

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE
SCIENTIFIQUE



UNIVERSITE DES FRERES MENTOURI CONSTANTINE 1
FACULTE DES SCIENCES EXACTES
DEPARTEMENT DE MATHEMATIQUE

N° d'ordre : 81/D3C/2019.

N° de série : 02/Math/2019.

THÈSE

présentée pour l'obtention du diplôme de

Doctorat 3^{ème} Cycle en Mathématiques

Option

Probabilité appliquée

intitulée

**Théorie des Valeurs Extrêmes :
Application au Calcul de Risques**

Présenté par :

Bahi Oussama

Dirigé par :

Dr. Dakhmouche Meghlaoui

Devant le jury :

Président :	Denche Mohamed	Prof.	UFM Constantine 1
Encadreur :	Dakhmouche Meghlaoui	M.C(A).	UFM Constantine 1
Examineur :	Benchettah Azzédine	Prof.	UBM Annaba
Examineur :	Ellagoune Fateh	Prof.	Université Guelma
Examineur :	Raynaud de Fitte Paul	Prof.	Université Rouen-France

Dédicaces

Celui qui veut voir l'arc-en-ciel doit apprendre à aimer la pluie
Paulo Coelho.

A tous croyants en savoir à travers toutes époques.

A tous émetteurs d'éclat de lumière dans les sentiers obscurs de la vérité.

Aux plus grands majestueux guerroyés par l'ignorance!

Aujourd'hui vous êtes l'objet de ma commémoration et glorification.

A la plus précieuse chose dans l'existence : mes parents, Maman ma première enseignante,
Papa le fidèle compagnon de mon esprit à l'éternité.

À mes frangins Asma et Ahmed, les petits bouts qui comblent mon âme. Merci pour votre soutien.

À celui qui était un frère avant d'être un ami, le long de mon cursus d'étude, Ridha! A toi mon immense amour, affection et gratitude.

A tous les membres de la grande famille, à tous les amis, certes je ne pourrais pas tous vous citer, mais ça ne m'empêche nulle part d'avouer l'ampleur d'influence de vos empreintes dans ma vie merci profondément.

À tous ceux qui me disaient un jour : quand est ce qu'on fêtera ta graduation, eh bien, réjouissez-vous maintenant! Je soutiens.

Remerciements

Ce travail de recherche a été accompli au laboratoire Modélisation Mathématique et Simulation (MMS) de l'Université Frère Mentouri Constantine 1, dirigé par Dr. DAKHMOUCHE Meghlaoui. Je le remercie vivement pour m'avoir accueillie comme doctorant dans son laboratoire.

Je tiens à remercier le Dr. Dakhmouche Meghlaoui de l'Université Frères Mentouri Constantine 1, de m'avoir permis de mener à terme ma thèse. Je le remercie aussi de m'avoir fait partager ses brillantes intuitions, ainsi que pour sa gentillesse, sa disponibilité à tout instant et pour les nombreux encouragements qu'il m'a prodigués.

J'exprime ma profonde gratitude au Pr. Denche Mohamed de l'Université Frères Mentouri Constantine 1, qui me fait l'honneur de présider mon jury de soutenance.

J'adresse tous mes remerciements au Pr. Benchettah Azzédine de l'Université Badji Mokhtar Annaba, et au Pr. Ellagoune Fateh, de l'Université de Guelma, de m'avoir fait l'honneur d'accepter d'expertiser ma thèse de Doctorat et d'être membre de mon jury de soutenance.

Je tiens plus spécialement à remercier le Pr. Paul Raynaud de Fitte, de l'Université de Rouen, de m'avoir accueilli au sein de son laboratoire et de m'avoir prodiguer des conseils très utiles. C'est grâce à lui que j'ai pu concilier avec bonheur entre recherche théorique et recherche appliquée. De plus, il me fait l'immense honneur d'expertiser mon travail et de participer à mon jury de soutenance.

Merci aussi à Madame Dakhmouche Rosa, Docteur à l'Université Frères Mentouri Constantine 1, qui m'a fait profiter de ces compétences en Anglais pour la rédaction de mes articles, en plus de ces constants encouragements durant la conception de ma thèse de Doctorat.

Je voudrais remercier tout particulièrement le Dr. Rezgui Imane membre de laboratoire MMS, pour sa relecture enrichissante de mes articles et ma thèse, en plus elle a toujours été disponible, à l'écoute de mes nombreuses questions, et s'est toujours intéressé à l'avancée de mes travaux.

Résumé

Ce travail se veut une contribution à la statistique univariée des valeurs extrêmes et ses applications au calcul du risque. Le but de ce travail est d'aborder en profondeur les techniques statistiques appliquées à la mesure et à la gestion des risques extrêmes. En effet, le résultat essentiel obtenu par Fisher-Tippett sur les lois limites possibles du maximum d'un échantillon, a apparemment provoqué l'idée que la théorie des valeurs extrêmes était quelque chose d'assez remarquable et très différent de la théorie classique de la limite centrale. Notre contribution essentielle consiste en l'étude des modèles *GPD* en présence de censure par intervalle. En effet, nous avons établi que les estimateurs du pseudo-maximum de vraisemblance conserve les propriétés de convergence et de normalité asymptotique. Nous avons appliqué les modèles *GPD* avec censure par intervalle pour comparer deux thérapies différentes utilisées dans le traitement du cancer du sein chez la femme. Ainsi, nous avons établi un test de type Wald qui nous a permis de comparer les deux traitements.

Mots-clés

Théorie des valeurs extrêmes ; Modèles GPD ; Estimation du maximum de vraisemblance ; Méthode des moments pondérés ; Gestion des risques ; Value-at-Risk ; Censure par intervalles ; Propriétés asymptotiques des pseudo-estimateurs ; Test de type Wald.

Abstract

This work is intended as a contribution to the univariate statistics of extreme values and its applications to the calculation of risk. The purpose of this dissertation is to take a deep look at the statistical techniques applied to the measurement and management of extreme risks. The essential result obtained by Fisher-Tippett on the possible existence of limit distributions of the maximum of a sample, apparently gave rise to the idea that the theory of extreme values was something quite remarkable and very different from the classical theory of the central limit. Our aim contribution is the study of *GPD* models with interval censoring. Indeed, we have established that the pseudo-maximum likelihood estimates retain the properties of convergence and asymptotic normality. We applied the *GPD* models with interval censoring to compare two different therapies used in the treatment of woman breast cancer. Finally, we have built a Wald type hypothesis testing that allows us to compare the two treatments.

Keywords and Phrases

Extreme values theory ; GPD models ; Maximum likelihood estimation method ; Method of probability weighted moments estimation ; Risk management ; Value-at-Risk ; Interval censoring ; Asymptotic properties of estimates ; Wald-type test.

ملخص:

يمثل هذا العمل مساهمة في النماذج الإحصائية للقيم المتطرفة وتطبيقاتها في حساب المخاطر. الغرض منه هو إلقاء نظرة أعمق على التقنيات الإحصائية المطبقة على قياس وإدارة المخاطر الشديدة. في الواقع، إن النتيجة الأساسية التي حصل عليها Fisher-Tippett بشأن نهاية التوزيعات الممكنة للقيم الحدية لعينة ما أثارت على ما يبدو فكرة أن نظرية القيم الحدية كانت شيئاً مختلف تماماً ومثيراً للاهتمام عن النظرية الكلاسيكية للنهاية المركزية. مساهمتنا الرئيسية هي دراسة نماذج GPD في وجود رقابة ضمن مجال. في الواقع، لقد أثبتنا أن مقدرات الاحتمالات الزائفة يحتفظون بخصائص التقارب والحالة الطبيعية المقاربة. استخدمنا نماذج GPD خاضعة للرقابة لمقارنة فعالية علاجين مختلفين لسرطان الثدي لدى النساء. وهكذا، أنشأنا اختبار من نوع Wald الذي سمح لنا بمقارنة العلاجين.

الكلمات المفتاحية:

نظرية القيم الحدية؛ نماذج GPD ؛ أقصى تقدير احتمال؛ طريقة تقدير احتمالية اللحظات الموزونة؛ إدارة المخاطر؛ القيمة المعرضة للخطر؛ الرقابة بفاصل زمني؛ خواص مقاربة للمقدرات؛ اختبار فرضية نوع Wald.

Table des matières

Introduction	13
1 Pourquoi la théorie des valeurs extrêmes ?	15
1.1 Problème simple de valeurs extrêmes	15
1.2 Outils graphiques pour l'identification de distributions	18
1.2.1 Quantile-Quantile plot	18
1.2.2 Log-Log plot	22
2 Approche probabiliste des valeurs extrêmes	24
2.1 Etude de quelques distributions de probabilités	24
2.1.1 Distribution de probabilité de Fréchet	25
2.1.2 Distribution de probabilité de Weibull	28
2.1.3 Distribution de probabilité de Gumbel	31
2.2 Classes de distributions à queue lourde	33
2.2.1 Distributions n'admettant aucun moments	34
2.2.2 Distributions sub-exponentielles	34
2.2.3 Fonctions α -stables	36
2.2.4 Fonctions à variation régulière	36
2.3 Distribution asymptotique d'un extrême	39
2.3.1 Distributions généralisées des valeurs extrêmes	39
2.3.2 Distribution du maximum d'un échantillon	41
2.4 Caractérisation des domaines d'attraction	44
2.4.1 Domaine de Fréchet	44

2.4.2	Domaine de Weibull	45
2.4.3	Domaine de Gumbel	46
2.5	Distribution conditionnelle des excès	47
2.5.1	Distribution de Paréto généralisée	47
3	Estimation dans les modèles <i>GPD</i>	52
3.1	Méthode des excès au dessus d'un seuil	52
3.1.1	Sélection du seuil	53
3.2	Méthode du maximum de vraisemblance	55
3.3	Méthode des moments ordinaires	56
3.4	Méthode des moments pondérés	57
3.5	Estimateur de Hill	58
3.5.1	Propriétés asymptotiques de l'estimateur de Hill	60
3.6	Estimateur de Pickands	62
3.6.1	Choix du nombre optimal de statistique d'ordre extrêmes	63
4	Calcul de risque	66
4.1	Evaluation du risque	66
4.2	Mesure synthétique du risque extrême	67
4.2.1	Définition mathématique de la <i>VaR</i>	68
4.2.2	<i>VaR</i> historique et <i>VaR</i> bootstrap	69
4.2.3	<i>VaR</i> Gaussienne	69
4.3	Estimation de quantiles extrêmes	70
4.3.1	Approche basée sur l'estimateur du maximum de vraisemblance	71
4.3.2	Approche basée sur l'estimateur de Hill	72
4.3.3	Approche basée sur l'estimateur des moments	73
5	Modèle <i>GPD</i> en présence de censure	74
5.1	Différents types de censures	74
5.2	Distribution d'une variable censurée	75
5.2.1	Variable censurée à droite	75
5.2.2	Variable censurée par intervalle	76
5.3	Asymptotic properties of maximum pseudo-likelihood of a <i>GPD</i> Model parameters estimators with interval censoring	77
5.3.1	Introduction	78
5.3.2	Parametric model with interval censoring	79

5.3.3	Efficiency of the maximum pseudo-likelihood estimator	80
5.3.4	Asymptotic normality of the maximum pseudo-likelihood estimator	82
5.3.5	GPD model with interval censoring	83
5.4	Analysis of breast cancer data in framework of a GPD model with interval censoring	86
5.4.1	Introduction	86
5.4.2	Parametric model with interval censoring	87
5.4.3	Asymptotic normality of the estimator	89
5.4.4	Analysis of breast cancer data	91

Bibliography		96
---------------------	--	-----------

Table des figures

1.1	\hat{F}_n et \hat{Q}_n des données sur la vitesse journalière maximale du vent dans la ville d'Albuquerque	16
1.2	Boxplot	17
1.3	QQ-plot exponentiel	19
1.4	Mesure quotidienne de la vitesse maximale du vent à Zaventem (Belgique) De 1985 à 1992.	20
1.5	Histogramme conditionnel des vitesses de vent supérieures à 82km/h.	21
1.6	QQ-plot exponentiel pour des vitesses de vent supérieures à 82km/h.	21
1.7	Illustration du graphique Log-Log	23
2.1	Fonction de densité de Fréchet : (a) pour différentes valeurs de μ , (b) pour différentes valeurs de σ	25
2.2	Fonction de survie de Fréchet pour différents valeurs de ξ	26
2.3	Fonction de densité de Weibull : (a) pour différentes valeurs de μ , (b) pour différentes valeurs de σ	29
2.4	Fonction de densité de Weibull pour différents valeurs de ξ	30
2.5	Fonction de densité de Gumbel : (a) pour différentes valeurs de μ , (b) pour différentes valeurs de σ	32
2.6	Comparaison d'une distribution normale à une distribution à queue lourde	34
2.7	Illustration de la différence entre la distribution exponentielle et une distribution à queue lourde (Fréchet)	35
2.8	Densité de probabilité des Valeurs Extrêmes avec différentes valeurs de ξ	41
2.9	Fonctions de survie des distributions <i>GEV</i> pour différentes valeurs de ξ	44
2.10	Densités de probabilité de <i>GPD</i> avec différentes valeurs de ξ	50
2.11	Fonction de survie de <i>GPD</i> avec différentes valeurs de ξ	51

3.1	Méthode des excès au dessus d'un seuil	53
5.1	(a) the QQ-plot of $\hat{\gamma}$ and (b) the QQ-plot of $\hat{\sigma}$	85
5.2	Frequency curves of the two empirical distributions of comfort time of the patients treated with chemotherapy and radiotherapy and the estimated comfort time of the patients treated with radiotherapy alone	93
5.3	Two survival functions of the estimated comfort time of the patients treated with chemotherapy and radiotherapy and the estimated comfort time of the patients treated with radiotherapy alone	95

Liste des tableaux

1.1	Résumé statistique	16
1.2	QQ-plots pour quelques distributions usuelles	20
2.1	Variation de S et K en fonction de ξ	28
2.2	Variation de S et K en fonction de ξ	31
5.1	Efficiency of estimators $\hat{\gamma}$ and $\hat{\sigma}$	85
5.2	Breast cancer data with two different treatments	92
5.3	GPD parameters estimation	94

Introduction

La gestion des risques est une procédure assez courante dans différents domaines, par exemple, l'hydrologie, la géologie, les assurances et la finance, et l'épidémiologie, etc. Pour gérer ce genre de risques, il nous faut disposer de méthodes assez robustes pour mesurer et contrôler le risque. Le résultat essentiel obtenu par Fisher-Tippett ([Fisher and Tippett \(1928\)](#)) sur les lois limites possibles du maximum d'un échantillon, a apparemment provoqué l'idée que la théorie des valeurs extrêmes était quelque chose d'assez remarquable, très différent de la théorie classique de la limite centrale. L'article de Gumbel ([Gumbel \(1958\)](#)) a longtemps été considéré comme l'œuvre principale de référence pour les applications de la théorie des valeurs extrêmes. Le manque d'intérêt aux valeurs extrêmes des probabilistes et des statisticiens se traduit par un nombre très limité de documents pertinents. La thèse de doctorat de Haan ([deHaan \(1970\)](#)) semble être le point de départ des développements théoriques de la théorie des valeurs extrêmes. Pour la première fois, les propriétés probabilistes et stochastiques des échantillons d'extrêmes ont été développées en une théorie cohérente et attrayante, comparable à la théorie des sommes de variables aléatoires. Les aspects statistiques ont dû attendre encore plus longtemps avant de capter l'attention nécessaire.

Notre thèse est structurée cinq chapitres. Dans le premier chapitre, on illustre à l'aide d'exemples concrets pourquoi et comment on peut étudier les valeurs extrêmes dans un ensemble de données. Par suite, on expose quelques techniques statistiques de type graphique qui fournissent assez d'informations concernant l'identification des distributions. Ces outils graphiques nous aideront à élaborer un modèle raisonnable pour décrire la population statistique sous étude. Dans le deuxième chapitre, on présente d'abord les

principales caractéristiques de quelques distributions d'extrêmes. On introduit ensuite la notion de distribution à queue lourde. On résume les résultats des travaux de Jenkinson sur les distributions généralisées des valeurs extrêmes. Puis on présente les différentes étapes de la démonstration du théorème de Fisher-Tippet qui établit l'existence de distributions limites pour le maximum d'un échantillon. La notion d'extrême peut être exprimée sous forme d'un excès au dessus d'un seuil élevé. La distribution limite d'un excès est donnée par le théorème de Balkema-de Hann-Pickands ont nous proposons une démonstration. Au troisième chapitre nous nous intéressons aux différentes méthodes d'estimation des paramètres (σ, ξ) d'un modèle *GPD*. Dans la littérature il existe beaucoup de méthodes d'estimation. Parmi ces méthodes, on s'est intéressé à celle du maximum de vraisemblance, celle des moments ordinaires et pondérés, à celle de Pickands et à l'estimateur de Hill essentiellement. Toutes ces méthodes partagent les mêmes propriétés de consistance et de normalité asymptotique. En général, il n'y a pas une meilleure méthode dans toutes les situations. Le chapitre quatre a été consacré à l'étude de la notion de risque et de son évaluation. L'outil calcul du risque le plus approprié est la Value-at-Risque. Le calcul de cette dernière est obtenu à travers l'estimation des quantiles extrêmes. Dans le dernier chapitre, on a introduit la notion de censure dans les modèles *GPD* et on s'est intéressé à l'estimation des (σ, ξ) . On a montré que les pseudo-estimateurs obtenus conservent les propriétés d'efficacité et de normalité asymptotique. Ce travail fait l'objet d'une publication internationale. Dans la dernière section du chapitre, on traite un corpus de données relatif au traitement du cancer du sein chez la femme. L'étude consiste en la comparaison entre deux types de traitements du cancer du sein. Et on voudrait comparer la durée du confort esthétique chez les patientes. Pour cela, on a construit un test d'hypothèse de type Wald et on a conclut que le traitement en radiothérapie seule est meilleur que le traitement combiné entre radiothérapie et chimiothérapie dans le sens où la durée du confort esthétique chez les patientes est plus longue. Cette dernière section à fait l'objet d'une publication internationale.

Chapitre 1

Pourquoi la théorie des valeurs extrêmes ?

Dans ce chapitre, nous nous intéressons au comportement des valeurs extrêmes d'un ensemble de données. Supposons que ces données soient une réalisation d'un n -échantillon X_1, X_2, \dots, X_n d'une variable aléatoire X de distribution F . Les données ordonnées seront alors notées par $X_{1,n}, X_{2,n}, \dots, X_{n,n}$, où $X_{1,n}$ et $X_{n,n}$ sont respectivement le minimum et le maximum. L'échantillon de données est généralement utilisé pour étudier la fonction de répartition inconnue F à travers la fonction de répartition empirique définie telle que

$$\widehat{F}_n(x) = \begin{cases} 0 & \text{si } x < x_{1,n} \\ \frac{i}{n} & \text{si } x \in [x_{i,n}, x_{i+1,n}[\\ 1 & \text{si } x \geq x_{n,n} \end{cases}$$

Cependant, dans certaines situations on étudie la fonction inverse de $Q = F^{-1}$, à travers la fonction quantile empirique définie telle que

$$\widehat{Q}(p) := \inf \{x : F(x) \geq p\}.$$

1.1 Problème simple de valeurs extrêmes

Afin d'illustrer nos propos, nous allons reprendre ici un exemple pratique traité dans l'ouvrage de Beirlant ([Beirlant et al. \(1996a\)](#)). On souhaite étudier les données quotidiennes sur la vitesse maximale du vent dans la ville d'Albuquerque. La fonction de distribution empirique des vitesses maximales du vent est représentée à la figure [1.1](#).

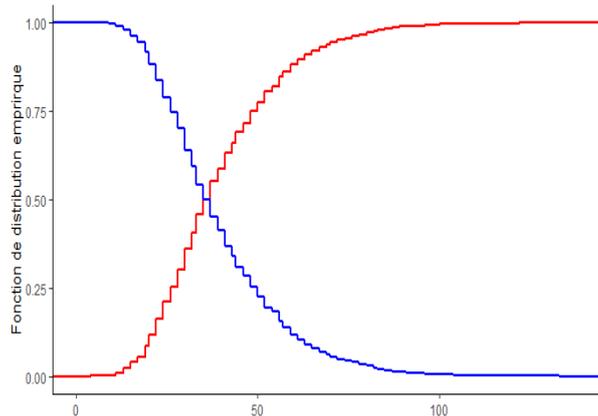


FIGURE 1.1 – \hat{F}_n et \hat{Q}_n des données sur la vitesse journalière maximale du vent dans la ville d’Albuquerque

Le résumé statistique de ces données est présenté dans le tableau suivant :

<i>Min</i>	<i>Q1</i>	<i>Med</i>	<i>Mean</i>	<i>Q3</i>	<i>Max</i>
0.50	26	35	39.39	50	139

TABLE 1.1 – Résumé statistique

En statistique classique, on s’intéresse souvent au comportement des données au voisinage de leur moyenne. Sur la base de la loi des grands nombres, la moyenne de l’échantillon ou moyenne empirique \bar{X} est utilisée comme estimateur consistant de la moyenne $E(X)$. De plus, le théorème de la limite centrale assure un comportement asymptotiquement normal de la moyenne de l’échantillon. Ce résultat peut être utilisé pour construire un intervalle de confiance pour $E(X)$ pour des tailles d’échantillons suffisamment grandes. Pour les données sur la vitesse du vent à Albuquerque, ces techniques conduisent à une vitesse maximale moyenne du vent de l’ordre de 39.39 miles par heure. De plus, l’intervalle de confiance à 95% pour $E[X]$ est $[38.78; 39.99]$.

Dans notre exemple, il est intéressant d’étudier la distribution des vitesses extrêmes du vent.

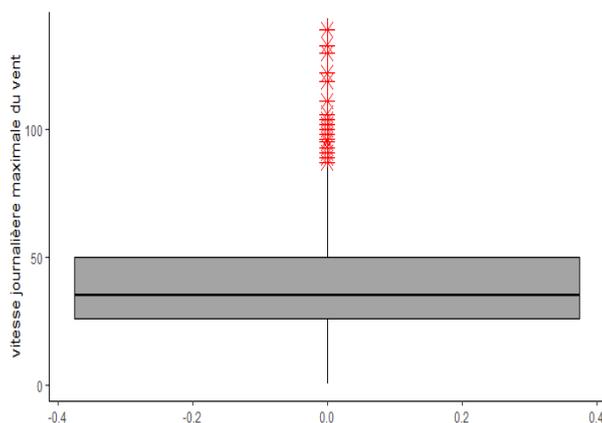


FIGURE 1.2 – Boxplot

Supposons qu'un hangar situé sur un site à Albuquerque ne puisse pas supporter des vitesses de vent supérieures à 40 miles par heure. Il serait alors très instructif d'estimer la probabilité p de l'évènement extrême ($X > 40$) par \hat{p} . Pour cela, on peut utiliser la fonction de répartition empirique \hat{F}_n telle que $\hat{p} = 1 - \hat{F}_n(40)$. Cependant, que se passera-t-il si l'un des moments d'ordre 1 ou 2 n'est pas fini ? Alors le théorème de la limite centrale ne s'applique plus et la théorie classique, dominée par la distribution normale, n'est plus appropriée. Par ailleurs, que se passera-t-il si l'on veut estimer $p = P(X > x_{n,n})$? Dans cette situation, l'estimateur \hat{p} défini plus haut donne la valeur 0. Les dommages causés au hangar par les vents extrêmes peuvent être importants, voire catastrophiques. De toute évidence, il ne suffit pas de supposer que ces valeurs extrêmes de la vitesse du vent sont impossibles pour résoudre le problème. Cependant, les techniques traditionnelles basées sur la fonction de répartition empirique ne fournissent pas d'information utile concernant ce type de question. Ces observations montrent qu'il est nécessaire de développer des techniques particulières basées sur les valeurs extrêmes d'un échantillon, sur des quantiles extrêmement élevés ou sur les petites probabilités de queue. Dans des situations pratiques, ces valeurs extrêmes présentent souvent un intérêt majeur. L'exemple de vitesse du vent ne fournit qu'une illustration, mais il existe de nombreuses autres situations où le raisonnement par valeurs extrêmes est primordial.

1.2 Outils graphiques pour l'identification de distributions

Dans cette section, nous allons présenter quelques techniques graphiques qui, combinées avec les informations a priori sur le problème étudié, peuvent aider à identifier le type de distribution adéquat. Ces outils graphiques nous aideront à élaborer un modèle raisonnable pour décrire la population statistique fondamentale. Nous n'insisterons pas sur les modèles globaux qui visent à décrire les données dans leur intégralité. On procédera cependant à des ajustements statistiques au-dessus de certains seuils élevés.

Nous rappellerons d'abord les techniques graphiques de la statistique traditionnelle telles que les tracés d'histogrammes et les estimations de densité. Par ailleurs, on insistera fortement sur la méthode Quantile-Quantile plot (QQ-plot) et celle du Log-Log plot qui seront souvent plus utiles pour notre objectif.

1.2.1 Quantile-Quantile plot

Le Quantile-Quantile plot ou plus succinctement QQ-plot, est une technique de représentation graphique apparue dans les années 1960 ([Kimball \(1960\)](#)), qui permet de mesurer graphiquement la proximité d'une distribution théorique à une distribution empirique. En effet, à partir de l'échantillon observé, on définit alors un certain nombre de quantiles $x_{i,n}$. Si l'échantillon est distribué suivant le modèle théorique choisi, on devrait avoir les quantiles $x_{i,n}$ observés égaux aux quantiles $x_{i,n}^*$ associés au modèle théorique.

On représente graphiquement l'ensemble des points $\{M_i(x_{i,n}^*, x_{i,n}), i = 1, 2, \dots, n\}$. Si la distribution théorique choisie est pertinente, le graphe de cet ensemble de points sera confondu avec la première diagonale. Cependant, dans la pratique la linéarité dans un graphique n'est toujours évidente. Il est possible de confirmer à l'aide du coefficient de corrélation ([Beirlant et al. \(2006a\)](#), [Embrechts et al. \(2013\)](#), [Thas \(2010\)](#)).

Nous allons illustrer l'idée d'un QQ-plot en prenant pour exemple le modèle exponentiel. La classe des distributions $Exp(\lambda)$ a pour fonction de survie

$$1 - F_\lambda(x) := \exp(-\lambda x).$$

On voudrait savoir si la distribution empirique \hat{F}_n des données dans le tableau [1.1](#) appartient à cette classe. La réponse doit se fonder sur les observations x_1, \dots, x_n . La fonction

quantile pour la distribution exponentielle se présente sous la forme simple suivante

$$Q_\lambda(p) = -\frac{1}{\lambda} \log(1 - p) \text{ pour } p \in (0, 1).$$

A partir de l'échantillon x_1, \dots, x_n , on estime la fonction quantile Q par la fonction quantile empirique \hat{Q}_n . Ensuite, on trace l'ensemble des points $\left\{ \left(-\log(1 - p), \hat{Q}_n(p) \right) \right\}$ pour un ensemble de valeurs de $p \in (0, 1)$. L'ensemble des valeurs $\left\{ \frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1 \right\}$ est un choix pratique pour les différentes valeurs de p .

Si l'ajustement des données au modèle exponentiel est correct alors le QQ-plot a la forme d'une droite. La pente de cette droite peut être utilisée comme un estimateur du paramètre λ^{-1} . En effet, si le modèle est correct, alors l'équation

$$Q_\lambda(p) = \frac{1}{\lambda} (-\log(1 - p)).$$

Remarque 1.2.1. *Un choix alternatif pour les différentes valeurs de p dans l'ensemble $\left\{ \frac{1-0.5}{n}, \frac{2-0.5}{n}, \dots, \frac{n-1-0.5}{n} \right\}$ permet d'éviter les problèmes de débordement à $p = 1$.*

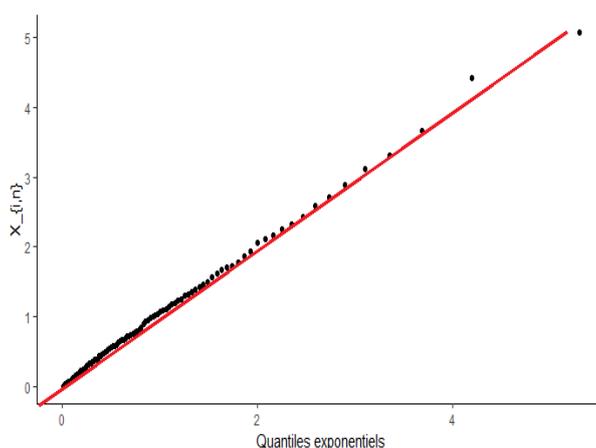


FIGURE 1.3 – QQ-plot exponentiel

La même méthodologie peut alors être appliquée afin de déterminer des QQ-plots pour le modèle *Normal*, le modèle *Paréto* ou autres. Dans le tableau suivant nous listons les QQ-plots de quelques distributions classiques

Distribution	$F(x)$	coordonnée
Normale	$\int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(u-\mu)^2}{2\sigma^2}\right) du$ $x \in \mathbb{R}; \mu \in \mathbb{R}, \sigma > 0$	$(\Phi^{-1}(p_{i,n}), x_{i,n})$
Log-normale	$\int_0^x \frac{1}{\sqrt{2\pi}\sigma u} \exp\left(-\frac{(\log u - \mu)^2}{2\sigma^2}\right) du$ $x \in \mathbb{R}; \mu \in \mathbb{R}, \sigma > 0$	$(\Phi^{-1}(p_{i,n}), \log x_{i,n})$
Exponentielle	$1 - \exp(-\lambda x)$ $x > 0; \lambda > 0$	$(-\log(1 - p_{i,n}), x_{i,n})$
Pareto	$1 - \left(1 + \xi \frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}$ $1 + \xi \frac{(x-\mu)}{\sigma} > 0, \xi \neq 0$	$(-\log(1 - p_{i,n}), \log x_{i,n})$
Weibull	$1 - \exp\left(1 + \xi \frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}$ $1 + \xi \frac{(x-\mu)}{\sigma} > 0, \xi \neq 0$	$(\log(-\log(1 - p_{i,n})), \log x_{i,n})$

TABLE 1.2 – QQ-plots pour quelques distributions usuelles

1.2.1.1 Etude d'un QQ-plot exponentiel

Pour un exemple pratique d'application de la méthode ci-dessus, prenons les mesures quotidiennes de la vitesse maximale du vent obtenues à Zaventem, en Belgique figure 1.4.

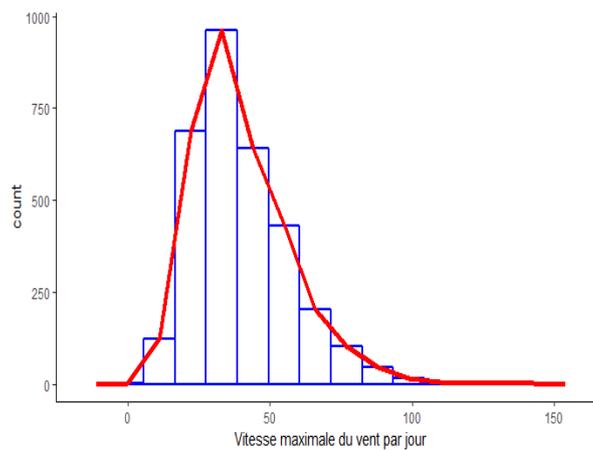


FIGURE 1.4 – Mesure quotidienne de la vitesse maximale du vent à Zaventem (Belgique) De 1985 à 1992.

On se limite aux données supérieures à $t = 82\text{km/h}$. On remarque que l'histogramme exhibe une forme exponentielle décroissante. figure 1.5

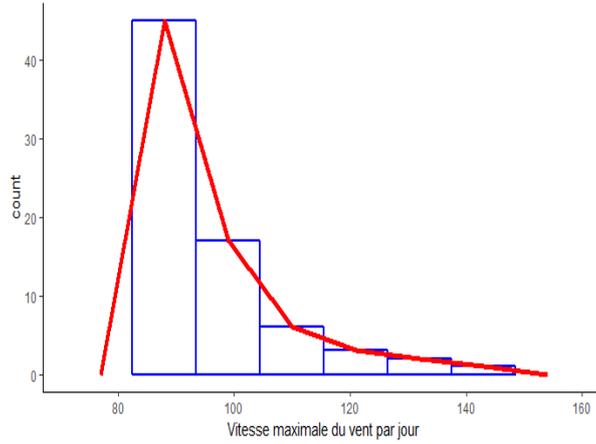


FIGURE 1.5 – Histogramme conditionnel des vitesses de vent supérieures à 82km/h.

Le QQ-plot exponentiel montre une droite avec une ordonnée à l'origine 82. Un comportement similaire, mais avec des valeurs différentes de λ a été montré pour la plupart des villes en Belgique. figure 1.6

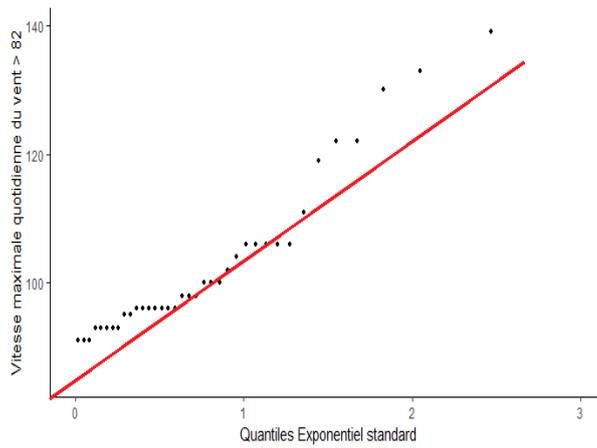


FIGURE 1.6 – QQ-plot exponentiel pour des vitesses de vent supérieures à 82km/h.

La qualité de l'ajustement peut être mesurée à l'aide du coefficient de corrélation

$$r_Q = \frac{\sum_{i=1}^n (x_{i,n} - \bar{x})(q_{i,n} - \bar{q})}{\sqrt{\sum_{i=1}^n (q_{i,n} - \bar{q})^2 \sum_{i=1}^n (x_{i,n} - \bar{x})^2}},$$

La quantité r_q satisfait toujours l'inégalité $0 \leq r_q \leq 1$. Effectivement, puisque les $x_{i,n}$ et les $q_{i,n}$ sont croissants, le coefficient de corrélation sera non négatif. De plus, $r_q = 1$ si et seulement si l'ajustement est parfait. Par conséquent, r_q peut être utilisé comme mesure de l'ajustement du modèle exponentiel aux données. Un test de signification peut être basé sur la statistique r_q et on rejette l'hypothèse de l'exponentialité lorsque la valeur de r_q diffère significativement de 1. Nous allons généraliser les résultats obtenus dans le cas d'un QQ-plot exponentiel pour pouvoir le définir pour d'autres distributions.

Notons Q la fonction quantile de la distribution standard dans un modèle paramétrique. Pour vérifier l'ajustement de la distribution empirique \widehat{Q}_n d'un corpus de données sur Q , on procède de la manière suivante. On trace le graphe $\left\{Q\left(\frac{i}{n+1}\right); \widehat{Q}_n\left(\frac{i}{n+1}\right) = x_{i,n}\right\}$, $i = 1, 2, \dots, n$; puis on vérifie la linéarité de ce graphe en le comparant à la première bissectrice. Une forte linéarité implique un bon ajustement.

Il est possible de procéder à un ajustement linéaire $y = \widehat{b} + \widehat{a}x$ sur le QQ-plot, avec

$$\widehat{a} = \frac{\sum_{i=1}^n (x_{i,n} - \bar{x}) Q\left(\frac{i}{n+1}\right)}{\sum_{i=1}^n \left(Q\left(\frac{i}{n+1}\right) - \bar{q}\right)^2},$$

$$\widehat{b} = \bar{x} - \widehat{a}\bar{q},$$

où $\bar{q} = \frac{1}{n} \sum_{i=1}^n Q\left(\frac{i}{n+1}\right)$.

Les QQ-plots peuvent être utilisés pour d'autres distributions que la distribution exponentielle. En effet, les QQ-plots peuvent être utilisés pour évaluer l'adéquation de tout modèle statistique à un corpus de données. Les QQ-plots les plus courants sont listés dans le tableau [1.2](#)

1.2.2 Log-Log plot

Pour les distributions sub-exponentiel (voir la sous section [2.2.2](#)) de moyenne μ , la fonction de survie $\overline{F}(u) = P(X > u)$ est donnée par :

$$\overline{F}(u) = P(X > u) = e^{-\frac{u}{\mu}}.$$

Et pour une distribution à variations régulières (voir la sous section 2.2.4) on a :

$$\bar{F}(u) = P(X > u) \approx C \int_u^\infty \frac{1}{\alpha - 1} u^{-\alpha+1}.$$

En considérant $\log(P(X > u))$, on obtient respectivement pour les deux types de distributions $-\frac{u}{\mu}$ et $\log(C') - (\alpha - 1)\log(u)$. Ainsi, en portant sur un graphique les valeurs de $\log(P(X > u))$ en fonction de $\log u$, on devrait obtenir une courbe linéaire pour une distribution de la classe de des distributions à variations régulières, et concave pour une distribution qui n'est pas à variations régulières c'est-à-dire une distribution de type subexponentiel ou exponentiel. En pratique, on calcule $\log [P(X > u)]$ à l'aide de la fonction de répartition empirique pour certaines valeurs de u . On trace ensuite le graphique Log-Log de la probabilité de la queue de la distribution et la décision est prise à partir de l'allure de la courbe (voir la figure 1.7).

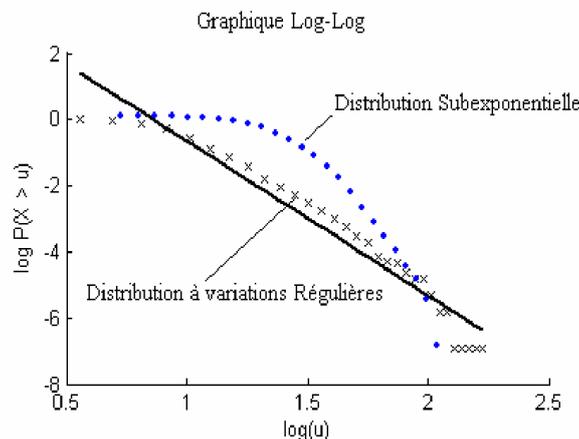


FIGURE 1.7 – Illustration du graphique Log-Log

Cette technique utilise le même principe que le graphique de QQ-plot. La différence entre les deux techniques et que pour le graphique Log-Log on n'a pas besoin d'estimer les paramètres de la distribution à ajuster, étape indispensable pour le QQ-plot où on doit calculer les estimateurs des quantiles à partir de la distribution considérée. Étant donnée que l'objectif est de déterminer une classe qui représente le mieux la série de données sans préciser une distribution particulière, la méthode Log-Log est plus adéquate pour répondre à ce but que le QQ-plot.

Chapitre 2

Approche probabiliste des valeurs extrêmes

L'étude des valeurs extrêmes est l'étude des queues de distributions. Elle repose sur l'analyse des plus grandes observations d'un échantillon. En ce sens, nous pouvons considérer la théorie des valeurs extrêmes comme la contrepartie de la théorie de la statistique classique, qui est principalement basée sur l'étude du comportement des valeurs autour de leur moyenne. Dans le cadre de la statistique classique les valeurs aberrantes ne sont pas prises en compte. Par contre, elles font l'objet de toutes les attentions en théorie des valeurs extrêmes. Le théorème des valeurs extrêmes ([Fisher and Tippett \(1928\)](#)) donne la distribution asymptotique du maximum d'un l'échantillon. Les distributions limites possibles obtenues sont paramétrées par l'indice des valeurs extrêmes ξ . Cet indice contrôle la lourdeur de la queue de distribution. La valeur de l'indice ξ permet de répartir les distributions limites en trois domaines d'attraction, Fréchet, Weibull et Gumbel. Cette théorie est apparue entre 1920 et 1940, grâce aux travaux de Fisher-Tippet [Fisher and Tippett \(1928\)](#), de Gumbel [Gumbel \(1935\)](#)) et de Gnedenko ([Gnedenko \(1943\)](#)). Par ailleurs, si on considère la valeur extrême d'une variable aléatoire comme un excès au dessus d'un seuil élevé, alors la distribution asymptotique de ce type d'excès, quand le seuil tend vers le point terminal, est donnée par le théorème de Pickands ([Pickands III et al. \(1975\)](#)).

2.1 Etude de quelques distributions de probabilités

Les relations donnant les principaux moments centrés et non centrés ainsi que les coefficients d'asymétrie et d'aplatissement des distributions de Fréchet, de Weibull et de Gumbel sont présentés dans cette section.

2.1.1 Distribution de probabilité de Fréchet

La distribution dite de Fréchet a été introduite par le mathématicien français Maurice René Fréchet (Fréchet (1928)). Il a développée cette distribution comme loi de valeurs extrêmes. Les travaux de Fréchet ont inspiré les recherches de Fisher-Tippett (Fisher and Tippett (1928)) et ceux de Gumbel en 1958.

La distribution de Fréchet est l'une des distributions les plus populaires dans de nombreux domaines, par exemple en hydrologie, la distribution de Fréchet est utilisée pour des évènements extrêmes tels que le maximum annuel des précipitations journalières ou le débit des rivières (Coles et al. (2001)). En finance et macroéconomie, la distribution de Fréchet est utilisée pour calculer la valeur à risque (Gilli et al. (2006)).

Définition 2.1.1. *On dit qu'une variable aléatoire X est de distribution de Fréchet de paramètres $\mu, \sigma > 0, \xi > 0$, si sa densité de probabilité s'écrit telle que :*

$$h_{\xi,\sigma,\mu}(x) = \begin{cases} 0 & \text{si } x < \frac{1}{\xi} \\ \frac{\xi}{\sigma} \left(\frac{x-\mu}{\sigma}\right)^{-1-\xi} \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}\right) & \text{si } x \geq \frac{1}{\xi}, \end{cases} \quad (2.1)$$

où ξ est le paramètre de forme, σ est le paramètre d'échelle et μ est le paramètre de position.

Dans la figure 2.1 nous allons visualiser l'effet de σ et de μ sur la densité de la distribution de Fréchet

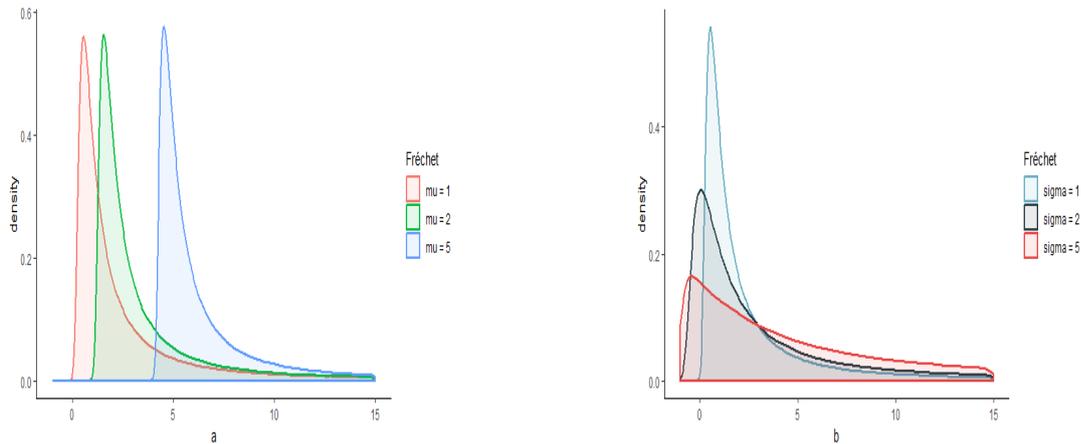


FIGURE 2.1 – Fonction de densité de Fréchet : (a) pour différentes valeurs de μ , (b) pour différentes valeurs de σ

Il est clair sur la figure 2.1(a) que la moyenne μ joue un rôle de translation. Par contre, sur la figure 2.1(b) le paramètre d'échelle σ contrôle l'aplatissement de la distribution.

Remarque 2.1.1. *On constate sur la figure 2.1 que la fonction de densité de Fréchet est bornée à gauche. De plus, la distribution de Fréchet admet un Kurtosis positif, donc sa médiane est supérieure à sa moyenne.*

Définition 2.1.2. *La fonction de répartition de Fréchet s'écrit telle que :*

$$H_{\xi,\sigma,\mu}(x) = \Pr(X \leq x) = \begin{cases} 0 & \text{si } x < \frac{1}{\xi} \\ \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) & \text{si } x \geq \frac{1}{\xi}, \end{cases} \quad (2.2)$$

Définition 2.1.3. *La fonction de survie de Fréchet s'écrit telle que :*

$$\bar{H}_{\xi,\sigma,\mu}(x) = 1 - H_{\xi,\sigma,\mu}(x) = \begin{cases} 0 & \text{si } x < \frac{1}{\xi} \\ 1 - \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) & \text{si } x \geq \frac{1}{\xi}. \end{cases}$$

Dans figure 2.2 nous visualisons l'effet du paramètre de queue ξ sur la queue de la fonction de survie de Fréchet.

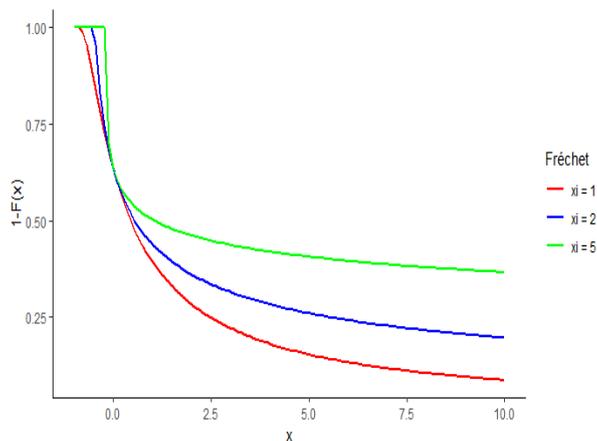


FIGURE 2.2 – Fonction de survie de Fréchet pour différents valeurs de ξ

Il est clair que la queue de la distribution de Fréchet s'alourdit d'autant plus que le paramètre ξ augmente.

Moments d'ordre k

Soit X une variable aléatoire distribuée selon une distribution Fréchet (2.1). Il est facile de montrer que la moyenne $E(X)$ et la variance $Var(X)$ de cette variable sont données respectivement par :

$$\begin{aligned} E(X) &= \mu + \frac{\xi}{\sigma} [1 - \Gamma(1 + \xi)] \\ Var(X) &= \left(\frac{\sigma}{\xi}\right)^2 [\Gamma(1 + 2\xi) - \Gamma^2(1 + \xi)]. \end{aligned}$$

Les moments d'ordre 3 et 4 est déterminé tel que :

$$\begin{aligned} \mu_3(X) &= -\left(\frac{\sigma}{\xi}\right)^3 [\Gamma(1 + 3\xi) - 3\Gamma(1 + 2\xi)\Gamma(1 + \xi) + 2\Gamma^3(1 + \xi)] \\ \mu_4(X) &= \left(\frac{\sigma}{\xi}\right)^4 [\Gamma(1 + 4\xi) - 4\Gamma(1 + \xi)\Gamma(1 + 3\xi) \\ &\quad + 6\Gamma^2(1 + \xi)\Gamma(1 + 2\xi) - 3\Gamma^4(1 + \xi)]. \end{aligned}$$

Coefficients de Skewness et de Kurtosis

En utilisant les résultats précédant, on peut déduire le coefficient S de Skewness ou d'asymétrie

$$S = \frac{\mu_3}{\mu_2^{3/2}} = -\frac{|\xi| [\Gamma(1 + 3\xi) - 3\Gamma(1 + 2\xi)\Gamma(1 + \xi) + 2\Gamma^3(1 + \xi)]}{\xi [\Gamma(1 + 2\xi) - \Gamma^2(1 + \xi)]^{3/2}},$$

ainsi que le coefficient K de Kurtosis ou d'aplatissement

$$K = \frac{\mu_4}{\mu_2^2} = \frac{[\Gamma(1 + 4\xi) - 4\Gamma(1 + \xi)\Gamma(1 + 3\xi) + 6\Gamma^2(1 + \xi)\Gamma(1 + 2\xi) - 3\Gamma^4(1 + \xi)]}{[\Gamma(1 + 2\xi) - \Gamma^2(1 + \xi)]^2}.$$

Il est clair que les coefficients S et K ne dépendent que du paramètre de forme ξ .

Dans le tableau 2.1 nous donnons les valeurs des coefficients d'asymétrie S et d'aplatissement K pour différentes valeurs du ξ .

ξ	S	K
0.05	0.8679651	4.267201
1	-2	87.72
3	-19.5848590	1005.532
5	-190.1132395	186183.1

TABLE 2.1 – Variation de S et K en fonction de ξ

2.1.2 Distribution de probabilité de Weibull

La distribution Weibull a été introduite par le mathématicien suédois Waloddi Weibull (Weibull (1951)). De nos jours, cette distribution est couramment utilisée dans l'analyse de survie et en fiabilité des systèmes. La distribution de Weibull peut également être utilisée dans de nombreux autres domaines, notamment la biologie, l'économie, l'industrie, l'hydrologie, etc. (Rinne (2008)).

Définition 2.1.4. *On dit qu'une variable aléatoire X est de distribution de Weibull de paramètres $\mu, \sigma > 0, \xi < 0$, si sa densité de probabilité s'écrit telle que :*

$$h_{\xi, \sigma, \mu}(x) = \begin{cases} 0 & \text{si } x > \frac{1}{\xi} \\ \frac{\xi}{\sigma} \left(\frac{x-\mu}{\sigma}\right)^{\xi-1} \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{\xi}\right) & \text{si } x \leq \frac{1}{\xi}, \end{cases} \quad (2.3)$$

où ξ est le paramètre de forme, σ le paramètre d'échelle et μ le paramètre de position.

Dans la figure 2.3 nous allons visualiser l'effet de σ et de μ sur la densité de la distribution de Weibull.

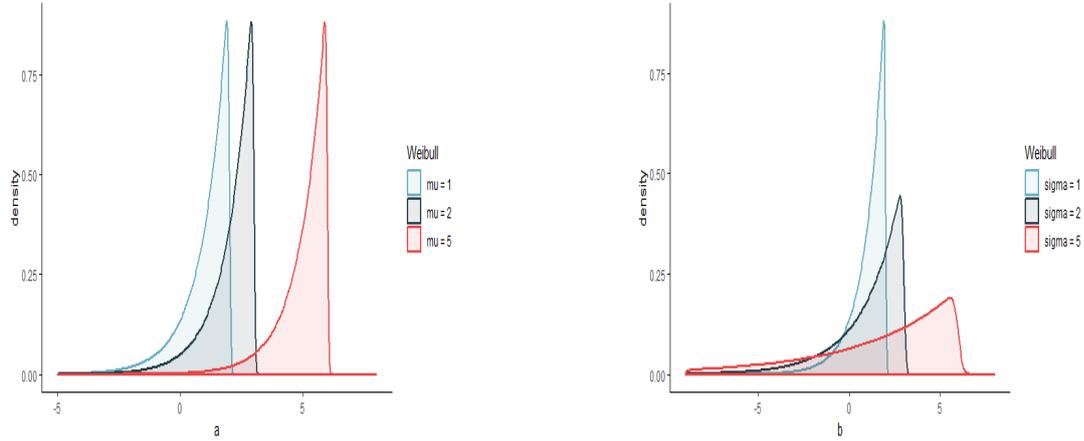


FIGURE 2.3 – Fonction de densité de Weibull : (a) pour différentes valeurs de μ , (b) pour différentes valeurs de σ

Il est clair sur la figure 2.3(a) que la moyenne μ joue un rôle de translation. Par contre, sur la figure 2.3(b) le paramètre d'échelle σ contrôle l'aplatissement de la distribution.

Remarque 2.1.2. *On constate sur la figure 2.3 que la fonction de densité de Weibull est bornée à droite. De plus, la distribution de Weibull admet un Kurtosis négatif, donc sa médiane est inférieure à sa moyenne.*

Définition 2.1.5. *La fonction de distribution de Weibull est donnée par la formule suivante :*

$$H_{\xi,\sigma,\mu}(x) = \Pr(X \leq x) = \begin{cases} 0 & \text{si } x > \frac{1}{\xi} \\ \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) & \text{si } x \leq \frac{1}{\xi}. \end{cases} \quad (2.4)$$

Définition 2.1.6. *La fonction de survie de Weibull est définie telle que :*

$$\bar{H}_{\xi,\sigma,\mu}(x) = 1 - H_{\xi,\sigma,\mu}(x) = \begin{cases} 0 & \text{si } x > \frac{1}{\xi} \\ 1 - \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) & \text{si } x \leq \frac{1}{\xi}, \end{cases}$$

Dans figure 2.4 nous visualisons l'effet du paramètre de queue ξ sur la queue de la fonction de survie de Weibull.

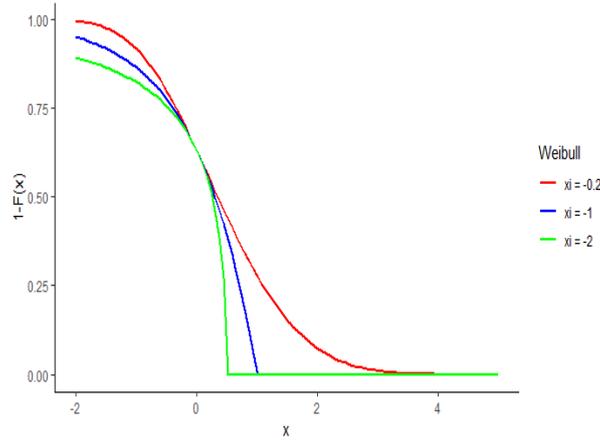


FIGURE 2.4 – Fonction de densité de Weibull pour différents valeurs de ξ

Il est clair que la queue de la distribution de Weibull s'alourdit d'autant plus que le paramètre ξ diminue.

Moments d'ordre k

Soit X une variable aléatoire distribuée suivant une loi de Weibull (2.3). On peut montrer (Britain) (1975)) que la moyenne $E(X)$ et la variance $Var(X)$ de cette variable sont données respectivement par :

$$E(X) = \mu + \frac{\xi}{\sigma} [1 - \Gamma(1 + \xi)]$$

$$Var(X) = \left(\frac{\sigma}{\xi}\right)^2 [\Gamma(1 + 2\xi) - \Gamma^2(1 + \xi)].$$

Les moments d'ordre 3 et 4 sont déterminés tels que :

$$\mu_3(X) = -\left(\frac{\sigma}{\xi}\right)^3 [\Gamma(1 + 3\xi) - 3\Gamma(1 + 2\xi)\Gamma(1 + \xi) + 2\Gamma^3(1 + \xi)]$$

$$\mu_4(X) = \left(\frac{\sigma}{\xi}\right)^4 [\Gamma(1 + 4\xi) - 4\Gamma(1 + \xi)\Gamma(1 + 3\xi) + 6\Gamma^2(1 + \xi)\Gamma(1 + 2\xi) - 3\Gamma^4(1 + \xi)].$$

Coefficients de Skewness et de Kurtosis

Utilisant les résultats précédant, on peut déduire le coefficient S de Skewness ou d'asymétrie

$$S = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = -\frac{|\xi| [\Gamma(1+3\xi) - 3\Gamma(1+2\xi)\Gamma(1+\xi) + 2\Gamma^3(1+\xi)]}{\xi [\Gamma(1+2\xi) - \Gamma^2(1+\xi)]^{\frac{3}{2}}},$$

ainsi que le coefficient K de Kurtosis ou d'aplatissement

$$K = \frac{\mu_4}{\mu_2^2} = \frac{[\Gamma(1+4\xi) - 4\Gamma(1+\xi)\Gamma(1+3\xi) + 6\Gamma^2(1+\xi)\Gamma(1+2\xi) - 3\Gamma^4(1+\xi)]}{[\Gamma(1+2\xi) - \Gamma^2(1+\xi)]^2},$$

Il est clair que les coefficients S et K ne dépendent que du paramètre de forme ξ .

Dans le tableau 2.2 nous donnons les valeurs des coefficients d'asymétrie S et d'aplatissement K pour différentes valeurs du ξ .

ξ	S	K
-0.05	1.473884	7.333494
-0.1	1.910339	10.978566
-0.2	3.535072	48.091512
-0.3	13.483552	-145.295346

TABLE 2.2 – Variation de S et K en fonction de ξ

2.1.3 Distribution de probabilité de Gumbel

La distribution Gumbel a été introduit par le mathématicien Américain Émil Julius Gumbel (Gumbel (2012)). Il a développée cette distribution comme loi de valeurs extrêmes. La distribution de Gumbel est l'une des distributions les plus populaires dans de nombreux domaines grâce à sa flexibilité.

Définition 2.1.7. *On dit qu'une variable aléatoire X est de distribution de Gumbel de paramètres μ et $\sigma > 0$, si sa densité de probabilité s'écrit telle que :*

$$f(x) = \frac{1}{\sigma} \exp \left\{ -\frac{(x-\mu)}{\sigma} - \exp \left(-\frac{(x-\mu)}{\sigma} \right) \right\}, \quad (2.5)$$

où σ est le paramètre d'échelle et μ le paramètre de position.

Dans la figure 2.5 nous allons visualiser l'effet de σ et de μ sur la densité de la distribution de Gumbel.

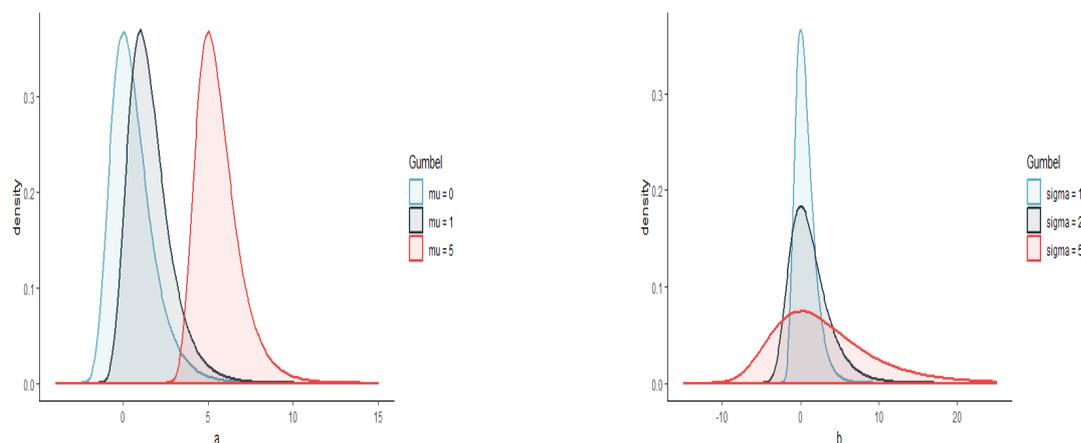


FIGURE 2.5 – Fonction de densité de Gumbel : (a) pour différentes valeurs de μ , (b) pour différentes valeurs de σ

Il est clair sur la figure 2.5(a) que la moyenne μ joue un rôle de translation. Par contre, sur la figure 2.5(b) le paramètre d'échelle σ contrôle l'aplatissement de la distribution.

Définition 2.1.8. La fonction de répartition s'écrit telle que :

$$F(x) = \exp \left\{ - \exp \left(- \frac{x - \mu}{\sigma} \right) \right\} \quad (2.6)$$

Moments d'ordre k

Soit X une variable aléatoire distribuée suivant une loi de Gumbel (2.5). On peut montrer (Gumbel (2012)) que la moyenne $E(X)$ et la variance $Var(X)$ de cette variable sont données respectivement par :

$$\begin{aligned} E(X) &= \mu + \sigma C \\ Var(X) &= \frac{\sigma^2 \pi^2}{6}, \end{aligned}$$

où $C = 0.5772..$ est la constante de Euler.

Coefficients de Skewness et de Kurtosis Les coefficients de Skewness ou d'asymétrie S et de Kurtosis ou d'aplatissement K sont donnés respectivement par :

$$\begin{aligned} S &= 1.139 \\ K &= 2.40. \end{aligned}$$

2.2 Classes de distributions à queue lourde

Dans cette section on introduit la notion de distribution à queue lourde. Les distributions à queues lourdes sont liées à la théorie des valeurs extrêmes et permettent de modéliser beaucoup de phénomènes que l'on trouve dans différentes disciplines telles que la finance, l'hydrologie, la climatologie épidémiologie, etc. Plusieurs définitions ont été proposées pour ces distributions en fonction de certains critères de classification. La caractérisation la plus simple et celle basée sur la comparaison avec la distribution normale (El Adlouni et al. (2007)).

Définition 2.2.1. *On dit qu'une distribution est à queue lourde si son Kurtosis est tel que :*

$$K = \frac{\mu_4}{\mu_2^2} > 3, \quad (2.7)$$

où μ_4 et μ_2 sont les moments centrés d'ordre respectifs 4 et 2 de X

Ce qui est équivalent à dire qu'une distribution est à queue lourde si et seulement si son coefficient d'aplatissement K est supérieur à celui de la distribution normale (Hubert and Bendjoudi (1996)). La différence entre la distribution normale et une distribution à queue lourde est illustrée dans la figure 2.6. La formulation dans l'équation (2.7) est très générale, et ne peut être valable que si le moment d'ordre 4 existe.

La famille des distributions à queue lourde peut être répartie en un ensemble de sous-familles, les distributions n'admettent aucun moments (E), les distributions sub-exponentielles (D), les distributions à variations régulières (\mathcal{R}_ρ), les distributions avec un comportement de Pareto (B), les distributions α -stables avec ($\alpha < 2$) (\mathcal{A}). Toutes ces classes sont emboîtées (BUND (2002)) telles que :

$$(\mathcal{A}) \subset (\mathcal{B}) \subset (\mathcal{R}_\rho) \subset (D) \subset (E)$$

Dans la figure 2.6 on représente les fonctions de densité de probabilité de la distribution normale et d'une distribution à queue plus lourde.

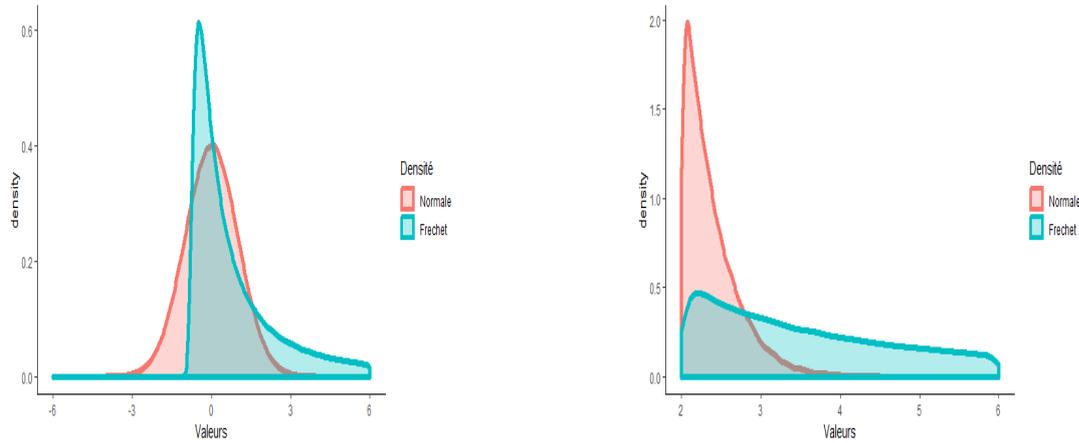


FIGURE 2.6 – Comparaison d’une distribution normale à une distribution à queue lourde

On remarque que la fonction de densité de la distribution normale s’aplatit plus rapidement que celle de Fréchet.

2.2.1 Distributions n’admettant aucun moments

Définition 2.2.2. Une distribution F est dite à queue lourde ou à queue épaisse si sa fonction de répartition vérifie

$$E(e^{\lambda x}) = \int_{\mathbb{R}} e^{\lambda x} dF(x) = \infty, \quad \text{pour tout } \lambda > 0.$$

La classe (E) englobe toutes les distributions qui sont caractérisées par la définition 2.2.2, (pour plus de détail voir [Foss et al. \(2011\)](#)). On note que la distribution normale n’appartient pas à cette classe parce que la probabilité au dépassement \bar{F} pour les extrêmes de cette classe, décroît moins rapidement que celle de la distribution normale. Dans ce sens, la classe (E) est celle des distributions ayant une queue plus lourde que celle de la distribution normale.

2.2.2 Distributions sub-exponentielles

La classe (D) des distributions sub-exponentielles est caractérisée par la définition suivante ([Beirlant et al. \(2006a\)](#)) :

Définition 2.2.3. On dit qu'une distribution est sub-exponentielle si :

$$\lim_{x \rightarrow \infty} \frac{P(X_1 + \dots + X_n > x)}{P(\max(X_1 + \dots + X_n) > x)} = 1 \quad (2.8)$$

C'est équivalent à dire que la somme de n distributions sub-exponentielles indépendantes et identiquement distribuées est extrême si et seulement si leur maximum est extrême. On peut démontrer que l'équation (2.8) implique que :

$$\lim_{x \rightarrow \infty} \frac{\overline{F}(x)}{e^{-\epsilon x}} = \infty, \quad \forall \epsilon > 0.$$

On rappelle que e^{-tx} est la forme de la queue de la distribution exponentielle. Comme son nom l'indique, la classe (D) contient les distributions telles que F décroît plus lentement que n'importe quelle distribution exponentielle. La différence entre la distribution Exponentielle et une distribution avec une queue plus lourde a été illustrée dans la figure 2.7.

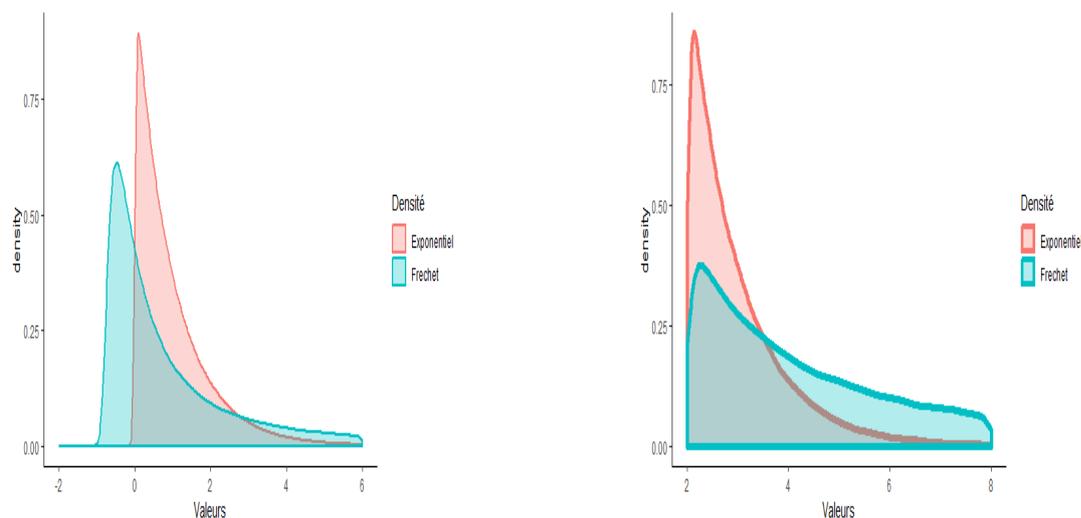


FIGURE 2.7 – Illustration de la différence entre la distribution exponentielle et une distribution à queue lourde (Fréchet)

Dans la figure 2.7 on présente les fonctions de densité de probabilité de la distribution exponentielle et d'une distribution à queue plus lourde. On remarque que la fonction de densité de la distribution exponentielle décroît rapidement que celle de la distribution Fréchet.

2.2.3 Fonctions α -stables

Les distributions \mathcal{A} des distributions α -stables constituent une classe très riche de distributions de probabilité capables de représenter différentes asymétries et des queues très lourdes. Cette classe a été caractérisée par (Lévy (1925)) dans ces travaux sur la somme de variables indépendantes et identiquement distribuées. L'absence de formules explicites des densités de ces distributions a limité leur utilisation. Les distributions de cette classe ont un comportement asymptotique de Pareto avec $0 < \alpha \leq 2$. Lorsque $\alpha = 2$ on retrouve la distribution normale, cependant pour $\alpha < 2$ le moment d'ordre $r \geq \alpha$ n'est pas fini, ces distributions ont donc une variance infinie, et par conséquent une queue très lourde. Pour $\alpha = 1$ on retrouve la distribution de Cauchy (avec une moyenne et une variance infinies). Cette classe a une grande importance dans la théorie des valeurs extrêmes, puisque les distributions stables peuvent être caractérisées à partir du théorème de la Limite Centrale Généralisé. En effet, le théorème de la limite centrale indique que la somme de n variables aléatoires indépendantes et identiquement distribuées de variance finie tend vers une distribution normale lorsque n tend vers l'infini. Une généralisation de ce théorème par (Gnedenko and Kolmogorov (1954)) indique que si la condition de variance finie n'est pas respectée, la seule distribution limite possible de la somme de n variables aléatoire indépendantes et identiquement distribuées, est une distribution stable. L'utilisation de cette famille de distributions était très limitée à cause des problèmes de calcul : leurs fonctions de densité de probabilité et de distributions n'étant pas explicites. Récemment plusieurs logiciels ont été proposés pour permettre la résolution de ces problèmes citetnolan1997maximum, et on trouve des applications des distributions stables dans plusieurs domaines tels que finances, physique et le trafic internet.

2.2.4 Fonctions à variation régulière

Nous abordons dans cette partie une classe de fonctions qui apparaissent dans un grand nombre d'applications dans l'ensemble des mathématiques. Cette classe est intimement liée à la classe des fonctions 'puissance'. On donne d'abord quelques généralités, ensuite on établit un certain nombre de propriétés fondamentales. Et on termine avec des propriétés qui sont particulièrement importantes dans la suite.

Définition 2.2.4. Soit f une fonction positive et mesurable sur \mathbb{R}_+ . On dit que f est à

variation régulière si et seulement s'il existe une constante réelle ρ telle que

$$\lim_{x \uparrow \infty} \frac{f(xt)}{f(x)} = t^\rho \quad \forall t > 0 \quad (2.9)$$

Définition 2.2.5. (Variation régulière du second ordre)

On dit que F est à variation régulière du second d'ordre si elle satisfait les conditions équivalentes suivantes :

- Il existe une constante réelle $\rho \leq 0$ et une fonction de signe constant $A^*(t) \rightarrow 0$ quand $t \rightarrow \infty$, telles que pour tout $x > 0$

$$\lim_{t \rightarrow \infty} \frac{(\overline{F}(tx)) / (\overline{F}(t) - x^{-1/\xi})}{A^*(t)} = x^{-1/\xi} \frac{x^\rho - 1}{\rho} \quad (2.10)$$

- Il existe une constante réelle $\rho \leq 0$ et une fonction de signe constant $A(t) \rightarrow 0$ quand $t \rightarrow \infty$, telles que pour tout $x > 0$

$$\lim_{t \rightarrow \infty} \frac{(U(tx)/U(t)) - x^\xi}{A(t)} = x^\xi \frac{x^\rho - 1}{\rho}. \quad (2.11)$$

On peut interpréter $\frac{x^\rho - 1}{\rho} = \log x$ dans le cas où $\rho = 0$.

Remarque 2.2.1. – L'ensemble des fonctions à variation régulière est noté \mathcal{R}_ρ , où ρ est l'indice de variation régulière.

- Pour $\rho = 0$, \mathcal{R}_0 est appelé ensemble des fonctions à variation lente.
- Dans la suite, tout élément de \mathcal{R}_0 sera noté par le symbole ℓ .

Nous donnons ci-après quelques exemples typiques de fonctions à variation lente

1. $\ell(x) = (\log x)_+^\alpha$ pour α arbitraire.

Plus loin, on omettra le signe (+); 'variation régulière' est en effet un concept asymptotique et donc ne dépend pas de ce qui se passe en des valeurs fixes.

2. $\ell(x) = \prod_{i=1}^k (\log_i x)^{\alpha_i}$ où $\log_1 x = \log x$ tandis que pour $n \geq 1$, $\log_{n+1} x := \log(\log_n x)$.

3. ℓ satisfaisant la condition $\ell(x) \rightarrow c \in (0, \infty)$.

4. $\ell(x) = \exp\{(\log x)^\beta\}$ où $\beta < 1$.

La classe \mathcal{R}_0 a beaucoup de propriétés qui seront constamment utilisées. Certaines démonstrations sont faciles, pour d'autres nous nous référons à la littérature.

Proposition 2.2.1. *Les fonctions à variation lente vérifient les propriétés suivantes :*

1. \mathcal{R}_0 est fermé par rapport à l'addition, la multiplication et la division.
2. Si ℓ est une fonction à variation lente alors ℓ^α est une fonction à variation lente $\forall \alpha \in \mathbb{R}$.
3. Si $\rho \in \mathbb{R}$, alors $f \in \mathcal{R}_\rho$ si et seulement si $f^{-1} \in \mathcal{R}_{-\rho}$.

Les deux résultats les plus importants sur les fonctions appartenant à \mathcal{R}_0 sont donnés dans le théorème suivant dû à Karamata.

Théorème 2.2.1. *(Karamata)*

1. *Théorème de la Convergence Uniforme*

Si $\ell \in \mathcal{R}_0$, alors

$$\lim_{x \uparrow \infty} \frac{\ell(xt)}{\ell(x)} = 1$$

est une convergence uniforme pour $t \in [a, b]$ où $0 < a < b < \infty$.

2. *Théorème de Représentation*

$\ell \in \mathcal{R}_0$ si et seulement si ℓ peut être exprimé sous la forme

$$\ell(x) = c(x) \exp \left\{ \int_1^x \frac{\epsilon(u)}{u} du \right\}$$

où $c(x) \rightarrow c \in (0, \infty)$ et $\epsilon(x) \rightarrow 0$ quand $x \rightarrow \infty$.

La deuxième partie du théorème ci-dessus permet la construction immédiate des éléments de \mathcal{R}_0 . Pour en référer ultérieurement, nous listons d'autres propriétés dans la proposition suivante.

Proposition 2.2.2. *Soit ℓ une fonction à variation lente. Alors*

- 1.

$$\lim_{x \uparrow \infty} \frac{\log \ell(x)}{\log x} = 0 \tag{2.12}$$

2. Pour tout $\delta > 0$ il existe x_δ tel que pour toute constante $A > 0$ et $x > x_\delta$

$$Ax^{-\delta} < \ell(x) < Ax^\delta$$

3. Si $f \in \mathcal{R}_\rho$ avec $\rho > 0$, alors $f(x) \xrightarrow{x \uparrow \infty} +\infty$, tandis que pour $\rho < 0$, $f(x) \xrightarrow{x \uparrow \infty} 0$.

2.3 Distribution asymptotique d'un extrême

Considérons un échantillon aléatoire $\{X_i, 1 \leq i \leq n\}$ d'une distribution F . L'étude statistique traditionnelle autour de la moyenne est basée sur le théorème de la limite centrale et donc souvent débouche sur la distribution normale comme base de l'inférence statistique. En effet, le théorème de la limite centrale classique stipule que la distribution de la quantité suivante

$$\frac{X_1 + X_2 + \dots + X_n - nE(X)}{\sqrt{nVar(X)}},$$

converge pour $n \rightarrow \infty$ vers la distribution normale standard. Plus généralement, le théorème central limite traite de la somme $S_n := X_1 + X_2 + \dots + X_n$ et essaie de trouver des constantes $a_n > 0$ et b_n telles que la variable $Y_n := a_n^{-1}(S_n - b_n)$ tend en loi vers une distribution non-dégénérée. Une fois que la limite est connue, elle peut être utilisée pour approcher l'encombrante distribution de la quantité Y_n .

Deux questions viennent à l'esprit. Une première question est de savoir quelles distributions peuvent apparaître dans la limite. Puis vient la question de savoir pour quelle distribution F une telle limite est atteinte. La réponse révèle que généralement que la distribution normale est obtenue comme une limite pour Y_n si les variables aléatoires $\{X_i, 1 \leq i \leq n\}$ sont indépendantes et identiquement distribuées. Cependant, si la distribution sous-jacente F possède une queue trop lourde, une autre distribution stable apparaît comme limite. Plus précisément, les distributions F de type Pareto avec une variance infinie donneront des limites non-normales pour Y_n . Alors les extrêmes exhibés par un tel échantillon vont affecter Y_n de telle sorte qu'un comportement asymptotique différent du comportement normal est obtenu.

2.3.1 Distributions généralisées des valeurs extrêmes

Jenkinson ([Jenkinson \(1955\)](#)) a proposé ce qu'il a appelé la distribution généralisée des valeurs extrêmes (*GEV*). Cette distribution a l'avantage de combiner les trois types

de lois des valeurs extrêmes étudiées dans la section 2.1 développées par Fisher-Tippett (Fisher and Tippett (1928)). La forme générale de la fonction de répartition d'une *GEV* s'exprime telle que :

$$H_{\mu,\sigma,\xi}(x) = \exp \left\{ - \left(1 + \frac{\xi}{\sigma}(x - \mu) \right)^{-\frac{1}{\xi}} \right\}, \quad -\infty < x < +\infty \quad (2.13)$$

où $\mu, \sigma > 0$ et ξ sont les paramètres de la distribution.

Chacun des trois types de loi des valeurs extrêmes est caractérisé par le signe du paramètre de forme ξ de la loi *GEV*. En effet, pour $\xi \rightarrow 0$, $H_{\mu,\sigma,\xi}(x)$ correspond à la loi de Gumbel donnée par l'équation (2.6) qui est en fait la limite, lorsque ξ tends vers zéro, de l'équation (2.13).

$$H_{\mu,\sigma}(x) = \exp \left\{ - \exp \left(- \frac{x - \mu}{\sigma} \right) \right\}. \quad (2.14)$$

Pour $\xi > 0$, $H_{\mu,\sigma,\xi}(x)$ correspond, à une paramétrisation près, à la loi Fréchet donnée par l'équation (2.2). Dans ce cas, le support de la loi de la variable aléatoire X est borné inférieurement par la valeur $\mu - \frac{\sigma}{\xi}$ avec $\sigma > 0$. Pour $\xi < 0$, $H_{\mu,\sigma,\xi}(x)$ correspond, à une reparamétrisation près, à la loi Weibull donnée par l'équation (2.4). Dans ce cas, le support de la loi de la variable aléatoire X est borné supérieurement par la valeur $\mu - \frac{\sigma}{\xi}$ avec $\sigma > 0$.

Ainsi, on distingue deux formes de fonction de répartition selon la valeur de ξ pour la loi *GEV*, équation (2.13) si $\xi \neq 0$ et équation (2.14) si $\xi = 0$). De la même façon, la fonction de densité de probabilité prend deux formes obtenues en dérivant (2.13) et (2.14) :

$$h_{\mu,\sigma,\xi}(x) = \frac{1}{\sigma} \exp \left\{ - \frac{(x - \mu)}{\sigma} - \exp \left(- \frac{(x - \mu)}{\sigma} \right) \right\}, \quad \text{si } \xi = 0 \quad (2.15)$$

$$h_{\mu,\sigma}(x) = \frac{1}{\sigma} \left(1 + \frac{\xi}{\sigma}(x - \mu) \right)^{-\frac{1}{\xi}-1} \exp \left\{ - \left(1 + \frac{\xi}{\sigma}(x - \mu) \right)^{-\frac{1}{\xi}} \right\}, \quad \text{si } \xi \neq 0 \quad (2.16)$$

Remarque 2.3.1. *On peut montrer qu'une distribution GEV possède*

1. *une espérance finie si et seulement si $\xi < 1$,*
2. *une variance finie si et seulement si $\xi < \frac{1}{2}$,*

3. un moment d'ordre k si et seulement si $\xi < \frac{1}{k}$.

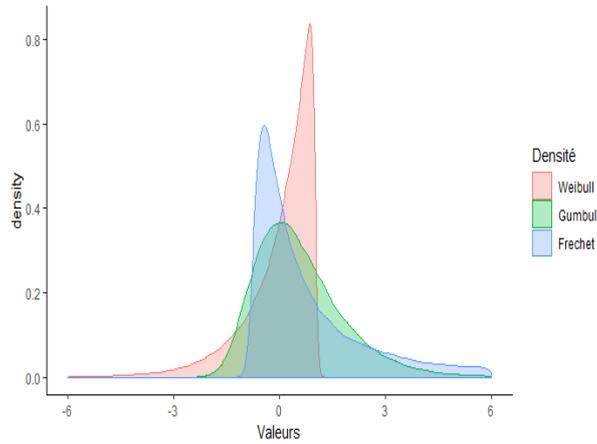


FIGURE 2.8 – Densité de probabilité des Valeurs Extrêmes avec différentes valeurs de ξ

On constate d'après le graphe 2.8 que la fonction de densité de Fréchet est bornée à gauche et il est décalé à gauche de la médiane. Aussi que la densité de Weibull est bornée à droite et il est décalé à droite de la médiane.

2.3.2 Distribution du maximum d'un échantillon

Maintenant on s'intéresse principalement au problème de la distribution du maximum d'un échantillon. Pour cela, on va examiner à la fois les limites possibles et les différentes façons de décrire les ensembles de distributions à partir desquelles la suite des maxima converge vers ces limites.

Considérons un échantillon aléatoire X_1, X_2, \dots, X_n d'une variable aléatoire de fonction de répartition F . Notons le maximum de cet échantillon par $X_{n,n}$ défini tel que

$$X_{n,n} = \max\{X_1, X_2, \dots, X_n\}.$$

Bien sûr, on pourrait tout aussi bien étudier le minimum plutôt que le maximum. De toute évidence, les résultats sur l'un des deux peuvent être immédiatement transférés à l'autre à travers la relation

$$X_{1,n} = -\max\{-X_1, -X_2, \dots, -X_n\}.$$

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes et identiquement distribuées, de fonction de répartition F . Il est naturel de s'intéresser à la distribution asymptotique de $X_{n,n}$. On vérifie aisément que

$$F_{X_{n,n}}(x) = P[X_{n,n} \leq x] = F^n(x),$$

et donc $F_{X_{n,n}}(x)$ converge vers 0 ou 1 selon la valeur de $F(x)$ et donc la loi limite de $X_{n,n}$ est dégénérée!

Par analogie avec le théorème central limite, on s'autorise alors une normalisation linéaire de $X_{n,n}$ et on étudie le comportement limite de la quantité

$$\frac{X_{n,n} - b_n}{a_n},$$

avec a_n et b_n des suites réelles bien choisies pour éviter la dégénérescence de la loi limite. Ainsi, le principal problème mathématique posé dans la théorie des valeurs extrêmes concerne la recherche de distributions des X_i pour lesquelles il existe deux suites de nombres réels $\{a_n > 0\}_{n \geq 1}$ et $\{b_n\}_{n \geq 1}$ telles que pour toutes les valeurs réelles x (où la limite est continue), on a

$$P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) \xrightarrow{n \rightarrow \infty} H(x). \quad (2.17)$$

La normalisation avec b_n et a_n paraît naturelle car sinon $X_{n,n} \rightarrow x_*$ *p.s.* De plus, le problème est double. Il faut trouver toutes les distributions possibles H pouvant apparaître comme limite dans (2.17), et essayer de caractériser les distributions F pour lesquelles il existe des suites $\{a_n\}_{n \geq 1}$ et $\{b_n\}_{n \geq 1}$ telles que la convergence dans (2.17) ait lieu pour de telles distributions limite.

Le problème de la limite a été résolu par Fisher-Tippett (Fisher and Tippett (1928)), Gnedenko (Gnedenko (1943)) et a ensuite été relancé et rationalisé par de Haan (deHaan (1970)).

Théorème 2.3.1. [Théorème de Fisher-Tippet-Gnedenko]

Soit X_1, X_2, \dots, X_n un échantillon aléatoire d'une loi F quelconque et soit $X_{n,n} = \max(X_1, X_2, \dots, X_n)$. S'il existe deux suites réelles $a_n > 0$ et b_n , et une constante $\xi \in \mathbb{R}$ telles que

$$P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) \xrightarrow{n \rightarrow \infty} H_\xi(x), \quad (2.18)$$

pour tout $x \in \mathbb{R}$ avec H_ξ non dégénérée, alors H_ξ appartient à la famille des lois GEV.

Remarque 2.3.2. La démonstration du théorème 2.3.1 est assez technique et laborieuse. Cependant, nous allons donner quelques indications sur quelques étapes essentielles. On utilise un résultat de la théorie des probabilités liant la convergence en distribution à la convergence en moyenne dû à Helly-Bray (Billingsley (1999)).

Soient $Y_n := a_n^{-1}(X_{n,n} - b_n)$ une suite de variables aléatoires et Y_ξ une autre variable aléatoire. On dit que $Y_n \xrightarrow{D} Y_\xi$ si et seulement si pour toute fonction z à support compact, on a

$$E(z(Y_n)) \xrightarrow{n \rightarrow \infty} \int_{-\infty}^{+\infty} z(u) dH_\xi(u),$$

où $H_\xi(u) := P(Y_\xi \leq u)$.

Ainsi, on peut écrire

$$E \left\{ z \left(a_n^{-1}(X_{n,n} - b_n) \right) \right\} = n \int_{-\infty}^{+\infty} z \left(\frac{x - b_n}{a_n} \right) F^{n-1}(x) dF(x). \quad (2.19)$$

Puisque la fonction de répartition F est croissante, on procède au changement de variable $F(x) = 1 - \frac{v}{n}$ dans l'intégrale (2.19).

En utilisant la fonction quantile de queue $U(y) = F^{-1}(1 - \frac{1}{y})$, et puisque on a le choix des suites a_n et b_n , l'intégrale (2.19) peut être exprimée telle que

$$\int_0^n z \left(\frac{U \left(\frac{n}{v} \right) - U_n}{a_n} \right) \left(1 - \frac{v}{n} \right)^{n-1} dv.$$

La limite de l'intégrale ci-dessus existe si les quantités $z \left(\frac{U \left(\frac{n}{v} \right) - U_n}{a_n} \right)$ et $\left(1 - \frac{v}{n} \right)^{n-1}$ convergent.

Il est évident que la quantité $\left(1 - \frac{v}{n} \right)^{n-1}$ admet une limite. Quant à la quantité $z \left(\frac{U \left(\frac{n}{v} \right) - U_n}{a_n} \right)$, elle converge si $\frac{U \left(\frac{n}{v} \right) - U_n}{a_n}$ converge. Cette dernière est obtenue à l'aide des propriétés des fonctions régulières sous la forme $h_\xi(u) = \frac{u^\xi - 1}{\xi}$.

Il suffit finalement d'écrire l'intégrale $\int_0^\infty z \left(h_\xi \left(\frac{1}{v} \right) \right) e^{-v} dv$ sous la forme $\int_{-\infty}^{+\infty} z(u) dH_\xi(u)$. Et on déduit facilement les différentes expressions de $H_\xi(u)$ pour $\xi > 0$, $\xi < 0$ et $\xi = 0$.

L'importance du paramètre ξ pour les distributions limites H_ξ est résumée visuellement dans la figure 2.9 ci-dessous :

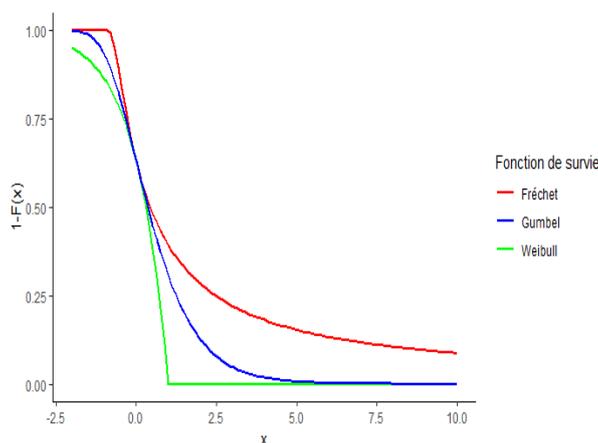


FIGURE 2.9 – Fonctions de survie des distributions GEV pour différentes valeurs de ξ

On constate d'après le graphe 2.9 que la distribution de Fréchet décroît moins vite que celles de Weibull et de Gumbel. C'est pour cela qu'on dit que la distribution de Fréchet est à queue lourde.

2.4 Caractérisation des domaines d'attraction

Un problème important est celui de la caractérisation des domaines d'attraction. Un domaine d'attraction est l'ensemble des distributions pour lesquelles la distribution limite du maximum converge vers la même limite GEV . La recherche de ce genre de domaines consiste en la détermination de conditions nécessaires et suffisantes d'appartenance à ce type d'ensemble. Différentes caractérisations des trois domaines d'attraction de Fréchet, Gumbel et de Weibull ont été proposées dans (Resnick (1987), Embrechts et al. (2013), De Haan and Ferreira (2007)). Ces caractérisations font appel aux classes de fonctions à variation régulière chapitre 2.2.4.

2.4.1 Domaine de Fréchet

Nous noterons dans toute la suite $Q = 1 - F = \overline{F}$ la fonction de survie de F . Souvent, on utilise l'inverse généralisée de F définie par $Q(p) = F^{\leftarrow}(p) = \inf x \in \mathbb{R}, F(x) \geq p, 0 < p \leq 1$. On définit également la fonction de quantile de queue par $U(x) = Q(1 - \frac{1}{x}), x > 1$.

La caractérisation du domaine de Fréchet a été énoncé par Gnedenko (Gnedenko (1943)) et démontré dans (Resnick (1987), Proposition 1.13).

Proposition 2.4.1. *La fonction de distribution F ayant pour point terminal x_* appartient au domaine d'attraction de Fréchet $D(H_\xi)$, $\xi > 0$ si et seulement si $x_* = +\infty$ et F est à variation régulière d'indice $\frac{-1}{\xi}$ à l'infini.*

$$\lim_{t \rightarrow \infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = x^{\frac{-1}{\xi}}.$$

Les suites de normalisation a_n et b_n sont données dans ce cas par :

$$a_n = F^{\leftarrow}\left(1 - \frac{1}{n}\right) = U(n) \quad \text{et} \quad b_n = 0, \forall n > 0. \quad (2.20)$$

Remarque 2.4.1. *On déduit de la proposition 2.2.1 que F appartient au domaine $D(H_\xi)$, $\xi > 0$ si et seulement si $x_* = \infty$ et $\overline{F}(x) = x^{\frac{-1}{\xi}} \ell_F(x)$ où ℓ_F est une fonction à variation lente à l'infini.*

2.4.2 Domaine de Weibull

Le résultat énoncé par (Gnedenko (1943)) et démontré par (Resnick (1987), Proposition 1.13), montre que l'on passe du domaine d'attraction de Fréchet à celui de Weibull par un simple changement de variable dans la fonction de distribution.

Proposition 2.4.2. *Une fonction de distribution F de point terminal x_* appartient au domaine d'attraction de Weibull $D(H_\xi)$, $\xi < 0$ si et seulement si x_* est fini et la fonction de répartition F^* définie par*

$$F^*(x) = \begin{cases} 0 & \text{si } x < 0, \\ F\left(x_* - \frac{1}{x}\right) & \text{si } x \geq 0. \end{cases}$$

appartient au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes $-\xi > 0$, c'est-à-dire que F^ est une fonction à variation régulière d'indice $\frac{1}{\xi}$ à l'infini et on note ($F^* \in \mathcal{R}_{\frac{1}{\xi}}$). Dans ce cas un choix possible pour les suites a_n et b_n est :*

$$a_n = x_* - F^{\leftarrow}\left(1 - \frac{1}{n}\right) \quad \text{et} \quad b_n = x_*.$$

Pour la démonstration de la proposition 2.4.2, le lecteur pourra se référer à (Resnick (1987), Proposition 1.13) ou à (Embrechts et al. (2013), Théorème 3.3.12).

Remarque 2.4.2. *On déduit de proposition 2.4.2 que F appartient au domaine de Weibull $D(H_\xi)$, $\xi < 0$ si et seulement si $x_* = \infty$ et $\bar{F}(x) = (x_* - x)^{\frac{-1}{\xi}} \ell((x_* - x)^{-1})$ où $\ell \in \mathcal{R}_0$. De manière équivalente, la fonction quantile $Q(1 - s)$ s'écrit :*

$$Q(1 - s) = x_* - s^{-\xi} \ell\left(\frac{1}{s}\right), \ell \in \mathcal{R}_0.$$

2.4.3 Domaine de Gumbel

La caractérisation des fonctions de répartition du domaine de Gumbel est plus complexe. Le résultat ci-dessous qui est démontré notamment dans (Resnick (1987), Proposition 1.4). Il permet alors de caractériser les fonctions $F \in D(H_\xi)$, avec $\xi = 0$.

Définition 2.4.1. *(Fonction de Von Mises)*

Une distribution F^ avec le point terminal x_* s'appelle une fonction de Von Mises si : $\exists z_0 < x_*$ tel que pour $z_0 < x < x_*$ et $c > 0$*

$$1 - F^* = c \exp \left\{ - \int_{z_0}^x \left(\frac{1}{f(u)} du \right) \right\}, \quad (2.21)$$

où $f(u) > 0$, $z_0 < u < x_$ et f est absolument continu sur (z_0, x_*) avec une densité $f'(u)$ et $\lim_{u \rightarrow x_*} f'(u) = 0$. On appelle f une fonction auxiliaire.*

Proposition 2.4.3. *La fonction de distribution F de point terminal x_* appartient au domaine d'attraction de Gumbel $D(H_\xi)$, $\xi = 0$ si et seulement s'il existe un fonction de Von Mises F^* telle que pour $x \in (z_0, x_*)$*

$$1 - F(x) = c(x)(1 - F^*(x)) = c(x) \exp \left(- \int_{z_0}^x \frac{1}{f(u)} du \right) \quad (2.22)$$

et $c(x) \rightarrow c > 0$.

2.5 Distribution conditionnelle des excès

On va introduire dans la suite, la notion d'un excès au dessus d'un seuil u . Ce dernier est suffisamment élevé pour que ces excès soient considérés comme des extrêmes.

Définition 2.5.1. On appelle excès de la variable aléatoire X au dessus d'un seuil $u \leq x_*$ où x_* est le point terminal de la distribution i.e $x_* = \sup\{x \mid F(x) \leq 1\}$ la variable aléatoire Y , qui prend ses valeurs dans $]0, x_* - u[$, définie par :

$$Y = \{X - u \mid X > u\}, \quad u \leq x_*.$$

Définition 2.5.2. On appelle distribution des excès de la variable aléatoire X au dessus d'un seuil $u \leq x_*$ la distribution de probabilité de la variable aléatoire $Y = \{X - u \mid X > u\}$, donnée par sa fonction de répartition F_u suivante :

$$\begin{aligned} F_u(y) &= P(X - u \leq y \mid X > u) = P(X \leq y + u \mid X > u) \\ &= \frac{P(\{X \leq y + u\} \cap \{X > u\})}{P(X > u)} = \frac{P(u < X \leq y + u)}{P(X > u)} \\ &= \frac{F(u + y) - F(u)}{1 - F(u)}. \end{aligned}$$

Alors

$$\forall y \in \mathbb{R} \quad F_u(y) = P(X - u \leq y \mid X > u) = \begin{cases} 0 & \text{si } y \leq 0 \\ \frac{F(u+y)-F(u)}{1-F(u)} & \text{si } 0 \leq y \leq x_* - u \\ 1 & \text{si } y \geq x_* - u \end{cases}$$

La fonction de survie de Y au dessus d'un seuil u est définie par :

$$P(Y > y) = \bar{F}_u(y) = 1 - F_u(y) = \frac{1 - F(y + u)}{1 - F(u)}.$$

2.5.1 Distribution de Paréto généralisée

La distribution de Paréto généralisée joue un rôle essentiel dans la modélisation des excès.

Définition 2.5.3. Soit X une variable aléatoire, on dit que X est une distribution généralisée de Paréto ou bien GPD, si sa fonction de répartition $G_{\xi,\sigma}$ est donnée par :

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - (1 + \xi \frac{y}{\sigma})^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ 1 - \exp(-\frac{y}{\sigma}) & \text{si } \xi = 0, \end{cases} \quad (2.23)$$

cette distribution est définie pour $1 + \xi \frac{y}{\sigma} > 0$

Remarque 2.5.1.

- L'ensemble de définition d'une GPD est :

$$\mathbb{R}_+ \text{ si } \xi \geq 0 \quad \text{ou} \quad [0, -\frac{\sigma}{\xi}] \text{ si } \xi < 0.$$

- La distribution GPD dépend de deux paramètres :
 - $\sigma > 0$ est une paramètre d'échelle.
 - $\xi \in \mathbb{R}$ est un paramètre de forme.

Théorème de Balkema-de Haan-Pickands

La distribution limite d'un excès est donnée par le théorème de Balkema-de Haan-Pickands (Balkema and De Haan (1974), Pickands III et al. (1975)). Soit X une variable aléatoire de fonction de distribution inconnue F . Soit x_* la limite supérieure finie ou infinie du support de la distribution F et soit $Y = X - u$ un excès au dessus d'un seuil élevé u . La fonction de distribution de l'excès Y est définie telle que :

$$F_u(x) = P\{X - u \leq x | X > u\} = P\{Y \leq x | X > u\} = \frac{F(x + u) - F(u)}{1 - F(u)}$$

où $0 \leq x < x_* - u$.

Théorème 2.5.1. Soit F une distribution appartenant au domaine d'attraction de Fréchet. Alors, il existe une fonction $\sigma(u)$ strictement positive et un réel ξ tels que

$$\lim_{u \rightarrow x_*} \sup_{0 \leq x < x_* - u} |F_u(x) - G_{\xi,\sigma(u)}(x)| = 0$$

où $\sigma(u)$ une fonction positive mesurable, et si seulement si $F \in D(H_\xi)$.

Démonstration. En effet, d'après le théorème de Fisher-Tippet, il est naturel de supposer que pour $u \rightarrow x_*$ on a :

$$F^n(x) \approx \exp \left(- \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right) \quad (2.24)$$

De plus, comme on s'intéresse aux valeurs de la queue de la distribution, alors $x > u$ pour u suffisamment grand implique $F(x) \approx 1$ et donc

$$\log F(x) \approx - (1 - F(x)).$$

En passant au logarithme dans l'équation (2.24), on aura

$$n \log F(x) \simeq - \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{-\frac{1}{\xi}},$$

ou bien

$$\log F(x) \simeq -\frac{1}{n} \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{-\frac{1}{\xi}}.$$

D'où

$$1 - F(x) \simeq \frac{1}{n} \left(1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right)^{-\frac{1}{\xi}}. \quad (2.25)$$

Par ailleurs, la fonction de survie des excès $Y = (X - u | X > u)$ est telle que

$$P(X - u > y | X > u) = \frac{1 - F(y + u)}{1 - F(u)}. \quad (2.26)$$

Par conséquent en utilisant la relation (2.25) dans la relation(2.26) on aura

$$\begin{aligned} P(X - u > y | X > u) &\approx \frac{\frac{1}{n} \left(1 + \xi \left(\frac{y+u-\mu}{\sigma} \right) \right)^{-\frac{1}{\xi}}}{\frac{1}{n} \left(1 + \xi \left(\frac{u-\mu}{\sigma} \right) \right)^{-\frac{1}{\xi}}} \\ &= \left(\frac{1 + \xi \left(\frac{u-\mu}{\sigma} \right) + \xi \frac{y}{\sigma}}{1 + \xi \left(\frac{u-\mu}{\sigma} \right)} \right)^{-\frac{1}{\xi}} \\ &= \left(1 + \xi \frac{y}{\sigma} \right)^{-\frac{1}{\xi}} \end{aligned}$$

où $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

Finalement, la fonction de survie des dépassements au dessus d'un seuil u est telle que :

$$P(X > u + y | X > u) = \frac{1 - F(u + y)}{1 - F(u)} \approx \left[1 + \frac{\xi y}{\tilde{\sigma}} \right]^{-\frac{1}{\xi}}$$

où $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

La distribution *GPD* regroupe trois distributions selon les valeurs du paramètre de forme.

Lorsque $\xi = 0$, c'est la distribution exponentielle :

$$G_{0, \tilde{\sigma}}(x) = 1 - \exp \left[-\frac{x}{\tilde{\sigma}} \right]$$

Lorsque $\xi > 0$, c'est la loi de Pareto usuelle :

$$G_{\xi, \tilde{\sigma}}(x) = 1 - \left[\frac{\tilde{\sigma}}{\xi x} \right]^{-\frac{1}{\xi}}$$

Lorsque $\xi < 0$, c'est la loi est de pareto de type *II* :

$$G_{\xi, \tilde{\sigma}}(x) = \left[1 + \frac{x}{\tilde{\sigma}} \right]^{-\frac{1}{\xi}}$$

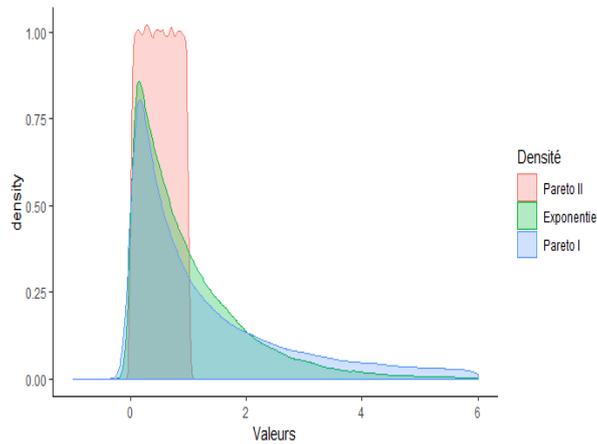


FIGURE 2.10 – Densités de probabilité de *GPD* avec différentes valeurs de ξ

On constate d'après le graphe 2.10 que les fonctions de densité sont bornées à gauche et ils sont décalés à droite de la médiane. Aussi que seulement la densité de Pareto usuelle est bornée à droite.

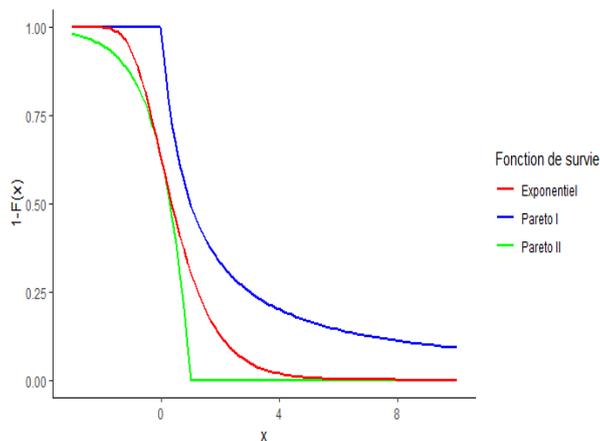


FIGURE 2.11 – Fonction de survie de *GPD* avec différentes valeurs de ξ

On constate d'après le graphe 2.11 les fonctions de survie que la distribution de Pareto I décroît moins vite que les distributions exponentielle et Pareto II. Ainsi, on dit que la distribution de Pareto I est à queue lourde.

Chapitre 3

Estimation dans les modèles GPD

Dans ce chapitre nous nous intéressons aux différentes méthodes d'estimation des paramètres σ, ξ d'un modèle statistique gouverné par une distribution $GPD_{\sigma, \xi}$. On trouve dans la littérature plusieurs méthodes d'estimation. Parmi ces méthodes, on peut citer la méthode du maximum de vraisemblance (Reiss and Thomas (2007)), la méthode des moments pondérés (Hosking and Wallis (1987)), l'estimateur de Hill (Hill (1975)), l'estimateur de Pickands (Pickands III et al. (1975)), et d'autres techniques d'estimation telles que l'estimateur de Peng (Peng (1998)), l'estimateur des moments Dekkers et al. (Dekkers et al. (1989a)). Toutes ces méthodes partagent les mêmes propriétés de consistance et de normalité asymptotique. Cependant, les simulations montrent qu'il y a de grandes différences entre ces estimateurs. Les méthodes les plus utilisées sont celles du maximum de vraisemblance, de Hill, de Pickands et des moments pondérés.

3.1 Méthode des excès au dessus d'un seuil

Pour développer les techniques d'estimation des paramètres dans un modèle GPD , on doit disposer d'un échantillon de valeurs extrêmes. Pour cela une technique de construction d'un tel échantillon existe. On considère un seuil u assez élevé ($u \rightarrow x_*$), alors la suite des excès au dessus du seuil u forme un échantillon d'extrêmes. La distribution limite de tels excès est donnée par le théorème de Balkema-de Hann-Pickands. Cette procédure est appelée méthode *POT* (Peack Over Threshold). Cette technique est illustrée à la figure 3.1

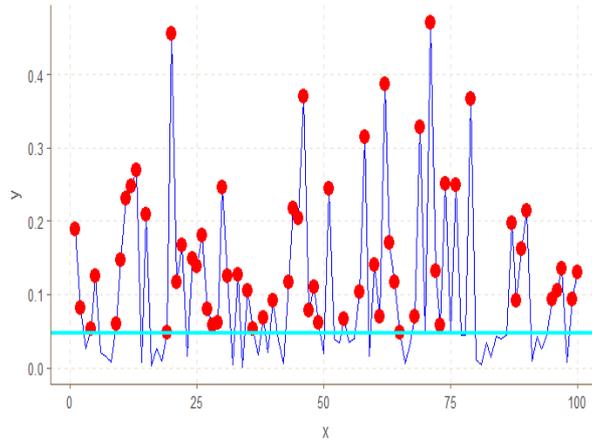


FIGURE 3.1 – Méthode des excès au dessus d'un seuil

3.1.1 Sélection du seuil

Avant de pouvoir estimer les paramètres du modèle, il faut construire un échantillon de maxima adéquat. Il nous faut trouver un seuil u de sélection des données extrêmes suffisamment élevé pour que l'approximation asymptotique (Balkema and De Haan (1974), Pickands III et al. (1975)) de la section 2.5 soit applicable (si nous choisissons un seuil trop bas, les estimations seront biaisées) et audessus duquel nous conservons assez de données pour des estimations précises (si le seuil est trop élevé, les écarts-types des estimateurs seront trop importants). La sélection du seuil se fait souvent à l'aide de la représentation graphique du "Mean-Excess plot".

Mean-Excess plot

Etant donnée un échantillon aléatoire X_1, X_2, \dots, X_n d'une distribution F et soit u un seuil donné. Le Mean Excess Plot est le graphe des points $(u, e(u))$ où $e(u)$ est la moyenne des excès au dessus du seuil u , appelée aussi la Mean Excess Function définie par :

$$e(u) = E(X - u | X > u).$$

Elle correspond à une fonction de u que l'on peut exprimer à l'aide de $1 - F$. On suppose que pour le modèle proposé, $E(X) < \infty$.

En pratique, la fonction des excès moyens $e(u)$ est estimée par $\hat{e}_n(u)$:

$$\hat{e}_n(u) = \frac{\sum_{i=1}^n (x_i - u) \mathbf{1}_{(x_i > u)}(x_i)}{\sum_{i=1}^n \mathbf{1}_{(x_i > u)}(x_i)}.$$

Un des principaux outils du choix du seuil est le graphe de la *sample mean excess function* $\hat{e}_n(u)$ ou le mean excess plot.

Définition 3.1.1. *Le mean excess plot ou ME-plot est défini par le nuage de points suivant :*

$$\{(u, \hat{e}_n(u)), x_{1:n} < u < x_{n:n}\}$$

où $x_{n:n}$ et $x_{1:n}$ sont respectivement le maximum et le minimum de l'échantillon.

Comment interprète-t-on ce graphe ? Il faut remarquer tout d'abord que la mean excess function d'une $GPD(\sigma, \xi)$, pour $\xi < 1$, est

$$E(X - u | X > u) = \frac{\xi}{1 - \xi} u + \frac{\sigma}{1 - \xi} \quad (3.1)$$

Dans ce cas, le seuil à retenir u , est celui pour lequel la moyenne des excès est approximativement linéaire. Plus les queues des distributions sont épaisses, plus la fonction $e(u)$ tend rapidement vers l'infini.

En effet, la fonction moyenne des excès relative à une distribution de loi Pareto de paramètre $\alpha > 0$, appartenant au domaine d'attraction de Fréchet d'indice des valeurs extrêmes $\xi = \frac{1}{\alpha}$, s'écrit comme suit :

$$e(u) = \frac{k + u}{\alpha - 1}$$

Dans la pratique, si la ME-plot semble avoir un comportement linéaire au-dessus d'une certaine valeur de u , cela signifie que les excès au-dessus de ce seuil suivent une GPD . Cette valeur correspond à un arbitrage acceptable entre le souhait d'avoir un seuil élevé et celui d'obtenir un échantillon d'excès de taille suffisante.

3.2 Méthode du maximum de vraisemblance

Dans le cas d'une distribution de Paréto Généralisée $G_{\xi, \sigma}$ (voir l'équation (2.23)), l'estimateur par la méthode de maximum de vraisemblance est basé sur la maximisation de la vraisemblance $L(\sigma, \xi) = \prod_{i=1}^{N_u} g_{\xi, \sigma}(y_i)$. Considérons un échantillon aléatoire des excès Y_1, \dots, Y_{N_u} *i.i.d.* Pour maximiser la log-vraisemblance $\ell(\sigma, \xi) = \log L(\sigma, \xi)$ nous allons procéder de la même manière que (Grimshaw (1993)). Maximiser la log-vraisemblance revient à résoudre les deux systèmes suivants :

$$\begin{cases} \frac{\partial \log L(y_1, y_2, \dots, y_{N_u}; \sigma)}{\partial \sigma} = 0; & \frac{\partial^2 \log L(y_1, y_2, \dots, y_{N_u})}{\partial \sigma^2} < 0 \\ \frac{\partial \log L(y_1, y_2, \dots, y_{N_u}; \sigma)}{\partial \xi} = 0; & \frac{\partial^2 \log L(y_1, y_2, \dots, y_{N_u})}{\partial \xi^2} < 0. \end{cases} \quad (3.2)$$

On distingue deux cas selon les valeurs de ξ :

1. Si $\xi \neq 0$ à condition que $1 + \frac{\xi y}{\sigma} > 0$

$$\begin{aligned} \log L(y_1, y_2, \dots, y_{N_u}; \sigma, \xi) &= \sum_{i=1}^{N_u} \log g_{\xi, \sigma}(y_i) \\ &= \sum_{i=1}^n \left(-\log \sigma + \left(-\frac{1 + \xi}{\xi} \right) \log \left(1 + \xi \frac{y_i}{\sigma} \right) \right) \\ &= -N_u \log \sigma - \left(\frac{1 + \xi}{\xi} \right) \sum_{i=1}^{N_u} \log \left(1 + \xi \frac{y_i}{\sigma} \right). \end{aligned}$$

Résoudre les deux systèmes (3.2) revient à résoudre les deux équation suivants :

$$\begin{aligned} -\frac{N_u}{\sigma} - \left(\frac{1 + \xi}{\xi} \right) \sum_{i=1}^{N_u} \left(\frac{-\xi y_i}{\sigma(\sigma + \xi y_i)} \right) &= 0 \\ \frac{1}{\xi^2} \sum_{i=1}^{N_u} \log \left(1 + \xi \frac{y_i}{\sigma} \right) - \left(1 + \frac{1}{\xi} \right) \sum_{i=1}^{N_u} \left(\frac{y_i}{\sigma + \xi y_i} \right) &= 0. \end{aligned}$$

On a deux systèmes des équations non linéaires n'admettent pas des solutions explicites. On fait appel alors à des méthodes numériques de type algorithmes de Newton-Raphson (Lindstrom and Bates (1988)).

2. Dans le cas ou $\xi = 0$

$$\begin{aligned}\log L(y_1, y_2, \dots, y_{N_u}; \sigma) &= \sum_{i=1}^{N_u} \log g_{\xi, \sigma}(y_i) = \sum_{i=1}^{N_u} \left(-\log \sigma - \frac{y_i}{\sigma} \right) \\ &= -N_u \log \sigma - \frac{1}{\sigma} \sum_{i=1}^{N_u} y_i.\end{aligned}$$

Résoudre les deux systèmes (3.2) revient à résoudre l'équation suivant :

$$-\frac{N_u}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^{N_u} y_i = 0,$$

on obtient :

$$\hat{\sigma} = \frac{1}{N_u} \sum_{i=1}^{N_u} y_i = \bar{Y}$$

3.3 Méthode des moments ordinaires

L'estimateur par la méthode des moments ([Hosking and Wallis \(1987\)](#)) est basé sur le fait que

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

En remplaçant dans les expressions précédentes μ_1 et μ_2 par leurs versions empiriques, nous obtenons l'estimateur des moments $(\hat{\xi}_{k_n}^{MOM}, \hat{\sigma}_{k_n}^{MOM})$ du couple (ξ, σ) définis comme suit

$$\hat{\xi}_{k_n}^{MOM} = \frac{1}{2} \left(1 - \frac{\bar{X}^2}{\bar{X}^2 - \bar{X}^2} \right) \quad \text{et} \quad \hat{\sigma}_{k_n}^{MOM} = \frac{1}{2} \bar{X} \left(\frac{\bar{X}^2}{\bar{X}^2 - \bar{X}^2} + 1 \right),$$

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 $(\hat{\xi}_{k_n}^{MOM}, \hat{\sigma}_{k_n}^{MOM})$ peut être établie sous la condition $\xi < \frac{1}{4}$. Pour affaiblir cette condition d'autres alternatives ont été proposées, telle que la méthode des moments pondérés (éventuellement généralisé)

3.4 Méthode des moments pondérés

Dans cette méthode nous proposons d'égaliser les moments empiriques d'un échantillon aux moments pondérés puisque à partir du deuxième ordre les moments empiriques d'un échantillon sont biaisés. Il peut arriver que certains moments n'existent pas, ne sont pas finis. Dans ce cas on utilise la Méthode des Moments pondérés ([Hosking and Wallis \(1987\)](#)). Définissons, avec r l'ordre du moment

$$w_r(\xi, \sigma) = E[X\bar{G}_{\xi, \sigma}^r(X)],$$

où $\bar{G}_{\xi, \sigma} = 1 - G_{\xi, \sigma}$ avec G la distribution GPD . Alors

$$w_r(\xi, \sigma) = \int_{-\infty}^{+\infty} x\bar{G}_{\xi, \sigma}^r d\bar{G}_{\xi, \sigma}(x) = \int_0^1 \bar{G}_{\xi, \sigma}^{-1}(y)y^r dy = \int_0^1 \frac{\sigma}{\xi}(y^{-\xi} - 1)y^r dy.$$

Nous obtenons grâce à la dernière formulation et après quelque calculs

$$w_r(\xi, \sigma) = \frac{\sigma}{(r+1)(r+1-\xi)},$$

ce qui nous permet d'obtenir avec $r = 0$ et $r = 1$

$$\begin{aligned} \sigma &= \frac{2w_0w_1}{w_0 - 2w_1}, \\ \xi &= 2 - \frac{w_0}{w_0 - 2w_1}. \end{aligned}$$

Pour calculer $\hat{\sigma}$ et $\hat{\xi}$, nous remplaçons le moment pondéré w_r par son estimateur empirique \hat{w}_r défini tel que :

$$\hat{w}_r(\xi, \sigma) = \frac{1}{n} \sum_{i=1}^n X_i \hat{F}^r(X_i), \quad r = 0, 1$$

où \hat{F} est la fonction de répartition empirique de l'échantillon X_1, \dots, X_n .

La normalité asymptotique de $(\hat{\xi}_{k_n}^{PWM}, \hat{\sigma}_{k_n}^{PWM})$ a été établie dans ([Hosking and Wallis \(1987\)](#)). Le domaine de validité nécessaire de cette normalité asymptotique est assez restrictif : $\xi \in (-1, \frac{1}{2})$; néanmoins, cette méthode présente un certain nombre d'avantages tant au niveau de la facilité du calcul que des performances à distance finie ([Hosking and Wallis \(1987\)](#)). Une généralisation de cette méthode permettant d'étendre le domaine de validité $\xi \in (-1, \frac{3}{2})$ a été proposée par ([Diebolt et al. \(2004\)](#) , [Diebolt et al. \(2007\)](#)).

3.5 Estimateur de Hill

Dans le domaine des distributions à queue épaisse, la statistique de la théorie des valeurs extrêmes se traduit par un problème d'estimation semi-paramétrique. En effet, si F appartient au domaine de Fréchet, alors sa fonction de survie est de la forme $x^{-\frac{1}{\xi}}\ell(x)$. Le paramètre d'intérêt est $\xi > 0$. La fonction ℓ fonction à variation lente, est un paramètre de nuisance. Dans le paragraphe suivant, nous présentons quelques estimateurs classiques pour l'indice de valeurs extrêmes ξ .

L'estimateur de Hill n'est utilisable que pour les distributions appartient au domaine de Fréchet. Il est introduit en 1975 par (Hill (1975)), l'estimateur de Hill est probablement l'estimateur le plus étudiée dans la littérature. Une fonction de répartition F appartient au domaine d'attraction de Fréchet $D(H_\xi)$, $\xi > 0$ si et seulement si sa fonction de survie $\bar{F} \in \mathcal{R}_\xi$, la fonction de survie de la distribution de ce domaine est la forme :

$$\bar{F}(x) = 1 - F(x) = x^{-\frac{1}{\xi}}\ell(x).$$

avec ℓ une fonction à variation lente. La fonction de survie \bar{F} comporte :

- Une partie paramétrique $x^{-\frac{1}{\xi}}$ ne dépendant que d'un paramètre réel,
- une partie non-paramétrique ℓ .

Pour un réel $t > 1$ il est clair que

$$\lim_{u \rightarrow +\infty} \frac{\bar{F}(tu)}{\bar{F}(u)} = t^{-\frac{1}{\xi}} \lim_{u \rightarrow +\infty} \frac{\ell(tu)}{\ell(u)} = t^{-\frac{1}{\xi}}, \quad (3.3)$$

on en déduit l'approximation

$$\bar{F}(tu) \simeq \bar{F}(u)t^{-\frac{1}{\xi}} \quad (3.4)$$

Dans l'approximation (3.4) nous remplaçons tu par x on trouve

$$\bar{F}(x) = \bar{F}(u) \left(\frac{x}{u}\right)^{-\frac{1}{\xi}}.$$

La densité $f(x)$ est comme suit

$$f(x) = \frac{1}{\xi} \frac{\bar{F}(u)}{u} \left(\frac{x}{u}\right)^{-\frac{1}{\xi}-1}.$$

L'expression de la fonction de vraisemblance est :

$$L(x_1, x_2, \dots, x_k; \xi) = \prod_{i=1}^k f(x_i)$$

Maximiser la vraisemblance L revient à résoudre le système suivant :

$$\left\{ \begin{array}{l} \frac{\partial \log L(x_1, x_2, \dots, x_k; \xi)}{\partial \xi} = 0; \\ \frac{\partial^2 \log L(x_1, x_2, \dots, x_k)}{\partial \xi^2} < 0, \end{array} \right. \quad (3.5)$$

en effet

$$\begin{aligned} \log L(x_1, x_2, \dots, x_k; \xi) &= \log \left(\prod_{i=1}^k f(x_i) \right) \\ &= \sum_{i=1}^k \left[\log \frac{1}{\xi} + \log \frac{\bar{F}(u)}{u} + \log \left(\frac{x_i}{u} \right)^{-\frac{1}{\xi}-1} \right] \\ &= -k \log \xi + k \log \frac{\bar{F}(u)}{u} + \left(-\frac{1}{\xi} - 1 \right) \sum_{i=1}^k \log \left(\frac{x_i}{u} \right) \end{aligned}$$

Résoudre les deux systèmes (3.5) revient à résoudre l'équation suivant :

$$-\xi k + \sum_{i=1}^k \log \frac{x_i}{u} = 0,$$

la solution de cette équation est de la forme

$$\hat{\xi}^H = \frac{1}{\xi} \sum_{i=1}^k \log \frac{x_i}{u},$$

On remplace x_i et u par leurs équivalents dans la statistique d'ordre on trouve la formule de l'estimateur de Hill

$$\hat{\xi}_{k_n}^H = \frac{1}{\xi} \sum_{i=1}^k [\log X_{n-i+1,n} - \log X_{n-k+1,n}]$$

D'autres méthodes pour la construction de l'estimateur de Hill sont données dans (Beirlant et al. (2006a)) et dans de (De Haan and Ferreira (2007)).

3.5.1 Propriétés asymptotiques de l'estimateur de Hill

Un grand nombre de travaux théoriques ont été consacré à l'étude des propriétés de l'estimateur de Hill. (Mason (1982)) a démontré la consistance faible et (Deheuvels et al. (1988)) ont établi la consistance forte. La normalité asymptotique est due entre autre à (Davis and Resnick (1984), Csörgö and Mason (1985), Haeusler and Teugels (1985) et Smith (1987)). Notons que la consistance (en probabilité ou presque sûrement) de l'estimateur $\hat{\xi}_{k_n}^H$ ne dépend que du comportement de k_n , alors que sa normalité asymptotique nécessite des conditions plus délicates sur la fonction de distribution F et donc sur la fonction de quantile de queue U .

Dans la définition 2.2.5, nous remarquons que les relations (2.10) et (2.11) impliquent que les fonctions $|A^*|$ et $|A|$ sont à variations régulières à l'infini d'indices respectifs $\frac{\rho}{\xi}$ et ρ (i.e $|A^*| \in \mathcal{R}_{\frac{\rho}{\xi}}$ et $|A| \in \mathcal{R}_{\rho}$), pour plus de détails on peut se référer à (Geluk and deHaan (1987)). Le paramètre ρ (appelé paramètre du second ordre) contrôle donc la vitesse de convergence de $\overline{F}(tx)/\overline{F}(t)$ vers $x^{-\frac{1}{\xi}}$ quand $t \rightarrow \infty$ ou de manière équivalente celle de $U(tx)/U(t)$ vers x^{ξ} quand $t \rightarrow \infty$ (et donc la vitesse de convergence de $\ell_U(tx)/\ell_U(t)$ vers 1). Par exemple plus cette convergence est rapide, plus la linéarité dans le « Paréto quantile plot » apparaît vite. En d'autres termes, une valeur de ρ proche de 0 implique une faible vitesse de convergence, et plus l'estimation de l'indice de queue ξ est difficile. Les relations (2.10) et (2.11) dans la définition 2.2.5 (peuvent être reformulées respectivement de la manière suivante :

$$\lim_{t \rightarrow \infty} \frac{(\overline{F}(tx)) / \overline{F}(t) + \xi^{-1} \log x}{A^*(t)} = x^{-\frac{1}{\xi}} \frac{x^{\rho} - 1}{\rho} \quad (3.6)$$

$$\lim_{t \rightarrow \infty} \frac{(U(tx)/U(t)) - \xi \log x}{A(t)} = x^{\xi} \frac{x^{\rho} - 1}{\rho}. \quad (3.7)$$

Ces conditions appelées « conditions du second ordre » sont satisfaites pour la plupart des distributions appartenant au domaine de Fréchet (Cas des distributions à queues lourdes). Par exemple dans le cas des modèles de Hall (ou classe de Hall) ci-dessous, ces conditions ont été un peu plus spécifiées. La classe de Hall est composée de l'ensemble

des distributions F telles que

$$1 - cx^{\frac{-1}{\xi}} \left(1 + dx^{\frac{\rho}{\xi}} + o(x^{\frac{\rho}{\xi}})\right) \quad \text{quand } x \rightarrow \infty \quad (3.8)$$

où $\xi > 0$, $\rho \leq 0$, $c > 0$, et $d \in \mathbb{R}^*$. Cette sous classe des distributions à queues lourdes (de type Paréto) contient les distributions de Paréto, Burr, Fréchet et t-Student fréquemment utilisées en assurance mathématique, comme des modèles pour des grands risques. La relation (3.8) peut être reformulée en utilisant la fonction quantile de queue U de la manière suivante $U(t) = c^\xi t^\xi (1 + \xi d c^{\rho t^\rho} + o(t^\rho))$ quand $t \rightarrow \infty$. Ainsi, pour le cas des modèles de Hall, les fonctions A et $A^*(t)$ sont respectivement équivalentes à $d\rho \xi c^{\rho t^\rho}$ quand $t \rightarrow \infty$.

Théorème 3.5.1. *Soit $k_n, n \geq 1$ une suite d'entiers telle que $1 \leq k_n < n, k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$ quand $n \rightarrow \infty$.*

- (Consistance faible) : Alors, $\widehat{\xi}_{k_n}^H$ converge en probabilité vers ξ .
- (Consistance forte) : Si de plus $k_n / \log \log n \rightarrow \infty$ quand $n \rightarrow \infty$, alors $\widehat{\xi}_{k_n}^H$ converge presque sûrement vers ξ .
- (Normalité asymptotique) : Si la condition (2.11) est satisfaite avec $\sqrt{k_n} A(\frac{n}{k_n}) \rightarrow \xi \in \mathbb{R}$ quand $n \rightarrow \infty$, alors $\sqrt{k_n} (\widehat{\xi}_{k_n}^H - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\xi}{1-\rho}, \xi^2\right)$.

En utilisant la normalité asymptotique de $\widehat{\xi}_{k_n}^H$, (Beirlant et al. (1996b)) trouvent un intervalle de confiance de $\widehat{\xi}_{k_n}^H$ où $\widehat{k}_{n\text{opt}}$ est une valeur optimale de k_n donnée par leur algorithme. Dans leurs travaux, ils mentionnent qu'il est nécessaire d'être prudent sur l'intervalle de confiance du fait que la variabilité du choix de k_n est ignorée et ce qui ne devait pas être le cas. Une méthode plus précise suggérée pour la construction d'un intervalle de confiance est le bootstrap (Danielsson et al. (1996), Gomes and Oliveira (2001)). Des méthodes de sélection du meilleur choix de k_n sont données dans la section 3.6.1.

Remarque 3.5.1. *D'après la normalité asymptotique de Hill, si $\sqrt{k_n} A(\frac{n}{k_n}) \rightarrow 0$, on peut écrire : $\sqrt{k_n} (\widehat{\xi}_{k_n}^H / \xi - 1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$. Ainsi, un intervalle de confiance au niveau $(1 - \alpha)$ pour l'estimation de ξ est donné par :*

$$\left(\widehat{\xi}_{k_n}^H \left(1 + k_n^{-1/2} z_{1-\alpha/2}\right)^{-1}, \widehat{\xi}_{k_n}^H \left(1 - k_n^{-1/2} z_{1-\alpha/2}\right)^{-1} \right),$$

où z_α est le quantile d'une distribution normale standard au niveau α . Par exemple pour une distribution de Paréto standard on a $F(x) = x^{-\frac{1}{\xi}}$, $x \geq 1$ et $\xi > 0$, la fonction à variations lentes associée est $\ell_F(x) = 1$ alors, il n'y a pas de biais asymptotique dans l'estimateur de Hill (la limite $\xi = 0$).

3.6 Estimateur de Pickands

Il est fondé sur le calcul des quantiles. Cet estimateur a été introduit par (Pickands III et al. (1975)) et revisité par (Drees (1995), Drees and Kaufmann (1998)) d'une part, (Dekkers et al. (1989a)) d'autre part, qui en ont établi sa consistance faible ainsi que sa normalité asymptotique. Cet estimateur est construit en utilisant trois statistiques d'ordres et a l'avantage d'être valable quelque soit le domaine d'attraction de la distribution. Il est défini de la manière suivante :

$$\widehat{\xi}_{K_n}^P = \frac{1}{2} \log \frac{X_{n-k_n+1,n} - X_{n-2k_n+1,n}}{X_{n-2k_n+1,n} - X_{n-4k_n+1,n}}, k_n = 1, \dots, [n/4]$$

Notons que l'estimateur de Pickands $\widehat{\xi}_{K_n}^P$, qui utilise l'information apportée par des distances entre deux statistiques ordonnées n'utilise pas le maximum de l'échantillon $X_{n,n}$, ce qui implique une perte d'information sur la queue de distribution. Une justification plus formelle de l'estimateur de Pickands a été fournie dans (Embrechts et al. (2013)) Les propriétés de l'estimateur de Pickands ont été principalement explorées dans (Dekkers et al. (1989b)). Plus précisément, ils ont prouvé les résultats suivants :

Théorème 3.6.1. *On suppose que $F \in D(H_\xi)$, $\xi \in \mathbb{R}$, la condition du second ordre (2.11) soit vérifiée. Soit $k_n, n \geq 1$ une suite d'entiers telle que $1 \leq k_n < [n/4]$, $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$ quand $n \rightarrow \infty$*

- (Consistance faible) : Alors, $\widehat{\xi}_{k_n}^P \xrightarrow{\mathbb{P}} \xi$ quand $n \rightarrow \infty$.
- (Consistance forte) : Si de plus $k_n / \log \log n \rightarrow \infty$ quand $n \rightarrow \infty$, alors $\widehat{\xi}_{k_n}^P \xrightarrow{p.s} \xi$
- (Normalité asymptotique) : On suppose que la fonction de queue U associée à F satisfait la condition du second ordre ci-dessous : ils existent une constante réelle $\rho \leq 0$ et une fonction $A(t) \rightarrow 0$ quand $t \rightarrow \infty$ telles que pour tout $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{A(t)} - D_\xi(x)}{A(t)} = \int_1^x s^{\xi-1} \int_1^s u^{\rho-1} du ds,$$

Alors pour $\sqrt{k_n}A(n/k_n) \rightarrow \xi \in \mathbb{R}$ quand $n \rightarrow \infty$,

où

$$\rho_P^2(\xi) = \frac{\xi^2 (2^{2\xi+1})}{4 (\log 2)^2 (2^\xi - 1)^2}$$

et

$$b_{\rho,\xi} = \begin{cases} \frac{4^{-\rho}\xi(4^{\xi+\rho}-1)-(2^\rho+1)(2^{\xi+\rho})-1}{\rho^{2\xi}(\xi+\rho)(2^{\rho+P}-1)} & \text{si } \rho < 0, \\ 1 & \text{si } \rho = 0. \end{cases}$$

Pour plus de détails on peut se référer dans (De Haan and Ferreira (2007)). Une justification plus formelle sur l'estimateur de Pickands est fournie dans (Embrechts et al. (2013)).

3.6.1 Choix du nombre optimal de statistique d'ordre extrêmes

Les résultats concernant les estimateurs de l'indice des valeurs extrêmes sont asymptotiques : ils sont obtenus lorsque $k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$. Choisir un nombre optimal de statistiques d'ordre extrêmes k_{opt} n'est pas facile à manipuler. Lorsque k est petit la variance de l'estimateur est grande et l'utilisation de k introduit un grand biais. L'équilibrage de ces composants (la variance et le biais) est important dans les applications de la théorie des valeurs extrêmes, parce que cela réduit l'erreur moyenne quadratique. Ce problème a été longuement abordé dans la littérature, voir par exemple (Balkema and De Haan (1974), Hall and Welsh (1985), De Haan and Peng (1998), Cheng and Peng (2001), Beirlant et al. (2006a)).

Méthode de Hall et Welsh

Hall et Welsh (Hall and Welsh (1985)) ont prouvé que l'erreur moyenne quadratique asymptotique de l'estimateur de Hill est minimale pour

$$k_{opt} \approx \left(\frac{c^{2\rho(\rho+1)^2}}{2d^2\rho^3} \right)^{\frac{1}{2\rho+1}} n^{\frac{2\rho}{2\rho+1}} \quad (3.9)$$

Si la fonction de répartition F satisfaisant la classe de Hall (3.9). Comme les paramètres sont inconnus, ce résultat ne peut pas être utilisé directement pour déterminer k_{opt} . L'erreur moyenne quadratique MSE est liée par le paramètre du second ordre ρ donc on ne

peut pas être utilisé dans les questions pratiques. Pour cette raison, nous présentons la méthode bootstrap.

Méthode Bootstrap

En 1990 (Hall (1990)) a utilisé la méthode du bootstrap dans l'estimation de l'indice de queue pour le nombre de statistiques d'ordre extrêmes. Son fondement théorique est minimisé MSE de l'IVE. Hall a précisé la taille n_1 de l'échantillon bootstrap qui doit être inférieure à la taille de l'échantillon total n . Et il avait prouvé que l'estimation n'était pas fiable quand $n_1 = n$. Nous avons besoin de retirer un sous-échantillon $\mathbf{X}_{n_1}^* = \{X_1, \dots, X_{n_1}\}$ de l'échantillon total $\mathbf{X}_n = \{X_1, \dots, X_n\}$ ($n_1 \ll n$) et $\mathbf{X}_{n_1}^*$ est appelé sous-échantillon bootstrap. Nous allons utiliser $X_{1,n_1} \leq X_{2,n_1} \leq \dots \leq X_{n_1,n_1}$ les statistiques d'ordre de l'échantillon $\mathbf{X}_{n_1}^*$ pour définir $\xi_{n_1}^*(k_1)$ tel que

$$\xi_{n_1}^*(k_1) := \frac{1}{k_1} \sum_{i=1}^{k_1} \log X_{n_1-i+1,n_1}^* - \log X_{n_1-k_1,n_1}^*. \quad (3.10)$$

On détermine k et k_1 en minimisant

$$MSE(n_1, k_1) = E \left(\left(\xi_{n_1}^*(k_1) - \xi_n(k) \right)^2 \mathbf{X}_n \right)$$

Et la relation est $k = k_1 \left(\frac{n}{n_1} \right)^\mu$. En fait, Hall avait supposé la relation de pouvoir de k et n était $k = cn^\mu$, ($0 < \mu < 1$) et le rapport de pouvoir n_1 et n était $n_1 = n^\beta$. Il a suggéré que μ et β étaient $1/2$. Cette méthode dépend de ξ et ρ et MSE est liée à k .

Méthode graphique

C'est la méthode la plus simple pour la détermination de k_n . Elle consiste à tracer le graphe

$$\left\{ (k_n, \hat{\xi}_{k_n}) : 1 \leq k_n < n \right\},$$

dans le but de trouver une valeur optimale de k_n . Ainsi le k_n optimal (notée k_{opt}) est choisi dans la première région où l'estimateur $\hat{\xi}_{k_n}$ devient stable c'est à dire le graphe devient horizontal. Cet estimateur est valable seulement dans le domaine d'attraction de Fréchet c'est à dire si $\xi > 0$. Pour généraliser aux autres domaines d'attractions, différents estimateurs ont été proposés, entre autres l'estimateur de Dekkers-Einmahl-de Hann et l'estimateur de Pickands. Nous illustrons cette méthode sur les données d'assurance de

montants de sinistres incendies (en millions de couronnes Danoises), survenus au Danemark (2492 données). La Figure 1.6.1 suggère une valeur estimée de 0,7 pour l'indice de queue sur ces données, et un nombre optimal d'extrêmes utilisé dans l'estimation de l'ordre de 300. 0 500

Chapitre 4

Calcul de risque

4.1 Evaluation du risque

Le risque est une notion difficile à cerner. Mais de façon générale, on peut dire que c'est une contingence indésirable, relativement anodine et peu probable. Le risque est généralement anodin, mais tout de même suffisamment nuisible pour être indésirable. En ce sens il se distingue notamment du danger qui suppose la possibilité d'un dommage grave voir létal. On dira par exemple de quelqu'un qui sort tête nue par temps froid qu'il court le risque d'attraper un rhume, par contre on dira qu'il se met en danger s'il traverse une autoroute.

Un risque est une contingence peu probable, ce qui constitue une autre différence par rapport au danger. On parle en effet de danger lorsque la probabilité d'occurrence et les conséquences sont importantes, tandis que le risque existe dès lors que sa probabilité d'occurrence n'est pas nulle. L'appréciation de ces différents critères est hautement subjective, ce qui peut justifier que dans les domaines scientifiques et techniques on recherche une définition quantifiable et rigoureuse du risque.

Toute situation, toute activité peut produire un événement profitable ou dommageable. Le risque est défini par la probabilité de survenue de cet événement et par l'ampleur de ses conséquences. Dans certains domaines, on ne prend en compte que les conséquences négatives, que les pertes et pas les gains ; on parle alors de risque aryétique.

Le risque est une notion importante notamment dans les domaines de la finance, des assurances, de l'environnement (risques industriels, risques majeurs), de l'industrie, du droit, de la santé, et bien sûr de la gestion, etc.

4.2 Mesure synthétique du risque extrême

Par exemple, le risque de marché en finance correspond à un risque de perte des portefeuilles des investisseurs dues aux variations des marchés financiers : actions, obligations, devises, matières premières, mais aussi marchés des produits dérivés (contrats à terme, options). Pour gérer le risque de marché, il faut donc mesurer de manière précise ce risque extrême. Néanmoins, le risk manager a besoin d'une mesure plus synthétique du risque encouru par son portefeuille d'investissement. On peut ainsi appréhender le risque à partir des profils de la dispersion des pertes et profits des actifs. Nous pouvons faire ressortir deux mesures de risque à partir de la distribution des rentabilités des actifs : la volatilité et la Value-at-Risk. Ces mesures ne captent pas la même chose.

En effet, la volatilité mesure la dispersion moyenne des rentabilités autour de leur moyenne.

Son expression en fonction du vecteur des rentabilités X est $\sigma = \sqrt{E[X - E(X)]^2}$. Cette expression du risque prend donc en compte toutes les rentabilités, positives ou non, extrêmes ou non.

La Value-at-Risk (ou VaR) est le quantile de perte déterminant la plus grande perte que peut subir un portefeuille avec une probabilité d'occurrence faible sur un horizon donné ; elle mesure ainsi le risque extrême.

La logique entre les deux mesures est donc différente. En effet, la volatilité peut être élevée et seulement capturer des risques moyens, certes importants, mais pas extrêmes. Tout l'enjeu d'une mesure du risque synthétique pertinente est d'estimer correctement la perte potentielle que peut subir un investissement. Par exemple, un actif peut avoir une volatilité de 20% mais ne pas connaître de pertes très fortes, avec un maximum par exemple de 15% de perte sur une journée. Au contraire, un actif avec une volatilité de 15% peut connaître des pertes, certes rares, de plus de 30%. Le second actif nous semble donc plus risqué, même si ses risques moyens s'avèrent être moins importants que ceux du premier.

En fait, déterminer le risque par le moment d'ordre 2 la volatilité présuppose que les moments suivants skewness et kurtosis ne nécessitent pas d'être incorporés dans une mesure de risque. Le paradigme sous-jacent est la normalité des rentabilités d'un actif financier (la loi Normale étant caractérisée par les deux premiers moments).

Utiliser la VaR permet de résoudre ce problème car un quantile de distribution n'est pas une mesure moyenne de la dispersion mais prend seulement en compte les pertes qui nous

intéressent ici, c'est-à-dire les pertes extrêmes. Ainsi, la volatilité n'est certainement pas la meilleure mesure de risque extrême. C'est pour cela que la VaR est utilisée par la plupart des 'Risk managers'. Nous allons maintenant définir et formaliser la *VaR* de manière plus précise.

4.2.1 Définition mathématique de la *VaR*

La *VaR* représente une perte potentielle dont il faut définir la probabilité α d'occurrence, appelée aussi niveau de confiance, ainsi que l'horizon h de durée d'investissement. Par exemple, une probabilité de 1% pour un horizon h de *un jour* revient à accepter que la perte potentielle ne dépasse la VaR que 2 à 3 fois par an (1 an = 251 jours ouvrés). L'horizon représente la durée sur laquelle nous estimons le risque, c'est-à-dire le temps nécessaire pour couvrir la position du portefeuille : 1 jour en trading, 1 mois en gestion, etc.

Soit Z_t la valeur en t d'un portefeuille et Z_{t+h} sa valeur à l'horizon h . La rentabilité $X_{t/t+h}$ du portefeuille investi à la date t et détenu sur un horizon de durée h a pour expression, $\forall t \geq 1$:

$$X_{t/t+h} = \ln \frac{Z_{t+h}}{Z_t} \approx \frac{Z_{t+h}}{Z_t} - 1$$

Supposons que $X_{t/t+h}$ est issu d'une variable aléatoire notée X . La $VaR(h, \alpha)$ est alors définie par :

$$P[X \leq VaR(h, \alpha)] = 1 - \alpha$$

c'est-à-dire la probabilité que la rentabilité du portefeuille soit inférieure à la *VaR* sur un horizon h est égale à $(1 - \alpha)$. Les praticiens parlent aussi du niveau de confiance $(1 - \alpha)$ qui correspond au quantile de probabilité α . Ainsi, une VaR à 99% par exemple correspond au quantile de probabilité $\alpha = 1 - 99\% = 1\%$.

Définition 4.2.1. Notons F la fonction de répartition de la variable aléatoire X^P des rentabilités du portefeuille de périodicité h . La *Value-at-Risk* de probabilité d'occurrence α et d'horizon d'investissement h est alors le quantile défini par :

$$VaR(h, \alpha) = F^{-1}(1 - \alpha) \tag{4.1}$$

Pour déterminer explicitement la *VaR*, il nous reste à caractériser la forme de F . La première solution consiste à estimer F par sa distribution empirique : nous avons alors la

VaR historique. Une seconde solution est obtenue en utilisant un modèle paramétrique : les praticiens utilisent classiquement la *VaR Gaussienne*. Mais d'autres paramétrisations (*Student, TVE*) se révéleront plus robustes.

4.2.2 *VaR* historique et *VaR* bootstrap

Pour la *VaR* historique, nous estimons F dans l'équation (4.1) par sa distribution empirique \hat{F} par une méthode non-paramétrique. Supposons que nous disposons d'un historique de N observations des rentabilités. La distribution empirique \hat{F} des rentabilités du portefeuille est alors caractérisée à partir du vecteur $(\hat{X}_{1|1+h}, \dots, \hat{X}_{N|N+h})'$ et $\hat{F}^{-1}(\alpha)$ correspond au quantile de probabilité α de $(\hat{X}_{1|1+h}, \dots, \hat{X}_{N|N+h})'$.

L'avantage de cette méthode est donc de ne pas imposer d'hypothèse sur la loi de distribution des rentabilités. Mais le problème qui se pose est la longueur N de l'historique qui, si elle est trop faible, ne fournit pas un calcul précis de la VaR. Dans ce cas on peut utiliser la technique du ré-échantillonnage ou bootstrap (Efron and Tibshirani (1994)). Nous construisons pour cela une version aléatoire de même taille du vecteur $(\hat{X}_{1|1+h}, \dots, \hat{X}_{N|N+h})'$ dans laquelle les rentabilités du portefeuille peuvent apparaître pour certaines aucune fois, pour d'autres une fois, pour d'autres encore deux fois, etc. Ce nouveau vecteur est une version rééchantillonnée ou bootstrap du vecteur d'origine. Nous pouvons alors déterminer la VaR historique de cet échantillon. Nous réalisons M répliques *bootstrap* et la VaR historique *bootstrap* correspond à la moyenne des M VaR historiques.

L'inconvénient de ces VaR non paramétriques est l'*over-fitting*, soit la dépendance trop forte à l'échantillon sur lequel elle a été déterminée, et donc la difficulté de cette VaR à être robuste dans le temps. C'est pour cela que les VaR paramétriques, certes moins ajustées aux données étudiées, permettent une meilleure prévision.

4.2.3 *VaR* Gaussienne

Historiquement, le choix du modèle paramétrique se porte sur la loi Normale. Supposons donc que les rentabilités du portefeuille $X_{t|t+h}$ sont issues d'une variable aléatoire Gaussienne. Alors, soit F la fonction de répartition d'une loi $\mathcal{N}(\mu, \sigma^2)$ avec μ comme moyenne des rentabilités et σ^2 comme la variance des rentabilités. Nous obtenons ainsi à

partir de l'équation (4.1) :

$$\Phi\left(\frac{VaR(h, \alpha) - \mu}{\sigma}\right) = 1 - \alpha$$

où Φ est la fonction de répartition de la loi Normale centrée réduite.

Définition 4.2.2. *La VaR Gaussienne est définie de la manière suivante :*

$$VaR(h, \alpha) = \mu + \sigma\Phi^{-1}(1 - \alpha)$$

Cette méthode repose sur des hypothèses irréalistes pour modéliser les rentabilités des actifs financiers puisque trois faits caractérisent leur distribution et ne sont pas captés par la modélisation Gaussienne :

1. l'asymétrie : l'occurrence de pertes extrêmes est plus forte que celle de profits extrêmes,
2. les queues épaisses,
3. le *clustering* : par exemple, les rentabilités élevées en valeur absolue sont suivies par des rentabilités élevées en valeur absolue.

Dans le calcul de risque, l'hypothèse de normalité est à bannir car seules les valeurs extrêmes sont prises en considération et on sait que leurs distributions sont de type *GEV* ou *GPD*. Mesurer le risque revient à calculer la *VaR* donc un quantile extrême.

4.3 Estimation de quantiles extrêmes

Définition 4.3.1. *Les quantiles extrêmes d'ordre $\alpha = \alpha_n \in (0, 1)$ et $\alpha_n \rightarrow 0$ quand $n \rightarrow \infty$ est définie par*

$$VaR(\alpha_n) = F^{\leftarrow}(1 - \alpha_n) = Q(1 - \alpha_n) = U\left(\frac{1}{\alpha}\right).$$

Dans l'analyse des extrêmes, on s'intéresse souvent à l'estimation des quantiles extrêmes qui jouent un rôle important dans le contexte de la gestion des risques, où il est crucial d'évaluer de manière adéquate le risque d'une grande perte, de sorte que les probabilités de les dépasser soient très faibles c'est-à-dire proches de zéro. Nous nous sommes

donc intéressés à l'estimation des quantiles extrêmes $VaR(\alpha_n)$. Plus généralement si la taille de l'échantillon tend vers l'infini, nous avons

$$\begin{aligned}
P(X_{n,n} < VaR(\alpha_n)) &= F^n(VaR(\alpha_n)) \\
&= (1 - \alpha_n)^n \\
&= \exp(n \log(1 - \alpha_n)) \\
&= \exp(-n\alpha_n(1 + o(1))) \quad \text{quand } \alpha_n \rightarrow 0
\end{aligned}$$

Dans la suite nous allons donner quelques méthodes d'estimation de quantiles extrêmes basées sur des estimateur de l'indice de queue.

4.3.1 Approche basée sur l'estimateur du maximum de vraisemblance

Cette méthode s'appuie sur le fait que la distribution des excès au dessus d'un seuil élevé u peut être approchée par une loi *GPD* généralisée : $F_u(x) \simeq G_{\xi, \sigma}(x)$. Soit N_u le nombre d'observations $X_{j_1}, X_{j_2}, \dots, X_{j_{N_u}}$ dépassant u et notons $Y_i = X_{j_i} - u$. On ajuste une loi *GPD* aux excès Y_1, \dots, Y_{N_u} pour obtenir les estimateurs du maximum de vraisemblance $\hat{\xi}$ et $\hat{\sigma}$. Une application directe du théorème 2.5.1 montre que $\bar{F}(x) = \bar{F}(u)\bar{F}_u(x - u)$ pour $x > u$, nous pouvons estimer la queue $\bar{F}(x) = 1 - F(x)$ par

$$\hat{\bar{F}}(x) = \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x - u}{\hat{\sigma}} \right)^{-1/\hat{\xi}_u}, \quad x \geq u. \quad (4.2)$$

Par inversion de (4.2), un estimateur du quantile $VaR(\alpha_n) = F^{\leftarrow}(1 - \alpha_n > u)$ est alors donné par

$$VaR(\alpha_n) = u + \frac{\hat{\sigma}}{\hat{\xi}} \left(\left(\frac{n}{N_u} \alpha_n \right)^{-\hat{\xi}_u} - 1 \right).$$

En pratique, le seuil u est en général la $(k_n + 1)$ ième plus grande observation ; c'est-à-dire $X_{n-k_n, n}$, qui correspond à la statistique d'ordre des observations situées au niveau de la queue de la distribution. On obtient dans ce cas $N_u = k$ et si on note $(\hat{\xi}_{k_n}^{POT}, \hat{\sigma}_{k_n}^{POT})$ les estimateurs $(\hat{\xi}_{k_n}^{MOM}, \hat{\sigma}_{k_n}^{MOM})$ ou $(\hat{\xi}_{k_n}^{PWM}, \hat{\sigma}_{k_n}^{PWM})$ du couple (ξ, σ) donne dans la sous

section 3.4, on obtient l'estimateur des quantiles extrêmes suivant :

$$VaR(\alpha_n) = X_{n-k_n, n} + \frac{\widehat{\sigma}_{k_n}^{POT}}{\widehat{\xi}_{k_n}^{POT}} \left(\left(\frac{n}{k_n} \alpha_n \right)^{-\widehat{\xi}_{k_n}^{POT}} - 1 \right) \text{ pour } \alpha_n < \frac{k_n}{n}.$$

4.3.2 Approche basée sur l'estimateur de Hill

Notons que parmi les estimateurs d'indices des valeurs extrêmes le plus connu est sans conteste l'estimateur de Hill, qui peut être interprète comme un estimateur de la pente du "Pareto Quantile Plot". Il apparait donc naturel d'estimer un quantile extrême a l'aide de cet estimateur de ξ en extrapolant le long de la droite du "Pareto quantile plot" d'équation

$$y = \log X_{n-k_n, n} + \widehat{\xi}_{k_n}^H \left(x + \log \frac{k_n + 1}{n + 1} \right),$$

de point d'origine $(\log((n+1)/(k_n+1)), \log X_{n-k_n, n})$, voir par exemple (Beirlant et al. (1996b)). Ceci conduit en prenant $x = -\log \alpha_n$ un estimateur le plus connu d'un quantile extrême $VaR(\alpha_n) = F^{\leftarrow}(1 - \alpha_n)$ introduit par (Weissman (1978)) :

$$VaR(\alpha_n) = X_{n-k_n, n} \left(\frac{k_n + 1}{(n + 1)\alpha_n} \right)^{\widehat{\xi}_{k_n}^H}. \quad (4.3)$$

Nous avons déjà vu que l'estimateur de Hill est valable uniquement dans le domaine de Fréchet (indices positifs). Autrement si on se restreint aux fonctions dans ce domaine pour lesquelles on a la caractérisation suivante $\bar{F}(x) = x^{-\frac{1}{\xi}} \ell_F(x)$, avec ℓ_F étant une fonction à variation lente a l'infini et $\xi > 0$. Cette caractérisation implique que

$$\begin{aligned} F^{\leftarrow}(1 - \alpha_n) &= U(1/\alpha_n) = \alpha_n^{-\xi} \ell_U(1/\alpha_n) \text{ avec } \alpha_n \leq 1/n, \\ F^{\leftarrow}(1 - \beta_n) &= U(1/\beta_n) = \beta_n^{-\xi} \ell_U(1/\beta_n) \text{ avec } \beta_n \leq 1/n, \end{aligned}$$

où $\ell_U(1/\beta_n)$ est une fonction a variation lente a l'infini. Comme $\frac{\ell_U(tx)}{\ell_U(t)} \rightarrow 1$ quand $t \rightarrow \infty$, alors pour β_n suffisamment petit, on a

$$F^{\leftarrow}(\alpha_n) \simeq F^{\leftarrow}(\beta_n) \left(\frac{\beta_n}{\alpha_n} \right)^{\xi}$$

En remplaçant $F^{\leftarrow}(\beta_n)$ par son estimateur naturel $X_{n-[n\beta_n],n}$ et ξ par un estimateur quelconque d'indices positifs note $\xi_{k_n}^+$, on obtient aussi l'estimateur général de type (Weissman (1978)) défini par

$$VaR(\alpha_n) = X_{n-[n\beta_n],n} \left(\frac{\beta_n}{\alpha_n} \right)^{\widehat{\xi}_{k_n}^+},$$

où $[x]$ désigne la partie entière de x . Remarquons que pour $\beta_n = (k_n + 1)/(n + 1)$ et $\widehat{\xi}_{k_n}^+ = \widehat{\xi}_{k_n}^H$ on retrouve l'estimateur de (4.3). Les propriétés asymptotiques de l'estimateur de Weissman sont discutées et un intervalle de confiance a été construit sous certaines conditions sur la fonction de distribution F , k_n et α_n dans (Paul et al. (1997), Matthys and Beirlant (2003), et Markovich (2005)).

4.3.3 Approche basée sur l'estimateur des moments

Plus généralement, nous introduisons ici une méthode d'estimation du quantile extrême pour $\xi \in \mathbb{R}$ que nous allons utiliser dans la suite. En utilisant l'estimateur des moments, (Dekkers et al. (1989b)) ont proposé d'estimer $VaR(\alpha_n) = U(\frac{1}{\alpha_n})$ pour $\alpha_n \rightarrow 0$ comme suit :

$$VaR(\alpha_n) = X_{n-k_n,n} + \widehat{a} \left(\frac{k_n \left(\frac{n}{k_n} \alpha_n \right)^{-\widehat{\xi}_{k_n}^M} - 1}{\widehat{\xi}_{k_n}^M} \right),$$

où $\widehat{a} \left(\frac{k_n}{n} \right) = X_{n-k_n,n} \widehat{\xi}_{k_n}^H \max(1 - \widehat{\xi}_{k_n}^M, 1)$. La normalité asymptotique de cet estimateur a été établie par (Dekkers et al. (1989b)) et par de (de Haan and Rootzén (1993)) sous diverses hypothèses, portant d'une part sur la queue et d'autre part sur α_n .

Chapitre 5

Modèle GPD en présence de censure

Dans ce chapitre nous nous intéressons à l'estimation des paramètres σ, ξ d'un modèle statistique gouverné par une distribution $GPD_{\sigma, \xi}$ censurées. Dans le contexte des données censurées à droite, différents estimateurs de l'indice des valeurs extrêmes ont été proposés par (Beirlant et al. (1996a) section 2.7) et (Reiss and Thomas (2007) section 6.1). Ils sont tous construits de la même façon : l'estimateur usuel (sans censure) est divisé par la proportion d'observations non censurées au dessus d'un certain seuil mais sans résultats asymptotiques. Récemment, (Einmahl et al. (2008)) 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 (σ, ξ) les paramètres de la GPD .

5.1 Différents types de censures

Une des caractéristiques des données de survie ou de queue de distribution est l'existence d'observations incomplètes. En effet, les données sont souvent recueillies partiellement, notamment à cause du phénomène de censure. Les données censurées proviennent du fait qu'on n'a pas accès à toute l'information. L'observation de la variable aléatoire X est soumise à diverses perturbations indépendantes ou non du phénomène étudié. Il est donc très important de disposer de bonnes méthodes d'analyse pour ce type de données que l'on rencontre fréquemment, essentiellement en fiabilité, en épidémiologie et en finance.

La censure des données se fait selon plusieurs mécanismes tels la censure à droite, la cen-

sure à gauche, la censure par intervalle. On se réfère à (Klein and Moeschberger (1997)) où ces types de censure sont présentés avec des exemples.

Définition 5.1.1. *Censure à droite*

Il y a censure à droite lorsque nous observons la censure C (et non pas la durée de vie d'intérêt T) et que nous savons que $T > C$. Ce modèle est le plus fréquent en pratique, il est par exemple adapté au cas où l'événement d'intérêt est le temps de survie à une maladie et où la date de fin de l'étude est préalablement fixée ; les patients vivants à la fin de l'étude fournissent des données censurées à droite. Les observations sont des répliques du couple $(T \wedge C, \Delta = 1_{\{T \leq C\}})$ où Δ vaut 1 quand l'observation est complète (elle correspond à une donnée de la variable d'intérêt) et vaut 0 si la donnée est censurée.

Définition 5.1.2. *Censure à gauche*

Il y a censure à gauche lorsque nous observons la censure C (et non pas la durée de vie T) et que nous savons que $T < C$. Un phénomène symétrique au précédent se produit, le patient a déjà subi l'événement avant où l'instant on commence l'étude. Ce modèle est par exemple adapté au cas où l'on s'intéresse à l'âge auquel un individu commence à accomplir une tâche. Tout ce qu'on sait chez l'individu censuré est que le véritable âge est inférieur à la valeur observée (l'âge au moment de l'étude par exemple).

Définition 5.1.3. *Censure par intervalle*

Dans ce cas de figure on est dans l'incapacité d'observer la variable d'intérêt X , cependant on sait que sa valeur se trouve dans un intervalle aléatoire (Z, Y) .

5.2 Distribution d'une variable censurée

La distribution d'une variable censurée est obtenue comme distribution conjointe d'un couple (X, Δ) où X est la variable d'observation et Δ la variable de censure.

5.2.1 Variable censurée à droite

On considère un couple de variables aléatoires positives (T, C) , où T est la variable d'intérêt et C est la variable de censure. Dans le modèle de censure considéré, l'observation est le couple $(X = \min(T, C), \Delta = 1_{\{T \leq C\}})$. T et C sont supposées indépendantes de fonctions de répartition respectives F et G et de fonctions de densité respectives f et g .

Si $\Delta = 0$

$$\begin{aligned}
P(X \leq t, \Delta = 0) &= P(T \wedge C \leq t, T > C) = P(C \leq t, T > C) \\
&= \int_{\{y \leq t, x > y\}} dP_{(T,C)}(x, y) = \int_0^t \int_y^{+\infty} dP_T(x) dP_C(y) \\
&= \int_0^t dP_C(y) \int_y^{+\infty} dP_T(x) = \int_0^t (1 - F(y))g(y) dy.
\end{aligned}$$

Si $\Delta = 1$

$$\begin{aligned}
P(X \leq t, \Delta = 1) &= P(T \wedge C \leq t, T \leq C) = P(T \leq t, T \leq C) \\
&= \int_{\{x \leq t, x \leq y\}} dP_{(T,C)}(x, y) = \int_0^t \int_y^{+\infty} dP_T(x) dP_C(y) \\
&= \int_0^t dP_T(x) \int_x^{+\infty} dP_C(y) = \int_0^t (1 - G(x))f(x) dx.
\end{aligned}$$

En dérivant par rapport à t , la densité du couple (X, Δ) est :

$$P(t, \delta) = \{f(t)^\delta (1 - F(t))^{1-\delta}\} \{g(t)^{1-\delta} (1 - G(t))^\delta\}$$

5.2.2 Variable censurée par intervalle

Soit X une variable aléatoire avec fonction de répartition F_X , défini dans l'espace de probabilité $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+}, P_\theta)_{\theta \in \Theta}$. En présence de la censure par intervalle au lieu d'observer la variable d'intérêt X on observe le triplet (Y, Z, Δ) . On suppose que la densité jointe de Y et Z ne dépend pas de θ , et que la variable d'intérêt X est indépendant du couple (Y, Z) . Le composant Δ est une variable aléatoire discrète définie telle que

$$\Delta = \begin{cases} 1 & \text{si } Y < X \leq Z \\ 2 & \text{si } X > Z \\ 3 & \text{si } X \leq Y \end{cases}$$

La distribution de probabilité jointe du triplet (Y, Z, Δ) est définie telle que

$$\begin{aligned}
P(y \leq Y < y + dy, z \leq Z < z + dz, \Delta = \delta) &= P(y \leq Y < y + dy, z \leq Z < z + dz) \\
&\quad P\{\Delta = \delta / (y \leq Y < y + dy, z \leq Z < z + dz)\}
\end{aligned}$$

Il est évident que

$$\lim_{dy \rightarrow 0, dz \rightarrow 0} \frac{P(y \leq Y < y + dy, z \leq Z < z + dz)}{dydz} = f_{(Y,Z)}(y, z)$$

où $f_{(Y,Z)}$ est la fonction de densité jointe de (Y, Z) . De plus, pour $\delta = 1, 2, 3$ on a

$$\lim_{dy \rightarrow 0, dz \rightarrow 0} P\{(\Delta = 1) / (y \leq Y < y + dy, z \leq Z < z + dz)\} = P(y < X \leq z)$$

$$\lim_{dy \rightarrow 0, dz \rightarrow 0} P\{(\Delta = 2) / (y \leq Y < y + dy, z \leq Z < z + dz)\} = P(X > z)$$

$$\lim_{dy \rightarrow 0, dz \rightarrow 0} P\{(\Delta = 3) / (y \leq Y < y + dy, z \leq Z < z + dz)\} = P(X \leq y)$$

La distribution de la variable aléatoire conditionnelle $(\Delta = \delta / Y = y, Z = z)$ peut être résumée comme :

$$\begin{aligned} P(\Delta = \delta | Y = y, Z = z) &= \lim_{dy \rightarrow 0, dz \rightarrow 0} P(\Delta = \delta / y \leq Y < y + dy, z \leq Z < z + dz) \\ &= P(y \leq X < z)^{\mathbf{1}_{\{\delta=1\}}(\delta)} P(X < y)^{\mathbf{1}_{\{\delta=2\}}(\delta)} P(X \geq z)^{\mathbf{1}_{\{\delta=3\}}(\delta)} \end{aligned}$$

Soit $f_{(Y,Z,\Delta)}$ la fonction de densité de (Y, Z, Δ) définie telle que

$$\lim_{dy \rightarrow 0, dz \rightarrow 0} P(y \leq Y < y + dy, z \leq Z < z + dz, \Delta = \delta) = f_{(Y,Z,\Delta)}(y, z, \delta)$$

où

$$\begin{aligned} f_{(Y,Z,\Delta)}(y, z, \delta) &= f_{(Y,Z)}(y, z) P(y \leq X < z)^{\mathbf{1}_{\{\delta=1\}}(\delta)} P(X \leq y)^{\mathbf{1}_{\{\delta=2\}}(\delta)} P(X > z)^{\mathbf{1}_{\{\delta=3\}}(\delta)} \\ &= f_{(Y,Z)}(y, z) (F_X(z) - F_X(y))^{\mathbf{1}_{\{\delta=1\}}(\delta)} F_X(y)^{\mathbf{1}_{\{\delta=2\}}(\delta)} (1 - F_X(z))^{\mathbf{1}_{\{\delta=3\}}(\delta)} \end{aligned}$$

5.3 Asymptotic properties of maximum pseudo-likelihood of a GPD Model parameters estimators with interval censoring

In this section, we are interested in the estimation of parameters of a GPD model with interval censoring. First, in a general parametric model with interval censoring we show that the parameter estimates obtained by the method of maximum pseudo-likelihood still possess the properties of efficiency and asymptotic normality. Then, using numerical methods relying on the Barzilai-Borwein algorithm we test the robustness of the parameter estimates of a simulated GPD sample.

5.3.1 Introduction

Survival or tail distribution data are often characterized by incomplete observations. Indeed, data are often collected partially, in particular because of the censoring phenomenon. The censored data come from the fact that we do not have access to the complete information. The observation of the random variable of interest X is frequently subject to various outside influences independent or not from the studied phenomenon. It is therefore very important to have good analytical methods for this kind of data, which is frequently found, mainly in reliability, epidemiology and finance (Beirlant et al. (2006b), Reiss and Thomas (2007), Thomas et al. (2016)). The most common types of censoring are the right censoring, the left censoring (Lawless (2003)) and the double censoring (Turnbull (1974)). More precisely, let C be the censoring variable. If we observe the couple $(\min(X, C), \Delta = \mathbf{1}_{\{X \leq C\}})$ instead of the variable of interest X , it is the right censoring. If we observe the couple $(\max(X, C), \Delta = \mathbf{1}_{\{X \geq C\}})$ instead of the variable of interest X , it is the left censoring; it is a double censoring when there is a right censoring and a left censoring simultaneously. Moreover, Kalbfleisch and Prentice (Klein and Moeschberger (1997)) introduced the concept of interval censoring. In this case, we cannot observe the variable of interest X , but we know that its value is in a random interval (Y, Z) . In the literature, many contributions have mainly been concerned with the properties of nonparametric estimators in models subject to interval censoring (Gentleman and Geyer (1994), Clifford Anderson-Bergman (2016), Kalbfleisch and Prentice (2002)). In this contribution, we are interested in parameters estimation in the parametric models with interval censoring. In this kind of model, the variable of interest X may not be observed. Thus, the statistical model will be poorly specified and the maximum likelihood method becomes weak. In fact, in the presence of interval censoring, a triplet (Y, Z, Δ) of independent random variables is observed. In this situation, the most appropriate estimation method is the pseudo-maximum likelihood method, which is more adapted to discrete variables (Gourieroux et al. (1984)). The idea in this method is to replace the density of the model with that of a poorly specified pseudo-model but with interesting properties relating to the first moment. The aim of our work is to show that the pseudo-maximum likelihood estimator still possesses the properties of convergence and asymptotic normality. Finally, the results obtained in the general framework are applied to a GPD model with interval

censoring.

5.3.2 Parametric model with interval censoring

Let $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+}, P_{\boldsymbol{\theta}})_{\boldsymbol{\theta} \in \Theta}$ be a parametric model associated with a random variable X with distribution function F_X . We suppose that the set Θ is compact in \mathbb{R}^d and contains the true value $\boldsymbol{\theta}_0$ of $\boldsymbol{\theta}$. Moreover, we assume that the model is identifiable, that is the map $\boldsymbol{\theta} \rightarrow P_{\boldsymbol{\theta}}$ is injective for all $\boldsymbol{\theta}$ in Θ . In the presence of interval censoring, instead of observing the variable of interest X we observe the triplet (Y, Z, Δ) with its values in $\mathcal{D} = \mathbb{R}_+^2 \times \{1, 2, 3\}$. The components Y and Z are two random variables absolutely continuous with respect to the Lebesgue measure. We assume that the joint density of Y and Z does not depend on $\boldsymbol{\theta}$, and that the variable of interest X is independent of the pair (Y, Z) . The component Δ is a discrete random variable defined such that

$$\Delta = \begin{cases} 1 & \text{if } Y < X \leq Z \\ 2 & \text{if } X > Z \\ 3 & \text{if } X \leq Y \end{cases}$$

Let $(Y_k, Z_k, \Delta_k)_{k=1, \dots, n}$ be a sample with the same distribution as (Y, Z, Δ) with probability density function $f_{(Y, Z, \Delta)}(y, z, \delta; \boldsymbol{\theta})$. To compute the likelihood function $L(\boldsymbol{\theta}) = \prod_{k=1}^n f_{(Y, Z, \Delta)}(y_k, z_k, \delta_k; \boldsymbol{\theta})$ of this sample we will proceed in the same way as Klein And Moeschberger ([6]). We then obtain :

$$L(\boldsymbol{\theta}) = \prod_{k=1}^n f_{(Y, Z)}(y_k, z_k) \prod_{k=1}^n U(y_k, z_k, \delta_k; \boldsymbol{\theta})$$

where $f_{(Y, Z)}(y_k, z_k)$ is the joint density function of the couple (Y, Z) , and

$$U(y_k, z_k, \delta_k; \boldsymbol{\theta}) = (F_X(y_k; \boldsymbol{\theta}) - F_X(z_k; \boldsymbol{\theta}))^{1_{\{\delta=1\}}} (1 - F_X(y_k; \boldsymbol{\theta}))^{1_{\{\delta=2\}}} F_X(z_k; \boldsymbol{\theta})^{1_{\{\delta=3\}}}$$

The maximum likelihood estimator of $\boldsymbol{\theta}$ is obtained by maximizing the log-likelihood $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta})$. But maximizing $\ell(\boldsymbol{\theta})$ is equivalent to maximizing the log-pseudo-likelihood

$$\tilde{\ell}(\boldsymbol{\theta}) = \sum_{k=1}^n \log U(y_k, z_k, \delta_k; \boldsymbol{\theta})$$

In this case, the maximum pseudo-likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta} \tilde{\ell}(\boldsymbol{\theta}) \quad (5.1)$$

It remains to check that the maximum pseudo-likelihood estimator still has the properties of efficiency and asymptotic normality.

We assume in the sequel that the function U satisfies the following conditions :

- (C₁) for all $(y, z, \delta) \in \mathcal{D}$, the function $U(y, z, \delta; \boldsymbol{\theta})$ is continuous at each $\boldsymbol{\theta} \in \Theta$,
- (C₂) there is a positive function $d : \mathcal{D} \rightarrow \mathbb{R}$, with $E[d(Y, Z, \Delta; \boldsymbol{\theta})] < \infty$ and such as $|\log U(Y, Z, \Delta; \boldsymbol{\theta})| \leq d(Y, Z, \Delta; \boldsymbol{\theta})$ a.s; $\forall \boldsymbol{\theta} \in \Theta$,
- (C₃) for all $(y, z, \delta) \in \mathcal{D}$, the function $U(y, z, \delta; \boldsymbol{\theta})$ is smooth and twice continuously differentiable with respect to $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$,
- (C₄) for all $(i, j) \in \{1, \dots, d\}^{\otimes 2}$, Fisher's information matrix $\mathcal{I}(\boldsymbol{\theta})$ exists,
- (C₅) for all $(i, j) \in \{1, \dots, d\}^{\otimes 2}$, the Hessian matrix $J(\boldsymbol{\theta})$ of $U(y, z, \delta; \boldsymbol{\theta})$ exists and is invertible.

5.3.3 Efficiency of the maximum pseudo-likelihood estimator

Let $(\hat{\boldsymbol{\theta}}_n)_{n \geq 1}$ be a sequence of maximum pseudo-likelihood estimators.

Proposition 5.3.1. *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the conditions C₁ et C₂ then*

$$\hat{\boldsymbol{\theta}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0$$

Démonstration. Let $M_n(\boldsymbol{\theta})$ and $M(\boldsymbol{\theta})$ be such that :

$$M_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{k=1}^n \log U(Y_k, Z_k, \Delta_k; \boldsymbol{\theta})$$

et

$$M(\boldsymbol{\theta}) = E(\log U(Y, Z, \Delta; \boldsymbol{\theta}))$$

From the equation (5.1) and the definition of $M_n(\boldsymbol{\theta})$, it is clear that $\hat{\boldsymbol{\theta}}_n$ maximizes $M_n(\boldsymbol{\theta})$. To conclude with the Van der Vaart's theorem (Van Der Vaart (1998), p.45), it remains to show that

$$\sup_{\boldsymbol{\theta} \in \Theta} |M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \xrightarrow{\mathbb{P}} 0 \quad (5.2)$$

and

$$\sup_{\boldsymbol{\theta} \in \mathcal{C}} M(\boldsymbol{\theta}) < M(\boldsymbol{\theta}_0) \quad (5.3)$$

where $\mathcal{C} = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \varepsilon\}$.

The function U is continuous by definition (C_1) and is bounded by construction (C_2), so the conditions of the Lemma of Newey and McFadden ([Newey and McFadden \(1994\)](#)) are satisfied, from where the condition (5.2).

It remains to check that the last condition of the theorem of Van Der Vaart is satisfied. Let us begin by showing that

$$M(\boldsymbol{\theta}) \leq M(\boldsymbol{\theta}_0), \quad \forall \boldsymbol{\theta} \in \Theta \quad (5.4)$$

From the definition of $M(\boldsymbol{\theta})$ it is clear that

$$M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) = E \left(\log \frac{U(Y, Z, \Delta; \boldsymbol{\theta})}{U(Y, Z, \Delta; \boldsymbol{\theta}_0)} \right)$$

Using Jensen's inequality, we can write

$$M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0) = E \left[\log \frac{U(Y, Z, \Delta; \boldsymbol{\theta})}{U(Y, Z, \Delta; \boldsymbol{\theta}_0)} \right] \leq \log E \left[\frac{U(Y, Z, \Delta; \boldsymbol{\theta})}{U(Y, Z, \Delta; \boldsymbol{\theta}_0)} \right]$$

Furthermore

$$\begin{aligned} \log E \left[\frac{U(Y, Z, \Delta; \boldsymbol{\theta})}{U(Y, Z, \Delta; \boldsymbol{\theta}_0)} \right] &= \log \left(\iint \sum_{\delta \in \{1,2,3\}} \frac{U(y, z, \delta; \boldsymbol{\theta})}{U(y, z, \delta; \boldsymbol{\theta}_0)} f_{Y,Z}(y, z) U(y, z, \delta; \boldsymbol{\theta}_0) dy dz \right) \\ &= \log \left(\iint \sum_{\delta \in \{1,2,3\}} f_{Y,Z,\Delta}(y, z, \delta; \boldsymbol{\theta}) dy dz \right) = 0 \end{aligned}$$

Whence the inequality (5.4).

Since $M(\boldsymbol{\theta})$ is continuous on the compact \mathcal{C} , there exists $\bar{\boldsymbol{\theta}} \in \mathcal{C}$ such that

$$\sup_{\boldsymbol{\theta} \in \mathcal{C}} M(\boldsymbol{\theta}) = M(\bar{\boldsymbol{\theta}})$$

The identifiability of the model, the strict concavity of the logarithm and the inequality (5.4) lead to :

$$M(\bar{\boldsymbol{\theta}}) < M(\boldsymbol{\theta}_0)$$

Finally, the three conditions of Van der Vaart's theorem are verified from where the proposition 1. \square

5.3.4 Asymptotic normality of the maximum pseudo-likelihood estimator

To facilitate the rest of the presentation, let us define the score function such that :

$$D(Y, Z, \Delta; \boldsymbol{\theta}) = \left(\frac{\partial}{\partial \theta_1} \log U(Y, Z, \Delta; \boldsymbol{\theta}), \dots, \frac{\partial}{\partial \theta_d} \log U(Y, Z, \Delta; \boldsymbol{\theta}) \right).$$

Lemma 5.3.1. *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the condition C_3 , then*

$$E [D(Y, Z, \Delta; \boldsymbol{\theta}_0)] = 0$$

Démonstration. From the relation (5.4) it is obvious that $\boldsymbol{\theta}_0$ maximizes $M(\boldsymbol{\theta})$. Since the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the condition C_3 , therefore for all $j \in \{1, \dots, d\}$ we have :

$$E \left(\frac{\partial}{\partial \theta_j} \log U(Y, Z, \Delta; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) = \frac{\partial}{\partial \theta_j} E \left(\log U(Y, Z, \Delta; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) = \left(\frac{\partial}{\partial \theta_j} M(\boldsymbol{\theta}) \right) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0,$$

whence the Lemma 2. \square

Lemma 5.3.2. *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the condition C_4 , then*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}_0) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0))$$

Démonstration. *Considering the Lemma 2, it is obvious that the covariance matrix of the vector $D(Y, Z, \Delta; \boldsymbol{\theta}_0)$ coincides with the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta}_0)$, so according to the central limit theorem we have*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}_0) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)) \quad (5.5)$$

Reconsider the sequence $(\hat{\boldsymbol{\theta}}_n)_{n \geq 1}$ of the maximum pseudo-likelihood estimators.

Proposition 5.3.2. *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the conditions $C_1, C_2, C_3, C_4,$ and $C_5,$ then*

$$\sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\theta}_0})$$

where

$$\Sigma_{\boldsymbol{\theta}_0} = J(\boldsymbol{\theta}_0)^{-1} \mathcal{I}(\boldsymbol{\theta}_0) J(\boldsymbol{\theta}_0)^{-1}$$

Démonstration. *The score function $D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}_n)$ vanishes at $\hat{\boldsymbol{\theta}}$ since $\hat{\boldsymbol{\theta}}$ maximizes $\log U(Y, Z, \Delta; \boldsymbol{\theta})$. Then, using the Taylor-Young formula in the neighborhood of $\boldsymbol{\theta}_0,$ we can write*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n D(Y_k, Z_k, \Delta_k; \hat{\boldsymbol{\theta}}_n) = \frac{1}{\sqrt{n}} \sum_{k=1}^n D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}_0) + J_n(\boldsymbol{\theta}_0) \sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + R_n = 0 \quad (5.6)$$

where $J_n(\boldsymbol{\theta}_0)$ is such as

$$J_n(\boldsymbol{\theta}_0) = \left(J_n^{i,j}(\boldsymbol{\theta}_0) \right)_{1 \leq i, j \leq d} = \left(\frac{1}{n} \sum_{k=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log U(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}_0) \right)_{1 \leq i, j \leq d}$$

$J_n^{i,j}(\boldsymbol{\theta}_0)$ is defined as empirical mean of a function of some random variables. Then, using the weak law of large numbers we can say that

$$J_n(\boldsymbol{\theta}_0) \xrightarrow{\mathbb{P}} J(\boldsymbol{\theta}_0) \quad (5.7)$$

Therefore, we can say that $J_n(\boldsymbol{\theta}_0)$ is almost surely invertible.

Finally, by Lemma 3 and using the equation (5.5), we can write that

$$\sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, J(\boldsymbol{\theta}_0)^{-1} \mathcal{I}(\boldsymbol{\theta}_0) J(\boldsymbol{\theta}_0)^{-1})$$

5.3.5 GPD model with interval censoring

In many applications, fitting the tail data is the main concern. The Generalized Pareto distribution (GPD) was developed as a distribution that can model tails of a wide variety of distributions (Reiss and Thomas (2007), Beirlant et al. (2006b)). The scarcity of observations in the tail of a distribution requires elaborating robust estimation methods. Parametric models with interval censoring may be appropriate for this kind of situation.

Let us retake the model considered in Section 2. Let X_1, X_2, \dots, X_n be a sample from a GPD distribution in the Fréchet MDA with distribution function $F(x; \sigma, \gamma)$ defined such as

$$F(x; \sigma, \gamma) = 1 - \left(1 + \gamma \frac{x}{\sigma}\right)^{-\frac{1}{\gamma}}$$

where $x > -\frac{\sigma}{\gamma}$.

We assume that the model under consideration is subject to interval censoring. Therefore, the parameters estimation can be carried out using the maximum pseudo-likelihood method.

Let $\tilde{\ell}(\sigma, \gamma)$ be the pseudo-likelihood function defined such as

$$\begin{aligned} \tilde{\ell}(\gamma, \sigma) = & \sum_{k=1}^n \left[1_{\{\delta=1\}} \log \left(\left(1 + \frac{\gamma}{\sigma} z_k\right)^{-\frac{1}{\gamma}} - \left(1 + \frac{\gamma}{\sigma} y_k\right)^{-\frac{1}{\gamma}} \right) \right. \\ & \left. - 1_{\{\delta=2\}} \frac{1}{\gamma} \log \left(1 + \frac{\gamma}{\sigma} y_k\right) + 1_{\{\delta=3\}} \log \left\{ 1 - \left(1 + \frac{\gamma}{\sigma} z_k\right)^{-\frac{1}{\gamma}} \right\} \right] \end{aligned} \quad (5.8)$$

The maximum pseudo-likelihood estimators $\hat{\sigma}$ and $\hat{\gamma}$ maximize equation (5.8). This optimization is carried out using the Barzilai-Borwein algorithm (BB) (Varadhan (2009)) which is found in the form of a package in the software R (R Development Core Team (2008a)). To check the asymptotic properties of the maximum pseudo-likelihood estimators we will proceed as follows.

Let X be the variable of interest of a $GPD(1, 0.2)$ distribution censored by interval. But in the presence of an interval censoring instead of observing the variable of interest X , we observe the triplet (Y, Z, Δ) . Thereby the sampling will be applied to the couple (Y, Z) . Let n_1, n_2 et n_3 be three increasing sample sizes. For each size, N samples of the couple (Y, Z) are replicated. The estimates of the parameters σ and γ of the GPD are obtained as empirical averages of the N estimates. The numerical results of this procedure are summarized in table 1.

N=10000	$n_1=100$	$n_2=500$	$n_3=1000$
$\hat{\gamma}$	0.2092073	0.1913203	0.1996281
MAE	0.1976629	0.10715128	0.07878693
MSE	0.06230731	0.01686642	0.009497064
$\hat{\sigma}$	1.011984	1.010003	1.009373
MAE	0.20418	0.1068269	0.0796333
MSE	0.06533427	0.01753554	0.00977675

TABLE 5.1 – Efficiency of estimators $\hat{\gamma}$ and $\hat{\sigma}$

In Table 1, we clearly see that the values of $\hat{\gamma}$ and $\hat{\sigma}$ converge to the true values $\gamma = 0.2$ and $\sigma = 1$ with very small mean square errors.

To check the asymptotic normality of $\hat{\gamma}$ and $\hat{\sigma}$, we simulate $N = 100$ samples of size $n = 100$, and we draw the QQ-plot for each parameter in Figure 1.

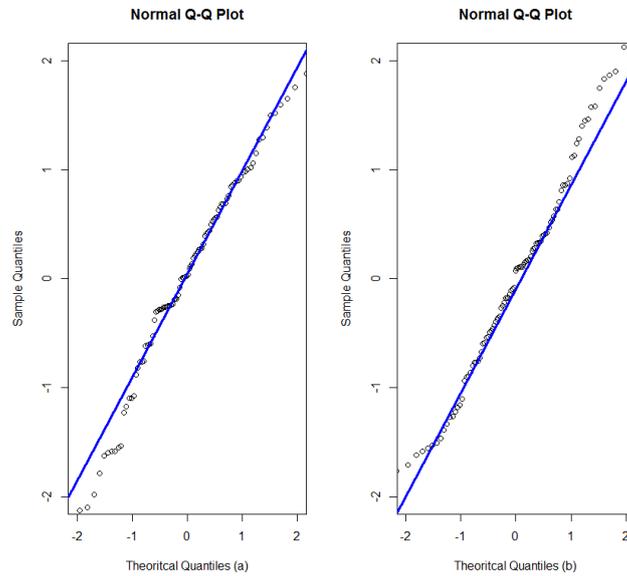


FIGURE 5.1 – (a) the QQ-plot of $\hat{\gamma}$ and (b) the QQ-plot of $\hat{\sigma}$

According to the two graphs in Figure 1, it is clear that the distributions of $\hat{\gamma}$ and $\hat{\sigma}$ are almost normal. Moreover, the asymptotic normality of $\hat{\gamma}$ and $\hat{\sigma}$ makes it possible to construct confidence intervals for each parameter such as

$$\gamma \in \left[\hat{\gamma} - q_{(1-\frac{\alpha}{2})} \sqrt{\frac{Var(\hat{\gamma})}{n}}; \hat{\gamma} + q_{(1-\frac{\alpha}{2})} \sqrt{\frac{Var(\hat{\gamma})}{n}} \right]$$

$$\sigma \in \left[\hat{\sigma} - q_{(1-\frac{\alpha}{2})} \sqrt{\frac{Var(\hat{\sigma})}{n}}; \hat{\sigma} + q_{(1-\frac{\alpha}{2})} \sqrt{\frac{Var(\hat{\sigma})}{n}} \right]$$

where $q_{(1-\frac{\alpha}{2})} = \phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ and where ϕ is the distribution function of the standard normal distribution.

5.4 Analysis of breast cancer data in framework of a GPD model with interval censoring

In this section, we are interested in a hypothesis testing problem within the framework of a GPD model with interval censoring. For this purpose, we first develop the calculation of the likelihood function using conditional probabilities to achieve the same expression proposed by Klein and Moeschberger. Next, we show that the properties of the maximum pseudo-likelihood estimates of the model parameters, and essentially the asymptotic normality, are preserved. Finally, we built a hypothesis testing to compare two types of breast cancer treatment as part of the model mentioned above.

5.4.1 Introduction

In practice, many problems are better formalized within censored models. In an overview of available literature, a lot of contributions have dealt with censored models on the right (Lawless (2003)), left (Samson et al. (2006)) or right and left (Turnbull (1974)). Furthermore, the concept of interval censoring has been introduced (Klein and Moeschberger (1997)) to generalize the different types of censoring. Statistical inference in such models has been developed using more generally non-parametric or semi-parametric methods (Gentleman and Geyer (1994), Clifford Anderson-Bergman (2016), Kalbfleisch and Prentice (2002), Lin et al. (2016)). In this work, we aim to develop a statistical inference within a parametric model by means of the study of a corpus of data (Lindsey and Ryan (1998)) concerning two types of breast cancer treatment. We want to compare these two treatments using a hypothesis testing strategy built around the tail indexes of empirical data distributions for both types of care. In the second section, we reformulate the likelihood function using a different calculation technique to achieve the same expression as in (Klein and Moeschberger (1997)). The estimates of the model parameters are obtained by maximizing a pseudo-likelihood function. Furthermore, we propose a proof of the asymptotic normality of model parameter estimates because this property is essential for the

construction of a hypothesis testing. In the last section, we define the variable of interest X as the duration of the aesthetic comfort of the patients breasts, that is the time elapsed between the beginning of the treatment and the appearance of breast retraction. Without treatment, breast retraction appears fairly quickly. However, treatment by radiotherapy or chemotherapy extends the patient's aesthetic comfort time. Therefore, we can assume that X is an excess above a threshold. Thereby, we reposition our study in the context of a GPD model. Afterwards, we consider a test statistic based on the difference between the tail indexes of the distributions and standardized using the results of the previous sections as in Wald (Wald (1943)). Finally, the rejection region is determined and the power of the test is calculated.

5.4.2 Parametric model with interval censoring

Let X be a random variable with probability distribution function F_X , defined on the probability space $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+}, P_{\theta})_{\theta \in \Theta}$. Let us suppose that the set Θ is compact in \mathbb{R}^d and contains the true value. Moreover, let us assume that the model is identifiable, that is the application $\theta \rightarrow P_{\theta}$ is injective for all θ in Θ . Data collection for some random phenomena is subject to interval censoring, so instead of observing the variable of interest X , we observe the triplet (Y, Z, Δ) which takes its values in $\mathcal{D} = \mathbb{R}_+^2 \times \{1, 2, 3\}$. The components Y and Z are two random variables absolutely continuous with respect to the Lebesgue measure. We assume that the joint density of Y and Z does not depend on θ , and that the variable of interest X is independent of the pair (Y, Z) . The component Δ is a discrete random variable defined such as

$$\Delta = \begin{cases} 1 & \text{if } Y < X \leq Z \\ 2 & \text{if } X > Z \\ 3 & \text{if } X \leq Y \end{cases}$$

The joint probability distribution of the triplet (Y, Z, Δ) is such that

$$P(y \leq Y < y + dy, z \leq Z < z + dz, \Delta = \delta) = P(y \leq Y < y + dy, z \leq Z < z + dz) P\{\Delta = \delta / (y \leq Y < y + dy, z \leq Z < z + dz)\}$$

It obvious that,

$$\lim_{dy \rightarrow 0, dz \rightarrow 0} \frac{P(y \leq Y < y + dy, z \leq Z < z + dz)}{dydz} = f_{(Y,Z)}(y, z)$$

where $f_{(Y,Z)}$ is the joint density function of the couple (Y, Z) . Moreover, for $\delta = 1, 2, 3$ we have

$$\begin{aligned}\lim_{dy \rightarrow 0, dz \rightarrow 0} P \{(\Delta = 1) / (y \leq Y < y + dy, z \leq Z < z + dz)\} &= P(y < X \leq z) \\ \lim_{dy \rightarrow 0, dz \rightarrow 0} P \{(\Delta = 2) / (y \leq Y < y + dy, z \leq Z < z + dz)\} &= P(X > z) \\ \lim_{dy \rightarrow 0, dz \rightarrow 0} P \{(\Delta = 3) / (y \leq Y < y + dy, z \leq Z < z + dz)\} &= P(X \leq y)\end{aligned}$$

The distribution of the conditional random variable $(\Delta = \delta / Y = y, Z = z)$ can be summarized as :

$$\begin{aligned}P(\Delta = \delta | Y = y, Z = z) &= \lim_{dy \rightarrow 0, dz \rightarrow 0} P(\Delta = \delta / y \leq Y < y + dy, z \leq Z < z + dz) \\ &= P(y \leq X < z)^{\mathbf{1}_{\{\delta=1\}}(\delta)} P(X < y)^{\mathbf{1}_{\{\delta=2\}}(\delta)} P(X \geq z)^{\mathbf{1}_{\{\delta=3\}}(\delta)}\end{aligned}$$

Let $f_{(Y,Z,\Delta)}$ be the density function of the triplet (Y, Z, Δ) defined such that

$$\lim_{dy \rightarrow 0, dz \rightarrow 0} P(y \leq Y < y + dy, z \leq Z < z + dz, \Delta = \delta) = f_{(Y,Z,\Delta)}(y, z, \delta)$$

where

$$\begin{aligned}f_{(Y,Z,\Delta)}(y, z, \delta) &= f_{(Y,Z)}(y, z) P(y \leq X < z)^{\mathbf{1}_{\{\delta=1\}}(\delta)} P(X \leq y)^{\mathbf{1}_{\{\delta=2\}}(\delta)} P(X > z)^{\mathbf{1}_{\{\delta=3\}}(\delta)} \\ &= f_{(Y,Z)}(y, z) (F_X(z) - F_X(y))^{\mathbf{1}_{\{\delta=1\}}(\delta)} F_X(y)^{\mathbf{1}_{\{\delta=2\}}(\delta)} (1 - F_X(z))^{\mathbf{1}_{\{\delta=3\}}(\delta)}\end{aligned}$$

Let $(Y_k, Z_k, \Delta_k)_{k=1, \dots, n}$ be a random sample of the triplet (Y, Z, Δ) whose distribution function depends on the parameter $\boldsymbol{\theta}$. The likelihood function of this sample is

$$\begin{aligned}L(\boldsymbol{\theta}) &= \prod_{k=1}^n f_{(Y,Z,\Delta)}(y_k, z_k, \delta_k) \\ &= \prod_{k=1}^n f_{(Y,Z)}(y_k, z_k) \prod_{k=1}^n U(y_k, z_k, \delta_k; \boldsymbol{\theta})\end{aligned}$$

where

$$U(y_k, z_k, \delta_k; \boldsymbol{\theta}) = (F_X(z_k; \boldsymbol{\theta}) - F_X(y_k; \boldsymbol{\theta}))^{\mathbf{1}_{\{\delta_k=1\}}(\delta_k)} (1 - F_X(z_k; \boldsymbol{\theta}))^{\mathbf{1}_{\{\delta_k=2\}}(\delta_k)} F_X(y_k; \boldsymbol{\theta})^{\mathbf{1}_{\{\delta_k=3\}}(\delta_k)}$$

The maximum likelihood estimator of $\boldsymbol{\theta}$ is obtained by maximizing the log-pseudo-likelihood $\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{k=1}^n \log U(y_k, z_k, \delta_k; \boldsymbol{\theta})$, so the maximum pseudo-likelihood estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta})$$

In the sequel, we are going to prove the asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}$.

5.4.3 Asymptotic normality of the estimator

Let us define the score function D to simplify the notation in the sequel, as

$$D(Y, Z, \Delta; \boldsymbol{\theta}) = \left(\frac{\partial}{\partial \theta_1} \log U(Y, Z, \Delta; \boldsymbol{\theta}), \dots, \frac{\partial}{\partial \theta_d} \log U(Y, Z, \Delta; \boldsymbol{\theta}) \right).$$

Moreover, let us assume that the function $U(y, z, \delta; \boldsymbol{\theta})$ is smooth and twice continuously differentiable with respect to $\boldsymbol{\theta}$ in a neighborhood of the true value for all $(y, z, \delta) \in \mathcal{D}$ (condition C_1).

Lemma 5.4.1. *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the condition C_1 , then*

$$E[D(Y, Z, \Delta; \boldsymbol{\theta})] = 0$$

Démonstration. In fact, for every $i = 1, \dots, d$, we have

$$\begin{aligned} E\left(\frac{\partial}{\partial \theta_i} \log U(Y, Z, \Delta; \boldsymbol{\theta}_i)\right) &= \iint \sum_{\delta \in \{1,2,3\}} \frac{\frac{\partial}{\partial \theta_i} U(y, z, \delta; \boldsymbol{\theta}_i)}{U(y, z, \delta; \boldsymbol{\theta}_i)} U(y, z, \delta; \boldsymbol{\theta}_i) f(y, z) dy dz \\ &= \frac{\partial}{\partial \theta_i} \iint \sum_{\delta \in \{1,2,3\}} f_{Y,Z,\Delta}(y, z, \delta; \boldsymbol{\theta}_i) dy dz = 0 \end{aligned}$$

□

Moreover, let us suppose that the information matrix $\mathcal{I}(\boldsymbol{\theta})$ exist for all $(i, j) \in \{1, \dots, d\}^{\otimes 2}$ (condition C_2).

Lemma 5.4.2. *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the condition C_2 , then*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}))$$

Proof. *Considering the Lemma 5.4.1, it is obvious that the covariance matrix of the vector $D(Y, Z, \Delta; \boldsymbol{\theta})$ coincides with the Fisher information matrix $\mathcal{I}(\boldsymbol{\theta})$, so according to the central limit theorem we have*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}))$$

Finally, let us assume that the function $U(y, z, \delta; \boldsymbol{\theta})$ is continuous for all $(y, z, \delta) \in \mathcal{D}$ and for all $\boldsymbol{\theta} \in \Theta$ (condition C_3), and that the Hessian matrix $J(\boldsymbol{\theta})$ exists and is invertible for all $(i, j) \in \{1, \dots, d\}^{\otimes 2}$ (condition C_4).

Proposition 5.4.1. *If the function $U(Y, Z, \Delta; \boldsymbol{\theta})$ satisfies the conditions C_1, C_2, C_3 and C_4 , then*

$$\sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, \Sigma_{\boldsymbol{\theta}})$$

where

$$\Sigma_{\boldsymbol{\theta}} = J(\boldsymbol{\theta})^{-1} \mathcal{I}(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1}$$

Proof. *The score function $D(Y_k, Z_k, \Delta_k; \hat{\boldsymbol{\theta}}_n)$ equal to zero at $\hat{\boldsymbol{\theta}}$ since $\hat{\boldsymbol{\theta}}$ maximizes $\log U(Y, Z, \Delta; \boldsymbol{\theta})$. Then, using the Taylor-Young formula in the neighborhood of the true value of $\boldsymbol{\theta}$, we can write*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n D(Y_k, Z_k, \Delta_k; \hat{\boldsymbol{\theta}}_n) = \frac{1}{\sqrt{n}} \sum_{k=1}^n D(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}) + J_n(\boldsymbol{\theta}) \sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) + R_n = 0 \quad (5.9)$$

where $J_n(\boldsymbol{\theta})$ is such as

$$J_n(\boldsymbol{\theta}) = \left(J_n^{i,j}(\boldsymbol{\theta}) \right)_{1 \leq i, j \leq d} = \left(\frac{1}{n} \sum_{k=1}^n \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log U(Y_k, Z_k, \Delta_k; \boldsymbol{\theta}) \right)_{1 \leq i, j \leq d}$$

$J_n^{i,j}(\boldsymbol{\theta})$ is the empirical mean of the function $H_{i,j} = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log U(Y_k, Z_k, \Delta_k; \boldsymbol{\theta})$, $1 \leq i, j \leq d$. Then, using the weak law of large numbers we claim that

$$J_n(\boldsymbol{\theta}) \xrightarrow{\mathbb{P}} J(\boldsymbol{\theta}) \quad (5.10)$$

Therefore, we can say that $J_n(\boldsymbol{\theta})$ is almost surely invertible.

Finally, by Lemma 5.4.2 and using the equation (5.9), we can write that

$$\sqrt{n} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, J(\boldsymbol{\theta})^{-1} \mathcal{I}(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1})$$

5.4.4 Analysis of breast cancer data

Statement of the problem

A study (Lindsey and Ryan (1998)) was performed to compare the state of patients with early breast cancer treated by radiotherapy and chemotherapy together with the state of those treated by radiotherapy alone in terms of aesthetic effects. The subjects of the study are patients who were treated at the Joint Center for Radiation Therapy in Boston between 1976 and 1980. The objective of this study is to compare the results obtained with the two treatments in terms of the deterioration of the aesthetic state of the breast of the patients. These patients are examined at clinical visits every 4 to 6 months. Monitoring intervals can be longer depending on the circumstances of each patient. Interval censoring data occur naturally in medical studies requiring periodic monitoring. Indeed, a person under periodic monitoring may miss visits and return in an altered state. The results of this study are presented in Table 1 where the intervals indicate the period of time during which the aesthetic state has deteriorated.

TABLE 5.2 – Breast cancer data with two different treatments

Therapy = RC		Status	Therapy = R		Status
Left	Right		Left	Right	
8	12	1	45	61	2
0	22	3	6	10	1
24	31	1	0	7	3
17	27	1	46	61	2
17	23	1	46	61	2
24	30	1	7	16	1
16	24	1	17	61	2
13	61	2	7	14	1
11	13	1	37	44	1
16	20	1	0	8	3
18	28	1	4	11	1
17	26	1	15	61	2
32	61	2	11	15	1
23	61	2	22	61	2
44	48	1	46	61	2
14	17	1	46	61	2
0	5	3	25	37	1
5	8	1	46	61	2
12	20	1	26	40	1
11	61	2	46	61	2
33	40	1	27	34	1
31	61	2	36	44	1
13	39	1	46	61	2
19	32	1	36	48	1
34	61	2	37	61	2
13	61	2	40	61	2
16	24	1	17	25	1
35	61	2	46	61	2
15	22	1	11	18	1
11	17	1	38	61	2
22	32	1	5	12	1
10	35	1	37	61	2
30	34	1	0	5	3
13	61	2	18	61	2
10	17	1	24	61	2
8	21	1	36	61	2
4	9	1	5	11	1
11	61	2	19	35	1
14	19	1	17	25	1
4	8	1	24	61	2
34	61	2	32	61	2
30	36	1	33	61	2
18	24	1	19	26	1
16	60	1	37	61	2
35	39	1	34	61	2
21	61	2	36	61	2
11	20	1			
48	61	2			

For example, if an observation is coded (6;14), then at 6 months the patient shows no aesthetic deterioration of the breast, but at 14 months the retraction of the breast is

present. The variable of interest X is the duration of the aesthetic comfort which is the time elapsed between the beginning of the treatment and the appearance of breast retraction. A preliminary analysis of the data allows us to assume that the probability distribution of X is of GPD type (Figure 5.2).

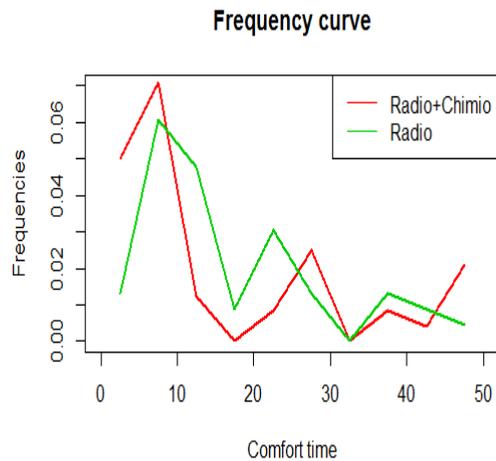


FIGURE 5.2 – Frequency curves of the two empirical distributions of comfort time of the patients treated with chemotherapy and radiotherapy and the estimated comfort time of the patients treated with radiotherapy alone

The esthetic deterioration of the breast is inevitable for a patient with breast cancer. So if the treatment is effective, the comfort time X for a treated patient tends to become longer. As a result, the tail of the X distribution for the treated patients tends to thicken in the case of effective treatment. Thus, to compare the effectiveness of the two types of treatment, we compare the distribution functions of the patients treated with chemotherapy and radiotherapy and radiotherapy alone using their tail indexes.

Estimation of model parameters

Let $\{X_i^{(0)}\}_{1 \leq i \leq n_1}$ be the comfort time of patients who have been treated with chemotherapy after initial radiotherapy and $\{X_j^{(1)}\}_{1 \leq j \leq n_2}$ be the comfort time of the patients who received only an equivalent dose of radiotherapy (Table 5.2), and let $F(x^{(m)}, \gamma_m, \sigma_m)$,

$m = 0$ or 1 , be their distribution functions defined as follows :

$$F(x^{(m)}; \gamma_m, \sigma_m) = 1 - \left(1 + \gamma_m \frac{x^{(m)}}{\sigma_m}\right)^{-\frac{1}{\gamma_m}} \quad \text{where } x^{(m)} > -\frac{\sigma_m}{\gamma_m}$$

The model under consideration is subject to interval censoring. Therefore, the parameters estimation can be carried out using the maximum pseudo-likelihood method (subsection 5.3.2). Let $\tilde{\ell}(\gamma_m, \sigma_m)$ be the pseudo-likelihood function defined such as

$$\begin{aligned} \tilde{\ell}(\gamma_m, \sigma_m) = & \sum_{k=1}^n \left[1_{\{\delta_k=1\}}(\delta_k) \log \left(\left(1 + \frac{\gamma_m}{\sigma_m} y_k\right)^{-\frac{1}{\gamma_m}} - \left(1 + \frac{\gamma_m}{\sigma_m} z_k\right)^{-\frac{1}{\gamma_m}} \right) \right. \\ & \left. - 1_{\{\delta_k=2\}}(\delta_k) \frac{1}{\gamma_m} \log \left(1 + \frac{\gamma_m}{\sigma_m} z_k\right) + 1_{\{\delta_k=3\}}(\delta_k) \log \left\{ 1 - \left(1 + \frac{\gamma_m}{\sigma_m} y_k\right)^{-\frac{1}{\gamma_m}} \right\} \right] \end{aligned}$$

To maximize $\tilde{\ell}(\gamma_m, \sigma_m)$, we use the numerical method of Nelder-Mead (Nelder and Mead (1965)). The results of this optimization are summarized in the following table :

TABLE 5.3 – GPD parameters estimation

Parameters	Estimation	Parameters	Estimation
γ_0	-0.4123641	γ_1	-0.6472023
σ_0	28.1896309	σ_1	73.9606626

The figure below shows the two survival function of GPD distribution with the estimated parameters :

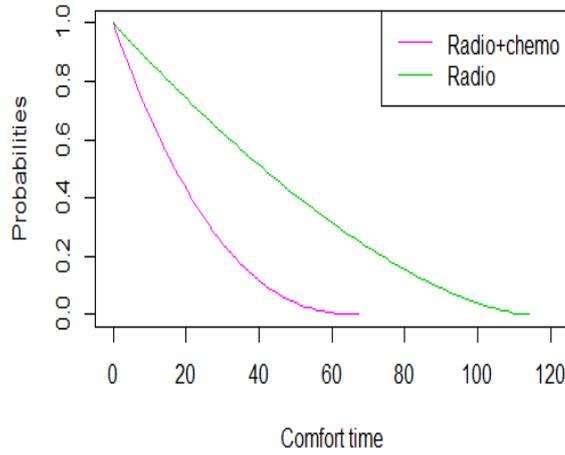


FIGURE 5.3 – Two survival functions of the estimated comfort time of the patients treated with chemotherapy and radiotherapy and the estimated comfort time of the patients treated with radiotherapy alone

Comparison between the two treatments

Figure 5.3 shows that the tail of the estimated comfort time distribution function of the patients treated with chemotherapy and radiotherapy is thinner than the tail of the distribution function of comfort time of the patients treated with radiotherapy alone. In other words, the aesthetic comfort time for the patients treated with chemotherapy and radiotherapy is shorter. That is, the treatment with chemotherapy and radiation does not lengthen the comfort time of the patients.

The comparison of the two treatments can be obtained simply by solving the testing problem of the null hypothesis $H_0 : \gamma = \gamma_0 = \gamma_1$ against the alternative hypothesis $H_1 : \gamma = \gamma_1 < \gamma_0$. There are several methods for determining a decision rule for this test. It is known that Neyman type tests and Wald type tests are equivalent for this kind of hypothesis (Toulemonde (2008)). However, for practical reasons we will proceed as in Wald (Wald (1943)) by standardizing the quantity $\hat{\gamma}_0 - \hat{\gamma}_1$ whose asymptotic variance $V = V_0 + V_1$ is given in the proposition 5.4.1 such that

$$V = \frac{J(\gamma_0)^{-1} \mathcal{I}(\gamma_0) J(\gamma_0)^{-1}}{n_0} + \frac{J(\gamma_1)^{-1} \mathcal{I}(\gamma_1) J(\gamma_1)^{-1}}{n_1} \quad (5.11)$$

Let us define the statistic T such as

$$T = \frac{\widehat{\gamma}_0 - \widehat{\gamma}_1}{\sqrt{\widehat{V}}}$$

where \widehat{V} is obtained by substituting γ_i by $\widehat{\gamma}_i$, $i=0,1$ in the relation (5.11).

Under the null hypothesis the statistic T has a standard normal distribution. Thus, the decision rule can be defined through the rejection region defined as :

$$R_c = \{T > q_{1-\alpha}\}$$

where the normal quantile $q_{1-\alpha} = \phi^{-1}(1 - \alpha)$ and where ϕ is the distribution function of the standard normal distribution.

Moreover, under the hypothesis H_1 , the statistic T is of normal distribution with mean $(\gamma_0 - \gamma_1)$ and variance \widehat{V} . Therefore, the power of this hypothesis testing can be expressed as

$$1 - \beta = P_{H_1}(R_c) = 1 - \phi\left(\frac{\phi^{-1}(1 - \alpha) - (\gamma_0 - \gamma_1)}{\sqrt{\widehat{V}}}\right)$$

With the help of our script written under ([R Development Core Team \(2008b\)](#)), the value of the statistic T and the power $1 - \beta$ of the test are calculated such that

$$T = 2.475412 \quad , \quad q_{1-\alpha} = 1.644854 \quad \text{and} \quad 1 - \beta = 0.7968884$$

as $T > q_{1-\alpha}$, we reject the hypothesis H_0 .

As a conclusion, we can state that the treatment with radiotherapy and chemotherapy together does not improve the aesthetic comfort of patients with breast cancer, but it tends to damage it.

Bibliographie

- Balkema, A. A., De Haan, L., 1974. Residual life time at great age. The Annals of probability, 792–804.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2006a. Statistics of extremes : theory and applications. John Wiley & Sons.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2006b. Statistics of extremes : theory and applications. John Wiley & Sons.
- Beirlant, J., Teugels, J. L., Vynckier, P., 1996a. Practical analysis of extreme values. Leuven University Press.
- Beirlant, J., Vynckier, P., Teugels, J. L., et al., 1996b. Excess functions and estimation of the extreme-value index. Bernoulli 2 (4), 293–318.
- Billingsley, P., 1999. Convergence of probability measures john wiley & sons. INC, New York 2 (2.4).
- Britain), N. E. R. C. G., 1975. Flood Studies Report in Five Volumes. NERC.
- BUND, T. B. O., 2002. Time variation in the tail behaviour of bund futures returns by thomas werner and christian upper.
- Cheng, S., Peng, L., 10 2001. Confidence intervals for the tail index. Bernoulli 7 (5), 751–760.
URL <https://projecteuclid.org:443/euclid.bj/1079399540>

- Clifford Anderson-Bergman, Y. Y., 2016. Computing the log concave npml for interval censored data. *Statistics and Computing* 26 (4), 813–826.
- Coles, S., Bawa, J., Trenner, L., Dorazio, P., 2001. An introduction to statistical modeling of extreme values. Vol. 208. Springer.
- Csörgő, S., Mason, D. M., 1985. Central limit theorems for sums of extreme values. In : *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 98. Cambridge University Press, pp. 547–558.
- Danielsson, J., Jansen, D. W., De vries, C. G., 1996. The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics-Theory and Methods* 25 (4), 711–720.
- Davis, R., Resnick, S., 1984. Tail estimates motivated by extreme value theory. *The Annals of Statistics*, 1467–1487.
- De Haan, L., Ferreira, A., 2007. *Extreme value theory : an introduction*. Springer Science & Business Media.
- de Haan, L., Rootzén, H., 1993. On the estimation of high quantiles. *Journal of Statistical Planning and Inference* 35 (1), 1–13.
- De Haan, L. d., Peng, L., 1998. Comparison of tail index estimators. *Statistica Neerlandica* 52 (1), 60–70.
- deHaan, L. F. M., 1970. On regular variation and its application to the weak convergence of sample extremes.
- Deheuvels, P., Haeusler, E., Mason, D. M., 1988. Almost sure convergence of the hill estimator. In : *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 104. Cambridge University Press, pp. 371–381.
- Dekkers, A. L., De Haan, L., et al., 1989a. On the estimation of the extreme-value index and large quantile estimation. *The Annals of Statistics* 17 (4), 1795–1832.
- Dekkers, A. L., Einmahl, J. H., De Haan, L., 1989b. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 1833–1855.

- Diebolt, J., Guillou, A., Rached, I., 2004. A new look at probability-weighted moments estimators. *Comptes Rendus Mathématique* 338 (8), 629–634.
- Diebolt, J., Guillou, A., Rached, I., 2007. Approximation of the distribution of excesses through a generalized probability-weighted moments method. *Journal of Statistical Planning and Inference* 137 (3), 841–857.
- Drees, H., 1995. Refined pickands estimators of the extreme value index. *The Annals of Statistics*, 2059–2080.
- Drees, H., Kaufmann, E., 1998. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications* 75 (2), 149–172.
- Efron, B., Tibshirani, R. J., 1994. *An introduction to the bootstrap*. CRC press.
- Einmahl, J. H., Fils-Villetard, A., Guillou, A., et al., 2008. Statistics of extremes under random censoring. *Bernoulli* 14 (1), 207–227.
- El Adlouni, S.-E., Bobée, B., Ouarda, T. B., 2007. Caractérisation des distributions à queue lourde pour l’analyse des crues. No. R929. INRS, Centre Eau, Terre et Environnement.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 2013. *Modelling extremal events : for insurance and finance*. Vol. 33. Springer Science & Business Media.
- Fisher, R. A., Tippett, L. H. C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In : *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 24. Cambridge University Press, pp. 180–190.
- Foss, S., Korshunov, D., Zachary, S., et al., 2011. *An introduction to heavy-tailed and subexponential distributions*. Vol. 6. Springer.
- Fréchet, M., 1928. Sur la loi de probabilité de l’écart maximum. In : *Annales de la société Polonaise de Mathématique*. [sn].
- Geluk, J. L., deHaan, L., 1987. *Regular variation, extensions and tauberian theorems*. CWI tracts.

- Gentleman, R., Geyer, C. J., 1994. Maximum likelihood for interval censored data : Consistency and computation. *Biometrika* 81, 618–623.
- Gilli, M., et al., 2006. An application of extreme value theory for measuring financial risk. *Computational Economics* 27 (2-3), 207–228.
- Gnedenko, B., 1943. Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of mathematics*, 423–453.
- Gnedenko, B., Kolmogorov, A., 1954. *Independent Random Variables*. Cambridge, Massachusetts : Addison-Wesley.
- Gomes, M. I., Oliveira, O., 2001. The bootstrap methodology in statistics of extremes—choice of the optimal sample fraction. *Extremes* 4 (4), 331–358.
- Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods : Applications to poisson models. *Econometrica* 52, 701–720.
- Grimshaw, S. D., 1993. Computing maximum likelihood estimates for the generalized pareto distribution. *Technometrics* 35 (2), 185–191.
- Gumbel, E. J., 1935. Les valeurs extrêmes des distributions statistiques. *Ann. Inst. Henri Poincaré* 5 (2), 115–158.
- Gumbel, E. J., 1958. *Statistics of extremes*, columbia univ. Press, New York, 201.
- Gumbel, E. J., 2012. *Statistics of extremes*. Courier Corporation.
- Haeusler, E., Teugels, J. L., 1985. On asymptotic normality of hill’s estimator for the exponent of regular variation. *The Annals of Statistics*, 743–756.
- Hall, P., 1990. Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *Journal of multivariate analysis* 32 (2), 177–203.
- Hall, P., Welsh, A. H., 03 1985. Adaptive estimates of parameters of regular variation. *Ann. Statist.* 13 (1), 331–341.
- URL <https://doi.org/10.1214/aos/1176346596>

- Hill, B. M., 1975. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 1163–1174.
- Hosking, J. R., Wallis, J. R., 1987. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics* 29 (3), 339–349.
- Hubert, P., Bendjoudi, H., 1996. Introduction à l'étude des longues séries pluviométriques. XIIème journées hydrologiques de l'Orstom, 10–11.
- Jenkinson, A. F., 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 81 (348), 158–171.
- Kalbfleisch, J. D., Prentice, R. L., 2002. *The statistical analysis of failure time data*. John Wiley et Sons, New York ; Chichester.
- Kimball, B. F., 1960. On the choice of plotting positions on probability paper. *Journal of the American Statistical Association* 55 (291), 546–560.
- Klein, J. P., Moeschberger, M. L., 1997. *Survival analysis : techniques for censored and truncated data*. Springer-Verlag Inc, Berlin ; New York.
- Lawless, J. F., 2003. *Statistical models and methods for lifetime data*. John Wiley et Sons, New York ; Chichester.
- Lévy, P., 1925. *Calcul des probabilités*. Paris : Gautier-Villars.
- Lin, J., Sinha, D., Lipsitz, S., Polpo, A., 2016. Semiparametric analysis of interval-censored survival data with median regression model. In : *Statistical Applications from Clinical Trials and Personalized Medicine to Finance and Business Analytics*. Springer, pp. 149–163.
- Lindsey, J. C., Ryan, L. M., 1998. Methods for interval-censored data. *Statistics in medicine* 17 (2), 219–238.
- Lindstrom, M. J., Bates, D. M., 1988. Newton—raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* 83 (404), 1014–1022.

- Markovich, N. M., 2005. High quantile estimation for heavy-tailed distributions. *Performance Evaluation* 62 (1-4), 178–192.
- Mason, D. M., 1982. Laws of large numbers for sums of extreme values. *The Annals of Probability*, 754–764.
- Matthys, G., Beirlant, J., 2003. Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica*, 853–880.
- Nelder, J. A., Mead, R., 1965. A simplex method for function minimization. *The computer journal* 7 (4), 308–313.
- Newey, W. K., McFadden, D., 1994. Large sample estimation and hypothesis testing. In : *Handbook of econometrics*, Vol. IV. Vol. 2 of *Handbooks in Econom.* North-Holland, Amsterdam, pp. 2111–2245.
- Paul, E., Claudia, K., Thomas, M., 1997. Modelling extremal events for insurance and finance.
- Peng, L., 1998. Asymptotically unbiased estimators for the extreme-value index. *Statistics & Probability Letters* 38 (2), 107–115.
- Pickands III, J., et al., 1975. Statistical inference using extreme order statistics. *the Annals of Statistics* 3 (1), 119–131.
- R Development Core Team, 2008a. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL <http://www.R-project.org>
- R Development Core Team, 2008b. *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL <http://www.R-project.org>
- Reiss, R.-D., Thomas, M., 2007. *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields*, 3rd Edition. Birkhäuser Verlag, Basel, with 1 CD-ROM (Windows).

- Resnick, S. I., 1987. Extreme values, regular variation, and point processes, volume 4 of applied probability. a series of the applied probability trust.
- Rinne, H., 2008. The Weibull distribution : a handbook. Chapman and Hall/CRC.
- Samson, A., Lavielle, M., Mentré, F., 2006. Extension of the saem algorithm to left-censored data in nonlinear mixed-effects model : Application to hiv dynamics model. *Computational Statistics & Data Analysis* 51 (3), 1562–1574.
- Smith, R. L., 1987. Estimating tails of probability distributions. *The annals of Statistics*, 1174–1207.
- Thas, O., 2010. Comparing distributions. Springer.
- Thomas, M., Lemaitre, M., Wilson, M. L., Viboud, C., Yordanov, Y., Wackernagel, H., Carrat, F., 2016. Applications of extreme value theory in public health. *PloS one* 11 (7), e0159312.
- Toulemonde, G., 2008. Estimation et tests en théorie des valeurs extrêmes. Ph.D. thesis, Université Pierre et Marie Curie-Paris VI.
- Turnbull, B. W., 1974. Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association* 69, 169–173.
- Van Der Vaart, A. W., 1998. Asymptotic statistics. Vol. 3 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.
URL <http://dx.doi.org/10.1017/CB09780511802256>
- Varadhan, e. a., 2009. Bb : An r package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of statistical software* 32 (4), 1–26.
- Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society* 54 (3), 426–482.
- Weihull, W., 1951. A statistical distribution function of wide applicability. *J Appl Mech* 18, 290–293.

Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association* 73 (364), 812–815.