



**HAL**  
open science

# Estimation et tests en théorie des valeurs extrêmes

Gwladys Toulemonde

► **To cite this version:**

Gwladys Toulemonde. Estimation et tests en théorie des valeurs extrêmes. Mathématiques [math]. Université Pierre et Marie Curie - Paris VI, 2008. Français. NNT: . tel-00348589

**HAL Id: tel-00348589**

**<https://theses.hal.science/tel-00348589>**

Submitted on 19 Dec 2008

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

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

THÈSE DE DOCTORAT  
DE L'UNIVERSITÉ PARIS VI - PIERRE ET MARIE CURIE

École Doctorale de Sciences Mathématiques de Paris Centre

Spécialité : Statistique Mathématique

Présentée par : Gwladys TOULEMONDE

Pour obtenir le grade de  
DOCTEUR DE L'UNIVERSITÉ PARIS VI

**Sujet de la thèse :**  
**Estimation et tests en théorie des valeurs extrêmes**

Soutenue le 30 octobre 2008

devant le jury composé de :

Mlle Armelle GUILLOU	Directrice de thèse
Mme Marie KRATZ	Rapporteur
M. Holger ROOTZÉN	Rapporteur
M. Paul DEHEUVELS	Examineur
M. Philippe NAVEAU	Examineur



## Remerciements

Il m'est particulièrement agréable aujourd'hui de remercier toutes les personnes qui m'ont aidé de près ou de loin à mener à bien ce travail, même si ces remerciements sont bien peu de choses au regard du nombre d'échanges fructueux dont j'ai eu la chance de bénéficier au cours de ces trois dernières années.

Je pense en premier lieu à Armelle Guillou, ma directrice de thèse qui a su m'encadrer mais aussi m'épauler sans relâche pendant ces trois ans. Sa rigueur scientifique, sa ténacité, son sens de l'écoute et son extrême patience sont autant de qualités sans lesquelles cette thèse n'aurait jamais vu le jour. Sincèrement MERCI!

Je suis très honorée que Marie Kratz et Holger Rootzén aient accepté la si difficile tâche de rapporteur. Merci sincèrement du temps et de l'énergie que vous avez consacré à la lecture de mon travail.

J'adresse toute ma gratitude à Philippe Naveau pour son implication dans mon doctorat à travers ses conseils et les nombreuses discussions particulièrement riches et variées que nous avons eues. Je pense également à Michael Falk avec qui j'ai eu l'immense honneur d'une première collaboration et qui par sa patience et son sens de l'écoute a su me donner plus de confiance en moi.

C'est aussi dans le bureau 8A27 que j'ai fait de précieuses rencontres. Je pense aux premiers temps de mon arrivée, époque que l'on peut qualifier « d'italienne » avec Samuela, Rosalba, Anna-Maria, Jean-Baptiste, Omar. Et puis Pierre. Merci Pierre pour ta générosité, ta présence et tes conseils qui ont fait de toi un véritable coach! Merci aussi à Ségo pour tous tes encouragements et ton aide sans limite. Je n'oublie pas Julien qui a également traversé avec moi toutes les étapes et cela depuis notre arrivée au DEA, merci pour ton soutien, pour ces discussions au tableau et ta bonne humeur. Merci aussi à Olivier et Philippe pour toutes ces discussions que nous avons eues (mathématiques? euh... pas toujours!) Une pensée pour Olive et nos nombreuses soirées soupe de pâtes au

bureau, merci pour ta patience et ton aide ! Et puis je n'oublie pas les filles, Claire et ses tongs, Clara et Jeanne ; un peu de douceur oh combien appréciable, ces pauses thés, discussions voyages, théâtre, partage de doutes aussi. Vos sourires et votre infinie gentillesse me manqueront.

Et d'autres avec qui je n'ai pas partagé de bureau mais qui par leurs sourires et leurs encouragements m'ont aidé à avancer, merci Salim, François-Xavier, Lahcen, Nabil, Boris, Véronique et tant d'autres.

Et quand découragée je venais frapper à la porte de Louise, toujours un sourire, que dis-je plutôt deux me redonnaient envie d'avancer et d'y croire. Merci également à Anne et à Pascal pour leur efficacité et leur gentillesse.

Enfin sans vouloir m'étendre sur ma vie en dehors du labo, je tiens à remercier mes amis ainsi que ma famille pour n'avoir jamais douté et m'avoir tant encouragée ! Merci à Laetitia et Lahcen avec qui j'ai fait mes premiers pas en statistiques. Merci aussi à Anissa et Mathilde qui m'ont beaucoup aidée notamment lors de mon arrivée à Paris. Je pense aussi à Nath et Sousou, pour s'être autant intéressées à mon travail et pour m'avoir soutenue notamment ces derniers mois ! Merci aussi à Billy et Esterina qui ont bien voulu me servir de public et qui m'ont tant rassurée. Je ne saurais oublier mon papa et ma maman qui ont toujours cru en moi, qui m'ont laissé faire mes choix et qui ont su m'aider et me soutenir chaque fois que j'en ai eu besoin. Un grand merci pour tout !!! Je tiens également à remercier ma soeur Virginie pour sa bienveillance. Merci également à Jean-Christophe pour toutes ces discussions intéressantes que nous avons eues sur nos travaux de recherche respectifs, sur le métier d'enseignant-chercheur aussi. Et puis Guillaume qui aura été d'un soutien infaillible. Un immense merci à toi même si tout grand qu'il soit, il reste dérisoire par rapport à tout ce que tu m'as apporté et m'apportes encore.

## Résumé

### Estimation et tests en théorie des valeurs extrêmes

Cette thèse se décompose en trois parties distinctes auxquelles s'ajoute une introduction. Dans un premier temps, nous nous intéressons à un test lisse d'ajustement à la famille de Pareto. Pour cela, nous proposons une statistique de test motivée par la théorie de LeCam sur la normalité asymptotique locale (LAN). Nous en établissons le comportement asymptotique sous l'hypothèse que l'échantillon provient d'une distribution de Pareto et sous des alternatives locales, nous plaçant ainsi dans le cadre LAN. Des simulations sont présentées afin d'étudier le comportement de la statistique de test à distance finie. Dans le chapitre suivant, nous nous plaçons dans le cadre de données censurées aléatoirement à droite. Nous proposons alors un estimateur des paramètres de la distribution de Pareto généralisée basé sur une première étape de l'algorithme de Newton-Raphson. Nous établissons la normalité asymptotique de cet estimateur. Par des simulations, nous illustrons son comportement à distance finie et le comparons à celui de l'estimateur du maximum de vraisemblance. Nous proposons enfin, dans un dernier chapitre, un modèle linéaire autorégressif adapté à la loi de Gumbel pour prendre en compte la dépendance dans les maxima. Nous établissons des propriétés théoriques de ce modèle et par simulations nous illustrons son comportement à distance finie. Enfin, comme des applications concrètes en sciences du climat motivaient ce modèle, nous l'avons utilisé pour modéliser des maxima de dioxyde de carbone et de méthane.

**Mots Clés :** Théorie des valeurs extrêmes, Tests d'ajustement, Test lisse de Neyman, Théorie de LeCam, Algorithme de Newton-Raphson, Censure aléatoire, Modèle autorégressif (AR).

## Abstract

### Estimation and tests in extreme value theory

This thesis is divided into three parts with an additional introduction. In the first part, we propose a smooth goodness-of-fit test for the Pareto distribution family. This test is based on LeCam's theory of local asymptotic normality (LAN). We establish the behaviour of our test statistic firstly under the null hypothesis that the sample follows a Pareto distribution and secondly under local alternatives using the LAN framework. We also expose some simulation results in order to study the finite sample behaviour of the test statistic. In the next chapter, we are interested in the topic of extreme value theory under random censoring. We propose an estimator of the two parameters of the generalized Pareto distribution based on the Newton-Raphson algorithm for which we establish the asymptotic normality. A simulation study illustrates its behaviour and also compares our estimator to the maximum likelihood estimator. In the last chapter, we deal with an autoregressive model adapted to the Gumbel distribution to take the dependence between maxima into account. We derive some theoretical properties of this model and present some simulation results in order to assess the quality of our model on finite samples. Finally, since this model is motivated by practical applications in atmospheric sciences, we fit it on daily and weekly maxima of carbon dioxide and methane.

**Key words :** Extreme Value Theory, Goodness-of-fit test, Neyman smooth test, LeCam's theory, Newton-Raphson algorithm, Random censoring, Autoregressive (AR) model.

# Table des matières

Remerciements	3
Résumé	5
Abstract	6
Publications et Conférences	11
<b>1 Introduction</b>	<b>13</b>
1.1 Domaines d'attraction . . . . .	13
1.2 Estimation des paramètres . . . . .	19
1.2.1 Estimateurs de l'indice des valeurs extrêmes $\gamma$ . . . . .	19
1.2.2 Estimateurs du couple $(\gamma, \sigma)$ . . . . .	21
1.3 Présentation des travaux de recherche . . . . .	24
<b>2 Un test de type Neyman pour les distributions de Pareto</b>	<b>27</b>
2.1 Introduction . . . . .	28
2.2 Notations et définitions . . . . .	31
2.3 Le comportement sous $H_0$ de la statistique de test proposée . . . . .	34
2.4 Le comportement sous $H_{1,n}$ de la statistique de test proposée . . . . .	37
2.4.1 Résultats préliminaires . . . . .	37
2.4.2 Puissance du test . . . . .	39



2.5	Simulations . . . . .	39
2.5.1	Paramètres et méthodes . . . . .	39
2.5.2	Résultats . . . . .	40
2.6	Remarques générales et lien avec la théorie des valeurs extrêmes . . . . .	42
2.7	Preuve des résultats . . . . .	43
2.7.1	Preuve de la Proposition 2.1 . . . . .	43
2.7.2	Preuve de la Proposition 2.2 . . . . .	48
2.7.3	Preuve de la Proposition 2.3 . . . . .	53
2.7.4	Preuve de la Proposition 2.4 . . . . .	54
2.7.5	Preuve de la Proposition 2.5 . . . . .	55
2.7.6	Preuve de la Proposition 2.6 . . . . .	57
2.7.7	Preuve de la Proposition 2.7 . . . . .	57
2.8	Annexes . . . . .	59
2.8.1	Définitions d'après Pfanzagl [56] . . . . .	59
2.8.2	Théorie de LeCam - Normalité Asymptotique Locale . . . . .	62
2.8.3	Simulation par la méthode de rejet . . . . .	64
<b>3</b>	<b>Estimation en présence de censure</b>	<b>67</b>
3.1	Introduction . . . . .	68
3.2	Résultat principal . . . . .	71
3.3	Exemples et résultats de simulations . . . . .	74
3.3.1	Différents exemples . . . . .	75
3.3.2	Simulations . . . . .	77
3.4	Preuve du théorème . . . . .	82
3.5	Annexes . . . . .	86
3.5.1	Annexe 1 : Calcul des dérivées premières et secondes . . . . .	86
3.5.2	Annexe 2 : Justification de la matrice $W$ . . . . .	88

---

3.5.3	Annexe 3 : Deux exemples d'intégrale à calculer . . . . .	90
<b>4</b>	<b>Un modèle linéaire autorégressif pour des maxima et applications</b>	<b>95</b>
4.1	Introduction . . . . .	96
4.2	Modèle autorégressif pour la distribution de Gumbel . . . . .	98
4.2.1	Description du modèle proposé : résultat théorique . . . . .	98
4.2.2	Estimation des paramètres du modèle . . . . .	101
4.2.3	Distribution jointe du vecteur $\mathbf{X}_h = (X_t, \dots, X_{t-h})^t$ , $h > 0$ . . . . .	103
4.2.4	Dépendance asymptotique . . . . .	103
4.3	Résultats de simulation . . . . .	104
4.4	Applications aux données de méthane et de dioxyde de carbone . . . . .	108
4.5	Généralisation à la distribution GEV . . . . .	118
4.6	Preuves des résultats . . . . .	119
4.6.1	Preuve de la Proposition 4.1 . . . . .	119
4.6.2	Preuve de la Proposition 4.2 . . . . .	121
4.6.3	Preuve de la Proposition 4.3 . . . . .	121
4.6.4	Preuve du Lemme 4.1 . . . . .	127
4.6.5	Preuve de la Proposition 4.4 . . . . .	127
4.6.6	Preuve de la Proposition 4.5 . . . . .	129
	<b>Conclusion et Perspectives</b>	<b>131</b>
	<b>Bibliographie</b>	<b>135</b>



## Publications et Conférences

### Articles publiés :

[1] J. Ryan, I. Carrière, K. Ritchie, R. Stewart, G. Toulemonde, J.F. Dartigues, C. Tzourio, M.L. Ancelin, Late-life depression and mortality : influence of gender and antidepressant use, *The British Journal of Psychiatry* (2008), 192, 12-18.

[2] M. Falk, A. Guillou, G. Toulemonde, A LAN based Neyman smooth test for Pareto distributions, *Journal of Statistical Planning and Inference* (2008), 138, 2867-2886.

### Articles soumis :

[3] J. Beirlant, A. Guillou, G. Toulemonde, Peaks-Over-Threshold modeling under random censoring.

[4] G. Toulemonde, A. Guillou, P. Naveau, M. Vrac, Autoregressive models for maxima and their applications to CH4.

### Conférences :

[1] XXXVIIIèmes Journées de la Statistique, SFDS, Clamart, France, 2006.

[2] 5th Conference on Extreme Value Analysis, Berne, Suisse, 2007.

[3] Statistical modeling of extremes in data assimilation and filtering approaches, Strasbourg, France, 2008.

[4] Journées MAS de la SMAI, Valeurs extrêmes et applications, Rennes, France, 2008.

**Séminaires :**

[5] Groupe de Travail des Thésards (GTT) du Laboratoire de Statistique Théorique et Appliquée, Université Paris 6, France, 2007.

[6] Séminaire de Probabilités et Statistique, Université Montpellier II, Montpellier SupAgro-INRA, CIRAD, France, 2007.

[7] Extreme Values Reading Group, Institut Pierre Simon Laplace, Paris, France, 2008.

# Chapitre 1

## Introduction

### 1.1 Domaines d'attraction

Au cours des dernières décennies, nous avons pu observer dans le monde scientifique, un intérêt renouvelé pour la modélisation et l'analyse statistique des événements rares. L'étude des valeurs extrêmes revient à l'étude des queues de distributions de fonctions, ou de façon équivalente, à l'analyse de la plus grande observation d'un échantillon. En ce sens, nous pouvons considérer la théorie des valeurs extrêmes comme la contrepartie de la théorie statistique classique, qui est principalement basée sur l'étude de la moyenne d'un échantillon plutôt que des observations extrêmes.

L'analyse des valeurs extrêmes repose principalement sur les distributions limites des extrêmes et leurs domaines d'attraction. Ces distributions apparaissent comme les seules distributions limites possibles du maximum d'un échantillon de variables aléatoires indépendantes et identiquement distribuées (iid)  $X_1, \dots, X_n$  de loi  $F$ . Plus spécifiquement, on dit que  $F$  est dans le domaine d'attraction  $D_\gamma$  si et seulement si il existe deux suites

normalisantes  $(a_n)$ ,  $a_n \in \mathbb{R}^+$  et  $(b_n)$ ,  $b_n \in \mathbb{R}$  telles que  $\forall x \in \mathbb{R}$  :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ a_n^{-1} \left( \max_{1 \leq i \leq n} X_i - b_n \right) \leq x \right\} = H_\gamma(x) \quad (1.1)$$

où

$$H_\gamma(x) = \begin{cases} \exp\left(- (1 + \gamma x)^{-\frac{1}{\gamma}}\right) & \text{pour tout } x \text{ tel que } 1 + \gamma x > 0, \text{ si } \gamma \neq 0, \\ \exp\left(- \exp(-x)\right) & \text{pour tout } x \in \mathbb{R}, \text{ si } \gamma = 0, \end{cases}$$

s'appelle la distribution des valeurs extrêmes généralisées (distribution GEV).

De ce théorème limite, il découle immédiatement que le comportement de la queue de distribution d'une fonction est complètement caractérisé par un unique paramètre, noté  $\gamma$ , et appelé indice des valeurs extrêmes. Le signe de ce paramètre est un indicateur essentiel sur la forme de la queue de distribution. Il faut donc distinguer les trois cas possibles.

Précisons que dans toute la suite, nous noterons par  $\tau_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$  le point terminal de  $F$  et rappelons qu'une fonction  $\ell$  est dite à variation lente si et seulement si elle peut s'écrire de la manière suivante :

$$\ell(x) = c(x) \exp \left\{ \int_a^x \varepsilon(u) \frac{du}{u} \right\}, \quad (x \geq a)$$

pour  $a > 0$ ,  $c(\cdot)$  une fonction mesurable telle que  $c(x) \rightarrow c \in (0, \infty)$  et  $\varepsilon(\cdot)$  une fonction telle que  $\varepsilon(x) \rightarrow 0$  quand  $x \rightarrow \infty$ . Une des caractéristiques souvent utilisées des fonctions à variation lente est que toute fonction mesurable sur  $(0, \infty)$  satisfaisant la condition de convergence suivante :

$$\lim_{x \rightarrow \infty} \frac{\ell(tx)}{\ell(x)} = 1, \text{ pour tout } t > 0$$

est à variation lente. Ces fonctions à variation lente sont utiles quand on veut définir la notion de fonction à variation régulière. On dit en effet que  $G$  est à variation régulière d'ordre  $\rho$ , noté  $G \in \mathcal{R}_\rho$ , si  $G(x) = x^\rho \ell(x)$  avec  $\ell$  une fonction à variation lente. De plus,

comme bien souvent nous n'avons besoin que d'un équivalent asymptotique des fonctions à variation lente, nous ne perdons rien à supposer la fonction  $c(\cdot)$  constante. Dans ce cas, nous parlerons de fonction à variation lente normalisée et alors, presque partout, nous avons

$$\varepsilon(x) = x\ell'(x)/\ell(x)$$

(cf. Bingham *et al.* [11], p. 15).

Nous sommes donc maintenant à même de décrire les trois situations possibles suivantes :

- $\gamma > 0$  correspond au domaine d'attraction de Fréchet. On a alors  $\tau_F = \infty$  et  $F$  est de type Pareto, c'est-à-dire  $1 - F(x) = x^{-1/\gamma}\ell_F(x)$  où  $\ell_F$  est une fonction à variation lente ;
- $\gamma = 0$  correspond au domaine d'attraction de Gumbel. Le point terminal  $\tau_F$  peut alors être fini ou non et les queues des lois appartenant à ce domaine décroissent de manière exponentielle ;
- $\gamma < 0$  correspond au domaine d'attraction de Weibull. On a alors  $\tau_F < \infty$  et  $1 - F(x) = (\tau_F - x)^{-1/\gamma}\ell_F((\tau_F - x)^{-1})$  où  $\ell_F$  est également une fonction à variation lente.

La Figure 1.1 ci-dessous illustre le comportement de différentes distributions GEV correspondant à différentes valeurs de  $\gamma$ . Les Tables 1.1 à 1.3 donnent quant à elles différents exemples de distributions standards dans ces trois domaines d'attraction.



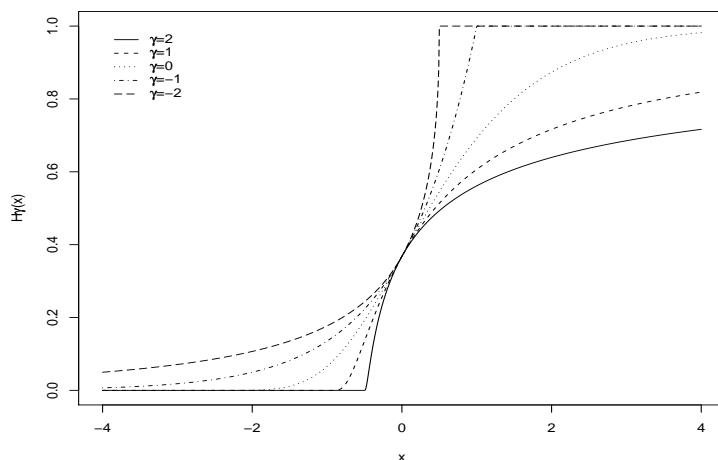


FIG. 1.1 – Distributions des valeurs extrêmes généralisées

Distribution	$1 - F(x)$	$\gamma$
Burr( $\beta, \tau, \lambda$ ), $\beta > 0, \tau > 0, \lambda > 0$	$\left(\frac{\beta}{\beta+x^\tau}\right)^\lambda$	$\frac{1}{\lambda\tau}$
Fréchet( $\frac{1}{\alpha}$ ), $\alpha > 0$	$1 - \exp(-x^{-\alpha})$	$\frac{1}{\alpha}$
Loggamma( $m, \lambda$ ), $m > 0, \lambda > 0$	$\frac{\lambda^m}{\Gamma(m)} \int_x^\infty (\log u)^{m-1} u^{-\lambda-1} du$	$\frac{1}{\lambda}$
Loglogistic( $\beta, \alpha$ ), $\beta > 0, \alpha > 1$	$\frac{1}{1+\beta x^\alpha}$	$\frac{1}{\alpha}$
Pareto( $\alpha$ ), $\alpha > 0$	$x^{-\alpha}$	$\frac{1}{\alpha}$

TAB. 1.1 – Quelques distributions de type Pareto associées à un indice positif

Distribution	$1 - F(x)$	$\gamma$
Gamma( $m, \lambda$ ), $m \in \mathbb{N}, \lambda > 0$	$\frac{\lambda^m}{\Gamma(m)} \int_x^\infty u^{m-1} \exp(-\lambda u) du$	0
Gumbel( $\mu, \beta$ ), $\mu \in \mathbb{R}, \beta > 0$	$\exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right)$	0
Logistic	$\frac{2}{1+\exp(x)}$	0
Lognormale( $\mu, \sigma$ ), $\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{1}{u} \exp\left(-\frac{1}{2\sigma^2}(\log u - \mu)^2\right) du$	0
Weibull( $\lambda, \tau$ ), $\lambda > 0, \tau > 0$	$\exp(-\lambda x^\tau)$	0

TAB. 1.2 – Quelques distributions associées à un indice nul

Distribution	$1 - F(x)$	$\gamma$
ReverseBurr( $\beta, \tau, \lambda, \tau_F$ ), $\beta > 0, \tau > 0, \lambda > 0$	$\left(\frac{\beta}{\beta + (\tau_F - x)^{-\tau}}\right)^\lambda$	$-\frac{1}{\lambda\tau}$
Uniforme(0, 1)	$1 - x$	$-1$

TAB. 1.3 – Quelques distributions associées à un indice négatif

Suivant la valeur de  $\gamma$ , une façon de caractériser l'appartenance au domaine d'attraction  $D_\gamma$  est la suivante.

**Théorème 1.1 :**

La distribution  $F$  appartient au domaine d'attraction  $D_\gamma$  si et seulement si il existe une fonction mesurable positive,  $\sigma(\cdot)$ , telle que pour tout  $1 + \gamma x > 0$ , on ait

$$\lim_{t \uparrow \tau_F} \frac{1 - F(t + x\sigma(t))}{1 - F(t)} = \begin{cases} (1 + \gamma x)^{-1/\gamma} & \text{si } \gamma \neq 0, \\ e^{-x} & \text{si } \gamma = 0. \end{cases}$$

Bien que très utilisée en pratique, cette approche des extrêmes basée sur la loi limite du maximum d'un échantillon de variables aléatoires iid a été fortement critiquée dans la littérature. En effet, elle ne tient compte que d'une seule observation, la plus grande. On a donc le sentiment de perdre de l'information et notamment toute celle contenue dans les autres grandes valeurs de l'échantillon. Une alternative a donc été proposée. Elle fait suite aux travaux de Balkema et de Haan [4], ainsi qu'à ceux de Pickands [57]. Elle repose sur la loi des excès au-delà d'un seuil fixe  $t$  et est communément appelée l'approche « pics au-delà d'un seuil » (approche POT, « Peaks-Over-Threshold »). L'idée est la suivante : partant d'un échantillon  $X_1, \dots, X_n$ , on se fixe un seuil  $t$  grand. On ne considère que les  $N_t$  observations dépassant ce seuil. On note  $Y_i, i = 1, \dots, N_t$ , les excès au-delà du seuil  $t$ , définis comme l'écart entre l'observation et  $t$ . La fonction de répartition des excès au-delà de  $t$  est alors donnée par

$$F_t(y) = \mathbb{P}(Y \leq y | X > t) = \mathbb{P}(X - t \leq y | X > t) = \frac{F(t + y) - F(t)}{1 - F(t)}.$$

**Théorème 1.2** (Pickands [57]) *F* appartient au domaine d'attraction  $D_\gamma$  si et seulement si il existe une fonction  $\sigma(\cdot)$  positive et un réel  $\gamma$  tels que la loi des excès  $F_t$  peut être uniformément approchée par une distribution de Pareto généralisée (GPD) notée  $G_{\gamma,\sigma}$ , i.e.

$$\lim_{t \uparrow \tau_F} \sup_{x \in (0, \tau_F - t)} |F_t(x) - G_{\gamma, \sigma(t)}(x)| = 0,$$

$$\text{où } G_{\gamma, \sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, x \geq 0 \text{ et } x < -\frac{\sigma}{\gamma} \text{ si } \gamma < 0 \\ 1 - \exp\left(-\frac{x}{\sigma}\right) & \text{si } \gamma = 0, x \geq 0. \end{cases}$$

Dans la Figure 1.2 ci-dessous, nous illustrons le comportement de différentes GPD pour  $\sigma = 1$  et différentes valeurs de  $\gamma$ . Il est intéressant de noter que le cas  $\gamma = 0$  correspond à la distribution exponentielle de moyenne  $\sigma$  et que le cas  $\gamma = -1$  correspond à la loi uniforme sur  $[0, \sigma]$ .

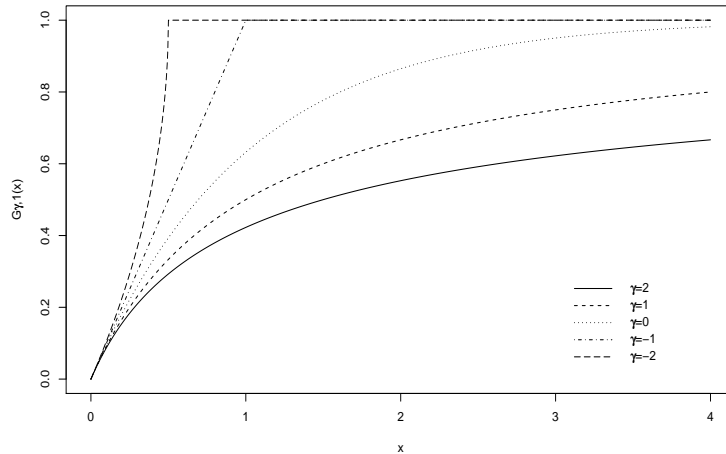


FIG. 1.2 – Distributions de Pareto généralisées  $G_{\gamma,1}$

Ces deux façons de caractériser l'appartenance à un domaine d'attraction (approches GEV et POT) ont en commun le problème de l'estimation de l'indice des valeurs extrêmes

$\gamma$  ou du couple  $(\gamma, \sigma)$ . Ce problème a été longuement étudié dans la littérature et nous présentons dans la section suivante quelques approches permettant de l'aborder.

## 1.2 Estimation des paramètres

### 1.2.1 Estimateurs de l'indice des valeurs extrêmes $\gamma$

#### a. Estimateur de Hill

Parmi les estimateurs les plus répandus, on trouve l'estimateur de Hill [40] défini de la façon suivante :

$$\hat{\gamma}_{X,k,n}^{(H)} = \frac{1}{k} \sum_{j=1}^k \log X_{n-j+1,n} - \log X_{n-k,n}$$

avec  $X_{1,n}, \dots, X_{n,n}$  les statistiques d'ordre associées à l'échantillon  $X_1, \dots, X_n$ . Cet estimateur est très populaire, pour différentes raisons. En premier lieu, si on note  $E_{j,t}$  les excès relatifs au-delà de  $t$ , i.e.  $E_{j,t} := \frac{X_j}{t}$  avec  $X_j > t$ , on peut facilement vérifier que  $\mathbb{P}(E_{j,t} > x | E_{j,t} > 1) \rightarrow x^{-1/\gamma}$ , quand  $t \rightarrow \infty$ ,  $x > 1$ . En formant la vraisemblance basée sur cette distribution limite, on vérifie facilement que l'estimateur de Hill n'est rien d'autre que l'estimateur du maximum de vraisemblance dans le cas où le seuil  $t = X_{n-k,n}$ . En second lieu, un côté très attrayant de l'estimateur de Hill est qu'il est possible de l'interpréter graphiquement. Ceci est particulièrement important pour les praticiens, qui préfèrent souvent des interprétations graphiques à des formules mathématiques. Plus précisément, si l'on considère le graphe de coordonnées

$$\left( \log \frac{n+1}{j}, \log X_{n-j+1,n} \right),$$

appelé communément « Pareto quantile plot », dans le cas de distributions de type Pareto, ce graphe sera approximativement linéaire, dans les points extrêmes, avec une pente  $\gamma$ . L'estimateur de Hill n'est alors rien d'autre qu'un estimateur naïf de cette pente et donc de  $\gamma$ .

Les propriétés asymptotiques de l'estimateur de Hill ont été établies par Mason [52] et Deheuvels *et al.* [25] pour la consistance et par Beirlant et Teugels [8] et Csörgö *et al.* [18], entre autres, pour la normalité. Mais le principal inconvénient de cet estimateur est qu'il n'est valable que dans le cas d'un indice positif. Différentes généralisations ont été proposées. Parmi elles, on peut citer l'estimateur des moments (Dekkers *et al.* [26]) ou encore l'estimateur  $UH$  (Beirlant *et al.* [9]) qui vont être brièvement présentés ci-dessous.

### b. Estimateur des moments

Dekkers *et al.* [26] ont proposé une extension de l'estimateur de Hill en l'estimateur des moments, qui lui est valable quelque soit le signe de l'indice et qui est défini comme suit :

$$\widehat{\gamma}_{X,k,n}^{(M)} = \widehat{\gamma}_{X,k,n}^{(H)} + 1 - \frac{1}{2} \left( 1 - \frac{\widehat{\gamma}_{X,k,n}^{(H)}}{S_{X,k,n}} \right)^{-1}$$

où

$$S_{X,k,n} = \frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1,n} - \log X_{n-k,n})^2. \quad (1.2)$$

La normalité de cet estimateur a été établie par Dekkers *et al.* [26] sous des conditions de régularité convenables. Le problème de cet estimateur est que l'on ne peut pas, contrairement à l'estimateur de Hill, l'interpréter graphiquement. Afin d'apporter une solution à ce problème, une généralisation du « Pareto quantile plot » a été proposée et a donné lieu à l'estimateur  $UH$  présenté ci-dessous.

### c. Estimateur $UH$

L'estimateur  $UH$  proposé par Beirlant *et al.* [9] est basé sur une extension du « Pareto quantile plot » en un « quantile plot généralisé » qui peut être décrit comme suit.

On considère la fonction  $UH$  définie par  $UH(x) := U(x)H(x)$  où  $U$  est la fonction queue suivante :

$$U(x) = \inf \left\{ y : F(y) \geq 1 - \frac{1}{x} \right\}, \quad x > 1$$

et

$$H(x) = \mathbb{E}(\log X - \log U(x) | X > U(x)).$$

L'estimateur empirique de cette fonction évaluée en  $x = n/k$  est le suivant :

$$UH_{X,k,n} = X_{n-k,n} \left( \frac{1}{k} \sum_{j=1}^k \log X_{n-j+1,n} - \log X_{n-k,n} \right) = X_{n-k,n} \widehat{\gamma}_{X,k,n}^{(H)}.$$

En remarquant que la fonction  $UH$  est à variation régulière à l'infini, d'indice  $\gamma$ , il est naturel de considérer le « quantile plot généralisé » dont les points ont pour coordonnées

$$\left( \log \left( \frac{n+1}{j} \right), \log UH_{X,j,n} \right), \quad j = 1, \dots, n.$$

Beirlant *et al.* [9] ont alors proposé l'estimateur suivant pour  $\gamma \in \mathbb{R}$  basé sur une approximation de la pente de ce « quantile plot généralisé » :

$$\widehat{\gamma}_{X,k,n}^{(UH)} = \frac{1}{k} \sum_{j=1}^k \log UH_{X,j,n} - \log UH_{X,k+1,n}.$$

Les propriétés de cet estimateur ont été par ailleurs établies dans ce même article.

Signalons enfin qu'une étude comparative entre ces différents estimateurs en terme d'erreur en moyenne quadratique (MSE) a été effectuée en 2005 par Beirlant *et al.* [5]. Par ailleurs, l'article de Csörgö et Viharos [19] permet quant à lui d'avoir une vue d'ensemble des différents estimateurs de l'indice dans la littérature, parmi lesquels, outre les estimateurs décrits dans cette section, d'autres estimateurs basés sur le « Pareto quantile plot » ou encore sur la méthode des noyaux.

### 1.2.2 Estimateurs du couple $(\gamma, \sigma)$

Suivant le type de problématique que l'on a, on peut avoir besoin non seulement d'un estimateur de  $\gamma$  mais aussi d'un estimateur de  $\sigma$ . C'est typiquement le cas par exemple si l'on s'intéresse à l'estimation d'un quantile extrême. Afin d'estimer ce couple  $(\gamma, \sigma)$ , différentes alternatives s'offrent à nous. En se rappelant le théorème de Pickands (Théorème 1.2) qui nous dit que les excès au-delà d'un seuil peuvent être uniformément

approchés par la distribution de Pareto généralisée, on peut proposer tout simplement d'utiliser cette distribution limite et de chercher des estimateurs des paramètres de cette distribution. Pour cela, la méthode la plus ancienne, et peut être la plus naïve, consiste à utiliser la méthode des moments, décrite ci-dessous.

### a. Estimateurs par la méthode des moments

Dans le cas d'une distribution  $\text{GPD}(\gamma, \sigma)$ , l'estimateur des moments (cf. Hosking et Wallis [41], Section 3.2) est basé sur le fait que

$$\mathbb{E} \left[ \left( 1 + \frac{\gamma X}{\sigma} \right)^r \right] = \frac{1}{1 - r\gamma} \text{ si } 1 - r\gamma > 0.$$

Il en découle l'expression des paramètres en fonction des deux premiers moments  $\mu_1$  et  $\mu_2$

$$\begin{aligned} \sigma &= \frac{1}{2} \mu_1 \left( \frac{\mu_1^2}{\mu_2 - \mu_1^2} + 1 \right) \\ \gamma &= \frac{1}{2} \left( 1 - \frac{\mu_1^2}{\mu_2 - \mu_1^2} \right). \end{aligned}$$

En remplaçant dans les expressions précédentes  $\mu_1$  et  $\mu_2$  par leurs versions empiriques, nous obtenons l'estimateur des moments du couple  $(\hat{\sigma}_n^{MOM}, \hat{\gamma}_n^{MOM})$  défini comme suit :

$$\hat{\sigma}_n^{MOM} = \frac{1}{2} \bar{X} \left( \frac{\bar{X}^2}{\bar{X}^2 - \bar{X}^2} + 1 \right), \quad \hat{\gamma}_n^{MOM} = \frac{1}{2} \left( 1 - \frac{\bar{X}^2}{\bar{X}^2 - \bar{X}^2} \right), \quad (1.3)$$

où  $\bar{X}$  et  $\bar{X}^2$  sont les estimateurs empiriques des moments d'ordre 1 et 2 de l'échantillon.

La normalité asymptotique du couple  $(\hat{\gamma}_n^{MOM}, \hat{\sigma}_n^{MOM})$  peut être établie sous la condition  $\gamma < 1/4$ . Pour affaiblir cette condition d'autres alternatives ont été proposées, telle la méthode des moments pondérés (éventuellement généralisés).

### b. Estimateurs par la méthode des moments pondérés

Hosking et Wallis [41] en 1987 ont également proposé d'utiliser les deux moments pondérés  $\nu_0$  et  $\nu_1$  suivants :

$$\nu_s = \mathbb{E}[X(1 - G_{\gamma, \sigma}(X))^s] \text{ pour } s \geq 0$$

pour estimer  $\gamma$  et  $\sigma$ . Le domaine de validité nécessaire à la normalité asymptotique de ces estimateurs est encore assez restrictif :  $\gamma \in (-1, 1/2)$ ; néanmoins, cette méthode présente un certain nombre d'avantages tant au niveau de la facilité du calcul que des performances à distance finie (cf. Hosking et Wallis [41]). Une généralisation de cette méthode permettant d'étendre le domaine de validité à  $\gamma \in (-1, 3/2)$  a été proposée par Rached [59] ainsi que Diebolt *et al.* [27], [28].

Les estimateurs d'indice de Hill, UH et des moments (cf. Section 1.2.1) peuvent également être utilisés pour en déduire des estimateurs de  $\sigma$ .

### c. Estimateurs de $\sigma$ basés sur des estimateurs de $\gamma$

Dans le cas de la méthode des moments ou de la méthode *UH* on peut proposer l'estimateur de  $\sigma$  suivant (cf. de Haan et Rootzén [38]) :

$$\hat{\sigma}_{X,k,n}^{(\cdot)} = \frac{X_{n-k,n} \sqrt{3(\hat{\gamma}_{X,k,n}^{(H)})^2 - S_{X,k,n}}}{\sqrt{3[\rho_1(\hat{\gamma}_{X,k,n}^{(\cdot)})]^2 - \rho_2(\hat{\gamma}_{X,k,n}^{(\cdot)})}},$$

avec  $S_{X,k,n}$  définie en Equation (1.2) et avec

$$\rho_1(\gamma) = \begin{cases} 1 & \text{si } \gamma \geq 0 \\ \frac{1}{1-\gamma} & \text{si } \gamma < 0 \end{cases}$$

$$\rho_2(\gamma) = \begin{cases} 2 & \text{si } \gamma \geq 0 \\ \frac{2}{(1-\gamma)(1-2\gamma)} & \text{si } \gamma < 0, \end{cases}$$

où  $\hat{\gamma}_{X,k,n}^{(\cdot)}$  peut être indifféremment  $\hat{\gamma}_{X,k,n}^{(M)}$  ou  $\hat{\gamma}_{X,k,n}^{(UH)}$ .

Une autre approche est possible, celle du maximum de vraisemblance, basée à nouveau sur la distribution limite GPD. Elle présente l'intérêt d'avoir de bonnes propriétés asymptotiques, mais a le désavantage de proposer des estimateurs non explicites, solutions d'un système de deux équations à deux inconnues. Ce dernier peut néanmoins



être résolu par des algorithmes, par exemple celui de Newton-Raphson.

#### d. Estimateurs par maximum de vraisemblance

Si on utilise l'approche POT, l'estimation du couple  $(\gamma, \sigma)$  devient alors naturelle, puisqu'elle découle de la maximisation de la fonction de log-vraisemblance. En effet, cette dernière dans le cas d'un échantillon  $\mathbf{X} = (X_1, \dots, X_n)$  issu d'une distribution de Pareto généralisée est donnée par :

$$\mathcal{L}(\mathbf{X}; \gamma, \sigma) = -n \log \sigma - \left( \frac{1}{\gamma} + 1 \right) \sum_{i=1}^n \log \left( 1 + \frac{\gamma}{\sigma} X_i \right).$$

En dérivant cette expression par rapport aux deux paramètres d'intérêt, nous obtenons un système de deux équations à deux inconnues  $\gamma$  et  $\sigma$ . C'est en résolvant ces équations que nous obtenons les estimateurs du maximum de vraisemblance pour lesquels Smith [66] a établi la normalité asymptotique sous la condition  $\gamma > -1/2$ .

### 1.3 Présentation des travaux de recherche

Avant d'ajuster un modèle sur des données, il est indispensable d'en tester l'adéquation. C'est pour cela que nous nous sommes intéressés à un test lisse d'ajustement pour la famille de Pareto. Cette famille a suscité une abondante littérature du fait qu'elle peut être utilisée pour modéliser de nombreux phénomènes physiques. Ce test lisse, initialement proposé par Neyman [55], a été généralisé au test d'adéquation à une loi de probabilité dépendant de paramètres inconnus par Thomas et Pierce [68] en 1979. La densité supposée sous  $H_0$ , notée  $f(x; \beta)$  (où  $\beta$  représente le couple de paramètres à estimer dans notre cas), est alors placée au sein d'une famille exponentielle de densité :

$$g_J(x; \theta, \beta) = f(x; \beta) \exp \left\{ \sum_{s=1}^J \theta_s \bar{F}^s(x; \beta) - K(\theta) \right\} \quad (1.4)$$

où  $\theta = (\theta_1, \dots, \theta_J)^t$ ,  $\bar{F} = 1 - F$ , avec  $F$  la fonction de répartition associée à la densité  $f$  et  $K(\theta)$  une constante de normalisation.

Nous déduisons donc de (1.4) que le test d'ajustement à la distribution de Pareto peut se ramener au test paramétrique suivant :  $H_0 : \theta = 0$  versus  $H_1 : \theta \neq 0$ . Nous proposons alors une statistique de test motivée par la théorie de LeCam sur la normalité asymptotique locale (LAN) pour laquelle nous établissons le comportement asymptotique sous l'hypothèse nulle que l'échantillon provient d'une distribution de Pareto. Puis pour évaluer la puissance de notre test, nous nous intéressons à son comportement asymptotique sous des alternatives locales, nous plaçant ainsi dans le cadre LAN. Des simulations sont présentées afin d'étudier le comportement de la statistique de test à distance finie. Ces résultats sont présentés dans le chapitre 2.

Dans le troisième chapitre, nous ne nous plaçons plus dans le cas d'une loi de Pareto stricte, mais dans le cas général de variables aléatoires iid d'une loi quelconque. Dans beaucoup d'applications pratiques, notamment en médecine et en assurance, les données s'avèrent censurées aléatoirement à droite. Le problème de l'analyse des valeurs extrêmes en présence de données censurées a été abordé dans la littérature très récemment (cf. Beirlant *et al.* [7], Einmahl *et al.* [31]); différents estimateurs d'indice ont été proposés et leur normalité asymptotique établie. En revanche, l'estimateur du maximum de vraisemblance résultant d'une généralisation de la vraisemblance au cadre censuré n'a pas été à ce jour étudié. Il est important de souligner à nouveau que, comme nous n'avons pas d'expression explicite de cet estimateur, il est très difficile d'en établir les propriétés asymptotiques. C'est pourquoi nous avons proposé dans ce chapitre, dans le cadre censuré, un nouvel estimateur du couple  $(\gamma, \sigma)$  basé sur une première étape de l'algorithme de Newton-Raphson, et pour lequel nous avons été à même d'établir la normalité asymptotique. De plus, par des simulations, nous avons illustré que l'estimateur proposé a un comportement, à distance finie, très similaire à celui du maximum de

vraisemblance.

Les chapitres 2 et 3 reposent sur l'hypothèse fondamentale que l'échantillon dont nous disposons est un échantillon de variables indépendantes. Il est clair que, en pratique, cette hypothèse n'est pas toujours satisfaite. Aussi avons nous cherché à prendre en compte dans le dernier chapitre la dépendance dans le cadre de séries temporelles en s'intéressant à l'étude d'un modèle linéaire autorégressif adapté à la distribution de Gumbel. Nous avons établi un certain nombre de propriétés théoriques pour ce modèle et avons également procédé à une étude de simulations pour évaluer la qualité du modèle à distance finie. Enfin, comme ce modèle trouvait son fondement dans des applications concrètes en sciences du climat, nous l'avons utilisé pour modéliser des maxima de dioxyde de carbone et de méthane.

Cette thèse offre de nombreuses perspectives, tant sur le plan pratique que théorique ; une liste non exhaustive sera présentée en conclusion.

# Chapitre 2

## Un test de type Neyman pour les distributions de Pareto

### Résumé

La distribution de Pareto est présente dans beaucoup de domaines et est notamment très utilisée pour modéliser des événements extrêmes. Dans l'esprit des travaux de Neyman [55] et de Thomas et Pierce [68], nous proposons un test lisse d'ajustement pour la famille de Pareto, lequel test est motivé par la théorie de LeCam sur la normalité asymptotique locale (LAN). Dans un premier temps, nous établissons le comportement asymptotique de la statistique de test proposée sous l'hypothèse nulle que l'échantillon provient d'une distribution de Pareto. Dans un second temps, nous nous intéressons à son comportement asymptotique sous des alternatives locales, nous plaçant ainsi dans le cadre LAN. Enfin, nous étudions par simulations le comportement de la statistique de test à distance finie.

## 2.1 Introduction

La distribution de Pareto a été initialement introduite par Vilfredo Pareto pour décrire la distribution des revenus des familles en Suisse. Elle a été par la suite largement appliquée dans des domaines comme l'assurance pour modéliser les déclarations de sinistres (Benktander [10]), en climatologie ou en hydrologie pour décrire des événements extrêmes (Katz et al. [45]), ainsi qu'en économie (Fisk [33]), en finance (Danielsson et de Vries [21]) ou encore en hydrogéologie (Gustafson et Fransson [36]). Pour plus de détails sur les différentes applications possibles de la distribution de Pareto, nous renvoyons le lecteur à Arnold [2]. Une autre raison justifiant l'intérêt porté à cette distribution est que, moyennant une paramétrisation adéquate, cette distribution peut être considérée comme un cas particulier de la distribution de Pareto généralisée, qui joue un rôle essentiel dans l'étude des excès au-delà d'un seuil en théorie des valeurs extrêmes.

Avant d'ajuster un modèle sur des données, il est indispensable d'en tester l'adéquation. Pour cela différents tests d'ajustement peuvent être utilisés dont le test du  $\chi^2$  de Pearson, premier test apparu en 1900. Depuis, différents autres tests ont été proposés (cf. D'Agostino et Stephens [20], Rayner et Best [61]). Vu l'intérêt qu'a suscité la loi de Pareto, il paraît naturel de chercher à construire un test d'ajustement pouvant être utilisé pour savoir si un échantillon de variables aléatoires indépendantes  $X_1, \dots, X_n$  est distribué selon une distribution de Pareto ou non. Pour garder un lien avec la théorie des valeurs extrêmes, nous utilisons la paramétrisation suivante pour la densité associée à la distribution de Pareto :

$$f(x, \beta) = \frac{1}{\sigma} \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}-1} \text{ où } \beta^t = (\sigma, \gamma) \in \Theta = (0, \infty) \times (0, \infty) \text{ et } x \in [0, \infty).$$

Il est clair que les paramètres sous-jacents au modèle sont inconnus dans la plupart des applications et doivent donc être estimés. Différentes techniques d'estimation ont déjà été

proposées (cf. Arnold et Press [3], Castillo et Hadi [14], Davison [23], Malik [51], Smith [65]). Suivant les domaines d'applications, certaines techniques sont plus ou moins populaires : par exemple, en hydrologie, l'approche la plus utilisée est la méthode des moments pondérés (Hosking et Wallis [41]). Cependant, tous domaines confondus, l'approche la plus standard est sans conteste la méthode du maximum de vraisemblance qui conduit à des estimateurs ayant de bonnes propriétés asymptotiques. Toutefois, le problème de ces estimateurs est le fait qu'ils n'admettent pas d'expression explicite. Nous présenterons donc, en Section 2.3, un estimateur ayant les mêmes propriétés asymptotiques que l'estimateur du maximum de vraisemblance mais qui présente l'avantage d'être plus facile à calculer.

Le problème des tests d'ajustement à une distribution de Pareto généralisée n'a pas été beaucoup étudié dans la littérature comme le soulignent Davison et Smith [24] en 1990. Même dans le cas particulier de la distribution de Pareto, ce problème n'a pas suscité une abondante littérature. Porter et al. [58] ont proposé dans ce cadre des statistiques de test basées sur la fonction de répartition empirique, sous l'hypothèse que le paramètre de forme est connu et ont exhibé des valeurs critiques associées à leurs statistiques de test. Plus récemment, Choulakian et Stephens [15] ont présenté des tests d'ajustement également basés sur la fonction de répartition empirique cette fois-ci pour la loi de Pareto généralisée et sous l'hypothèse d'un paramètre de forme éventuellement inconnu. Ces statistiques de test basées sur la fonction de répartition empirique sont consistantes pour la plupart des alternatives. Néanmoins, certaines études empiriques (Kopecky et Pierce [46], Miller et Quesenberry [54], Rayner et Best [60]) ont montré que les tests lisses de type Neyman sont plus puissants que certains tests usuels tels ces tests basés sur la fonction de répartition empirique ou le test du  $\chi^2$  de Pearson.

Initialement, Neyman [55] a introduit en 1937 les tests lisses d'ajustement pour tester l'uniformité et les a développé pour améliorer le test du  $\chi^2$  de Pearson. Depuis, de nombreuses généralisations ont été proposées (cf. Rayner et Best [61], [62], Thomas et

Pierce [68]). L'idée première est de placer la densité supposée sous  $H_0$  au sein d'une famille exponentielle de densité :

$$g_J(x, \theta, \beta) = f(x, \beta) \exp \left\{ \sum_{s=1}^J \theta_s \bar{F}^s(x, \beta) - K(\theta) \right\} \quad (2.1)$$

où  $\theta = (\theta_1, \dots, \theta_J)^t$ ,  $\bar{F} = 1 - F$ , avec  $F$  la fonction de répartition associée à la densité  $f$  et  $K(\theta)$  une constante de normalisation. Nous pouvons facilement constater que  $K(\theta)$  dépend uniquement de  $\theta$  et a pour expression

$$K(\theta) = \log \left\{ \int_0^1 \exp \left( \sum_{s=1}^J \theta_s t^s \right) dt \right\}.$$

Notons que dans (2.1), nous aurions pu remplacer  $\bar{F}^s$  par  $F^s$ , mais nous avons choisi d'utiliser l'expression plus simple de la fonction de survie de Pareto :

$$\bar{F}(x, \beta) = \left( 1 + \frac{\gamma x}{\sigma} \right)^{-\frac{1}{\gamma}} \text{ où } \beta^t = (\sigma, \gamma) \in \Theta = (0, \infty) \times (0, \infty) \text{ et } x \in [0, \infty).$$

Nous déduisons de (2.1) que le test d'ajustement à la distribution de Pareto peut se ramener au test paramétrique suivant :

$$H_0 : \theta = 0$$

$$\text{versus } H_1 : \theta \neq 0.$$

Dans la Section 2.3, nous présenterons le comportement asymptotique sous  $H_0$  de la statistique de test proposée. Comme en pratique les paramètres sont inconnus, nous aurons besoin d'un estimateur pour  $\beta$ . Nous présenterons alors un estimateur  $\hat{\beta}_n$  plus facile à calculer que l'estimateur du maximum de vraisemblance mais ayant les mêmes propriétés asymptotiques. Avec cet estimateur, nous obtiendrons la distribution limite de la statistique de test sous  $H_0$  en utilisant des résultats dus à Pfanzagl [56]. Puis, dans la Section 2.4, nous étudierons le comportement asymptotique de la statistique de test sous des alternatives locales nous plaçant ainsi dans le cadre LAN. Ces dernières sont définies de la

façon suivante :

$$H_{1,n} : \theta = \theta_n = \frac{\delta}{\sqrt{n}}(1 + o(1)) \quad (2.2)$$

où  $\delta = (\delta_1, \dots, \delta_J)^t \neq 0$  avec  $\delta_i \in \mathbb{R}$ , pour tout  $i = 1, \dots, J$ . La notation  $\delta(1 + o(1))$  doit être interprétée comme un vecteur  $(\delta_1(1 + r_1(n)), \dots, \delta_J(1 + r_J(n)))^t$ , où les fonctions  $r_s(n)$  convergent vers 0 quand  $n \rightarrow \infty$ . Dans la section suivante, nous étudierons par simulations le comportement de la statistique de test à distance finie. Enfin nous présenterons quelques conclusions et perspectives pour ce premier chapitre. Les preuves des résultats seront renvoyées en Section 2.7.

## 2.2 Notations et définitions

Nous commençons cette section en introduisant quelques notations que nous utiliserons tout au long de ce chapitre. La famille de mesures de probabilité associée à la densité  $f$  de Pareto est notée  $\{\mathbb{P}_{H_0, \beta}; \beta \in \Theta\}$ . Nous notons  $\xrightarrow{\mathbb{P}_{H_0, \beta}}$  (respectivement  $\xrightarrow{\mathbb{P}_{H_{1,n}}}$ ) la convergence en probabilité sous  $H_0$  (respectivement sous  $H_{1,n}$ ).

Définissons  $\ell^{(\cdot)}(\cdot, \beta)$  et  $\ell^{(\mu\nu)}(\cdot, \beta)$  comme suit :

$$\ell^{(\cdot)}(\cdot, \beta) = \begin{pmatrix} \ell^{(1)}(\cdot, \beta) \\ \ell^{(2)}(\cdot, \beta) \end{pmatrix} = \begin{pmatrix} \frac{\partial \log f(\cdot, \beta)}{\partial \sigma} \\ \frac{\partial \log f(\cdot, \beta)}{\partial \gamma} \end{pmatrix}$$

$$\ell^{(\mu\nu)}(\cdot, \beta) = \frac{\partial^2 \log f(\cdot, \beta)}{\partial \beta[\mu] \partial \beta[\nu]} \text{ pour } \mu, \nu = 1, 2, \text{ avec } \beta[1] = \sigma \text{ et } \beta[2] = \gamma.$$

Des calculs directs nous donnent

$$\ell^{(\cdot)}(x, \beta) = \begin{pmatrix} -\frac{1}{\sigma} + (1 + \gamma) \frac{x}{\sigma^2} \left(1 + \frac{\gamma x}{\sigma}\right)^{-1} \\ \frac{1}{\gamma^2} \log \left(1 + \frac{\gamma x}{\sigma}\right) - \frac{1 + \gamma}{\gamma \sigma} x \left(1 + \frac{\gamma x}{\sigma}\right)^{-1} \end{pmatrix} \quad (2.3)$$



et

$$\ell^{(11)}(x, \beta) = \frac{1}{\sigma^2} - 2\frac{1+\gamma}{\sigma^3}x \left(1 + \frac{\gamma x}{\sigma}\right)^{-1} + \frac{1+\gamma}{\sigma^4}\gamma x^2 \left(1 + \frac{\gamma x}{\sigma}\right)^{-2} \quad (2.4)$$

$$\ell^{(12)}(x, \beta) = \frac{1}{\sigma^2}x \left(1 + \frac{\gamma x}{\sigma}\right)^{-1} - \frac{1+\gamma}{\sigma^3}x^2 \left(1 + \frac{\gamma x}{\sigma}\right)^{-2} \quad (2.5)$$

$$\ell^{(22)}(x, \beta) = -\frac{2}{\gamma^3} \log \left(1 + \frac{\gamma x}{\sigma}\right) + \frac{2}{\gamma^2\sigma}x \left(1 + \frac{\gamma x}{\sigma}\right)^{-1} + \frac{1+\gamma}{\gamma\sigma^2}x^2 \left(1 + \frac{\gamma x}{\sigma}\right)^{-2}. \quad (2.6)$$

Nous allons maintenant expliciter la matrice d'information de Fisher

$$\begin{aligned} I_{\beta\beta} &= \mathbb{E}_{H_0, \beta} \begin{pmatrix} (\ell^{(1)}(X, \beta))^2 & \ell^{(1)}(X, \beta)\ell^{(2)}(X, \beta) \\ \ell^{(1)}(X, \beta)\ell^{(2)}(X, \beta) & (\ell^{(2)}(X, \beta))^2 \end{pmatrix} \\ &= \begin{pmatrix} 1/\{\sigma^2(1+2\gamma)\} & 1/\{\sigma(1+\gamma)(1+2\gamma)\} \\ 1/\{\sigma(1+\gamma)(1+2\gamma)\} & 2/\{(1+\gamma)(1+2\gamma)\} \end{pmatrix} \end{aligned}$$

ainsi que son inverse

$$I_{\beta\beta}^{-1} = \begin{pmatrix} 2\sigma^2(1+\gamma) & -\sigma(1+\gamma) \\ -\sigma(1+\gamma) & (1+\gamma)^2 \end{pmatrix}.$$

La méthodologie utilisée pour l'étude du comportement de la statistique de test sous  $H_0$  est basée sur des résultats de Pfanzagl [56]. Il peut alors être utile de rappeler quelques définitions notamment concernant la convergence uniforme locale; ceci est proposé en Annexe 2.8.1. L'étude sous  $H_{1,n}$  repose sur l'utilisation de la théorie LAN dont quelques rappels sont donnés en Annexe 2.8.2.

La construction du test repose également sur la théorie LAN. L'idée est de chercher un test capable de détecter au mieux les alternatives dites locales. Pour s'assurer que nous sommes bien dans le cadre de la normalité asymptotique locale et connaître la suite centrale associée, nous commençons par calculer le logarithme du rapport de vraisemblance suivant

$$\begin{aligned} L_n &= \log \prod_{i=1}^n \frac{g_J(X_i, \frac{\delta}{\sqrt{n}}(1+o(1)), \beta)}{f(X_i, \beta)} \\ &= \sum_{i=1}^n \left[ \log \left( g_J \left( X_i, \frac{\delta}{\sqrt{n}}(1+o(1)), \beta \right) \right) - \log(f(X_i, \beta)) \right]. \end{aligned}$$

Via (2.1) et (2.2),  $L_n$  s'exprime aussi selon

$$\begin{aligned}
L_n &= \sum_{i=1}^n \sum_{s=1}^J \frac{\delta_s}{\sqrt{n}} (1 + r_s(n)) \bar{F}^s(X_i, \beta) - nK \left( \frac{\delta}{\sqrt{n}} (1 + o(1)) \right) \\
&= \sum_{s=1}^J \delta_s (1 + r_s(n)) \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \bar{F}^s(X_i, \beta) - \frac{1}{s+1} \right) \\
&\quad + \sqrt{n} \sum_{s=1}^J \frac{\delta_s}{s+1} (1 + r_s(n)) - nK \left( \frac{\delta}{\sqrt{n}} (1 + o(1)) \right) \\
&= \sum_{s=1}^J \delta_s (1 + r_s(n)) \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \bar{F}^s(X_i, \beta) - \frac{1}{s+1} \right) \\
&\quad + \sqrt{n} \sum_{s=1}^J \frac{\delta_s}{s+1} (1 + r_s(n)) - n \left[ \frac{(\delta(1 + o(1)))^t}{\sqrt{n}} \frac{\partial K(\theta)}{\partial \theta} \Big|_{\theta=0} \right. \\
&\quad \left. + \frac{1}{2} \frac{1}{\sqrt{n}} (\delta(1 + o(1)))^t \frac{\partial^2 K(\theta)}{\partial \theta \partial \theta} \Big|_{\theta=0} \frac{1}{\sqrt{n}} \delta(1 + o(1)) + o \left( \left\| \frac{\delta}{\sqrt{n}} (1 + o(1)) \right\|^2 \right) \right],
\end{aligned}$$

en utilisant un développement de Taylor à l'ordre 2 de la fonction  $K(\cdot)$  dans la dernière égalité.

Définissons le vecteur colonne  $Z_n(\beta)$  comme suit

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \bar{F}^s(X_i, \beta) - \frac{1}{s+1} \right] \Big|_{s=1, \dots, J}$$

et notons que la matrice

$$I := \frac{\partial^2 K(\theta)}{\partial \theta^2} \Big|_{\theta=0}$$

correspond à la variance de  $Z_n(\beta)$  (cf. (2.14)). Le théorème central limite implique que la quantité aléatoire  $Z_n(\beta)$  converge en distribution vers une gaussienne centrée de matrice de variance-covariance  $I$ .

Comme pour  $s = 1, \dots, J$ , les fonctions  $r_s(n)$  convergent vers 0 et

$$\frac{\partial K(\theta)}{\partial \theta} \Big|_{\theta=0} = \frac{1}{s+1} \Big|_{s=1, \dots, J},$$

nous obtenons que

$$L_n = \delta^t Z_n(\beta) - \frac{1}{2} \delta^t I \delta + o_{\mathbb{P}_{H_0, \beta}}(1). \tag{2.7}$$

De plus, la quantité  $L_n$  converge en distribution vers une gaussienne d'espérance  $-\frac{1}{2} \delta^t I \delta$  et

de variance  $\delta^t I \delta$ . D'après les rappels de l'Annexe 2.8.2, il est clair que nous nous inscrivons bien dans le cadre des modèles LAN. La suite centrale  $Z_n(\beta)$  tient une place importante dans ce modèle (cf. Strasser [67], Chapitre 13). Il s'agit de la principale quantité aléatoire contenue dans le logarithme du rapport de vraisemblance et par conséquent la statistique de test de  $H_0$  contre  $H_{1,n}$  devrait être basée sur cette quantité.

Dans la prochaine section nous nous placerons sous  $H_0$  et nous étudierons le comportement de la statistique de test.

### 2.3 Le comportement sous $H_0$ de la statistique de test proposée

Nous supposons dans cette section que l'échantillon  $X_1, \dots, X_n$  est issu d'une distribution de Pareto avec paramètre inconnu  $\beta = (\sigma, \gamma)^t \in \Theta$ . Une première idée serait d'utiliser l'estimateur du maximum de vraisemblance. Cependant cet estimateur, uniquement défini comme la solution d'un système de deux équations à deux inconnues, ne possède pas d'expression explicite. Parfois même, il ne peut pas du tout être obtenu, comme le montrent Hosking et Wallis [41] empiriquement. Nous proposons donc d'utiliser un estimateur de  $\beta$  ayant les mêmes propriétés que celui du maximum de vraisemblance mais plus facile à implémenter. Cet estimateur dont la structure est définie dans le théorème 7.5.9 de Pfanzagl [56] rappelé en Annexe 2.8.1 sera noté  $\widehat{\beta}_n$ . Il s'agit de mettre en oeuvre l'algorithme de Newton-Raphson mais de ne procéder qu'à la première itération. L'estimateur initial de la procédure noté  $\widetilde{\beta}_n$  doit être  $\sqrt{n}$ -consistant (localement uniformément). L'estimateur  $\widehat{\beta}_n$  se définit alors de la façon suivante :

$$\widehat{\beta}_n = \widetilde{\beta}_n + I_{\widetilde{\beta}_n \widetilde{\beta}_n}^{-1} \frac{1}{n} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \widetilde{\beta}_n). \quad (2.8)$$

L'estimateur  $I_{\widetilde{\beta}_n \widetilde{\beta}_n}^{-1}$  est obtenu par la méthode de « plug-in » qui consiste à remplacer le pa-

paramètre  $\beta$  par son estimateur initial  $\tilde{\beta}_n$  dans la matrice  $I_{\beta\beta}^{-1}$ . En effet nous pouvons déduire du théorème des transformations continues une version uniforme locale (cf. Annexe 2.8.1). L'estimateur  $\tilde{\beta}_n$  est consistant pour  $\beta$ , localement uniformément sur  $\Theta$ , par conséquent en appliquant la version uniforme locale du théorème des transformations continues, nous déduisons que l'estimateur  $I_{\tilde{\beta}_n\tilde{\beta}_n}^{-1}$  est consistant pour  $I_{\beta\beta}^{-1}$ , localement uniformément sur  $\Theta$ . La proposition qui suit est un résultat clé pour établir le comportement sous  $H_0$  de la statistique de test définie dans la Proposition 2.2. Elle établit la convergence en distribution, localement uniformément, de l'estimateur  $\hat{\beta}_n$ . Nous noterons la convergence en distribution par  $\rightsquigarrow$ .

**Proposition 2.1** *Sous l'hypothèse  $H_0$ , nous pouvons établir que*

$$\sqrt{n}(\hat{\beta}_n - \beta) - I_{\beta\beta}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \xrightarrow{\mathbb{P}_{H_0, \beta}} 0 \text{ localement uniformément sur } \Theta \quad (2.9)$$

$$\text{et } \sqrt{n}(\hat{\beta}_n - \beta) \rightsquigarrow \mathcal{N}(0, I_{\beta\beta}^{-1}) \text{ localement uniformément sur } \Theta. \quad (2.10)$$

Nous avons déjà expliqué à la fin de la Section 2.2 la raison pour laquelle la statistique de test doit dépendre de la quantité aléatoire  $Z_n(\beta)$ . Le paramètre  $\beta$  étant en pratique inconnu, il est tout naturel de construire la statistique de test comme une fonction de  $Z_n(\hat{\beta}_n)$ . Cette statistique de test est appelée  $\Psi_J^2$  et son expression exacte, tout comme son comportement asymptotique sous  $H_0$ , sont donnés dans la proposition ci-dessous.

**Proposition 2.2** *Soient  $Z_n(\hat{\beta}_n)$  et  $\Sigma_J(\beta)$  définis comme suit :*

$$Z_n(\hat{\beta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \left(1 - F(X_i, \hat{\beta}_n)\right)^s - \frac{1}{s+1} \right] \Big|_{s=1, \dots, J},$$

$$\Sigma_J(\beta) = \left[ \frac{uv}{(u+v+1)(u+1)(v+1)} - \frac{uv(1+\gamma)(uv+\gamma+(u+1)(v+1))}{(v+\gamma+1)(u+\gamma+1)(u+1)^2(v+1)^2} \right] \Big|_{u,v=1, \dots, J}.$$

*Sous l'hypothèse  $H_0$  que  $X_1, \dots, X_n$  sont des variables aléatoires indépendantes et identiquement distribuées issues d'une distribution de Pareto, la statistique  $Z_n(\hat{\beta}_n)$  converge en distribution vers une gaussienne de vecteur moyenne nul et de matrice de variance-covariance  $\Sigma_J(\beta)$  :  $Z_n(\hat{\beta}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_J(\beta))$ . De plus, la statistique de test  $\Psi_J^2 :=$*

$Z_n^t(\widehat{\beta}_n)\Sigma_J^{-1}(\widehat{\beta}_n)Z_n(\widehat{\beta}_n)$  converge en distribution vers une loi du  $\chi^2$  à  $J$  degrés de liberté :  $\Psi_J^2 \rightsquigarrow \chi_J^2$ . Par conséquent l'hypothèse nulle sera rejetée pour les grandes valeurs de  $\Psi_J^2$ .

**Remarque 2.1** Quand nous cherchons à tester  $\theta = 0$  versus  $\theta \neq 0$ , il existe trois tests asymptotiquement équivalents : le test de Wald, le test du score et le test du rapport de vraisemblance. On peut montrer que notre statistique de test est de la forme de celle du test du score, communément utilisé dans la théorie des tests lisses d'ajustement. Dans la mise en œuvre du test de Neyman, l'utilisation du test du score apparaît être le plus pratique, du fait qu'il ne nécessite pas le calcul de l'estimateur du paramètre  $\theta$  par maximum de vraisemblance, contrairement aux deux autres tests. En effet, le paramètre  $\theta$  étant de dimension  $J$ , il est pratique de ne pas avoir à l'estimer, d'autant plus qu'il n'a pas de réelle interprétation dans le modèle (cf. Rayner et Best [61], pour plus de détails).

Pour la construction de l'estimateur  $\widehat{\beta}_n$ , il reste à choisir une méthode d'estimation pour l'estimateur initial  $\widetilde{\beta}_n$ . À titre d'illustration, nous proposons l'utilisation de l'estimateur des moments (cf. Hosking et Wallis [41], Section 3.2). Cet estimateur des moments est basé sur le fait que

$$\mathbb{E}_{H_0, \beta} \left[ \left( 1 + \frac{\gamma X}{\sigma} \right)^r \right] = \frac{1}{1 - r\gamma} \text{ si } 1 - r\gamma > 0.$$

Nous pouvons directement déduire l'expression des paramètres en fonction des deux premiers moments  $\mu_1$  et  $\mu_2$

$$\begin{aligned} \sigma &= \frac{1}{2}\mu_1 \left( \frac{\mu_1^2}{\mu_2 - \mu_1^2} + 1 \right) \\ \gamma &= \frac{1}{2} \left( 1 - \frac{\mu_1^2}{\mu_2 - \mu_1^2} \right). \end{aligned}$$

En remplaçant  $\mu_1$  et  $\mu_2$  par leurs versions empiriques, nous obtenons l'estimateur des moments  $\widetilde{\beta}_n = (\widetilde{\sigma}_n, \widetilde{\gamma}_n)^t$  défini par

$$\widetilde{\sigma}_n = \frac{1}{2}\overline{X} \left( \frac{\overline{X}^2}{\overline{X^2} - \overline{X}^2} + 1 \right), \quad \widetilde{\gamma}_n = \frac{1}{2} \left( 1 - \frac{\overline{X}^2}{\overline{X^2} - \overline{X}^2} \right), \quad (2.11)$$

$$\text{où } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ et } \overline{X^2} = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

La proposition suivante montre que  $\tilde{\beta}_n$  est un candidat possible pour l'estimation initiale de  $\beta$ . En effet, cette proposition montre qu'il est  $\sqrt{n}$ -consistant localement uniformément pour  $\beta$ .

**Proposition 2.3** *Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes et identiquement distribuées issues d'une distribution de Pareto de paramètre  $\beta$ . L'estimateur des moments  $\tilde{\beta}_n$  défini en (2.11) est  $\sqrt{n}$ -consistant pour  $\beta = (\sigma, \gamma)^t$ , localement uniformément sur  $\mathcal{B} \subset \Theta$  où  $\mathcal{B} = (0, \infty) \times (0, 1/4)$ .*

La condition  $\gamma \in (0, 1/4)$ , certes restrictive, est due à l'utilisation de la méthode d'estimation par les moments. Une façon de l'affaiblir serait de considérer d'autres estimateurs tel l'estimateur des moments pondérés (cf. Hosking et Wallis [41], Section 3.3).

Dans les simulations de la Section 2.5, cet estimateur des moments sera choisi comme estimateur initial. Avant cela, il convient d'étudier le comportement de la statistique de test  $\Psi_J^2$  sous les alternatives locales  $H_{1,n}$ . Cette étude fait l'objet de la section suivante.

## 2.4 Le comportement sous $H_{1,n}$ de la statistique de test proposée

Commençons par expliciter certains résultats préliminaires nécessaires à l'étude du comportement de la statistique de test sous  $H_{1,n}$ .

### 2.4.1 Résultats préliminaires

Dans ce qui suit, les variables aléatoires  $X_1, \dots, X_n$  sont indépendantes et identiquement distribuées de loi de Pareto (lorsque l'on se place sous  $H_0$ ) ou de densité commune  $g_J$  (lorsque l'on se place sous  $H_{1,n}$ ).

La proposition suivante est basée sur le concept de contiguïté (cf. LeCam [49]). Les arguments de contiguïté sont des outils permettant d'obtenir la distribution limite d'une statistique sous des lois (ici les alternatives) à partir de la distribution sous d'autres lois (ici celle associée à l'hypothèse nulle).

**Proposition 2.4** *Pour toute statistique  $T_n = T(X_1, \dots, X_n; \beta)$ ,  $T_n \xrightarrow{\mathbb{P}_{H_0, \beta}} 0$  si et seulement si  $T_n \xrightarrow{\mathbb{P}_{H_{1,n}}} 0$ . Ceci est équivalent à dire que  $\mathbb{P}_{H_0, \beta}$  et  $\mathbb{P}_{H_{1,n}}$  sont mutuellement contigües.*

La Proposition 2.4 nous apprend que lorsque nous aurons une représentation asymptotique sous  $H_0$ , alors elle sera également valable sous  $H_{1,n}$  (et inversement). Les deux propositions suivantes sont des résultats intermédiaires indispensables pour établir le comportement asymptotique de la statistique de test  $\Psi_J^2$  sous  $H_{1,n}$ .

**Proposition 2.5** *Notons*

$$I := \frac{uv}{(u+v+1)(u+1)(v+1)} \Big|_{u,v=1,\dots,J} \quad (2.12)$$

et

$$I_\beta := \left( \frac{-u}{(u+1+\gamma)(u+1)\sigma} \quad \frac{-u}{(u+1+\gamma)(u+1)^2} \right) \Big|_{u=1,\dots,J}.$$

Alors nous avons sous  $H_{1,n}$

$$\begin{pmatrix} Z_n(\beta) \\ \sqrt{n}(\hat{\beta}_n - \beta) \end{pmatrix} \rightsquigarrow \mathcal{N} \left[ \begin{pmatrix} I\delta \\ I_{\beta\beta}^{-1} I_\beta^t \delta \end{pmatrix}, \begin{pmatrix} I & I_\beta I_{\beta\beta}^{-1} \\ I_{\beta\beta}^{-1} I_\beta^t & I_{\beta\beta}^{-1} \end{pmatrix} \right].$$

A présent une application de la méthode delta conduit au résultat suivant.

**Proposition 2.6** *Sous  $H_{1,n}$ , nous avons la convergence en distribution suivante*

$$\sqrt{n} \left( 1 - \frac{\hat{\gamma}_n \sigma}{\hat{\sigma}_n \gamma} \right) \rightsquigarrow \mathcal{N} \left( -\delta^t \frac{u^2(1+\gamma)(1+2\gamma)}{\gamma(u+1+\gamma)(u+1)^2} \Big|_{u=1,\dots,J}, \frac{(1+\gamma)^2(1+2\gamma)}{\gamma^2} \right).$$

Maintenant, afin de pouvoir évaluer la qualité de notre test, nous devons calculer sa puissance.

## 2.4.2 Puissance du test

Pour évaluer la puissance de la statistique de test  $\Psi_J^2$ , nous étudions le comportement asymptotique de cette dernière sous les alternatives locales  $H_{1,n}$ . C'est l'objet de la proposition suivante.

**Proposition 2.7** *Sous l'hypothèse  $H_{1,n} : \theta = \theta_n = \frac{\delta}{\sqrt{n}}(1 + o(1))$  avec  $\delta$  défini dans (2.2), la statistique  $Z_n(\widehat{\beta}_n)$  converge en distribution vers une gaussienne  $\mathcal{N}(\Sigma_J(\beta)\delta, \Sigma_J(\beta))$ , et  $\Psi_J^2 = Z_n^t(\widehat{\beta}_n)\Sigma_J^{-1}(\widehat{\beta}_n)Z_n(\widehat{\beta}_n)$  converge en distribution vers une loi du  $\chi^2$  non centrée à  $J$  degrés de liberté avec un paramètre de décentrage égal à  $\delta^t\Sigma_J(\beta)\delta$ .*

Remarquons que ce paramètre de décentrage dépend en partie de la distance entre les alternatives et l'hypothèse nulle.

Avec un risque de première espèce  $\alpha$ , la puissance du test est alors égale à  $1 - \chi_{J,\delta^t\Sigma_J(\beta)\delta}^2(x_{1-\alpha})$  où  $\chi_{J,\delta^t\Sigma_J(\beta)\delta}^2$  correspond à la fonction de répartition de la distribution du  $\chi^2$  à  $J$  degrés de liberté avec un paramètre de décentrage égal à  $\delta^t\Sigma_J(\beta)\delta$  et où  $x_{1-\alpha}$  correspond au quantile d'ordre  $1 - \alpha$  associé à la distribution du  $\chi^2$  centré à  $J$  degrés de liberté.

## 2.5 Simulations

### 2.5.1 Paramètres et méthodes

Nous allons dans cette section étudier le comportement, à distance finie, de la statistique de test  $\Psi_J^2$ . Pour cela nous avons simulé 1 000 échantillons d'observations indépendantes et identiquement distribuées de densité  $g_J$  définie en (2.1). Quatre tailles d'échantillons ont été choisies,  $n = 20, 50, 100, 1\,000$ . Nous avons fixé les valeurs de  $\delta_i, 1 \leq i \leq J$ , à  $\delta_1 = \dots = \delta_J = \delta \in \{0, 0.1, 2, 5\}$ . Le cas  $\delta = 0$  correspond à l'hypothèse nulle.

Soient  $\gamma = 0.1$  et  $\sigma = 1$ . Dans toutes nos simulations, l'estimateur  $\widehat{\beta}_n$  est construit à partir de l'estimateur des moments  $\widetilde{\beta}_n$  (défini en (2.11)). Dans le cas où  $\delta \neq 0$ , quatre



valeurs différentes de  $J$  ont été prises dans le but d'étudier l'impact de ce facteur. Le trait plein correspond à  $J = 1$ , les tirets à  $J = 2$ , les pointillés à  $J = 3$  tandis que  $J = 4$  est représenté par une alternance de tirets et de points.

Pour simuler des échantillons de densité  $g_J$ , nous avons utilisé une méthode de rejet (cf. Annexe 2.8.3) puisque la fonction de répartition associée était inconnue. Pour chaque configuration, 1 000 échantillons de variables aléatoires indépendantes et identiquement distribuées ont été générés. Pour chacun de ces échantillons, nous avons mis en œuvre le test et obtenu le seuil de signification associé ( $p$ -valeur). Comme la  $p$ -valeur est (asymptotiquement, quand la taille de l'échantillon augmente) uniformément distribuée sur  $(0, 1)$  sous l'hypothèse nulle  $\delta = 0$ , les quantile plots ont ici été construits pour représenter l'écart à l'uniformité. La Figure 1 représente pour chaque configuration les quantile plots des 1 000 réalisations indépendantes de la  $p$ -valeur asymptotique  $p = 1 - \chi_J^2(\Psi_J^2)$  correspondant à la statistique de test  $\Psi_J^2$ , où  $\chi_J^2$  est la fonction de répartition de la distribution du  $\chi^2$  à  $J$  degrés de liberté. Les 1 000  $p$ -valeurs sont ordonnées sur l'axe vertical,  $p_{1:1000} \leq \dots \leq p_{1000:1000}$  et elles sont associées aux points  $(i/1001, p_{i:1000})$ ,  $i \leq 1000$ , sur l'axe des abscisses. La ligne horizontale correspond au niveau 5%. Ceci se justifie par le fait qu'une réalisation de la  $p$ -valeur inférieure à ce niveau conduit usuellement au rejet de l'hypothèse nulle. La proportion d'échantillons pour laquelle on rejette l'hypothèse nulle au risque de première espèce 5% se lit alors à l'intersection de cette droite de niveau et du quantile plot.

## 2.5.2 Résultats

Comme nous l'avons déjà mentionné, la  $p$ -valeur est (asymptotiquement, quand la taille de l'échantillon augmente) uniformément distribuée sur  $(0, 1)$  sous l'hypothèse nulle  $\delta = 0$ . Le quantile plot devrait donc dans ce cas être proche de la ligne droite reliant les points  $(0, 0)$  et  $(1, 1)$ . Le quantile plot correspondant montre que cela est vérifié avec une légère

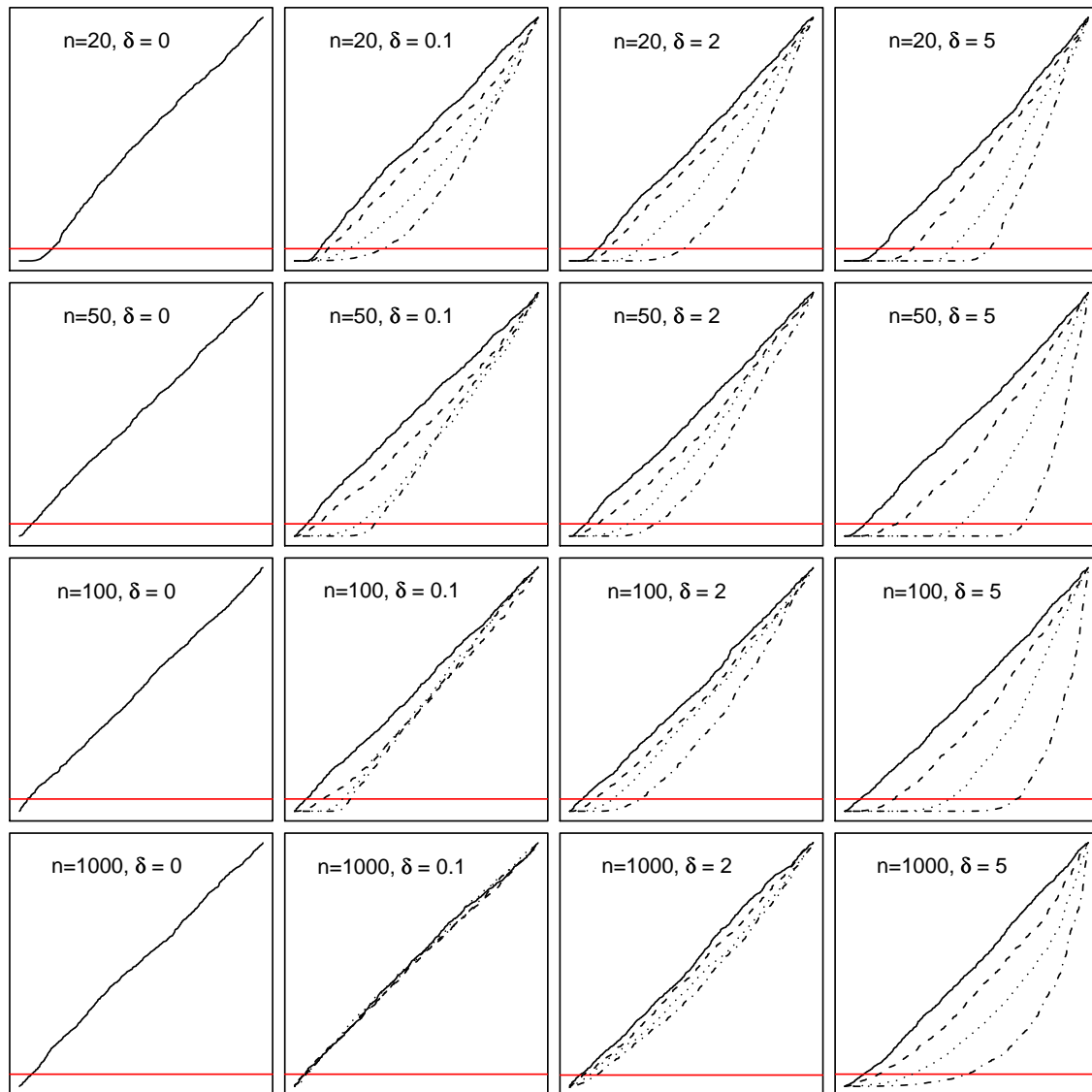


FIG. 2.1 – Quantile plots de 1 000 réalisations indépendantes de la  $p$ -valeur asymptotique correspondant à la statistique de test  $\Psi_J^2$ . Sont fixés  $\gamma = 0.1$ ,  $\sigma = 1$  et  $\delta_1 = \dots = \delta_J = \delta \in \{0, 0.1, 2, 5\}$ . Le trait plein correspond à  $J = 1$ , les tirets à  $J = 2$ , les pointillés à  $J = 3$  tandis que  $J = 4$  est représenté par une alternance de tirets et de points. La ligne horizontale correspond au niveau 5% et les axes des abscisses et des ordonnées vont de 0 à 1.

tendance à rejeter l'hypothèse nulle pour de très petits échantillons ( $n = 20$ ). Le risque de première espèce est donc bien conservé.

Par ailleurs, la puissance du test face aux alternatives locales augmente avec  $\delta$ . Cela se visualise par les quantile plots de la  $p$ -valeur qui ont tendance à rester en dessous de la ligne de niveau 5% quand la valeur de  $\delta$  augmente. Nous nous attendions à ce résultat car l'écart entre l'hypothèse alternative et l'hypothèse nulle dépend en partie de  $\delta$ . À une taille d'échantillon fixée, plus  $\delta$  est grand et plus l'hypothèse alternative est éloignée de l'hypothèse nulle. Il est donc naturel de mieux détecter les alternatives associées à une forte valeur de  $\delta$ .

Nous observons également que la puissance du test augmente avec  $J$ , avec une différence particulièrement marquée pour une grande valeur de  $\delta$  (cf. par exemple  $\delta = 5$ ) mais pas pour un  $\delta$  proche de 0 (correspondant à l'hypothèse nulle). Néanmoins le problème de la sélection d'une valeur optimale pour ce paramètre n'est pas traité dans ce chapitre. Cette question a déjà été plusieurs fois abordée dans la littérature, cf. par exemple Thomas et Pierce [68] ou Kopecky et Pierce [46] pour des valeurs spécifiques justifiées empiriquement, ou bien Inglot *et al.* [42] pour une règle de sélection de ce paramètre associée à la statistique de test.

## 2.6 Remarques générales et lien avec la théorie des valeurs extrêmes

Dans ce chapitre nous avons proposé un test lisse d'ajustement pour la distribution de Pareto. Ce test a été motivé par la théorie LAN proposée par LeCam [49]. Cette distribution apparaissant comme un cas particulier de la distribution de Pareto généralisée, il s'agit d'un modèle essentiel dans l'étude des évènements extrêmes. En effet, elle apparaît comme la distribution limite des excès au-delà d'un seuil élevé en théorie des valeurs extrêmes.

Cette approche est communément appelée approche POT (Peaks-Over-Threshold). Par conséquent une extension intéressante de notre test serait de pouvoir considérer le cas « approximativement Pareto » et non « exactement Pareto ». Un autre problème important que l'on rencontre en théorie des valeurs extrêmes est le choix du seuil. Selon Davison et Smith [24], ce type de test pourrait être utile dans l'élaboration d'une méthode de sélection de seuil dans l'approche POT. C'est également ce que suggèrent Choulakian et Stephens [15] en proposant leur test. L'idée est alors de choisir un seuil initial suffisamment petit puis de l'augmenter au fur et à mesure jusqu'au non-rejet de la distribution de Pareto.

## 2.7 Preuve des résultats

### 2.7.1 Preuve de la Proposition 2.1

Cette preuve est basée sur le Théorème 7.5.9. de Pfanzagl [56]. Elle peut être divisée en deux parties.

Première partie : Cinq conditions doivent être vérifiées. Dans le livre de Pfanzagl [56], elles sont appelées Conditions 7.5.1, 7.5.2, 7.5.3, i) et ii) du Théorème 7.5.5. La première condition consiste à vérifier que  $\mathbb{E}_{H_{0,\beta}}(\ell^{(\cdot)}(X, \beta)) = 0$  avec  $\beta \in \Theta$ . En utilisant la transformation  $X_i = F^{-1}(1 - U_i, \beta) = \frac{\sigma}{\gamma}(U_i^{-\gamma} - 1)$  avec  $U_1, \dots, U_n$  un échantillon de variables aléatoires indépendantes et identiquement distribuées issu de la loi uniforme sur  $(0, 1)$ , il est évident que la Condition 7.5.1 est remplie.

Pour satisfaire la seconde, nous devons vérifier que

$$\mathbb{E}_{H_{0,\beta}}(\ell^{(\mu\nu)}(X, \beta)) + \mathbb{E}_{H_{0,\beta}}(\ell^{(\mu)}(X, \beta)\ell^{(\nu)}(X, \beta)) = 0 \text{ pour } \mu, \nu = 1, 2 \text{ et } \beta \in \Theta.$$

Cela se fait facilement avec le même changement de variable que précédemment. La troisième condition nécessite que les fonctions  $\ell^{(\nu)}(\cdot, \beta)$  pour  $\nu = 1, 2$  et  $\beta \in \Theta$  soient linéairement  $\mathbb{P}_{H_{0,\beta}}$ -indépendantes. Pour cela, il suffit de remarquer que  $c_1\ell^{(1)}(x, \beta) +$

$c_2 \ell^{(2)}(x, \beta) = 0$  pour  $\mathbb{P}_{H_0, \beta}$ -presque tout  $x$  implique que  $c_1 = c_2 = 0$ .

Pour la première partie de la preuve de cette proposition, il reste à montrer que les deux Conditions i) et ii), rappelées ci-dessous, sont satisfaites pour tout  $\beta_0 \in \Theta$ .

**Condition i) du Théoreme 7.5.5 de Pfanzagl [56].** Nous devons vérifier qu'il existe un voisinage  $U(\sigma_0, \gamma_0)$  de  $\beta_0$  tel que

$\beta \mapsto \ell^{(\mu\nu)}(x, \beta)$  est continue sur  $U(\sigma_0, \gamma_0)$  pour  $\mu, \nu = 1, 2$ , pour tout  $x$

et  $\sup_{\beta_1 \in U(\sigma_0, \gamma_0)} |\ell^{(\mu\nu)}(\cdot, \beta_1)|$  est intégrable par rapport à la distribution de Pareto ( $\beta$ )

localement uniformément en  $\beta_0$ .

Pour établir ce dernier point, il est suffisant, d'après l'Annexe 2.8.1, de prouver qu'il existe  $\eta > 0$  tel que

$$\sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E}_{H_0, \beta} \left( \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} |\ell^{(\mu\nu)}(X, \beta_1)|^{1+\eta} \right) < \infty.$$

Au vu de (2.4), (2.5) et (2.6), il suffit d'établir ce résultat pour  $\ell^{(22)}(X, \beta_1)$ . Nous avons

$$\begin{aligned} & \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E}_{H_0, \beta} \left( \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} |\ell^{(22)}(X, \beta_1)|^{1+\eta} \right) \\ &= \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E}_{H_0, \beta} \left( \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \left| \frac{\partial^2}{\partial \gamma_1 \partial \gamma_1} \log f(X, \beta_1) \right|^{1+\eta} \right). \end{aligned}$$

Avec le changement de variable  $X = \frac{\sigma}{\gamma}(U^{-\gamma} - 1)$ , où  $U$  est une variable aléatoire uniformément distribuée sur  $(0, 1)$ , il est équivalent d'étudier

$$\begin{aligned} & \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \left| \frac{\partial^2}{\partial \gamma_1 \partial \gamma_1} \log f \left( \frac{\sigma}{\gamma}(U^{-\gamma} - 1), \beta_1 \right) \right|^{1+\eta} \right) \\ &= \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \left| -\frac{2}{\gamma_1^3} \log \left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right) \right. \right. \\ & \quad \left. \left. + \frac{2}{\gamma_1^2 \sigma_1} \frac{\frac{\sigma}{\gamma}(U^{-\gamma} - 1)}{1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma})} + \frac{1 + \gamma_1}{\gamma_1 \sigma_1^2} \frac{\left( \frac{\sigma}{\gamma} \right)^2 (1 - U^{-\gamma})^2}{\left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right)^2} \right|^{1+\eta} \right). \end{aligned}$$

Si l'on choisit  $\eta = 1$ , il suffit alors de considérer les termes  $\sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \{(1 - U^{-\gamma})^4\} / \{(1 - (\gamma_1 \sigma / (\sigma_1 \gamma))(1 - U^{-\gamma}))^4\}$ ,  $\sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \log^2 \{1 - (\gamma_1 \sigma / (\sigma_1 \gamma))(1 - U^{-\gamma})\}$

et  $\sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \{\log\{1 - (\gamma_1 \sigma / (\sigma_1 \gamma))(1 - U^{-\gamma})\} \times \{(1 - U^{-\gamma})^2\} / \{(1 - (\gamma_1 \sigma / (\sigma_1 \gamma))(1 - U^{-\gamma}))^2\}\}$ .

Tout comme  $\sigma$  et  $\gamma$ , les paramètres  $\sigma_1$  et  $\gamma_1$  sont dans le voisinage de  $\sigma_0$  et  $\gamma_0$ . Par conséquent nous pouvons supposer qu'il existe  $\varepsilon \in (0, 1)$  tel que  $(\gamma_1 \sigma) / (\sigma_1 \gamma) \in (1 - \varepsilon, 1 + \varepsilon)$ .

Concernant le premier terme, nous avons, pour un  $\varepsilon$  fixé,

$$\begin{aligned} C_1 &= \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \left\{ \frac{1 - U^{-\gamma}}{1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma})} \right\}^4 \right) \\ &\leq \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \left\{ \frac{1 - U^{-\gamma}}{1 - (1 - \varepsilon)(1 - U^{-\gamma})} \right\}^4 \right). \end{aligned}$$

Pour  $u \in (0, 1)$ , posons  $h_{1, \varepsilon}(u) := \{1 - u^{-\gamma}\} / \{1 - (1 - \varepsilon)(1 - u^{-\gamma})\}$ . Comme  $\gamma$  est positif,  $h_{1, \varepsilon}(\cdot)$  est une fonction croissante et négative.

Utilisant le fait que  $\varepsilon \in (0, 1)$ , nous déduisons que

$$\sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \left\{ \frac{1 - U^{-\gamma}}{1 - (1 - \varepsilon)(1 - U^{-\gamma})} \right\}^4 \right) \leq h_{1, \varepsilon}^4(0) = \frac{1}{(1 - \varepsilon)^4} < \infty.$$

Concernant le second terme,

$$\begin{aligned} C_2 &= \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \log^2 \left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right) \right) \\ &\leq \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \log^2 (1 - (1 + \varepsilon)(1 - U^{-\gamma})) \right) \\ &= \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \left[ -\gamma \log U + \log(1 + \varepsilon(-U^{-\gamma} + 1)) \right]^2 \right) \\ &\leq \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \left[ -\gamma \log U + \log(1 + \varepsilon) \right]^2 \right) \\ &\leq \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} (2\gamma^2 + 2\gamma \log(1 + \varepsilon)) + \log^2(1 + \varepsilon) \\ &< \infty. \end{aligned}$$

Concernant le troisième terme,

$$\begin{aligned}
C_3 &= \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \left\{ \log \left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right) \right\} \right. \\
&\quad \left. \times \left\{ \frac{(1 - U^{-\gamma})^2}{\left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right)^2} \right\} \right) \\
&\leq \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \left\{ \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \log \left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right) \right\} \right. \\
&\quad \left. \times \left\{ \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \frac{(1 - U^{-\gamma})^2}{\left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right)^2} \right\} \right).
\end{aligned}$$

Alors, une application directe de l'inégalité de Cauchy-Schwarz implique que

$$\begin{aligned}
C_3 &\leq \sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \left( \sqrt{\mathbb{E} \left( \left\{ \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \log \left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right) \right\}^2 \right)} \right. \\
&\quad \left. \times \sqrt{\mathbb{E} \left( \left\{ \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \frac{(1 - U^{-\gamma})^2}{\left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right)^2} \right\}^2 \right)} \right) \\
&\leq \sqrt{\sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \left\{ \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \log \left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right) \right\}^2 \right)} \\
&\quad \times \sqrt{\sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} \left( \left\{ \sup_{(\sigma_1, \gamma_1) \in U(\sigma_0, \gamma_0)} \frac{(1 - U^{-\gamma})^2}{\left( 1 - \frac{\gamma_1 \sigma}{\sigma_1 \gamma} (1 - U^{-\gamma}) \right)^2} \right\}^2 \right)}.
\end{aligned}$$

Par conséquent comme  $C_1$  et  $C_2$  sont finis,  $C_3$  l'est aussi.

**Condition ii) du Théorème 7.5.5 de Pfanzagl [56].** Pour  $\nu = 1, 2$ , la famille de fonctions  $x \mapsto [\ell^{(\nu)}(x, \beta)]^2$  doit être intégrable par rapport à la distribution de Pareto de paramètre  $\beta$ , localement uniformément en  $\beta_0$ , c'est-à-dire que nous devons vérifier qu'il existe  $\eta > 0$  tel que,

$$\sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E}_{H_0, \beta} \left( |\ell^{(\nu)}(X, \beta)|^{2(1+\eta)} \right) < \infty.$$

Soit  $\eta = 1/2$ . Nous avons

$$\begin{aligned} [\ell^{(1)}(X, \beta)]^3 &= -\frac{1}{\sigma^3} + \frac{(1+\gamma)^3}{\sigma^6} X^3 \left(1 + \frac{\gamma X}{\sigma}\right)^{-3} - \frac{3(1+\gamma)^2}{\sigma^5} X^2 \left(1 + \frac{\gamma X}{\sigma}\right)^{-2} \\ &\quad + \frac{3(1+\gamma)}{\sigma^4} X \left(1 + \frac{\gamma X}{\sigma}\right)^{-1}. \end{aligned}$$

Avec  $X = (\sigma/\gamma)(U^{-\gamma} - 1)$ ,  $U \sim U(0, 1)$ , on obtient que  $[\ell^{(1)}(X, \beta)]^3$  est aussi égal à

$$-\frac{1}{\sigma^3} - \frac{(1+\gamma)^3}{\gamma^3 \sigma^3} (U^\gamma - 1)^3 - \frac{3(1+\gamma)^2}{\gamma^2 \sigma^3} (U^\gamma - 1)^2 - \frac{3(1+\gamma)}{\gamma \sigma^3} (U^\gamma - 1).$$

De la même manière on a

$$\begin{aligned} [\ell^{(2)}(X, \beta)]^3 &= \frac{1}{\gamma^6} \log^3 \left(1 + \frac{\gamma X}{\sigma}\right) - \frac{(1+\gamma)^3}{\sigma^3 \gamma^3} X^3 \left(1 + \frac{\gamma X}{\sigma}\right)^{-3} \\ &\quad + \frac{3(1+\gamma)^2}{\sigma^2 \gamma^4} X^2 \left(1 + \frac{\gamma X}{\sigma}\right)^{-2} \log \left(1 + \frac{\gamma X}{\sigma}\right) \\ &\quad - \frac{3(1+\gamma)}{\sigma \gamma^5} X \left(1 + \frac{\gamma X}{\sigma}\right)^{-1} \log^2 \left(1 + \frac{\gamma X}{\sigma}\right) \end{aligned}$$

qui est avec le même changement de variable égal à

$$-\frac{1}{\gamma^3} \log^3(U) + \frac{(1+\gamma)^3}{\gamma^6} (U^\gamma - 1)^3 - \frac{3(1+\gamma)^2}{\gamma^5} (U^\gamma - 1)^2 \log(U) + \frac{3(1+\gamma)}{\gamma^4} (U^\gamma - 1) \log^2(U).$$

En utilisant les inégalités de Hölder nous obtenons que

$$\begin{aligned} \mathbb{E} \left( (U^\gamma - 1)^2 \log(U) \right) &\leq \left( \mathbb{E} (U^\gamma - 1)^{2 \cdot \frac{3}{2}} \right)^{2/3} \times \left( \mathbb{E} (\log^3(U)) \right)^{1/3} \\ \mathbb{E} \left( (U^\gamma - 1) \log^2(U) \right) &\leq \left( \mathbb{E} (U^\gamma - 1)^3 \right)^{1/3} \times \left( \mathbb{E} (\log^{2 \cdot \frac{3}{2}}(U)) \right)^{2/3}. \end{aligned}$$

Comme  $\sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} (U^\gamma - 1)^3$  et  $\sup_{(\sigma, \gamma) \in V(\sigma_0, \gamma_0)} \mathbb{E} (\log^3(U))$  sont finis, la Condition ii) est vérifiée. Ceci achève la première partie de la preuve de la Proposition 2.1.

Deuxième partie : Nous allons utiliser l'estimateur particulier  $\widehat{\beta}_n$  de  $\beta$  donné dans l'équation (2.8). Une application du Théorème 7.5.9 de Pfanzagl [56] entraîne que

$$\sqrt{n}(\widehat{\beta}_n - \beta) - I_{\beta\beta}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \xrightarrow{\mathbb{P}_{H_0, \beta}} 0 \text{ localement uniformément en } \beta_0.$$

De plus, par le théorème central limite (voir les rappels de l'Annexe 2.8.1),



$(1/\sqrt{n}) \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \rightsquigarrow \mathcal{N}(0, I_{\beta\beta})$  localement uniformément en  $\beta_0$ . Nous déduisons alors que

$$\sqrt{n}(\widehat{\beta}_n - \beta) \rightsquigarrow \mathcal{N}(0, I_{\beta\beta}^{-1}) \text{ localement uniformément en } \beta_0.$$

Comme ces deux convergences sont valables pour tout  $\beta_0$  dans  $\Theta$ , la preuve de la Proposition 2.1 est achevée.

### 2.7.2 Preuve de la Proposition 2.2

Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes et identiquement distribuées de fonction de répartition  $F(\cdot, \beta)$ . Sous  $H_0$  nous avons  $X_i = F^{-1}(1 - U_i, \beta) = (\sigma/\gamma)(U_i^{-\gamma} - 1)$ ,  $i = 1, \dots, n$ , où  $U_1, \dots, U_n$  sont des variables aléatoires indépendantes et identiquement distribuées selon une distribution uniforme sur  $(0, 1)$ . Alors

$$\begin{aligned} Z_n(\widehat{\beta}_n) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \left(1 - F(X_i, \widehat{\beta}_n)\right)^s - \frac{1}{s+1} \right] \Big|_{s=1, \dots, J} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \left(1 - F(X_i, \beta)\right)^s - \frac{1}{s+1} + \left(1 - F(X_i, \widehat{\beta}_n)\right)^s - \left(1 - F(X_i, \beta)\right)^s \right] \Big|_{s=1, \dots, J} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ U_i^s - \frac{1}{s+1} + \left( U_i^{-\gamma} + 1 - U_i^{-\gamma} - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} (1 - U_i^{-\gamma}) \right)^{-s/\widehat{\gamma}_n} - (U_i^{-\gamma})^{-s/\gamma} \right] \Big|_{s=1, \dots, J} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ U_i^s - \frac{1}{s+1} \right] \Big|_{s=1, \dots, J} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \left( U_i^{-\gamma} + (1 - U_i^{-\gamma}) \left(1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma}\right) \right)^{-s/\gamma} - (U_i^{-\gamma})^{-s/\gamma} \right] \Big|_{s=1, \dots, J} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \left( U_i^{-\gamma} + (1 - U_i^{-\gamma}) \left(1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma}\right) \right)^{-s/\widehat{\gamma}_n} \right. \\ &\quad \quad \left. - \left( U_i^{-\gamma} + (1 - U_i^{-\gamma}) \left(1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma}\right) \right)^{-s/\gamma} \right] \Big|_{s=1, \dots, J} \\ &=: A + B + C. \end{aligned}$$

Nous allons étudier ces trois termes séparément. Le terme  $B$  peut s'écrire :

$$\begin{aligned} B &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \left( U_i^{-\gamma} + (1 - U_i^{-\gamma}) \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \right)^{-s/\gamma} - \left( U_i^{-\gamma} \right)^{-s/\gamma} \right] \Bigg|_{s=1, \dots, J} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i^s \left[ \left( 1 + (U_i^\gamma - 1) \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \right)^{-s/\gamma} - 1 \right] \Bigg|_{s=1, \dots, J}. \end{aligned}$$

En utilisant un développement de Taylor sur une fonction de la forme  $f(x) = (1+x)^\alpha$  et en appliquant ensuite la loi des grands nombres, nous obtenons que

$$\begin{aligned} B &= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i^s \left[ \frac{s}{\gamma} (1 - U_i^\gamma) \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \right] \Bigg|_{s=1, \dots, J} + o_{\mathbb{P}_{H_0, \beta}}(1) \\ &= \sqrt{n} \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \frac{s}{(s + \gamma + 1)(s + 1)} \Bigg|_{s=1, \dots, J} + o_{\mathbb{P}_{H_0, \beta}}(1). \end{aligned}$$

Concernant le terme  $C$ , un développement de Taylor de  $f(x) = (1+x)^\alpha$  nous conduit à

$$\begin{aligned} C &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ \left( U_i^{-\gamma} + (1 - U_i^{-\gamma}) \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \right)^{-s/\widehat{\gamma}_n} - \left( U_i^{-\gamma} + (1 - U_i^{-\gamma}) \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \right)^{-s/\gamma} \right] \Bigg|_{s=1, \dots, J} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[ U_i^{s\gamma/\widehat{\gamma}_n} \left( 1 + (U_i^\gamma - 1) \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \right)^{-s/\widehat{\gamma}_n} - U_i^s \left( 1 + (U_i^\gamma - 1) \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \right)^{-s/\gamma} \right] \Bigg|_{s=1, \dots, J} \\ &= \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i^{s\gamma/\widehat{\gamma}_n} \left( 1 + \frac{s}{\widehat{\gamma}_n} (1 - U_i^\gamma) \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \right) \right. \\ &\quad \left. - \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i^s \left( 1 + \frac{s}{\gamma} (1 - U_i^\gamma) \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \right) \right] \Bigg|_{s=1, \dots, J} + o_{\mathbb{P}_{H_0, \beta}}(1) \\ &= \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( U_i^{s\gamma/\widehat{\gamma}_n} - U_i^s \right) \right. \\ &\quad \left. + \sqrt{n} \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \frac{1}{n} \sum_{i=1}^n U_i^\gamma \left[ -\frac{s}{\widehat{\gamma}_n} \exp \left( \frac{s}{\widehat{\gamma}_n} \gamma \log U_i \right) + \frac{s}{\gamma} \exp \left( \frac{s}{\gamma} \gamma \log U_i \right) \right] \right. \\ &\quad \left. - \sqrt{n} \left( 1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma} \right) \frac{1}{n} \sum_{i=1}^n \left[ -\frac{s}{\widehat{\gamma}_n} \exp \left( \frac{s}{\widehat{\gamma}_n} \gamma \log U_i \right) + \frac{s}{\gamma} \exp \left( \frac{s}{\gamma} \gamma \log U_i \right) \right] \right] \Bigg|_{s=1, \dots, J} \\ &\quad + o_{\mathbb{P}_{H_0, \beta}}(1). \end{aligned}$$

En utilisant un développement de Taylor des fonctions  $f(x) = x \exp(-x\gamma \log U_i)$  et  $f(x) =$

$\exp(x)$ , nous obtenons que

$$\begin{aligned}
C &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [U_i^{s\gamma/\hat{\gamma}_n} - U_i^s] \Big|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i^s \left( U_i^{s(\gamma/\hat{\gamma}_n - 1)} - 1 \right) \Big|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n U_i^s \left[ \exp \left( s \left( \frac{\gamma}{\hat{\gamma}_n} - 1 \right) \log(U_i) \right) - 1 \right] \Big|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\
&= \sqrt{n}s \left( \frac{\gamma}{\hat{\gamma}_n} - 1 \right) \frac{1}{n} \sum_{i=1}^n U_i^s \log(U_i) \Big|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\
&= \frac{s}{(s+1)^2} \frac{1}{\hat{\gamma}_n} \sqrt{n}(\hat{\gamma}_n - \gamma) \Big|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1)
\end{aligned}$$

par la loi des grands nombres.

D'après les développements précédents, nous pouvons écrire que

$$\begin{aligned}
Z_n(\hat{\beta}_n) &= \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( U_i^s - \frac{1}{s+1} \right) + \sqrt{n} \left( 1 - \frac{\hat{\gamma}_n \sigma}{\hat{\sigma}_n \gamma} \right) \frac{s}{(s+\gamma+1)(s+1)} \right. \\
&\quad \left. + \frac{s}{(s+1)^2} \frac{1}{\hat{\gamma}_n} \sqrt{n}(\hat{\gamma}_n - \gamma) \right] \Big|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1) \\
&=: [A'(s) + B'(s) + C'(s)] \Big|_{s=1,\dots,J} + o_{\mathbb{P}_{H_0,\beta}}(1). \tag{2.13}
\end{aligned}$$

Nous pouvons maintenant calculer la variance asymptotique de  $Z_n(\hat{\beta}_n)$ . Dans ce qui suit, l'espérance, la variance et la covariance asymptotiques sont respectivement notées  $\mathbb{E}_a$ ,  $\text{Var}_a$  et  $\text{Cov}_a$ . Nous avons

$$\begin{aligned}
\text{Cov}_a(A'(u), A'(v)) &= \text{Cov}_a \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( U_i^u - \frac{1}{u+1} \right), \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( U_i^v - \frac{1}{v+1} \right) \right) \\
&= \text{Cov}_a(U_i^u, U_i^v) = \frac{1}{u+v+1} - \frac{1}{(u+1)(v+1)} \\
&= \frac{uv}{(u+v+1)(u+1)(v+1)}.
\end{aligned}$$

Nous pouvons alors remarquer que

$$\text{Cov}_a(A'(u), A'(v)) \Big|_{u,v=1,\dots,J} = \frac{\partial^2 K(\theta)}{\partial \theta \partial \theta} \Big|_{\theta=0} = I. \tag{2.14}$$

De façon similaire

$$\begin{aligned}
& \mathbb{Cov}_a(B'(u), B'(v)) \\
&= \mathbb{Cov}_a\left(\sqrt{n}\left(1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma}\right) \frac{u}{(u + \gamma + 1)(u + 1)}, \sqrt{n}\left(1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma}\right) \frac{v}{(v + \gamma + 1)(v + 1)}\right) \\
&= \mathbb{Var}_a\left(\sqrt{n}\left(1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma}\right)\right) \frac{uv}{(u + \gamma + 1)(v + \gamma + 1)(u + 1)(v + 1)} \\
&= \mathbb{Var}_a\left(\sqrt{n}\left(\frac{\gamma(\widehat{\sigma}_n - \sigma) - (\widehat{\gamma}_n - \gamma)\sigma}{\gamma \widehat{\sigma}_n}\right)\right) \frac{uv}{(u + \gamma + 1)(v + \gamma + 1)(u + 1)(v + 1)}.
\end{aligned}$$

De (2.9) nous déduisons que :

$$\begin{aligned}
\sqrt{n}(\widehat{\gamma}_n - \gamma) &= \frac{1}{\sqrt{n}}\left(\frac{1}{\gamma^2}(1 + \gamma)^2(1 + 2\gamma) \sum_{i=1}^n (U_i^\gamma - 1) - \frac{1}{\gamma}(1 + \gamma)^2 \sum_{i=1}^n \log(U_i) + n(1 + \gamma)\right) \\
&+ o_{\mathbb{P}_{H_0, \beta}}(1)
\end{aligned} \tag{2.15}$$

et

$$\begin{aligned}
\sqrt{n}(\widehat{\sigma}_n - \sigma) &= \frac{1}{\sqrt{n}}\left(-\frac{\sigma}{\gamma^2}(1 + \gamma)^2(1 + 2\gamma) \sum_{i=1}^n (U_i^\gamma - 1) + \frac{\sigma}{\gamma}(1 + \gamma) \sum_{i=1}^n \log(U_i) - 2n\sigma(1 + \gamma)\right) \\
&+ o_{\mathbb{P}_{H_0, \beta}}(1).
\end{aligned} \tag{2.16}$$

Par conséquent,

$$\begin{aligned}
\sqrt{n}\left(\frac{\gamma(\widehat{\sigma}_n - \sigma) - (\widehat{\gamma}_n - \gamma)\sigma}{\gamma \widehat{\sigma}_n}\right) &= \frac{1}{\sqrt{n}} \frac{1}{\gamma \widehat{\sigma}_n} \left[ (1 + \gamma)^3(1 + 2\gamma) \frac{\sigma}{\gamma^2} \sum_{i=1}^n (1 - U_i^\gamma) \right. \\
&\left. + \frac{\sigma}{\gamma}(1 + \gamma)(1 + 2\gamma) \sum_{i=1}^n \log(U_i) - n\sigma(1 + \gamma)(1 + 2\gamma) \right] + o_{\mathbb{P}_{H_0, \beta}}(1).
\end{aligned} \tag{2.17}$$

Des calculs directs conduisent à l'expression suivante :

$$\mathbb{Cov}_a(B'(u), B'(v)) = \frac{(1 + \gamma)^2(1 + 2\gamma)}{\gamma^2} \frac{uv}{(u + \gamma + 1)(v + \gamma + 1)(u + 1)(v + 1)}.$$

Comme la variance de  $\sqrt{n}(\widehat{\gamma}_n - \gamma)$  est  $(1 + \gamma)^2$  selon la Proposition 2.1, nous avons

$$\mathbb{Cov}_a(C'(u), C'(v)) = \frac{uv(1 + \gamma)^2}{\gamma^2(u + 1)^2(v + 1)^2}.$$

Il reste à calculer les covariances entre les termes  $A'$ ,  $B'$  et  $C'$ .

Les relations (2.15) et (2.17) impliquent :

$$\mathbb{Cov}_a \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( U_i^s - \frac{1}{s+1} \right), \sqrt{n}(\hat{\gamma}_n - \gamma) \right) = \frac{s(1+\gamma)(s-\gamma)}{(s+1)^2(s+\gamma+1)},$$

$$\mathbb{Cov}_a \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( U_i^s - \frac{1}{s+1} \right), \sqrt{n} \left( 1 - \frac{\hat{\gamma}_n \sigma}{\hat{\sigma}_n \gamma} \right) \right) = -\frac{s^2(1+\gamma)(1+2\gamma)}{\gamma(s+1)^2(s+\gamma+1)},$$

$$\mathbb{Cov}_a \left( \sqrt{n} \left( 1 - \frac{\hat{\gamma}_n \sigma}{\hat{\sigma}_n \gamma} \right), \sqrt{n}(\hat{\gamma}_n - \gamma) \right) = -\frac{1}{\gamma}(1+\gamma)(1+2\gamma).$$

Nous en déduisons que :

$$\begin{aligned} & \mathbb{Cov}_a(A'(u), B'(v)) + \mathbb{Cov}_a(A'(v), B'(u)) \\ &= -\frac{u^2 v(1+\gamma)(1+2\gamma)}{\gamma(u+1)^2(u+\gamma+1)(v+\gamma+1)(v+1)} \\ & \quad - \frac{v^2 u(1+\gamma)(1+2\gamma)}{\gamma(v+1)^2(v+\gamma+1)(u+\gamma+1)(u+1)}, \end{aligned}$$

$$\begin{aligned} & \mathbb{Cov}_a(A'(u), C'(v)) + \mathbb{Cov}_a(A'(v), C'(u)) \\ &= \frac{uv(1+\gamma)(2uv+u+v-2\gamma^2-2\gamma)}{\gamma(u+1)^2(v+1)^2(u+\gamma+1)(v+\gamma+1)}, \end{aligned}$$

$$\begin{aligned} & \mathbb{Cov}_a(B'(u), C'(v)) + \mathbb{Cov}_a(B'(v), C'(u)) \\ &= -\frac{uv(1+\gamma)(1+2\gamma)[(u+1)(v+\gamma+1) + (v+1)(u+\gamma+1)]}{\gamma^2(v+\gamma+1)(u+\gamma+1)(u+1)^2(v+1)^2}, \end{aligned}$$

ce qui permet de calculer la variance asymptotique de  $Z_n(\hat{\beta}_n)$  selon

$$\mathbb{Var}_a(Z_n(\hat{\beta}_n)) = \left[ \frac{uv}{(u+v+1)(u+1)(v+1)} - \frac{uv(1+\gamma)(uv+\gamma+(u+1)(v+1))}{(v+\gamma+1)(u+\gamma+1)(u+1)^2(v+1)^2} \right] \Bigg|_{u,v=1,\dots,J}.$$

En combinant le théorème central limite et les représentations (2.13), (2.15) et (2.17),

nous obtenons :

$$Z_n(\widehat{\beta}_n) \rightsquigarrow \mathcal{N}(0, \Sigma_J(\beta)).$$

Nous pouvons alors conclure par (2.10) et par continuité la preuve de la Proposition 2.2.

### 2.7.3 Preuve de la Proposition 2.3

Soit  $(\sigma, \gamma) \in \mathcal{B} \subset \Theta$  avec  $\mathcal{B} = (0, \infty) \times (0, 1/4)$ . Posons  $\mu_r = \mathbb{E}_{H_0, \beta}(X^r) < \infty$ ,  $r = 1, 2, 3, 4$ .

Nous avons  $\mu_1 = \frac{\sigma}{1-\gamma}$ ,  $\mu_2 = \frac{2\sigma}{1-2\gamma}\mu_1$ ,  $\mu_3 = \frac{3\sigma}{1-3\gamma}\mu_2$  et  $\mu_4 = \frac{4\sigma}{1-4\gamma}\mu_3$ .

Il est possible d'exprimer  $\beta$  comme une fonction de  $\mu_1$  et  $\mu_2$ .

Dans ce but, nous écrivons  $\beta[\nu] = h_\nu(\mu_1, \mu_2)$  avec  $h_1 : (0, \infty) \times (0, \infty) \mapsto (0, \infty)$  et  $h_2 : (0, \infty) \times (0, \infty) \mapsto (0, \infty)$ , et obtenons

$$\begin{aligned} \beta[1] = \sigma &= h_1(\mu_1, \mu_2) = \frac{1}{2}\mu_1 \left( \frac{\mu_1^2}{\mu_2 - \mu_1^2} + 1 \right) \\ \beta[2] = \gamma &= h_2(\mu_1, \mu_2) = \frac{1}{2} \left( 1 - \frac{\mu_1^2}{\mu_2 - \mu_1^2} \right) \end{aligned}$$

avec  $\mu_2 - \mu_1^2 > 0$ . Pour tout couple  $(y_1, y_2) \in (0, \infty) \times (0, \infty)$ ,  $y \mapsto h_\nu(y, y_2)$  et  $y \mapsto h_\nu(y_1, y)$  sont deux fonctions continues.

Comme par la loi forte des grands nombres,  $\frac{1}{n} \sum_{i=1}^n X_i^r \xrightarrow{p.s.} \mathbb{E}_{H_0, \beta}(X^r) = \mu_r$ ,  $r = 1, 2$ , l'estimateur des moments  $\widetilde{\beta}_n[\nu] = h_\nu\left(\frac{1}{n} \sum_{i=1}^n X_i, \frac{1}{n} \sum_{i=1}^n X_i^2\right)$ , pour  $\nu = 1, 2$ , définit une suite d'estimateurs fortement consistants.

Les dérivées partielles de  $h_1$  et  $h_2$  existent, sont continues et  $H(\sigma, \gamma)$ , définie comme

$$H(\sigma, \gamma) = \frac{\partial h_\mu}{\partial y_\nu}(y_1, y_2) \Big|_{(y_1, y_2) = \left(\frac{\sigma}{1-\gamma}, \frac{2\sigma^2}{(1-\gamma)(1-2\gamma)}\right)}, \quad \mu, \nu = 1, 2,$$

est de rang 2.

La version uniforme locale du théorème central limite (cf. Annexe 2.8.1) implique

$$\sqrt{n} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n X_i - \mu_1 \\ \frac{1}{n} \sum_{i=1}^n X_i^2 - \mu_2 \end{pmatrix} \rightsquigarrow \mathcal{N}(0, \Sigma(\beta)) \text{ localement uniformément sur } \mathcal{B}$$

avec

$$\begin{aligned}\Sigma(\beta) &= \mathbb{V}\text{ar}_{H_0, \beta} \begin{pmatrix} X \\ X^2 \end{pmatrix} = \begin{pmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 \end{pmatrix} \\ &= \begin{pmatrix} \frac{\sigma^2}{(1-\gamma)^2(1-2\gamma)} & \frac{4\sigma^3}{(1-\gamma)^2(1-2\gamma)(1-3\gamma)} \\ \frac{4\sigma^3}{(1-\gamma)^2(1-2\gamma)(1-3\gamma)} & \frac{20\sigma^4 - 44\gamma\sigma^4}{(1-\gamma)^2(1-2\gamma)^2(1-3\gamma)(1-4\gamma)} \end{pmatrix}.\end{aligned}$$

Nous pouvons alors appliquer la version uniforme locale de la méthode delta (cf. Annexe 2.8.1) et obtenons

$$\sqrt{n}(\tilde{\beta}_n - \beta) \rightsquigarrow \mathcal{N}(0, H(\beta)\Sigma(\beta)H(\beta)^t) \text{ localement uniformément sur } \mathcal{B}$$

avec

$$H(\beta) = \begin{pmatrix} (3 - 4\gamma)(1 - \gamma) & -\frac{1}{2\sigma}(1 - \gamma)(1 - 2\gamma)^2 \\ \frac{2}{\sigma}(1 - \gamma)^2(1 - 2\gamma) & -\frac{1}{2\sigma^2}(1 - \gamma)^2(1 - 2\gamma)^2 \end{pmatrix}.$$

Nous en déduisons directement la  $\sqrt{n}$ -consistance localement uniformément de  $\tilde{\beta}_n$  pour  $\beta$ .

Dans la suite, nous noterons la convergence en distribution sous  $H_0$  (respectivement sous  $H_{1,n}$ ) par  $\overset{H_0}{\rightsquigarrow}$  (respectivement par  $\overset{H_{1,n}}{\rightsquigarrow}$ ).

#### 2.7.4 Preuve de la Proposition 2.4

Ce résultat est une conséquence du premier lemme de LeCam rappelé en Annexe 2.8.2. En utilisant les mêmes notations posons  $dP_n = \prod_{i=1}^n f(X_i, \beta)$  (correspondant à l'hypothèse  $H_0$ ) et  $dQ_n = \prod_{i=1}^n g_J(X_i, \theta_n, \beta)$  (correspondant aux alternatives  $H_{1,n}$ ). Le logarithme du rapport de vraisemblance  $dQ_n/dP_n$  est alors égal à  $L_n$  dont une expression est donnée dans (2.7). Nous savons que la quantité  $L_n$  converge en distribution sous  $H_0$  vers une distribution gaussienne  $\mathcal{N}(-\frac{1}{2}\delta^t I\delta, \delta^t I\delta)$ .

Nous pouvons alors vérifier l'assertion *iii*). En effet, si  $\log V$  suit une distribution gaussienne  $\mathcal{N}(-\frac{1}{2}\delta^t I \delta, \delta^t I \delta)$  alors  $\mathbb{E}(V) = 1$ . Par *i*),  $Q_n$  est donc contigüe par rapport à  $P_n$ , ce qui est noté  $Q_n \triangleleft P_n$ .

En interchangeant dans *ii*) les rôles de  $P_n$  et de  $Q_n$ ,  $\log U$  suit une distribution gaussienne  $\mathcal{N}(-\frac{1}{2}\delta^t I \delta, \delta^t I \delta)$  impliquant que  $\mathbb{P}(U > 0) = 1$ . Par conséquent, par *i*),  $P_n \triangleleft Q_n$ .

Les suites  $P_n$  et  $Q_n$  sont alors mutuellement contigües :  $P_n \triangleleft \triangleright Q_n$ , ce qui est équivalent à dire que  $T_n \xrightarrow{P_n} 0$  si et seulement si  $T_n \xrightarrow{Q_n} 0$ .

### 2.7.5 Preuve de la Proposition 2.5

Par le théorème central limite multivarié et par la Proposition 2.1, nous obtenons que :

$$\begin{pmatrix} Z_n(\beta) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \end{pmatrix} \xrightarrow{H_0} \mathcal{N} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I & I_\beta \\ I_\beta^t & I_{\beta\beta} \end{pmatrix} \right].$$

Sous  $H_0$  le logarithme du rapport de vraisemblance  $L_n = \log \prod_{i=1}^n (g_J(X_i, \theta_n, \beta)) / (f(X_i, \beta))$  converge en distribution vers une gaussienne  $\mathcal{N}(-\frac{1}{2}\delta^t I \delta, \delta^t I \delta)$ . De plus,  $L_n$  est une fonction de  $Z_n(\beta)$  (cf. (2.7)). Par conséquent, le vecteur

$$\begin{pmatrix} Z_n(\beta) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \\ \log \prod_{i=1}^n \frac{g_J(X_i, \theta_n, \beta)}{f(X_i, \beta)} \end{pmatrix}$$



est asymptotiquement gaussien. Calculons les termes de covariance asymptotique associés :

$$\begin{aligned} \mathbb{Cov}_a \left( \log \prod_{i=1}^n \frac{g_J(X_i, \theta_n, \beta)}{f(X_i, \beta)}, Z_n(\beta) \right) &= \mathbb{Cov} \left( \delta^t Z_n(\beta) - \frac{1}{2} \delta^t I \delta, Z_n(\beta) \right) \\ &= \delta^t \mathbb{V}ar (Z_n(\beta)) = \delta^t I \end{aligned}$$

$$\begin{aligned} \text{et } \mathbb{Cov}_a \left( \log \prod_{i=1}^n \frac{g_J(X_i, \theta_n, \beta)}{f(X_i, \beta)}, \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \right) \\ &= \mathbb{Cov} \left( \delta^t Z_n(\beta) - \frac{1}{2} \delta^t I \delta, \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \right) \\ &= \delta^t \mathbb{Cov} \left( Z_n(\beta), \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \right) \\ &= \delta^t I_\beta. \end{aligned}$$

Par conséquent, nous avons

$$\begin{aligned} \begin{pmatrix} Z_n(\beta) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \\ \log \prod_{i=1}^n \frac{g_J(X_i, \theta_n, \beta)}{f(X_i, \beta)} \end{pmatrix} \stackrel{H_0}{\rightsquigarrow} \mathcal{N}_{J+3} \left[ \begin{pmatrix} 0 \\ 0 \\ -\frac{1}{2} \delta^t I \delta \end{pmatrix}, \begin{pmatrix} I & I_\beta & I \delta \\ I_\beta^t & I_{\beta\beta} & I_\beta^t \delta \\ \delta^t I & \delta^t I_\beta & \delta^t I \delta \end{pmatrix} \right] \\ \text{et } \begin{pmatrix} Z_n(\beta) \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) \end{pmatrix} \stackrel{H_{1,n}}{\rightsquigarrow} \mathcal{N}_{J+2} \left[ \begin{pmatrix} I \delta \\ I_\beta^t \delta \end{pmatrix}, \begin{pmatrix} I & I_\beta \\ I_\beta^t & I_{\beta\beta} \end{pmatrix} \right], \end{aligned}$$

en utilisant le troisième lemme de LeCam (cf. Annexe 2.8.2).

À présent, nous rappelons que la Proposition 2.1 nous donne la représentation asymptotique suivante :

$$\sqrt{n}(\hat{\beta}_n - \beta) = I_{\beta\beta}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) + o_{\mathbb{P}_{H_0, \beta}}(1).$$

Nous terminons alors la preuve de la Proposition 2.5 en remarquant que cette représentation asymptotique reste valable sous  $H_{1,n}$  (cf. Proposition 2.4) :

$$\sqrt{n}(\hat{\beta}_n - \beta) = I_{\beta\beta}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \beta) + o_{\mathbb{P}_{H_{1,n}}}(1).$$

### 2.7.6 Preuve de la Proposition 2.6

Cette preuve est basée sur la méthode delta avec la fonction  $h : \mathbb{R}^2 \mapsto \mathbb{R}$  définie par

$$h \begin{pmatrix} x \\ y \end{pmatrix} = \frac{y\sigma}{x\gamma},$$

dont le gradient  $\dot{h}$  vaut

$$\dot{h} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y\sigma/(x^2\gamma) \\ \sigma/(x\gamma) \end{pmatrix}.$$

La méthode delta repose sur le développement suivant.

$$\sqrt{n}(h(\widehat{\beta}_n) - h(\beta)) = \dot{h}(\beta)^t \sqrt{n}(\widehat{\beta}_n - \beta) + o_{\mathbb{P}_{H_{1,n}}}(1). \quad (2.18)$$

Comme  $\sqrt{n}(\widehat{\beta}_n - \beta)$  est borné en probabilité d'après la Proposition 2.5, cette équation s'obtient facilement par un développement de Taylor.

Par conséquent, puisque  $\sqrt{n}(\widehat{\beta}_n - \beta) \stackrel{H_{1,n}}{\rightsquigarrow} \mathcal{N}(I_{\beta\beta}^{-1}I_{\beta}^t\delta, I_{\beta\beta}^{-1})$  (cf. Proposition 2.5), nous avons

$$\sqrt{n}(h(\widehat{\beta}_n) - h(\beta)) \stackrel{H_{1,n}}{\rightsquigarrow} \mathcal{N}\left(\dot{h}(\beta)^t I_{\beta\beta}^{-1} I_{\beta}^t \delta, \dot{h}(\beta)^t I_{\beta\beta}^{-1} \dot{h}(\beta)\right),$$

et donc

$$\sqrt{n} \left(1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma}\right) \stackrel{H_{1,n}}{\rightsquigarrow} \mathcal{N}\left(-\delta^t \frac{u^2(1+\gamma)(1+2\gamma)}{\gamma(u+1+\gamma)(u+1)^2} \Big|_{u=1,\dots,J}, \frac{(1+\gamma)^2(1+2\gamma)}{\gamma^2}\right).$$

### 2.7.7 Preuve de la Proposition 2.7

Selon la Proposition 2.4, la représentation asymptotique de  $Z_n(\widehat{\beta}_n)$  sous  $H_{1,n}$  est la même que celle précédemment obtenue, c'est-à-dire :

$$\begin{aligned} Z_n(\widehat{\beta}_n) &= Z_n(\beta) + \sqrt{n} \left(1 - \frac{\widehat{\gamma}_n \sigma}{\widehat{\sigma}_n \gamma}\right) \frac{s}{(s+\gamma+1)(s+1)} \Big|_{s=1,\dots,J} \\ &\quad + \sqrt{n}(\widehat{\gamma}_n - \gamma) \frac{1}{\widehat{\gamma}_n} \frac{s}{(s+1)^2} \Big|_{s=1,\dots,J} + o_{\mathbb{P}_{H_{1,n}}}(1). \end{aligned} \quad (2.19)$$

En utilisant les Propositions 2.5 et 2.6, et avec  $I$  défini dans (2.12), nous obtenons

$$\begin{aligned} \mathbb{E}_{H_{1,n}}(Z_n(\beta)) &= I\delta = \frac{uv}{(u+v+1)(u+1)(v+1)} \Big|_{u,v=1,\dots,J} \delta, \\ \mathbb{E}_{H_{1,n}} \left( \sqrt{n} \left( 1 - \frac{\hat{\gamma}_n \sigma}{\hat{\sigma}_n \gamma} \right) \frac{v}{(v+\gamma+1)(v+1)} \Big|_{v=1,\dots,J} \right) \\ &= - \frac{u^2 v (1+\gamma)(1+2\gamma)}{\gamma(u+1+\gamma)(v+1+\gamma)(u+1)^2(v+1)} \Big|_{u,v=1,\dots,J} \delta, \\ \text{et } \mathbb{E}_{H_{1,n}} \left( \sqrt{n}(\hat{\gamma}_n - \gamma) \frac{v}{\hat{\gamma}_n(v+1)^2} \Big|_{v=1,\dots,J} \right) &= \frac{uv(1+\gamma)(u-\gamma)}{\gamma(u+1+\gamma)(u+1)^2(v+1)^2} \Big|_{u,v=1,\dots,J} \delta. \end{aligned} \quad (2.20)$$

Des calculs directs mènent à l'expression suivante pour l'espérance asymptotique de  $Z_n(\hat{\beta}_n)$  sous  $H_{1,n}$  :

$$\begin{aligned} \mathbb{E}_{a;H_{1,n}}(Z_n(\hat{\beta}_n)) &= \left[ \frac{uv}{(u+v+1)(u+1)(v+1)} \right. \\ &\quad \left. - \frac{uv(1+\gamma)(uv+\gamma+(u+1)(v+1))}{(v+\gamma+1)(u+\gamma+1)(u+1)^2(v+1)^2} \right] \Big|_{u,v=1,\dots,J} \delta \\ &= \Sigma_J(\beta)\delta. \end{aligned} \quad (2.21)$$

En combinant (2.18), (2.19) et les Propositions 2.4 et 2.5, il est trivial de remarquer que les termes de covariances sont identiques sous  $H_{1,n}$  et sous  $H_0$ . En effet, en examinant le troisième lemme de LeCam (cf. Annexe 2.8.2), nous constatons que seuls les termes d'espérance diffèrent. Par conséquent,  $\text{Var}_{a;H_{1,n}}(Z_n(\hat{\beta}_n)) = \Sigma_J(\beta)$ . Par ailleurs, d'après (2.19), la quantité  $Z_n(\hat{\beta}_n)$  peut être décomposée en trois termes plus un terme de reste. Selon les Propositions 2.5 et 2.6 et le théorème de Slutsky, le vecteur formé par ces trois termes est asymptotiquement gaussien sous  $H_{1,n}$ . Ainsi, sous  $H_{1,n}$ ,  $Z_n(\hat{\beta}_n)$  converge en distribution vers une gaussienne  $\mathcal{N}_J(\Sigma_J(\beta)\delta, \Sigma_J(\beta))$ . Par conséquent,  $Z_n^t(\hat{\beta}_n)\Sigma_J^{-1}(\beta)Z_n(\hat{\beta}_n) \xrightarrow{H_{1,n}} \chi_{J,\delta^t\Sigma_J(\beta)\delta}^2$ , ce qui permet de déduire la Proposition 2.7 par la Proposition 2.5 et par continuité.

## 2.8 Annexes

### 2.8.1 Définitions d'après Pfanzagl [56]

Soient  $X, X_1, \dots, X_n$  des variables aléatoires indépendantes et identiquement distribuées dont la distribution dépend d'un paramètre  $\zeta \in \mathcal{S}$  et est notée  $F_\zeta$ . Dans cette section, sauf mention contraire, l'espace  $\mathcal{S}$  sera supposé localement compact. Il s'agit donc d'un espace séparé (dit de Hausdorff) dont tout point admet un voisinage compact. Notons  $\mathbb{P}_\zeta$  la probabilité sous  $F_\zeta$  définie sur l'espace probabilisable  $(\mathcal{X}, \mathcal{A})$  et  $\mathbb{E}_\zeta$  l'espérance associée.

Avant de donner quelques définitions et théorèmes de convergence, il est important de bien comprendre ce qu'on entend par « localement uniformément » :

« localement uniformément en  $\zeta_0$  » signifie « uniformément sur un voisinage de  $\zeta_0$  ».

« localement uniformément sur  $S$  » signifie « localement uniformément en  $\zeta_0$  pour tout  $\zeta_0 \in S$  ».

**Définition 2.1** Une famille de fonctions  $h(\cdot, \zeta)$  est  $\mathbb{P}_\zeta$ -intégrable localement uniformément en  $\zeta_0$  s'il existe un voisinage  $V(\zeta_0)$  tel que

$$\lim_{M \rightarrow \infty} \sup_{\zeta \in V(\zeta_0)} \mathbb{E}_\zeta (|h(X, \zeta)| \mathbf{1}_{\{|h(X, \zeta)| > M\}}) = 0.$$

**Remarque 2.2** S'il existe  $\eta > 0$  tel que  $\sup_{\zeta \in V(\zeta_0)} \mathbb{E}_\zeta (|h(X, \zeta)|^{1+\eta}) < \infty$  alors  $h(\cdot, \zeta)$  est  $\mathbb{P}_\zeta$ -intégrable localement uniformément en  $\zeta_0$ . En effet, comme  $\mathbb{E}_\zeta (|h(X, \zeta)| \mathbf{1}_{\{|h(X, \zeta)| > M\}}) \leq M^{-\eta} \mathbb{E}_\zeta (|h(X, \zeta)|^{1+\eta})$ , il suffit que  $\sup_{\zeta \in V(\zeta_0)} \mathbb{E}_\zeta (|h(X, \zeta)|^{1+\eta})$  soit finie avec  $\eta > 0$  pour conclure.

**Définition 2.2** Posons  $Y_n = H(X_1, \dots, X_n; \zeta)$ , où la fonction  $H$  est à valeurs dans  $\mathbb{R}^q$  et soit  $Q_{n, \zeta}$  la loi de  $Y_n$ . La suite  $Y_n$ ,  $n \in \mathbb{N}$ , converge en distribution vers  $Q_\zeta$  localement

uniformément sur  $\mathcal{S}$  si pour tout  $\zeta_0 \in \mathcal{S}$ , il existe un voisinage  $V(\zeta_0)$  de  $\zeta_0 \in \mathcal{S}$  tel que

$$\lim_{n \rightarrow \infty} \sup_{\zeta \in V(\zeta_0)} |Q_{n,\zeta}(h) - Q_\zeta(h)| = 0 \text{ pour tout } h \in \mathcal{C}$$

où  $\mathcal{C}$  représente la classe des fonctions bornées et continues  $h : \mathbb{R}^q \mapsto \mathbb{R}$ .

**Définition 2.3** Une quantité aléatoire  $Q(X_1, \dots, X_n; \zeta)$  est  $\sqrt{n}$ -consistante localement uniformément sur  $\mathcal{S}$  pour  $Q(\zeta)$  si pour tout  $\zeta_0$ , il existe un voisinage  $V(\zeta_0)$  de  $\zeta_0 \in \mathcal{S}$  tel que pour tout  $\varepsilon > 0$ , il existe  $M_\varepsilon > 0$  tel que

$$\sup_{\zeta \in V(\zeta_0)} \mathbb{P}_\zeta(\sqrt{n} \|Q(X_1, \dots, X_n; \zeta) - Q(\zeta)\| > M_\varepsilon) < \varepsilon \text{ pour tout } n \in \mathbb{N}$$

où  $\|\cdot\|$  est la norme Euclidienne usuelle.

Remarquons également qu'une quantité aléatoire  $Q(X_1, \dots, X_n; \zeta)$  est  $\sqrt{n}$ -consistante localement uniformément sur  $\mathcal{S}$  pour  $Q(\zeta)$  si la quantité  $\sqrt{n}(Q(X_1, \dots, X_n; \zeta) - Q(\zeta))$  converge en distribution localement uniformément sur  $\mathcal{S}$  vers une loi ayant une variance finie.

**Théorème 2.1 (Version uniforme du théorème central limite)** Soit  $\mathcal{S}$  un espace quelconque et  $f(\cdot, \zeta)$  une fonction mesurable de  $(\mathcal{X}, \mathcal{A})$  dans  $(\mathbb{R}^q, \mathcal{B}_{\mathbb{R}^q})$  avec  $\mathcal{B}_{\mathbb{R}^q}$  la tribu des boréliens de  $\mathbb{R}^q$  telle que

$$\mathbb{E}_\zeta(f(X, \zeta)) = 0 \text{ pour } \zeta \in \mathcal{S}$$

et pour  $\zeta \in \mathcal{S}$ ,  $\|f(\cdot, \zeta)\|^2$  est uniformément  $\mathbb{P}_\zeta$ -intégrable sur  $\mathcal{S}$ . Soit

$$\Sigma(\zeta) := \mathbb{E}_\zeta(f(X, \zeta)f(X, \zeta)^t).$$

Alors

$$n^{-1/2} \sum_{i=1}^n f(X_i, \zeta) \rightsquigarrow \mathcal{N}(0, \Sigma(\zeta)) \quad \text{uniformément sur } \mathcal{S}.$$

**Théorème 2.2 (Méthode delta)** Soit  $T$  un ouvert de  $\mathbb{R}^q$ . Soit  $\widehat{\zeta}_n \in T$  un estimateur tel que

$$\sqrt{n}(\widehat{\zeta}_n - \zeta) \rightsquigarrow \mathcal{N}(0, \Sigma(\zeta)) \quad \text{localement uniformément en } \zeta_0.$$

Supposons que  $h : T \rightarrow \mathbb{R}^p$ ,  $p \leq q$ , ait ses dérivées partielles continues sur  $\mathcal{S}$  et notons  $H_{ij}(t) := h_i^{(j)}(t_1, \dots, t_q)$ ,  $i = 1, \dots, p$ ,  $j = 1, \dots, q$ , où  $H(\zeta_0)$  est de rang  $p$ .

Alors

$$\sqrt{n}(h(\widehat{\zeta}_n) - h(\zeta)) \rightsquigarrow \mathcal{N}(0, H(\zeta)\Sigma(\zeta)H(\zeta)^t) \quad \text{localement uniformément en } \zeta_0.$$

**Définition 2.4** Une quantité aléatoire  $Q(X_1, \dots, X_n; \zeta)$  converge en probabilité ( $\xrightarrow{\mathbb{P}_\zeta}$ ) localement uniformément sur  $\mathcal{S}$  vers  $Q(\zeta)$  si pour tout  $\zeta_0 \in \mathcal{S}$ , il existe un voisinage  $V(\zeta_0)$  de  $\zeta_0$  tel que

$$\lim_{n \rightarrow \infty} \sup_{\zeta \in V(\zeta_0)} \mathbb{P}_\zeta(\|Q(X_1, \dots, X_n; \zeta) - Q(\zeta)\| > \varepsilon) = 0 \quad \text{pour } \varepsilon > 0.$$

Nous dirons qu'une quantité aléatoire  $Q(X_1, \dots, X_n; \zeta)$  est consistante localement uniformément sur  $\mathcal{S}$  pour  $Q(\zeta)$  si  $Q(X_1, \dots, X_n; \zeta)$  converge en probabilité localement uniformément sur  $\mathcal{S}$  vers  $Q(\zeta)$ .

**Proposition 2.8 (Version uniforme locale du théorème des transformations continues)** Soit  $g : \mathbb{R}^k \mapsto \mathbb{R}^m$  une fonction continue en tout point de l'ensemble  $\mathcal{S}$  tel que  $\mathbb{P}(\zeta \in \mathcal{S}) = 1$ . Si  $\widehat{\zeta}_n$  converge en probabilité vers  $\zeta$  localement uniformément sur  $\mathcal{S}$ , alors  $g(\widehat{\zeta}_n)$  converge en probabilité vers  $g(\zeta)$  localement uniformément sur  $\mathcal{S}$ .

La preuve de cette proposition est une adaptation directe de celle du Théorème 2.3 de van der Vaart [69].

**Théorème 2.3 (Théorème 7.5.9. de Pfanzagl [56])** Soit  $\mathcal{S}$  un ouvert de  $\mathbb{R}^p$  et  $\{\mathbb{P}_\zeta : \zeta \in \mathcal{S}\}$  une famille de mesures de probabilité dominées par une mesure  $\mu$   $\sigma$ -finie. Soit  $f(\cdot, \zeta)$  une densité de  $\mathbb{P}_\zeta$  par rapport à  $\mu$ . Notons  $\ell(\cdot, \zeta) = \log f(\cdot, \zeta)$ . Les premières et deuxièmes dérivées partielles de  $\ell(x, \zeta)$  sont notées  $\ell^{(i)}(x, \zeta)$  et  $\ell^{(ij)}(x, \zeta)$ ,  $i, j = 1, \dots, p$ . Nous noterons  $\ell^{(\cdot)}(x, \zeta)$  le vecteur colonne des  $\ell^{(i)}(x, \zeta)$ ,  $i = 1, \dots, p$ . Soit  $L(\zeta)$  la  $p \times p$ -matrice formée des éléments :

$$L_{ij}(\zeta) := \mathbb{E}_\zeta(\ell^{(i)}(X, \zeta)\ell^{(j)}(X, \zeta)), \quad i, j = 1, \dots, p.$$

Supposons que la famille  $\{\mathbb{P}_\zeta : \zeta \in \mathcal{S}\}$  satisfasse les conditions suivantes :

- $\mathbb{E}_\zeta (\ell^{(\cdot)}(X, \zeta)) = 0$  avec  $\zeta \in \mathcal{S}$ .
- $\mathbb{E}_\zeta (\ell^{(ij)}(X, \zeta)) + \mathbb{E}_\zeta (\ell^{(i)}(X, \zeta)\ell^{(j)}(X, \zeta)) = 0$  pour  $i, j = 1, \dots, p$  et  $\zeta \in \mathcal{S}$ .
- Les fonctions  $\ell^{(i)}(\cdot, \zeta)$  pour  $i = 1, \dots, p$  et  $\zeta \in \mathcal{S}$  sont linéairement  $\mathbb{P}_\zeta$ -indépendantes.
- Il existe un voisinage  $U(\zeta_0)$  de  $\zeta_0$  tel que

$\zeta \mapsto \ell^{(ij)}(x, \zeta)$  est continue sur  $U(\zeta_0)$  pour  $i, j = 1, \dots, p$ , pour tout  $x$

et  $\sup_{\zeta_1 \in U(\zeta_0)} |\ell^{(ij)}(\cdot, \zeta_1)|$  est  $\mathbb{P}_\zeta$ -intégrable localement uniformément en  $\zeta_0$ .

- Pour  $i = 1, \dots, p$ , la famille de fonctions  $x \mapsto [\ell^{(i)}(x, \zeta)]^2$  est  $\mathbb{P}_\zeta$ -intégrable, localement uniformément en  $\zeta_0$ .

Considérons maintenant une quantité aléatoire  $\tilde{\zeta}_n \in \mathbb{R}^p$   $\sqrt{n}$ -consistante localement uniformément pour  $\zeta$  en  $\zeta_0$ . Soit  $\Lambda(X_1, \dots, X_n)$  un estimateur consistant pour  $\Lambda(\zeta) = L(\zeta)^{-1}$  localement uniformément en  $\zeta_0$ . Alors l'estimateur

$$\hat{\zeta}_n := \tilde{\zeta}_n + \Lambda(X_1, \dots, X_n) \frac{1}{n} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \tilde{\zeta}_n) \quad (2.22)$$

vérifie la représentation asymptotique suivante

$$\sqrt{n}(\hat{\zeta}_n - \zeta) - \Lambda(\zeta) \frac{1}{\sqrt{n}} \sum_{i=1}^n \ell^{(\cdot)}(X_i, \zeta) \xrightarrow{\mathbb{P}_\zeta} 0 \text{ localement uniformément en } \zeta_0$$

ainsi que la convergence en distribution suivante

$$\sqrt{n}(\hat{\zeta}_n - \zeta) \rightsquigarrow \mathcal{N}(0, \Lambda(\zeta)) \text{ localement uniformément en } \zeta_0.$$

## 2.8.2 Théorie de LeCam - Normalité Asymptotique Locale

Dans la construction de la statistique de test proposée dans ce chapitre ainsi que dans son étude de comportement asymptotique sous des alternatives locales, nous avons utilisé le concept de normalité asymptotique locale. Les rappels donnés dans cette section sont

principalement issus de van der Vaart [69].

Une suite de modèles statistiques est « localement asymptotiquement normale » si asymptotiquement le logarithme du rapport de vraisemblance associé admet un certain développement quadratique. Une définition plus formelle est donnée ci-dessous :

**Définition 2.5** *La suite de modèles statistiques  $\{P_{n,\theta} : \theta \in \Theta\}$  est localement asymptotiquement normale (LAN) en  $\theta$  s'il existe des matrices  $r_n$  et  $I_\theta$  ainsi que des vecteurs aléatoires  $\Delta_{n,\theta}$  tel que  $\Delta_{n,\theta} \xrightarrow{P_{n,\theta}} \mathcal{N}(0, I_\theta)$  et pour toute suite convergente  $h_n \rightarrow h$*

$$\log \frac{dP_{n,\theta+r_n^{-1}h_n}}{dP_{n,\theta}} = h^t \Delta_{n,\theta} - \frac{1}{2} h^t I_\theta h + o_{P_{n,\theta}}(1).$$

Sous la normalité asymptotique locale, nous avons donc

$$\log \frac{dP_{n,\theta+r_n^{-1}h_n}}{dP_{n,\theta}} \xrightarrow{P_{n,\theta}} \mathcal{N} \left( -\frac{1}{2} h^t I_\theta h, h^t I_\theta h \right).$$

Par conséquent il va être possible d'utiliser le premier lemme de LeCam rappelé ci-dessous et d'en déduire que  $P_{n,\theta}$  et  $P_{n,\theta+r_n^{-1}h_n}$  sont mutuellement contigües (noté  $P_{n,\theta} \triangleleft P_{n,\theta+r_n^{-1}h_n}$ ). Soient  $P_n$  et  $Q_n$  des suites de mesures de probabilité sur des espaces probabilisables  $(\mathcal{X}_n, \mathcal{A}_n)$ .

**Premier lemme de LeCam :** *Les assertions suivantes sont équivalentes :*

- (i)  $Q_n$  est contigüe par rapport à  $P_n$ , ce qui est noté  $Q_n \triangleleft P_n$ .
- (ii) Il existe une sous-suite  $m(n)$  telle que si  $dP_{m(n)}/dQ_{m(n)} \xrightarrow{Q_{m(n)}} U$ , alors  $\mathbb{P}(U > 0) = 1$ .
- (iii) Il existe une sous-suite  $m(n)$  telle que si  $dQ_{m(n)}/dP_{m(n)} \xrightarrow{P_{m(n)}} V$ , alors  $\mathbb{E}(V) = 1$ .
- (iv) Pour toute statistique  $T_n : \Omega_n \mapsto \mathbb{R}^k$  : si  $T_n \xrightarrow{P_n} 0$ , alors  $T_n \xrightarrow{Q_n} 0$ .

La quatrième assertion, la plus utilisée, est très importante car elle assure que si une représentation asymptotique est valable sous  $P_n$  alors elle l'est aussi sous  $Q_n$ .

Les arguments de contigüité sont des outils pour obtenir la distribution limite d'une suite



de statistiques sous des lois  $Q_n$  à partir de la distribution limite sous des lois  $P_n$ . Le plus souvent,  $P_n$  représente l'hypothèse nulle alors que  $Q_n$  représente l'hypothèse alternative. Le troisième lemme de LeCam décrit ci-dessous peut donc être utilisé pour trouver le comportement d'une statistique sous  $Q_n$  :

**Troisième lemme de LeCam :** Soit  $X_n : \Omega_n \mapsto \mathbb{R}^k$  une suite de vecteurs aléatoires.

Supposons

$$\begin{pmatrix} X_n \\ \log \frac{dQ_n}{dP_n} \end{pmatrix} \stackrel{P_n}{\rightsquigarrow} \mathcal{N}_{k+1} \left( \begin{pmatrix} \mu \\ -\frac{1}{2}\sigma^2 \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^t & \sigma^2 \end{pmatrix} \right),$$

$$\text{alors } X_n \stackrel{Q_n}{\rightsquigarrow} \mathcal{N}_k(\mu + \tau, \Sigma).$$

Nous pouvons remarquer que la matrice de variance-covariance asymptotique de  $X_n$  reste identique sous  $P_n$  et sous  $Q_n$ . C'est en fait le vecteur moyenne qui diffère en fonction de la covariance asymptotique  $\tau$  entre  $X_n$  et le rapport de log-vraisemblance.

### 2.8.3 Simulation par la méthode de rejet

Cette technique de simulation permet de simuler une loi de densité  $g$  quelconque à partir d'une densité  $f$  facile à simuler. Cette méthode particulière de simulation introduite par von Neumann [70] repose sur la proposition suivante :

**Proposition 2.9** Soit  $\mu$  une mesure positive sur  $\mathbb{R}^k$ ,  $f$  et  $g$  deux densités de probabilité sur  $(\mathbb{R}^k, \mu)$  telles qu'il existe une constante  $c$  vérifiant :

$$\forall x \in \mathbb{R}^k, cf(x) \geq g(x).$$

Soit  $X$  une variable aléatoire de densité  $f$  par rapport à  $\mu$  et  $U$  une variable aléatoire de loi uniforme sur  $[0, 1]$ , indépendante de  $X$ . Alors la loi conditionnelle de  $X$  sachant l'évènement  $(cUf(X) < g(X))$  a pour densité  $g$  par rapport à  $\mu$ .

L'algorithme de simulation est le suivant

**Répéter**

**Simuler  $X$  de densité  $f$**

**Simuler  $U$  de loi uniforme sur  $[0, 1]$**

**jusqu'à obtenir  $\{cUf(X) < g(X)\}$ .**

Le nombre de passages dans la boucle suit la loi géométrique  $\mathcal{G}(\frac{1}{c})$ , d'espérance  $c$ . On aura donc intérêt à choisir  $f$  assez proche de  $g$  de sorte que la constante  $c = \max(g(x)/f(x))$  soit la plus petite possible.

Pour simuler la loi de densité  $g_J$  définie en (2.1), cette méthode peut être appliquée. En effet, si nous imposons que les  $\delta_i$ ,  $i = 1, \dots, J$  soient égaux à une unique valeur  $\delta > 0$ , nous obtenons l'inégalité suivante

$$\exp \left\{ \sum_{s=1}^J \theta_s \bar{F}^s(x, \beta) - K(\theta) \right\} \leq \exp \left( \frac{J\delta}{\sqrt{n}} \right) = c,$$

ainsi, pour  $n$ ,  $\beta$  et  $\theta$  fixés et pour tout  $x \in [0, \infty)$ ,

$$g_J(x, \theta, \beta) \leq c f(x, \beta).$$



# Chapitre 3

## Estimation en présence de censure

### Résumé

Dans ce troisième chapitre, nous nous plaçons dans le cadre de données censurées aléatoirement à droite. Dans ce contexte, différents estimateurs de l'indice des valeurs extrêmes ont été proposés par Beirlant *et al.* [7]. Ils sont tous construits de la même façon : l'estimateur usuel (sans censure) est divisé par la proportion d'observations non censurées au-delà d'un certain seuil. Récemment, Einmahl *et al.* [31] ont établi la normalité asymptotique de tels estimateurs. Une autre façon d'aborder le problème consiste à utiliser l'approche POT et à adapter au contexte des données censurées la vraisemblance. Nous pouvons ainsi obtenir des estimateurs du maximum de vraisemblance pour  $(\gamma_1, \sigma_{1,t})$ , les paramètres de la GPD. Cependant les propriétés asymptotiques de ces derniers sont très difficiles à obtenir. Nous proposons donc dans ce chapitre un estimateur du couple  $(\gamma_1, \sigma_{1,t})$  basé sur une seule étape de l'algorithme de Newton-Raphson lorsque les données sont censurées aléatoirement à droite. Nous établissons la normalité asymptotique de ces estimateurs et montrons par le biais de simulations que, à distance finie, ils se comportent de façon similaire à ceux du maximum de vraisemblance.

### 3.1 Introduction

Nous allons nous intéresser dans ce chapitre au problème de l'estimation de l'indice des valeurs extrêmes et cela en présence de données censurées aléatoirement à droite. Ce problème est très récent dans la littérature. Il a été mentionné dans le livre de Reiss et Thomas [63] mais n'a été réellement abordé qu'en 2007 par Beirlant *et al.* [7] et l'année suivante par Einmahl *et al.* [31]. Pourtant ses applications sont multiples, dont une, naturelle, concerne le domaine médical. Dans ce contexte, nous nous intéressons par exemple à la durée de survie de patients tous sujets à une même maladie. Ces patients sont suivis pendant un certain intervalle de temps et à la fin de la période d'observation, le point est fait sur leur état. Pour ceux décédés pendant la période d'observation, nous disposons exactement de leur date de décès. En revanche, pour ceux toujours en vie à la fin de l'étude, nous n'avons aucune indication sur leur future date de décès. En conséquence, ces données sont dites censurées aléatoirement à droite. Un tel exemple a été considéré dans Beirlant *et al.* [7] ainsi que dans Einmahl *et al.* [31].

Formellement, dire que nous sommes en présence de données censurées aléatoirement à droite revient à dire que l'échantillon  $X_1, \dots, X_n$  (des durées réelles de vie) n'est pas observé mais est censuré par un deuxième échantillon  $Y_1, \dots, Y_n$ . Ces deux échantillons sont formés de variables aléatoires indépendantes et identiquement distribuées de loi  $F$  et  $G$  respectivement et sont supposés indépendants entre eux. Les variables que nous observons sont d'une part les  $Z_i$  définies par  $Z_i = X_i \wedge Y_i$ ,  $i = 1, \dots, n$ , et d'autre part les indicateurs de censure  $\delta_i = \mathbb{1}_{\{X_i \leq Y_i\}}$ ,  $i = 1, \dots, n$ . Autrement dit, nous savons si la donnée observée a été censurée ou non. Il est clair que les  $Z_i$  sont des variables indépendantes, de loi  $H$  liée à  $F$  et  $G$  par la relation  $\overline{H}(z) := 1 - H(z) = \overline{F}(z)\overline{G}(z)$ . Le point terminal de  $H$  sera noté dans toute la suite  $\tau_H = \sup\{x : H(x) < 1\}$ .

Nous supposons dans la suite l'hypothèse très générale que  $F$  et  $G$  sont deux distri-

butions dans un domaine d'attraction des extrêmes, d'indices respectifs  $\gamma_1$  et  $\gamma_2$ . Ceci implique que l'indice des valeurs extrêmes de  $H$  existe et il sera noté dans toute la suite  $\gamma$ .

Beirlant *et al.* [7] ont proposé différents estimateurs de  $\gamma_1$ , l'indice des valeurs extrêmes associé à  $F$ . Ces derniers sont tous construits de façon similaire, à partir d'un estimateur non adapté à la censure, par exemple l'estimateur de Hill [40], l'estimateur des moments de Dekkers *et al.* [26] ou encore l'estimateur  $UH$  (Beirlant *et al.* [9]). Ces estimateurs, basés sur les observations  $Z_i$ , estiment par conséquent l'indice  $\gamma$  de  $H$ . Il s'agit alors de les modifier de façon à estimer  $\gamma_1$  et non  $\gamma$ . Une façon de procéder (cf. Beirlant *et al.* [7]) consiste à diviser ces estimateurs usuels (non adaptés à la censure) par la proportion de données non censurées au-delà d'un seuil  $t$ , c'est-à-dire à utiliser

$$\widehat{\gamma}_{Z,t}^{(c,\cdot)} := \frac{\widehat{\gamma}_{Z,t}^{(\cdot)}}{\frac{1}{N_t} \sum_{j=1}^n \delta_j \mathbb{1}_{\{Z_j > t\}}} \quad (3.1)$$

avec  $N_t$  le nombre d'excès au-delà de  $t$ . Plusieurs estimateurs adaptés à la censure sont possibles ( $\widehat{\gamma}_{Z,t}^{(c,H)}$ ,  $\widehat{\gamma}_{Z,t}^{(c,M)}$ , ou encore  $\widehat{\gamma}_{Z,t}^{(c,UH)}$ ) selon le choix de la méthode d'estimation de départ  $\widehat{\gamma}_{Z,t}^{(\cdot)}$ . Récemment, Einmahl *et al.* [31] ont établi, de façon unifiée, la normalité asymptotique de tout estimateur de l'indice des valeurs extrêmes écrit sous la forme (3.1) dans le cas où le seuil choisi,  $t$ , est aléatoire et égal à  $Z_{n-k,n}$ , la  $n - k$ -ième statistique d'ordre de l'échantillon  $Z_1, \dots, Z_n$ .

Comme nous l'avons vu dans le chapitre d'introduction, ces estimateurs  $\widehat{\gamma}_{Z,t}^{(\cdot)}$  découlent de la première caractérisation d'appartenance à un domaine d'attraction. Nous pourrions cependant envisager d'utiliser la méthode POT pour construire de nouveaux estimateurs. Il est alors nécessaire d'adapter la vraisemblance à la censure. Nous pouvons ainsi proposer d'autres estimateurs, non seulement de l'indice mais également du deuxième paramètre  $\sigma_{1,t}$  de la GPD, distribution limite des excès  $E_j = Z_j - t$ , sachant  $Z_j > t$ , au-delà d'un seuil  $t$  quand  $t \rightarrow \tau_H$ . Dans ce cas, la vraisemblance s'exprime selon (cf. Andersen *et al.*

[1], p. 411) :

$$L(\gamma_1, \sigma_{1,t}) = \prod_{j=1}^{N_t} [g_{\gamma_1, \sigma_{1,t}}(E_j)]^{\delta_j} [1 - G_{\gamma_1, \sigma_{1,t}}(E_j)]^{1-\delta_j} \quad (3.2)$$

où  $1 - G_{\gamma_1, \sigma_{1,t}}(x) = \left(1 + \frac{\gamma_1}{\sigma_{1,t}}x\right)^{-\frac{1}{\gamma_1}}$  et  $g_{\gamma_1, \sigma_{1,t}}$  est la densité associée.

En dérivant cette vraisemblance par rapport à ses deux paramètres, nous obtenons un système de deux équations à deux inconnues. Comme toujours dans ce type de contexte, nous ne pouvons pas en déduire explicitement les estimateurs du maximum de vraisemblance, mais pouvons les obtenir numériquement par des méthodes de type Newton-Raphson. Malgré ces difficultés numériques, les estimateurs du maximum de vraisemblance sont en général appréciés pour leurs bonnes propriétés asymptotiques. Cependant, établir ces dernières demeure extrêmement difficile. Il s'agit d'un sujet d'actualité qui, dans le cas où il n'y a pas de censure, a donné lieu à plusieurs publications (cf. Smith [66], Drees [29], Drees *et al.* [30]). Dans le cas censuré, cette normalité est encore plus problématique à établir et n'a pas trouvé de solution à ce jour. Nous proposons donc, dans ce chapitre, de nous intéresser à des estimateurs de  $(\gamma_1, \sigma_{1,t})$  appelés estimateurs « one-step », obtenus après une seule itération de la procédure de Newton-Raphson, notés  $(\widehat{\gamma}_{Z,t}^{(c,OS)}, \widehat{\sigma}_{Z,t}^{(c,OS)})$ . Nous montrons également, par simulations, qu'à distance finie ces estimateurs se comportent de façon similaire à ceux du maximum de vraisemblance.

Dans la suite nous supposerons que les indices  $\gamma_1$  et  $\gamma_2$  associés respectivement à  $F$  et  $G$  sont de même signe et que les points terminaux de  $F$  et  $G$  sont identiques. Ceci donne lieu à trois cas possibles :

$$\left\{ \begin{array}{l} \text{cas 1 : } \gamma_1 > 0, \gamma_2 > 0, \\ \text{cas 2 : } \gamma_1 < 0, \gamma_2 < 0, \tau_F = \tau_G, \\ \text{cas 3 : } \gamma_1 = \gamma_2 = 0, \tau_F = \tau_G. \end{array} \right.$$

Notons que les autres cas (en particulier les mixtures entre domaines d'attraction) ne sont pas intéressants. En effet, ils correspondent (du moins asymptotiquement) à un taux de

censure proche de 0 et donc à la théorie existante (par exemple si  $\gamma_1 < 0$  et  $\gamma_2 > 0$ ) ou à un taux de censure proche de 1 (par exemple si  $\gamma_1 > 0$  et  $\gamma_2 < 0$ ) pour lequel une bonne estimation de l'indice semble impossible. Ceci justifie donc pleinement que seuls ces trois cas soient étudiés dans la suite du chapitre. Remarquons que dans les deux premiers cas, il est facile de vérifier que  $\gamma = \frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}$ . Dans le dernier cas,  $\gamma$  est nul et par convention nous le noterons également  $\frac{\gamma_1\gamma_2}{\gamma_1+\gamma_2}$ .

Le résultat principal de ce chapitre est la normalité asymptotique du vecteur  $(\widehat{\gamma}_{Z,t}^{(c,OS)}, \widehat{\sigma}_{Z,t}^{(c,OS)})'$  correctement renormalisé; une démonstration en sera proposée en Section 3.4 et les calculs techniques associés détaillés en annexe. Plusieurs exemples seront détaillés en Section 3.3. Enfin, une étude par simulations sera présentée afin d'illustrer la proximité entre les estimateurs « one-step » que nous proposons ici et pour lesquels nous avons établi des résultats asymptotiques, et ceux du maximum de vraisemblance obtenus en maximisant (3.2).

## 3.2 Résultat principal

Notons  $U_F$  (resp.  $U_G$ ) la fonction queue de  $F$  (resp. de  $G$ ) définie par  $U_F(x) := F^{\leftarrow}(1 - \frac{1}{x}) = \inf\{y : F(y) \geq 1 - \frac{1}{x}\}$  (resp.  $U_G(x) := G^{\leftarrow}(1 - \frac{1}{x})$ ). Nous supposons l'existence de deux fonctions positives  $a_F$  et  $a_G$  telles que

$$\lim_{x \rightarrow \infty} \frac{U_F(xu) - U_F(x)}{a_F(x)} = \int_1^u v^{\gamma_1-1} dv =: h_{\gamma_1}(u) \quad \text{pour tout } u > 0 \quad (3.3)$$

$$\lim_{x \rightarrow \infty} \frac{U_G(xu) - U_G(x)}{a_G(x)} = \int_1^u v^{\gamma_2-1} dv =: h_{\gamma_2}(u) \quad \text{pour tout } u > 0. \quad (3.4)$$

Nous dirons que la fonction queue  $U_F$  (resp.  $U_G$ ) satisfait la condition des valeurs extrêmes  $C_{\gamma_1}(a_F)$  (resp.  $C_{\gamma_2}(a_G)$ ) si (3.3) (resp. (3.4)) est vérifiée avec la fonction  $a_F$  (resp.  $a_G$ ). Il s'agit là d'une condition du premier ordre qui n'est pas suffisante pour établir un résultat de normalité. Pour de tels résultats, il est indispensable de supposer une condition



du second ordre qui peut se formaliser comme suit (cf. de Haan et Stadtmüller [39]).

Supposons pour tout  $u > 0$  que :

$$\frac{U_F(xu) - U_F(x)}{a_F(x)} - h_{\gamma_1}(u) \sim a_{2,F}(x)k_F(u), \quad x \rightarrow \infty \quad (3.5)$$

$$\frac{U_G(xu) - U_G(x)}{a_G(x)} - h_{\gamma_2}(u) \sim a_{2,G}(x)k_G(u), \quad x \rightarrow \infty \quad (3.6)$$

où  $a_{2,F}, a_{2,G} \rightarrow 0$  sont des fonctions à variation régulière  $\mathcal{R}_{\rho_1}, \mathcal{R}_{\rho_2}$  respectivement, avec  $\rho_1 \leq 0$  et  $\rho_2 \leq 0$  et

$$\begin{aligned} k_F(u) &= A_F h_{\gamma_1+\rho_1}(u) + c_F \int_1^u t^{\gamma_1-1} h_{\rho_1}(t) dt \\ k_G(u) &= A_G h_{\gamma_2+\rho_2}(u) + c_G \int_1^u t^{\gamma_2-1} h_{\rho_2}(t) dt, \end{aligned}$$

où  $A_F, A_G, c_F$  et  $c_G$  sont des constantes convenables.

Notons que si  $\rho_1$  (resp.  $\rho_2$ ) est strictement négatif, alors un choix approprié de  $a_F$  (resp. de  $a_G$ ) implique une simplification de la fonction limite  $k_F$  (resp.  $k_G$ ) alors égale à  $(A_F + \rho_1^{-1}c_F)h_{\gamma_1+\rho_1}(u)$  (resp.  $(A_G + \rho_2^{-1}c_G)h_{\gamma_2+\rho_2}(u)$ ) (cf. Proposition 3.2 de Beirlant *et al.* [6]).

Suivant le problème considéré, il est parfois plus facile de travailler, non pas avec une fonction queue, mais avec la fonction de répartition elle-même. Dans ce cas, sous les conditions du premier et second ordre (3.3)-(3.6) et lorsque  $U_F$  et  $U_G$  sont continues, nous pouvons dire que

$$\frac{1}{\chi_1(t)} \left\{ \frac{1 - F(t + v \sigma_{1,t})}{1 - F(t)} - \eta_{\gamma_1}(v) \right\} \longrightarrow \psi_1(v), \quad t \rightarrow \tau_F \quad (3.7)$$

$$\frac{1}{\chi_2(t)} \left\{ \frac{1 - G(t + v \sigma_{2,t})}{1 - G(t)} - \eta_{\gamma_2}(v) \right\} \longrightarrow \psi_2(v), \quad t \rightarrow \tau_G \quad (3.8)$$

où  $\chi_1(t) = a_{2,F}(U_F^-(t))$ ,  $\chi_2(t) = a_{2,G}(U_G^-(t))$ ,  $\eta_{\gamma}(v) = (1 + \gamma v)^{-1/\gamma}$ ,  $\psi_1(v) = \eta_{\gamma_1}^{1+\gamma_1}(v)k_F(\eta_{\gamma_1}^{-1}(v))$  et  $\psi_2(v) = \eta_{\gamma_2}^{1+\gamma_2}(v)k_G(\eta_{\gamma_2}^{-1}(v))$ . Nous référons le lecteur à Beirlant *et al.* [6] (p. 93-94) pour plus de détails.

Avant d'énoncer notre résultat principal, nous avons encore besoin d'introduire les principaux termes de biais de nos estimateurs donnés par

$$B_{1,t}(\gamma_1, \sigma_{1,t}) := \begin{cases} \frac{1}{\gamma_1 \sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}(x)}{1 + \frac{\gamma_1}{\sigma_{1,t}}(x-t)} dx - \frac{1}{\gamma_1 \sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}(x)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(x-t))^2} dx \\ \quad - \frac{1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}^1(x)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(x-t))^2} dx & \text{si } \gamma_1 \neq 0 \\ \frac{1}{\sigma_{1,t}^2} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} (x-t) \overline{H}(x) dx - \frac{1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \overline{H}^1(x) dx & \text{si } \gamma_1 = 0 \end{cases}$$

et

$$B_{2,t}(\gamma_1, \sigma_{1,t}) := \begin{cases} -\frac{\overline{H}^1(t)}{\overline{H}(t)} + \frac{1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}(x)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(x-t))^2} dx \\ \quad + \frac{\gamma_1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}^1(x)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(x-t))^2} dx & \text{si } \gamma_1 \neq 0 \\ -\frac{\overline{H}^1(t)}{\overline{H}(t)} + \frac{1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \overline{H}(x) dx & \text{si } \gamma_1 = 0 \end{cases}$$

avec  $\overline{H}^1(t) = \mathbb{P}(Z > t, \delta = 1) = \mathbb{P}(t < X \leq Y)$ .

Nous sommes maintenant à même d'énoncer le résultat de normalité.

**Théorème 3.1** *Supposons que  $U_F$  et  $U_G$  soient continues et satisfassent les conditions du premier et second ordre (3.3)-(3.6). Avec  $N_t$  le nombre d'excès au-delà de  $t$ , sous l'hypothèse que la fonction à variation lente associée à  $F$  est normalisée et que, lorsque  $t \rightarrow \tau_H$ , les termes de biais sont tels que*

$$\sqrt{N_t} B_{1,t}(\gamma_1, \sigma_{1,t}) \xrightarrow{\mathbb{P}} 0 \quad (3.9)$$

$$\sqrt{N_t} B_{2,t}(\gamma_1, \sigma_{1,t}) \xrightarrow{\mathbb{P}} 0 \quad (3.10)$$

$$\text{et } \frac{\sigma_{2,t}}{\sigma_{1,t} + \sigma_{2,t}} \longrightarrow C \in (0, 1) \quad \text{si } \gamma_1 = \gamma_2 = 0 \text{ et } \tau_F = \tau_G, \quad (3.11)$$

alors, pour  $\gamma > -\frac{1}{2}$ , nous avons le comportement asymptotique suivant :

$$\sqrt{N_t} \begin{pmatrix} \widehat{\gamma}_{Z,t}^{(c,OS)} - \gamma_1 \\ \frac{\widehat{\sigma}_{Z,t}^{(c,OS)}}{\sigma_{1,t}} - 1 \end{pmatrix} \xrightarrow{d} \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right), \text{ quand } t \rightarrow \tau_H \quad (3.12)$$

avec

$$\Sigma := \begin{cases} \begin{pmatrix} \left( \frac{\gamma_1}{\gamma} \right)^3 (1 + \gamma)^2 & - \left( \frac{\gamma_1}{\gamma} \right)^2 (1 + \gamma) \\ - \left( \frac{\gamma_1}{\gamma} \right)^2 (1 + \gamma) & 2 \frac{\gamma_1}{\gamma} (1 + \gamma) \end{pmatrix} & \text{si } \gamma_1 \gamma_2 > 0 \text{ et } \tau_F = \tau_G, \\ \begin{pmatrix} C^{-3} & -C^{-2} \\ -C^{-2} & 2C^{-1} \end{pmatrix} & \text{si } \gamma_1 = \gamma_2 = 0 \text{ et } \tau_F = \tau_G. \end{cases}$$

Nous allons, dans la section suivante, étudier plus en détail les hypothèses du Théorème 3.1 sur quelques exemples de lois.

### 3.3 Exemples et résultats de simulations

Dans cette section, nous allons, dans un premier temps, considérer les trois exemples suivants :

- une Burr censurée par une Burr ;
- une ReverseBurr censurée par une ReverseBurr ;
- une Logistique censurée par une Logistique.

Notons que ces trois exemples couvrent les trois cas possibles considérés et permettent de montrer ce que peuvent représenter les hypothèses (3.9), (3.10) et (3.11) du Théorème 3.1. Dans chacun des cas, nous allons calculer explicitement les principaux termes de

biais  $B_{1,t}(\gamma_1, \sigma_{1,t})$  et  $B_{2,t}(\gamma_1, \sigma_{1,t})$ , ce qui nous amènera à trouver pour ces exemples des conditions suffisantes pour que les hypothèses (3.9) et (3.10) du Théorème 3.1 soient satisfaites. Dans l'exemple du cas 3, nous montrons que la condition (3.11) est également satisfaite.

Enfin, nous allons illustrer, par simulations, le comportement de nos estimateurs « one-step » par rapport à ceux du maximum de vraisemblance, résultant de la maximisation de la vraisemblance (3.2).

### 3.3.1 Différents exemples

• **Exemple 1** :  $X \sim \text{Burr}(\beta_1, \tau_1, \lambda_1)$  et  $Y \sim \text{Burr}(\beta_2, \tau_2, \lambda_2)$ ,  $\beta_i, \tau_i, \lambda_i > 0$ ,  $i = 1, 2$ , et  $\tau_1 \neq 1$ .

Dans ce cas, nous avons

$$\begin{aligned} 1 - F(x) &= \left( \frac{\beta_1}{\beta_1 + x^{\tau_1}} \right)^{\lambda_1} = x^{-\tau_1 \lambda_1} \beta_1^{\lambda_1} (1 + \beta_1 x^{-\tau_1})^{-\lambda_1}, x > 0 \\ 1 - G(x) &= \left( \frac{\beta_2}{\beta_2 + x^{\tau_2}} \right)^{\lambda_2} = x^{-\tau_2 \lambda_2} \beta_2^{\lambda_2} (1 + \beta_2 x^{-\tau_2})^{-\lambda_2}, x > 0 \end{aligned}$$

et alors  $\gamma_1 = \frac{1}{\lambda_1 \tau_1}$ ,  $\gamma = \frac{1}{\lambda_1 \tau_1 + \lambda_2 \tau_2}$  et  $\sigma_{1,t} = \gamma_1 t \left( 1 + \frac{\beta_1}{\tau_1} t^{-\tau_1} (1 + o(1)) \right)$ .

Des calculs assez laborieux nous conduisent aux principaux termes de biais suivants :

$$\begin{aligned} B_{1,t}(\gamma_1, \sigma_{1,t}) &= \begin{cases} -\frac{\gamma^2}{\gamma_1^2} \frac{\beta_1(1-\tau_1)(1+\gamma-\gamma^2\tau_1)}{\tau_1(1+\gamma\tau_1)(1+\gamma\tau_1+\gamma)(1+\gamma)(1+2\gamma)} t^{-\tau_1} & \text{si } \tau_1 \leq \tau_2 \\ o(t^{-\tau_2}) & \text{si } \tau_1 > \tau_2 \end{cases} \\ B_{2,t}(\gamma_1, \sigma_{1,t}) &= \begin{cases} -\frac{\gamma}{\gamma_1} \frac{\beta_1(1-\tau_1)(1+\gamma)}{\tau_1(1+\gamma\tau_1+\gamma)(1+2\gamma)} t^{-\tau_1} & \text{si } \tau_1 \leq \tau_2 \\ o(t^{-\tau_2}) & \text{si } \tau_1 > \tau_2. \end{cases} \end{aligned}$$

Il en résulte qu'une condition suffisante pour (3.9) et (3.10) est :

$$\sqrt{n\overline{H}(t)} t^{-\tau} \longrightarrow 0 \quad \text{ou de façon équivalente} \quad \sqrt{n} t^{-\tau - \frac{1}{2\gamma}} \longrightarrow 0, \quad \text{quand } t \rightarrow \infty,$$

où  $\tau = \min(\tau_1, \tau_2)$ .

• **Exemple 2** :  $X \sim \text{ReverseBurr}(\beta_1, \tau_1, \lambda_1, \tau_H)$  et  $Y \sim \text{ReverseBurr}(\beta_2, \tau_2, \lambda_2, \tau_H)$ ,  
 $\beta_i, \tau_i, \lambda_i, \tau_H > 0, i = 1, 2$ .

Dans ce cas

$$1 - F(x) = \left( \frac{\beta_1}{\beta_1 + (\tau_H - x)^{-\tau_1}} \right)^{\lambda_1} = (\tau_H - x)^{\tau_1 \lambda_1} \beta_1^{\lambda_1} \left( 1 + \beta_1 (\tau_H - x)^{\tau_1} \right)^{-\lambda_1}, \quad x < \tau_H$$

$$1 - G(x) = \left( \frac{\beta_2}{\beta_2 + (\tau_H - x)^{-\tau_2}} \right)^{\lambda_2} = (\tau_H - x)^{\tau_2 \lambda_2} \beta_2^{\lambda_2} \left( 1 + \beta_2 (\tau_H - x)^{\tau_2} \right)^{-\lambda_2}, \quad x < \tau_H$$

d'où  $\gamma_1 = -\frac{1}{\tau_1 \lambda_1}, \gamma = -\frac{1}{\tau_1 \lambda_1 + \tau_2 \lambda_2}$  et  $\sigma_{1,t} = -\gamma_1 (\tau_H - t) \left( 1 - \frac{\beta_1}{\tau_1} (\tau_H - t)^{\tau_1} (1 + o(1)) \right)$ .

Ceci nous conduit aux nouveaux termes principaux de biais suivants :

$$B_{1,t}(\gamma_1, \sigma_{1,t}) = \begin{cases} \frac{\gamma^2 \beta_1 (1 + \tau_1) (1 + \gamma + \gamma^2 \tau_1)}{\gamma_1^2 \tau_1 (1 - \gamma \tau_1) (1 - \gamma \tau_1 + \gamma) (1 + \gamma) (1 + 2\gamma)} (\tau_H - t)^{\tau_1} & \text{si } \tau_1 \leq \tau_2 \\ o((\tau_H - t)^{\tau_2}) & \text{si } \tau_1 > \tau_2 \end{cases}$$

$$B_{2,t}(\gamma_1, \sigma_{1,t}) = \begin{cases} \frac{\gamma \beta_1 (1 + \tau_1) (1 + \gamma)}{\gamma_1 \tau_1 (1 - \gamma \tau_1 + \gamma) (1 + 2\gamma)} (\tau_H - t)^{\tau_1} & \text{si } \tau_1 \leq \tau_2 \\ o((\tau_H - t)^{\tau_2}) & \text{si } \tau_1 > \tau_2. \end{cases}$$

De façon similaire à l'exemple précédent, nous en déduisons qu'une condition suffisante pour (3.9) et (3.10) est

$$\sqrt{n\overline{H}(t)} (\tau_H - t)^\tau \longrightarrow 0 \quad \text{ou de façon équivalente} \quad \sqrt{n} (\tau_H - t)^{\tau - \frac{1}{2\gamma}} \longrightarrow 0, \quad \text{quand } t \rightarrow \tau_H,$$

où  $\tau = \min(\tau_1, \tau_2)$ .

• **Exemple 3** :  $X, Y \sim \text{Logistique}$ .

Dans ce cas

$$1 - F(x) = 1 - G(x) = \frac{2}{1 + e^x}, \quad x > 0$$

avec  $\gamma_1 = \gamma = 0$  et  $\sigma_{1,t} = 1$ . Cela implique que

$$B_{1,t}(\gamma_1, \sigma_{1,t}) = \frac{1}{9} e^{-t} (1 + o(1))$$

et

$$B_{2,t}(\gamma_1, \sigma_{1,t}) = \frac{1}{3} e^{-t} (1 + o(1)).$$

Cette fois-ci nous pouvons en déduire une condition nécessaire et suffisante pour (3.9) et (3.10), qui est

$$\sqrt{n\bar{H}(t)} e^{-t} \longrightarrow 0 \quad \text{ou de façon équivalente} \quad \sqrt{n} e^{-2t} \longrightarrow 0, \quad \text{quand } t \rightarrow \infty.$$

Comme  $\sigma_{1,t} = \sigma_{2,t} = 1$ , (3.11) est clairement satisfaite.

### 3.3.2 Simulations

Afin d'illustrer ces trois exemples, nous avons simulé 300 échantillons de taille 500 issus des distributions suivantes :

- une Burr (10, 4, 1) censurée par une Burr (10, 1, 0.5) :  $\gamma_1 = 0.25, \gamma_2 = 2$ ;
- une ReverseBurr (1, 8, 0.5, 10) censurée par une ReverseBurr (10, 1, 2.5, 10) :  $\gamma_1 = -0.25, \gamma_2 = -0.4$ ;
- une Logistique censurée par une autre Logistique :  $\gamma_1 = \gamma_2 = 0$ .

Ces trois choix spécifiques de paramètres correspondent à trois taux différents de censure dans la queue : 11% pour l'exemple de la Burr, 38% pour celui de la ReverseBurr et enfin 50% pour la Logistique.

Dans ces simulations, les estimateurs initiaux utilisés dans la construction des estimateurs « one-step » sont les estimateurs des moments adaptés à la censure introduits par Beirlant

et al. [7] et définis ci-dessous :

$$\hat{\gamma}_{Z,t}^{(c,M)} := \frac{\bar{H}_n(t)}{\bar{H}_n^1(t)} \left( \hat{\gamma}_{Z,t}^{(H)} + 1 - \frac{1}{2} \left( 1 - \frac{(\hat{\gamma}_{Z,t}^{(H)})^2}{S_{Z,t}} \right)^{-1} \right)$$

et

$$\hat{\sigma}_{Z,t}^{(c,M)} = \frac{\bar{H}_n(t)}{\bar{H}_n^1(t)} \frac{t \sqrt{3(\hat{\gamma}_{Z,t}^{(H)})^2 - S_{Z,t}}}{\sqrt{3[\rho_1(\hat{\gamma}_{Z,t}^{(M)})]^2 - \rho_2(\hat{\gamma}_{Z,t}^{(M)})}},$$

où  $\hat{\gamma}_{Z,t}^{(H)}$ ,  $\hat{\gamma}_{Z,t}^{(M)}$  et  $S_{Z,t}$  ne sont rien d'autres que  $\hat{\gamma}_{Z,k,n}^{(H)}$ ,  $\hat{\gamma}_{Z,k,n}^{(M)}$  et  $S_{Z,k,n}$ , définis dans le chapitre d'introduction, quand le seuil  $t$  est fixé égal à  $Z_{n-k,n}$ . Quant aux quantités  $\bar{H}_n(t)$  et  $\bar{H}_n^1(t)$ , il s'agit des versions empiriques de  $\bar{H}(t)$  et  $\bar{H}^1(t)$  respectivement.

Nous illustrons dans la Figure 3.1 le comportement (en terme de médiane et d'erreur en moyenne quadratique (MSE)) des estimateurs de  $\gamma_1$  et dans la Figure 3.2 ceux du paramètre d'échelle renormalisé. Les deux estimateurs adaptés à la censure considérés sont les estimateurs « one-step » (en pointillé) et ceux du maximum de vraisemblance (trait plein), obtenus en maximisant (3.2). La droite horizontale représente toujours la vraie valeur du paramètre. Ces graphes correspondent successivement aux trois distributions considérées : Burr (haut), ReverseBurr (milieu) et Logistique (bas). Nous observons que les estimateurs « one-step » ont un comportement similaire à celui des estimateurs par maximum de vraisemblance pour lesquels nous rappelons que nous ne disposons pas du comportement asymptotique. En particulier, si nous regardons les résultats pour la distribution de Burr(10, 4, 1) censurée par une distribution de Burr(10, 1, 0.5), il est même très difficile de différencier les deux courbes. En terme de MSE, nous pouvons dire que les deux minima de courbes sont proches l'un de l'autre, sauf dans le cas de la Logistique pour le paramètre d'échelle renormalisé où il y a un écart significatif (également en terme de médiane), le « one-step » étant considérablement meilleur. Ceci dit nous avons pu constater que cela était dû au taux de censure élevé (50%) puisque si nous traçons ces estimateurs dans le cas standard où il n'y a pas de censure, nous retrouvons alors une très grande proximité entre les deux courbes, comme cela apparaît sur la Figure 3.3.

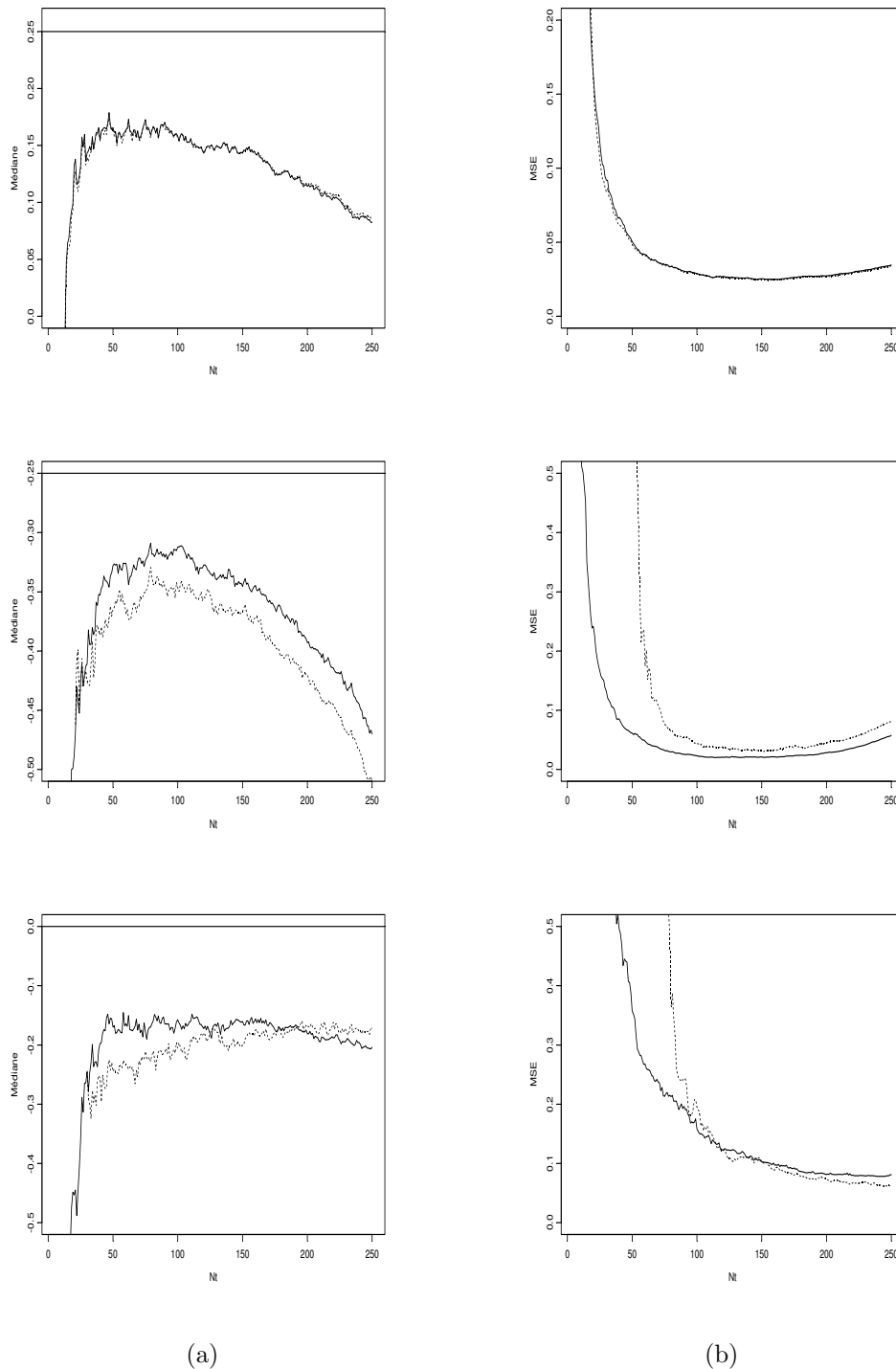


FIG. 3.1 – (a) Médiane pour  $\hat{\gamma}_{Z,t}^{(c, \cdot)}$  et (b) l'erreur en moyenne quadratique empirique pour une Burr (10, 4, 1) censurée par une Burr (10, 1, 0.5) (en haut), pour une ReverseBurr (1, 8, 0.5, 10) censurée par une ReverseBurr (10, 1, 2.5, 10) (au milieu) et pour une Logistique censurée par une autre Logistique (en bas). La courbe en trait plein est l'estimateur du maximum de vraisemblance et celle en pointillé est l'estimateur « one-step ». La droite horizontale représente  $\gamma_1$ .



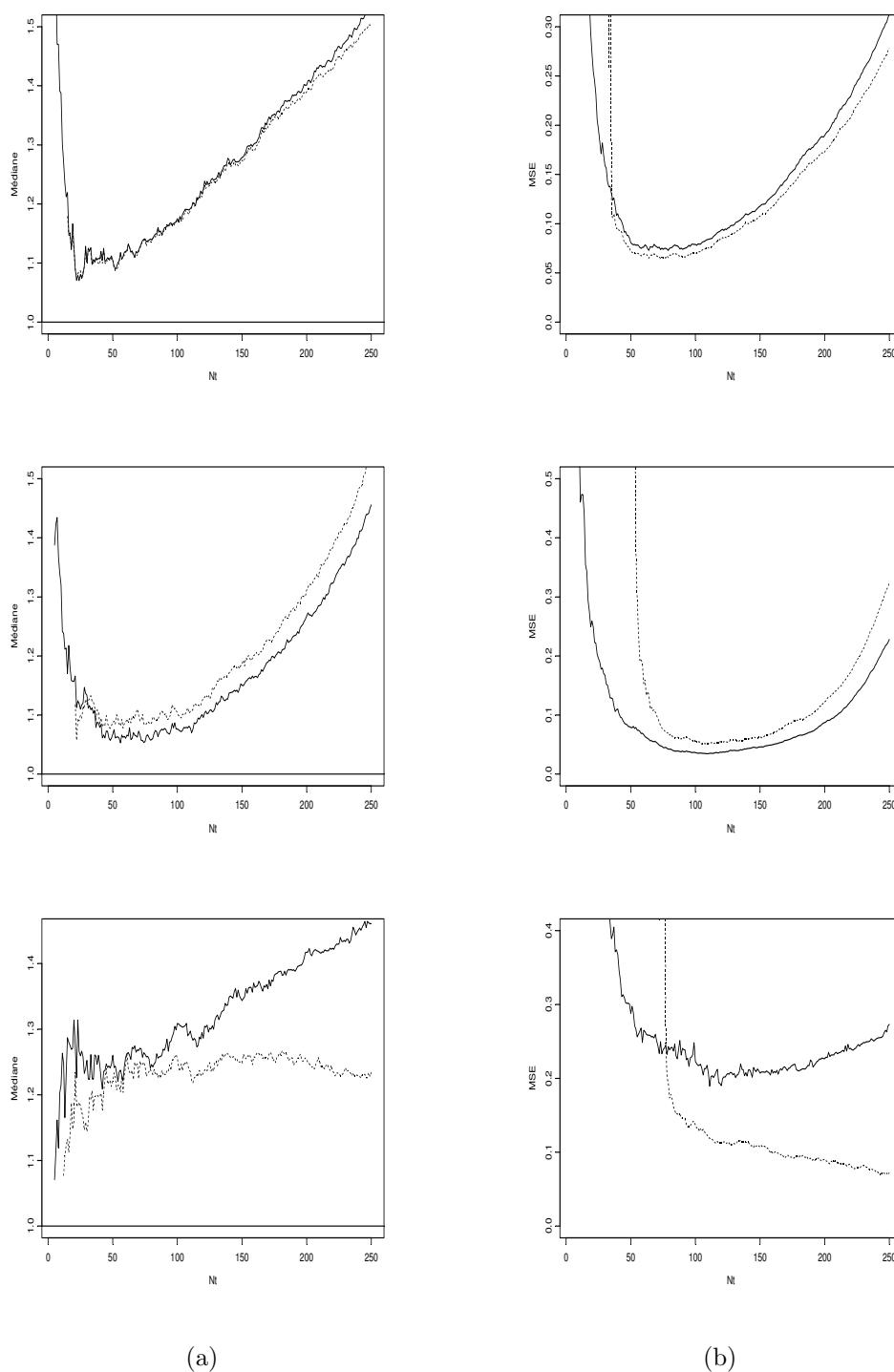


FIG. 3.2 – (a) Médiane pour  $\hat{\sigma}_{Z,t}^{(c,.)}$  et (b) l'erreur en moyenne quadratique empirique pour une Burr (10, 4, 1) censurée par une Burr (10, 1, 0.5) (en haut), pour une ReverseBurr (1, 8, 0.5, 10) censurée par une ReverseBurr (10, 1, 2.5, 10) (au milieu) et pour une Logistique censurée par une autre Logistique (en bas). La courbe en trait plein est l'estimateur du maximum de vraisemblance et celle en pointillé est l'estimateur « one-step ».

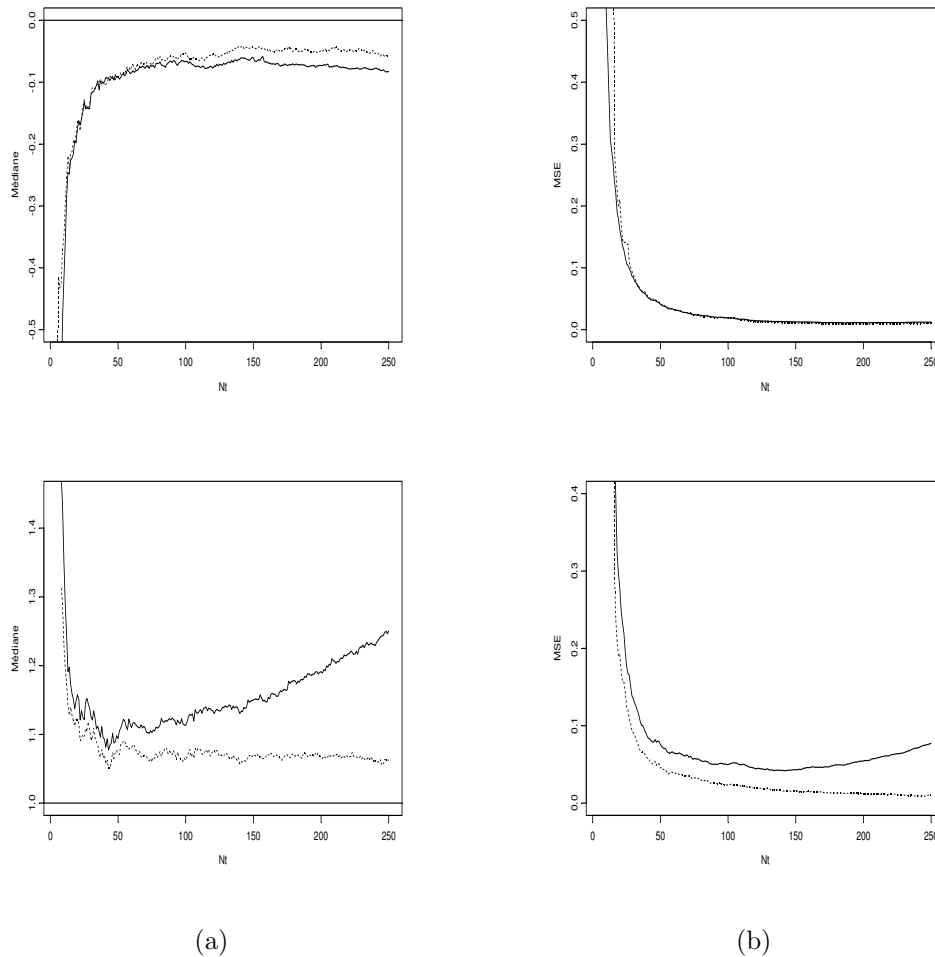


FIG. 3.3 – Cas sans censure : en haut : (a) médiane pour les estimateurs de  $\gamma_1$  et (b) l'erreur en moyenne quadratique empirique associée et en bas : (a) médiane pour les estimateurs du paramètre d'échelle renormalisé et (b) l'erreur en moyenne quadratique empirique associée pour une distribution Logistique. La courbe en trait plein est l'estimateur du maximum de vraisemblance et celle en pointillé est l'estimateur « one-step ». La ligne horizontale correspond à la vraie valeur du paramètre estimé.



$$\begin{aligned}
& \sqrt{N_t} \begin{pmatrix} \widehat{\gamma}_{Z,t}^{(c,OS)} - \widehat{\gamma}_{Z,t}^{(c,M)} \\ \frac{\widehat{\sigma}_{Z,t}^{(c,OS)}}{\sigma_{1,t}} - \frac{\widehat{\sigma}_{Z,t}^{(c,M)}}{\sigma_{1,t}} \end{pmatrix} \\
&= \frac{1}{\left(\frac{1}{N_t} L''_{11}\right) \left(\frac{\sigma_{1,t}^2}{N_t} L''_{22}\right) - \left(\frac{\sigma_{1,t}}{N_t} L''_{12}\right)^2} \begin{pmatrix} \frac{\sigma_{1,t}^2}{N_t} L''_{22} & -\frac{\sigma_{1,t}}{N_t} L''_{12} \\ -\frac{\sigma_{1,t}}{N_t} L''_{12} & \frac{1}{N_t} L''_{11} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{N_t}} L'_1 \\ -\frac{\sigma_{1,t}}{\sqrt{N_t}} L'_2 \end{pmatrix} \quad (3.14)
\end{aligned}$$

où, pour alléger les notations, nous n'avons pas spécifié que toutes les dérivées de (3.14) étaient évaluées en  $(\widehat{\gamma}_{Z,t}^{(c,M)}, \widehat{\sigma}_{Z,t}^{(c,M)})$ .

**Étape 3 :** En vertu de (3.14), des calculs directs nous donnent :

$$\begin{aligned}
-\frac{1}{N_t} L''_{11}(\gamma_1, \sigma_{1,t}) &\xrightarrow{\mathbb{P}} \begin{cases} 2 \left(\frac{\gamma}{\gamma_1}\right)^3 \frac{1}{(1+\gamma)(1+2\gamma)} & \text{si } \gamma_1 \neq 0 \\ 2C^3 & \text{si } \gamma_1 = 0 \end{cases} \\
-\frac{\sigma_{1,t}^2}{N_t} L''_{22}(\gamma_1, \sigma_{1,t}) &\xrightarrow{\mathbb{P}} \begin{cases} \frac{\gamma}{\gamma_1} \frac{1}{1+2\gamma} & \text{si } \gamma_1 \neq 0 \\ C & \text{si } \gamma_1 = 0 \end{cases} \\
\text{et } -\frac{\sigma_{1,t}}{N_t} L''_{12}(\gamma_1, \sigma_{1,t}) &\xrightarrow{\mathbb{P}} \begin{cases} \left(\frac{\gamma}{\gamma_1}\right)^2 \frac{1}{(1+\gamma)(1+2\gamma)} & \text{si } \gamma_1 \neq 0 \\ C^2 & \text{si } \gamma_1 = 0. \end{cases}
\end{aligned}$$

Cependant il faut être prudent car les convergences précédentes ne sont valides que dans le cas où les dérivées sont évaluées en les vraies valeurs des paramètres  $(\gamma_1, \sigma_{1,t})$ . Or nous avons besoin de ces mêmes convergences lorsque ces dérivées secondes sont évaluées en  $(\widehat{\gamma}_{Z,t}^{(c,M)}, \widehat{\sigma}_{Z,t}^{(c,M)})$ . Toutefois il est assez facile de les déduire en observant que  $\frac{\gamma_1 E_j / \sigma_{1,t}}{1 + \frac{\gamma_1}{\sigma_{1,t}} E_j} \leq 1$ .

À titre d'exemple, nous avons

$$\begin{aligned}
& \frac{1}{N_t} \frac{1}{\widehat{\sigma}_{Z,t}^{(c,M)}} \sum_{j=1}^{N_t} \frac{E_j}{1 + \frac{\widehat{\gamma}_{Z,t}^{(c,M)}}{\widehat{\sigma}_{Z,t}^{(c,M)}} E_j} - \frac{1}{N_t} \frac{1}{\sigma_{1,t}} \sum_{j=1}^{N_t} \frac{E_j}{1 + \frac{\gamma_1}{\sigma_{1,t}} E_j} \\
&= \frac{1}{N_t} \frac{1}{\widehat{\sigma}_{Z,t}^{(c,M)}} \sum_{j=1}^{N_t} \left( \frac{E_j}{1 + \frac{\widehat{\gamma}_{Z,t}^{(c,M)}}{\widehat{\sigma}_{Z,t}^{(c,M)}} E_j} - \frac{E_j}{1 + \frac{\gamma_1}{\sigma_{1,t}} E_j} \right) + \frac{1}{N_t} \left( \frac{1}{\widehat{\sigma}_{Z,t}^{(c,M)}} - \frac{1}{\sigma_{1,t}} \right) \sum_{j=1}^{N_t} \frac{E_j}{1 + \frac{\gamma_1}{\sigma_{1,t}} E_j} \\
&= -\frac{1}{\widehat{\gamma}_{Z,t}^{(c,M)}} \left( \frac{\widehat{\tau}_{1,t}}{\tau_{1,t}} - 1 \right) \frac{\tau_{1,t}}{\widehat{\tau}_{1,t}} \frac{\widehat{\tau}_{1,t}}{\widehat{\tau}_{1,t}} \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \frac{\widehat{\tau}_{1,t} E_j}{1 + \widehat{\tau}_{1,t} E_j} \right)^2 + \frac{1}{\gamma_1} \left( \frac{\sigma_{1,t}}{\widehat{\sigma}_{Z,t}^{(c,M)}} - 1 \right) \frac{1}{N_t} \sum_{j=1}^{N_t} \frac{\tau_{1,t} E_j}{1 + \tau_{1,t} E_j} \\
&\xrightarrow{\mathbb{P}} 0,
\end{aligned}$$

où  $\tau_{1,t} = \frac{\gamma_1}{\sigma_{1,t}}$ ,  $\widehat{\tau}_{1,t} = \frac{\widehat{\gamma}_{Z,t}^{(c,M)}}{\widehat{\sigma}_{Z,t}^{(c,M)}}$  et  $\widehat{\tau}_{1,t}$  est compris entre  $\tau_{1,t}$  et  $\widehat{\tau}_{1,t}$ .

**Étape 4 :** En développant le membre de droite de (3.13) autour de  $(\gamma_1, \sigma_{1,t})$  (comme dans la preuve du Théorème 4.2 de Lehmann [50]) et compte tenu de (3.9) et (3.10), nous devons maintenant établir la normalité asymptotique de

$$V := \begin{pmatrix} -\sqrt{N_t} \left[ \frac{1}{N_t} L'_1(\gamma_1, \sigma_{1,t}) - B_{1,t}(\gamma_1, \sigma_{1,t}) \right] \\ -\sqrt{N_t} \left[ \frac{\sigma_{1,t}}{N_t} L'_2(\gamma_1, \sigma_{1,t}) - B_{2,t}(\gamma_1, \sigma_{1,t}) \right] \end{pmatrix}.$$

Celle-ci, comme nous le démontrons en Annexe 2 en utilisant l'expression des dérivées premières de l'Annexe 1, va se déduire de la normalité asymptotique de

$$W := \sqrt{\frac{n}{\overline{H}(t)}} \begin{pmatrix} \overline{H}_n(t) - \overline{H}(t) \\ \overline{H}_n^1(t) - \overline{H}^1(t) \\ \frac{1}{n} \sum_{j=1}^n \log \left( 1 + \frac{\gamma_1}{\sigma_{1,t}} (Z_j - t) \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}(x)}{1 + \frac{\gamma_1}{\sigma_{1,t}} (x - t)} dx \\ \frac{1}{n} \sum_{j=1}^n \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}} (Z_j - t)} \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}(x)}{(1 + \frac{\gamma_1}{\sigma_{1,t}} (x - t))^2} dx \\ \frac{1}{n} \sum_{j=1}^n \delta_j \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}} (Z_j - t)} \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}^1(x)}{(1 + \frac{\gamma_1}{\sigma_{1,t}} (x - t))^2} dx \end{pmatrix}$$

dans le cas où  $\gamma_1 \neq 0$  et de

$$W := \sqrt{\frac{n}{\overline{H}(t)}} \begin{pmatrix} \overline{H}_n(t) - \overline{H}(t) \\ \overline{H}_n^1(t) - \overline{H}^1(t) \\ \frac{1}{\sigma_{1,t}} \left[ \frac{1}{n} \sum_{j=1}^n (Z_j - t) \mathbb{1}_{\{Z_j > t\}} - \int_t^{\tau_H} \overline{H}(x) dx \right] \\ \frac{1}{\sigma_{1,t}^2} \left[ \frac{1}{2n} \sum_{j=1}^n (Z_j - t)^2 \mathbb{1}_{\{Z_j > t\}} - \int_t^{\tau_H} (x - t) \overline{H}(x) dx \right] \\ \frac{1}{\sigma_{1,t}} \left[ \frac{1}{n} \sum_{j=1}^n \delta_j (Z_j - t) \mathbb{1}_{\{Z_j > t\}} - \int_t^{\tau_H} \overline{H}^1(x) dx \right] \end{pmatrix}$$

sinon.

**Étape 5 :** Nous déduisons du théorème central limite multivarié (les couples  $(Z_i, \delta_i)$ ,  $i = 1, \dots, n$  étant indépendants) la matrice de variance-covariance asymptotique pour  $W$  dans le cas  $\gamma_1 \neq 0$ , à savoir :

$$\begin{pmatrix} 1 & \frac{\gamma}{\gamma_1} & \gamma & \frac{\gamma_1}{\gamma} a & a \\ \frac{\gamma}{\gamma_1} & \frac{\gamma}{\gamma_1} & \frac{\gamma^2}{\gamma_1} & a & a \\ \gamma & \frac{\gamma^2}{\gamma_1} & 2\gamma^2 & b & \frac{\gamma}{\gamma_1} b \\ \frac{\gamma_1}{\gamma} a & a & b & c & \frac{\gamma}{\gamma_1} c \\ a & a & \frac{\gamma}{\gamma_1} b & \frac{\gamma}{\gamma_1} c & \frac{\gamma}{\gamma_1} c \end{pmatrix},$$

où  $a = \gamma \gamma_1^{-1} (1 - (1 + \gamma)^{-1})$ ,  $b = \gamma (1 - (1 + \gamma)^{-2})$ , et  $c = 1 - 2(1 + \gamma)^{-1} + (1 + 2\gamma)^{-1}$ .

Dans le cas d'un indice  $\gamma_1 = 0$ , en utilisant l'hypothèse (3.11), nous obtenons la matrice

de variance-covariance asymptotique suivante pour  $W$  :

$$\begin{pmatrix} 1 & C & C & C^2 & C^2 \\ C & C & C^2 & C^3 & C^2 \\ C & C^2 & 2C^2 & 3C^3 & 2C^3 \\ C^2 & C^3 & 3C^3 & 6C^4 & 3C^4 \\ C^2 & C^2 & 2C^3 & 3C^4 & 2C^3 \end{pmatrix}.$$

Il est important de signaler que le calcul de ces matrices est assez complexe. Il nécessite de décomposer chacune des intégrales en plusieurs morceaux, et d'utiliser les résultats de convergence (3.7) et (3.8) combinés avec le théorème de convergence dominée afin de pouvoir conclure. Pour illustrer ces techniques, nous avons donné, dans l'Annexe 3, deux exemples d'intégrale, l'une dans le cas d'un indice négatif, l'autre dans le cas d'un indice nul, le cas d'un indice positif étant un peu plus facile à appréhender.

**Dernière Étape :** De l'étape précédente et des résultats de l'Annexe 2, nous déduisons par des calculs directs que la matrice de variance-covariance asymptotique de  $V$  dans le cas  $\gamma_1 \neq 0$  est :

$$\begin{pmatrix} 2 \left( \frac{\gamma}{\gamma_1} \right)^3 \frac{1}{(1+\gamma)(1+2\gamma)} & \left( \frac{\gamma}{\gamma_1} \right)^2 \frac{1}{(1+\gamma)(1+2\gamma)} \\ \left( \frac{\gamma}{\gamma_1} \right)^2 \frac{1}{(1+\gamma)(1+2\gamma)} & \frac{\gamma}{\gamma_1} \frac{1}{1+2\gamma} \end{pmatrix}$$

tandis qu'elle vaut, dans le cas  $\gamma_1 = 0$ , la quantité suivante :

$$\begin{pmatrix} 2C^3 & C^2 \\ C^2 & C \end{pmatrix}.$$

## 3.5 Annexes

### 3.5.1 Annexe 1 : Calcul des dérivées premières et secondes

Des calculs directs conduisent aux dérivées premières suivantes :

$$\left\{ \begin{array}{l} L'_1(\gamma_1, \sigma_{1,t}) := \frac{\partial \log L(\gamma_1, \sigma_{1,t})}{\partial \gamma_1} = \frac{1}{\gamma_1^2} \sum_{j=1}^{N_t} \log\left(1 + \frac{\gamma_1}{\sigma_{1,t}} E_j\right) - \frac{1}{\gamma_1} \sum_{j=1}^{N_t} \left(\frac{1}{\gamma_1} + \delta_j\right) \frac{\gamma_1 E_j / \sigma_{1,t}}{1 + \frac{\gamma_1}{\sigma_{1,t}} E_j} \\ L'_2(\gamma_1, \sigma_{1,t}) := \frac{\partial \log L(\gamma_1, \sigma_{1,t})}{\partial \sigma_{1,t}} = -\frac{1}{\sigma_{1,t}} \sum_{j=1}^{N_t} \delta_j + \frac{1}{\sigma_{1,t}} \sum_{j=1}^{N_t} \left(\frac{1}{\gamma_1} + \delta_j\right) \frac{\gamma_1 E_j / \sigma_{1,t}}{1 + \frac{\gamma_1}{\sigma_{1,t}} E_j} \end{array} \right.$$

et aux dérivées secondes :

$$\left\{ \begin{array}{l} L''_{11}(\gamma_1, \sigma_{1,t}) := \frac{\partial^2 \log L(\gamma_1, \sigma_{1,t})}{\partial^2 \gamma_1} \\ = -\frac{2}{\gamma_1^3} \sum_{j=1}^{N_t} \log\left(1 + \frac{\gamma_1}{\sigma_{1,t}} E_j\right) + \frac{2}{\gamma_1^3} \sum_{j=1}^{N_t} \frac{\gamma_1 E_j / \sigma_{1,t}}{1 + \frac{\gamma_1}{\sigma_{1,t}} E_j} + \frac{1}{\gamma_1^2} \sum_{j=1}^{N_t} \left(\frac{1}{\gamma_1} + \delta_j\right) \frac{(\gamma_1 E_j / \sigma_{1,t})^2}{(1 + \frac{\gamma_1}{\sigma_{1,t}} E_j)^2} \\ L''_{22}(\gamma_1, \sigma_{1,t}) := \frac{\partial^2 \log L(\gamma_1, \sigma_{1,t})}{\partial^2 \sigma_{1,t}} \\ = \frac{1}{\sigma_{1,t}^2} \sum_{j=1}^{N_t} \delta_j - \frac{2}{\sigma_{1,t}^2} \sum_{j=1}^{N_t} \left(\frac{1}{\gamma_1} + \delta_j\right) \frac{\gamma_1 E_j / \sigma_{1,t}}{1 + \frac{\gamma_1}{\sigma_{1,t}} E_j} + \frac{1}{\sigma_{1,t}^2} \sum_{j=1}^{N_t} \left(\frac{1}{\gamma_1} + \delta_j\right) \frac{(\gamma_1 E_j / \sigma_{1,t})^2}{(1 + \frac{\gamma_1}{\sigma_{1,t}} E_j)^2} \\ L''_{12}(\gamma_1, \sigma_{1,t}) := \frac{\partial^2 \log L(\gamma_1, \sigma_{1,t})}{\partial \sigma_{1,t} \partial \gamma_1} \\ = -\frac{1}{\gamma_1} \frac{1}{\sigma_{1,t}} \sum_{j=1}^{N_t} \left(\frac{1}{\gamma_1} + \delta_j\right) \frac{(\gamma_1 E_j / \sigma_{1,t})^2}{(1 + \frac{\gamma_1}{\sigma_{1,t}} E_j)^2} + \frac{1}{\gamma_1} \frac{1}{\sigma_{1,t}} \sum_{j=1}^{N_t} \delta_j \frac{\gamma_1 E_j / \sigma_{1,t}}{1 + \frac{\gamma_1}{\sigma_{1,t}} E_j}. \end{array} \right.$$

Dans le cas où  $\gamma_1 = 0$ , ces dérivées s'écrivent :

$$\left\{ \begin{array}{l} L'_1(0, \sigma_{1,t}) := \frac{1}{2} \sum_{j=1}^{N_t} \frac{E_j^2}{\sigma_{1,t}^2} - \sum_{j=1}^{N_t} \delta_j \frac{E_j}{\sigma_{1,t}} \\ L'_2(0, \sigma_{1,t}) := -\frac{1}{\sigma_{1,t}} \sum_{j=1}^{N_t} \delta_j + \frac{1}{\sigma_{1,t}^2} \sum_{j=1}^{N_t} E_j \end{array} \right.$$



et

$$\left\{ \begin{array}{l} L''_{11}(0, \sigma_{1,t}) := -\frac{2}{3} \sum_{j=1}^{N_t} \left( \frac{E_j}{\sigma_{1,t}} \right)^3 + \sum_{j=1}^{N_t} \delta_j \left( \frac{E_j}{\sigma_{1,t}} \right)^2 \\ L''_{22}(0, \sigma_{1,t}) := \frac{1}{\sigma_{1,t}^2} \sum_{j=1}^{N_t} \delta_j - \frac{2}{\sigma_{1,t}^3} \sum_{j=1}^{N_t} E_j \\ L''_{12}(0, \sigma_{1,t}) := -\frac{1}{\sigma_{1,t}} \sum_{j=1}^{N_t} \left( \frac{E_j}{\sigma_{1,t}} \right)^2 + \frac{1}{\sigma_{1,t}} \sum_{j=1}^{N_t} \delta_j \frac{E_j}{\sigma_{1,t}}. \end{array} \right.$$

### 3.5.2 Annexe 2 : Justification de la matrice $W$

En utilisant l'annexe précédente, il est clair que si  $\gamma_1 \neq 0$ , nous avons

$$\begin{aligned} & -\sqrt{N_t} \left[ \frac{1}{N_t} L'_1(\gamma_1, \sigma_{1,t}) - B_1(\gamma_1, \sigma_{1,t}) \right] \\ &= -\sqrt{\frac{\overline{H}(t)}{\overline{H}_n(t)}} \sqrt{\frac{n}{\overline{H}(t)}} \left\{ \frac{1}{\gamma_1^2} \left( \frac{1}{n} \sum_{j=1}^n \log \left( 1 + \frac{\gamma_1}{\sigma_{1,t}} (Z_j - t) \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}(z)}{1 + \frac{\gamma_1}{\sigma_{1,t}}(z-t)} dz \right) \right. \\ & \quad - \frac{1}{\gamma_1^2} \left( \frac{1}{n} \sum_{j=1}^n \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}}(Z_j - t)} \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}(z)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(z-t))^2} dz \right) \\ & \quad - \frac{1}{\gamma_1} \left( \frac{1}{n} \sum_{j=1}^n \delta_j \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}}(Z_j - t)} \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}^1(z)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(z-t))^2} dz \right) \\ & \quad \left. + [\overline{H}_n(t) - \overline{H}(t)] \left[ -\frac{1}{\gamma_1 \sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}(z)}{1 + \frac{\gamma_1}{\sigma_{1,t}}(z-t)} dz + \frac{1}{\gamma_1 \sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}(z)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(z-t))^2} dz \right. \right. \\ & \quad \left. \left. + \frac{1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}^1(z)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(z-t))^2} dz \right] \right\} \end{aligned}$$

et

$$\begin{aligned}
& -\sqrt{N_t} \left[ \frac{\sigma_{1,t}}{N_t} L'_2(\gamma_1, \sigma_{1,t}) - B_2(\gamma_1, \sigma_{1,t}) \right] \\
&= -\sqrt{\frac{\overline{H}(t)}{\overline{H}_n(t)}} \sqrt{\frac{n}{\overline{H}(t)}} \left\{ -\left(\overline{H}_n^1(t) - \overline{H}^1(t)\right) \right. \\
&\quad + \frac{1}{\gamma_1} \left( \frac{1}{n} \sum_{j=1}^n \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}}(Z_j - t)} \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}(z)}{\left(1 + \frac{\gamma_1}{\sigma_{1,t}}(z - t)\right)^2} dz \right) \\
&\quad + \left( \frac{1}{n} \sum_{j=1}^n \delta_j \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}}(Z_j - t)} \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}^1(z)}{\left(1 + \frac{\gamma_1}{\sigma_{1,t}}(z - t)\right)^2} dz \right) \\
&\quad \left. + [\overline{H}_n(t) - \overline{H}(t)] \left[ \frac{\overline{H}^1(t)}{\overline{H}(t)} - \frac{1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}(z)}{\left(1 + \frac{\gamma_1}{\sigma_{1,t}}(z - t)\right)^2} dz \right. \right. \\
&\quad \quad \left. \left. - \frac{\gamma_1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \frac{\overline{H}^1(z)}{\left(1 + \frac{\gamma_1}{\sigma_{1,t}}(z - t)\right)^2} dz \right] \right\}
\end{aligned}$$

alors que dans le cas où  $\gamma_1 = 0$ , nous avons

$$\begin{aligned}
& -\sqrt{N_t} \left[ \frac{1}{N_t} L'_1(\gamma_1, \sigma_{1,t}) - B_1(\gamma_1, \sigma_{1,t}) \right] \\
&= -\sqrt{\frac{\overline{H}(t)}{\overline{H}_n(t)}} \sqrt{\frac{n}{\overline{H}(t)}} \left\{ \frac{1}{\sigma_{1,t}^2} \left( \frac{1}{2n} \sum_{j=1}^n (Z_j - t)^2 \mathbb{1}_{\{Z_j > t\}} - \int_t^{\tau_H} (x - t) \overline{H}(x) dx \right) \right. \\
&\quad - \frac{1}{\sigma_{1,t}} \left( \frac{1}{n} \sum_{j=1}^n \delta_j (Z_j - t) \mathbb{1}_{\{Z_j > t\}} - \int_t^{\tau_H} \overline{H}^1(x) dx \right) \\
&\quad \left. - (\overline{H}_n(t) - \overline{H}(t)) \left( \frac{1}{\sigma_{1,t}^2} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} (x - t) \overline{H}(x) dx - \frac{1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \overline{H}^1(x) dx \right) \right\}
\end{aligned}$$

et

$$\begin{aligned}
& -\sqrt{N_t} \left[ \frac{\sigma_{1,t}}{N_t} L'_2(\gamma_1, \sigma_{1,t}) - B_2(\gamma_1, \sigma_{1,t}) \right] \\
&= -\sqrt{\frac{\overline{H}(t)}{\overline{H}_n(t)}} \sqrt{\frac{n}{\overline{H}(t)}} \left\{ -\left(\overline{H}_n^1(t) - \overline{H}^1(t)\right) + \frac{1}{\sigma_{1,t}} \left( \frac{1}{n} \sum_{j=1}^n (Z_j - t) \mathbb{1}_{\{Z_j > t\}} - \int_t^{\tau_H} \overline{H}(x) dx \right) \right. \\
&\quad \left. + (\overline{H}_n(t) - \overline{H}(t)) \left( \frac{\overline{H}^1(t)}{\overline{H}(t)} - \frac{1}{\sigma_{1,t}} \frac{1}{\overline{H}(t)} \int_t^{\tau_H} \overline{H}(x) dx \right) \right\}.
\end{aligned}$$

### 3.5.3 Annexe 3 : Deux exemples d'intégrale à calculer

Nous avons à calculer un certain nombre d'intégrales. Pour cela des découpages judicieux doivent être effectués et nous allons donner deux exemples types ci-dessous.

**(a) Exemple d'intégrale dans le cas où  $\gamma_1$  et  $\gamma_2$  sont négatifs avec  $\tau_F = \tau_G$**

Nous allons ci-dessous présenter en détail le calcul de l'intégrale suivante :

$$\text{Var} \left( \sqrt{\frac{n}{\overline{H}(t)}} \left[ \frac{1}{n} \sum_{j=1}^n \delta_j \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}}(Z_j - t)} \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}^1(x)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(x - t))^2} dx \right] \right)$$

parce que d'une part c'est la plus difficile à appréhender et d'autre part elle fait apparaître explicitement la condition  $\gamma > -1/2$  indispensable à la validité de nos résultats. D'après de Haan et Ferreira [37] (Théorème 1.2.5, p. 21), si la convergence

$$\lim_{t \rightarrow \tau_H} \frac{\overline{F}(t + \sigma_{1,t}x)}{\overline{F}(t)} = (1 + \gamma_1 x)^{-1/\gamma_1}$$

a lieu pour un  $\sigma_{1,t} > 0$ , alors elle est aussi satisfaite pour  $\sigma_{1,t} = -\gamma_1(\tau_H - t)$ . De ce fait, dans toute la suite nous supposons  $\sigma_{1,t}$  ainsi défini.

La variance qui nous intéresse peut donc se réécrire selon

$$\begin{aligned} & \text{Var} \left( \sqrt{\frac{n}{\overline{H}(t)}} \left[ \frac{1}{n} \sum_{j=1}^n \delta_j \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}}(Z_j - t)} \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}^1(x)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(x - t))^2} dx \right] \right) \\ &= - \int_t^{\tau_H} \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}}(x - t)} \right)^2 \frac{d\overline{H}^1(x)}{\overline{H}(t)} - \overline{H}(t) \left( \int_t^{\tau_H} \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}}(x - t)} \right) \frac{d\overline{H}^1(x)}{\overline{H}(t)} \right)^2 \\ &= - \int_t^{\tau_H} \left( 1 - \frac{\tau_H - t}{\tau_H - x} \right)^2 \frac{d\overline{H}^1(x)}{\overline{H}(t)} + o(1) \\ &= \int_t^{\tau_H} \frac{\overline{H}^1(x)}{\overline{H}(t)} d \left( \left( 1 - \frac{\tau_H - t}{\tau_H - x} \right)^2 \right) + o(1) \\ &= - \int_t^{\tau_H} \int_x^{\tau_H} \frac{\overline{G}(u)}{\overline{H}(t)} d\overline{F}(u) d \left( \left( 1 - \frac{\tau_H - t}{\tau_H - x} \right)^2 \right) + o(1) \\ &= - \int_t^{\tau_H} \frac{\overline{G}(u)}{\overline{H}(t)} \int_t^u d \left( \left( 1 - \frac{\tau_H - t}{\tau_H - x} \right)^2 \right) d\overline{F}(u) + o(1) \\ &= - \int_t^{\tau_H} \frac{\overline{G}(u)}{\overline{H}(t)} \left( 1 - \frac{\tau_H - t}{\tau_H - u} \right)^2 d\overline{F}(u) + o(1) \end{aligned}$$

$$\begin{aligned}
&= - \int_t^{\tau_H} \frac{\overline{G}(u)}{\overline{H}(t)} \left(1 - \frac{\tau_H - t}{\tau_H - u}\right)^2 \left[ \frac{1}{\gamma_1} (\tau_H - u)^{-\frac{1}{\gamma_1} - 1} \ell_F \left( \frac{1}{\tau_H - u} \right) \right. \\
&\quad \left. + (\tau_H - u)^{-\frac{1}{\gamma_1}} \ell_F \left( \frac{1}{\tau_H - u} \right) \frac{\ell'_F \left( \frac{1}{\tau_H - u} \right)}{\ell_F \left( \frac{1}{\tau_H - u} \right)} \frac{1}{(\tau_H - u)^2} \right] du + o(1) \\
&= - \int_t^{\tau_H} \frac{\overline{H}(u)}{\overline{H}(t)} \left(1 - \frac{\tau_H - t}{\tau_H - u}\right)^2 \frac{1}{\gamma_1} (\tau_H - u)^{-1} du \\
&\quad - \int_t^{\tau_H} \frac{\overline{H}(u)}{\overline{H}(t)} \left(1 - \frac{\tau_H - t}{\tau_H - u}\right)^2 (\tau_H - u)^{-1} \varepsilon \left( \frac{1}{\tau_H - u} \right) du + o(1)
\end{aligned}$$

avec  $\varepsilon \left( \frac{1}{\tau_H - u} \right) := \frac{\ell'_F \left( \frac{1}{\tau_H - u} \right)}{\ell_F \left( \frac{1}{\tau_H - u} \right)} \frac{1}{\tau_H - u}$  en utilisant le fait que  $\ell_F$  est une fonction à variation lente normalisée. Rappelons que nous avons la convergence vers 0 de  $\varepsilon \left( \frac{1}{\tau_H - u} \right)$  lorsque  $u \rightarrow \tau_H$  (cf. Bingham *et al.* [11], page 15).

En effectuant un changement de variable, nous avons alors

$$\begin{aligned}
&\text{Var} \left( \sqrt{\frac{n}{\overline{H}(t)}} \left[ \frac{1}{n} \sum_{j=1}^n \delta_j \left( 1 - \frac{1}{1 + \frac{\gamma_1}{\sigma_{1,t}}(Z_j - t)} \right) \mathbb{1}_{\{Z_j > t\}} - \frac{\gamma_1}{\sigma_{1,t}} \int_t^{\tau_H} \frac{\overline{H}^1(x)}{(1 + \frac{\gamma_1}{\sigma_{1,t}}(x - t))^2} dx \right] \right) \\
&= -\frac{1}{\gamma_1} \int_1^\infty \frac{\overline{H}(\tau_H - \frac{\tau_H - t}{x})}{\overline{H}(t)} (1 - x)^2 x^{-1} dx \\
&\quad - \int_1^\infty \frac{\overline{H}(\tau_H - \frac{\tau_H - t}{x})}{\overline{H}(t)} (1 - x)^2 x^{-1} \varepsilon \left( \frac{x}{\tau_H - t} \right) dx + o(1) \\
&=: T_{1,t} + T_{2,t}.
\end{aligned}$$

Rappelons que  $\frac{\overline{H}(\tau_H - \frac{\tau_H - t}{x})}{\overline{H}(t)} \rightarrow x^{\frac{1}{\gamma}}$  quand  $t$  tend vers  $\tau_H$ . Par ailleurs, en utilisant les bornes de Potter (cf. Bingham *et al.* [11], p. 25), pour toutes constantes  $A > 1$  et  $\eta > 0$ , nous avons l'inégalité suivante :

$$\frac{\overline{H}(\tau_H - \frac{\tau_H - t}{x})}{\overline{H}(t)} = x^{\frac{1}{\gamma}} \frac{\ell_H \left( \frac{x}{\tau_H - t} \right)}{\ell_H \left( \frac{1}{\tau_H - t} \right)} \leq Ax^{\frac{1}{\gamma} + \eta}, \quad x > 1$$

pour tout  $t$  suffisamment grand.

Par ailleurs si  $\gamma > -1/2$ , en choisissant  $\eta$  suffisamment petit, l'intégrale suivante est finie

$$-\frac{1}{\gamma_1} \int_1^\infty x^{\frac{1}{\gamma} + \eta} (1 - x)^2 x^{-1} dx.$$

Ceci nous permet d'utiliser le théorème de convergence dominée et ainsi de déduire que

$$\lim_{t \rightarrow \tau_H} T_{1,t} = -\frac{1}{\gamma_1} \int_1^\infty x^{\frac{1}{\gamma}} (1-x)^2 x^{-1} dx = \frac{\gamma}{\gamma_1} \left( 1 - \frac{2}{1+\gamma} + \frac{1}{1+2\gamma} \right).$$

Concernant  $T_{2,t}$ , en utilisant le fait que  $\varepsilon \left( \frac{x}{\tau_H - t} \right)$  tend vers 0 quand  $t \rightarrow \tau_H$  et l'argument précédent, nous pouvons également utiliser la convergence dominée pour déduire que  $T_{2,t}$  tend vers 0.

**(b) Exemple d'intégrale dans le cas où  $\gamma_1$  et  $\gamma_2$  sont nuls**

De façon similaire à l'exemple précédent, nous avons choisi l'intégrale la plus complexe à appréhender, à savoir :

$$\begin{aligned} \text{Var} \left( \sqrt{\frac{n}{\overline{H}(t)}} \frac{1}{\sigma_{1,t}} \left[ \frac{1}{n} \sum_{j=1}^n \delta_j (Z_j - t) \mathbb{1}_{Z_j > t} - \int_t^{\tau_H} \overline{H}^1(x) dx \right] \right) \\ = \frac{1}{\sigma_{1,t}^2} \frac{-1}{\overline{H}(t)} \int_t^{\tau_H} (x-t)^2 d\overline{H}^1(x) - \overline{H}(t) \left( \frac{1}{\sigma_{1,t}} \frac{-1}{\overline{H}(t)} \int_t^{\tau_H} (x-t) d\overline{H}^1(x) \right)^2 \\ = \frac{2\sigma_{H,t}^2}{\sigma_{1,t}^2} \int_0^{\frac{\tau_H-t}{\sigma_{H,t}}} z \frac{\overline{H}^1(t + \sigma_{H,t}z)}{\overline{H}(t)} dz + o(1). \end{aligned}$$

Cette dernière intégrale peut se réécrire sous la forme

$$\begin{aligned} \int_0^{\frac{\tau_H-t}{\sigma_H}} z \frac{\overline{H}^1(t + \sigma_{H,t}z)}{\overline{H}(t)} dz &= - \int_0^{\frac{\tau_H-t}{\sigma_{H,t}}} \frac{z}{\overline{H}(t)} \int_{t+z\sigma_{H,t}}^{\tau_H} \overline{G}(u) d\overline{F}(u) dz \\ &= - \int_t^{\tau_H} \frac{\overline{G}(u)}{\overline{H}(t)} \int_0^{\frac{u-t}{\sigma_{H,t}}} z dz d\overline{F}(u) \\ &= -\frac{1}{2} \left( \frac{\sigma_{2,t}}{\sigma_{H,t}} \right)^2 \int_0^{\frac{\tau_H-t}{\sigma_{2,t}}} \frac{\overline{G}(t + \sigma_{2,t}z)}{\overline{G}(t)} z^2 d \frac{\overline{F}(t + \sigma_{2,t}z)}{\overline{F}(t)} \\ &= -\frac{1}{2} \left( \frac{\sigma_{2,t}}{\sigma_{H,t}} \right)^2 \left[ \int_0^{\frac{\tau_H-t}{\sigma_{2,t}}} \left\{ \frac{\overline{G}(t + \sigma_{2,t}z)}{\overline{G}(t)} - e^{-z} \right\} z^2 d \frac{\overline{F}(t + \sigma_{2,t}z)}{\overline{F}(t)} \right. \\ &\quad \left. - \int_0^{\frac{\tau_H-t}{\sigma_{2,t}}} \left\{ \frac{\overline{F}(t + \sigma_{2,t}z)}{\overline{F}(t)} - e^{-\frac{\sigma_{2,t}}{\sigma_{1,t}}z} \right\} d(z^2 e^{-z}) \right. \\ &\quad \left. - \int_0^{\frac{\tau_H-t}{\sigma_{2,t}}} e^{-\frac{\sigma_{2,t}}{\sigma_{1,t}}z} d(z^2 e^{-z}) \right] \\ &=: T_{3,t} + T_{4,t} + T_{5,t}. \end{aligned}$$

Nous allons traiter ces trois termes séparément.

Des calculs directs conduisent à

$$T_{5,t} = \frac{\sigma_{H,t}}{\sigma_{1,t}} + o(1)$$

et par conséquent

$$\frac{2\sigma_{H,t}^2}{\sigma_{1,t}^2} T_{5,t} = \frac{2\sigma_{H,t}^3}{\sigma_{1,t}^3} + o(1),$$

ce qui correspond au résultat attendu. Reste alors à montrer la négligeabilité des deux autres termes.

Concernant le terme  $T_{4,t}$ , nous avons

$$\begin{aligned} T_{4,t} &= \frac{1}{2} \left( \frac{\sigma_{2,t}}{\sigma_{H,t}} \right)^2 \int_0^{\frac{\tau_{H-t}}{\sigma_{2,t}}} \left\{ \frac{\overline{F}(t + \sigma_{2,t}z)}{\overline{F}(t)} - e^{-\frac{\sigma_{2,t}}{\sigma_{1,t}}z} \right\} (2ze^{-z} - z^2e^{-z}) dz \\ &= \frac{1}{2} \left( \frac{\sigma_{2,t}}{\sigma_{H,t}} \right)^2 \frac{\sigma_{1,t}}{\sigma_{2,t}} \int_0^{\frac{\tau_{H-t}}{\sigma_{1,t}}} \left\{ \frac{\overline{F}(t + \sigma_{1,t}z)}{\overline{F}(t)} - e^{-z} \right\} \left( 2\frac{\sigma_{1,t}}{\sigma_{2,t}} ze^{-\frac{\sigma_{1,t}}{\sigma_{2,t}}z} - \left( \frac{\sigma_{1,t}}{\sigma_{2,t}} \right)^2 z^2 e^{-\frac{\sigma_{1,t}}{\sigma_{2,t}}z} \right) dz. \end{aligned}$$

Par conséquent

$$|T_{4,t}| \leq \frac{1}{2} \frac{\sigma_{2,t}\sigma_{1,t}}{\sigma_{H,t}^2} \sup_{z \in \left(0, \frac{\tau_{H-t}}{\sigma_{1,t}}\right)} \left| \frac{\overline{F}(t + \sigma_{1,t}z)}{\overline{F}(t)} - e^{-z} \right| \times 4 \frac{\sigma_{2,t}}{\sigma_{1,t}} (1 + o(1))$$

et nous pouvons en déduire que

$$\frac{2\sigma_{H,t}^2}{\sigma_{1,t}^2} |T_{4,t}| \leq 4 \left( \frac{\sigma_{2,t}}{\sigma_{1,t}} \right)^2 (1 + o(1)) \sup_{z \in \left(0, \frac{\tau_{H-t}}{\sigma_{1,t}}\right)} \left| \frac{\overline{F}(t + \sigma_{1,t}z)}{\overline{F}(t)} - e^{-z} \right| \longrightarrow 0$$

lorsque  $t \rightarrow \tau_H$ , en utilisant la condition (3.11).

Concernant maintenant le terme  $T_{3,t}$ , nous avons

$$\begin{aligned} \left| 2 \frac{\sigma_{H,t}^2}{\sigma_{1,t}^2} T_{3,t} \right| &= \left( \frac{\sigma_{2,t}}{\sigma_{1,t}} \right)^2 \left| \int_0^{\frac{\tau_{H-t}}{\sigma_{2,t}}} \left\{ \frac{\overline{G}(t + \sigma_{2,t}z)}{\overline{G}(t)} - e^{-z} \right\} z^2 d \frac{\overline{F}(t + \sigma_{2,t}z)}{\overline{F}(t)} \right| \\ &\leq - \left( \frac{\sigma_{2,t}}{\sigma_{1,t}} \right)^2 \sup_{z \in \left(0, \frac{\tau_{H-t}}{\sigma_{2,t}}\right)} \left| \frac{\overline{G}(t + \sigma_{2,t}z)}{\overline{G}(t)} - e^{-z} \right| \int_0^{\frac{\tau_{H-t}}{\sigma_{2,t}}} z^2 d \frac{\overline{F}(t + \sigma_{2,t}z)}{\overline{F}(t)}. \end{aligned}$$

Comme le supremum tend vers zéro, il suffit de montrer que

$$Q_t := - \left( \frac{\sigma_{2,t}}{\sigma_{1,t}} \right)^2 \int_0^{\frac{\tau_{H-t}}{\sigma_{2,t}}} z^2 d \frac{\overline{F}(t + \sigma_{2,t}z)}{\overline{F}(t)}$$

est bornée. À cette fin, nous allons utiliser (3.7). Remarquons que

$$\begin{aligned}
Q_t &= 2 \int_0^{\frac{\tau_H-t}{\sigma_{1,t}}} \frac{\overline{F}(t + \sigma_{1,t}z)}{\overline{F}(t)} z dz \\
&= 2 \chi_1(t) \int_0^{\frac{\tau_H-t}{\sigma_{1,t}}} z \psi_1(z) \frac{1}{\psi_1(z) \chi_1(t)} \left[ \frac{\overline{F}(t + \sigma_{1,t}z)}{\overline{F}(t)} - e^{-z} \right] dz + 2 \int_0^{\frac{\tau_H-t}{\sigma_{1,t}}} z e^{-z} dz \\
&= 2 \chi_1(t) \int_0^\infty f_t(z) dz + 2 + o(1).
\end{aligned}$$

Comme  $f_t(z) \rightarrow z\psi_1(z)$  lorsque  $t \rightarrow \tau_H$  et  $|f_t(z)| \leq 2z|\psi_1(z)|$  qui est intégrable, nous déduisons que  $Q_t \rightarrow 2$  par le théorème de convergence dominée.

# Chapitre 4

## Un modèle linéaire autorégressif pour des maxima et applications

### Résumé

Dans la plupart des applications en statistique et particulièrement en sciences de l'environnement, les données dont nous disposons et que nous cherchons à modéliser ne sont pas toujours indépendantes. C'est pourquoi les modèles statistiques que nous proposons doivent pouvoir prendre en compte des formes de dépendance temporelle et/ou spatiale. Dans l'étude des séries temporelles, les processus linéaires autorégressifs sont très souvent utilisés. Ces derniers ont l'avantage de considérer la dépendance temporelle entre l'observation présente et les observations passées. Cependant, il n'y a pas dans la littérature de processus autorégressif qui soit linéaire et adapté à la distribution des valeurs extrêmes généralisée. Dans ce chapitre, nous proposons un modèle autorégressif d'ordre un adapté à la distribution de Gumbel. Pour ce modèle nous déclinons un certain nombre de propriétés théoriques. Nous illustrons ensuite par simulations les qualités, à distance finie, du modèle et de l'estimation des paramètres associés. Enfin, après



l'avoir utilisé dans deux applications pratiques concernant la modélisation, d'une part, de maxima journaliers et hebdomadaires de méthane (CH<sub>4</sub>) et d'autre part, de dioxyde de carbone (CO<sub>2</sub>) mesurés à Gif-sur-Yvette (France), une extension du modèle à la distribution GEV est proposée.

## 4.1 Introduction

Dans les applications statistiques, il n'est pas rare que seules des données consistant en des maxima ne soient disponibles. Les séries temporelles, constituées de ces maxima journaliers, hebdomadaires ou annuels, présentent de façon générale un phénomène de persistance dans le temps dont il faut absolument tenir compte (Coles *et al.* [17]). En effet, la valeur au temps présent d'un phénomène observé dépend bien souvent des valeurs aux temps passés. C'est typiquement le cas si on s'intéresse par exemple à la température maximale du jour  $j$ , qui clairement n'est pas indépendante de celle du jour précédent.

Rappelons que dans le chapitre d'introduction, nous avons vu que, sous l'hypothèse d'appartenance au domaine d'attraction des valeurs extrêmes, les maxima d'échantillons ont pour distribution limite la distribution GEV définie comme suit :

$$H_{\gamma,\mu,\sigma}(x) = \exp \left\{ - \left[ 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\gamma}} \right\} \quad (4.1)$$

où  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ ,  $\gamma \in \mathbb{R}$  et  $x$  est tel que  $1 + \gamma(x - \mu)/\sigma > 0$ , les paramètres  $\mu$ ,  $\sigma$  et  $\gamma$  représentant respectivement les paramètres de localisation, d'échelle et de forme. Dans le cas où  $\gamma$  est nul, il s'agit de la distribution de Gumbel définie par

$$H_{\mu,\sigma}(x) = \exp \left\{ - \exp \left( - \frac{x - \mu}{\sigma} \right) \right\}, \quad -\infty < x < +\infty.$$

Ce résultat de convergence est également valable pour des processus stationnaires sous une condition peu restrictive sur la dépendance présente dans les données (cf. Leadbetter *et al.* [48]). Pour gérer la dépendance entre maxima, il est possible d'utiliser une approche

par blocs. Dans le cas où la taille de ces blocs est suffisamment grande (maxima annuels par exemple), on considère alors que les maxima obtenus sont indépendants entre eux et on s'affranchit donc de tout problème de dépendance. Seulement, il arrive souvent en pratique que nous ayons à considérer des maxima sur de plus courtes périodes (maxima hebdomadaires ou journaliers), afin de disposer d'un nombre suffisant de données pour faire de l'inférence. Il devient alors peu raisonnable de supposer que ces maxima puissent être indépendants entre eux. Une approche possible, illustrée par exemple dans Kysely [47], consiste alors à utiliser un modèle autorégressif d'ordre un (avec un bruit gaussien). Cela paraît d'autant plus naturel que dans le domaine de l'analyse des séries temporelles, une des approches traditionnelles consiste en l'utilisation de modèles linéaires et particulièrement des modèles autorégressifs et/ou à moyenne mobile (ARMA pour Auto Regressive Moving Average). Toutefois, un bruit gaussien ne paraît pas adapté lorsque les données consistent en des maxima qui peuvent donc être considérés comme issus d'une distribution GEV. Dans le cas de distributions à queues lourdes ( $\gamma > 0$ ), des modèles linéaires de type ARMA basés sur l'utilisation de variables aléatoires  $\alpha$ -stables ont été proposés (cf. Embrechts *et al.* [32], Section 7.2 et également Mikosch *et al.* [53] et Samorodnitsky et Taqqu [64], Section 7.12). Cependant, en dehors du cadre des distributions à queues lourdes, aucun modèle autorégressif linéaire permettant des données distribuées selon la distribution GEV n'existe. La distribution GEV n'étant pas stable pour l'addition mais pour l'opérateur *max*, une alternative consiste à utiliser des processus max-ARMA qui sont des processus max-stables (cf. e.g. Davis et Resnick [22]). Cependant nous préférons privilégier les processus linéaires dont l'interprétation est plus simple que celle des processus max-ARMA. Plus particulièrement, nous traitons dans ce chapitre le cas de la distribution de Gumbel, d'une grande importance en théorie des valeurs extrêmes. En effet, cette distribution apparaît comme la distribution limite des maxima de la plupart des distributions standards (comme la distribution gaussienne) et est particulièrement utilisée dans un grand nombre d'applications notamment en sciences environnementales.

Nous proposons donc un modèle linéaire autorégressif adapté à la distribution de Gumbel pour lequel nous établissons un certain nombre de propriétés théoriques. En Section 4.3, des résultats de simulation, tant au niveau du modèle que des paramètres associés, sont proposés afin d'illustrer la validité du modèle. Comme ce dernier a été motivé par des applications concrètes en géoscience, nous l'appliquons dans la Section 4.4 à des maxima de méthane (CH<sub>4</sub>) et de dioxyde de carbone (CO<sub>2</sub>). Enfin, une extension au cas de la distribution GEV est proposée en Section 4.5. Les preuves des résultats théoriques figurent en Section 4.6.

## 4.2 Modèle autorégressif pour la distribution de Gumbel

### 4.2.1 Description du modèle proposé : résultat théorique

C'est sur la base d'une relation linéaire entre la distribution de Gumbel et le logarithme d'une variable  $\alpha$ -stable positive que le modèle proposé dans ce chapitre est construit. Si  $X$  est une variable de loi Gumbel de paramètres  $\mu$  et  $\sigma$  indépendante de  $S$ , cette dernière étant une variable  $\alpha$ -stable positive définie par sa transformé de Laplace

$$\mathbb{E}(\exp(-uS)) = \exp(-u^\alpha), \quad u \geq 0, \quad \alpha \in (0, 1), \quad (4.2)$$

alors la somme  $X + \sigma \log S$  est aussi de loi Gumbel de paramètres  $\mu$  et  $\sigma/\alpha$ .

Cette propriété a déjà été utilisée, notamment dans des travaux de Fougères *et al.* [35] qui ont proposé différents modèles de mélanges basés sur ces distributions  $\alpha$ -stables positives pour des données extrêmes. Ils étudient alors des modèles temporels basés sur des processus linéaires stables parmi lesquels le modèle autorégressif caché défini de la manière suivante :

$$X_t = \mu_t + G_t + \sigma \log(H_t)$$

avec  $G_t$  une variable de loi Gumbel( $0, \sigma$ ) et  $H_t = \sum_{i=0}^{\infty} \rho^i S_{t-i}$  où  $\rho \in (0, 1)$  et où les  $S_i$  sont des variables aléatoires  $\alpha$ -stables positives. De la définition de  $H_t$  découle l'équation de récurrence suivante qui met en exergue le modèle autorégressif caché :

$$H_t = \rho H_{t-1} + S_t.$$

De la même manière, le modèle proposé dans ce chapitre et exposé dans la proposition suivante est également basé sur cette relation clé entre la distribution de Gumbel et le logarithme d'une variable  $\alpha$ -stable positive. Il permet en outre une structure autorégressive directement sur des variables aléatoires distribuées selon une loi de Gumbel.

**Proposition 4.1** *Soient  $S_t, t \in \mathbb{Z}$ , des variables aléatoires indépendantes et de même loi que  $S$  définie en (4.2). Soit  $\{X_t, t \in \mathbb{Z}\}$  un processus stochastique défini par la relation de récurrence suivante :*

$$X_t = \alpha X_{t-1} + \alpha \sigma \log S_t + \mu(1 - \alpha) \quad (4.3)$$

où  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_*^+$ .

L'équation (4.3) a une unique solution strictement stationnaire donnée par

$$X_t = \mu + \sigma \sum_{j=0}^{\infty} \alpha^{j+1} \log S_{t-j} \quad (4.4)$$

et  $X_t$  suit alors une loi de Gumbel  $(\mu, \sigma), \forall t \in \mathbb{Z}$ .

Pour ce processus  $\{X_t, t \in \mathbb{Z}\}$ , nous pouvons facilement calculer la covariance entre  $X_t$  et  $X_{t-h}$ , laquelle est une fonction croissante de  $\alpha$ ; plus précisément nous avons  $\text{Cov}(X_t, X_{t-h}) = \text{Var}(X_0)\alpha^{|h|}$ .

La Figure 4.1 présente une simulation de  $X_t$  pour  $t = 1, \dots, 200$  et différentes valeurs de  $\alpha \in \{0.2, 0.4, 0.6\}$  et la Figure 4.2 présente le « scatter plot » de  $X_{t+1}$  contre  $X_t$ , basé sur la même série. Ces graphiques confirment que la dépendance à court terme est directement liée à la valeur de  $\alpha$  : plus  $\alpha$  est grand et plus la dépendance est forte.

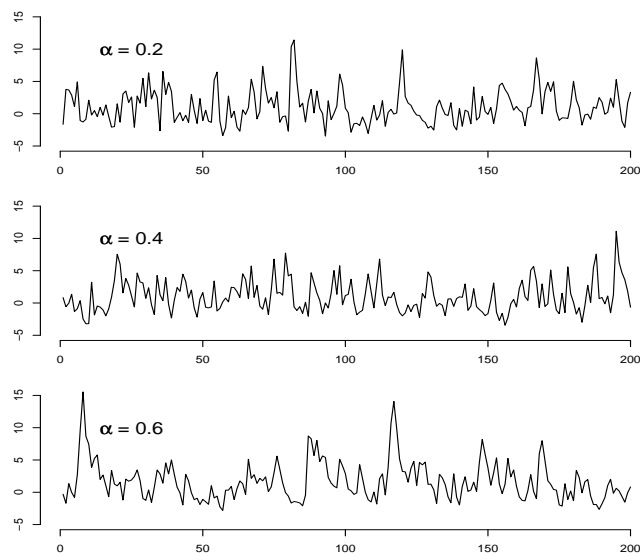


FIG. 4.1 – Une simulation de  $X_t$  pour  $t = 1, \dots, 200$  et différentes valeurs de  $\alpha$ . Nous posons  $\mu = 0$  et  $\sigma = 2$ .

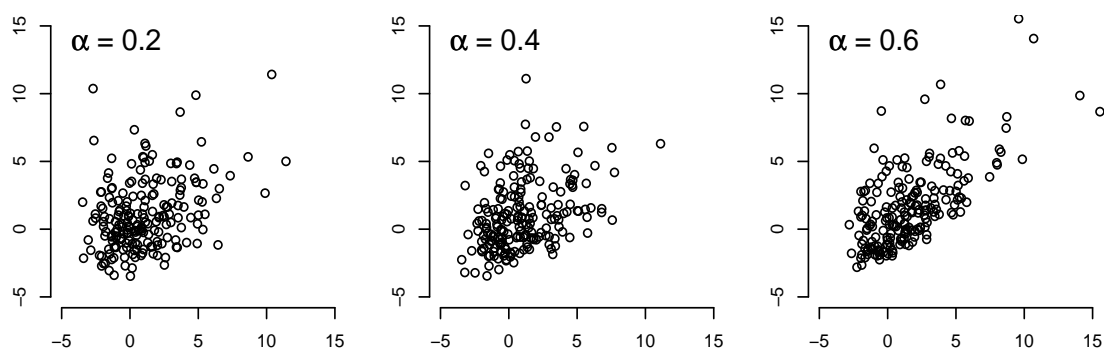


FIG. 4.2 – « Scatter plots » de  $(X_t, X_{t+1})$  pour  $t = 1, \dots, 200$  et différentes valeurs de  $\alpha$ . Nous posons  $\mu = 0$  et  $\sigma = 2$ .

### 4.2.2 Estimation des paramètres du modèle

Trois paramètres  $\mu$ ,  $\sigma$  et  $\alpha$  sont présents dans le modèle proposé. En pratique ils sont inconnus et doivent donc être estimés. Supposons que nous disposons des observations  $\{X_1, \dots, X_T\}$ . Comme nous connaissons d'après la Proposition 4.1 la distribution de  $X_t$ , il est tout à fait envisageable de considérer la méthode des moments. Comme  $X_t$  suit une loi de Gumbel, de paramètres  $\mu$  et  $\sigma$ , pour tout entier  $t$ , nous déduisons les deux équations suivantes :

$$\mu = \mathbb{E}(X_0) - \frac{\delta \sqrt{6 \text{Var}(X_0)}}{\pi} \quad (4.5)$$

$$\sigma = \frac{\sqrt{6 \text{Var}(X_0)}}{\pi} \quad (4.6)$$

où  $\delta$  est la constante d'Euler. Le processus défini en (4.3) est ergodique ; nous pouvons donc calculer ses caractéristiques (moyenne, variance, fonction d'autocovariance) à partir d'une seule trajectoire du processus, c'est-à-dire à partir d'une observation du processus. Par conséquent, des estimateurs naturels de  $\mu$  et  $\sigma$  sont ceux basés sur la méthode des moments :

$$\hat{\mu} = \bar{X} - \frac{\delta \sqrt{6s}}{\pi} \quad (4.7)$$

$$\hat{\sigma} = \frac{\sqrt{6s}}{\pi} \quad (4.8)$$

où  $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$  et  $s^2 = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2$ .

Concernant maintenant l'estimation de  $\alpha$ , nous pouvons utiliser la méthode des moindres carrés, ce qui nous conduit à nous intéresser à

$$\arg \min_r \sum_{t=1}^{T-1} ([X_{t+1} - \mathbb{E}(X_0)] - r [X_t - \mathbb{E}(X_0)])^2 = \frac{\sum_{t=1}^{T-1} (X_t - \mathbb{E}(X_0))(X_{t+1} - \mathbb{E}(X_0))}{\sum_{t=1}^{T-1} (X_t - \mathbb{E}(X_0))^2}.$$

En apportant une légère modification au dénominateur de façon à faire apparaître explicitement  $s^2$  et en remplaçant  $\mathbb{E}(X_0)$  par son estimateur naturel  $\bar{X}$ , nous obtenons

l'estimateur de  $\hat{\alpha}$  suivant :

$$\hat{\alpha} = \frac{\frac{1}{T} \sum_{t=1}^{T-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{s^2}. \quad (4.9)$$

Ces estimateurs sont consistants, comme l'indique la proposition suivante.

**Proposition 4.2** *Les estimateurs  $\hat{\mu}$ ,  $\hat{\sigma}$  et  $\hat{\alpha}$  respectivement définis en (4.7), (4.8) et (4.9) convergent presque sûrement vers la vraie valeur des paramètres.*

Nous allons maintenant nous attacher à établir un résultat de normalité asymptotique pour ces estimateurs renormalisés. Ceci fait l'objet de la proposition suivante.

**Proposition 4.3** *Le vecteur*

$$\sqrt{T} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \\ \hat{\alpha} - \alpha \end{pmatrix}$$

*est asymptotiquement gaussien d'espérance nulle et de matrice de variance-covariance égale à*

$$\begin{pmatrix} \frac{\pi^2 \sigma^2}{6} \frac{1+\alpha}{1-\alpha} - \frac{12\delta\sigma^2\zeta(3)(1+\alpha+\alpha^2)}{\pi^2(1-\alpha^2)} + \frac{11\delta^2\sigma^2(1+\alpha^2)}{10(1-\alpha^2)} & \frac{6\sigma^2\zeta(3)(1+\alpha+\alpha^2)}{\pi^2(1-\alpha^2)} - \frac{11\delta\sigma^2(1+\alpha^2)}{10(1-\alpha^2)} & -\alpha\sigma\delta \\ \frac{6\sigma^2\zeta(3)(1+\alpha+\alpha^2)}{\pi^2(1-\alpha^2)} - \frac{11\delta\sigma^2(1+\alpha^2)}{10(1-\alpha^2)} & \frac{11\sigma^2(1+\alpha^2)}{10(1-\alpha^2)} & \alpha\sigma \\ -\alpha\sigma\delta & \alpha\sigma & 1 - \alpha^2 \end{pmatrix} \quad (4.10)$$

*où  $\delta$  est la constante d'Euler et  $\zeta(\cdot)$  la fonction de Riemann.*

Notons que les termes de la matrice de variance-covariance (4.10) sont des fonctions de  $\alpha$ . Plus précisément, les termes de variance associés aux estimateurs  $\hat{\mu}$  et  $\hat{\sigma}$  sont des fonctions croissantes de  $\alpha$  alors que la variance associée à  $\hat{\alpha}$  est une fonction décroissante de  $\alpha$ .

Avant d'appliquer ce modèle sur des données réelles, nous allons présenter d'autres résultats théoriques concernant cette fois-ci la loi jointe du vecteur  $\mathbf{X}_h = (X_t, \dots, X_{t-h})^t$ ,  $h > 0$ , ainsi que des coefficients de dépendance asymptotique.

### 4.2.3 Distribution jointe du vecteur $\mathbf{X}_h = (X_t, \dots, X_{t-h})^t$ , $h > 0$

Comme la fonction caractéristique d'un vecteur aléatoire caractérise complètement sa distribution, la proposition suivante est essentielle pour l'étude de la loi jointe du vecteur  $\mathbf{X}_h = (X_t, \dots, X_{t-h})^t$ ,  $h > 0$ .

**Lemme 4.1** *Soit  $\{X_t, t \in \mathbb{Z}\}$  défini de la même manière que dans la Proposition 4.1.*

*La fonction caractéristique de  $\mathbf{X}_h = (X_t, \dots, X_{t-h})^t$ ,  $h > 0$ , est donnée par*

$$\mathbb{E}(e^{i\langle u, \mathbf{X}_h \rangle}) = e^{i\mu \sum_{j=0}^h u_j} \Gamma \left( 1 - i\sigma \sum_{j=0}^h u_j \alpha^{h-j} \right) \prod_{j=0}^{h-1} \frac{\Gamma \left( 1 - i\sigma \sum_{k=0}^j u_k \alpha^{j-k} \right)}{\Gamma \left( 1 - i\sigma \sum_{k=0}^j u_k \alpha^{j-k+1} \right)}.$$

D'après cette représentation, nous retrouvons que le processus  $\{X_t, t \in \mathbb{Z}\}$  est effectivement strictement stationnaire.

### 4.2.4 Dépendance asymptotique

En modélisation des extrêmes multivariés, nous nous intéressons souvent à la dépendance dans les queues. En particulier il convient de savoir si nous sommes dans le cas de l'indépendance asymptotique ou non. Pour mesurer cette dépendance, nous pouvons calculer deux paramètres  $\chi$  et  $\bar{\chi}$  définis dans la proposition ci-dessous.

**Proposition 4.4** *Soit  $\{X_t, t \in \mathbb{Z}\}$  défini comme dans la Proposition 4.1. Alors*

1. *Le paramètre de dépendance de queue défini par Joe [43] en 1993 selon*

$$\chi = \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X_{t-1} > x, X_t > x)}{\mathbb{P}(X_{t-1} > x)}$$

*est égal à zéro, ce qui correspond au cas de l'indépendance asymptotique.*

2. *Le paramètre de dépendance  $\bar{\chi}$  défini par Coles et al. [17] selon*

$$\bar{\chi} = \lim_{x \rightarrow \infty} \frac{2 \log \mathbb{P}(X_{t-1} > x)}{\log \mathbb{P}(X_{t-1} > x, X_t > x)} - 1$$

*vaut  $\alpha/(2 - \alpha) \in (0, 1)$ . C'est une mesure qui augmente avec  $\alpha$  et qui est donc d'autant plus forte que la dépendance est importante.*



Notons que le paramètre  $\chi$  est directement lié à l'indice extrémal, souvent noté  $\theta$ , et que  $\bar{\chi}$  est lié au coefficient de dépendance de queue usuellement noté  $\eta$ . Pour des précisions sur les mesures de dépendance dans les queues, nous renvoyons le lecteur à Coles [16], Coles *et al.* [17], Fougères [34] et Joe [44] entre autres.

La Proposition 4.4 nous indique que nous sommes dans le cas de l'indépendance asymptotique, c'est-à-dire que la probabilité que les deux composantes  $X_t$  et  $X_{t-1}$  soient simultanément grandes est négligeable par rapport à la probabilité que l'une des composantes soit grande. En d'autres termes, la fonction de répartition bivariée de  $X_t$  et  $X_{t-1}$  appartient au domaine d'attraction de l'indépendance. Toutefois, il s'agit uniquement d'indépendance asymptotique, le paramètre  $\bar{\chi}$  indiquant une dépendance « quasi-linéaire » d'autant plus forte que le paramètre  $\alpha$  est proche de 1.

### 4.3 Résultats de simulation

Afin d'étudier le comportement des estimateurs proposés  $\hat{\mu}$ ,  $\hat{\sigma}$  et  $\hat{\alpha}$  à distance finie, nous avons simulé 1000 échantillons de taille  $n = 50, 100, \dots, 1000$  pour différentes valeurs de  $\alpha \in \{0.2, 0.5, 0.8\}$ . Nous avons posé  $\mu = 0$  et  $\sigma = 2$ . Pour chacun de ces échantillons, les paramètres ont été estimés. Comme nous avons généré 1000 échantillons, nous pouvons calculer pour chaque paramètre la moyenne des estimations mais aussi les premier et troisième quartiles. Les résultats sont représentés sur un même graphe ayant comme axe des abscisses les tailles d'échantillon. Les valeurs des moyennes  $y$  sont représentées par la ligne en tiret et celles des premiers et troisièmes quartiles sont représentés par les lignes en tiret-pointillé (voir Figure 4.3). Cette figure témoigne de la bonne convergence des estimateurs. Nous observons que les estimations sont assez bonnes pour des échantillons suffisamment grands (avec plus de 300 observations). L'estimation des deux paramètres de la Gumbel ( $\mu$  et  $\sigma$ ) est meilleure quand la dépendance n'est pas trop forte. Cela peut

s'expliquer par le fait qu'alors la quantité d'information est plus importante. En effet, si  $\alpha$  est proche de 1 et que donc la dépendance est très forte,  $X_t$  apporte moins d'information par rapport à  $X_{t-1}$  que si la dépendance est faible. Ces résultats sont par conséquent tout à fait cohérents avec ce que l'on avait pu observer en Section 4.2.2 concernant la matrice de variance-covariance asymptotique (4.10).

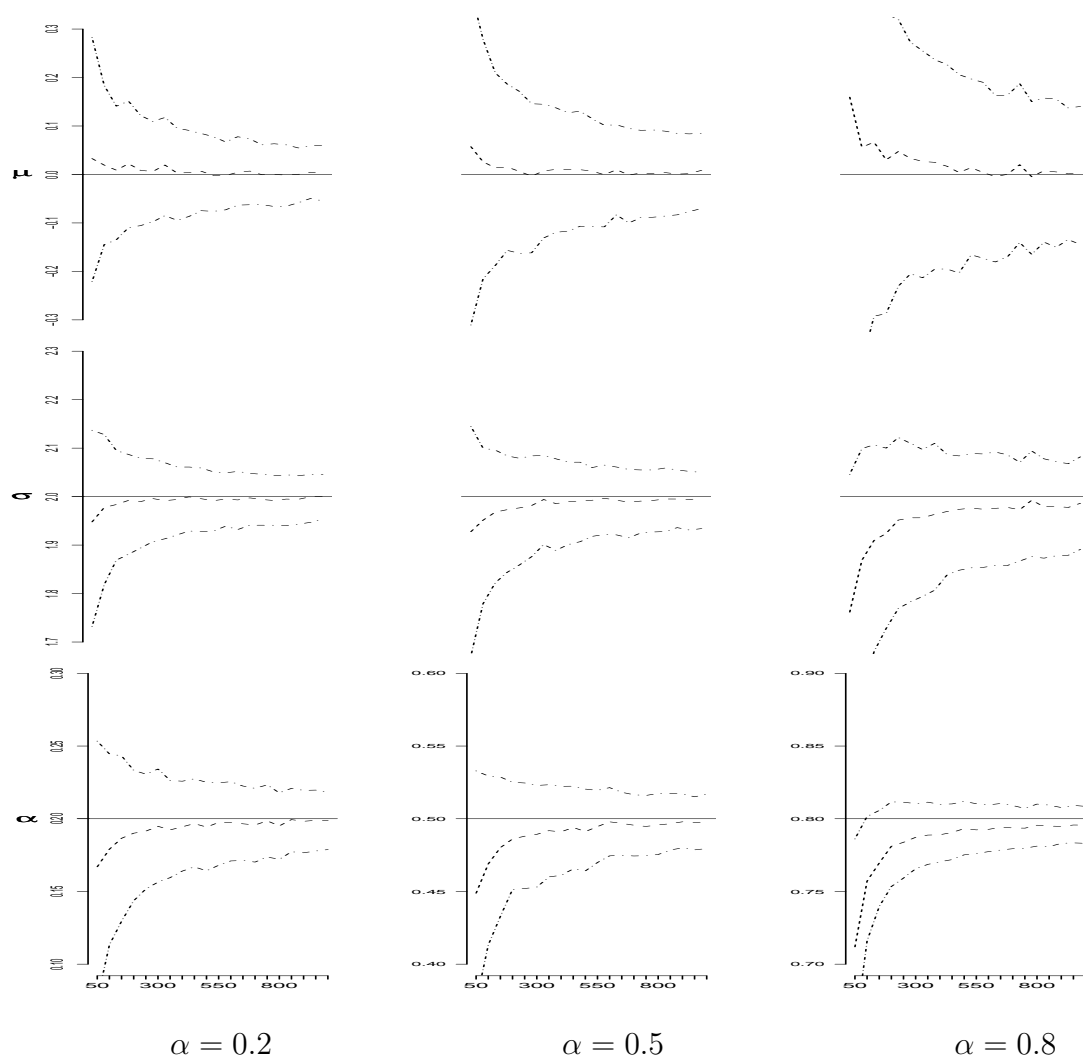


FIG. 4.3 – Moyenne (ligne en tiret), premier et troisième quartiles (ligne en tiret-pointillé) des estimateurs de  $\mu$ ,  $\sigma$  et  $\alpha$  (basés sur 1000 échantillons) pour différentes tailles d'échantillons  $n \in \{50, 100, \dots, 1000\}$  données en abscisse. Pour ces simulations  $\mu = 0$ ,  $\sigma = 2$ .

Après avoir étudié le comportement des estimateurs en fonction de la taille de l'échantillon ( $n$  allant de 50 à 1000), nous nous intéressons au comportement de ces derniers sur des échantillons de très grande taille ( $n = 5000$ ). La Figure 4.4 illustre les résultats obtenus ; nous remarquons qu'ils sont également bien cohérents avec les résultats asymptotiques décrits en Section 4.2.2.

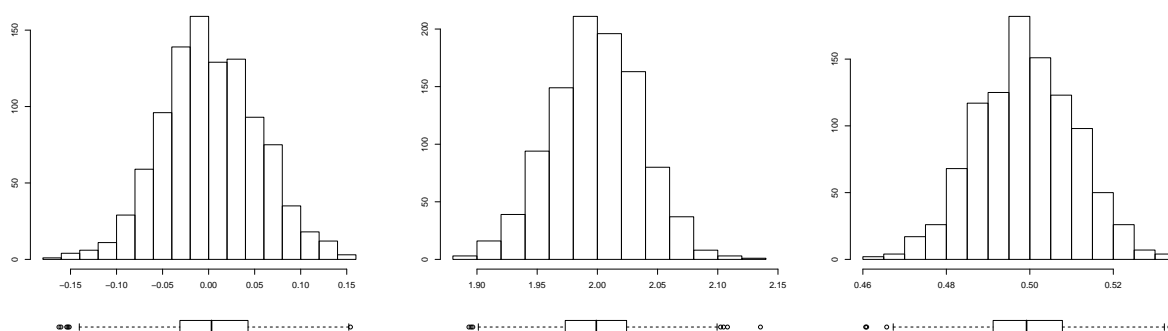


FIG. 4.4 – Histogrammes et boîtes à moustaches basés sur 1000 échantillons correspondant aux estimateurs des trois paramètres,  $\hat{\mu}$ ,  $\hat{\sigma}$  et  $\hat{\alpha}$ . Pour ces simulations  $\mu = 0$ ,  $\sigma = 2$ ,  $\alpha = 0.5$  et  $n=5000$ .

Nous allons terminer cette étude de simulations en revenant sur un problème fréquent en pratique qui est celui où nous disposons d'un échantillon issu d'une loi de Gumbel. En pratique, les modèles gaussiens sont souvent utilisés y compris pour ce type de données. Notre but ci-dessous est d'illustrer la perte que l'utilisation du modèle gaussien génère au gain obtenu en ajustant avec un modèle de Gumbel. Bien entendu, il est tout à fait normal qu'avec ce dernier les résultats soient bien meilleurs. Ceci est illustré en Figure 4.5 où nous comparons donc le modèle de Gumbel proposé et le modèle classique gaussien. Dans les deux cas, les données sont générées selon le modèle de Gumbel. Pour cela, nous avons simulé  $X_1, \dots, X_{1000}$  selon le modèle décrit en (4.3) avec  $\alpha = 0.5$ ,  $\mu = 0$  et  $\sigma = 2$  et nous nous sommes intéressés à la distribution de  $X_{t+1}|X_t = x$ ,  $t \gg 1000$ . Nous avons alors dû générer une valeur en  $X_t = x$  selon une Gumbel  $(\mu, \sigma)$ .

Dans un premier temps, plaçons nous dans le cas du modèle de Gumbel défini en (4.3) (Modèle 1 sur la Figure 4.5). Nous estimons  $\mu$ ,  $\sigma$  et  $\alpha$  selon (4.7), (4.8) et (4.9) basés sur les simulations  $X_1, \dots, X_{1000}$  et nous présentons un histogramme de  $\widehat{X}_{t+1}|X_t = x$  basé sur les 1000 évaluations de  $\widehat{X}_{t+1} = \widehat{\alpha}x + \widehat{\alpha}\widehat{\sigma} \log S_{t+1} + \widehat{\mu}(1 - \widehat{\alpha})$  avec  $S_{t+1}$  une variable aléatoire  $\widehat{\alpha}$ -stable positive définie par (4.2). Sur cet histogramme, nous avons tracé la densité, représentée en tiret sur la Figure 4.5, associée au logarithme d'une variable  $\widehat{\alpha}$ -stable. Le trait plein correspond quant à lui au vrai modèle, c'est-à-dire à la densité associée au logarithme d'une variable  $\alpha$ -stable. En effet, lorsque les paramètres sont connus, nous avons

$$\begin{aligned} \mathbb{P}(X_{t+1} \leq y|X_t = x) &= \mathbb{P}(\alpha X_t + \alpha\sigma \log S_{t+1} + \mu(1 - \alpha) \leq y|X_t = x) \\ &= \mathbb{P}(\alpha x + \alpha\sigma \log S_{t+1} + \mu(1 - \alpha) \leq y) \\ &= \mathbb{P}\left(\log S_{t+1} \leq \frac{y - (\alpha x + \mu(1 - \alpha))}{\alpha\sigma}\right) \end{aligned}$$

avec  $S_{t+1}$  une variable aléatoire  $\alpha$ -stable positive définie par (4.2). La vraie densité et la densité estimée sont très proches l'une de l'autre et c'est pourquoi il est difficile de les distinguer sur la figure correspondante.

Dans un second temps, considérons un modèle gaussien (Modèle 2 sur la Figure 4.5). Comme précédemment, nous estimons  $\alpha$  d'après (4.9) et obtenons la même estimation que pour le Modèle 1 car l'estimateur de  $\alpha$  est indépendant du modèle. Ensuite, nous présentons un histogramme de  $\widehat{X}_{t+1}|X_t = x$  basé sur les 1000 évaluations de  $\widehat{X}_{t+1} = \widehat{\alpha}x + \theta_{t+1} + \widehat{\nu}(1 - \widehat{\alpha})$  avec  $\widehat{\nu} = \overline{X}$  et  $\theta_{t+1}$  une variable aléatoire gaussienne d'espérance nulle et de variance  $\widehat{\sigma}_\theta^2 = (\overline{X^2} - \overline{X}^2) \times (1 - \widehat{\alpha}^2)$ . Le trait plein sur l'histogramme correspond au vrai modèle, c'est-à-dire à la densité associée au logarithme d'une variable  $\alpha$ -stable. Cette figure illustre clairement que notre modèle a le sérieux avantage de tenir compte de l'asymétrie de la distribution ce qui n'est pas le cas du second modèle basé sur l'hypothèse gaussienne classique.

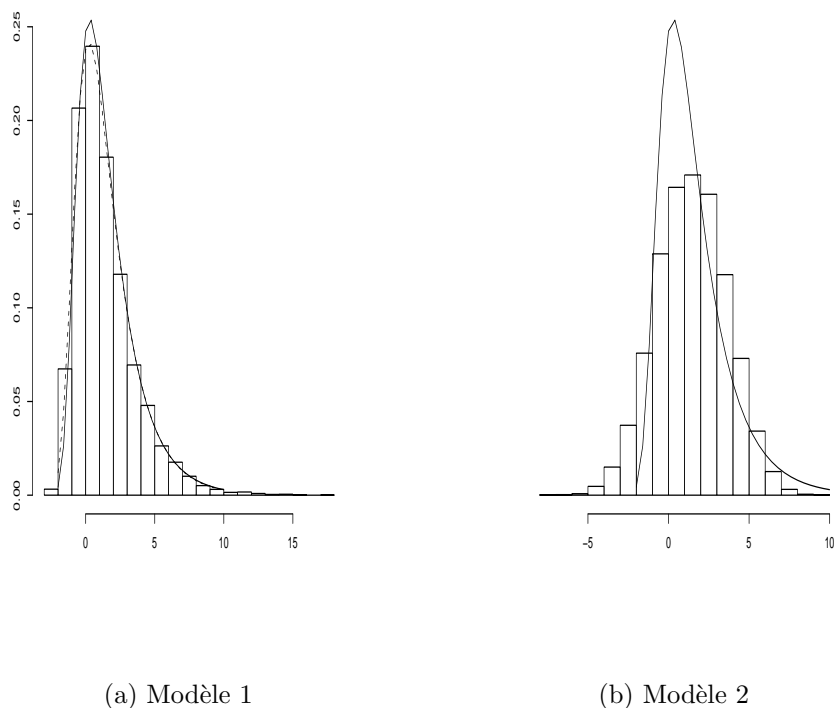


FIG. 4.5 – Histogrammes conditionnels de  $\widehat{X}_{t+1}$  pour le modèle Gumbel (Modèle 1) et pour le modèle gaussien (Modèle 2).

## 4.4 Applications aux données de méthane et de dioxyde de carbone

L'unité de mesure pour le méthane (CH<sub>4</sub>) est le ppb (« part per billion »), c'est-à-dire partie par milliard et celle pour le dioxyde de carbone (CO<sub>2</sub>) est le ppm (« part per million »), c'est-à-dire partie par million. Ces mesures représentent donc des proportions dans l'air. Plus précisément, nous disposons de maxima journaliers et hebdomadaires de méthane de l'année 2002 à l'année 2006 et de dioxyde de carbone de l'année 1981 à l'année 2001. Ces mesures ont été effectuées à Gif-sur-Yvette, dans l'Essonne en France. Comme les données de dioxyde de carbone présentaient une très nette tendance linéaire, celle-ci a été prise en compte en ajustant une régression linéaire et en la retirant ensuite des données.

### a. Les maxima journaliers et hebdomadaires de méthane

Les maxima journaliers et hebdomadaires de méthane sont représentés sur la Figure 4.6. Cette dernière nous permet également de visualiser la présence de valeurs manquantes (35% des maxima journaliers et 16% des maxima hebdomadaires).

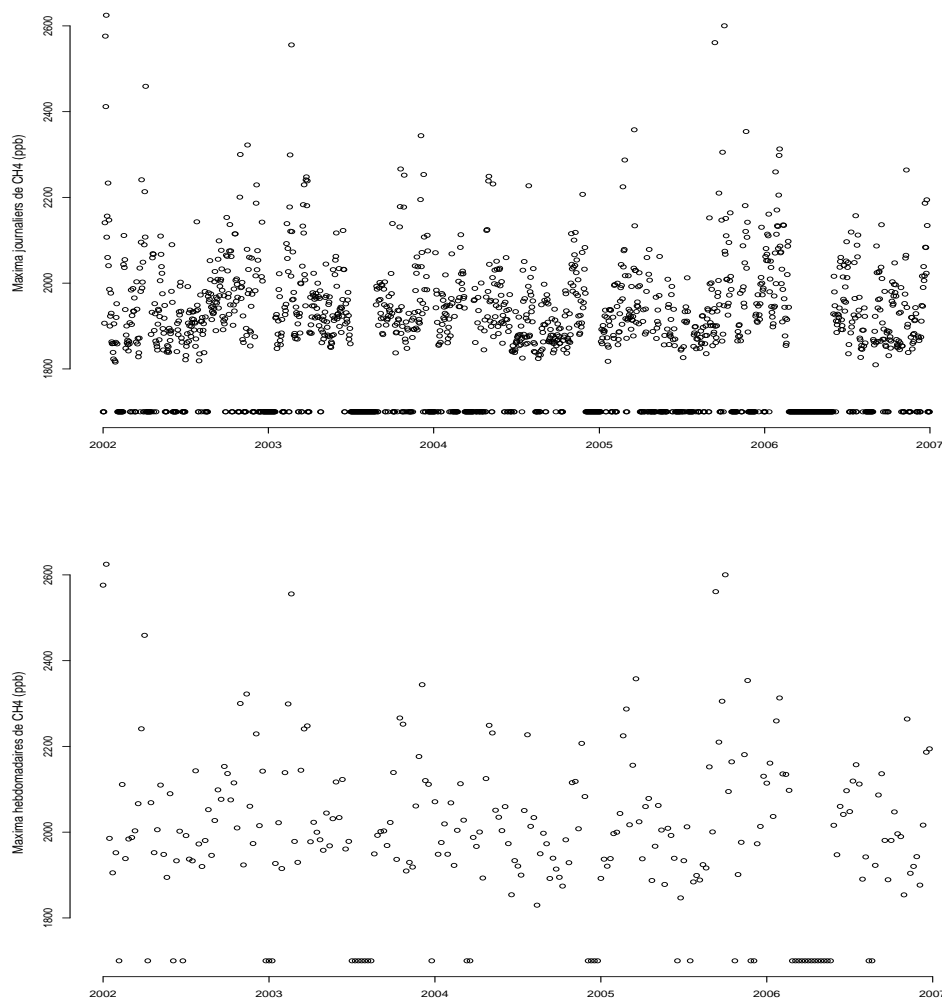


FIG. 4.6 – Séries des maxima journaliers (en haut) et hebdomadaires (en bas) de  $CH_4$  à Gif-sur-Yvette avec valeurs manquantes.

La Figure 4.7 présente quant à elle les « QQ-plots » associés aux séries de maxima journaliers et hebdomadaires de méthane. Les paramètres de la loi de Gumbel ont été estimés selon la méthode des moments. Au vu de ces graphes, il est tout à fait raisonnable d'ajuster aux données une telle distribution.

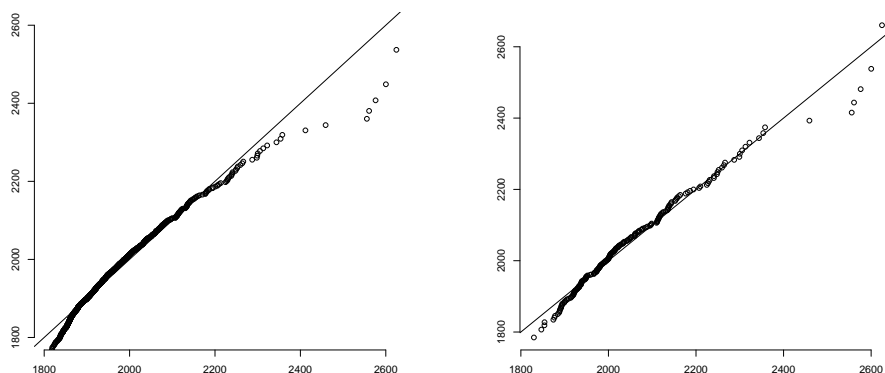


FIG. 4.7 – « QQplot » pour la loi de Gumbel des maxima journaliers (à gauche) et hebdomadaires (à droite) de  $CH_4$ . Les paramètres correspondants sont estimés par la méthode des moments. L'unité est le ppb.

Une caractéristique intéressante de ces données réside dans la dépendance temporelle présente. Pour visualiser cette dépendance, nous présentons en Figure 4.8 les « scatter plots » des données successives. Nous constatons alors que ces graphiques ne sont pas si éloignés de ceux présentés en Figure 4.2 avec  $\alpha = 0.4$  ou  $\alpha = 0.6$ . Plus précisément, nous pouvons calculer un estimateur de  $\alpha$  pour ces données réelles en utilisant (4.9) et obtenons  $\hat{\alpha}_{\text{Jour}} = 0.54$  et  $\hat{\alpha}_{\text{Semaine}} = 0.35$ .

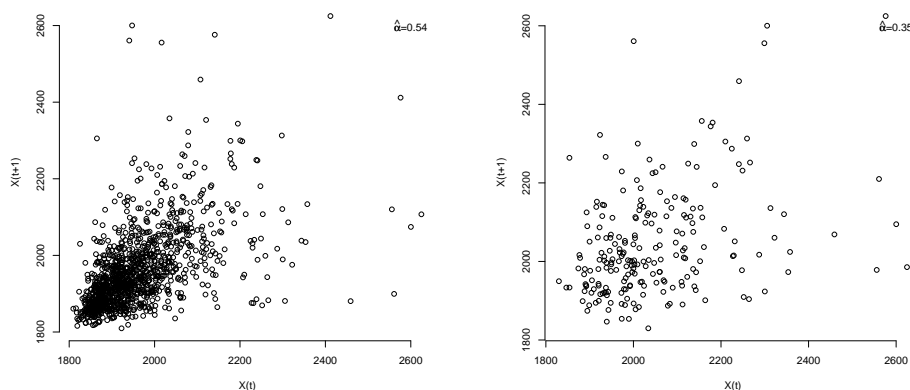


FIG. 4.8 – « Scatter plots » des valeurs successives, i.e.  $(X_t, X_{t+1})$ , correspondant aux maxima journaliers (à gauche) et hebdomadaires (à droite) de  $CH_4$ . L'unité est le ppb.

Il convient dans un premier temps de s'assurer que l'ajustement du modèle proposé dans ce chapitre et défini dans la Proposition 4.1 est raisonnable et plus adéquat que le modèle classique gaussien par exemple. C'est dans ce but que nous présentons les histogrammes en Figure 4.9. Ceux hachurés correspondent aux données observées de maxima de méthane ; en haut, les maxima journaliers et en bas les hebdomadaires. Ensuite nous effectuons 1000 estimations de  $X_{t+1}|X_t = x$ ,  $t = 1, \dots, T-1$  selon le modèle présenté en Proposition 4.1 avec  $x$  prenant tour à tour chaque valeur observée. Pour cela nous avons estimé les paramètres à partir de la totalité des données et selon (4.7), (4.8) et (4.9). Nous obtenons ainsi les histogrammes non hachurés de la Figure 4.9 (a). De manière analogue, nous effectuons 1000 estimations de  $X_{t+1}|X_t = x$ ,  $t = 1, \dots, T-1$ , selon le modèle gaussien, i.e. avec  $\widehat{X}_{t+1} = \widehat{\alpha}x + \theta_{t+1} + \widehat{\nu}(1 - \widehat{\alpha})$  où  $\widehat{\nu} = \overline{X}$  et  $\theta_{t+1}$  est une variable aléatoire gaussienne d'espérance nulle et de variance  $\widehat{\sigma}_\theta^2 = (\overline{X^2} - \overline{X}^2) \times (1 - \widehat{\alpha}^2)$ . Cela conduit alors aux histogrammes non hachurés de la Figure 4.9 (b). Au vu de ces histogrammes, nous pouvons conclure que le modèle proposé est beaucoup mieux adapté qu'un modèle gaussien classique que ce soit pour les maxima journaliers ou hebdomadaires. L'une des raisons pour cela est la prise en compte de la structure asymétrique de la distribution des



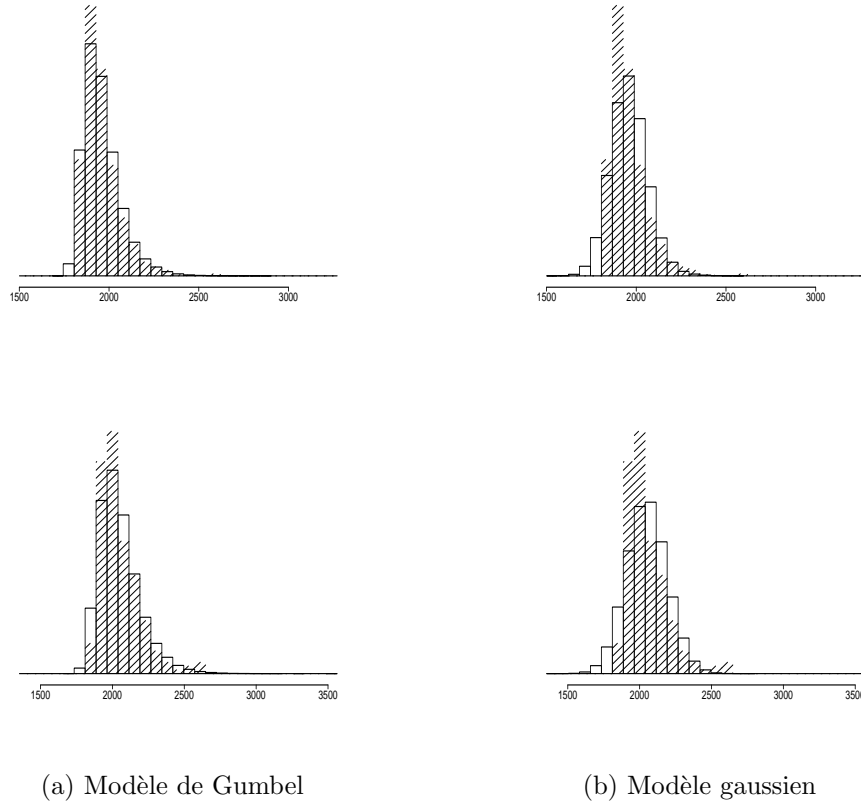


FIG. 4.9 – *Histogrammes des prévisions à un jour des maxima journaliers (en haut) et hebdomadaires (en bas) de  $CH_4$  selon (a) le modèle proposé et selon (b) le modèle gaussien. Les histogrammes hachurés représentent les données de maxima journaliers et hebdomadaires de  $CH_4$ .*

données par notre modèle.

Pour valider plus précisément le modèle proposé, nous avons tout d'abord estimé les trois paramètres présents dans le modèle,  $\mu$ ,  $\sigma$  et  $\alpha$  sur une première période, soit ici de 2002 jusqu'à mi-2006. Dans un second temps, pour la seconde partie de l'année 2006, nous avons calculé 1000  $\widehat{X}_{t+1} = \widehat{\alpha}x_t + \widehat{\alpha}\widehat{\sigma} \log S_{t+1} + \widehat{\mu}(1 - \widehat{\alpha})$  avec  $x_t$  la valeur observée au temps  $t$  et  $S_{t+1}$  une variable aléatoire  $\widehat{\alpha}$ -stable positive définie par (4.2). Enfin, nous en avons déduit les quartiles empiriques de la distribution de  $\widehat{X}_{t+1}|X_t = x_t$ . La Figure 4.10 représente la prévision de maxima journaliers (à gauche) et hebdomadaires (à droite) de

méthane à Gif-sur-Yvette durant le second semestre de 2006. Nous pouvons sur cette figure observer la médiane des estimations en tiret, les premier et troisième quartiles délimitant la zone grisée ainsi que les valeurs réellement observées en trait plein. Au premier abord, la Figure 4.10 concernant la prévision à un jour des maxima journaliers est plutôt bonne. En revanche, celle des maxima hebdomadaires est moins satisfaisante, mais cela n'a rien de très surprenant : en effet, il faut garder en mémoire que les estimations des paramètres sont basées sur une taille d'échantillon assez faible (inférieure à 300) et par ailleurs, pour les maxima hebdomadaires, la dépendance est moins forte (cf. Figure 4.8).

Quoiqu'il en soit, le message que nous avons cherché à faire passer ici, c'est que bien que notre modèle peine à prévoir les pics, il se comporte en dehors de cela plutôt bien. Nous n'avons pas la prétention d'affirmer que notre approche résout tous les problèmes, mais elle apporte une première solution dans un contexte où aucune autre n'existait à ce jour.

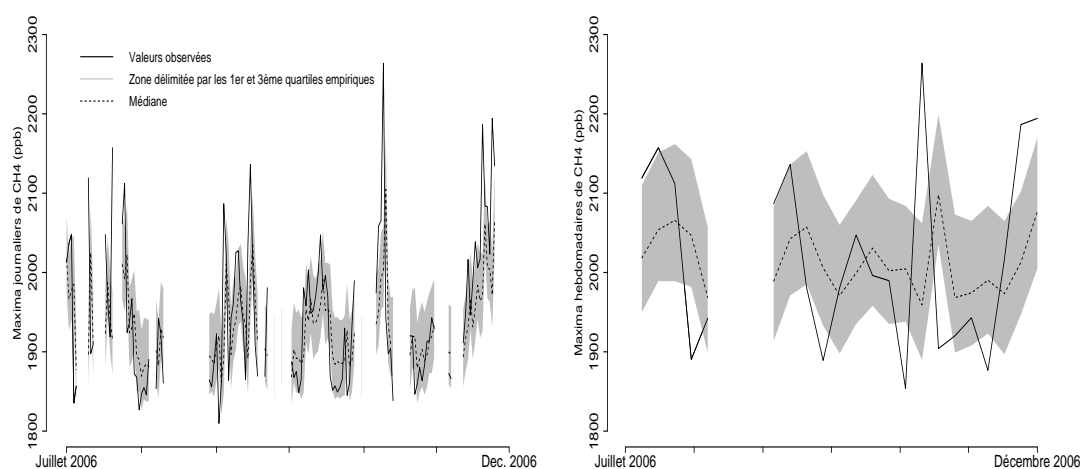


FIG. 4.10 – Validation de la prévision à un jour des maxima journaliers de  $CH_4$  (à gauche) et hebdomadaires (à droite) sur le second semestre 2006.

## b. Les maxima journaliers et hebdomadaires de dioxyde de carbone

Nous disposons de données sur 21 années et nous avons moins de valeurs manquantes

que pour le CH<sub>4</sub>. Plus précisément, le pourcentage de valeurs manquantes est de 11.8% pour les maxima journaliers et de 7.8% pour les maxima hebdomadaires. Nous pouvons visualiser ces données sur la Figure 4.11. Au vu de la Figure 4.12, il est assez raisonnable de supposer que ces maxima proviennent également d'une distribution de Gumbel. De

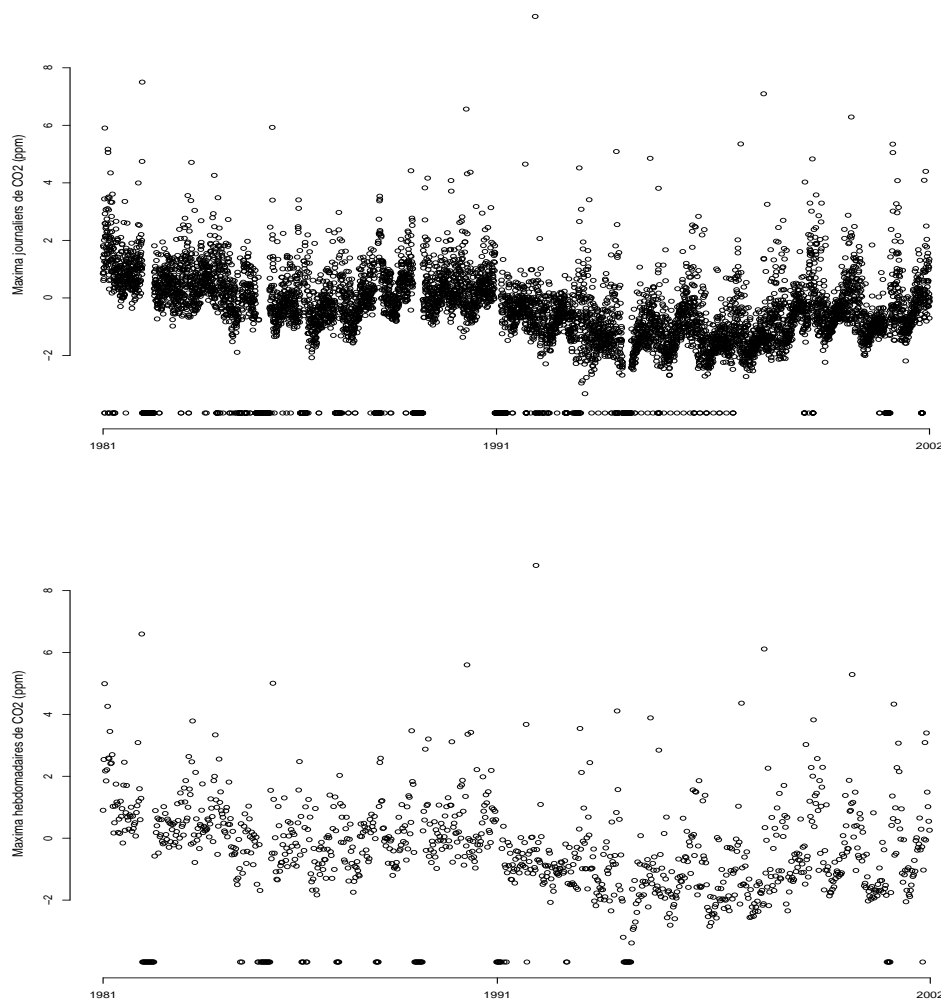


FIG. 4.11 – Séries des maxima journaliers (en haut) et hebdomadaires (en bas) de CO<sub>2</sub> à Gif-sur-Yvette avec valeurs manquantes.

plus nous sommes en présence d'une très forte dépendance temporelle comme en témoigne la Figure 4.13. Les histogrammes de la Figure 4.14 illustrent l'avantage de l'utilisation du modèle proposé par rapport à un modèle gaussien bien qu'ils soient moins convaincants

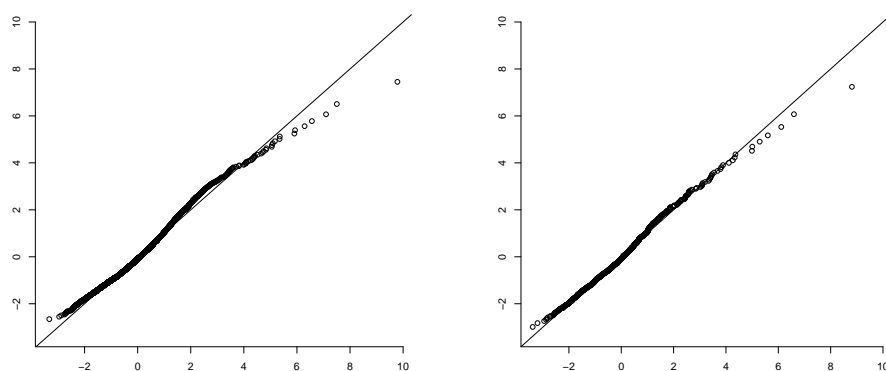


FIG. 4.12 – « QQplot » pour la loi de Gumbel des maxima journaliers (à gauche) et hebdomadaires (à droite) de  $CO_2$ . Les paramètres correspondants sont estimés par la méthode des moments. L'unité est le ppm.

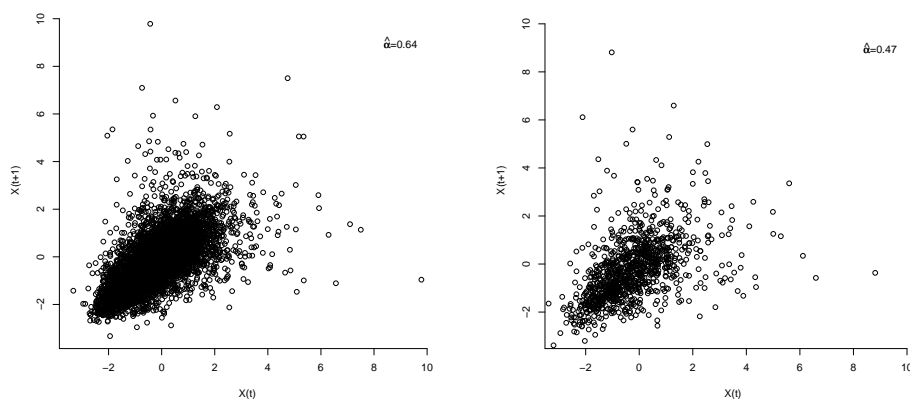


FIG. 4.13 – « Scatter plots » des valeurs successives, i.e.  $(X_t, X_{t+1})$ , correspondant aux maxima journaliers (à gauche) et hebdomadaires (à droite) de  $CO_2$ . L'unité est le ppm.



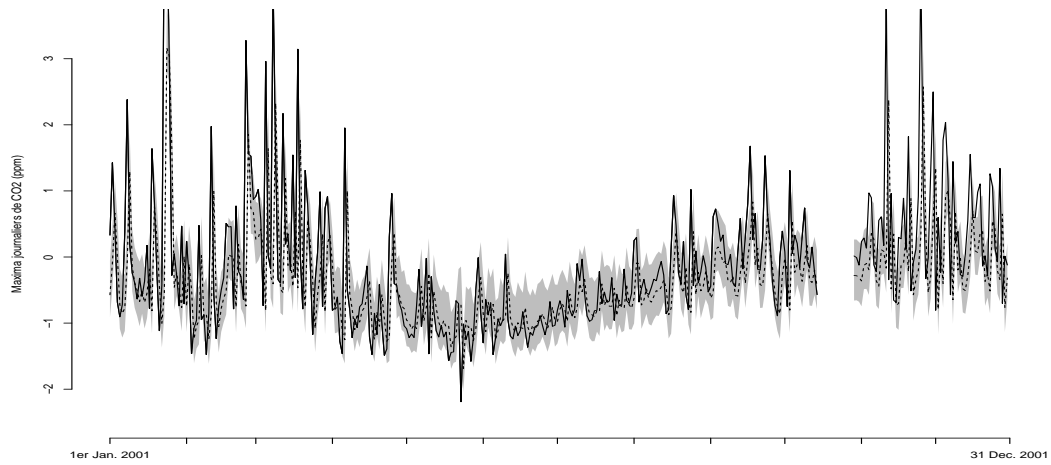


FIG. 4.15 – Validation de la prévision à un jour des maxima journaliers de  $CO_2$  sur l'année 2001. La légende est la même que pour la Figure 4.10.

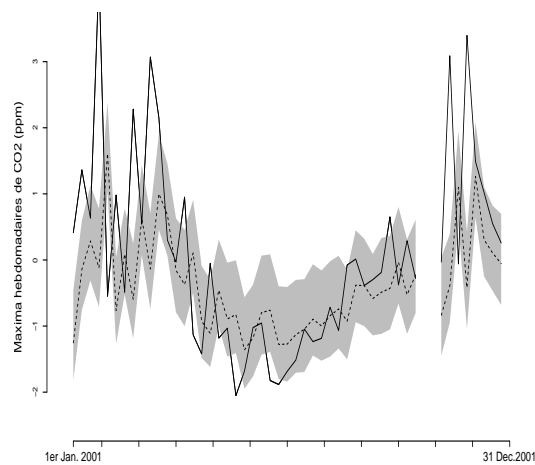


FIG. 4.16 – Validation de la prévision à un jour des maxima hebdomadaires de  $CO_2$  sur l'année 2001. La légende est la même que pour la Figure 4.10.

Nous allons montrer dans la section suivante que nous pouvons l'étendre au cas d'une distribution GEV.

## 4.5 Généralisation à la distribution GEV

Comme nous allons le voir dans la remarque suivante, un simple changement de variable est suffisant pour étendre notre modèle à tous types de maxima. Nous proposons alors un modèle de récurrence pour la distribution généralisée des valeurs extrêmes (GEV).

**Remarque 4.1** Soit  $E$  une variable aléatoire distribuée selon la distribution GEV de paramètres  $\mu$ ,  $\sigma$  et  $\gamma$ . Si  $\gamma$  est strictement négatif, alors  $-\log(\mu - \sigma/\gamma - E)$  suit une distribution de Gumbel de paramètre de localisation  $\log(-\gamma/\sigma)$  et de paramètre d'échelle égal à  $-\gamma$ .

De la même manière, si  $\gamma$  est strictement positif, alors  $\log(E - \mu + \sigma/\gamma)$  suit une distribution de Gumbel de paramètre de localisation égal à  $\log(\sigma/\gamma)$  et de paramètre d'échelle  $\gamma$ .

Nous sommes donc à même d'étendre notre modèle au cas d'une distribution GEV.

**Proposition 4.5** Soient  $S_t$ ,  $t \in \mathbb{Z}$ , des variables aléatoires indépendantes et de même loi que  $S$  définie en (4.2). Soit  $\{X_t, t \in \mathbb{Z}\}$  un processus défini par la relation de récurrence suivante

$$X_t = \mu - \frac{\sigma}{\gamma} + \left( X_{t-1} - \mu + \frac{\sigma}{\gamma} \right)^\alpha \times S_t^{\alpha\gamma} \times \left( \frac{\sigma}{\gamma} \right)^{1-\alpha} \quad (4.11)$$

où  $(\mu, \sigma, \gamma) \in \mathbb{R} \times \mathbb{R}_*^+ \times \mathbb{R}_*$ .

L'équation (4.11) a une unique solution strictement stationnaire donnée par

$$X_t = \mu - \frac{\sigma}{\gamma} + \frac{\sigma}{\gamma} \prod_{j=0}^{\infty} (S_{t-j})^{\gamma\alpha^{j+1}} \quad (4.12)$$

et  $X_t$  est distribuée selon une GEV( $\mu, \sigma, \gamma$ ),  $\forall t \in \mathbb{Z}$ .

En pratique, ce modèle peut être utilisé lorsque les maxima ne sont pas distribués selon une loi de Gumbel mais selon une distribution GEV, et de la même manière il sera alors possible d'estimer les différents paramètres.

## 4.6 Preuves des résultats

### 4.6.1 Preuve de la Proposition 4.1

En posant  $Y_t = X_t - (\mu + \delta\sigma)$  et  $\varepsilon_t = \alpha\sigma \log S_t - \delta\sigma(1 - \alpha)$  où  $\delta$  est la constante d'Euler, (4.3) peut se réécrire de la manière suivante :

$$Y_t = \alpha Y_{t-1} + \varepsilon_t. \quad (4.13)$$

Il s'agit là d'un modèle bien connu : un processus autorégressif à temps discret d'ordre un (AR(1)) avec des variables aléatoires  $\varepsilon_t$  qui sont indépendantes et identiquement distribuées d'espérance nulle et de variance  $\alpha^2\sigma^2\sigma_\varepsilon^2$ . En effet, selon Zolotarev [71] (Section 3.6),  $\mathbb{E}(\log S) = \delta(1/\alpha - 1)$  et  $\text{Var}(\log S) = (\pi^2/6) \times (1/\alpha^2 - 1) =: \sigma_\varepsilon^2$ . Si nous nous référons aux résultats standards concernant les processus AR(1) détaillés par exemple dans Brockwell et Davis [13] (Section 3.1), nous avons dans le cas où  $|\alpha| < 1$  et sous l'hypothèse que  $\{Y_t\}$  est stationnaire, d'une part que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( Y_t - \sum_{j=0}^n \alpha^j \varepsilon_{t-j} \right)^2 = 0 \quad (4.14)$$

et, d'autre part que  $\sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$  est convergent en moyenne quadratique. Par conséquent, le processus  $\{Y_t\}$  a une unique solution stationnaire du second-ordre

$$Y_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}. \quad (4.15)$$

Comme il est clair que (4.15) est équivalent à (4.4), le processus  $\{X_t\}$  défini en (4.4) est l'unique solution stationnaire du second-ordre de (4.3).

Dans le but d'obtenir la distribution de  $X_t$ , nous nous sommes intéressés à la fonction



caractéristique de  $\log S$ . D'après Zolotarev [71] (page 117), la fonction caractéristique de  $\log S$  est définie par

$$\mathbb{E}(\exp(iu \log S)) = \frac{\Gamma\left(1 - \frac{iu}{\alpha}\right)}{\Gamma(1 - iu)}. \quad (4.16)$$

D'après (4.14) nous avons

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( X_t - \mu - \sigma \sum_{j=0}^n \alpha^{j+1} \log S_{t-j} - \delta \sigma \alpha^{n+1} \right)^2 = 0, \quad (4.17)$$

ce qui implique que  $\mu + \sigma \sum_{j=0}^n \alpha^{j+1} \log S_{t-j} + \delta \sigma \alpha^{n+1}$  converge en probabilité et donc en distribution vers  $X_t$ , quand  $n \rightarrow \infty$ .

Par conséquent, en utilisant le lemme du Portmanteau (cf. van der Vaart [69], Lemme 2.2)

$$\mathbb{E}(e^{iuX_t}) = \lim_{n \rightarrow \infty} \mathbb{E} \left( e^{iu(\mu + \sigma \sum_{j=0}^n \alpha^{j+1} \log S_{t-j} + \delta \sigma \alpha^{n+1})} \right) = \lim_{n \rightarrow \infty} e^{iu\mu} e^{iu\delta \sigma \alpha^{n+1}} \mathbb{E} \left( \prod_{j=0}^n e^{iu\sigma \alpha^{j+1} \log S_{t-j}} \right).$$

Comme les variables  $S_t$ ,  $t \in \mathbb{Z}$ , sont indépendantes, nous avons

$$\begin{aligned} \mathbb{E}(e^{iuX_t}) &= \lim_{n \rightarrow \infty} e^{iu\mu} e^{iu\delta \sigma \alpha^{n+1}} \prod_{j=0}^n \mathbb{E} \left( e^{iu\sigma \alpha^{j+1} \log S_{t-j}} \right) \\ &= \lim_{n \rightarrow \infty} e^{iu\mu} e^{iu\delta \sigma \alpha^{n+1}} \prod_{j=0}^n \frac{\Gamma(1 - iu\sigma \alpha^j)}{\Gamma(1 - iu\sigma \alpha^{j+1})} \\ &= \lim_{n \rightarrow \infty} e^{iu\mu} e^{iu\delta \sigma \alpha^{n+1}} \frac{\Gamma(1 - iu\sigma)}{\Gamma(1 - iu\sigma \alpha^{n+1})} \\ &= e^{iu\mu} \Gamma(1 - iu\sigma). \end{aligned}$$

La fonction caractéristique de  $X_t$  correspond donc exactement à celle d'une variable aléatoire de loi Gumbel  $(\mu, \sigma)$ .

De plus, comme les variables  $S_t$ , pour tout entier  $t$ , sont indépendantes et identiquement distribuées, le processus  $X_t$  est non seulement stationnaire du second ordre mais aussi strictement stationnaire (cf. 4.1.1 dans Bosq et Lecoutre [12]).

### 4.6.2 Preuve de la Proposition 4.2

Les estimateurs  $\hat{\mu}$  et  $\hat{\sigma}$  définis en (4.7) et (4.8) des paramètres  $\mu$  et  $\sigma$  eux-mêmes définis en (4.5) et (4.6) sont basés sur la méthode des moments. Par le théorème ergodique,

$$\begin{aligned}\bar{X} &= \frac{1}{T} \sum_{t=1}^T X_t \xrightarrow{p.s.} \mathbb{E}(X_0) \\ s^2 &= \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2 \xrightarrow{p.s.} \text{Var}(X_0).\end{aligned}$$

Par conséquent par continuité, il découle que  $\hat{\mu} \xrightarrow{p.s.} \mu$  et  $\hat{\sigma} \xrightarrow{p.s.} \sigma$ .

Des arguments similaires peuvent être utilisés pour montrer la convergence de  $\hat{\alpha}$ . En effet, par le théorème ergodique,

$$\frac{1}{T} \sum_{t=1}^{T-1} (X_t - \bar{X})(X_{t+1} - \bar{X}) \xrightarrow{p.s.} \text{Cov}(X_0, X_1) = \alpha \text{Var}(X_0).$$

Par conséquent, un estimateur naturel de  $\alpha$  est :

$$\hat{\alpha} = \frac{\frac{1}{T} \sum_{t=1}^{T-1} (X_t - \bar{X})(X_{t+1} - \bar{X})}{s^2} \xrightarrow{p.s.} \alpha.$$

### 4.6.3 Preuve de la Proposition 4.3

Soit  $\alpha \in (0, 1)$ . Définissons par  $\{Y_t\}$  le processus moyenne mobile suivant :

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \quad (4.18)$$

où  $\psi_j = \sigma \alpha^{j+1}$  pour  $j \geq 0$  et 0 sinon. Soit  $\varepsilon_t = \log S_t - \frac{\delta}{\alpha}(1-\alpha)$  avec  $S_t$  définie comme dans la Proposition 4.1. Nous en déduisons que les variables aléatoires  $\varepsilon_t$  sont indépendantes et identiquement distribuées de moyenne nulle et de variance égale à  $\sigma_\varepsilon^2 = (\pi^2/6) \times (1/\alpha^2 - 1)$ .

Pour établir la Proposition 4.3 nous allons avoir besoin du Lemme suivant, qui, combiné avec la méthode delta nous permettra de conclure.

**Lemme 4.2** Soit  $\{Y_t\}$  défini par (4.18). Le vecteur aléatoire

$$\sqrt{T} \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T Y_t \\ \frac{1}{T} \sum_{t=1}^T (Y_t^2 - \mathbb{E}Y_t^2) \\ \frac{1}{T} \sum_{t=1}^T (Y_t Y_{t+1} - \mathbb{E}Y_t Y_{t+1}) \end{pmatrix} \quad (4.19)$$

converge vers une distribution gaussienne de vecteur moyenne  $\mathbf{0}$  et de matrice de variance-covariance  $\Sigma_{\mathbf{Y}}$  égale à

$$\begin{pmatrix} \xi_0 \frac{1+\alpha}{1-\alpha} & \frac{2\sigma^3 \zeta(3)(1+\alpha+\alpha^2)}{1-\alpha^2} & \frac{2\sigma^3 \zeta(3)\alpha(1+\alpha+\alpha^2)}{1-\alpha^2} \\ \frac{2\sigma^3 \zeta(3)(1+\alpha+\alpha^2)}{1-\alpha^2} & \frac{22}{5} \xi_0^2 \frac{1+\alpha^2}{1-\alpha^2} & \frac{4\alpha \xi_0^2}{1-\alpha^2} \left(1 + \frac{3}{5}(1+\alpha^2)\right) \\ \frac{2\sigma^3 \zeta(3)\alpha(1+\alpha+\alpha^2)}{1-\alpha^2} & \frac{4\alpha \xi_0^2}{1-\alpha^2} \left(1 + \frac{3}{5}(1+\alpha^2)\right) & \frac{\xi_0^2}{1-\alpha^2} [(1+\alpha^2)\left(\frac{12}{5}\alpha^2 + 1\right) + \alpha^2(3-\alpha^2)] \end{pmatrix} \quad (4.20)$$

où  $\zeta(\cdot)$  est la fonction de Riemann et  $\xi_k = \xi(k)$  avec  $\xi(\cdot)$  la fonction d'autocovariance de  $\{Y_t\}$ .

Nous cherchons à établir le comportement limite de

$$\sqrt{T} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \\ \hat{\alpha} - \alpha \end{pmatrix}$$

avec  $\hat{\mu}$ ,  $\hat{\sigma}$  et  $\hat{\alpha}$  définis en (4.7), (4.8) et (4.9). Dans ce but, nous allons étudier celui de

$$\sqrt{T} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \\ \hat{\alpha} - \alpha \end{pmatrix} \text{ avec } \tilde{\alpha} = \frac{\frac{1}{T} \sum_{t=1}^T (Y_t Y_{t+1} - \bar{Y}^2)}{\frac{1}{T} \sum_{t=1}^T (Y_t - \bar{Y})^2}.$$

Pour cela nous allons combiner le Lemme 4.2 avec la méthode delta appliquée aux fonctions

$$\begin{cases} \phi_1(\mathbf{z}) & := z_1 - \frac{\delta\sqrt{6}}{\pi} \sqrt{z_2 - z_1^2} \\ \phi_2(\mathbf{z}) & := \frac{\sqrt{6}}{\pi} \sqrt{z_2 - z_1^2} \\ \phi_3(\mathbf{z}) & := \frac{z_3 - z_1^2}{z_2 - z_1^2} \end{cases}$$

où  $\mathbf{z}^t = (z_1, z_2, z_3)$ .

Les dérivées partielles de  $\phi_1$ ,  $\phi_2$  et  $\phi_3$  existent, sont continues et  $H(\mu, \sigma, \alpha)$  défini comme

suit

$$H(\mu, \sigma, \alpha) = \frac{\partial \phi_\mu(\mathbf{z})}{\partial z_\nu} \Big|_{(z_1, z_2, z_3) = (0, \mathbb{E}(Y^2), \mathbb{E}(Y_0 Y_1))}, \quad \mu, \nu = 1, 2, 3$$

est égal à

$$\begin{pmatrix} 1 & \frac{-\sigma\delta}{2\xi_0} & 0 \\ 0 & \frac{\sigma}{2\xi_0} & 0 \\ 0 & \frac{-\alpha}{\xi_0} & \frac{1}{\xi_0} \end{pmatrix}$$

et est de rang 3.

Par conséquent nous obtenons par la méthode delta la normalité asymptotique de

$$\sqrt{T} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \\ \hat{\alpha} - \alpha \end{pmatrix}$$

avec un vecteur moyenne égal à  $\mathbf{0}$  et une matrice de variance-covariance égale à  $H(\mu, \sigma, \alpha) \Sigma_{\mathbf{Y}} H(\mu, \sigma, \alpha)^t$  où  $\Sigma_{\mathbf{Y}}$  est définie dans le Lemme 4.2. Des calculs directs nous conduisent alors à la matrice (4.10).

Par ailleurs, comme il est possible de montrer que  $\sqrt{T}(\hat{\alpha} - \tilde{\alpha}) = o_{\mathbb{P}}(1)$  avec  $\hat{\alpha}$  défini en (4.9), nous déduisons par le théorème de Slutsky que le vecteur

$$\sqrt{T} \begin{pmatrix} \hat{\mu} - \mu \\ \hat{\sigma} - \sigma \\ \hat{\alpha} - \alpha \end{pmatrix}$$

converge vers la même distribution limite.

Reste alors à établir le lemme.

### Preuve du Lemme 4.2

Dans une première partie, nous allons calculer les éléments de  $\Sigma_{\mathbf{Y}}$ . Comme il est possible de montrer que  $\mathbb{E}(\varepsilon_t^4) = \eta\sigma_\varepsilon^4 < \infty$  avec

$$\eta = 3 + \frac{12}{5} \frac{1 + \alpha^2}{1 - \alpha^2}$$

et comme  $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ , nous en déduisons directement la matrice de variance-covariance asymptotique du vecteur  $\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T Y_t^2, \frac{1}{\sqrt{T}} \sum_{t=1}^T Y_t Y_{t+1}\right)^t$  en utilisant la Proposition 7.3.1 dans Brockwell et Davis [13].

Par des techniques similaires, il est également possible de calculer les autres termes de la matrice  $\Sigma_{\mathbf{Y}}$ . Par exemple, une application directe du lemme de Kronecker nous conduit à

$$\lim_{T \rightarrow \infty} T \text{Cov} \left( \frac{1}{T} \sum_{t=1}^T Y_t, \frac{1}{T} \sum_{s=1}^T Y_s \right) = \lim_{T \rightarrow \infty} \xi_0 + 2 \sum_{k=1}^T \left(1 - \frac{k}{T}\right) \xi_k = \sum_{k=-\infty}^{\infty} \xi_k.$$

Finalement, nous obtenons que

$$\begin{aligned} & \lim_{T \rightarrow \infty} T \text{Var} \begin{pmatrix} \frac{1}{T} \sum_{t=1}^T Y_t \\ \frac{1}{T} \sum_{t=1}^T Y_t^2 \\ \frac{1}{T} \sum_{t=1}^T Y_t Y_{t+1} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{k=-\infty}^{\infty} \xi_k & \sum_{k=-\infty}^{\infty} \mathbb{E}(Y_0 Y_k^2) & \sum_{k=-\infty}^{\infty} \mathbb{E}(Y_0 Y_k Y_{k+1}) \\ \sum_{k=-\infty}^{\infty} \mathbb{E}(Y_0 Y_k^2) & (\eta - 3)\xi_0^2 + 2 \sum_{k=-\infty}^{\infty} \xi_k^2 & (\eta - 3)\xi_0 \xi_1 + 2 \sum_{k=-\infty}^{\infty} \xi_k \xi_{k+1} \\ \sum_{k=-\infty}^{\infty} \mathbb{E}(Y_0 Y_k Y_{k+1}) & (\eta - 3)\xi_0 \xi_1 + 2 \sum_{k=-\infty}^{\infty} \xi_k \xi_{k+1} & (\eta - 3)\xi_1^2 + \sum_{k=-\infty}^{\infty} (\xi_k^2 + \xi_{k+1} \xi_{k-1}) \end{pmatrix}. \end{aligned}$$

Comme  $\mathbb{E}(\varepsilon_t^3) = 2\zeta(3)(1/\alpha^3 - 1)$  et  $\xi_k = \xi_0 \alpha^{|k|}$ , nous obtenons pour  $\Sigma_{\mathbf{Y}}$  l'expression (4.20).

Dans la seconde partie de la preuve du Lemme 4.2, nous devons montrer la convergence vers une distribution gaussienne. Pour cela, définissons la suite « tronquée » suivante :

$$\mathbf{W}_{\mathbf{m},t} = \begin{pmatrix} Y_{m,t} \\ Y_{m,t}^2 - \mathbb{E}Y_{m,t}^2 \\ Y_{m,t} Y_{m,t+1} - \mathbb{E}Y_{m,t} Y_{m,t+1} \end{pmatrix}$$

où

$$Y_{m,t} = \sum_{j=-m}^m \psi_j \varepsilon_{t-j} \quad (4.21)$$

et  $\psi_j$  défini comme précédemment.

L'idée est de montrer dans un premier temps la normalité asymptotique de  $T^{-1/2} \sum_{t=1}^T \mathbf{W}_{\mathbf{m},t}$  et ensuite de faire tendre  $m$  vers l'infini. Pour cela, nous devons montrer

que toute combinaison linéaire des trois composantes de  $T^{-1/2} \sum_{t=1}^T \mathbf{W}_{\mathbf{m},t}$  est gaussienne. Notons que pour tout  $\lambda \in \mathbb{R}^3$ , la suite  $\{\lambda \mathbf{W}_{\mathbf{m},t}\}$  est strictement stationnaire  $(2m+1)$ -dépendante. De plus

$$\lim_{T \rightarrow \infty} T^{-1} \text{Var} \left( \sum_{t=1}^T \lambda^t \mathbf{W}_{\mathbf{m},t} \right) = \lambda^t \Sigma_{\mathbf{Y}_m} \lambda$$

où  $\Sigma_{\mathbf{Y}_m}$  est définie comme

$$\begin{pmatrix} \sum_{k=-\infty}^{\infty} \xi_{m,k} & \sum_{k=-\infty}^{\infty} \mathbb{E}(Y_{m,0} Y_{m,k}^2) & \sum_{k=-\infty}^{\infty} \mathbb{E}(Y_{m,0} Y_{m,k} Y_{m,k+1}) \\ \sum_{k=-\infty}^{\infty} \mathbb{E}(Y_{m,0} Y_{m,k}^2) & (\eta-3)\xi_{m,0}^2 + 2 \sum_{k=-\infty}^{\infty} \xi_{m,k}^2 & (\eta-3)\xi_{m,0}\xi_{m,1} + 2 \sum_{k=-\infty}^{\infty} \xi_{m,k}\xi_{m,k+1} \\ \sum_{k=-\infty}^{\infty} \mathbb{E}(Y_{m,0} Y_{m,k} Y_{m,k+1}) & (\eta-3)\xi_{m,0}\xi_{m,1} + 2 \sum_{k=-\infty}^{\infty} \xi_{m,k}\xi_{m,k+1} & (\eta-3)\xi_{m,1}^2 + \sum_{k=-\infty}^{\infty} (\xi_{m,k}^2 + \xi_{m,k+1}\xi_{m,k-1}) \end{pmatrix}$$

avec  $\xi_{m,k} = \xi_m(k)$  où  $\xi_m(\cdot)$  est la fonction d'autocovariance de  $\{Y_{m,t}\}$ .

Par conséquent, nous pouvons directement appliquer le Théorème 6.4.2 dans Brockwell et Davis [13] et nous obtenons que

$$T^{-1/2} \sum_{t=1}^T \lambda^t \mathbf{W}_{\mathbf{m},t} \xrightarrow{d} \Theta_m \text{ avec } \Theta_m \sim \mathcal{N}(\mathbf{0}, \lambda^t \Sigma_{\mathbf{Y}_m} \lambda)$$

pour tout vecteur  $\lambda \in \mathbb{R}^3$  tel que  $\lambda^t \Sigma_{\mathbf{Y}_m} \lambda > 0$ .

Cela implique que

$$T^{-1/2} \sum_{t=1}^T \mathbf{W}_{\mathbf{m},t} \xrightarrow{d} \Omega_m \text{ avec } \Omega_m \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{Y}_m}). \quad (4.22)$$

La dernière étape dans la preuve de ce lemme est de montrer que la normalité asymptotique décrite par (4.22) reste satisfaite si  $\mathbf{W}_{\mathbf{m},t}$  est remplacé par  $\mathbf{W}_t$  défini comme suit :

$$\mathbf{W}_t = \begin{pmatrix} Y_t \\ Y_t^2 - \mathbb{E}Y_t^2 \\ Y_t Y_{t+1} - \mathbb{E}Y_t Y_{t+1} \end{pmatrix}.$$

L'idée pour cela est d'établir le résultat pour  $\mathbf{W}_t$  en faisant tendre  $m$  vers l'infini. En utilisant principalement le théorème de convergence dominée, il est possible de montrer que  $\Sigma_{\mathbf{Y}_m}$  converge vers  $\Sigma_{\mathbf{Y}}$  quand  $m$  tend vers l'infini, ce qui implique que  $\Omega_m \xrightarrow{d} \Omega$  quand  $m$  tend vers l'infini avec  $\Omega \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{Y}})$ .

La preuve du lemme peut maintenant se terminer de la façon suivante. Nous appliquons la Proposition 6.3.9 de Brockwell et Davis [13] comme dans la preuve de leur Proposition 7.3.3. Pour cela, il faut encore vérifier une condition. D'après la Proposition 7.3.3, pour  $p = 0, 1$ , la limite

$$\lim_{m \rightarrow \infty} \limsup_{T \rightarrow \infty} \mathbb{P} \left( \sqrt{T} \left| \frac{1}{T} \sum_{t=1}^T (Y_{m,t} Y_{m,t+p} - \mathbb{E}(Y_{m,t} Y_{m,t+p})) - \frac{1}{T} \sum_{t=1}^T (Y_t Y_{t+p} - \mathbb{E}(Y_t Y_{t+p})) \right| > \varepsilon \right)$$

est nulle. De la même manière, nous devons montrer que

$$\lim_{m \rightarrow \infty} \limsup_{T \rightarrow \infty} \mathbb{P} \left( \sqrt{T} \left| \frac{1}{T} \sum_{t=1}^T Y_{m,t} - \frac{1}{T} \sum_{t=1}^T Y_t \right| > \varepsilon \right) = 0. \quad (4.23)$$

En utilisant l'inégalité de Bienaymé-Tchebychev, nous avons que

$$\begin{aligned} & \mathbb{P} \left( T^{1/2} \left| T^{-1} \sum_{t=1}^T Y_{m,t} - T^{-1} \sum_{t=1}^T Y_t \right| > \varepsilon \right) \\ & \leq \varepsilon^{-2} T \mathbb{V}ar \left( T^{-1} \sum_{t=1}^T Y_{m,t} - T^{-1} \sum_{t=1}^T Y_t \right) \\ & = \varepsilon^{-2} T \left[ \mathbb{V}ar \left( T^{-1} \sum_{t=1}^T Y_{m,t} \right) + \mathbb{V}ar \left( T^{-1} \sum_{t=1}^T Y_t \right) - 2 \mathbb{C}ov \left( T^{-1} \sum_{t=1}^T Y_{m,t}, T^{-1} \sum_{t=1}^T Y_t \right) \right]. \end{aligned}$$

Les limites des deux variances et de la covariance du membre de droite de l'inégalité précédente peuvent facilement être calculées. En effet

$$\lim_{m \rightarrow \infty} \lim_{T \rightarrow \infty} T \mathbb{V}ar \left( T^{-1} \sum_{t=1}^T Y_{m,t} \right) = \lim_{T \rightarrow \infty} T \mathbb{V}ar \left( T^{-1} \sum_{t=1}^T Y_t \right) = \xi_0 \frac{1 + \alpha}{1 - \alpha}$$

et de manière similaire, nous avons que

$$\lim_{m \rightarrow \infty} \lim_{T \rightarrow \infty} T \mathbb{C}ov \left( T^{-1} \sum_{t=1}^T Y_{m,t}, T^{-1} \sum_{t=1}^T Y_t \right) = \xi_0 \frac{1 + \alpha}{1 - \alpha}$$

d'où

$$\lim_{m \rightarrow \infty} \limsup_{T \rightarrow \infty} \varepsilon^{-2} T \mathbb{V}ar \left( T^{-1} \sum_{t=1}^T Y_{m,t} - T^{-1} \sum_{t=1}^T Y_t \right) = 0.$$

Ceci permet de déduire (4.23) et achève ainsi la preuve du Lemme 4.2.

#### 4.6.4 Preuve du Lemme 4.1

Pour tout  $h > 0$ , soient

$$\mathbf{X}_h = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-k} \\ \vdots \\ X_{t-h} \end{pmatrix} = \begin{pmatrix} \alpha^h X_{t-h} + \sigma \sum_{j=0}^{h-1} \alpha^{j+1} \log S_{t-j} + \mu(1-\alpha) \sum_{j=0}^{h-1} \alpha^j \\ \alpha^{h-1} X_{t-h} + \sigma \sum_{j=1}^{h-1} \alpha^j \log S_{t-j} + \mu(1-\alpha) \sum_{j=0}^{h-2} \alpha^j \\ \vdots \\ \alpha^{h-k} X_{t-h} + \sigma \sum_{j=k}^{h-1} \alpha^{j-k+1} \log S_{t-j} + \mu(1-\alpha) \sum_{j=0}^{h-k-1} \alpha^j \\ \vdots \\ X_{t-h} \end{pmatrix}$$

et  $u = (u_0, u_1, \dots, u_h)^t$ .

Il suffit de calculer la fonction caractéristique associée à  $\mathbf{X}_h$  :

$$\begin{aligned} \mathbb{E}(e^{i\langle u, \mathbf{X}_h \rangle}) &= \mathbb{E}(e^{iu_0 X_t + iu_1 X_{t-1} + \dots + iu_k X_{t-k} + \dots + iu_h X_{t-h}}) \\ &= \mathbb{E}\left(e^{iu_0 (\alpha^h X_{t-h} + \sigma \sum_{j=0}^{h-1} \alpha^{j+1} \log S_{t-j} + \mu(1-\alpha) \sum_{j=0}^{h-1} \alpha^j)} \right. \\ &\quad \left. + \dots + iu_k (\alpha^{h-k} X_{t-h} + \sigma \sum_{j=k}^{h-1} \alpha^{j-k+1} \log S_{t-j} + \mu(1-\alpha) \sum_{j=0}^{h-k-1} \alpha^j) + \dots + iu_h X_{t-h}\right) \\ &= \mathbb{E}\left(e^{iX_{t-h} \sum_{j=0}^h u_j \alpha^{h-j}}\right) \mathbb{E}\left(e^{i\sigma \sum_{j=0}^{h-1} (\sum_{k=0}^j \alpha^{j-k+1} u_k) \log S_{t-j}}\right) e^{i\mu \sum_{j=0}^h u_j (1-\alpha^{h-j})} \\ &= e^{i\mu \sum_{j=0}^h u_j \alpha^{h-j}} \Gamma\left(1 - i\sigma \sum_{j=0}^h u_j \alpha^{h-j}\right) \prod_{j=0}^{h-1} \mathbb{E}\left(e^{i\sigma (\sum_{k=0}^j \alpha^{j-k+1} u_k) \log S_{t-j}}\right) e^{i\mu \sum_{j=0}^h u_j (1-\alpha^{h-j})} \\ &= e^{i\mu \sum_{j=0}^h u_j} \Gamma\left(1 - i\sigma \sum_{j=0}^h u_j \alpha^{h-j}\right) \prod_{j=0}^{h-1} \frac{\Gamma\left(1 - i\sigma \sum_{k=0}^j u_k \alpha^{j-k}\right)}{\Gamma\left(1 - i\sigma \sum_{k=0}^j u_k \alpha^{j-k+1}\right)}. \end{aligned}$$

#### 4.6.5 Preuve de la Proposition 4.4

La quantité suivante est essentielle dans le calcul des deux paramètres de dépendance  $\chi$  et  $\bar{\chi}$  :

$$\begin{aligned} \mathbb{P}(X_{t-1} > x, X_t > x) &= \mathbb{P}(X_{t-1} > x, \alpha X_{t-1} + \alpha\sigma \log S_t + \mu(1-\alpha) > x) \\ &= \int_x^\infty \mathbb{P}\left(\log S_t > \frac{1}{\alpha\sigma} (x - \alpha y - \mu(1-\alpha)) \mid X_{t-1} = y\right) dH_{\mu,\sigma}(y) \\ &= \int_x^\infty \mathbb{P}\left(S_t > \exp\left(\frac{1}{\alpha\sigma} (x - \alpha y - \mu(1-\alpha))\right)\right) dH_{\mu,\sigma}(y) \end{aligned}$$



où  $S_t$  est une variable  $\alpha$ -stable positive et  $X_t$  une variable aléatoire de loi de Gumbel  $(\mu, \sigma)$ . Par conséquent

$$\begin{aligned} \mathbb{P}(X_{t-1} > x, X_t > x) &= \int_0^{\exp(-\frac{1}{\sigma}(x-\mu))} \mathbb{P}\left(S_t > z \exp\left(\frac{x-\mu}{\alpha\sigma}\right)\right) \exp(-z) dz \\ &= \exp\left(-\frac{x-\mu}{\alpha\sigma}\right) \int_0^{\exp(-\frac{x-\mu}{\alpha\sigma}(\alpha-1))} \exp\left(-u \exp\left(-\frac{x-\mu}{\alpha\sigma}\right)\right) \mathbb{P}(S_t > u) du \\ &= \exp\left(-\frac{x-\mu}{\alpha\sigma}\right) \int_0^\infty \exp\left(-u \exp\left(-\frac{x-\mu}{\alpha\sigma}\right)\right) \mathbb{1}_{0 \leq u \leq e^{\frac{x-\mu}{\alpha\sigma}(1-\alpha)}} \mathbb{P}(S_t > u) du. \end{aligned}$$

Notons maintenant que

$$\frac{1}{\mathbb{P}(X_{t-1} > x)} \underset{x \rightarrow \infty}{\sim} \exp\left(\frac{x-\mu}{\sigma}\right).$$

Dans un premier temps, nous nous intéressons au paramètre de dépendance  $\chi$  :

$$\begin{aligned} \chi &= \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X_{t-1} > x, X_t > x)}{\mathbb{P}(X_{t-1} > x)} \\ &= \lim_{x \rightarrow \infty} \exp\left(\frac{x-\mu}{\alpha\sigma}(\alpha-1)\right) \int_0^\infty \exp\left(-u \exp\left(-\frac{x-\mu}{\alpha\sigma}\right)\right) \mathbb{1}_{0 \leq u \leq e^{\frac{x-\mu}{\alpha\sigma}(1-\alpha)}} \mathbb{P}(S_t > u) du \\ &= \lim_{x \rightarrow \infty} \int_0^1 \exp\left(-\omega \exp\left(-\frac{x-\mu}{\sigma}\right)\right) \mathbb{P}\left(S_t > \omega e^{\frac{x-\mu}{\alpha\sigma}(1-\alpha)}\right) d\omega. \end{aligned}$$

Comme  $S_t$  est une variable  $\alpha$ -stable positive, nous avons  $\mathbb{P}(S_t > x) = x^{-\alpha} \ell(x)$  où  $\ell$  est une fonction à variation lente (cf. Bingham *et al.* [11], Section 8.3.5, page 349), ce qui implique que

$$\chi = \lim_{x \rightarrow \infty} \int_0^1 \exp\left(-\omega \exp\left(-\frac{x-\mu}{\sigma}\right)\right) \omega^{-\alpha} \ell\left(\omega e^{\frac{x-\mu}{\alpha\sigma}(1-\alpha)}\right) d\omega e^{-\frac{x-\mu}{\sigma}(1-\alpha)}.$$

En effectuant un développement de Taylor, nous obtenons que  $\exp(-\omega \exp(-(x-\mu)/\sigma)) = 1 - \omega e^{-\omega\kappa} e^{-\frac{x-\mu}{\sigma}}$  avec  $\kappa \in \left(0, e^{-\frac{x-\mu}{\sigma}}\right)$ , et par conséquent

$$\begin{aligned} \chi &= (1 + o(1)) \lim_{x \rightarrow \infty} \int_0^1 \omega^{-\alpha} \ell\left(\omega e^{\frac{x-\mu}{\alpha\sigma}(1-\alpha)}\right) d\omega e^{-\frac{x-\mu}{\sigma}(1-\alpha)} \\ &\sim \frac{1}{1-\alpha} \lim_{x \rightarrow \infty} \frac{\ell\left(e^{\frac{x-\mu}{\alpha\sigma}(1-\alpha)}\right)}{e^{\frac{1}{\sigma}(x-\mu)(1-\alpha)}} \\ &= 0, \end{aligned}$$

comme  $\alpha \in (0, 1)$ , l'approximation provenant du théorème de Karamata.

Nous nous intéressons maintenant au paramètre de dépendance  $\bar{\chi}$  :

$$\begin{aligned}
\bar{\chi} &= \lim_{x \rightarrow \infty} \frac{2 \log \mathbb{P}(X_{t-1} > x)}{\log \mathbb{P}(X_{t-1} > x, X_t > x)} - 1 \\
&= \lim_{x \rightarrow \infty} \frac{-2(x - \mu)/\sigma}{-(x - \mu)/(\alpha\sigma) + \log \left( \int_0^\infty \exp \left( -u \exp \left( -\frac{x - \mu}{\alpha\sigma} \right) \mathbb{1}_{0 \leq u \leq e^{\frac{x - \mu}{\alpha\sigma}(1 - \alpha)}} \mathbb{P}(S_t > u) du \right)} - 1 \\
&= \lim_{x \rightarrow \infty} \left[ \frac{1}{2\alpha} - \frac{\sigma}{2(x - \mu)} \log \left( \int_0^1 \exp \left( -\omega \exp \left( -\frac{x - \mu}{\sigma} \right) \right) \mathbb{P} \left( S_t > \omega e^{\frac{x - \mu}{\alpha\sigma}(1 - \alpha)} \right) d\omega \right. \right. \\
&\quad \left. \left. \times e^{\frac{x - \mu}{\alpha\sigma}(1 - \alpha)} \right) \right]^{-1} - 1 \\
&= \lim_{x \rightarrow \infty} \left[ \frac{1}{2} - \frac{\sigma}{2(x - \mu)} \log \left( \int_0^1 \exp \left( -\omega \exp \left( -\frac{x - \mu}{\sigma} \right) \right) \mathbb{P} \left( S_t > \omega e^{\frac{x - \mu}{\alpha\sigma}(1 - \alpha)} \right) d\omega \right) \right]^{-1} - 1.
\end{aligned}$$

Comme précédemment, nous avons  $\mathbb{P}(S_t > x) = x^{-\alpha} \ell(x)$  où  $\ell$  est une fonction à variation lente. Par conséquent, nous obtenons que

$$\begin{aligned}
\bar{\chi} &= \lim_{x \rightarrow \infty} \left[ \frac{1}{2} - \frac{\sigma}{2(x - \mu)} \log \left( \int_0^1 \exp \left( -\omega \exp \left( -\frac{x - \mu}{\sigma} \right) \right) \omega^{-\alpha} \ell \left( \omega e^{\frac{x - \mu}{\alpha\sigma}(1 - \alpha)} \right) d\omega \right. \right. \\
&\quad \left. \left. \times e^{-\frac{x - \mu}{\sigma}(1 - \alpha)} \right) \right]^{-1} - 1 \\
&= \lim_{x \rightarrow \infty} \left[ 1 - \frac{\alpha}{2} - \frac{\sigma}{2(x - \mu)} \log \left( \int_0^1 \exp \left( -\omega \exp \left( -\frac{x - \mu}{\sigma} \right) \right) \omega^{-\alpha} \ell \left( \omega e^{\frac{x - \mu}{\alpha\sigma}(1 - \alpha)} \right) d\omega \right) \right]^{-1} - 1 \\
&= \left[ 1 - \frac{\alpha}{2} \right]^{-1} - 1 \\
&= \frac{\alpha}{2 - \alpha}.
\end{aligned}$$

#### 4.6.6 Preuve de la Proposition 4.5

Nous pouvons réécrire (4.11) de la manière suivante

$$\log \left( X_t - \mu + \frac{\sigma}{\gamma} \right) = \alpha \log \left( X_{t-1} - \mu + \frac{\sigma}{\gamma} \right) + \alpha \gamma \log S_t + (1 - \alpha) \log \left( \frac{\sigma}{\gamma} \right)$$

avec  $X_t$  une variable aléatoire distribuée selon la distribution GEV de paramètres  $\mu$ ,  $\sigma$  et  $\gamma$  où  $(\mu, \sigma, \gamma) \in \mathbb{R} \times \mathbb{R}_*^+ \times \mathbb{R}_*$ .

En utilisant la Remarque 4.1 et en distinguant bien le cas où  $\gamma$  est strictement négatif du cas où  $\gamma$  est strictement positif, nous pouvons directement déduire le résultat de cette proposition de la même manière que celui de la Proposition 4.1.

Par exemple, si  $\gamma > 0$ , nous pouvons poursuivre la réécriture de (4.11) de la manière suivante

$$Z_t = \alpha Z_{t-1} + \alpha\gamma \log S_t + (1 - \alpha) \log \left( \frac{\sigma}{\gamma} \right)$$

avec  $Z_t$  une variable aléatoire distribuée selon la distribution de Gumbel  $(\log \frac{\sigma}{\gamma}, \gamma)$ . Nous reconnaissons ici l'équation (4.3) de la Proposition 4.1.

# Conclusion et Perspectives

Dans cette thèse nous avons abordé différents aspects de la théorie des valeurs extrêmes univariées allant du cas indépendant au cas dépendant en passant par la prise en compte des données censurées.

Après avoir dans un premier chapitre rappelé certaines des notions clés en théorie des valeurs extrêmes, nous nous sommes intéressés dans le deuxième chapitre à l'élaboration d'un test d'adéquation à une loi de probabilité souvent rencontrée en théorie des valeurs extrêmes, la loi de Pareto. L'élaboration de ce test a été motivée par la théorie de LeCam bien connue en statistique mathématique. Nous avons alors établi le comportement asymptotique de la statistique de test proposée sous l'hypothèse nulle que l'échantillon provient d'une distribution de Pareto. Dans un second temps, nous avons établi son comportement asymptotique sous des alternatives locales, nous plaçant ainsi dans le cadre LAN. Des simulations ont également été présentées afin d'apprécier le comportement de la statistique de test à distance finie.

Dans le troisième chapitre de notre thèse, nous nous sommes intéressés à un problème récent en théorie des valeurs extrêmes qui est celui de la présence de données censurées. Ce problème est très fréquent dans plusieurs domaines d'applications, comme en médecine ou en actuariat, où souvent les données dont nous disposons sont censurées aléatoirement à droite. Plusieurs estimateurs ont déjà été étudiés dans ce contexte, mais celui résultant

d'une généralisation de la vraisemblance au cadre censuré n'avait pas été introduit malgré les bonnes propriétés des estimateurs du maximum de vraisemblance. L'une des raisons pour cela est sans doute que l'un des principaux défauts de ce type d'estimateurs est le fait qu'on ne dispose pas de leurs expressions explicites. De ce fait, leurs propriétés asymptotiques sont très difficiles à établir. Pour palier ce problème, nous avons proposé dans ce chapitre, toujours dans le cadre censuré, des nouveaux estimateurs des paramètres de la distribution de Pareto généralisée basés sur une première étape de l'algorithme de Newton-Raphson. Leurs comportements asymptotiques ont été ensuite établis. Enfin, par des simulations, nous avons illustré le fait que les estimateurs proposés avaient un comportement, à distance finie, très similaire à ceux du maximum de vraisemblance.

Dans les chapitres précédents, nous avons toujours supposé que les observations dont nous disposions étaient indépendantes. Or, bien sûr, dans beaucoup d'applications pratiques, cette hypothèse est mise à défaut. Nous nous sommes donc placés dans le quatrième chapitre de notre thèse dans le cadre de données dépendantes et avons cherché à prendre en compte la dépendance dans les séries temporelles. Pour cela, nous avons proposé un modèle linéaire autorégressif adapté aux maxima et plus particulièrement à la distribution de Gumbel. Nous avons établi quelques propriétés théoriques de notre modèle et une étude par simulations nous a permis d'illustrer la qualité du modèle proposé. Enfin, comme celui-ci est issu d'une réelle problématique en sciences du climat, nous l'avons utilisé pour modéliser des maxima de dioxyde de carbone et de méthane.

Si parmi ces différents travaux de recherche, certains sont motivés par une problématique appliquée et d'autres sont à caractère plus mathématique, tous ont en commun de soulever des perspectives que nous envisageons d'ores et déjà d'explorer. En voici une liste non exhaustive :

- L'un des problèmes clés avec tous les estimateurs d'indice est le problème du biais. Dans le cadre sans censure, pour certains estimateurs (Hill,  $UH$ ), des techniques de réduction de biais ont été proposées. Elles sont basées sur des modèles de régression exponentielle. Dans le cadre de l'approche POT, à notre connaissance, le biais des estimateurs n'a jamais été étudié. Le Chapitre 3 donne une perspective pour atteindre cet objectif, du moins dans le cas sans censure. En effet, il est possible dans ce cas là de remplacer les fonctions de répartition par leurs versions empiriques, l'estimateur de  $\gamma_1$  par un estimateur débiaisé. Reste alors le problème de  $\sigma_{1,t}$  pour lequel nous ne disposons pas d'estimateur débiaisé. En revanche, comme nous l'avons vu, nous avons un lien entre  $\sigma_{1,t}$  d'une part, et  $\gamma_1$  et le seuil  $t$  d'autre part. Ce lien peut donc être utilisé dans un premier temps et conduire ainsi à une estimation de  $B_{1,t}(\gamma_1, \sigma_{1,t})$  et  $B_{2,t}(\gamma_1, \sigma_{1,t})$ . En retranchant alors ces biais estimés de nos estimateurs, nous devrions obtenir des estimations à biais réduits de  $(\gamma_1, \sigma_{1,t})$ .
- Une fois les termes principaux de biais estimés, nous pouvons également nous intéresser au problème de la sélection du paramètre  $k$  représentant le nombre de statistiques d'ordre à utiliser dans la construction de nos estimateurs. L'idée naturelle est alors de reprendre l'approche standard consistant à choisir  $k$  comme la valeur qui minimise l'AMSE (erreur en moyenne quadratique asymptotique), étant donné que nous aurons estimé au préalable le biais.
- D'un point de vue pratique, il serait également bon de pouvoir inclure dans notre modèle des covariables. En effet, il est légitime de penser que, par exemple dans les applications médicales, il y a beaucoup de covariables qui interviennent et qu'il serait utile d'en tenir compte. Nous pouvons par exemple raisonnablement penser qu'un bébé réagira de façon différente à une même opération qu'un adulte. Par conséquent, l'âge devra être pris en compte dans une telle étude statistique.

- Le modèle que nous avons développé dans le Chapitre 4 mériterait d'être étendu au cadre multivarié et spatial. En effet, en sciences du climat par exemple, nous disposons souvent de mesures localisées en plusieurs endroits et il serait donc fort intéressant de pouvoir prendre en compte l'aspect spatial dans l'analyse de nos données.

- Classiquement en géophysique, les modèles à espace d'états dont nous disposons sont de la forme :

$$\begin{cases} Y_t = F_t X_t + \varepsilon_t \\ X_t = G_t X_{t-1} + \eta_t \end{cases}$$

où  $\eta_t$  et  $\varepsilon_t$  sont des bruits gaussiens. Si  $Y_t$  est un maximum, alors il devrait être de loi GEV. Ceci est impossible avec un modèle linéaire à espace d'états dont le bruit est gaussien. Nous devons donc définir de nouveaux modèles à espace d'états admettant cette contrainte de loi et les étudier. Une piste pour aborder ce problème est d'utiliser les résultats du Chapitre 4, le modèle linéaire adapté aux maxima proposé pouvant être considéré comme l'équation dynamique d'un modèle à espace d'états. Ce travail est directement lié au problème de l'assimilation de données en théorie des valeurs extrêmes, un champ de recherche actuellement en plein essor.

# Bibliographie

- [1] Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N., 1993. *Statistical models based on counting processes*, Springer, New York.
- [2] Arnold, B.C., 1983. *Pareto distributions*, International Co-operative Publishing House, Burtonsville, MD.
- [3] Arnold, B.C., Press, S.J., 1989. Bayesian estimation and prediction for Pareto data. *J. Amer. Statist. Assoc.*, **84**, 1079-1084.
- [4] Balkema, A., de Haan, L., 1974. Residual life at great age. *Ann. Probab.*, **2**, 792-804.
- [5] Beirlant, J., Dierckx, G., Guillou, A., 2005. Estimation of the extreme-value index and generalized quantile plots. *Bernoulli*, **11**, 6, 949-970.
- [6] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of extremes, theory and applications*, Wiley, New York.
- [7] Beirlant, J., Guillou, A., Dierckx, G., Fils-Villetard, A., 2007. Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes*, **10**, 151-174.
- [8] Beirlant, J., Teugels, J.L., 1987. Asymptotic normality of Hill's estimator. *Extreme value theory*, 51, 148-155, Springer, New York.
- [9] Beirlant, J., Vynckier, P., Teugels, J.L., 1996. Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.*, **91**, 1659-1667.



- 
- [10] Benktander, G., 1970. Schaderteilung nenvach Grösse in der Nicht-Leben-Versicherung. *Mitt. Verein. Schweiz. Versich.-Mathr.*, **70**, 268-284.
- [11] Bingham, N.H., Goldie, C.M., Teugels, J.L., 1987. *Regular Variation*, Cambridge University Press, Cambridge.
- [12] Bosq, D., Lecoutre, J.P., 1992. *Analyse et prévision des séries chronologiques, méthodes paramétriques et non paramétriques*, Masson, Paris.
- [13] Brockwell, P.J., Davis, R.A., 1987. *Time series : theory and methods*, Springer Series in Statistics, Springer-Verlag, New York.
- [14] Castillo, E., Hadi, A.S., 1997. Fitting the generalized Pareto distribution to data. *J. Amer. Statist. Assoc.*, **92**, 1604-1620.
- [15] Choulakian, V., Stephens, M.A., 2001. Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, **43**, 478-484.
- [16] Coles, S., 2001. *An introduction to statistical modeling of extreme values*, Springer Series in Statistics, Springer-Verlag, London.
- [17] Coles, S., Heffernan, J., Tawn, J., 1999. Dependence measures for extreme value analyses. *Extremes*, **2**, 339-65.
- [18] Csörgö, S., Deheuvels, P., Mason, D.M., 1985. Kernel estimates of the tail index of a distribution. *Ann. Statist.*, **13**, 1050-1077.
- [19] Csörgö, S., Viharos, L., 1998. Estimating the tail index. In *Asymptotic methods in probability and statistics*, North-Holland, Amsterdam, 833-881.
- [20] D'Agostino, R.B., Stephens, M.A., 1986. *Goodness-of-fit techniques*, Marcel Dekker, New York.
- [21] Danielsson, J., de Vries, C.G., 1997. Tail index and quantile estimation with very high frequency data. *J. Empir. Finance*, **4**, 241-257.
- [22] Davis, R.A., Resnick, S.I., 1989. Basic properties and prediction of max-ARMA processes. *Adv. in Appl. Probab.*, **21**, 781-803.

- 
- [23] Davison, A.C., 1984. Modelling excesses over high thresholds, with an application. In J. Tiago de Oliveira (Ed.), *Statistical Extremes and Applications*, Reidel, Dordrecht, 461-482.
- [24] Davison, A.C., Smith, R.L., 1990. Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B*, **52**, 393-442.
- [25] Deheuvels, P., Haeusler, E., Mason, D.M., 1988. Almost sure convergence of the Hill estimator. *Math. Proc. Cambridge Philos. Soc.*, **104**, 2, 371-381.
- [26] Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L., 1989. A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, **17**, 1833-1855.
- [27] Diebolt, J., Guillou, A., Rached, I., 2004. A new look at probability-weighted moments estimators. *C. R. Acad. Sci. de Paris*, **338**, Série I, 8, 629-634.
- [28] Diebolt, J., Guillou, A., Rached, I., 2007. Approximation of the distribution of excesses through a generalized probability-weighted moments method. *J. Statist. Plann. Inference*, **137**, 3, 841-857.
- [29] Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Statist.*, **25**, 187-210.
- [30] Drees, H., Ferreira, A., de Haan, L., 2004. On maximum likelihood estimation of the extreme value index. *Ann. Applied Probab.*, **14**, 1179-1201.
- [31] Einmahl, J.H.J., Fils-Villetard, A., Guillou, A., 2008. Statistics of extremes under random censoring. *Bernoulli*, **14**, 1, 207-227.
- [32] Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. *Modelling extremal events*, Springer-Verlag, Berlin.
- [33] Fisk, P.R., 1961. The Graduation of Income Distributions. *Econometrica*, **29**, 171-185.

- 
- [34] Fougères, A-L., 2004. Multivariate extremes. In *Extreme Values in Finance, Telecommunications, and the Environment*, Finkenstädt B and Rootzén H (eds), Boca Raton, CRC, Boca Raton, Florida, 373-388.
- [35] Fougères, A-L., Nolan, J.P., Rootzén, H, 2005. Mixture models for extremes. *Soumis*.
- [36] Gustafson, G., Fransson, A., 2005. The use of the Pareto distribution for fracture transmissivity assessment. *Hydrogeology Journal*, **14**, 15-20.
- [37] de Haan, L., Ferreira, A., 2006. *Extreme value theory : An introduction*, Springer, New York.
- [38] de Haan, L., Rootzén, H., 1993. On the estimation of high quantiles. *J. Statist. Plann. Inference*, **35**, 1-13.
- [39] de Haan, L. and Stadtmüller, U., 1996. Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A*, **61**, 381-395.
- [40] Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163-1174.
- [41] Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**, 339-349.
- [42] Inglot, T., Kallenberg, W.C.M., Ledwina, T., 1997. Data driven smooth tests for composite hypotheses. *Ann. Statist.*, **25**, 1222-1250.
- [43] Joe, H., 1993. Parametric families of multivariate distributions with given margins. *J. Multivariate Anal.*, **46**, 262-282.
- [44] Joe, H., 1997. *Multivariate models and dependence concepts*, Monographs on Statistics and Applied Probability, 73, Chapman & Hall, London.
- [45] Katz, R.W., Parlange, M.B., Naveau, P., 2002. Statistics of extremes in hydrology. *Advances in Water Resources*, **25**, 1287-1304.
- [46] Kopecky, K.J., Pierce, D.A., 1979. Efficiency of smooth goodness-of-fit tests. *J. Amer. Statist. Assoc.*, **74**, 393-397.

- 
- [47] Kysely, J., 2002. Probability estimates of extreme temperature events : stochastic modelling approach vs. extreme value distributions. *Studia Geophysica et Geodaetica*, **46**, 93-112.
- [48] Leadbetter, M. R., Lindgren, G., Rootzén, H., 1983. *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York.
- [49] LeCam, L., 1960. Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Statist.*, **3**, 37-98.
- [50] Lehmann, E.L., 1991. *Theory of point estimation*, Brooks-Cole publishing company.
- [51] Malik, H.J., 1966. Estimation of the parameters of the Pareto distribution. *Metrika*, **15**, 126-132.
- [52] Mason, D.M., 1982. Laws of large numbers for sums of extreme values. *Ann. Probab.*, **10**, 3, 754-764.
- [53] Mikosch, T., Gadrich, T., Klüppelberg, C., Adler, R.J., 1995. Parameter estimation for ARMA models with infinite variance innovations. *Ann. Statist.*, **23**, 1, 305-326.
- [54] Miller, F.L., Quesenberry, C.P., 1979. Power studies of tests for uniformity, II. *Comm. Statist. B-Simulation Comput.*, **8**, 271-290.
- [55] Neyman, J., 1937. Smooth tests for goodness of fit. *Skand. Aktuarietidskr.*, **20**, 149-199.
- [56] Pfanzagl, J., 1994. *Parametric Statistical Theory*, Walter de Gruyter & Co., Berlin.
- [57] Pickands, J., 1975. Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119-131.
- [58] Porter, J.E., Coleman, J.W., Moore, A.H., 1992. Modified KS, AD, and C-VM Tests for the Pareto Distribution with Unknown Location and Scale Parameters. *IEEE transactions on reliability*, **41**, 112-117.

- 
- [59] Rached, I., 2004. *Méthode des moments généralisés pondérés*. Thèse de l'Université de Marne-La-Vallée.
- [60] Rayner, J.C.W., Best, D.J., 1986. Neyman-type smooth tests for location-scale families. *Biometrika*, **73**, 437-446.
- [61] Rayner, J.C.W., Best, D. J., 1989. *Smooth tests of goodness of fit*, Oxford University Press, New York.
- [62] Rayner, J.C.W., Best, D. J., 1990. Smooth tests of goodness of fit, an overview. *Internat. Statist. Rev.*, **58**, 9-17.
- [63] Reiss, R.-D., Thomas, M., 1997. *Statistical analysis of extreme values. From insurance, finance, hydrology and other fields*, Birkhäuser Verlag, Basel.
- [64] Samorodnitsky, G., Taqqu, M.S., 1994. *Stable non-Gaussian random processes*, Chapman & Hall, New York.
- [65] Smith, R.L., 1984. Threshold methods for sample extremes, with an application. In J. Tiago de Oliveira (Ed.), *Statistical Extremes and Applications*, Reidel, Dordrecht, 621-638.
- [66] Smith, R.L., 1987. Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174-1207.
- [67] Strasser, H., 1985. *Mathematical Theory of Statistics*, De Gruyter, Berlin.
- [68] Thomas, D.R., Pierce, D.A., 1979. Neyman's smooth goodness-of-fit test when the hypothesis is composite. *J. Amer. Statist. Assoc.*, **74**, 441-445.
- [69] van der Vaart, A.W., 1998. *Asymptotic Statistics*, Cambridge University Press, Cambridge.
- [70] von Neumann, J., 1951. Various techniques used in connection with random digits. Monte Carlo methods. *National Bureau of Standards AMS*, **12**, 36-38

- 
- [71] Zolotarev, V.M., 1986. *One-Dimensional Stable Distributions*. American Mathematical Society Translations of Mathematical Monographs, Vol. 65. American Mathematical Society, Providence, RI. Translation of the original 1983 (in Russian).