5.24e-004	7.33e-004	0.93	0.57	0.09	0.69
22	MRK	2.75	1.73	-0.03	-0.08	8.29e-006	1.05e-004	-4.46e-004	-2.84e-004	0.12	0.51	0.23	0.83
23	MSFT	2.68	1.66	0.02	0.02	1.92e-005	2.45e-004	-5.25e-004	-3.08e-004	0.12	0.16	0.06	0.45
24	PFE	3.22	1.78	0.02	0.01	2.06e-006	8.26e-005	-7.74e-005	1.07e-004	0.10	0.33	0.11	0.37
25	PG	2.24	1.56	0.05	0.10	2.96e-005	1.95e-004	1.05e-004	7.43e-004	0.15	0.16	0.01	0.70
26	T	3.03	1.72	0.04	0.03	5.40e-006	1.49e-004	-3.33e-004	-3.25e-004	0.40	0.50	0.08	0.61
27	UTX	3.06	1.79	0.02	-0.01	3.93e-006	7.92e-005	4.58e-004	6.74e-004	0.31	0.99	0.00	0.19
28	VZ	3.01	1.79	0.06	0.14	4.53e-006	8.05e-005	-1.70e-004	-9.75e-005	0.37	0.73	0.05	0.56
29	WMT	2.71	1.67	0.06	0.11	1.15e-005	1.73e-004	-1.62e-004	-5.13e-005	0.88	0.76	0.05	1.00
30	XOM	3.04	1.80	-0.11	-0.25	1.81e-006	4.48e-005	3.10e-004	4.67e-005	0.05	0.99	0.25	0.77

TAB. 3.1: Résultats d'estimation et des tests pour les 30 composantes de l'indice DJIA. Les graphes des valeurs des $\bar{\alpha}$, α_{stab} , $\bar{\beta}$ et β_{stab} avec les titres correspondants sont dans les Figures 3.1 et 3.2.

3.2.2 Données tri-variées

Nous choisissons trois composantes représentatives, AIG, C et DIS, qui ont les valeurs de $\bar{\alpha}_i$ très proches de la moyenne. Notons X_A , X_C et X_D les variables aléatoires correspondantes. Supposons que le vecteur (X_A, X_C, X_D) forme un v.a. de loi à queue régulière. Considérons les v.a. 2-dimensionnels (X_A, X_C) , (X_A, X_D) et (X_C, X_D) . Ce sont des projections du v.a. (X_A, X_C, X_D) transformées par T_2 définie par (1.2.14), i.e.

$$\begin{aligned} T_k : \mathbb{R}^d \setminus \{\mathbf{x} \mid x_1 = \dots = x_k = 0\} &\rightarrow \mathbb{R}^k, \\ \mathbf{x} &\mapsto \mathbf{x}_k. \end{aligned}$$

Ici $d = 3, k = 2$. D'après la proposition 1.2.3. chaque projection suit une loi à queue régulière de même indice α que celui de (X_A, X_C, X_D) . Les valeurs de paramètre estimées sur les combinaisons des données bivariées et tri-variées sont présentées dans le Tableau 3.2. Les paramètres α sont plus élevés que ceux estimés sur les données unidimensionnelles. D'après les exemples présentés dans le deuxième chapitre, nous avons besoin d'un plus grand échantillon pour estimer les paramètres d'une loi multidimensionnelle. Sinon l'estimateur proposé a toujours une tendance à surestimer le paramètre α . Néanmoins les valeurs de α estimées sur les données ayant la même dimension sont proches les unes des autres. Cela indique que ces trois composantes forment un vecteur aléatoire de loi à queue régulière.

Un avantage de notre méthode est que l'estimateur de la mesure spectrale ne dépend pas de α . Donc les résultats d'estimation de mesure spectrale donnent des informations significatives malgré les différences entre les paramètres α estimés sur des données ayant la dimension différente. Notons les mesures spectrales estimées sur les données bivariées respectivement par $\hat{\sigma}_{A,C}$, $\hat{\sigma}_{A,D}$ et $\hat{\sigma}_{C,D}$. La Figure 3.3 présente les graphes des nuages des données bivariées et les histogrammes des points du support de la mesure spectrale correspondante. Nous observons que ces mesures spectrales sont loin d'être uniformes. Le graphe des données tri-variées est présenté dans la Figure 3.4 a). Les diagrammes b) et c) présentent le graphe de nuage et les histogrammes des points du support de la mesure spectrale estimée. Les histogrammes sont en coordonnées sphériques : $\theta \in [0, \pi]$ représente le zénith, $\phi \in [0, 2\pi)$ représente l'azimut. Soit σ une mesure uniforme sur la sphère unité S^2 . La densité de σ en coordonnées sphériques est $f(\theta, \phi) = \frac{\sin \theta}{4\pi}$. Le diagramme d) dans la Figure 3.4 présente la densité de σ , qui nous indique ainsi que la mesure spectrale du v.a. (X_A, X_C, X_D) n'a pas l'allure d'une loi uniforme. Cela coïncide avec les résultats sur les données bivariées. La structure de mesure spectrale présentée ici exclut le modèle de lois avec mesure spectrale uniforme. Donc les lois stables isotropiques ne sont pas applicables pour modéliser ces données multidimensionnelles ainsi que leur norme, malgré le fait que le modèle de lois stables est validé par les tests dans le cas unidimensionnel. Dans ce cas nous n'avons pas une loi simple pour la norme du vecteur aléatoire. Ce problème est étudié dans le paragraphe suivant.

3.2.3 Amplitude des données

Si nous considérons nos données comme 1255 réalisations d'un vecteur aléatoire 30-dimensionnel, alors nous obtenons 1255 réalisations de la norme du vecteur.

code	α	σ	δ
(AIG, C)	4.07	1.61e-006	(-0.31, 2.89)e-004
(AIG, DIS)	3.75	3.48e-006	(-0.31, -0.19)e-004
(C, DIS)	3.43	9.27e-006	(2.89, -0.19)e-004
(AIG, C, DIS)	3.84	5.93e-006	(-0.31, 2.89, -0.19)e-004

TAB. 3.2: Résultats d'estimation des paramètres des lois à queue régulière pour les données de trois composantes de l'indice DJIA : AIG, C et DIS.

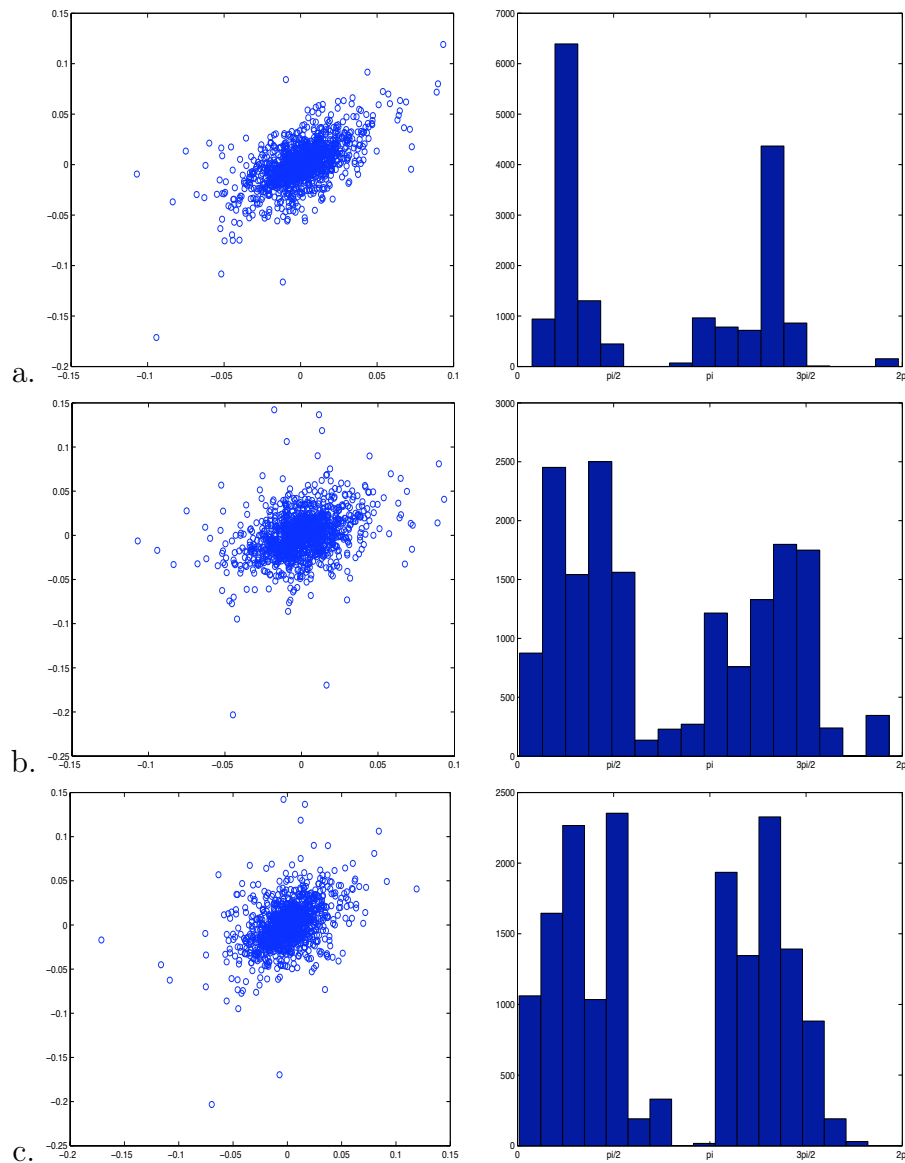


FIG. 3.3: Comparaison deux par deux des données de AIG, Citigroup et DIS, (AIG, C) (a), (AIG, DIS) (b) et (C, DIS) (c). Les nuages de données bivariées (gauche), les histogrammes des points du support de la mesure spectrale estimée correspondante $\hat{\sigma}_{A,C}$, $\hat{\sigma}_{A,D}$ et $\hat{\sigma}_{C,D}$ (droite).

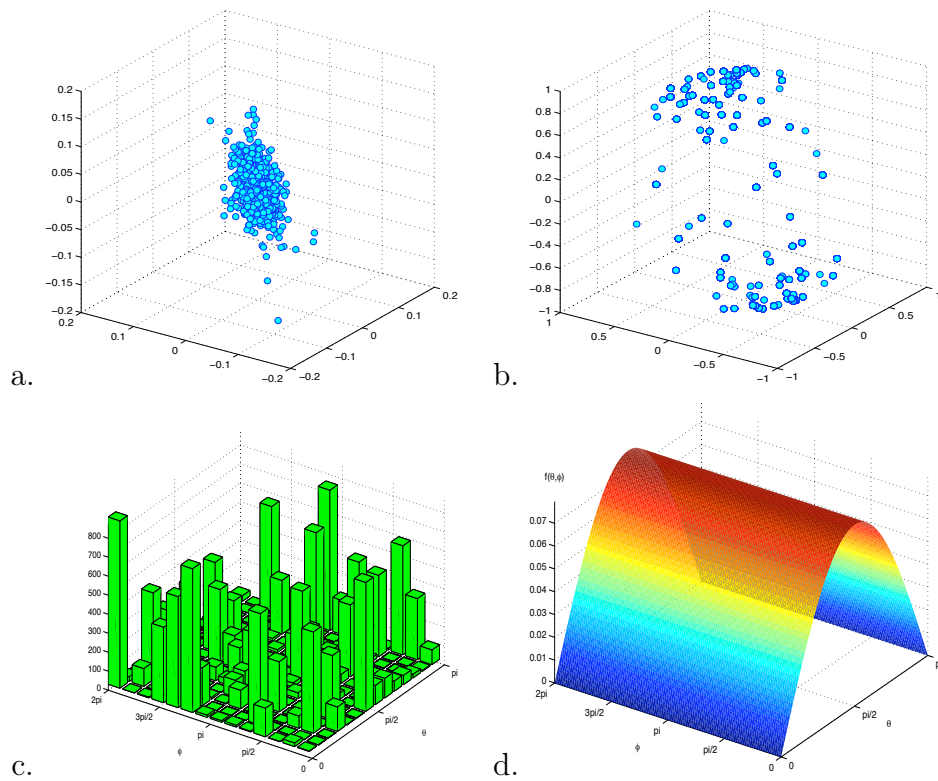


FIG. 3.4: Résultats d'estimation de la mesure spectrale des données 3-dimensionnelles de AIG, Citigroup et DIS. Le nuage des données 3-dimensionnelles (AIG, C, DIS) (a), le support de la mesure spectrale estimée (b), les histogrammes des points du support de la mesure spectrale estimée en coordonnées sphériques (c), la densité d'une mesure uniforme sur la sphère S^2 en coordonnées sphériques (d).

Soit $X = (X^{(1)}, \dots, X^{(d)})$ un v.a. d -dimensionnel. L'amplitude de X est définie par

$$R = \|X\| = \sqrt{(X^{(1)})^2 + \dots + (X^{(d)})^2}.$$

Si X a une loi α -stable isotropique, alors $R^2 \stackrel{\mathcal{L}}{=} AT$ où $A \sim \mathcal{S}_1(\alpha/2, (1, \gamma), 0)$ et T est une variable aléatoire de loi de χ^2 de degré de liberté d . Dans ce cas on peut déduire une forme explicite pour la fonction de répartition de R . Ce modèle est utilisé dans [Nol05]. Par rapport au résultat précédemment obtenu nous ne pouvons pas appliquer le modèle de lois stables : la mesure spectrale n'étant pas uniforme, ceci implique que même si le v.a. X est stable il n'est pas isotropique, donc R n'a pas une forme simple pour la fonction de répartition, raison pour laquelle nous n'utilisons pas le modèle de lois stables dans cette situation.

Par ailleurs si X est un v.a. à queue régulière, alors $R = \|X\|$ est une variable aléatoire positive de loi à queue régulière de même indice caractéristique d'après (1.1.5). Nous estimons les paramètres d'une loi 30-dimensionnelle à queue régulière d'où est issu l'échantillon, et nous utilisons ces paramètres pour modéliser l'amplitude des données. Les résultats d'estimation sont $\bar{\alpha}_X = 4.46$ et $\bar{\sigma}_X = 4.93 \times 10^{-5}$.

Nous testons les queues de distribution pour l'amplitude. La loi de type Paréto tronquée est utilisée. Les résultats dans le Tableau 3.3 montrent que la loi de type Paréto est souhaitable pour les queues de la distribution de ce jeu de données. Mais cette loi ne donne pas le corps de distribution.

La famille de lois stables non gaussiennes est un sous-ensemble des lois à queue régulière ayant l'indice de queue α inférieur à 2. Si la valeur estimée de α est supérieure à 2, le modèle de lois stables n'est plus applicable. C'est-à-dire que l'approche développée dans cette thèse nous aide à diagnostiquer le caractère stable ou non-stable d'une distribution ajustable à un jeu de données. Dans ce cas où $\alpha > 2$ nous ne sommes capable de fournir que des informations concernant uniquement la queue de la distribution. Nous sommes donc devant la deuxième perspective annoncée au commencement du chapitre : selon le problème, trouver la loi à queue régulière qui ajuste au mieux un jeu de données.

modèle	$q_{0.9}$	$q_{0.95}$	$q_{0.99}$
Paréto	0.589	0.718	0.661

TAB. 3.3: Les p -valeurs obtenues du test des quantiles pour l'amplitude des données.

3.3 Deuxième application : étude des perturbations planétaires des comètes du nuage de Oort

3.3.1 Introduction

Le mouvement des comètes est l'un des phénomènes les plus difficiles à modéliser dans la mécanique céleste. Il est connu que cette dynamique est fortement chaotique. Il s'agit là, d'un cas particulier du problème des N -corps. Nous considérons ici six corps célestes, le Soleil, les planètes Jupiter, Saturne, Uranus et Neptune et une comète. Les trajectoires des comètes sont influencées par les perturbations planétaires au cours des "rencontres proches" avec les planètes. Ces perturbations constituent le mécanisme principal qui influence les trajectoires des comètes. L'intégration numérique directe de ce problème n'est pas possible à cause du temps de calcul.

A la recherche d'une approche alternative, nous pouvons profiter du fait que les perturbations planétaires sur les comètes du nuage de Oort sont non corrélées. En fait, les périodes orbitales de telles comètes sont beaucoup plus grandes que celles des planètes, ainsi lorsque la comète retourne, les phases des dernières peuvent être prises au hasard. Par conséquent, il est intéressant de modéliser ces perturbations d'une manière statistiquement fiable et moins coûteuse en temps de calcul.

Nous donnons dans cette section une description statistique d'un grand ensemble des perturbations planétaires calculées numériquement [RVF02]. Un modèle de lois à queue régulière est proposé pour la modélisation. Ce modèle peut être utile pour une meilleure compréhension du comportement des perturbations planétaires dans notre système solaire. Il peut également servir à économiser le temps de calcul de simulation de la dynamique des comètes. La sous-section 3.3.2 présente la compilation des données. Les résultats obtenus sont présentés et interprétés dans la sous-section 3.3.3.

3.3.2 Données

Les données des perturbations planétaires qu'on utilise ici sont obtenues par des méthodes numériques [RVF02]. La Figure 3.5a présente le mouvement képlérien elliptique d'une comète P_k par rapport au soleil P_0 . La notation a représente le demi-grand axe. La notation q représente la distance du périhélie au soleil. Les éléments orbitaux sont présentés dans la Figure 3.5b. La direction du périhélie \vec{u}_0 et la direction du centre de gravité \vec{G}/G sont reliées par $\vec{v}_0 = \vec{G}/G \wedge \vec{u}_0$. Les notations i, ω et Ω représentent respectivement l'inclinaison, l'argument du périhélie et la longitude du noeud. Ces trois angles dépendent évidemment du choix du repère orthonormé. Ainsi la dynamique d'une comète est décrite par le vecteur de paramètres suivant $(z, q, \cos i, \omega, \Omega)^T$ où z représente $-\frac{1}{a}$.

Les perturbations planétaires sont données par la variation des paramètres orbitaux d'une comète entre les valeurs qu'ils ont avant d'entrer dans la région des planètes du système solaire $(z_i, q_i, \cos i_i, \omega_i, \Omega_i)^T$ et les valeurs finales $(z_f, q_f, \cos i_f, \omega_f, \Omega_f)^T$ quand la comète est de retour sur l'orbite barycentrique képlérienne. Ainsi la perturbation planétaire est $(\Delta z = z_f - z_i, \Delta q = q_f - q_i, \Delta \cos i = \cos i_f - \cos i_i, \Delta \omega = \omega_f - \omega_i, \Delta \Omega = \Omega_f - \Omega_i)^T$. L'étude présente admet que ces perturbations dépendent principalement de q_i et $\cos i_i$

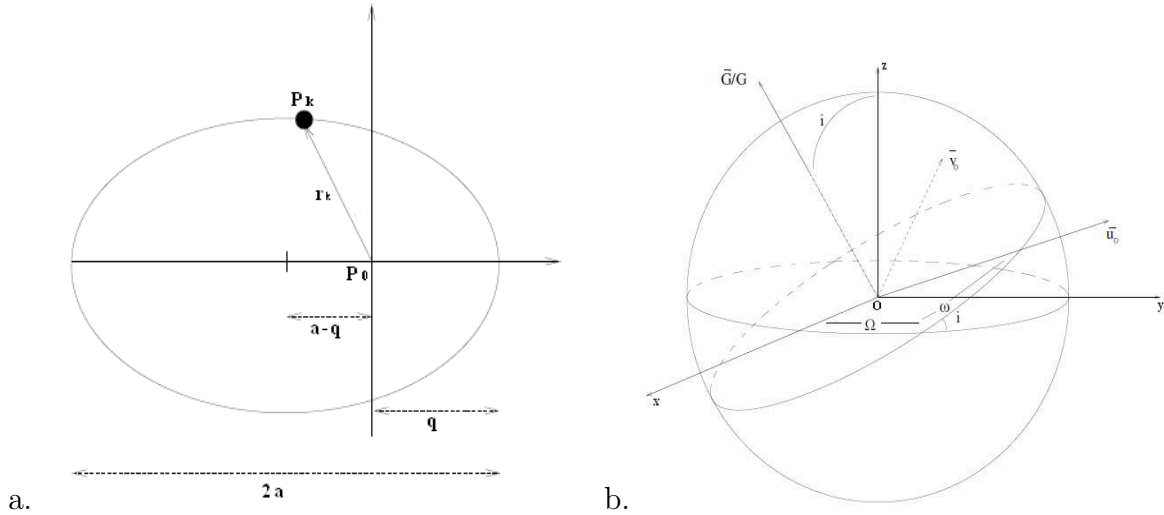


FIG. 3.5: a. Ellipse du mouvement képlérien. b. Éléments d'orbite.

[Fer81], ainsi chaque perturbation est associée au couple $(\cos i_i, q_i)$. De plus, puisque l'énergie orbitale est la quantité principale qui est influencée par les perturbations planétaires, on considère ici seulement ce genre de perturbation.

Par conséquent, nos données sont composées par un ensemble des triplets $(\cos i_i, q_i, \Delta z)$. Nous considérons ces perturbations comme des données spatiales. Les points $(\cos i_i, q_i)$ dans l'espace $K = [-1, 1] \times [0, 32]$ représentent les positions des données. Les valeurs Δz dans \mathbb{R}^1 représentent les marques attachées à chaque emplacement. Dans la suite on appellera Δz tout simplement la marque des perturbations.

3.3.3 Résultats

Quantiles empiriques

L'absence de stationnarité des marques des perturbations impose la partition de l'espace des positions en un nombre fini de cellules. Considérons la partition $K = \bigcup_{i=1}^n K_i$ en cellules K_i disjointes et de même volume. La taille du volume doit être assez grande pour contenir assez de perturbations. En même temps, le volume doit être suffisamment petit pour admettre les hypothèses de stationnarité des marques des perturbations à l'intérieur de la cellule. Après plusieurs essais nous avons opté pour une partition en cellules carrées, ayant toutes le même volume de 0.1×0.1 UA, telle que chaque cellule contienne environ 1500 perturbations.

Nous nous sommes intéressés à trois questions concernant la loi des marques des perturbations. Les deux premières sont liées à la queue et à la symétrie des lois des données. La troisième question est liée à un problème plus délicat. Il est connu que les positions des perturbations suivent la loi uniforme dans K . Les valeurs des perturbations dépendent beaucoup de leurs positions. Néanmoins, on a très peu de connaissances sur la loi spatiale

des perturbations. Ainsi, la troisième question est la suivante : existe-t-il une structure spatiale soulignée par les marques des perturbations ?

L'analyse exploratoire que nous proposons pour les données des perturbations est basée sur le calcul de quantiles empiriques. Il existe plusieurs raisons qui motivent un tel choix. Premièrement, il n'y a pas beaucoup de connaissances a priori concernant ces perturbations. Nous savons seulement qu'elles sont distribuées autour de zéro et qu'elles se situent de manière uniforme dans K . Cela implique que très peu d'hypothèses par rapport aux données peuvent être faites. Pour appliquer une telle analyse, la seule hypothèse nécessaire concerne les conditions de validité du théorème central limite.

A cet effet, les q -quantiles empiriques sont calculés dans chaque cellule. Les calculs des q -quantiles centraux indiquent que les marques des perturbations sont distribuées autour de l'origine, et qu'aucune structure spatiale particulière n'est observable dans l'espace des positions des perturbations.

En revanche, la situation est complètement différente pour les q -quantiles extrêmes : $q_{0.01}, q_{0.05}, q_{0.95}, q_{0.99}$. Nous observons que ces quantiles prennent des valeurs importantes autour du demi-grand axe de chaque planète. Afin de vérifier si ces valeurs peuvent indiquer des distributions à queue lourde, les différences $\hat{z}_q - \hat{n}_q$ sont calculées. Le premier terme de cet indicateur représente le q -quantile empirique. Le deuxième terme est le q -quantile théorique de la loi normale avec la moyenne et l'écart-type donnés respectivement par $\hat{z}_{0.50}$ et $(\hat{z}_{0.84} - \hat{z}_{0.16})/2$. Ainsi pour les q -quantiles extrêmes, la différence $\hat{z}_q - \hat{n}_q$ peut indiquer une distribution à queue lourde.

La Figure 3.6 présente les différences entre les valeurs estimées de $\hat{z}_{0.99}$ et $\hat{n}_{0.99}$. On observe que cette différence prend des valeurs importantes pour les perturbations qui se situent autour d'une orbite de planète et quand le cosinus de l'angle d'inclinaison est proche de 1. Toutes ces valeurs forment une structure spatiale en forme de flèche. Cette forme se situe autour de l'orbite de planète, pointe de la droite vers la gauche et disparaît lorsque le cosinus de l'angle d'inclinaison s'approche de -1 . La netteté de cette flèche dépend de la taille de la planète la plus proche : plus la planète est grande, plus nette est la forme de la flèche. Cela peut être observé par le changement des valeurs de l'indicateur de différence par rapport à la taille de la planète. Ces observations correspondent dans un certain sens à nos attentes : les perturbations sont plus importantes quand la comète traverse l'orbite d'une planète géante.

Puisque ces phénomènes sont observés pour les q -quantiles extrêmes, ils indiquent que les queues de la loi sont une caractéristique importante des données. Ainsi le modèle statistique pour les données doit être capable de capturer ces caractéristiques des marques.

Les quantiles empiriques peuvent être également utilisés comme un indicateur simple de la symétrie de la distribution des données. Si la différence $\hat{z}_q - |\hat{z}_{1-q}|$ tend vers 0, alors cela indique plutôt la symétrie. La Figure 3.7 présente le résultat de ces différences calculées autour de Jupiter. Les valeurs obtenues sont assez faibles dans les régions considérées. Néanmoins, il existe certaines régions, notamment autour de l'orbite de la planète où nous pouvons supposer que les lois des données sont légèrement asymétriques. Cependant, les perturbations ont des valeurs numériquement assez faibles. Ainsi l'évaluation de la symétrie utilisant l'indicateur proposé doit être effectuée avec prudence.

Il est raisonnable d'espérer une réponse plus fiable à cette question en utilisant un modèle statistique. Manifestement, un tel modèle devrait aussi être capable de capturer

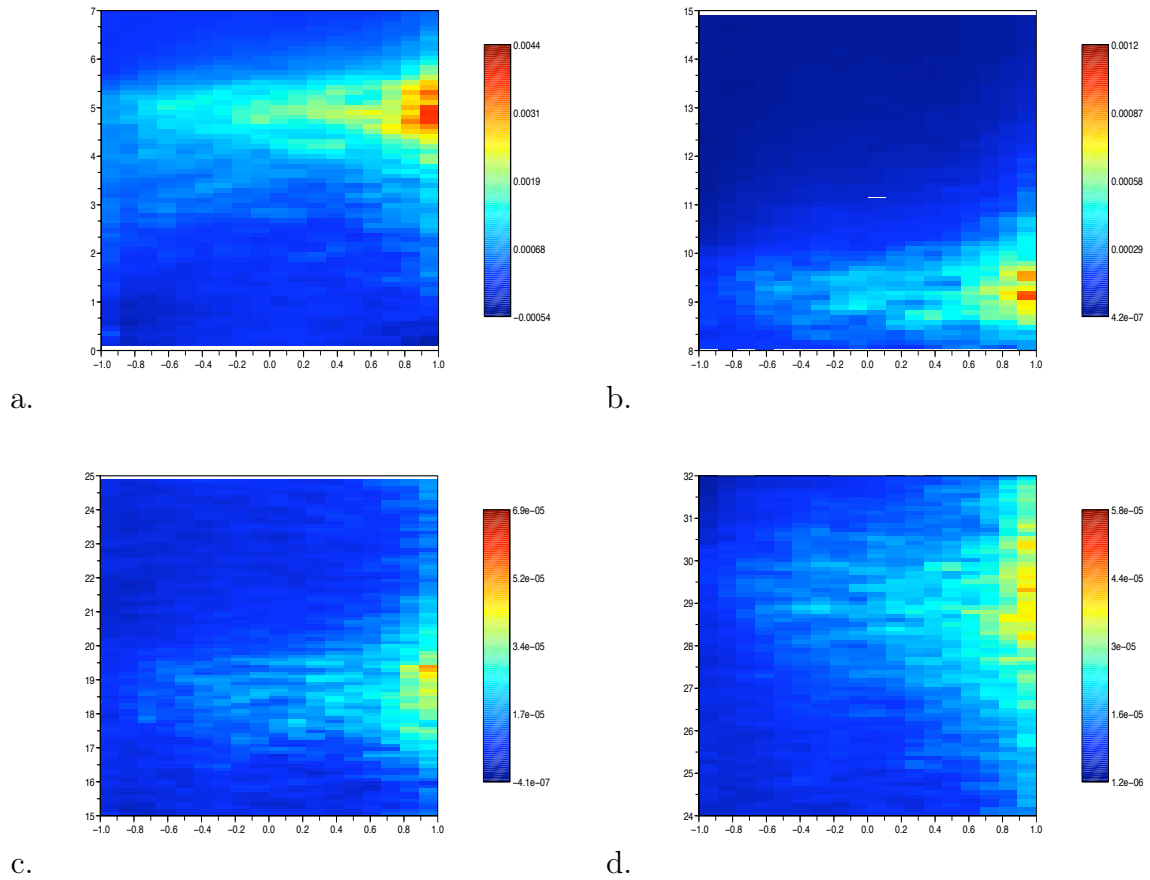


FIG. 3.6: Différences des quantiles empiriques $\hat{z}_{0.99} - \hat{n}_{0.99}$ pour les perturbations autour des grandes planètes : a) Jupiter, b) Saturne, c) Uranus et d) Neptune. Pour chaque diagramme l'axe des ordonnées représente la distance initiale du périhélie (UA), et l'axe des abscisses représente le cosinus de l'angle d'inclinaison. Les demi-grands axes des quatre planètes géantes sont : $a_J = 5.2$ UA, $a_S = 9.6$ UA, $a_U = 19.2$ UA, $a_N = 30.1$ UA.

la symétrie de la distribution des données.

Le théorème central limite pour les statistiques d'ordre permet de construire des tests d'hypothèses. Puisque notre analyse nous amène vers le modèle des lois à queue lourde, avant toute chose, nous souhaitons écarter l'hypothèse la plus simple, c'est-à-dire la normalité. L'essai a été effectué pour les données de chaque cellule, en considérant les moyennes et les écart-types empiriques comme paramètres de la loi normale. Les p -valeurs ont été calculées en utilisant la loi de χ^2 . Dans ce contexte, l'hypothèse de normalité locale des marques des perturbations est globalement rejetée.

La Figure 3.8 présente le résultat des tests de normalité du quantile empirique $\hat{z}_{0.95}$ calculé autour de l'orbite de Jupiter. Nous observons qu'il y a des régions où la normalité ne peut pas être rejetée. Néanmoins, les régions où cette hypothèse est rejetée indiquent que la normalité ne peut être assumée en totalité. Par conséquent, un modèle statistique paramétrique doit être capable de tenir compte de cette situation, aussi.

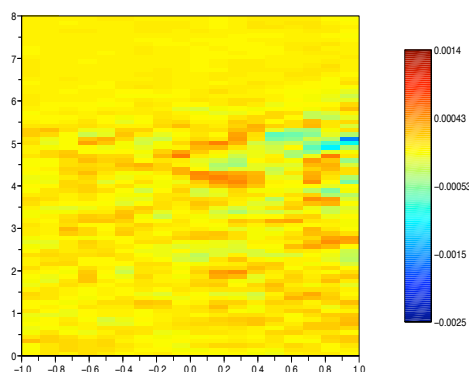


FIG. 3.7: Analyse de la symétrie en utilisant la différence des quantiles $\hat{z}_{0.99} - |\hat{z}_{0.01}|$ pour les marques des perturbations autour de Jupiter.

Le seul paramètre utilisé dans cette analyse exploratoire est la partition du domaine des positions K . Une question se pose : les résultats obtenus dépendent de la structure induite par les données, ou bien ils ne sont que la conséquence de la partition en cellules ? Pour répondre à cette question, une procédure de type bootstrap et un test de permutation ont été mis en oeuvre [DH97].

L'échantillon de bootstrap est pris aléatoirement en choisissant de manière uniforme 20% dans l'ensemble de données initiales. Les différences $\hat{z}_{0.99} - \hat{n}_{0.99}$ ont été calculées pour cet ensemble de données. Cette opération a été répétée 100 fois. A la fin de la procédure, les moyennes empiriques de ces différences ont été calculées. Les résultats sont présentés dans la Figure 3.9 a). Comme attendu, la même structure que celle de la Figure 3.6 a) est obtenue.

Le test de permutation suit les mêmes étapes que la procédure de bootstrap sauf que les perturbations sont préalablement perméutées. Cette transformation se passe comme suit : pour chaque perturbation, sa marque est conservée, tandis que sa position est échangée avec la position d'une autre perturbation choisie au hasard. Cette procédure devrait

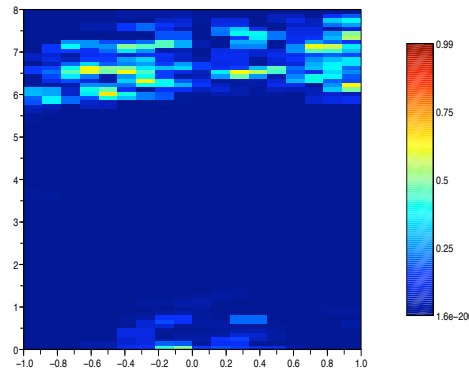


FIG. 3.8: p -valeurs obtenues pour le test de normalité des quantiles empiriques $\hat{z}_{0.95}$ autour de l'orbite de Jupiter.

détruire toute structure existante dans les données. Dans ce cas, nous espérons que la procédure de type bootstrap ne montrera pas la présence d'une structure spatiale. La Figure 3.9b) montre le résultat de ce test effectué dans le voisinage de l'orbite de Jupiter. Après la permutation, une procédure de type bootstrap est appliquée pour estimer les moyennes des indicateurs de différence $\hat{z}_{0.99} - \hat{n}_{0.99}$. Le résultat confirme nos attentes : aucune structure particulière n'est observée. Cela indique que les résultats de l'analyse sont dus principalement à la distribution des données et ne dépendent pas de la partition en cellules du domaine des positions des perturbations.

En même temps, la permutation est également un test de vérification de la méthodologie exploratoire proposée. Cette méthodologie permet de caractériser la structure spatiale induite par les données, en fonction d'un paramètre de précision. Quand une telle structure n'existe pas, la méthode ne détecte rien.

Inférence statistique en utilisant les lois à queue régulière

Les observations empiriques des lois des marques des perturbations indiquent une loi à queue lourde avec une asymétrie parfois non réglable. Le caractère leptocurtique des lois des perturbations a été observé surtout dans les voisinages des orbites des planètes. Pour répondre à cette évidence empirique, des lois à queue régulière sont choisies pour la modélisation.

La même partition en cellules que pour l'analyse exploratoire est utilisée. L'algorithme présenté dans l'annexe A est effectué pour les données dans chaque cellule.

Les résultats d'estimation des paramètres α et γ sur les données des perturbations autour de l'orbite de quatre planètes sont présentés dans les Figures 3.10 et 3.11. Nous observons dans la Figure 3.10 des régions formées par les cellules avec une valeur estimée de α inférieure à 2. Ce type de région se situe autour de l'orbite de chaque grande planète. La forme de cette région est moins pointue que la région obtenue à l'aide des quantiles empiriques. Pourtant, il y a une corrélation entre les deux résultats. Ces deux résultats

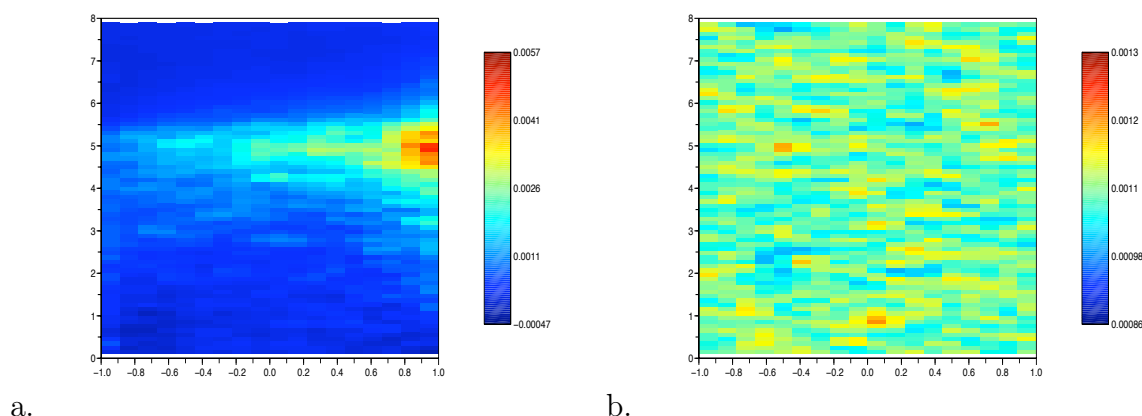


FIG. 3.9: Validation de l’analyse basée sur l’indicateur de différence $\hat{z}_{0.99} - \hat{n}_{0.99}$ calculé autour de l’orbite de Jupiter : a) procédure de bootstrap, b) test de permutation.

indiquent que le caractère de type queue lourde de la loi des perturbations présente une forme spatiale. Cette forme spatiale se situe autour de l’orbite des grandes planètes. Le paramètre d’échelle γ indique le “poids” des queues des distributions. Dans la Figure 3.11 il peut être observé que les valeurs les plus importantes forment une forme spatiale similaire à celle formée par l’indice de queue, c’est-à-dire une bande autour de l’orbite de la grande planète correspondante.

La symétrie de la distribution des données peut être analysée en examinant les résultats d’estimation du paramètre β présentés dans la Figure 3.12. Les résultats obtenus n’indiquent pas la présence de symétrie ou d’asymétrie en tant que caractéristique dominante ou sous-entendant une structure spatiale.

Les résultats d’estimation des paramètres de position δ sont présentés dans la Figure 3.13. Ce paramètre représente l’espérance de la loi des perturbations. Ici on utilise tout simplement la moyenne empirique comme estimateur. Les résultats indiquent qu’une translation des perturbations est possible autour de l’orbite des grandes planètes correspondantes.

Afin de vérifier ces résultats, un test statistique en utilisant le théorème central limite des statistiques d’ordre a a été construit. La Figure 3.14 présente le résultat du test qui vérifie bien que les quantiles $\hat{z}_{0.99}$ autour de l’orbite de Jupiter viennent d’une loi stable avec le paramètre α correspondant, tandis que les quantiles en dehors de cette région viennent d’une loi de type Paréto. Il peut être observé que des valeurs élevées pour les p -valeurs sont réparties sur l’ensemble de la région : pour 81.5% des cellules, nous ne pouvons pas rejeter l’hypothèse nulle. Ce résultat montre une meilleure caractérisation des queues des lois des perturbations que le test de l’hypothèse de normalité effectué dans la sous-section précédente. Le test précédent indique que les lois des perturbations présentent un comportement de type stable ou de type Paréto. Si les perturbations sont proches de l’orbite d’une grande planète, elles ont plutôt un comportement stable. La Figure 3.15 présente les p -valeurs du test de χ^2 mis en oeuvre pour les perturbations

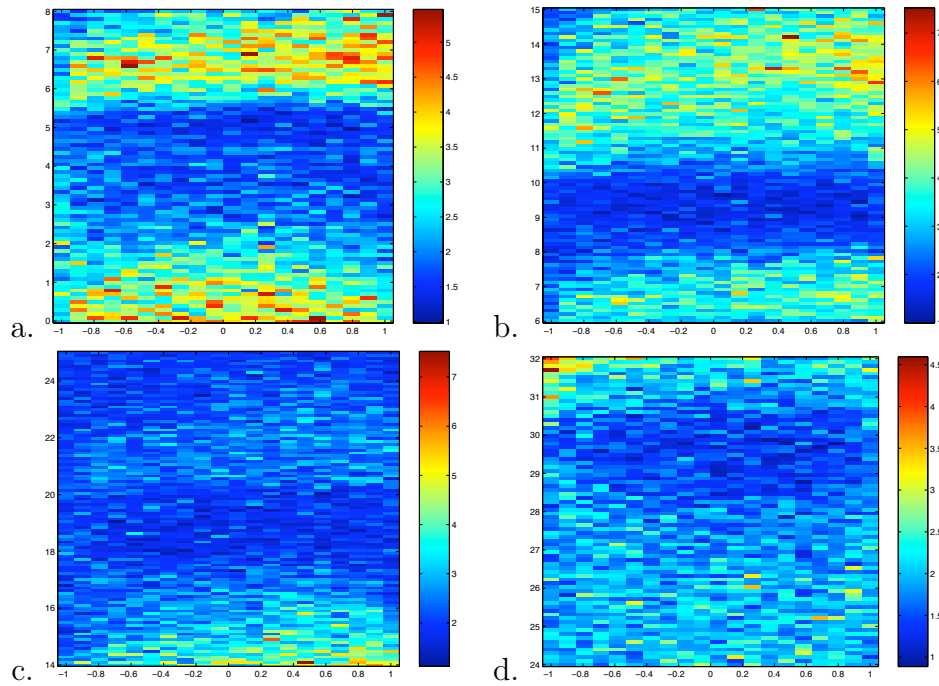


FIG. 3.10: Résultats d'estimation de l'indice de queue α sur les marques des perturbations autour des grandes planètes : a) Jupiter, b) Saturne, c) Uranus et d) Neptune.

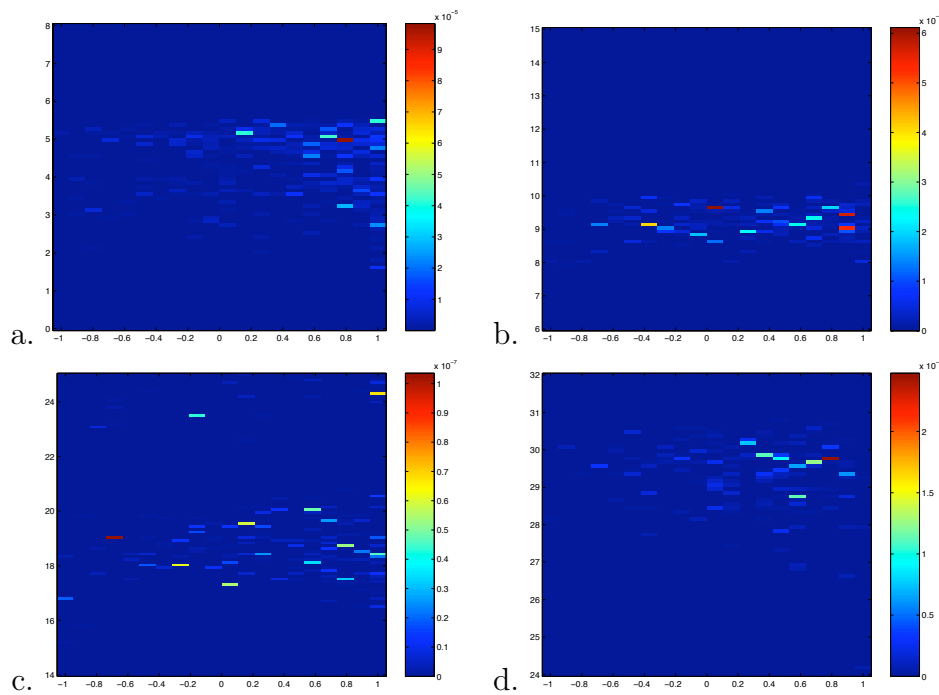


FIG. 3.11: Résultats d'estimation du paramètre d'échelle γ sur les marques des perturbations autour des grandes planètes : a) Jupiter, b) Saturne, c) Uranus et d) Neptune.

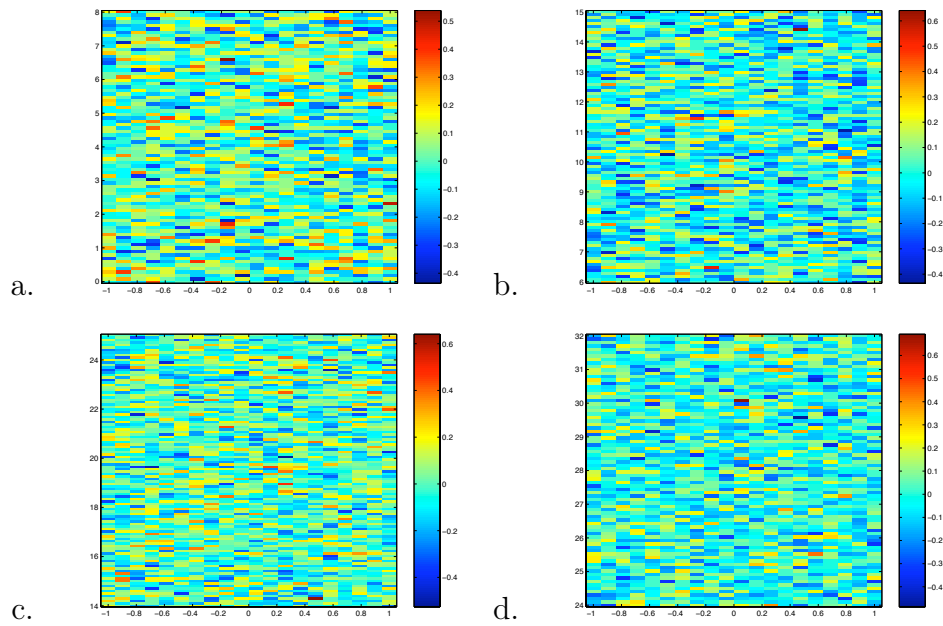


FIG. 3.12: Résultats d'estimation du paramètre β sur les marques des perturbations autour des grandes planètes : a) Jupiter, b) Saturne, c) Uranus et d) Neptune.

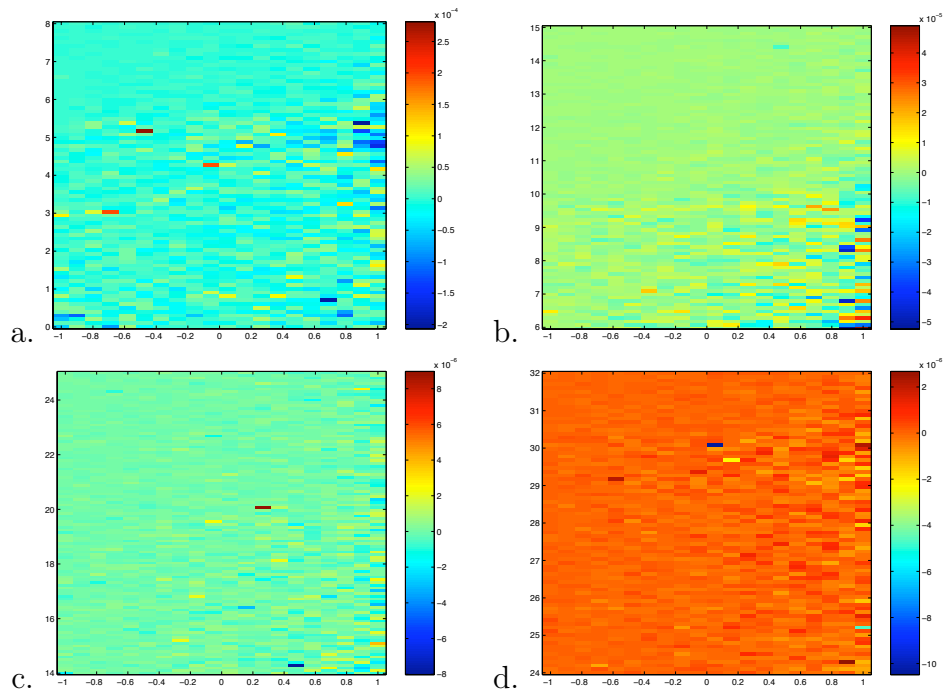


FIG. 3.13: Résultats d'estimation du paramètre de position δ sur les marques des perturbations autour des grandes planètes : a) Jupiter, b) Saturne, c) Uranus et d) Neptune.

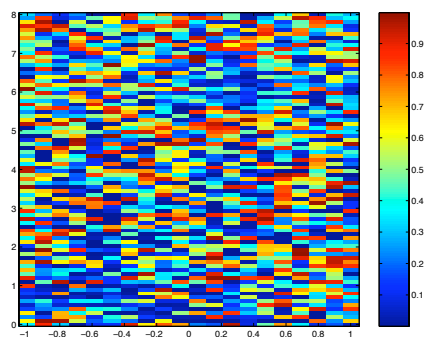


FIG. 3.14: Les p -valeurs calculées autour de l'orbite de Jupiter pour le test des quantiles empiriques $\hat{z}_{0.99}$ des lois à queue régulière.

avec $\alpha < 2$. Dans ce cas, nous modélisons en utilisant les lois stables. Ce test permet de vérifier les perturbations non seulement pour la queue, mais aussi pour le corps de leurs distributions. Nous observons que dans presque toutes ces régions, l'hypothèse de loi stable n'est pas rejetée.

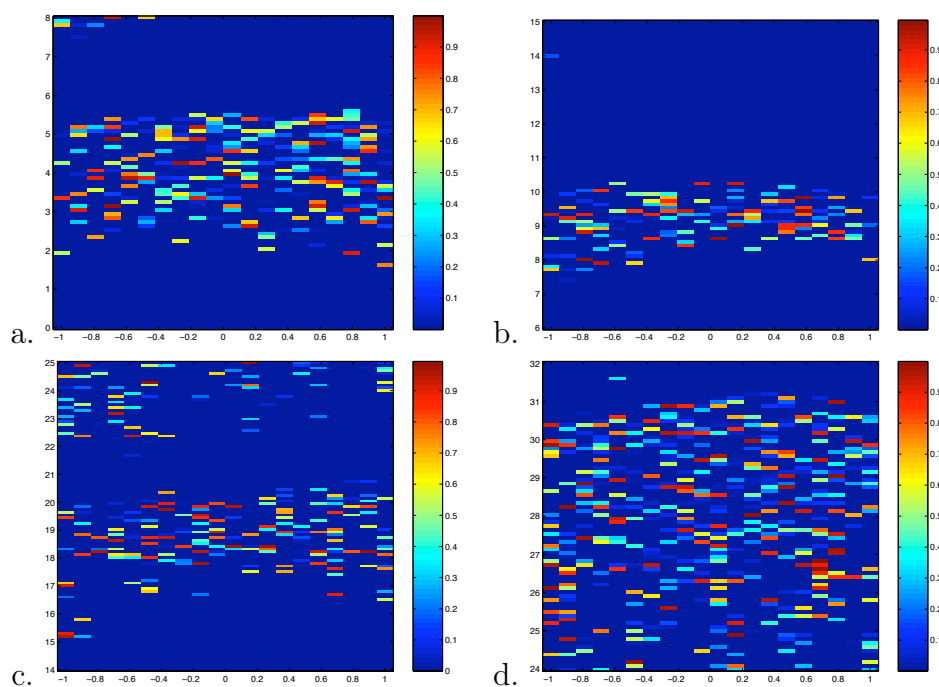


FIG. 3.15: Les p -valeurs du test statistique de χ^2 pour les perturbations autour des grandes planètes avec $\alpha < 2$: a) Jupiter, b) Saturne, c) Uranus et d) Neptune.

Pour les perturbations avec $\alpha \geq 2$, nous considérons la famille des distributions dont la densité est définie par (3.1.2). Le test statistique du χ^2 est appliqué pour les perturbations avec $\alpha \geq 2$. L'hypothèse nulle est que les perturbations proviennent d'une loi donnée par (3.1.2). Les p -valeurs obtenues sont présentées dans la Figure 3.16. On constate que pour

la majorité des cellules concernées cette hypothèse n'est pas rejetée.

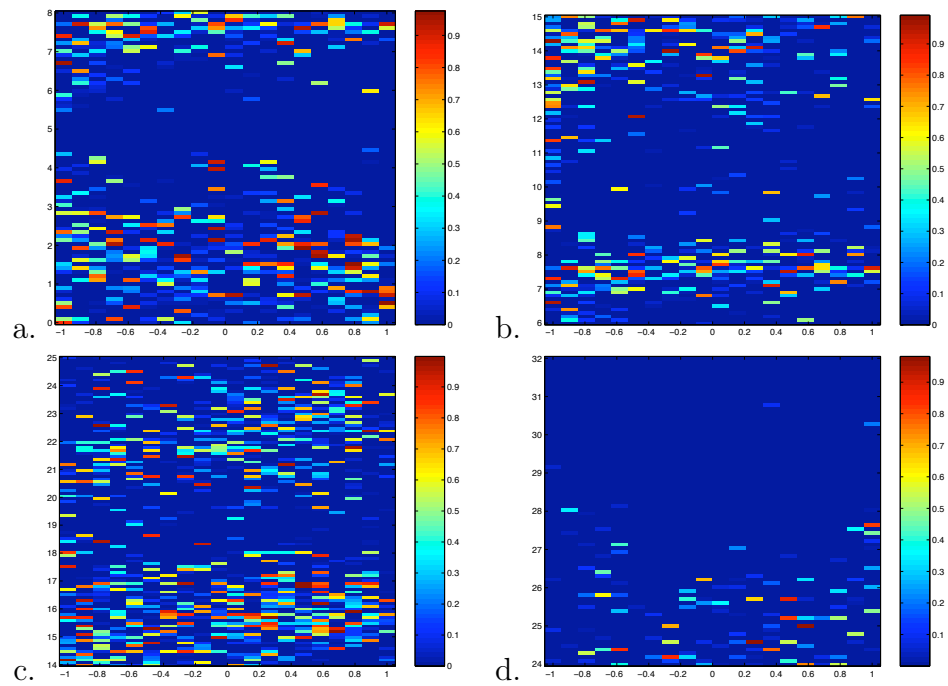


FIG. 3.16: Les p -valeurs du test statistique de χ^2 pour les perturbations autour des grandes planètes avec $\alpha \geq 2$: a) Jupiter, b) Saturne, c) Uranus et d) Neptune.

Conclusion et perspectives

Nous avons présenté dans cette thèse une méthode pour estimer l'indice caractéristique et la mesure spectrale d'une loi à queue régulière dans un cône abstrait. C'est une application statistique utilisant le lien entre la variation régulière et la représentation de LePage d'un élément aléatoire stable dans un cône. Elle nous amène des méthodes d'estimation des paramètres pour des différentes familles de lois. Les lois α -stables et les lois max-stables sont des exemples typiques. Un des points forts de cette méthode est que l'estimateur de la mesure spectrale normalisée ne dépend pas de l'indice de la queue. Nous avons prouvé que les estimateurs proposés ont une loi asymptotiquement normale. La vitesse de convergence dépend des exposants de premier et second ordre de la queue de la loi marginale. Par exemple pour estimer les paramètres des lois α -stables dans \mathbb{R}^d , la vitesse de convergence optimale peut atteindre $N^{1/3}$ pour les estimateurs de l'indice α et de la mesure spectrale normalisée, et $N^{1/4}$ pour l'estimateur de la masse totale de mesure spectrale, où N représente la taille d'échantillon.

Nous avons proposé un nouvel estimateur de la masse totale de mesure spectrale sous l'hypothèse que l'échantillon est issu d'une loi appartenant au domaine d'attraction normal d'une loi stable. On souhaite affaiblir cette hypothèse afin de donner un estimateur général pour la masse totale de la mesure spectrale d'une loi à queue régulière.

Dans le cas où la mesure spectrale est absolument continue par rapport à la mesure de Lebesgue, une estimation de sa densité peut être obtenue par la méthode de noyau appliquée à une suite de vecteurs aléatoires i.i.d. dont la loi commune converge vers la mesure spectrale. La consistance de cet estimateur est démontrée sous la condition forte de variation régulière. On souhaite trouver la vitesse de convergence pour un choix de fenêtre optimal.

Nous avons effectué la mise en oeuvre de l'algorithme d'estimation précité dans le langage de programmation MATLAB. Trois différentes méthodes d'estimation des paramètres des lois stables ont été comparées par l'étude de simulation. Les estimateurs proposés sont asymptotiquement sans biais, et ils ont l'avantage évident sur le temps de calcul. Il est clair que dans un cadre paramétrique, les estimateurs développés dans cette thèse n'avaient "aucune chance" face à un estimateur paramétrique comme le maximum de vraisemblance. Pour compléter cette étude de simulation, on pourrait comparer avec d'autres estimateurs non-paramétriques sur des données issues d'une loi à queue régulière.

Nous avons étudié trois sortes de transformations par lesquelles la propriété de variation régulière est préservée. Des conditions suffisantes pour cette préservation ont été présentées. Ces transformations ont été utilisées pour simuler des vecteurs aléatoires appartenant au domaine d'attraction d'une loi stable. Les propriétés des projections ont été utilisées pour étudier le comportement de la mesure spectrale d'une loi à queue régulière

multivariée.

La famille de lois à queue régulière peut être utilisée dans de nombreux domaines d'application des événements rares ou des valeurs extrêmes. Nous avons présenté dans le dernier chapitre deux applications de modèles de lois à queue régulière en utilisant la méthode d'estimation proposée. Une direction intéressante est d'améliorer cette méthode afin de pouvoir donner une description plus complète pour les données réelles. En moyennant les résultats d'estimation sur des permutations aléatoires des échantillons on améliore numériquement la convergence des estimateurs. Il serait intéressant de démontrer théoriquement cette amélioration.

Dans la partie d'application, une autre question ouverte devant nous est de trouver une alternative aux lois stables pour $\alpha > 2$ qui ajuste au mieux un jeu de données. Cette question peut être considérée aussi dans le cas multidimensionnel.

Annexe A

Bilan sur les estimateurs

Le tableau suivant forme l'algorithme de l'estimation des paramètres des lois à queue régulière présentée dans la section 2.1. Nous résumons ensuite des résultats principaux concernant la consistance et la normalité asymptotique des estimateurs.

A.1 Liste des variables et calcul des estimateurs

Variables et notations	
ξ_1, \dots, ξ_N	une suite des e.a. i.i.d. de loi satisfaisant la condition de VR(α, σ)
N	la taille d'échantillon
$r \in (0, 1)$	le rapport de regroupement
$n = \lfloor N^r \rfloor$	le nombre des groupes
$m = \lfloor N^{1-r} \rfloor$	la taille des groupes
$G_{m,i} = \{\xi_{(i-1)m+1}, \dots, \xi_{im}\}$	le i ème groupe, $i = 1, \dots, n$
$M_{m,i}^{(1)} = \max\{\ \xi\ \mid \xi \in G_{m,i}\}$	la plus grande norme dans le i ème groupe
$\xi_{m,i} : \ \xi_{m,i}\ = M_{m,i}^{(1)}$	l'élément dont la norme est la plus grande dans le i ème groupe
$M_{m,i}^{(2)} = \max\{\ \xi\ \mid \xi \in G_{m,i} \setminus \{\xi_{m,i}\}\}$	la deuxième grande norme dans le i ème groupe
Estimateurs	
$\hat{\alpha}_N = \frac{S_n}{n - S_n}$	où $S_n = \sum_{i=1}^n \frac{M_{m,i}^{(2)}}{M_{m,i}^{(1)}}$
$\widehat{\sigma(S)}_N = \left(\frac{1}{n\Gamma(1 - \frac{t}{\alpha})} \sum_{i=1}^n q_{m,i}^t \right)^{\frac{\alpha}{t}}$	où $q_{m,i} = \frac{M_{m,i}^{(1)}}{m^{1/\alpha}}$ et $0 < t < \frac{\alpha r}{2}$
$\hat{\sigma}_N(\cdot) = \frac{1}{n} \sum_{i=1}^n \delta_{\theta_{m,i}}(\cdot)$ (cas multidimensionnel)	où $\theta_{m,i} = \frac{\xi_{m,i}}{\ \xi_{m,i}\ }$
$\hat{\beta}_N = \frac{2}{n} \sum_{i=1}^n \mathbb{I}_{\{1\}} \theta_{m,i} - 1$ (cas unidimensionnel)	

A.2 Consistance

Condition de régularité (1.1.6) :

L'e.a. ξ satisfait la *condition de régularité* si $\forall \lambda > 0$ et pour tout $B \in \mathcal{B}(S)$ avec $\sigma(\partial B) = 0$

$$\lim_{n \rightarrow \infty} n \mathbf{P} \left\{ \frac{\xi}{\|\xi\|} \in B, \|\xi\| > \lambda b_n \right\} = \sigma(B) \lambda^{-\alpha}.$$

- (Théorème 2.1.4.) Si ξ satisfait la condition de régularité alors on a

$$\hat{\alpha}_N \xrightarrow[N \rightarrow \infty]{p.s.} \alpha \text{ quand } N \rightarrow \infty.$$

- (Lemme 2.1.7.) Si ξ satisfait la condition de régularité alors on a pour chaque i

$$\theta_{m,i} \Rightarrow \sigma \text{ quand } m \rightarrow \infty.$$

- (Corollaire 2.1.9.) Si ξ satisfait la condition de régularité alors on a

$$\hat{\sigma}_N \xrightarrow[N \rightarrow \infty]{p.s.} \tilde{\sigma} \text{ quand } N \rightarrow \infty.$$

- (Théorème 2.1.6.) Si ξ satisfait la condition de régularité avec $b_n = n^{1/\alpha}$, alors pour $0 < t < \frac{\alpha r}{2}$ on a

$$\widehat{\sigma(S)}_N - \sigma(S) \xrightarrow[N \rightarrow \infty]{p.s.} 0 \text{ quand } N \rightarrow \infty.$$

A.3 Normalité asymptotique

Relation asymptotique du second ordre (2.1.40) :

L'e.a. ξ satisfait la *relation asymptotique du second ordre* si

$$\mathbf{P}\{\|\xi\| > x\} = \sigma(S)x^{-\alpha} + cx^{-\rho} + o(x^{-\rho}) \text{ quand } x \rightarrow \infty,$$

avec $0 < \alpha < \rho \leq \infty$ où c est une constante.

- (Théorème 2.1.10.) Si la loi de ξ satisfait la condition de régularité (1.1.6) et la relation asymptotique de second ordre (2.1.40) et si on choisit

$$n = N^{2\zeta/(1+2\zeta)-\varepsilon}, \quad m = N^{1/(1+2\zeta)+\varepsilon},$$

où $\zeta = (\rho - \alpha)/\alpha$ et ε est une constante strictement positive et arbitrairement petite, alors on a

$$\frac{\sqrt{n} \left(\frac{1}{n} S_n - \frac{\alpha}{\alpha + 1} \right)}{\left(\frac{1}{n} \sum_{i=1}^n \left(\frac{M_{m,i}^{(2)}}{M_{m,i}^{(1)}} \right)^2 - \left(\frac{1}{n} S_n \right)^2 \right)^{1/2}} \Rightarrow \mathcal{N}(0, 1).$$

- (Théorème 2.1.12.) Si la loi de ξ satisfait la condition de régularité (1.1.6) et la relation asymptotique de second ordre (2.1.40) avec $\rho > \alpha + 1$. Si on choisit

$$r = \frac{3\alpha - 4(\rho - 1) + \sqrt{16(\rho - 1)^2 - 8\alpha(\rho - 1) - 7\alpha^2}}{2\alpha} - \varepsilon, \quad \text{si } \rho \leq \frac{11}{8}\alpha + 1,$$

$$r = \frac{1}{2} - \varepsilon, \quad \text{si } \rho > \frac{11}{8}\alpha + 1,$$

où ε est une constante suffisamment petite, alors dans deux cas

$$0 < t < \frac{\alpha r}{4} \wedge 1 \quad \text{si } \alpha + 1 < \rho \leq \frac{11}{8}\alpha + 1 \quad \text{ou } \rho \geq \frac{3}{2}\alpha + 1,$$

$$0 < t < \frac{3\alpha + 2 - 2\rho}{2} \wedge 1 \quad \text{si } \frac{11}{8}\alpha + 1 < \rho < \frac{3}{2}\alpha + 1,$$

on a

$$\frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n q_{m,i}^t - \Gamma \left(1 - \frac{t}{\alpha} \right) \sigma(S)^{t/\alpha} \right)}{\left(\frac{1}{n} \sum_{i=1}^n q_{m,i}^{2t} - \left(\frac{1}{n} \sum_{i=1}^n q_{m,i}^t \right)^2 \right)^{1/2}} \Rightarrow \mathcal{N}(0, 1).$$

Relation forte du second ordre (2.1.75) :

L'e.a. ξ satisfait la relation forte du second ordre si $\forall B \in \mathcal{B}(S)$ tel que $\sigma(\partial B) = 0$

$$\mathbf{P} \left\{ \frac{\xi}{\|\xi\|} \in B, \|\xi\| > x \right\} = \sigma(B)x^{-\alpha} + cx^{-\rho} + o(x^{-\rho}) \quad \text{quand } x \rightarrow \infty,$$

avec $0 < \alpha < \rho \leq \infty$ où c est une constante.

- (Théorème 2.1.14.) Si la loi de ξ satisfait la relation forte du second ordre (2.1.75) et si on choisit

$$n = N^{2\zeta/(1+2\zeta)-\varepsilon}, \quad m = N^{1/(1+2\zeta)+\varepsilon},$$

où $\zeta = \min(\frac{\rho-\alpha}{\alpha}, 1)$ et ε est une constante strictement positive et arbitrairement petite, alors $\forall B \in \mathcal{B}(S)$ tel que $\sigma(\partial B) = 0$

$$\frac{\sqrt{n}(\hat{\sigma}_N(B) - \sigma(B))}{\left(\frac{1}{n} \sum_{i=1}^n (\mathbb{1}_B(\theta_{m,i}))^2 - \left(\frac{1}{n} \sum_{i=1}^n \mathbb{1}_B(\theta_{m,i}) \right)^2 \right)^{1/2}} \Rightarrow \mathcal{N}(0, 1).$$

Annexe B

Programmes MATLAB

Nous fournissons ici le code des fonctions qui est utilisé dans les sections 1.3 et 2.1 pour simuler les données des lois α -stables et max-stables, et pour estimer les paramètres. Les fonctions sont : “mvrfit.m”, “stabsym1d.m”, “stabiso2d.m”, “maxstabuni.m” et “maxstab2p.m”. Le texte commencé par % est considéré comme le commentaire.

B.1 Programme pour estimer les paramètres : mvrfit.m

```
% Fonction qui estime les paramètres des données provenant d'une loi à
% queue régulière multivariée en utilisant l'algorithme présenté dans la
% section 2.1. La dimension maximale admissible est 30.

% Les sorties sont :
% alpha : nombre réel positif,
% masse : nombre réel positif,
% sigma : matrice  $n \times \dim$  qui contient les coordonnées des points du support
% de la mesure spectrale estimée,
% delta : vecteur  $1 \times \dim$  qui contient les coordonnées de la moyenne des
% données.

% L'argument obligatoire est la matrice A.
% Les arguments optionnels sont :
% ispara1 : la paramétrisation de [Samorodnitsky 94],
% ismax : la norme définie par le maximum des coordonnées du vecteur.

% Les variables sont :
% N : la taille d'échantillon,
% ddim : la dimension des données,
% r : le rapport de regroupement,
% n : le nombre des groupes,
% m : la taille des groupes,
% delta : la moyenne des données,
% nnorme : la norme du vecteur,
```

```

% s :  $S_n$ ,
% ma :  $M_{n,i}^{(1)}$ ,
% sigma : la mesure spectrale estimée, dans le cas unidimensionnel sigma
% apporte la valeur du paramètre d'asymétrie  $\beta$ ,
% alpha : l'indice de la queue,
% t :  $0 < t < \min(\alpha r/2, 1)$ .

```

```
function [alpha,masse,sigma,delta] = mvrfit(A,varargin)
```

```

para = 0; % Paramétrisation par défaut
ismax = 0; % Norme par défaut
nvarargin = size(varargin,2);
for i = 1 :nvarargin
    % Si l'utilisateur précise la paramétrisation
    if strcmp(varargin{i},'ispara1')
        para = 1;
    end
    % Si l'utilisateur précise la norme
    if strcmp(varargin{i},'ismax')
        ismax = 1;
    end
end

[N,ddim] = size(A);
if ddim > 30
    error('The dimension must be less than 30.')
end
r = 0.6;
n = floor(N^r);
m = floor(N/n);
if m < 2
    error('The size of sample is too small.')
end
delta = mean(A);
if ismax == 1
    nnorme = max(A')';
else
    nnorme = sqrt(sum(A.^2,2));
end
s = 0;
ma = zeros(1,n);
sigma = zeros(n,ddim);
for j = 0 : n-1
    % Ordonner les éléments dans le groupe j+1
    [v,pos] = sort(nnorme(j*m+1 :(j+1)*m));

```

```

sigma(j+1, :) = A(j*m+pos(m), :)/nnorme(j*m+pos(m));
s = s+v(m-1)/v(m);
ma(j+1) = v(m);
end
alpha = s/(n-s);

% Centrer les données et répéter la procédure d'estimation, si la valeur de
%  $\alpha$  pré-estimée est plus grande que 1
if alpha>1
    if ismax == 1
        nnorme = max((A-ones(N,1)*delta)')';
    else
        nnorme = sqrt(sum((A-ones(N,1)*delta).^2,2));
    end
    s = 0;
    ma = zeros(1,n);
    sigma = zeros(n,ddim);
    for j = 0 : n-1
        [v,pos] = sort(nnorme(j*m+1 :(j+1)*m));
        sigma(j+1, :) = A(j*m+pos(m), :)/nnorme(j*m+pos(m));
        s = s+v(m-1)/v(m);
        ma(j+1) = v(m);
    end
    alpha = s/(n-s);
end

t = min(alpha*r/2-0.0001,0.9999);
q = sum((ma/(m^(1/alpha))).^t)/n;
masse = (q/gamma(1-t/alpha))^(alpha/t);

if ddim == 1
    sigma = length(find(sigma>0))*2/n-1;
end

if para == 1 && alpha < 2
    if alpha == 1
        ca = 2/pi;
    else
        ca = (1-alpha)/(gamma(2-alpha)*cos(pi*alpha/2));
    end

    if ddim == 1
        masse = (masse/ca)^(1/alpha);
    else
        masse = masse/ca;
    end
end

```



```

    end
end

```

B.2 Programmes pour simuler les vecteurs aléatoires

B.2.1 stabsym1d.m

```

% Fonction qui simule n données de loi stable symétrique unidimensionnelle de
% paramètres (alpha, masse) en utilisant la série de LePage. Cette fonction
% est utilisée dans la sous-section 1.3.1 pour simuler les lois  $\mathcal{S}_1(0.75, (0.5, 1), 0)$ 
% et  $\mathcal{S}_1(0.75, (0, 1), 0)$ .

```

```

% La sortie est le vecteur qui contient n données simulées y.

```

```

% L'argument optionnel est :
% ispara1 : la paramétrisation de [Samorodnitsky 94],

```

```

function y = stabsym1d(alpha,masse,n,varargin)

```

```

para = 0; % Paramétrisation par défaut
% Si l'utilisateur précise la paramétrisation
if strcmp(varargin{1},'ispara1')
    para = 1;
end

```

```

nn = 50; % Nombre de termes de la somme partielle
if para == 0
    c = masse^(1/alpha);
else
    if alpha == 1
        ca = 2/pi;
    else
        ca = (1-alpha)/(gamma(2-alpha)*cos(pi*alpha/2));
    end
    c = ca^(1/alpha)*masse;
end

```

```

p = 1/2;
r = zeros(n,1);
xx = zeros(n,1);

```

```

for i = 1 :nn
    ra1 = rand(n,1);

```

```

ra2 = rand(n,1);

lambda = -log(1-ra1);
r = r+lambda;

e = -sign(ra2-p);

xx = xx+r.^(-1/alpha).*e;
end
y = c*xx';

```

B.2.2 stabiso2d.m

% Fonction qui simule n données de loi alpha-stable isotopique bivariée en utilisant
 % la série de LePage. Cette fonction est utilisée dans la sous-section 1.3.1 pour
 % simuler les lois 0.75 et 1.5-stables isotopiques.

% La sortie est le vecteur $2 \times n$ qui contient n données simulées xy.

```

function xy = stabiso2d(alpha,n)
xy = zeros(2,n);
nn = 10; % Nombre de termes de la somme partielle
r = zeros(n,1);
xx = zeros(n,1);
yy = zeros(n,1);

for i = 1 :nn
    ra1 = rand(n,1);
    ra2 = rand(n,1);

    lambda = -log(1-ra1);
    r = r+lambda;

    ex = cos(ra2*2*pi);
    ey = sin(ra2*2*pi);

    xx = xx+r.^(-1/alpha).*ex;
    yy = yy+r.^(-1/alpha).*ey;
end
xy(1, :) = xx';
xy(2, :) = yy';

```

B.2.3 maxstabuni.m

% Fonction qui simule n données de loi max-stable bivariée de mesure spectrale
% uniforme en utilisant la série de LePage. Cette fonction est utilisée dans l'exemple 1
% de la sous-section 1.1.4 et l'exemple 5 de la sous-section 1.3.3.

% La sortie est le vecteur 2×n qui contient n données simulées xy.

```
function xy = maxstabuni(alpha,masse,n)
xy = zeros(2,n);
nn = 30; % Nombre de termes de la somme partielle

for j = 1 :n
    ra1 = rand(1,nn);
    ra2 = rand(1,nn);

    lambda = -log(1-ra1);

    ey = ones(1,nn);
    ex = ra2*2;

    pos = find(ex>1);
    ey(pos) = 2-ex(pos);
    ex(pos) = 1;

    px = find(ex == 1);
    py = find(ey == 1);

    x = zeros(1,nn);
    y = zeros(1,nn);

    for i = 1 :px(1)
        x(i) = sum(lambda(1 :i))^(1/alpha)*(masse^(1/alpha))*ex(i);
    end

    for i = 1 :py(1)
        y(i) = sum(lambda(1 :i))^(1/alpha)*(masse^(1/alpha))*ey(i);
    end

    xy( :,j) = [max(x);max(y)];
end
```

B.2.4 maxstab2p.m

% Fonction qui simule n données de loi max-stable bivariée de mesure spectrale
 % concentrée sur deux points (1,a) et (b,1) de masse p et q en utilisant la série
 % de LePage. Cette fonction est utilisée dans l'exemple 1 de la sous-section 1.1.4
 % et l'exemple 6 de la sous-section 1.3.3.

% La sortie est le vecteur 2×n qui contient n données simulées xy.

```
function xy = maxstab2p(alpha,a,b,p,q,n)

xy = zeros(2,n);
nn = 30; % Nombre de termes de la somme partielle
masse = p+q;
p1 = p/masse;

for j = 1 :n
    ra1 = rand(1,nn);
    ra2 = rand(1,nn);

    lambda = -log(1-ra1);

    ex = ones(1,nn)*b;
    ey = ones(1,nn);

    pos = find(ra2<p1);
    ex(pos) = 1;
    ey(pos) = a;

    px = find(ex == 1);
    py = find(ey == 1);

    x1 = sum(lambda(1 :px(1)))^(-1/alpha)*(masse^(1/alpha))*ex(px(1));
    y1 = sum(lambda(1 :px(1)))^(-1/alpha)*(masse^(1/alpha))*ey(px(1));

    x2 = sum(lambda(1 :py(1)))^(-1/alpha)*(masse^(1/alpha))*ex(py(1));
    y2 = sum(lambda(1 :py(1)))^(-1/alpha)*(masse^(1/alpha))*ey(py(1));

    xy( :,j) = [max(x1,x2);max(y1,y2)];
end
```


Bibliographie

- [AB88] V. Akgiray and G. G. Booth. The stable-law model of stock returns. *Journal of Business & Economic Statistics*, 6(1) :51–57, 1988.
- [AG80] A. Araujo and E. Giné. *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, New York, 1980.
- [BDM02] B. Basrak, R. A. Davis, and T. Mikosch. Regular variation of garch processes. *Stoch., Proc. Appl.*, 99 :95–116, 2002.
- [BGT87] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, Cambridge, 1987.
- [Bil68] P. Billingsley. *Convergence of Probability Measures*. Wiley, 2d ed edition, 1968.
- [Bre65] L. Breiman. On some limit theorems similar to the arc-sin law. *Theory of Probability and its Applications*, 10 :323–331, 1965.
- [Bre68] L. Breiman. *Probability*. Massachusetts. Addison-Wesley, Reading, 1968.
- [BRZ88] Z. D. Bai, C. Radhakrishna Rao, and L. C. Zhao. Kernel estimators of density function of directional data. *J. Multivariate Anal.*, 27(1) :24–39, 1988.
- [CMS76] J. Chambers, C. Mallows, and B. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354) :340–344, 1976.
- [CR95] B. N. Cheng and S. T. Rachev. Multivariate stable securities in financial markets. *Mathematical Finance*, 54 :133–153, 1995.
- [Dav81] H. A. David. *Order statistics*. John Wiley and Sons, 1981.
- [DE00] Y. Davydov and V. Egorov. Functional limit theorems for induced order statistics of a sample from a domain of attraction of α -stable law, $\alpha \in (0, 2)$. *Asymptotics in Statistics and Probability : Papers in Honor of George Gregory Roussas*, pages 85–116, 2000.
- [Dev83] L. Devroye. The equivalence of weak, strong and complete convergence in \mathbb{R}^d for kernel density estimates. *The Annals of Statistics*, 11(3) :896–904, 1983.

- [DH97] A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, 1997.
- [DJ06] J. F. Delmas and B. Jourdain. *Modèles aléatoires. Application aux sciences de l'ingénieur et du vivant*, volume 57 of *Mathématiques et Applications (SMAI)*. Springer, 2006.
- [DMZ08] Y. Davydov, I. Molchanov, and S. Zuyev. Strictly stable distributions on convex cones. *Electron. J. Probab.*, 13 :259–321, 2008.
- [DP99] Y. Davydov and V. Paulauskas. On the estimation of the parameters of multivariate stable distributions. *Acta Applicandae Mathematicae : An International Survey Journal on Applying Mathematics and Mathematical Applications*, 58(1) :107–124, 1999.
- [DPR00] Y. Davydov, V. Paulauskas, and A. Račkauskas. More on p -stable convex sets in banach spaces. *Journal of Theoretical Probability*, 13(1) :39–64, 2000.
- [DuM71] W. H. DuMouchel. *Stable Distributions in Statistical Inference*. PhD thesis, Department of Statistics, Yale University, 1971.
- [DuM73] W. H. DuMouchel. On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics*, 1(5) :948–957, 1973.
- [DVJ03] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes*. Springer, New York, 2003.
- [Fel71] W. Feller. *An Introduction to Probability Theory and its Applications*. Wiley, New York, 1971.
- [Fer81] J. Fernandez. New and evolved comets in the solar system. *Astronomy and Astrophysics*, 96(1) :26–35, 1981.
- [FR68] E. F. Fama and R. Roll. Some properties of symmetric stable distributions. *Journal of the American Statistical Association*, 63(323) :817–836, 1968.
- [FR71] E. F. Fama and R. Roll. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, 66(334) :331–338, 1971.
- [Fri72] B. Fristedt. Expansions for the density of the absolute value of a strictly stable vector. *Ann. Math. Stat.*, 43(2) :669–672, 1972.
- [GH85] E. Giné and M. G. Hahn. Characterization and domains of attraction of p -stable random compact sets. *Ann. Probab.*, 13(2) :447–468, 1985.
- [GHZ82] E. Giné, M. G. Hahn, and J. Zinn. Limit theorems for random sets : An application of probability in banach space results. *Probability in Banach spaces IV, Proc. Semin., Oberwolfach*, 990 :112–135, 1982.

- [Hil75] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3 :1163–1174, 1975.
- [HL06] H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publications de l'Institut Mathématique*, 80(94) :121–140, 2006.
- [Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301) :13–30, 1963.
- [JM06] A. H. Jessen and T. Mikosch. Regularly varying functions. *Publications de L'Institut Mathématique, Nouvelle Serie*, (94) :171–192, 2006.
- [JMRS07] M. Jacobsen, T. Mikosch, J. Rosinski, and G. Samorodnitsky. Inverse problems for regular variation of linear filters, a cancellation property for σ -finite measures, and identification of stable laws. 2007.
- [LWZ81] R. LePage, M. Woodroffe, and J. Zinn. Convergence to a stable distribution via order statistics. *Ann. Prob.*, 9 :624–632, 1981.
- [McC86] J.H. McCulloch. Simple consistent estimations of stable distribution parameters. *Commun. Statist.-Simila.*, 15(4) :1109–1136, 1986.
- [McC00] J. H. McCulloch. Estimation of the bivariate stable spectral representation by the projection method. *Computational Economics, Springer*, 16(1) :47–62, 2000.
- [Mik03] T. Mikosch. Modeling dependence and tails of financial time series. *Extreme Values in Finance, Telecommunications, and the Environment*, pages 185–286, 2003.
- [MN94] R. Modarres and J. P. Nolan. A method for simulating stable random vectors. *Computational Statistics*, 9 :11–19, 1994.
- [Mol05] I. S. Molchanov. *Theory of Random Sets*. Springer, London, 2005.
- [MR93] S. Mittnik and S. T. Rachev. Modeling asset returns with alternative stable distributions. *Econometric reviews*, 12(3) :261–330, 1993.
- [Nag01] A. V. Nagaev. On nonparametric estimation of the poisson spectral measure of a stable law. *Journal of Mathematical Sciences*, 106(2) :2854–2859, 2001.
- [Nol97] J. P. Nolan. Numerical calculation of stable densities and distribution functions. *Stochastic Models*, 13(4) :759–774, 1997.
- [Nol05] J. P. Nolan. Multivariate stable densities and distribution functions : general and elliptical case. Deutsche Bundesbank's Annual Fall Conference, 2005.
- [NPM01] J. P. Nolan, A. K. Panorska, and J. H. McCulloch. Estimation of stable spectral measures. *Mathematical and Computer Modelling*, 34 :1113–1122, 2001.

- [NR95] J. P. Nolan and B. Rajput. Calculation of multidimensional stable densities. *Communications in Statistics-Simulation and Computation*, 24(3) :551–566, 1995.
- [NS95] C. L. Nikias and M. Shao. *Signal processing with alpha-stable distributions and applications*. Wiley-Interscience, New York, 1995.
- [Pau03] V. Paulauskas. A new estimator for a tail index. *Acta Appl. Math*, 79 :55–67, 2003.
- [Pre72a] S. J. Press. *Applied Multivariate Analysis*. Holt Rinehart and Winston, New York, 1972.
- [Pre72b] S. J. Press. Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, 67(340) :842–846, 1972.
- [Rac91] S. T. Rachev. *Probability metrics and the stability of stochastic models*. Wiley, Chichester, 1991.
- [Rei93] R. D. Reiss. *A Course on Point Process*. Springer-Verlag, Berlin, 1993.
- [Res87] S. I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, Berlin, 1987.
- [Res92] S. I. Resnick. *Adventures in Stochastic Processes*. Birkhauser, Boston, 1992.
- [Res06] S. I. Resnick. *Heavy-Tail Phenomena : Probabilistic And Statistical Modeling*. Springer, New York, 2006.
- [Ros70] H. P. Rosenthal. On the subspaces of l^p ($p > 2$) spanned by sequences of independent random variables. *Israel J. Math.*, 8 :273–303, 1970.
- [RS97] S. Resnick and C. Starica. Smoothing the hill estimator. *Advances in Applied Probability*, 29 :271–293, 1997.
- [Rva62] E. Rvaceva. On domains of attraction of multi-dimensional distributions. *Select. Transl. Math. Statist. and Probab.*, 2 :183–205, 1962.
- [RVF02] H. Rickman, G. B. Valsecchi, and C. Froeschl. From the oort cloud to observable short-period comets-i. *The initial stage of cometary capture*. *MNRAS*, 325 :1303–1311, 2002.
- [RW91] S. I. Resnick and E. Willekens. Moving averages with random coefficients and random coefficient autoregressive models. *Communications in statistics. Stochastic models*, 7 :511–525, 1991.
- [RX93] S. T. Rachev and H. Xin. Test for association of random variables in the domain of attraction of multivariate stable law. *Probability and mathematical statistics*, 14 :125–141, 1993.

-
- [SK74] B. W. Stuck and B. Kleiner. A statistical analysis of telephone noise. *Bell Syst. Tech. J.*, 53(7) :1263–1320, 1974.
- [ST94] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York, 1994.
- [Wat83] G. S. Watson. *Statistics on spheres*. Wiley, New York, 1983.
- [Wer96] R. Weron. On the chambers-mallows-stuck method for simulating skewed stable random variables. *Stat. and Prob. Lett.*, 28(2) :165–171, 1996.
- [Zol86] V. M. Zolotarev. *One-dimensional Stable Distributions*, volume 65. American Mathematical Society, translation from the original 1983 russian edition. edition, 1986.