

UNIVERSITE SAAD DAHLAB DE BLIDA 1

Faculté des Sciences

Département de Mathématiques

THESE DE DOCTORAT

en Mathématiques

**LES MESURES DE RISQUES ET LES
DISTRIBUTIONS À QUEUES LOURDES**

Par

Mohamed LAIDI

Devant le jury composé de :

Mustapha CHELLALI	Président	Prof.	Univ.Blida 1
Abdelaziz RASSOUL	Directeur de thèse	Prof.	ENSH, Blida
Hamid OULD ROUIS	Co-Directeur	MCA	Univ. Blida 1
Nadia OUKID	Examinatrice	Prof.	Univ.Blida 1
Ghania SAIDI	Examinatrice	Prof.	ENSSEA Koléa
Diffalah LAISSAOUI	Examineur	MCA	Univ. Médéa

Blida, 30 Mars 2021

Dédicace

A la mémoire de mon cher et regretté père à qui je dois persévérance et volonté, acharnement et fierté,

à ma tendre maman qui m'a toujours aidé par sa présence, ses conseils et ses prières,

à ma chère épouse pour ses encouragements, sa patience et sa compréhension,

à mes adorables enfants Bouchra, Abderrahmane et Anes,

à ma sœur Zineb et son mari Taib qui ont tous été là pour moi ainsi que leurs enfants,

à tous mes amis et mes professeurs de la primaire jusqu'à ce jour,

je dédie mon modeste travail qui, j'espère, trouvera bonne réception, fera honneur et donnera plaisir à tous ceux qui le lisent, l'écoutent ou le discutent.

Remerciements

Il me sera très difficile de remercier tout le monde car c'est grâce à l'aide de nombreuses personnes que j'ai pu mener cette thèse à son terme. Je souhaite remercier en premier lieu mon directeur de thèse, M. RASSOUL Abdelaziz, Professeur des Universités à l'Ecole Nationale Supérieure de l'Hydraulique pour m'avoir accueilli au sein de son équipe. Je lui suis également reconnaissant pour le temps conséquent qu'il m'a accordé, ses qualités pédagogiques et scientifiques, sa franchise et sa sympathie. J'ai beaucoup appris à ses côtés et je lui adresse ma gratitude pour tout cela.

J'adresse de chaleureux remerciements à mon co-directeur de thèse, M. OULD ROUIS Hamid, Maître de conférences à l'Université de Saad Dahleb Blida 1, pour son attention de tout instant sur mes travaux, pour ses conseils avisés et son écoute qui ont été prépondérants pour la bonne réussite de cette thèse. Son énergie et sa confiance ont été des éléments moteurs pour moi. J'ai pris un grand plaisir à travailler avec lui.

Je remercie également M. CHELLALI. M, Professeur des Universités à l'Université de Saad Dahleb Blida 1, pour avoir bien voulu chapeauter le jury de ma thèse.

Mes remerciements s'adressent aussi
à Mme OUKID. N, Professeur des Universités à l'Université de Saad Dahleb Blida 1,
à Mme SAIDI. Gh, Professeur des Universités à ENSSEA de Koléa
ainsi qu'à M. LAISSAOUI. D, Maître de conférences à l'Université de Médéa,
qui ont aimablement accepté de juger ce travail.

ملخص

يقدم هذا العمل دراسة لمقاييس المخاطر في حالة التوزيعات الثقيلة الذيل للقيم المتطرفة. يقترح تحسيناً لمقدر كلاسيكي لأحد مقاييس المخاطر المسماة الأمل الشرطي للذيل (CTE) وذلك عن طريق تقليل التحيز. تم من خلال هذه الأطروحة إجراء دراسة للسلوك المقارب، ولا سيما التقارب الطبيعي لهذا المقدّر الجديد، بالإضافة إلى المقارنة مع المقدّر الكلاسيكي، مما يثبت تحسناً واضحاً في جودة التقدير. كلمات مفتاحية: مقياس خطر، ذيل التوزيع، الأمل الشرطي، القيمة المتطرفة، المقدر، السلوك المقارب.

Résumé

Ce travail présente une étude des mesures du risque dans le cas des distributions à queues lourdes des valeurs extrêmes.

Il propose une amélioration à un estimateur classique de l'une des mesures de risque appelée Espérance Conditionnelle de queue (CTE) en réduisant le biais.

Dans cette thèse, une étude du comportement asymptotique, en particulier de la convergence normale de ce nouvel estimateur, a été menée, en plus d'une comparaison avec l'estimateur classique, ce qui prouve une nette amélioration de la qualité de l'estimation.

Mots clés : mesure du risque, queue de distribution, espérance conditionnelle, valeur extrême, estimateur, comportement asymptotique.

Abstract

This work presents a study of risk measures in the case of heavy tailed extremes values distributions. It proposes an improvement to a classical estimator of one of the risk measures called Conditional Tail Expectation (CTE) by reducing the bias. In this thesis, a study of the asymptotic behavior, in particular of the normal convergence of this new estimator, was carried out, in addition to a comparison with the classical estimator, which proves a clear improvement of the quality of the estimate.

Keswords : risk measure, distribution's tail, conditional expectation, extreme value, estimator, Asymptotic behavior.

Notations

va	: Variable aléatoire.
f	: Densité de probabilité d'une va .
F	: Fonction de répartition (fdr).
μ	: Espérance ou moyenne d'une va .
$\bar{F}(t)$: Fonction de survie.
F_n	: Fonction de répartition empirique.
va	: Variable aléatoire.
TCL	: Théorème Centrale Limite.
TVE	: Théorie des valeurs extrêmes.
iid	: Indépendantes et identiquement distribuées.
(Ω, F, \mathbb{P})	: Espace probabilisé.
\mathbb{R}	: Ensemble des valeurs réelles.
σ^2	: Variance d'une variable aléatoire.
S_n	: Somme arithmétique.
\bar{X}	: Moyenne arithmétique.
\xrightarrow{p}	: Converge en probabilité.
\xrightarrow{L}	: Converge en loi.
$\xrightarrow{p.s.}$: Convergence presque sûre.
Q	: Fonction de quantile.
Q_n	: Quantile empirique.
$p.s$: Presque sûre.
$\mathcal{N}(0, 1)$: Loi normale standard.
$X_{n:n}$: Maximum de X_1, \dots, X_n .
$X_{1:n}$: Minimum de X_1, \dots, X_n .
$X_{k:n}$: $k^{\text{ème}}$ statistique d'ordre.
X_1, \dots, X_n	: Echantillon de taille n de X .
x_F	: Point terminal.
$sup(A)$: Sup de l'ensemble A .
t_s	: Quantile d'ordre s .

u	: Seuil.
$\mathcal{R}v_\rho$: Ensemble des fonctions à variation régulière au voisinage de ∞ d'indice ρ .
GEV	: Distribution de valeurs extrêmes généralisée.
GPD	: Distribution de Paréto généralisée.
POT	: Peak Over Threshold.
$\mathcal{D}(\cdot)$: Domaine d'attraction du maximum.
\mathcal{H}_γ	: Famille de la lois de valeurs extrêmes généralisée.
ℓ	: Fonction à variation lente.
$\mathbf{1}\{A\}$: La fonction indicatrice de l'ensemble A .
MSE	: L'erreur moyenne quadratique.
BM	: Block maxima.

Table des matières

Notations	5
Introduction générale	13
I Valeurs Extrêmes	17
1 Valeurs extrêmes univariées	18
1.1 Comportement asymptotique des plus grandes valeurs d'un échantillon . . .	18
1.1.1 Comportement asymptotique du maximum d'un échantillon	19
1.1.2 Méthode des maxima par blocs (BM)	29
1.1.3 Loi des excès	30
1.1.4 Lien entre GEV et loi des excès	32
2 Caractérisation des lois extrêmes	34
2.1 Classes de distributions à queue lourde	35
2.1.1 Classification sans comparaison à la loi normale	35
2.1.2 Classification avec comparaison à la loi normale	38
2.2 Caractérisation des lois extrêmes	49
2.2.1 Domaine d'attraction de Fréchet	49
2.2.2 Domaine d'attraction de Weibull	49
2.2.3 Domaine d'attraction de Gumbel	50
2.2.4 Caractérisation générale	50
3 Estimation d'un quantile et indice de queue :	54
3.1 Estimation d'un quantile : cas classique	55
3.2 Estimation d'un quantile : cas extrême	56
3.2.1 Approximation d'un quantile extrême	57
3.2.2 Estimation d'un quantile extrême	59
3.3 Etude pratique : maximas annuelles des températures	72

II	Mesures de Risques	74
4	Mesures de risques	75
4.1	Première mesure de risque	76
4.2	Approche axiomatique des mesures de risque et mesures de risque usuelles	76
4.2.1	Liens entre mesures de risque.	82
4.2.2	Mesures de risque de Wang	82
4.2.3	Mesures de risque par distorsion	83
4.3	Estimation des mesures de risque définies par distorsion	84
4.3.1	Estimation de la Value-at-Risk	85
4.4	Une nouvelle mesure de risque : le Conditional Tail Moment	87
III	Travaux de recherche	89
5	Improved Estimator of the Conditional Tail Expectation in the case of heavy-tailed losses	90
5.1	Introduction and Motivation	90
5.2	Main result and its practical implementation	94
5.3	Simulation study	95
5.4	Conclusion	99
5.5	Proof of Theorem 5.1	100
5.6	Confidence interval of improved Estimator	102
	Conclusion et Perspectives	104
IV	Annexes	105
A	Définitions et propriétés indispensables	106
A.1	Convergence de variables aléatoires	106
A.1.1	Modes de convergences	106
A.1.2	Liens entre modes de convergences	108
A.1.3	Remplacement de la convergence en loi par la convergence presque sûre	108
A.1.4	Lois fondamentales de la statistique	109
B	Statistique d'ordre	110
B.1	Fonction de répartition empirique et statistiques d'ordre	110

B.1.1	Fonction de répartition et densité du maximum	111
B.1.2	Point extrême	111
B.1.3	Fonction quantile	112
B.1.4	Distributions exactes des statistiques d'ordre	114
C	Propriétés statistiques des lois limites des valeurs extrêmes	115

Table des figures

1	Vue aérienne d'Oude Tonge, au nord de la Zélande, le 3 février 1953, après le raz-de-marée qui a causé 2 500 morts, 47 000 habitations inondées et 10 000 détruites, 200 000 hectares de terres inondées, 30 000 têtes de bétail noyées. Source : Wikipédia.	13
1.1	Densités (à gauche) et répartitions (à droite) G_γ des valeurs extrêmes.	24
1.2	Simulation de 1000 loi exponentielles standard en gris. En haut de la figure, les points rouge sont les valeurs extrême, les points blues sont les moyennes empiriques. Au milieu : la distributions de la moyenne empirique. En bas : la distribution des valeurs extrêmes.	26
1.3	Simulation de 1000 loi de Pareto de paramètres $alpha = 4$ et d'échelle 1 en gris. En haut de la figure, les points rouge sont les valeurs extrême, les points blues sont les moyennes empiriques. Au milieu : la distributions de la moyenne empirique. En bas : la distribution des valeurs extrêmes.	27
1.4	Simulation de 300 loi Uniformes standards en gris. En haut de la figure, les points rouge sont les valeurs extrême, les points blues sont les moyennes empiriques. Au milieu : la distributions de la moyenne empirique. En bas : la distribution des valeurs extrêmes.	28
1.5	Comportement du maximum par blocs de taille 100 de 10000 v.a Normales standard	29
1.6	Les excès Y_1, Y_2, \dots de X au delà du seuil u	30
1.7	La méthode POT pour une échantillon de 10000 lois normales standards avec un seuil égal à 1.	31
2.1	Différentes classes de distributions de queue de très légère (E) à très lourde (A).	39
2.2	Comparaison du comportement des queues : distribution exponentielle (trait plein), distribution de Pareto (trait pointillé).	45

3.1	Convergence en loi du quantile d'une exponentielle standard vers une loi normale en générant 1000 échantillons de tailles 10000. A gauche de la figure : la courbe de la distribution empirique et de celle de la théorique. A droite : l'absolu erreur entre la distribution empirique et la théorique . . .	56
3.2	Comprtement de l'estimateur de Hill en fonction de k_n	68
3.3	les maximas annuelles des températures en °F de la ville de New York de l'année 1938 à l'année 2017.	72
3.4	Etude des valeurs extrêmes des températures maximales annuelles en degré Fahrenheit (°F) de la ville de New York de l'année 1938 à l'année 2017 par approche \mathcal{GEV}	73
4.1	Comparaison entre la VaR de la somme de deux variables aléatoires et la somme de chacune de leur VaR.	80
5.1	Behavior of the new estimator \bar{C}_n and old estimator \tilde{C}_n of <i>CTE</i> in respect to the variation of k for Fréchet distribution.	98
5.2	Behavior of the new estimator \bar{C}_n and old estimator \tilde{C}_n of <i>CTE</i> in respect to the variation of k for Burr distribution.	99
5.3	95%-Confidence interval of improved Estimator $\bar{C}_n(t)$ in respect of sample size for Fréchet distribution with $\alpha = 1.5$	103

Liste des tableaux

1.1	Lois usuelles et leur domaine d'attracion.	25
4.1	Principales mesures de risque par distorsion (où α est une constante comprise entre 0 et 1).	84
4.2	Lien entre le CTM et certaines mesures de risque lorsque la répartition de X est continue.	88
5.1	Comparison between the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in terms of bias and RMSE respecting to the variation of sample size n based on Fréchet distribution.	96
5.2	Comparison between the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in terms of bias and RMSE respecting to the variation of sample size n based on Fréchet distribution.	97
5.3	Comparison between the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in terms of bias and RMSE respecting to the variation of sample size n based on Burr distribution.	97
5.4	Comparison between the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in terms of bias and RMSE respecting to the variation of sample size n based on Burr distribution.	98
C.1	Propriétés statistiques des lois limites des valeurs extrêmes	116

Introduction générale

”La loi des grands nombres et la distribution gaussienne, fondements de l’étude statistique des grandeurs moyennes, échouent à rendre compte des événements rares ou extrêmes. Pour ce faire, des outils statistiques plus adaptés existent . . . mais ne sont pas toujours utilisés !”

Rama Cont ¹

Les événements extrêmes s’intensifient d’une année à une autre. Parmi ces événements, nous citons les inondations, les fortes tremblements de terre, les ouragans, les cyclones, les krachs boursiers etc. Ces événements extrêmes causent souvent des dégâts matériels et humaines.



FIGURE 1 – Vue aérienne d’Oude Tonge, au nord de la Zélande, le 3 février 1953, après le raz-de-marée qui a causé 2 500 morts, 47 000 habitations inondées et 10 000 détruites, 200 000 hectares de terres inondées, 30 000 têtes de bétail noyées. Source : Wikipédia.

Afin d’éviter ce genre de catastrophes, il est primordiale de faire une prévention de leurs risques. Un des outils utilisés pour se protéger ou on va dire pour être en mesure de minimiser l’impact de ces risques est la théorie des valeurs extrêmes (TVE). Cette théorie

1. Rama Cont est professeur de finance mathématique à l’ Université d’Oxford .

s'intéresse aux événements rares. La rareté d'un événement vient de sa très faible probabilité de réalisation. Malgré cette rareté, Les conséquences des événements rares sont souvent désastreuses et peuvent causer beaucoup de pertes dans différents domaines : finance, assurance, santé, écologie etc. Cela incite à modéliser et à prévoir l'occurrence de ces phénomènes. Cette incitation a fait que les scientifiques ont donné beaucoup d'importance à cette théorie qui connaît une croissance de plus en plus forte du nombre de travaux aussi bien sur l'aspect théorique que sur l'aspect pratique.

En fait, la TVE complète la théorie statistique classique qui utilise la tendance centrale au lieu de celle des extrêmes. La tendance centrale argumentée par le fameux Théorème de la Limite Centrale (TCL) nécessite un nombre assez important de réalisations du phénomène étudié pour être applicable, chose qui n'ai guère satisfaite dans le cas des événements rares. Comme on l'a déjà mentionné, un événement rare est de faible probabilité, donc en pratique si on veut étudier un phénomène, on tombera sur des échantillons de faibles tailles et même dans certain cas sur des échantillons de taille nulle, aucune donnée !

L'ingéniosité de la TVE est d'utiliser les valeurs fréquentes, ça veut dire qui se trouve en abondance, comme une base de données pour faire une extrapolation en dehors du domaine central vers le domaine extrême où les données sont rarissimes. La TVE a trouvé place dans différents domaines pour anticiper les risques financiers, actuariels, etc. et dans divers secteurs où la plus grande et la plus petite valeur d'un échantillon sont plus parlantes que ses valeurs centrales et décrivent mieux le comportement de ces événements.

Parmi ces secteurs, nous citons :

- L'industrie où l'on s'intéresse à la fiabilité des machines, des produits, des structures et même de celle des travailleurs.
- La finance où la prédiction de crises monétaires (risque sur les marchés) et la gestion de portefeuille sont devenus les premiers soucis des économistes et les actuaires.
- La météorologie où il faut prévoir les tsunamis, les cyclones , les ouragans et les inondations.
- L'hydrologie où l'on s'intéresse à la prévision des grandes crues pour la conception des aménagements des cours d'eau afin de protéger les zones urbaines inondables et même parfois les champs agricoles.

Cette théorie à vue le jour avec les travaux de Fréchet [36], Fisher et Tippett [35], qui ont montré que sous certaines conditions, les seules distributions limites des extrêmes sont les lois de Fréchet, Gumbel et Weibull, constituant ainsi trois domaines d'attraction où chaque domaine porte le nom de la distribution qui le caractérise (voir Embrechts et al. [34], De Haan et Ferreira [27]). Une généralisation par paramétrisation des trois lois limites des extrêmes en une distribution unique donne naissance à ce qu'on appelle : Generalized

Extreme Value distribution (GEV). Cette généralisation est due à Von Mises [73]. L'indice de queue ou indice des valeurs extrêmes est l'un des paramètres les plus importants étudié dans la TVE car il décrit la lourdeur de la queue de la distribution. Dans la littérature, plusieurs estimateurs de cet indice sont proposés et la plupart sont basés sur la statistique d'ordre pour assurer un meilleur ajustement à la queue de la distribution car seules les données extrêmes sont utilisées ; nous citons Smith [69, 70], Pickands [62], Hill [45] ou encore Dekkers et al. [31]. Un événement extrême est souvent en étroite relation avec un risque. Ce risque doit d'être quantifier pour pouvoir estimer son impact et sa gravité. Une telle quantification est dite mesure de risque. La première mesure de risques est la variance mais la plus connue est la Value-at-Risk (VaR) qui est introduite dans les années 90 par Morgan [57]. Cette dernière mesure de risque souffre d'une faiblesse majeure qui est d'être non cohérente (voir [8]).

“Value at Risk is to finance what body temperature is to a patient, an indication of bad health but not an instrument telling us what is wrong and for less a clue on how to get the patient system healthy again”.

Paul Embrechts

Cette faiblesse diminuant l'efficacité de cette mesure de risque a poussé les chercheurs à définir de nouvelles mesures de risque. Nous citons : Tail-Value-at-Risk, Conditional Tail Expectation sur laquelle nous avons une contribution dans son estimation (voir [51]), Conditional-VaR, Expected shortfall, etc.

L'objectif de cette thèse, est d'étudier les mesures de risque dans le cas des distributions à queues lourdes.

Organisation de la thèse

Cette thèse s'organise en quatre parties.

- ▶ Partie I s'intéresse à la théorie des valeurs extrêmes univariées. Elle commence par le comportement asymptotique des plus grandes valeurs d'un échantillon, puis la classification des distribution selon leur lourdeur de la queue ainsi que la caractérisation des différents domaines d'attraction, et enfin l'estimation d'un quantile et l'indice de queue en passant du cas classique au cas extrême
- ▶ Partie II présente la première définition d'une mesure de risque ainsi que les définitions axiomatique des mesures de risque usuelles et la mesure de risque génératrice dite mesure par distorsion. elle donne les différentes méthodes d'estimations de la mesure la plus

fréquente qui est la Value-at-Risk (VeR) et puis un cadre plus générale qui est celui des mesures définies par distorsion.

- Partie III est le cœur de cette thèse. Elle présente les travaux de recherche menés pendant les quatre années. Elle commence par l'article publié dont l'intitulé est : Improved Estimator of the Conditional Tail Expectation in the case of heavy-tailed losses. Cet article est une amélioration par réduction de biais de l'estimateur classique introduit par Lie et al. (2010) [54] de la mesure de risque Conditional tail expectation (CTE) dans le cas de fonction de distribution de perte à queue lourde. Cette amélioration a donné naissance à un nouveau estimateur dont sa normalité asymptotique est prouvée théoriquement et par simulation.

Ce premier travail est suivi par une conférence internationale qui entre dans la continuité du premier travail et le complète. C'est une conférence qui fait l'objet la construction d'un intervalle de confiance à notre nouveau estimateur en se basant au fait qu'on montré sa normalité asymptotique.

Une deuxième conférence internationale qui présente des résultats préliminaires sur l'estimation sous données censurées d'une nouvelle mesure de risque dite Conditional Tail Moment (CTM) dont elle a été définie par El Methni et al. [56] et qui représente une généralisation de la CTE. Comme Ce travail reste inachevé nous allons citer cette mesure de risque sans entrer dans les détails.

- Partie IV cumule les différentes définitions nécessaires qu'il faut savoir pour être à l'aise avec la TVE et les mesures de risque.

Première partie
Valeurs Extrêmes

Chapitre 1

Introduction à la théorie des valeurs extrêmes univariées

Sommaire

1.1 Comportement asymptotique des plus grandes valeurs d'un échantillon	18
1.1.1 Comportement asymptotique du maximum d'un échantillon	19
1.1.2 Méthode des maxima par blocs (BM)	29
1.1.3 Loi des excès	30
1.1.4 Lien entre GEV et loi des excès	32

Résumé

Ce chapitre a comme objectif l'introduction à la théorie des valeurs extrêmes à une dimension. Il est inspiré des ouvrages [11, 27, 63, 24]. Il se focalise au comportement de(s) plus grande(s) valeur(s) d'un grand échantillon de variables aléatoires (taille de l'échantillon tendant vers l'infini) ou d'une autre manière c'est une étude asymptotique des grandes valeurs et en particulier le maximum. Cette étude donne les différentes lois limites possibles dites domaines d'attraction.

1.1 Comportement asymptotique des plus grandes valeurs d'un échantillon

Afin de voir la nécessité d'étudier les valeurs extrêmes d'un échantillon, examinons l'exemple suivant qui représente les températures ambiantes en degré Celsius (°C) dans un certain endroit :

$$-48, 5, 7, -43, 127, -65, 166, 51, 115, -67.$$

En se basant sur des approches statistiques classiques, on doit calculer la moyenne de ces températures qui est dans notre cas 24.8 °C. Cette température moyenne est agréable à vivre mais est-ce le même constat pour les températures des données et en particulier les plus petites et les plus grandes? la réponse est bien sûr non, ce qui est contraire à celle donnée en calculant la moyenne. Pour ne pas tomber dans ce genre de situation, on doit donc étudier les valeurs les plus petites et les plus grandes de l'échantillon, ces valeurs sont dites valeurs extrêmes dont toute une théorie est y consacrée : théorie des valeurs extrêmes. La théorie des valeurs extrêmes s'intéresse à répondre aux deux questions suivantes :

Question 1 : Quelle est la probabilité p d'événement rare supérieur (ou inférieur) aux événements enregistrés précédemment ?

Question 2 : Quelle est l'amplitude x de l'événement qui est rarement dépassée par les événements enregistrés ?

La première question se soucie du problème de calcul d'une probabilité proche de zéro, associée à un événement extrême. La deuxième question cherche à quantifier la valeur appelée quantile d'un événement extrême, c'est à dire d'un événement dont la probabilité de survenir est faible par définition. Les premiers qui ont travaillé et essayé de répondre à ces questions sont FISHER et TIPPETT [1928][35] et GNEDENKO [1943][38] dont ils ont pu donner les résultats fondamentaux de la théorie sur la convergence en loi de la valeur maximale d'une suite de variables aléatoires indépendantes et identiquement distribuées en suivant leurs intuition que l'information pertinente se trouve dans les extrémités d'une série statistique, ces résultats sont l'objet du paragraphe 1.1.1. Leurs résultats sont complétés par le travail de PICKANDS [1975][62] sur la convergence en loi des excès au-dessus d'un seuil dont il est traité au paragraphe 1.1.3.

1.1.1 Comportement asymptotique du maximum d'un échantillon

Soit X_1, X_2, \dots, X_n un échantillon de n variables aléatoires (v.a) indépendantes de même loi que X (on dira que c'est une suite de variables aléatoires indépendantes et identiquement distribuées (iid)) où X est une variable aléatoire continue de fonction de répartition F . La fonction de survie \bar{F} associée est donnée par

$$\bar{F}(x) = \mathbb{P}(X > x) = 1 - F(x).$$

On va s'intéresser au comportement asymptotique du maximum de l'échantillon qu'on va le noter :

$$X_{n,n} = \max(X_1, X_2, \dots, X_n) = \max_{i=1, \dots, n} X_i$$

dont sa répartition est notée $F_{X_{n,n}}$ ¹. C'est à dire le comportement du maximum quand la taille de l'échantillon n tend vers l'infini. Comme les v.a sont iid alors on peut calculer $F_{X_{n,n}}$ sans difficulté en se basant sur les notions de base de probabilité.

$$\begin{aligned}
 F_{X_{n,n}}(x) &= \mathbb{P}(X_{n,n} \leq x) \\
 &= \mathbb{P}(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\
 &= \prod_{i=1}^n \mathbb{P}(X_i \leq x) \\
 &= \mathbb{P}^n(X \leq x) \\
 &= F^n(x).
 \end{aligned} \tag{1.1}$$

En passant à la limite quand n tend vers l'infini et en prenant en compte que par définition

$$\begin{cases} 0 \leq F(x) < 1 & \text{si } x < x_F \\ F(x) = 1 & \text{si } x \geq x_F \end{cases}$$

où

$$x_F = \sup\{x \in \mathbb{R}, F(x) < 1\}$$

$$\begin{aligned}
 \lim_{n \rightarrow +\infty} F_{X_{n,n}}(x) &= \lim_{n \rightarrow +\infty} F^n(x) \\
 &= \begin{cases} 0 & \text{si } x < x_F \\ 1 & \text{si } x \geq x_F \end{cases}
 \end{aligned} \tag{1.2}$$

L'équation 1.2 ne convient pas à répondre car elle a abouti à une loi dégénérée² du maximum et on doit donc remédier à ce problème. L'idée est de faire une renormalisation du maximum qui est similaire à la transformation linéaire dans le Théorème de la Limite Centrale (TCL) dont l'énoncé est :

Théorème 1.1 (TCL)

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F d'espérance μ et d'écart-type σ qui existent avec $\sigma \neq 0$. Soit S_n la somme $S_n = X_1 + X_2 + \dots + X_n$. Alors

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{en loi}} N(0, 1), \quad n \rightarrow +\infty \tag{1.3}$$

1. Cette notation fait appel à la notion de la statistique d'ordre (Voir l'annexe B).

2. On dit que la va X est dégénérée (ou la loi de X est dégénérée) s'il existe un réel a tel que

$$\mathbb{P}(X = a) = 1.$$

Cette normalisation du maximum a été donnée par GNEDENKO[1943] [38] dont l'énoncé du Théorème :

Théorème 1.2

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F . S'il existe deux suites réelles normalisantes $(a_n)_{n \geq 1} > 0$, $(b_n)_{n \geq 1}$ et une loi non-dégénérée G_γ telles que :

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{X_{n,n} - b_n}{a_n} \leq x \right) = \lim_{n \rightarrow +\infty} F^n(a_n x + b_n) = G_\gamma(x),$$

alors

$$G_\gamma(x) = \begin{cases} \exp \left(-(1 + \gamma x)_+^{-1/\gamma} \right) & \text{si } \gamma \neq 0 \\ \exp(-\exp(-x)) & \text{si } \gamma = 0 \end{cases} \quad (1.4)$$

où $\gamma \in \mathbb{R}$ et $y_+ = \max(0, y)$.

Preuve

Supposons qu'il existe une suite $((a_n, b_n), n \geq 1)$ telle que $a_n > 0$ et la suite de terme $a_n^{-1}(M_n - b_n)$ converge en loi vers une limite W non constante. D'après le Lemme de Portmanteau (Voir A.1), on sait que pour toute fonction g continue bornée, on a

$$\lim_{n \rightarrow \infty} \mathbb{E} [g(a_n^{-1}(M_n - b_n))] = \mathbb{E} [g(W)].$$

Supposons que la v.a X_1 possède une densité de probabilité $f > 0$. Alors la loi de M_n à comme densité $nF^{n-1}(x)f(x)$. On a donc

$$I_n = \mathbb{E} [g(a_n^{-1}(M_n - b_n))] = \int_{\mathbb{R}} g \left(\frac{x - b_n}{a_n} \right) nF^{n-1}(x)f(x)dx.$$

Le fait que $f > 0$ nous permet de confirmer que F est inversible et d'inverse continue. On pose pour $t > 1$

$$U(t) = F^{-1} \left(1 - \frac{1}{t} \right).$$

On peut vérifier que

$$U(t) = x \Leftrightarrow 1 - \frac{1}{t} = F(x) \Leftrightarrow \mathbb{P}(X > x) = \frac{1}{t}.$$

On fait le changement de variable $x = U(\frac{n}{v})$. On obtient

$$I_n = \int_{\mathbb{R}} g \left(\frac{U(n/v) - b_n}{a_n} \right) \left(1 - \frac{v}{n} \right)^{n-1} \mathbb{1}_{]0, n]}(v)dv.$$

Il est à noter que

$$\lim_{n \rightarrow \infty} \left(1 - \frac{v}{n} \right)^{n-1} \mathbb{1}_{]0, n]}(v) = \exp(-v) \mathbb{1}_{\{v > 0\}}.$$

Supposons que la suite $\frac{U(n/v) - b_n}{a_n}$ converge pour tout $v > 0$. Cette dernière hypoythèse conduit à prouver que pour tout $w > 0$,

$$\frac{U(wx) - U(x)}{a(x)} \xrightarrow{x \rightarrow \infty} h(w),$$

où h est une fonction croissante non constante, $a(x) = a_{[x]}$ pour $x \geq 1$, et $[x]$ est la partie entière de x . En suivant une démarche prise des méthodes de l'analyse fonctionnelle, on aboutit à

$$l(w'w) = l(w)l(w'),$$

où l est une fonction strictement positive mesurable localement bornée. Les solutions non nulles de cette équation fonctionnelle sont : $l(w) = w^\gamma, \gamma \in \mathbb{R}$. En faisant une discussion sur γ et en prenant $U(n) = b_n$ on aura

$$\lim_{n \rightarrow \infty} \frac{U(n/v) - b_n}{a_n} = h\left(\frac{1}{v}\right) = \begin{cases} \frac{v^{-\gamma} - 1}{\gamma} & \text{si } \gamma \neq 0 \\ -\log(v) & \text{si } \gamma = 0. \end{cases}$$

Donc, en posant

$$y = \begin{cases} \frac{v^{-\gamma} - 1}{\gamma} & \text{si } \gamma \neq 0 \\ \log(v) & \text{si } \gamma = 0. \end{cases}$$

par convergence dominée, on aura finalement

$$\begin{aligned} \lim_{n \rightarrow \infty} I_n &= \int g\left(\frac{v^{-\gamma} - 1}{\gamma}\right) e^{-\gamma} \mathbf{1}_{\{v > 0\}} dv \\ &= \int g(y) \mathbf{1}_{\{1 + \gamma y > 0\}} d\left(e^{-(1 + \gamma y)^{-1/\gamma}}\right) \end{aligned}$$

Ce qui veut dire que la loi G_γ est bien la fonction de répartition de la limite.

Ce Théorème donne la loi Généralisée des Valeurs Extrêmes (\mathcal{GEV}) qui a une importance majeure dans la théorie des valeurs extrêmes semblable à celle de TCL dans la statistique classique. Le mot généralisée vient du fait qu'il existe une version moins compacte et qui est en fait la version originale de ce Théorème. La version citée auparavant est dite de **von Mises** dont la forme originale est la suivante :

Théorème 1.3 (Théorème de Fisher-Tipett-Gnedenko)

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F . S'il existe deux suites réelles normalisantes $(a_n)_{n \geq 1} > 0$ $(b_n)_{n \geq 1}$ et une loi non-dégénérée G_γ telles que :

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow +\infty} F^n(a_n x + b_n) = G_\gamma(x),$$

alors à une translation et un changement d'échelle près, la fonction de répartition de la limite est du type des trois classes suivantes :

Loi de GUMBEL : $\Lambda(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$ et $\gamma = 0$

Loi de FRÉCHET : $\Phi_\gamma(x) = \begin{cases} \exp(-x^{-\gamma}) & , x \geq 0 \\ 0 & , x < 0 \end{cases}$ et $\gamma > 0$

Loi de WEIBULL : $\Psi_\gamma(x) = \begin{cases} \exp(-(-x^{-\gamma})) & , x \leq 0 \\ 0 & , x > 0 \end{cases}$ et $\gamma < 0$

Remarque 1.1

On peut remarqué que dans les deux Théorèmes TCL et GEV qu'il y a une correspondance dans les écritures

$$\begin{array}{ccc} b_n & \xleftrightarrow{\text{joue le rôle du paramètre d'échelle}} & n\mu \\ a_n & \xleftrightarrow{\text{joue le rôle du paramètre de position}} & \sigma\sqrt{n} \\ G_\gamma & \xleftrightarrow{\text{joue le rôle de la loi limite}} & N(0, 1) \end{array}$$

Le seul problème qu'on peut soulever est que la loi GEV G_γ dépend de γ , ce qui veut dire qu'elle n'est pas unique. Ce paramètre s'appelle l'indice des valeurs extrêmes, il caractérise le comportement et la forme de la queue de distribution de F , pour cette raison il est appelé aussi indice de queue de la distribution. Fort heureusement, la non unicité de la loi limite ne mène pas à une infinité de cas, ces cas sont appelés domaines d'attraction et qui sont du nombre de trois :

Cas $\gamma = 0$: c'est le domaine d'attraction de GUMBEL [1958].

Si F a un indice de queue nulle on dit qu'elle appartient au domaine d'attraction de Gumbel et on écrit $F \in \text{DA}(\text{Gumbel})$.

Cas $\gamma > 0$: c'est le domaine d'attraction de FRÉCHET [1927].

Si F a un indice de queue positif on dit qu'elle appartient au domaine d'attraction de Fréchet et on écrit $F \in \text{DA}(\text{Fréchet})$.

Cas $\gamma < 0$: c'est le domaine d'attraction de WEIBULL [1951].

Si F a un indice de queue négatif on dit qu'elle appartient au domaine d'attraction de Weibull et on écrit $F \in \text{DA}(\text{Weibull})$.

On représente dans la figure 1.1 les trois densités (à gauche de la figure) ainsi que les répartitions (à droite de la figure) des différentes domaines d'attraction pour les indices de queue $\gamma_{\text{Gumbel}} = 0$, $\gamma_{\text{Fréchet}} > 0$ et $\gamma_{\text{Weibull}} < 0$.

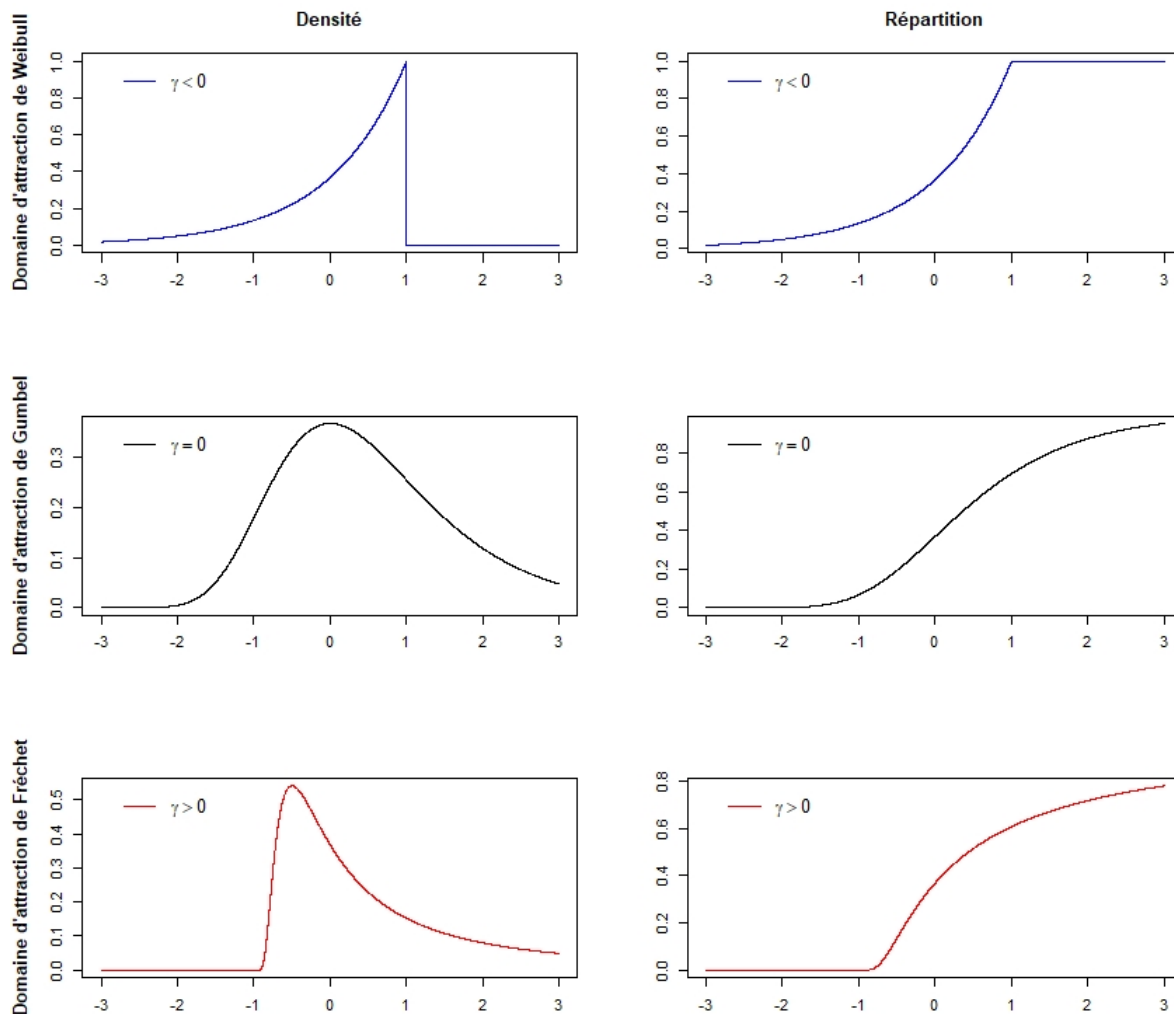


FIGURE 1.1 – Densités (à gauche) et répartitions (à droite) G_γ des valeurs extrêmes.

On a constaté que le domaine qui contient les lois les plus usuelles, telles que la loi Normale, la loi Exponentielle la loi Gamma, est celui de Gumbel. Ces lois ont des fonctions de survie qui décroissent rapidement d'une façon exponentielle, elles sont appelées lois à queue légère. Les lois à queue lourde quand à elles, dont la décroissance de la fonction de survie est polynomiale, appartiennent au domaine d'attraction de Fréchet ; on trouve la loi de Fréchet, la loi de Student, la loi de Paréto. Le domaine d'attraction de Weibull contient que les lois dont le point terminal x_F est fini. le tableau 1.1 résume ce constat [34].

Domaine d'attraction	Weibull ($\gamma < 0$)	Gumbel ($\gamma = 0$)	Fréchet ($\gamma > 0$)
Lois	Uniforme Beta	Gumbel Exponentielle Log-Normale Gamma Weibull Logistique Normale	Fréchet Student Burr Chi-deux Pareto Cauchy

TABLE 1.1 – Lois usuelles et leur domaine d'attraction.

Remarque 1.2

La loi de Weibull fait l'exception de ne pas appartenir à son domaine d'attraction, elle appartient à celui de Gumbel!

Comme la théorie des valeurs extrêmes repose sur les trois distributions, il faut bien connaître leurs propriétés statistiques (voir l'annexe C).

Exemples*1. Domaine d'attraction de Gumbel : la loi exponentielle standard*

$$F(x) = 1 - \exp(-x), x > 0$$

appartient au domaine d'attraction de Gumbel d'indice de valeurs extrême γ nulle

car en prenant
$$\begin{cases} a_n = 1 \\ b_n = \ln(n) \end{cases}$$

$$\begin{aligned} \lim_{n \rightarrow \infty} F^n(a_n x + b_n) &= \lim_{n \rightarrow \infty} [1 - \exp(-(a_n x + b_n))]^n \\ &= \lim_{n \rightarrow \infty} [1 - \exp(-(x + \ln(n)))]^n \\ &= \lim_{n \rightarrow \infty} \left[1 - \frac{\exp(-x)}{n} \right]^n \\ &= \exp(-\exp(-x)) \end{aligned}$$

La figure suivante nous montre ce résultat par simulation.

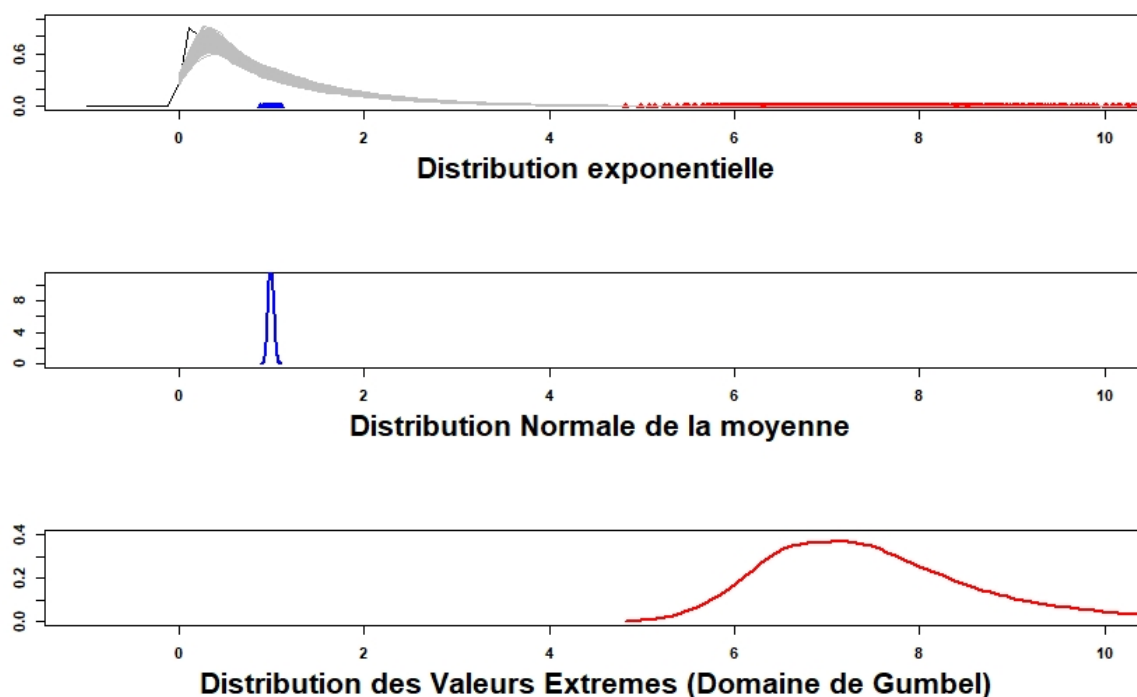


FIGURE 1.2 – Simulation de 1000 loi exponentielles standard en gris. En haut de la figure, les points rouge sont les valeurs extrême, les points bleus sont les moyennes empiriques. Au milieu : la distributions de la moyenne empirique. En bas : la distribution des valeurs extrêmes.

2. *Domaine d'attraction de Fréchet : la loi de Pareto standard*

$$F(x) = 1 - x^{-\alpha}, x > 1, \alpha > 0$$

appartient au domaine de Fréchet avec $\gamma = \frac{1}{\alpha}$. En effet en prenant

$$\begin{cases} a_n = \frac{1}{\alpha} n^{\frac{1}{\alpha}} \\ b_n = n^{\frac{1}{\alpha}} \end{cases}$$

$$\begin{aligned} \lim_{n \rightarrow \infty} F^n(a_n x + b_n) &= \lim_{n \rightarrow \infty} \left[1 - \left(\frac{1}{\alpha} n^{\frac{1}{\alpha}} x + n^{\frac{1}{\alpha}} \right)^{-\alpha} \right]^n \\ &= \lim_{n \rightarrow \infty} \left[1 - \frac{1}{n} \left(\frac{1}{\alpha} x + 1 \right)^{-\alpha} \right]^n \\ &= \exp - \left(1 + \frac{1}{\alpha} x \right)^{-\alpha} \end{aligned}$$

L'illustration suivante donne ce résultat par simulation.

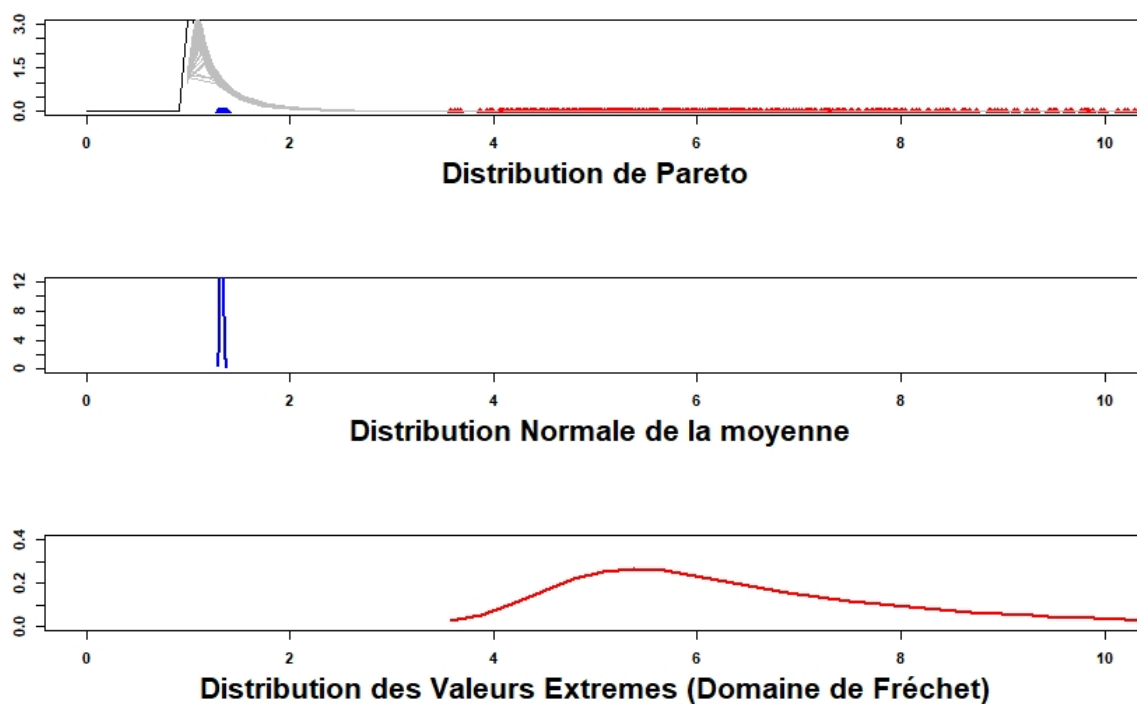


FIGURE 1.3 – Simulation de 1000 loi de Pareto de paramètres $\alpha = 4$ et d'échelle 1 en gris. En haut de la figure, les points rouge sont les valeurs extrême, les points bleus sont les moyennes empiriques. Au milieu : la distributions de la moyenne empirique. En bas : la distribution des valeurs extrêmes.

3. Domaine d'attraction de Weibull : la loi uniforme standard

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \in [0, 1] \\ 1 & \text{si } x > 1 \end{cases}$$

appartient au domaine de Weibull avec $\gamma = -1$ car en prenant

$$\begin{cases} a_n = -\frac{1}{n} \\ b_n = 1 \end{cases}$$

$$\begin{aligned} \lim_{n \rightarrow \infty} F^n(a_n x + b_n) &= \lim_{n \rightarrow \infty} \left[F\left(1 - \frac{1}{n}x\right) \right]^n \\ &= \begin{cases} 1 & \text{si } x < 0 \\ \lim_{n \rightarrow \infty} \left[1 - \frac{1}{n}x\right]^n & \text{si } x \in [0, n] \end{cases} \\ &= \begin{cases} 1 & \text{si } x < 0 \\ \exp(-x) & \text{si } x \geq 0 \end{cases} \end{aligned}$$

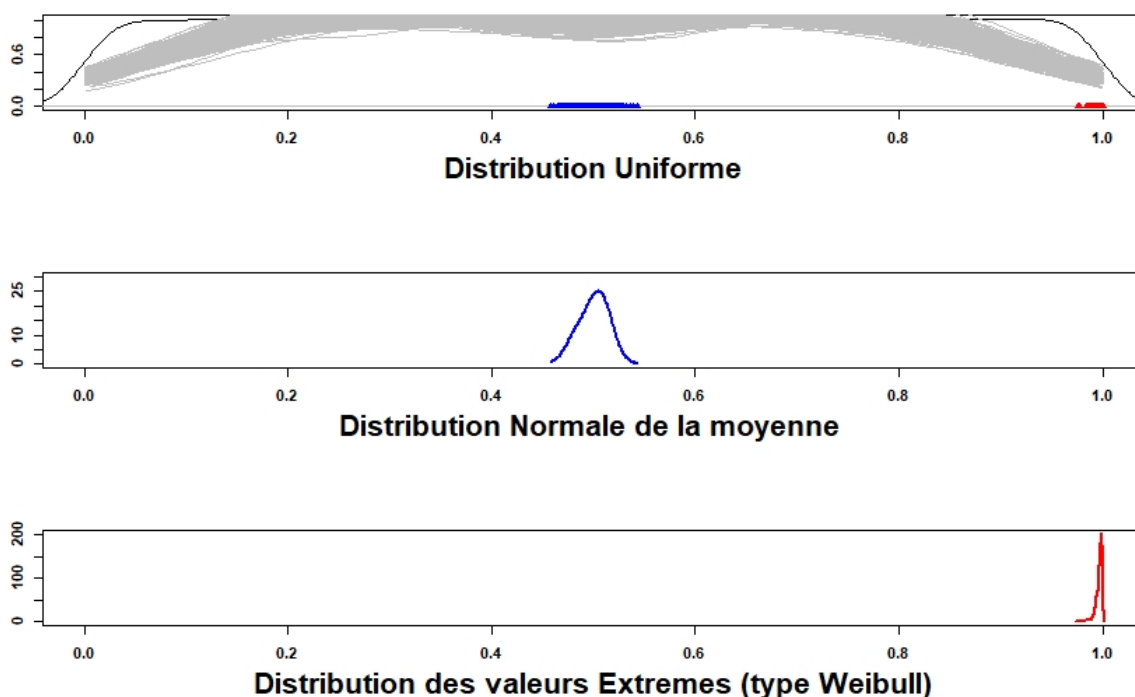


FIGURE 1.4 – Simulation de 300 loi Uniformes standards en gris. En haut de la figure, les points rouge sont les valeurs extrême, les points blues sont les moyennes empiriques. Au milieu : la distributions de la moyenne empirique. En bas : la distribution des valeurs extrêmes.

1.1.2 Méthode des maxima par blocs (BM)

Théoriquement, on ne peut appliquer le théorème de Fisher-Tippet que si on a un grand nombre d'observations i.i.d. pour donner la loi approchée du maximum. En pratique, si on dispose de n observations (x_1, \dots, x_n) , on procède comme suit :

1. regrouper les données en k blocs de taille l ,
2. calculer le maximum sur chaque bloc :

$$m_i = \max(x_{(i-1)l+1}, \dots, x_{il}) \text{ pour } i \in 1, \dots, k,$$

3. approcher ensuite la loi de la variable aléatoire M_i par une loi GEV,
4. estimer les paramètres de la loi GEV en utilisant l'échantillon (m_1, \dots, m_k) ,
5. chercher à satisfaire des conditions d'optimalité en jouant sur la taille des blocs l , qui doit être assez grande pour que l'approximation par la loi GEV soit réaliste, et sur le nombre de blocs k qui doit être assez grand pour avoir assez d'informations pour estimer les trois paramètres de la GEV.

Exemple 1.1

Dans cet exemple on a généré 10000 valeurs issues de la loi normale qu'on a partagé en 100 blocs de taille 100 chacun et dont on a suivi la démarche précédente. Les résultats sont les suivants :

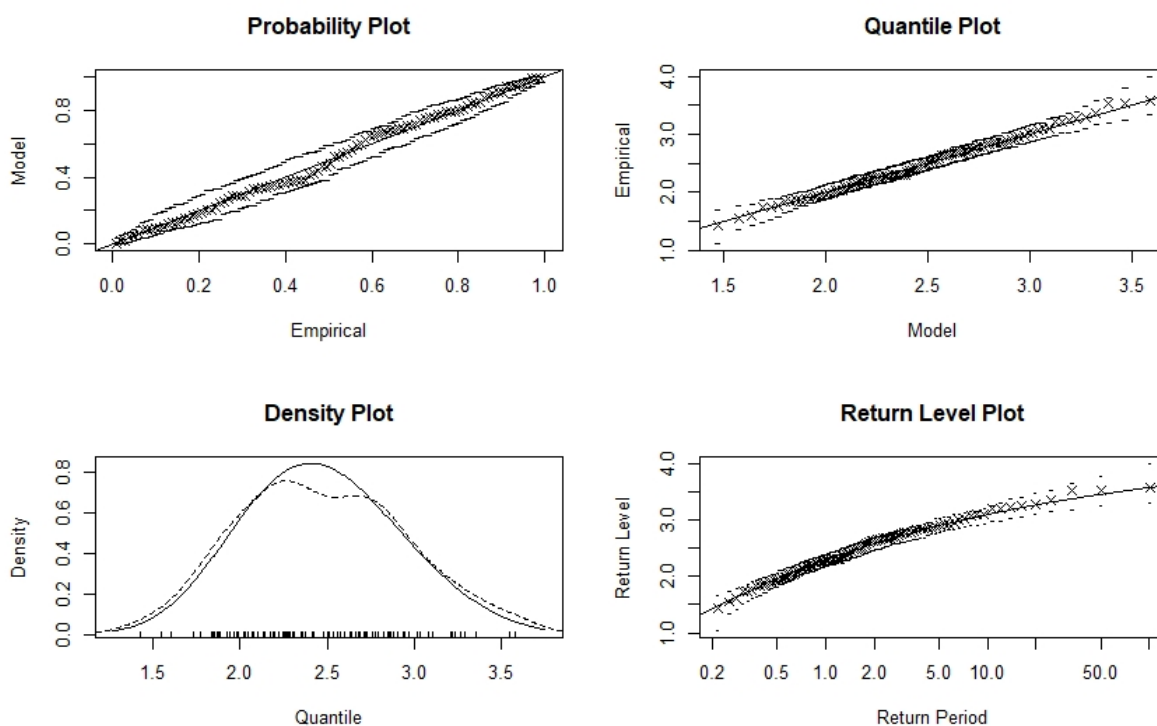


FIGURE 1.5 – Comportement du maximum par blocs de taille 100 de 10000 v.a Normales standard

1.1.3 Loi des excès

La méthode du dépassement du seuil " Peaks-Over-Threshold " en anglais, notée POT, à comme stratégie l'utilisation des statistiques d'ordre supérieur de l'échantillon. En se fixant un seuil, on garde que les valeurs de la statistique qui le dépassent. L'écart entre l'observation et le seuil est défini comme l'excès au-delà du seuil. Cela se traduit mathématiquement par : si on a une suite de variables aléatoires *i.i.d* X_1, X_2, \dots, X_n de fonction de répartition F et de point terminal x_F . en se fixant un seuil u déterministe tel que $u < x_F$ et en considérant les N_u observations qui dépassent le seuil u , et qui sont $X_{i_1}, X_{i_2}, \dots, X_{i_{N_u}}$. Alors l'excès au-delà du seuil u est défini par $Y_j = X_{i_j} - u, j = 1, \dots, N_u$. Pour bien comprendre la procédure, la figure 1.6 illustre la variable des excès.

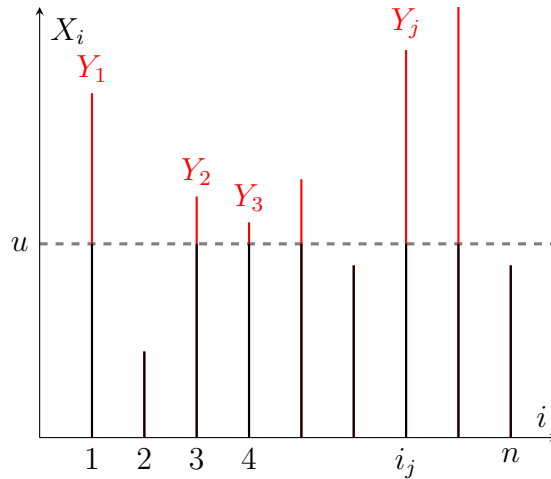


FIGURE 1.6 – Les excès Y_1, Y_2, \dots de X au delà du seuil u .

On note F_u la fonction de répartition des excès au-delà du seuil u . Cette fonction de répartition est en fait une probabilité conditionnelle, car comme on l'a mentionné précédemment, on a imposé qu'une condition soit vérifiée et qui est que la variable observée X doit dépasser le seuil u , ce qui se traduit par l'événement $(X > u)$. La loi des excès Y pour une suite de v.a.i.i.d X_1, X_2, \dots, X_n s'écrit

$$F_u(y) = \mathbb{P}(Y \leq y | X > u).$$

Cette répartition peut s'écrire en fonction de la répartition de X en utilisant la définition

de probabilité conditionnelle

$$\begin{aligned}
 F_u(y) &= \mathbb{P}(Y \leq y | X > u) = \frac{\mathbb{P}(Y \leq y, X > u)}{\mathbb{P}(X > u)} = \frac{\mathbb{P}(Y \leq y, X - u > 0)}{\mathbb{P}(X > u)} \\
 &= \frac{\mathbb{P}(X - u \leq y, X - u > 0)}{\mathbb{P}(X > u)} = \begin{cases} 0 & \text{si } y < 0 \\ \frac{\mathbb{P}(0 < X - u \leq y)}{\mathbb{P}(X > u)} & \text{si } y \geq 0 \end{cases} \\
 &= \begin{cases} 0 & \text{si } y < 0 \\ \frac{\mathbb{P}(u < X \leq y + u)}{\mathbb{P}(X > u)} & \text{si } y \geq 0 \end{cases} = \begin{cases} 0 & \text{si } y < 0 \\ \frac{F(y+u) - F(u)}{1 - F(u)} & \text{si } y \geq 0 \end{cases}
 \end{aligned}$$

$1 - F(u)$ est dite la fonction de survie, elle est notée $\bar{F}(u)$. On peut l'introduire dans la formule précédente pour le loi des excès, et on aura

$$\bar{F}_u(y) = 1 - F_u(y) = \begin{cases} 1 & \text{si } y < 0 \\ \frac{\bar{F}(y+u)}{\bar{F}(u)} & \text{si } y \geq 0 \end{cases}$$

Exemple 1.2

Dans cet exemple on a généré 10000 valeurs issues de la loi normale qu'on a partager en 100 blocs de taille 100 chacun et dont on a suivi la démarche POT. Les résultats sont les suivants :

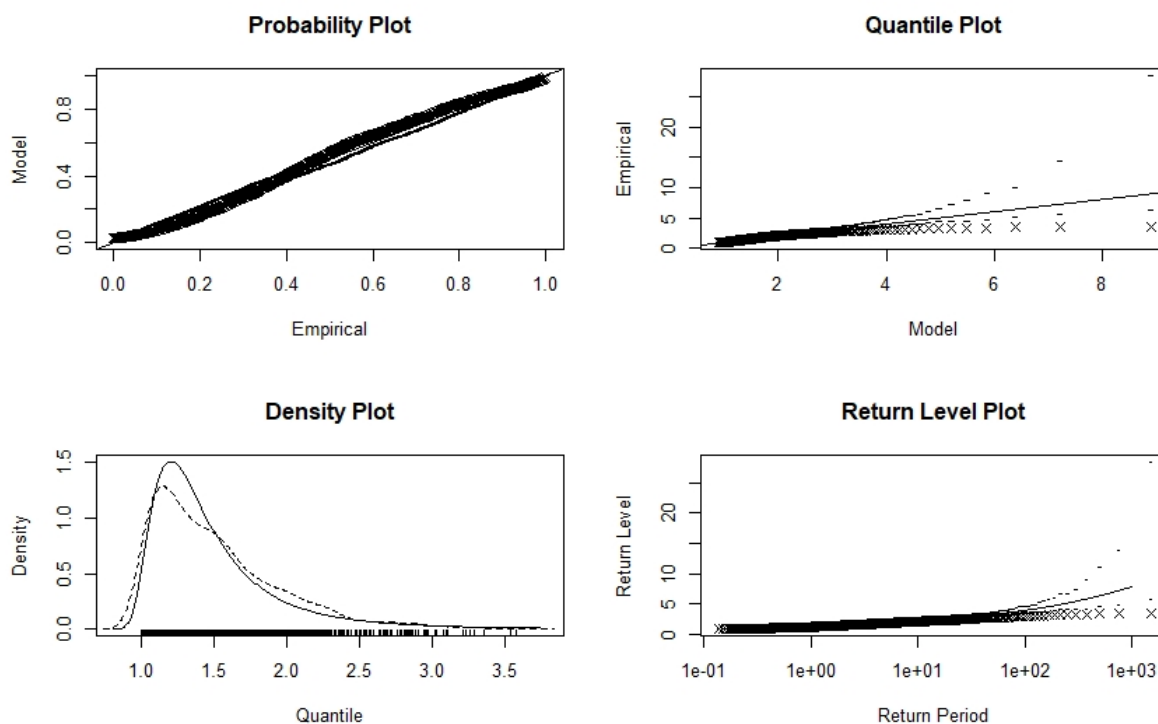


FIGURE 1.7 – La méthode POT pour une échantillon de 10000 lois normales standards avec un seuil égal à 1.

1.1.4 Lien entre GEV et loi des excès

Ce paragraphe montre qu'il y a en réalité une équivalence entre la loi GEV et celle des excès. On peut montrer ce qu'on a avancé en utilisant le Théorème 1.2, pour n assez grand

$$F^n(u) \approx \exp\left(-\left(1 + \gamma\left(\frac{u - b_n}{a_n}\right)\right)^{-1/\gamma}\right) \text{ avec } a_n > 0, (b_n, \gamma) \in \mathbb{R}^2.$$

Ce qui implique en passant au logarithme

$$n \ln(F(u)) \approx -\left(1 + \gamma\left(\frac{u - b_n}{a_n}\right)\right)^{-1/\gamma} \quad (1.5)$$

Si le seuil u est assez élevé alors $F(u)$ est proche de 1 et dans ce cas le développement limité de la fonction de répartition donne

$$\ln(F(u)) = \ln(1 - (1 - F(u))) \approx -(1 - F(u))$$

En remplaçant dans la formule 1.5 on aura pour u est assez grand

$$1 - F(u) \approx \frac{1}{n} \left(1 + \gamma\left(\frac{u - b_n}{a_n}\right)\right)^{-1/\gamma}.$$

En procédant de la même manière que précédemment, pour $y > 0$ on a,

$$1 - F(u + y) \approx \frac{1}{n} \left(1 + \gamma\left(\frac{u + y - b_n}{a_n}\right)\right)^{-1/\gamma}.$$

En utilisant la propriété de la fonction de survie, on obtient :

$$F_u(y) \approx 1 - \left(1 + \gamma\frac{y}{\sigma}\right)^{-1/\gamma} \text{ avec } \sigma = a_n + \gamma(u - b_n).$$

Lorsque la fonction de répartition s'approche de son point terminale y_F , les travaux de Balkema et de Haan [9], Pickands [62] vont plus loin dans la précision de cette approximation et énoncent :

Théorème 1.4 (Balkema et de Haan [9], Pickands [62])

F appartient au domaine d'attraction de G_γ si et seulement si il existe $\sigma > 0$ et $\gamma \in \mathbb{R}$ tels que la loi des excès F_u peut être uniformément approchée par une loi de Pareto généralisée (GPD) notée $\mathcal{G}_{\gamma,\sigma}$, i.e.

$$\lim_{y \rightarrow y_F} \sup_{y \in]0, y_F - u[} |F_u(y) - \mathcal{G}_{\gamma,\sigma}(y)| = 0$$

où

$$\mathcal{G}_{\gamma,\sigma}(y) = 1 - \left(1 + \gamma\frac{y}{\sigma}\right)^{-1/\gamma}$$

est définie pour $y \geq 0$ si $\gamma \geq 0$ ou $[0, -\sigma/\gamma[$ si $\gamma < 0$.

Remarque 1.3

On peut voir dans la formule qu'on ne peut pas remplacer γ par le zéro, donc il faut passer à la limite pour traiter ce cas et on trouve

$$\lim_{\gamma \rightarrow 0} \mathcal{G}_{\gamma, \sigma}(y) = \mathcal{G}_{0, \sigma}(y) = \lim_{\gamma \rightarrow 0} 1 - \left(1 + \gamma \frac{y}{\sigma}\right)^{-1/\gamma} = 1 - \exp\left(-\frac{1}{\sigma}y\right), y \geq 0$$

qui n'est rien d'autre que la loi exponentielle de paramètre $\frac{1}{\sigma}$.

Le Théorème 1.4 fait la suggestion d'approcher la loi des excès d'un échantillon lorsque le seuil choisit est assez grand par une loi GPD. Cette idée est bien traitée dans l'article de Beirland *et al.* [14]

Chapitre 2

Caractérisation des lois extrêmes

Sommaire

2.1	Classes de distributions à queue lourde	35
2.1.1	Classification sans comparaison à la loi normale	35
2.1.2	Classification avec comparaison à la loi normale	38
	Classe E : Distributions avec des moments exponentiels inexistants	40
	Classe D : Distributions Subexponentielles	40
	Classe C : Distributions à variations régulières d'indice $\alpha > 0$	40
	Classe B : Comportement de Pareto avec $\alpha > 0$	48
	Classe A : Distributions α -Stables	48
2.2	Caractérisation des lois extrêmes	49
2.2.1	Domaine d'attraction de Fréchet	49
2.2.2	Domaine d'attraction de Weibull	49
2.2.3	Domaine d'attraction de Gumbel	50
2.2.4	Caractérisation générale	50

Introduction

Ce chapitre répond à la question suivante : étant donnée une loi H de type extrême quels sont les tests à établir pour que la loi du maximum de la suite de variables aléatoires i.i.d. de loi F converge vers H ? La réponse est donnée pour les trois domaines d'attraction de Fréchet, Gumbel et de Weibull dans [27], [63], [34] et [27]. Les tests qui caractérisent l'appartenance de la loi du maximum à un domaine d'attraction font appel à la notion de fonctions à variations régulières et entre autres les fonctions à queues lourdes qui sont une notion importante dans le sujet étudié dans cette thèse. Pour cette raison nous les étudions dans ce qui suit.

2.1 Classes de distributions à queue lourde

L'utilisation du terme «distribution à queue lourde» varie selon le domaine d'intérêt, mais il est souvent considéré comme correspondant à une absence de moments exponentiels (positifs). Nous donnons deux caractérisations des distributions à queue lourde, la première est introduite juste après ce paragraphe, et la deuxième qui est une caractérisation basée sur la comparaison de la distribution avec la loi de référence, la loi normale.

2.1.1 Classification sans comparaison à la loi normale

Définition 2.1

Une distribution F sur \mathbb{R} est dite à support non borné à droite si pour tout x , $\overline{F}(x) > 0$.

Définition 2.2

Une distribution F est dite à queue lourde (à droite) si et seulement si

$$\int_{\mathbb{R}} e^{\lambda x} dF(x) = \infty \text{ pour tout } \lambda > 0. \quad (2.1)$$

Définition 2.3

Une distribution F est dite à queue légère si et seulement si

$$\int_{\mathbb{R}} e^{\lambda x} dF(x) < \infty \text{ pour certain } \lambda > 0. \quad (2.2)$$

i.e. si et seulement si elle n'est pas à queue lourde.

Il est clair que pour toute distribution F à queue légère sur la demi droite positive $\mathbb{R}^+ = [0, +\infty[$, tous les moments d'ordre k sont finis.

$$\int_{\mathbb{R}} x^k dF(x) < \infty \text{ pour tout } k > 0.$$

On dira qu'une fonction non négative (tendant généralement vers zéro) est à queue lourde si elle n'est pas dominée par une fonction exponentielle décroissante. Plus précisément nous énonçons la définition suivante.

Définition 2.4

On dira d'une fonction $f \geq 0$ qu'elle est à queue lourde si et seulement si

$$\limsup_{x \rightarrow \infty} f(x)e^{\lambda x} = \infty \text{ pour tout } \lambda > 0. \quad (2.3)$$

Théorème 2.1

Pour toute distribution F , les assertions suivantes sont équivalentes :

1. F est à queue lourde.

2. La fonction \bar{F} est à queue lourde.

Il est clair que la propriété de queue lourde d'une fonction revient à la forme de sa queue. Le Théorème 2.1 montre en particulier qu'une distribution est à queue lourde si et seulement si sa fonction de queue est une fonction à queue lourde. Nous introduisons la définition suivante pour l'extension de ce dernier théorème.

Définition 2.5

Pour toute fonction F , la fonction $R(x) = -\ln \bar{F}(x)$ est dite la fonction hasard de la distribution. Si la fonction hasard (fonction de risque ou danger) est différentiable alors sa dérivée $r(x) = R'(x)$ est dite taux de hasard (taux de risque).

Théorème 2.2

Pour toute distribution F , les assertions suivantes sont équivalentes :

1. F est à queue lourde.
2. La fonction \bar{F} est à queue lourde.
3. La fonction de hasard R correspondante satisfait $\liminf_{x \rightarrow \infty} R(x)/x = 0$.
4. Pour certains (tout) T fixe avec $T > 0$, la fonction $F(x, x + T]$ est à queue lourde.

Preuve

1) \Rightarrow 4). Supposons que la fonction $F(x, x + T]$ n'est pas à queue lourde. Alors

$$c = \sup_{x \in \mathbb{R}} F(x, x + T] e^{\lambda' x} < \infty \text{ pour certains } \lambda' > 0,$$

et donc pour tous $\lambda < \lambda'$

$$\begin{aligned} \int_0^\infty e^{\lambda x} dF(x) &\leq \sum_{n=0}^\infty e^{\lambda(n+1)T} F(nT, nT + T] \\ &\leq c \sum_{n=0}^\infty e^{\lambda(n+1)T} e^{-\lambda' nT} = ce^{\lambda T} \sum_{n=0}^\infty e^{(\lambda - \lambda')nT} < \infty. \end{aligned}$$

Il s'ensuit que l'intégrale définie en 2.1 qu'il est finie pour tout $\lambda \in (0, \lambda')$, ce qui implique que la distribution F ne peut pas être à queue lourde d'où la contradiction, ce qui fait que l'implication est vraie.

4) \Rightarrow 2). Cette implication découle de l'inégalité $\bar{F}(x) \geq F(x, x + T]$.

2) \Rightarrow 3). Supposons qu'au contraire, la limite inférieure \liminf dans 3) est (strictement) positive. Alors il existe $x_0 > 0$ et $\varepsilon > 0$ tel que $R(x) \geq \varepsilon x$ pour tous $x \geq x_0$ ce qui implique que $\bar{F}(x) \leq e^{-\varepsilon x}$ qui est en contradiction avec 2).

3) \Rightarrow 1). Supposons qu'au contraire, la fonction F est à queue légère. Il découle alors de 2.2 en utilisant par exemple l'inégalité de Chebyshev que pour certains $\lambda > 0$ et $c > 0$, nous avons $\bar{F}(x) \leq ce^{-\lambda x}$ pour tous x . Ce qui implique que $\liminf_{x \rightarrow \infty} R(x)/x \geq \lambda$ qui contredit 3).

Lemme 2.1

Soit F une distribution absolument continue de densité f . Supposons que la distribution F est à queue lourde. Alors la fonction f est aussi à queue lourde.

Preuve

Supposons que f n'est pas à queue lourde; alors ils existent $\lambda' > 0$ et x_0 tels que

$$c = \sup_{x > x_0} f(x)e^{\lambda'x} < \infty,$$

et donc pour tous $\lambda \in (0, \lambda')$

$$\int_{\mathbb{R}} e^{\lambda x} dF(x) \leq e^{\lambda x_0} + c \int_{x_0}^{\infty} e^{(\lambda - \lambda')x} dx < \infty.$$

Il s'ensuit que l'intégrale définie dans 2.1 est finie pour tout λ tel que $0 < \lambda < \lambda'$, ce qui contredit le fait que F est à queue lourde.

Remarque 2.1

La réciproque est fautive, en général. Pour se convaincre, considérons la fonction densité continue par morceaux

$$f(x) = \sum_{n=1}^{\infty} \mathbb{1}_{[n, n+2^{-n}]}(x).$$

nous avons d'une part

$$\limsup_{x \rightarrow \infty} f(x)e^{\lambda x} = \infty \text{ pour tout } \lambda > 0,$$

ainsi la fonction f est à queue lourde. D'une autre part, pour tout $\lambda \in (0, \ln 2)$,

$$\int_0^{\infty} e^{\lambda x} f(x) dx < \sum_{n=1}^{\infty} e^{\lambda(n+2^{-n})} 2^{-n} = \sum_{n=1}^{\infty} e^{\lambda(n+2^{-n}) - n \ln 2} < \infty,$$

donc F est à queue légère.

Nous donnons maintenant quelques exemples de distributions à queue lourde.

Exemple 2.1

La distribution de Pareto sur \mathbb{R}^+ de paramètre d'échelle $a > 0$ et de forme $\alpha > 0$ qui a comme fonction de queue \bar{F}

$$\bar{F}(x) = \left(\frac{a}{x+a} \right)^{\alpha}.$$

Il est clair que

$$\bar{F}(x) \sim \left(\frac{x}{a} \right)^{-\alpha} \text{ quand } x \rightarrow \infty.$$

La distribution de Pareto a des moments finis pour tout ordre strictement inférieur à α et des moments infinis pour les ordres qui sont supérieurs à α .

Exemple 2.2

La distribution de Burr sur \mathbb{R}^+ de paramètres strictement positifs α, β, a qui a comme fonction de queue \bar{F}

$$\bar{F}(x) = \left(\frac{a}{x^\beta + a} \right)^\alpha.$$

On peut constater que cette distribution est une généralisation de la distribution précédente. Il est clair que

$$\bar{F}(x) \sim a^\alpha x^{-\alpha\beta} \text{ quand } x \rightarrow \infty.$$

Donc La distribution de Burr a des moments finis pour tout ordre strictement inférieur à $\alpha\beta$ et des moments infinis pour les ordres qui sont supérieurs à $\alpha\beta$.

Exemple 2.3

La distribution de Weibull sur \mathbb{R}^+ de paramètres strictement positifs α, β qui a comme fonction de queue \bar{F}

$$\bar{F}(x) = \exp - \left(\frac{x}{\beta} \right)^\alpha.$$

Cette distribution est à queue lourde si et seulement si $\alpha < 1$. Tous les moments de cette distribution sont finis.

Nous résumons ces exemples non exhaustifs sous forme de tableaux des distributions à queue légère et à queue lourde avec de nouveaux cas :

— Distributions à queues légères

Nom	Paramètres	densité
Exponentielle	$\lambda > 0$	$\lambda \exp - \lambda x$
Gamma	$\alpha > 0, \beta > 0$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
Weibull	$\alpha \geq 1, \beta > 0$	$\alpha \beta x^{\alpha-1} e^{-\beta x^\alpha}$

— Distributions à queues lourdes

Nom	Paramètres	densité
Log-normale	$\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sigma x \sqrt{2\pi}} \exp - \frac{(\ln x - \mu)^2}{2\sigma}$
Pareto	$\alpha > 0, a > 0$	$\frac{\alpha}{a+x} \left(\frac{a}{a+x} \right)^\alpha$
Weibull	$0 < \alpha < 1, \beta > 0$	$\alpha \beta x^{\alpha-1} e^{-\beta x^\alpha}$
Burr	$\alpha > 0, \beta > 0, \gamma > 0$	$\frac{\alpha \gamma \beta^\alpha x^{\gamma-1}}{(\beta + x^\gamma)^{\alpha+1}}$

2.1.2 Classification avec comparaison à la loi normale

Définition 2.6

Étant donnée une variable aléatoire réelle X d'espérance μ et d'écart type σ , on définit son coefficient d'aplatissement ou kurtosis non normalisé β_2 comme le moment d'ordre

quatre de la variable centrée réduite :

$$\beta_2 = \mathbb{E} \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right]$$

lorsque cette espérance existe. On a donc :

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{(\mu_2 - \mu_1^2)^2},$$

avec μ_i le moment d'ordre i .

Définition 2.7

Une distribution est à queue lourde si et seulement si son coefficient d'aplatissement est supérieur à celui de la loi Normale (pour laquelle $\beta_2 = 3$).

Remarque 2.2

Cette caractérisation générale ne peut être appliquée que si le moment d'ordre 4 existe ce qui empêche toute discrimination entre les distributions qui ont un moment d'ordre 4 infini.

On peut faire un classement selon la queue droite de certaines distributions comme suit, mais malheureusement on ne peut pas les classer toutes en se basant sur ce critère. Le classement suivant est fait par [76].

1. les distributions avec des moments exponentielles inexistantes (E),
2. les distributions Subexponentielles (D),
3. les distributions à variations régulières (C),
4. les distributions avec un comportement de Pareto (B),
5. les distributions α -Stables avec $\alpha < 2$ (A).

Ce sont des classes emboîtées et on a : $A \subset B \subset C \subset D \subset E$ et nous pouvons les représenter comme suit

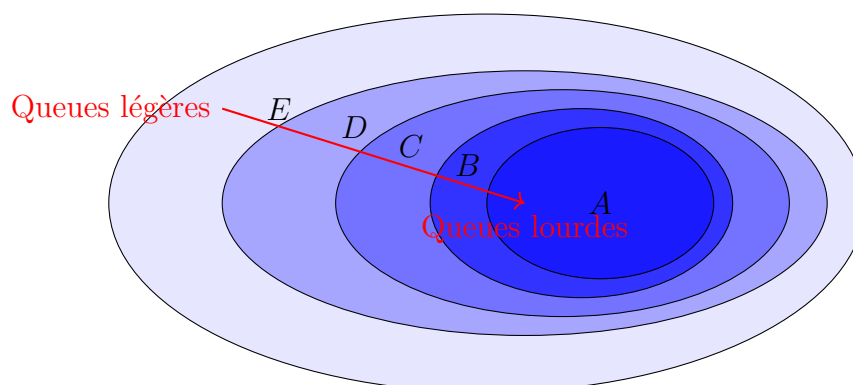


FIGURE 2.1 – Différentes classes de distributions de queue de très légère (E) à très lourde (A).

La classe B et la classe C sont très importants dû leur lien à la théorie des valeurs extrêmes.

Classe E : Distributions avec des moments exponentiels inexistants

C'est la classe des distributions avec des moments exponentiels inexistants, c'est-à-dire les distributions telles que

$$E(e^X) = \infty.$$

la classe E est celle des distributions ayant une queue plus lourde que celle de la loi normale car la probabilité de dépassement pour les extrêmes décroît moins vite que celle de la loi normale.

Classe D : Distributions Subexponentielles

Cette classe est caractérisée par la définition suivante introduite par Beirlant et al. 2006 [12] :

Définition 2.8

Une distribution est dite subexponentielle si :

$$\lim_{x \rightarrow \infty} \frac{P(X_1 + \dots + X_n > x)}{P(\max(X_1, \dots, X_n) > x)} = 1. \quad (2.4)$$

Cela veut dire que si on dispose de n variables aléatoires subexponentielles indépendantes et identiquement distribuées leur somme est extrême si et seulement si leur maximum est extrême.

L'équation 2.4 implique que :

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x)}{e^{-\varepsilon x}} = \infty \text{ quelque soit } \varepsilon > 0. \quad (2.5)$$

L'équation 2.5 nous suggère que la classe D contient les distributions dont leur queue décroît plus lentement que n'importe quelle loi exponentielle car comme on peut le constater le dénominateur de cette dernière équation n'est rien d'autre que l'expression de la queue de la loi exponentielle.

Classe C : Distributions à variations régulières d'indice $\alpha > 0$

C'est une classe de distributions définie par (Goldie et Klüppelberg 1998) [41] dont les distributions vérifient la condition suivante (Embrechts, Klüppelberg et Mikosch 2013 [34]) :

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha} \quad (2.6)$$

Ce qui veut dire que le comportement des valeurs extrêmes des distributions de cette classe est le même que celui de la loi de Pareto dont la probabilité de dépassement de ses valeurs extrêmes décroît géométriquement ou suivant une fonction puissance.

Le paramètre α dans est appelé indice de queue ou indice des valeurs extrêmes (Tail index). C'est un paramètre très important dans la théorie des valeurs extrêmes et sert à classer les distributions selon le comportement de la fonction de probabilité au dépassement des valeurs extrêmes.

Dûe l'importance de cette classe dans notre travail et en particulier notre thèse nous développons plus cette partie et nous donnons des définitions et des propriétés que nous voyons nécessaires pour la suite.

Distributions à Variations Régulières

Dans cette partie, on traite la classe \mathcal{C} des fonctions qui apparait dans un vaste nombre d'applications dans la totalité de mathématiques. Ici, on va définir quelques généralités sur ces fonctions avec certaines de leurs propriétés les plus importantes. Ceux qui sont intéressés par la théorie de variation régulière, peuvent consulter par exemple Bingham et al. [15], Embrechts et al. [34], Beirlant et al. [11], de Haan et Ferreira [27] et Resnick [63].

Définition 2.9 (Fonctions à variation régulière et à variation lente)

- Une fonction mesurable $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est à variation régulière à ∞ avec l'index $\rho \in \mathbb{R}$, et on note $V \in \mathcal{RV}_\rho$, si

$$\lim_{x \rightarrow \infty} \frac{V(tx)}{V(x)} = t^\rho, \quad t > 0 \quad (2.7)$$

on appelle ρ l'exposant de variation ou l'indice de variation régulière.

- Une fonction mesurable $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est à variation lente à ∞ avec l'index $\rho = 0$ et on note $V \in \mathcal{RV}_0$, si

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

- Une fonction à variation régulière d'indice $\rho \in \mathbb{R}$ peut toujours s'écrire sous la forme $V(x) = x^\rho \ell(x)$, où $\ell \in \mathcal{RV}_0$

Les exemples typiques des fonctions à variations lentes sont les fonctions convergent vers une constante positive et en particulier les fonctions positives constantes, ainsi que les fonctions logarithmes.

- Les fonctions x^ρ (qui est considérée comme la fonction de référence pour les fonctions à variations régulières), $\ln(x)$, $\ln(1+x)$, $\ln \ln(1+x)$ et $\exp[(\ln x)^\alpha]$, $0 < \alpha < 1$ sont à variations régulières.
- Les fonctions $\exp(x)$, $\sin(x+1)$, $\exp(\log(x))$ et $\exp(\log(1+x))$ ne sont pas à variations régulières.

Nous donnons dans ce qui suit, quelques propriétés élémentaires, des fonctions à variations lentes.

Proposition 2.1 (Propriétés de fonction à variation lente)

1. \mathcal{RV}_0 est fermé sous l'addition, la multiplication et la division.
2. Si ℓ est à variation lente, $\lim_{x \rightarrow \infty} (\ln \ell(x)) / \ln x = 0$.
3. Si ℓ est à variation lente, alors ℓ^α est à variation lente pour tout $\alpha \in \mathbb{R}$.
4. Si ℓ est à variation lente et $\rho > 0$ alors

$$\lim_{x \rightarrow \infty} x^\rho \ell(x) = \infty, \quad \lim_{x \rightarrow \infty} x^{-\rho} \ell(x) = 0.$$

Le premier résultat utile que nous allons énoncer est le Théorème de la convergence uniforme.

Théorème 2.3 (Convergence uniforme)

Si $V \in \mathcal{RV}_\rho$, pour $0 < a \leq b < \infty$, alors la relation 2.7 se tient uniformément pour

1. $x \in [a, b]$ si $\rho = 0$
2. $x \in (0, b]$ si $\rho > 0$
3. $x \in [a, \infty)$ si $\rho < 0$

Preuve

Voir Bingham et al. [15] Théorème 1.5.2, page 22.

Un autre résultat important concerne la représentation des fonctions à variations régulières.

Théorème 2.4 (Représentation de Karamata)

1. $\ell \in \mathcal{RV}_0$, si et seulement si peut être représentée sous la forme

$$\ell(x) = c(x) \exp \left\{ \int_1^x \frac{r(t)}{t} dt \right\}, \quad x > 0 \tag{2.8}$$

où c, r sont des fonctions mesurables, et

$$\lim_{x \rightarrow \infty} c(x) = c_0 \in (0, \infty), \text{ et } \lim_{t \rightarrow \infty} r(t) = 0 \tag{2.9}$$

2. Une fonction $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ à variation régulière avec l'index ρ si et seulement si V a la représentation

$$V(x) = c(x) \exp \left\{ \int_1^x t^{-1} \rho(t) dt \right\}, \quad x > 0 \tag{2.10}$$

où c satisfait 2.9 et $\lim_{t \rightarrow \infty} \rho(t) = \rho$.

Preuve

Voir Resnick [64], Corollaire 2.1, page 29.

Quelques autres propriétés utiles sont rassemblés dans la Proposition suivante :

Proposition 2.2

1. Si $V \in \mathcal{RV}_{\rho_1}$, $-\infty \leq \rho \leq \infty$, alors $\lim_{x \rightarrow \infty} \log V(x) / \log x = \rho$, donc

$$\lim_{x \rightarrow \infty} V(x) \rightarrow \begin{cases} \infty & \text{si } \rho > 0 \\ 0 & \text{si } \rho < 0 \end{cases}$$

2. On suppose que V est non décroissante, $V(\infty) = \infty$, et $V \in \mathcal{RV}_{\rho}$, $0 \leq \rho \leq \infty$. Alors

$$V^{-1} \in \mathcal{RV}_{1/\rho}$$

3. **Inégalité de Potter** : On suppose que $V \in \mathcal{RV}_{\rho}$, $\rho \in \mathbb{R}$ et pour tout $\varepsilon > 0$. Alors il existe t_0 tel que pour $x \geq 1$ et $t \geq t_0$

$$(1 - \varepsilon)x^{\rho - \varepsilon} < \frac{V(tx)}{V(x)} < (1 + \varepsilon)x^{\rho + \varepsilon}$$

4. Si $V \in \mathcal{RV}_{\rho}$, $\rho \in \mathbb{R}$ et $(a_n), (b_n)$ sont deux suites strictement positives et tendent vers l'infini et $b_n \sim ca_n$ lorsque $n \rightarrow \infty$ avec $0 < c < \infty$, alors $V(b_n) \sim c^{\rho}V(a_n)$. Ce résultat reste vrai si $\rho \neq 0$ pour $c = 0$ ou ∞ . Des résultats analogues sont valables en remplaçant les suites par des fonctions.

5. Si $V_1 \in \mathcal{RV}_{\rho_1}$, et $V_2 \in \mathcal{RV}_{\rho_2}$, $\rho_2 < \infty$ et $\lim_{x \rightarrow \infty} V_2(x) = \infty$, alors

$$V_1 \circ V_2 \in \mathcal{RV}_{\rho_1 \rho_2}.$$

6. Soient deux fonctions croissantes V_1, V_2 telles que $V_1, V_2 \in \mathcal{RV}_{\rho}$, $0 < \rho < \infty$. alors pour $0 \leq c \leq \infty$,

$$V_1(x) \sim cV_2(x), x \rightarrow \infty$$

est équivalent à

$$V_1^{-1}(x) \sim c^{-\rho^{-1}}V_2^{-1}(x), x \rightarrow \infty.$$

7. Si $V \in \mathcal{RV}_{\rho}$, $\rho \neq 0$, alors il existe une fonction V^* qui soit absolument continue et strictement monotone et vérifiant

$$V(x) \sim V^*(x), \text{ lorsque } x \rightarrow \infty.$$

Preuve

Voir Resnick [64], Proposition 2.6, page 32.

L'inégalité de Potter peut être reformulée comme suit :

Proposition 2.3

Soient $V \in \mathcal{RV}_\rho$, $\delta_1 > 0$, $\delta_2 > 0$. Alors il existe $t_0 \in \mathbb{R}$ tel que, pour tout $x \geq 1$ et pour tout $t \geq t_0$, avec $tx \geq t_0$,

$$(1 - \delta_1)x^\rho \min(x^{\delta_2}, x^{-\delta_2}) < \frac{V(tx)}{V(t)} < (1 + \delta_1)x^\rho \max(x^{\delta_2}, x^{-\delta_2}).$$

Cette dernière Proposition sert à démontrer ce qui suit.

Proposition 2.4

Soit $V \in \mathcal{RV}_\rho$ et soient $(u_n), (v_n)$ deux suites vérifiant $u_n \rightarrow \infty$ et $v_n \sim u_n$ quand $n \rightarrow \infty$. Alors $V(v_n) \sim V(u_n)$ quand $n \rightarrow \infty$.

Définition 2.10 (Fonction à variation rapide)

Une fonction mesurable $V : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est à variation rapide ou à variation régulière d'indice ∞ et on note $V \in \mathcal{RV}_\infty$ si pour tout $t > 0$

$$\lim_{x \rightarrow \infty} \frac{V(tx)}{V(x)} = x^\infty = \begin{cases} 0 & \text{si } x < 1 \\ 1 & \text{si } x = 1 \\ \infty & \text{si } x > 1 \end{cases}$$

De même, $V \in \mathcal{RV}_{-\infty}$ si

$$\lim_{x \rightarrow \infty} \frac{V(tx)}{V(x)} = x^{-\infty} = \begin{cases} \infty & \text{si } x < 1 \\ 1 & \text{si } x = 1 \\ 0 & \text{si } x > 1 \end{cases}$$

Un exemple élémentaire d'une fonction à variation rapide, $V \in \mathcal{RV}_{-\infty}$ est $V(x) = \exp(-x)$ qui est bien la queue de distributions d'une exponentielle standard. Le lecteur peut consulter les ouvrages de Resnick [64] et Embrechts et al. [34] pour une discussion sur la liste des propriétés des fonctions à variation rapide.

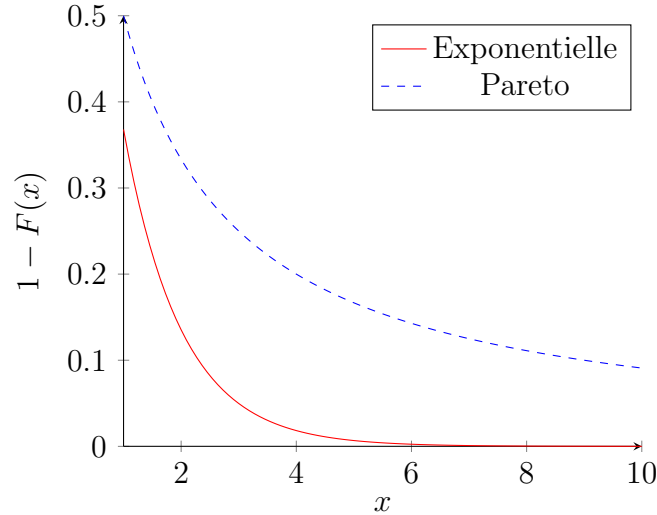


FIGURE 2.2 – Comparaison du comportement des queues : distribution exponentielle (trait plein), distribution de Pareto (trait pointillé).

Conditions du Premier et du Second ordre

Dans le cadre des modèles à queue lourde où les conditions équivalentes suivantes sont satisfaites :

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-1/\gamma} \iff \lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma \quad (2.11)$$

pour tout $x > 0$, avec U est la fonction de queue de quantile, l'équation 2.11 est connue sous le nom de la condition du premier ordre des fonctions à variations régulières avec l'indice $-1/\gamma$ (ou l'indice γ) et $\gamma > 0$, ce qui signifie $\bar{F} \in \mathcal{RV}_{-1/\gamma}$ ou $U \in \mathcal{RV}_\gamma$. Cependant, l'étude des propriétés des estimateurs des paramètres de la queue d'une distribution nécessite en général plus de conditions que celle du premier ordre, en particulier l'étude de la normalité asymptotique. Dans ce cas, une condition du second ordre des fonctions à variations régulières est nécessaire en spécifiant le taux de convergence dans l'équation 2.11. Les premiers qui ont donné la définition de cette condition sont de Haan et Stadtmaller [28], et reprise par Geluk et al. [37], de Haan et Ferreira [27], et Resnick [64], Neves [59] et Neves et al. [7] pour une version légèrement différente du second ordre étendu des fonctions à variations régulières. On trouve la définition de la condition du troisième ordre dans Gomes et al. [42], Fraga Alves et al. ([6],[5]) et celle d'ordre supérieur dans Wang et Cheng [74].

Définition 2.11 (Condition du second ordre)

On dit que la fonction de queue de U est à variation régulière du second ordre avec le paramètre du premier ordre $\gamma > 0$ et le paramètre du second ordre $\rho \leq 0$, on écrit $U \in 2\mathcal{RV}_{\gamma,\rho}$, s'il existe une fonction $A^*(t) \rightarrow 0$ qui ne change pas de signe au voisinage

de ∞ , telle que

$$\lim_{t \rightarrow \infty} \frac{U(tx)/U(t) - x^\gamma}{A^*(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, \quad x > 0 \quad (2.12)$$

où $|A^*| \in \mathcal{RV}_\rho$ est appelée la fonction auxiliaire de U .

Le Corollaire suivant exprime la condition du second ordre des fonctions à variations régulières en fonction de \bar{F} .

Corollaire 2.1

Pour tout $x > 0$ avec $\rho \leq 0$ et $A(t) = A^*(1/(1 - F(t)))$, la relation 2.12 est équivalente à

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)/\bar{F}(t) - x^{-1/\gamma}}{A(t)} = x^{-1/\gamma} \frac{x^\rho - 1}{\gamma\rho} \quad (2.13)$$

Preuve

Voir la preuve du Lemme de Vervaat [27] qui est une version générale du Corollaire 2.1.

Remarque 2.3

Les équations 2.12 et 2.13 sont aussi équivalentes à :

$$\lim_{t \rightarrow \infty} \frac{\log(\bar{F}(tx)) - \log(\bar{F}(t)) + \gamma^{-1} \log x}{A^*(t)} = \frac{x^\rho - 1}{\rho},$$

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

Qui sont aussi équivalentes à

Il existe une constante réelle $\rho \leq 0$ et une fonction $A^{**}(s) \rightarrow 0$ quand $s \rightarrow 0$, telles que pour tout $x > 0$

$$\lim_{s \rightarrow 0} \frac{(Q(1 - sx))/Q(1 - s) - x^{-\gamma}}{A^{**}(s)} = x^{-\gamma} \frac{x^\rho - 1}{\rho}$$

$|A^{**}|$ est à variation régulière d'indice $-\rho$ en zéro et $A^{**}(s) = A(1/s)$.

Remarque 2.4

Les conditions du second ordre sont satisfaites pour la plupart des lois appartenant au domaine d'attraction de Fréchet (Cas des distribution à queues lourdes).

Exemple 2.4 (Classe de Hall)

Les conditions du second ordre ont été plus développées dans le cas des modèles de Hall (ou classe de Hall) qui est une classe composée de l'ensemble des distributions F telles que

$$F(x) = 1 - cx^{-1/\gamma} (1 + dx^{\rho/\gamma} + o(x^{\rho/\gamma})) \quad \text{quand } x \rightarrow \infty \quad (2.14)$$

où $\gamma > 0, \rho \leq 0, c > 0$, et $d \in \mathbb{R}^*$.

Cette sous classe des distributions à queues lourdes (de type Pareto) joue un rôle important dans la discussion des estimateurs d'un indice de valeurs extrêmes positif γ . Elle contient les lois de Pareto, Burr, Fréchet et t-Student qui sont fréquemment utilisées en assurance comme des modèles mathématique pour des grands risques.

La relation 2.14 peut être écrite en utilisant la fonction quantile de queue $U(\cdot)$ de la manière suivante

$$U(t) = c^\gamma t^\gamma (1 + \gamma d c^\rho t^\rho + o(t^\rho)) \text{ quand } t \rightarrow \infty$$

et peut être aussi reformulée avec la fonction quantile $Q(\cdot)$

$$Q(1 - s) = c^\gamma s^{-\gamma} (1 + \gamma d c^\rho s^{-\rho} + o(s^{-\rho})) \text{ quand } s \rightarrow 0$$

Ainsi, pour la classe de Hall, les fonctions $A(\cdot)$ et $A^*(\cdot)$ sont respectivement équitantes à $d\rho\gamma c^\rho t^\rho$ et $d\rho\gamma t^{\rho/\gamma}$ lorsque $t \rightarrow \infty$.

Malgré la difficulté théorique de l'estimation du paramètre du second ordre ρ , plusieurs estimateurs qui fonctionnent bien dans la pratique ont été récemment introduits ainsi que leurs propriétés asymptotiques. On se réfère ici à Gomes et al. [42], Fraga Alves et al. [6] et Caeiro et Gomes [19]. Ciuperca et Mercadier [23] ont fait des généralisations des estimateurs dans [42] et [6]. Goegebeur et al. [40] ont introduit une nouvelle classe d'estimateurs de noyau basées sur l'échelle log-excesses. Nous pouvons citer aussi des estimateurs proposés par Wet et al. [30], ainsi que par Caeiro et Gomes [20] qui ont utilisé dans leur estimation la méthode de la réduction du biais. Worms et Worms [78] ont introduit un estimateur de moment de probabilité pondérée ρ . La classe des estimateurs semi paramétriques du paramètre du second ordre ρ a été traité dans certains livres de référence dans le domaine de la statistique des extrêmes (voir Beirlant et al. [10], de Haan et Ferreira [27], Resnick [64], ...).

Exemple 2.5 (Classe de Hall et Weiss)

Supposons que pour $x \leq 1, \alpha > 0, \rho < 0$,

$$1 - F(x) = \frac{1}{2} x^{-\alpha} (1 + x^\rho).$$

Alors pour $t \rightarrow \infty$

$$\begin{aligned} \frac{1 - F(tx)}{1 - F(t)} - x^{-\alpha} &= x^{-\alpha} \left\{ \frac{1 + (tx)^\rho}{1 + t^\rho} - 1 \right\} \\ &\sim x^{-\alpha} (x^\rho - 1) t^\rho, \end{aligned}$$

et ainsi nous pouvons définir $A(t) = \rho t^\rho$ et

$$\lim_{t \rightarrow \infty} \frac{1 - F(t)}{|A(t)|} = \begin{cases} 0 & \text{si } |\rho| < \alpha, \\ |\rho|^{-1} & \text{si } |\rho| = \alpha, \\ \infty & \text{si } |\rho| > \alpha. \end{cases}$$

Exemple 2.6 (Distribution de Burr)

On considère la distribution de Burr (β, τ, λ) , donnée par

$$F(x) = 1 - \left(\frac{\beta}{\beta + x^\tau} \right)^\lambda, \quad x > 0, \quad \beta, \tau, \lambda > 0$$

La fonction de queue de quantile est

$$U(t) = \beta^{1/\tau} t^{1/\tau\lambda} - \tau^{-1} \beta^{1/\tau} t^{1/(\tau\lambda)-1/\lambda} (1 + o(1)).$$

La distribution de Burr est donc bien un élément de la classe de Hall. Par conséquent, $U \in 2\mathcal{RV}_{\gamma, \theta}$ avec $\gamma = 1/\tau\lambda$, $\rho = -1/\lambda$ et la fonction auxiliaire $A^*(t) = (\lambda\tau)^{-1} t^{-1/\lambda}$

Classe B : Comportement de Pareto avec $\alpha > 0$

C'est la classe de distributions de qui ont un comportement de Pareto dont la répartition est donnée par

$$F(x) = 1 - \left(\frac{u}{x} \right)^\alpha, \quad x \geq u \text{ et } u > 0. \quad (2.15)$$

D'où

$$\bar{F}(x) = \left(\frac{u}{x} \right)^\alpha, \quad x \geq u \text{ et } u > 0.$$

Pour ce type de distributions, les moments d'ordre k ne sont finis que si $k < \alpha$ car

$$\begin{aligned} \mathbb{E}(X^k) &= \int_{\mathbb{R}} x^k dF(x) = \int_{\mathbb{R}} x^k d \left(1 - \left(\frac{u}{x} \right)^\alpha \mathbb{1}_{[u, \infty[}(x) \right) \\ &= \alpha u^\alpha \int_u^\infty x^{k-\alpha-1} dx. \end{aligned} \quad (2.16)$$

Classe A : Distributions α -Stables

Cette classe appelée classe des distributions α -Stables ou stable tout court, a été caractérisée par Lévy (1925) [53] dans ces travaux sur la somme de variables indépendantes et identiquement distribuées. Le comportement asymptotique de cette classe de distributions est de Pareto avec $0 < \alpha \leq 2$ avec comme cas particulier la loi Normale pour $\alpha = 2$. Par contre pour $\alpha < 2$ les moments d'ordres strictement supérieurs à α de cette classe de distributions ne sont pas finis et en particulier leurs variances ce qui veut dire qu'elles ont des queues très lourdes. Les distributions stables peuvent être caractérisées à partir du Théorème de la Limite Centrale Généralisé qui est une généralisation du Théorème de la Limite Centrale et qui indique que le comportement asymptotique de la somme de variables aléatoires indépendantes et identiquement distribuées de variance finie est normal (La somme à une distribution Normale). Le Théorème de la Limite Centrale Généralisé vient pour répondre au cas où la variance est infinie et c'est Gnedenko et Kolmogorov (1954)[39] qui ont fait cette généralisation et ont conclu que si la condition de variance

finie n'est pas assurée, la seule loi limite possible de la somme de n variables aléatoires (iid), est une loi stable. Cette classe de distributions a trouvée sa place avec l'évolution de l'outil informatique car elle avait un problème qui est la non explicité de ses densités et de ses distributions, cette inexplicité a causé dans le passé des problèmes de calcul et une limitation d'utilisation malgré la richesse de cette classe par des lois de probabilité pouvant représenter différentes asymétries et des queues très lourdes.

2.2 Caractérisation des lois extrêmes

La caractérisation se fait suivant les domaines d'attraction, mais il existe une caractérisation générale qui englobe tout ces domaines. Nous commençons par le cas spécifique pour chaque domaine puis nous donnons la généralisation.

On réintroduit la fonction U défini par

$$U(t) = F^{-1} \left(1 - \frac{1}{t} \right), t > 1,$$

où F^{-1} est l'inverse généralisé de F .

2.2.1 Domaine d'attraction de Fréchet

On définit également la fonction de quantile de queue par $U(x) = Q(1 - 1/x), x > 1$. Le résultat ci-dessous énoncé dans Gnedenko (1943) [38] et dont on trouvera une démonstration dans Resnick (1987) [63], Proposition 1.13, assure que toute fonction appartenant au domaine d'attraction de Fréchet est une fonction à variation régulière.

Théorème 2.5

La fonction de distribution F ayant pour point terminal x_F appartient au domaine d'attraction de Fréchet $D(\mathcal{H}_\gamma), \gamma > 0$ si et seulement si $x_F = +\infty$ et \bar{F} est à variation régulière d'indice $-1/\gamma$ à l'infini ie.

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = x^{-1/\gamma}$$

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

$$a_n = F^{-1}(1 - 1/n) = U(n) \text{ et } b_n = 0, \forall n > 0$$

2.2.2 Domaine d'attraction de Weibull

Le résultat suivant (voir Gnedenko (1943) [38] ou Resnick (1987) [63], 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 répartition.

Théorème 2.6

Une fonction de distribution F avec un point terminal x_F appartient au domaine d'attraction de Weibull $D(\mathcal{H}_\gamma)$, $\gamma < 0$ si et seulement si x_F est fini et la fonction de répartition F^* définie par

$$F^*(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ F(x_F - 1/x) & \text{si } x > 0 \end{cases}$$

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

$$a_n = x_F - F^{-1}(1 - 1/n) \text{ et } b_n = x_F$$

2.2.3 Domaine d'attraction 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, permet de décrire que $F \in D(\mathcal{H}_\gamma)$, $\gamma = 0$

Théorème 2.7

La fonction de distribution F avec un point terminal x_F appartient au domaine d'attraction de Gumbel $D(\mathcal{H}_\gamma)$, $\gamma = 0$ si et seulement si il existe un réel z , $z < x_F \leq \infty$ tel que

$$\bar{F}(x) = c(x) \exp\left(-\int_z^x \frac{g(t)}{a(t)} dt\right), z < x < x_F$$

où $c(\cdot)$ et $g(\cdot)$ sont des fonctions mesurables positives satisfaisant $c(x) \rightarrow c > 0$, $g(x) \rightarrow 1$ lorsque $x \uparrow x_F$, et $a(\cdot)$ est une fonction positive absolument continue (par rapport à la mesure de Lebesgue), de densité $a'(\cdot)$ avec $\lim_{x \uparrow x_F} a'(x) = 0$. Dans ce cas, un choix possible pour les suites a_n et b_n est $a_n = F^{\leftarrow}(1 - 1/n) = U(n)$ et $b_n = a(a_n)$. Un choix possible de la fonction $a(\cdot)$ est

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} dt, x < x_F$$

La fonction $a(\cdot)$ est appelée fonction auxiliaire.

2.2.4 Caractérisation générale

Les calculs de la démonstration du Théorème 1.1.1 nous fait la suggestion que si $F \in D(H(\xi))$, on a alors

$$\lim_{s \rightarrow +\infty} \frac{U(sw) - U(s)}{a(s)} = \frac{w^\xi - 1}{\xi}$$

à un changement d'échelle près. Si, en particulier, $x, y > 0$ et $y \neq 1$, on a

$$\lim_{s \rightarrow +\infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \lim_{s \rightarrow +\infty} \frac{U(sx) - U(s)}{a(s)} \frac{a(s)}{U(sy) - U(s)}$$

$$\begin{aligned} \lim_{s \rightarrow +\infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} &= \lim_{s \rightarrow +\infty} \frac{U(sx) - U(s)}{a(s)} \frac{a(s)}{U(sy) - U(s)} \\ &= \begin{cases} \frac{x^\xi - 1}{y^\xi - 1} & \text{si } \xi \neq 0, \\ \frac{\ln(x)}{\ln(y)} & \text{si } \xi = 0. \end{cases} \end{aligned}$$

En fait la Proposition suivante, voir [34] Théorème 3.4.5 pour la démonstration, assure que cette condition est suffisante pour que $F \in D(H(\xi))$.

Proposition 2.5

Soit $\gamma \in \mathbb{R}$. Il y a équivalence entre $F \in D(H(\gamma))$ et pour tout $x > 0, y > 0, y \neq 1$,

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\gamma - 1}{y^\gamma - 1} & \text{si } \gamma \neq 0, \\ \frac{\ln(x)}{\ln(y)} & \text{si } \gamma = 0. \end{cases}$$

Nous utiliserons ce résultat pour construire l'estimateur de Pickands de l'indice de queue au paragraphe 3.3.

Exemple 2.7

En prenant la loi exponentielle de paramètre $\lambda > 0$, dont la densité est donnée par

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

et de fonction de répartition F :

$$F(x) = \begin{cases} 1 - \exp(-\lambda x) & \text{si } x \geq 0, \\ 0 & \text{si } x < 0. \end{cases}$$

avec

$$F^{-1}(y) = \frac{\ln(1-y)}{-\lambda}, y \in [0, 1[.$$

En calculant la fonction U associée, pour $t > 1$

$$\begin{aligned} U(t) &= \frac{\ln(1 - (1 - \frac{1}{t}))}{-\lambda} \\ &= \frac{-1}{\lambda} \ln(\frac{1}{t}) \\ &= \frac{\ln(t)}{\lambda}. \end{aligned}$$

Calculons la limite :

$$\begin{aligned} \lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} &= \lim_{s \rightarrow \infty} \frac{\frac{\ln(sx)}{\lambda} - \frac{\ln(s)}{\lambda}}{\frac{\ln(sy)}{\lambda} - \frac{\ln(s)}{\lambda}} \\ &= \lim_{s \rightarrow \infty} \frac{\ln(sx) - \ln(s)}{\ln(sy) - \ln(s)} \\ &= \frac{\ln(x)}{\ln(y)}. \end{aligned}$$

ça veut dire que $F \in D(H(0))$, qui est le domaine d'attraction de Gumbel.

Exemple 2.8

Prenons la loi uniforme sur $[0, \theta]$, sa densité est donnée par

$$f(x) = \begin{cases} \frac{1}{\theta} & \text{si } x \in [0, \theta], \\ 0 & \text{si } x \notin [0, \theta]. \end{cases}$$

Sa répartition est

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{\theta} & \text{si } x \in [0, \theta], \\ 1 & \text{si } x > \theta. \end{cases}$$

Avec

$$F^{-1}(y) = \theta y, y \in [0, 1].$$

D'où, pour $t > 1$

$$U(t) = \theta \left(1 - \frac{1}{t}\right),$$

et

$$\begin{aligned} \lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} &= \lim_{s \rightarrow \infty} \frac{\theta \left(1 - \frac{1}{sx}\right) - \theta \left(1 - \frac{1}{s}\right)}{\theta \left(1 - \frac{1}{sy}\right) - \theta \left(1 - \frac{1}{s}\right)} \\ &= \lim_{s \rightarrow \infty} \frac{1 - \frac{1}{x}}{1 - \frac{1}{y}} \\ &= \frac{x^{-1} - 1}{y^{-1} - 1}, \end{aligned}$$

ça veut dire que $F \in D(H(-1))$, qui est le domaine d'attraction de Weibull.

Exemple 2.9

Soit la loi de Cauchy standard dont la densité est

$$f(x) = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R}.$$

Sa répartition est

$$F(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2},$$

et

$$F^{-1}(y) = \tan \left(\pi \left(y - \frac{1}{2} \right) \right), y \in [0, 1].$$

D'où, pour $t > 1$

$$\begin{aligned} U(t) &= \tan \left(\pi \left(\frac{1}{2} - \frac{1}{t} \right) \right), \\ &= \frac{1}{\tan \left(\frac{\pi}{t} \right)}, \end{aligned}$$

et

$$\begin{aligned} \lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} &= \lim_{s \rightarrow \infty} \frac{\frac{1}{\tan\left(\frac{\pi}{sx}\right)} - \frac{1}{\tan\left(\frac{\pi}{s}\right)}}{\frac{1}{\tan\left(\frac{\pi}{sy}\right)} - \frac{1}{\tan\left(\frac{\pi}{s}\right)}} \\ &= \frac{x-1}{y-1}, \end{aligned}$$

ça veut dire que $F \in D(H(1))$, qui est le domaine d'attraction de Fréchet.

Chapitre 3

Estimation d'un quantile et indice de queue :

Sommaire

3.1	Estimation d'un quantile : cas classique	55
3.2	Estimation d'un quantile : cas extrême	56
3.2.1	Approximation d'un quantile extrême	57
	Première méthode : Loi des valeurs extrêmes	57
	Deuxième méthode : Loi de Pareto Généralisée	57
	Cas des distributions à queues lourdes, domaine d'attraction de Fréchet	58
3.2.2	Estimation d'un quantile extrême	59
	Estimation en utilisant l'approximation EVD	59
	Estimation en utilisant l'approximation <i>GPD</i>	59
	Estimation pour le domaine d'attraction de Fréchet	65
3.3	Etude pratique : maximas annuelles des températures	72

Résumé

Nous nous intéressons dans ce chapitre au problème de l'estimation du quantile et en particulier le quantile extrême. Soient donc à considérer n variables aléatoires indépendantes X_1, \dots, X_n et de même fonction de répartition F ainsi qu'une suite (α_n) convergeant vers 0 lorsque n tend vers l'infini. Notre but est d'estimer le quantile extrême $q(\alpha_n)$ d'ordre α_n . Le premier obstacle qui nous rencontre dans notre quête est que la valeur $q(\alpha_n)$ se trouve dans la queue de F ce qui fait qu'elle ne peut pas être estimée par l'estimateur empirique classique \hat{F}_n . Pour résoudre ce problème, il est nécessaire de passer par l'approximations

de la quantité $q(\alpha)$ lorsque $\alpha \rightarrow 0$, et à partir de ces dernières faire des estimations par différentes méthodes.

Théoriquement, il y a trois types de quantiles en fonction de leur ordre α_n où n est la taille de l'échantillon. La définition suivante nous donne ces trois types.

Définition 3.1

Les trois catégories de quantiles selon leur ordre α_n sont :

1. Un quantile classique si $n\alpha_n \rightarrow \infty$.
2. Un quantile intermédiaire si $n\alpha_n \rightarrow \alpha \in [1, \infty[$.
3. Un quantile extrême si $n\alpha_n \rightarrow \alpha \in [0, 1[$.

Remarque 3.1

Nous pouvons étudier le cas classique et intermédiaire en même temps car dans les deux cas nous pouvons estimer les quantiles en se servant de la fonction de survie empirique.

Commençons à présenter le cas classique.

3.1 Estimation d'un quantile : cas classique

Nous considérons que nous sommes dans le cas classique dans l'estimation du quantile $q(\alpha_n)$ si $n\alpha_n \rightarrow \infty$ où α_n est une suite d'éléments de $]0, 1[$. Nous rappelons que si $\alpha \in [0, 1]$, $q(\alpha) = F^{-1}(1 - \alpha) = \overline{F}^{-1}(\alpha)$.

On se donne un échantillon de taille n pour une variable aléatoire X . la statistique d'ordre associée est donnée par $X_{1,n} \leq \dots \leq X_{n,n}$. Un estimateur naturel du quantile dans ce cas est donné par l'inversion de la fonction de répartition empirique F_n , il est défini comme suit

$$\hat{q}_n(\alpha_n) = \hat{F}_n^{-1}(1 - \alpha_n) = X_{n - \lfloor n\alpha_n \rfloor, n},$$

où $\lfloor x \rfloor$ est la partie entière (par défaut) de x .

Et on a le théorème suivant qui nous donne un résultat sur le comportement asymptotique de cet estimateur.

Théorème 3.1

Soient X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F . Soit α_n une suite dans $]0, 1[$ telle que $\alpha_n \rightarrow \alpha \in [0, 1[$ et $n\alpha_n \rightarrow \infty$. Si la fonction $q = \overline{F}^{-1}$ est dérivable de dérivée q' et si la fonction $|q'(1/\cdot)|$ est à variations régulières alors

$$\frac{\sqrt{n}}{\sqrt{\alpha_n} q'(\alpha_n)} (q(\alpha_n) - \hat{q}_n(\alpha_n)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1 - \alpha) \quad (3.1)$$

Exemple 3.1

Soit un échantillon de loi exponentielle standard. On a $\bar{F}(x) = \exp(-x)$ donc $q(\alpha) = \ln(1/\alpha)$ qui est une fonction dérivable de dérivée $|q'(\alpha)| = \alpha$ dont elle est à variation régulière d'indice 1. La figure 3.1 illustre la normalité asymptotique du quantile empirique en prenant comme suite

$$\alpha_n = \frac{\lfloor n\alpha \rfloor}{n},$$

ce qui fait que l'équation (3.1) devient

$$\sqrt{\lfloor n\alpha \rfloor} (X_{n-\lfloor n\alpha \rfloor:n} - \ln(n/\lfloor n\alpha \rfloor)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1 - \alpha) \quad (3.2)$$

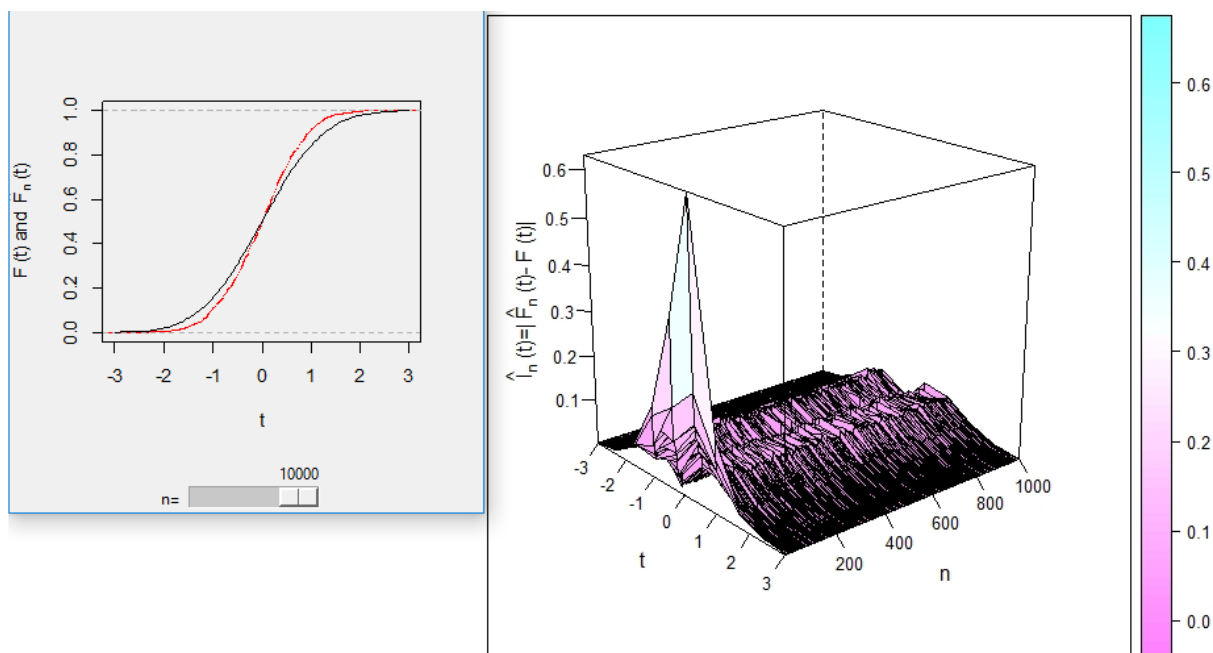


FIGURE 3.1 – Convergence en loi du quantile d'une exponentielle standard vers une loi normale en générant 1000 échantillons de tailles 10000. A gauche de la figure : la courbe de la distribution empirique et de celle de la théorique. A droite : l'absolu erreur entre la distribution empirique et la théorique

3.2 Estimation d'un quantile : cas extrême

Considérons maintenant le cas extrême, c'est-à-dire que la suite (α_n) converge vers 0 quand n tend vers l'infini. Soient X_1, \dots, X_n des variables aléatoires indépendantes et de même fonction de répartition F . Le but de ce paragraphe est de donner une estimation du quantile extrême d'ordre α_n qui est exprimé par

$$q(\alpha_n) = \bar{F}^{-1}(\alpha_n).$$

Le problème majeur dans ce cas, est qu'on ne peut pas utiliser l'estimateur empirique classique \widehat{F}_n de F car la valeur $q(\alpha_n)$ est située dans la queue de distribution de F .

Nous commençons par donner des approximations possibles de quantile extrême, puis donner des estimations de $q(\alpha)$ quand $\alpha \rightarrow 0$.

3.2.1 Approximation d'un quantile extrême

Nous allons introduire dans ce paragraphe deux méthodes de base pour approcher $q(\alpha)$ quand α tend vers 0. La première méthode utilise la loi des valeurs extrêmes par contre la deuxième méthode utilise la loi de Pareto Généralisée.

Première méthode : Loi des valeurs extrêmes

Nous savons d'après le théorème fondamental de la théorie des valeurs extrêmes que si une fonction de répartition F appartient à un domaine d'attraction alors il existe deux suites $(a_n) > 0$ et (b_n) telles que

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H_\gamma(x) = \begin{cases} \exp\left(- (1 + \gamma x)_+^{-1/\gamma}\right) & \text{si } \gamma \neq 0, \\ \exp(-e^{-x}) & \text{si } \gamma = 0. \end{cases}$$

Ce qui est équivalent à

$$\lim_{n \rightarrow \infty} n \overline{F}(a_n x + b_n) = -\ln H_\gamma(x)$$

Ainsi, lorsque $y = a_n x + b_n$ s'approche du point terminal x_F de F , ($\lim_{n \rightarrow \infty} a_n x + b_n = x_F$), alors nous pouvons faire l'approximation suivante

$$\overline{F}(y) \approx \frac{-1}{n} \ln H_\gamma\left(\frac{y - b_n}{a_n}\right) = \frac{1}{n} \left(1 + \gamma \frac{y - b_n}{a_n}\right)_+^{-1/\gamma} \quad (3.3)$$

qui a un sens par passage à la limite lorsque $\gamma = 0$. En inversant (3.3), nous aurons une approximation de $q(\alpha)$ lorsque α voisine le 0 donnée par la formule suivante

$$q(\alpha) \approx b_n + \frac{a_n}{\gamma} ((n\alpha)^{-\gamma} - 1) \quad (3.4)$$

Deuxième méthode : Loi de Pareto Généralisée

Le théorème de Pickands nous assure que pour un seuil u voisinant le point terminal x_F de F , la fonction de répartition des dépassements pour $x > 0$ définie par

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

peut être approchée par la loi de Pareto Généralisée dont sa fonction de répartition est donnée par

$$\mathcal{G}_{\gamma, a(u)}(x) = 1 + \ln\left(H_\gamma\left(\frac{x}{a(u)}\right)\right),$$

où a est une fonction positive.

En faisant le changement de variable $y = x + u$ et en tendant y vers le point terminal de F , nous pouvons avoir l'approximation

$$\bar{F}(y) \approx -\bar{F}(u) \ln \left(H_\gamma \left(\frac{y-u}{a(u)} \right) \right). \quad (3.5)$$

En procédant de la même manière que précédemment pour donner une approximation de $q(\alpha)$ lorsque α voisine le 0 et ça en inversant (3.5). Nous trouvons

$$q(\alpha) \approx u + \frac{a(u)}{\gamma} \left(\left(\frac{\alpha}{\bar{F}(u)} \right)^{-\gamma} - 1 \right) \quad (3.6)$$

Lorsque l'on suppose $\gamma = 0$, deux variantes de cette méthode ont été présentées par Breiman et al. [18] en 1990 sous les appellations Exponential Tail (ET) et Quadratique Tail (QT).

Remarque 3.2

On peut remarquer qu'il y a une similitude entre l'approximation (3.6) et (3.4) et ça en prenant $n = 1/\bar{F}(u) \rightarrow \infty$ quand $u \rightarrow x_F$, $b_n = u$ et $a_n = a(u)$. Malgré cette similitude, le choix des méthodes d'estimation des paramètres inconnus dans les deux approximations est différent.

Cas des distributions à queues lourdes, domaine d'attraction de Fréchet

Le domaine d'attraction de Fréchet est très utilisé dans la modélisations des événements rares. Nous nous intéressons ici à l'estimation de quantiles extrêmes pour des lois dont on sait qu'elles appartiennent au domaine d'attraction de Fréchet. On rappelle que dans ce cas la fonction de répartition est de la forme

$$F(x) = x^{-1/\gamma} L(x),$$

avec $\gamma > 0$ et L une fonction à variations lentes. où d'une façon équivalente, \bar{F} est une fonction à variations régulières d'indice $-1/\gamma$. Nous savons aussi d'après la Proposition 2.2 que le quantile associé est donné par

$$q(\alpha) = \alpha^{-\gamma} \ell(\alpha^{-1})$$

où ℓ est une autre fonction à variations lentes. Pour estimer $q(\alpha)$ il suffit donc d'estimer γ et la fonction ℓ . En pratique, il est difficile d'estimer la fonction à variations lentes ℓ . Pour surpasser cet obstacle, on utilise l'idée suivante : pour $t > 1$,

$$\frac{q(\alpha)}{q(t\alpha)} = t^\gamma \frac{\ell(\alpha^{-1})}{\ell(t^{-1}\alpha^{-1})},$$

et comme ℓ est une fonction à variations lentes, on obtient l'approximation

$$q(\alpha) \approx q(t\alpha)t^\gamma \text{ pour } \alpha \text{ proche de } 0.$$

Le quantile $q(\alpha)$ est fonction de γ et $q(t\alpha)$. On a besoin donc d'estimer γ et $q(t\alpha)$ ce qui nous mène à faire un déplacement sur l'axe des abscisses en passant des valeurs α de la fonction $q(\alpha)$ vers les valeurs pour $t\alpha$ de cette fonction. L'astuce est de faire tendre t vers l'infini (suffisamment grand) de telle sorte que $t\alpha$ soit un ordre de quantile classique et de cette manière on peut l'estimer par la fonction de répartition empirique.

3.2.2 Estimation d'un quantile extrême

L'estimation d'un quantile extrême se fait en se servant de l'une des trois approximations présentées dans le paragraphe précédent. Metons nous dans la situation d'un quantile extrême $q(\alpha_n)$ que nous allons estimer à partir d'un échantillon X_1, \dots, X_n .

Estimation en utilisant l'approximation EVD

La base de cette méthode est l'approximation

$$\mathbf{P} [\max (X_1, \dots, X_n) \leq x] \approx H_\gamma \left(\frac{x - b_n}{a_n} \right)$$

Cette approximation devient de mieux en mieux meilleure avec l'augmentation de la taille n de l'échantillon. Les estimations des paramètres a_n, b_n et γ se font à l'aide d'un échantillon de maxima de n variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F en procédant de la manière suivante :

1. Choisir une valeur $k \in \{1, \dots, n\}$ qui soit assez grande pour utiliser l'approximation.
2. On divise l'échantillon X_1, \dots, X_n en $m = \lfloor n/k \rfloor$ sous-échantillons de taille k .
3. On calcul les maximums des m échantillons construits.
4. On suppose que les m maxima sont i.i.d. et leur loi est une loi EVD.
5. On estime les paramètres de loi des valeurs extrêmes supposée qui sont a_k, b_k et γ .

Cette méthode est dite la méthode des "maxima par blocs". C'est une méthode qui nécessite un nombre assez grand pour que l'étape (4) soit vérifiée.

Estimation en utilisant l'approximation GPD

On sait que pour un seuil u proche du point terminal de F , on peut approcher $q(\alpha_n)$ avec $\alpha_n \rightarrow 0$ par

$$u + \frac{a(u)}{\gamma} \left[\left(\frac{\alpha_n}{\bar{F}(u)} \right)^{-\gamma} - 1 \right]$$

Donc pour estimer $q(\alpha_n)$ on procédera de la manière suivante :

1. Choisir un seuil u suffisamment grand
2. Estimer les paramètres γ et $a(u)$ de la loi de Pareto Généralisée suivie approximativement par les excès au dessus du seuil u . (Pratiquement, on choisit u de telle sorte que $\bar{F}(u) = k_n/n$ (i.e $u = q(k_n/n)$) où $k_n \in \{1, \dots, n\}$ est une suite intermédiaire, c'est-à-dire telle que $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$)
3. On prend pour seuil $\hat{u}_n = \hat{F}_n^{-1}(k_n/n) = X_{n-k_n, n}$.
4. On construit les k_n variables aléatoires des excès au dessus de \hat{u}_n

$$Y_1 = X_{n-k_n+1, n} - X_{n-k_n, n}, \dots, Y_{k_n} = X_{n, n} - X_{n-k_n, n}.$$

5. On suppose que les variables Y_1, \dots, Y_{k_n} sont i.i.d de loi Pareto généralisée de paramètres γ et $a(u)$ (Théorème de Pickands).
6. On estime γ et $a(k_n)$ obtenus à partir de l'échantillon des excès par $\hat{\gamma}(k_n)$ et $\hat{a}(k_n)$ respectivement.
7. On estime $q(\alpha_n)$ par

$$\hat{q}_n(\alpha_n) = X_{n-k_n+1, n} + \frac{\hat{a}(k_n)}{\hat{\gamma}(k_n)} \left[\left(\frac{n\alpha_n}{k_n} \right)^{-\hat{\gamma}(k_n)} - 1 \right]$$

Cette démarche repose essentiellement sur le choix du seuil à dépasser et l'estimation des paramètres de la loi *GPD*, ce qui nous pousse à donner les méthodes les plus d'importantes utilisées pour y parvenir. Nous citons pour l'estimation des paramètres de la loi *GPD* :

1. la méthode du maximum de vraisemblance,
2. la méthode des moments,
3. la méthode des moments pondérés,
4. la méthode des percentiles.

Considérons un échantillon de variables aléatoires X_1, \dots, X_n indépendantes et de même loi *GPD* de paramètres $\gamma \in \mathbb{R}$ et $\sigma > 0$, estimer γ et σ .

1. **Estimation des paramètres par la méthode du maximum de vraisemblance** : La fonction de log-vraisemblance obtenue à partir de la loi *GPD* $\mathcal{G}_{\gamma, \sigma}$ est donnée par

$$\ln(\mathcal{L}(\gamma, \sigma)) = -n \ln(\sigma) - \left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^n \ln\left(1 + \frac{\gamma}{\sigma} X_i\right),$$

avec $1 + \frac{\gamma}{\sigma} X_i > 0$ pour $i = 1, \dots, n$ sinon $\mathcal{L}(\gamma, \sigma) = -\infty$.

$$\begin{cases} \frac{\partial}{\partial \sigma} (\ln(\mathcal{L}(\gamma, \sigma))) &= -\frac{n}{\sigma} + \left(\sum_{i=1}^n \frac{1}{\sigma^2 + \sigma \gamma X_i} X_i \right) (\gamma + 1) &= 0 \\ \frac{\partial}{\partial \gamma} (\ln(\mathcal{L}(\gamma, \sigma))) &= -\left(1 + \frac{1}{\gamma}\right) \sum_{i=1}^n \frac{\frac{1}{\sigma} X_i}{\left(1 + \frac{\gamma}{\sigma} X_i\right)} + \frac{1}{\gamma^2} \sum_{i=1}^n \frac{\ln\left(1 + \frac{\gamma}{\sigma} X_i\right)}{\left(1 + \frac{\gamma}{\sigma} X_i\right)} &= 0 \end{cases}$$

En pratique, nous avons recours aux méthodes numériques pour résoudre ce système, ce qui veut dire qu'on ne peut pas avoir des expressions explicites des estimateurs des équations de vraisemblance. En considérons le cas $\gamma > -1/2$ les estimateurs du maximum de vraisemblance sont asymptotiquement gaussiens et efficaces, voir Smith [70]. Malgré cela, Davison et Smith [25] ont fait remarquer que l'estimateur du maximum de vraisemblance pose des problèmes numériques et il est peu performant sur des échantillons de petite taille ($n < 500$). Il vaut mieux donc utiliser autres méthodes que celle du maximum de vraisemblance. Parmi ces méthodes alternatives, nous introduisons les méthodes suivantes.

2. **Méthode des moments** Cette méthode a été introduite en 1987 par Hosking et Wallis [46] pour estimer les paramètres de la loi *GPD*. Les moments d'ordre un et deux existent pour une variable X de loi *GPD* $\mathcal{G}_{\gamma,\sigma}$ si $\gamma < 1/2$. Dans ce cas :

$$\mathbb{E}(X) = \frac{\sigma}{1-\gamma} \text{ et } \text{Var}(X) = \frac{\sigma^2}{(1-\gamma)^2(1-2\gamma)}.$$

Ce qui nous donne la possibilité d'exprimer les paramètres γ et σ de la loi $\mathcal{G}_{\gamma,\sigma}$ en fonction de l'espérance et la variance de X , comme suit :

$$\gamma = \frac{1}{2} \left(1 - \frac{\mathbb{E}(X)^2}{\text{Var}(X)} \right) \text{ et } \sigma = \frac{\mathbb{E}(X)}{2} \left(1 + \frac{\mathbb{E}(X)^2}{\text{Var}(X)} \right).$$

Il suffit maintenant de remplacer $\mathbb{E}(X)$ et $\text{Var}(X)$ par leurs estimateurs empiriques :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ et } S_{n-1}^2(X) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Ce qui nous donne les estimateurs des moments de γ et σ :

$$\hat{\gamma}_n^M = \frac{1}{2} \left(1 - \frac{\bar{X}^2}{S_{n-1}^2(X)} \right) \text{ et } \hat{\sigma}_n^M = \frac{\bar{X}}{2} \left(1 + \frac{\bar{X}^2}{S_{n-1}^2(X)} \right).$$

Ces estimateur ont un défaut majeur qui est d'avoir un domaine de validé très restreint car ils sont asymptotiquement gaussiens que si $\gamma < 1/4$ (voir [46]).

3. **Méthode des moments pondérés** C'est une amélioration de la méthode précédente qui a été proposée par les mêmes auteurs dans la même année, pour estimer les paramètres de la loi *GPD*. Cette méthode vient pour remédier au problème des échantillons de petites tailles, malheureusement elle souffre d'un inconvénient qui est de ne être utilisable que si l'indice des valeurs extrêmes γ est inférieur (strictement) à 1. Donnons dans un premier temps la définition des moments pondérés.

Définition 3.2

Soit X une variable aléatoire de fonction de répartition F . Si X est intégrable, le moment pondéré d'ordres $r \in \mathbb{N}$ et $s \in \mathbb{N}$ de X est

$$\text{WM}_X(r, s) = \mathbb{E} [XF^r(X)(1 - F(X))^s]$$

Dans la suite, on prendra $r = 0$. Le Théorème suivant donne la valeur du moment pondéré d'une loi $GPD \mathcal{G}_{\gamma,\sigma}$.

Théorème 3.2

Soit X une variable aléatoire de loi de Pareto Généralisé de paramètres γ et σ . Si $\gamma < 1$, on a pour tout $s \in \mathbb{N}$

$$WM_X(0, s) = \frac{\sigma}{(s+1)(s+1-\gamma)}$$

Il est possible d'écrire γ et σ en fonction de deux moments pondérés d'ordres s différents et on aura :

Proposition 3.1

Pour $s \in \mathbb{N}$, en posant $\mu_s = WM_X(0, s)$, on a

$$\gamma = \frac{4\mu_1 - \mu_0}{2\mu_1 - \mu_0} \text{ et } \sigma = \frac{2\mu_0\mu_1}{\mu_0 - 2\mu_1} \quad (3.7)$$

Preuve

Il suffit de calculer μ_0 et μ_1 puis résoudre le système suivant

$$\begin{cases} \mu_0 &= \sigma/(1-\gamma) \\ \mu_1 &= \sigma/2(2-\gamma) \end{cases}$$

De la même façon que pour la méthode des moments classique, on propose d'estimer γ et σ en remplaçant dans (3.7) les moments μ_0 et μ_1 par des estimateurs adéquats. Nous devons donc estimer le moment pondéré

$$\mu_s = \mathbb{E} [X (1 - \mathcal{G}_{\gamma,\sigma}(X))^s]$$

où X est une variable aléatoire de loi de Pareto Généralisée de paramètres γ et σ . Pour ce faire, nous prenons un échantillon X_1, \dots, X_n de n copies indépendantes de X . En premier lieu, nous remplaçons l'espérance par l'espérance empirique. Nous obtenons

$$\tilde{\mu}_s = \frac{1}{n} \sum_{i=1}^n X_i [1 - \mathcal{G}_{\gamma,\sigma}(X_i)]^s = \frac{1}{n} \sum_{i=1}^n X_{i,n} [1 - \mathcal{G}_{\gamma,\sigma}(X_{i,n})]^s.$$

Le fait que la fonction de répartition $\mathcal{G}_{\gamma,\sigma}$ reste inconnue rend la variable aléatoire $\tilde{\mu}_s$ inutile pour être utilisée comme un estimateur de l'espérance. Pour remédier à ce problème, nous remplaçons dans l'expression de $\tilde{\mu}_s$ la variable aléatoire $1 - \mathcal{G}_{\gamma,\sigma}(X_{i,n})$ par son espérance. Cette espérance est calculée en utilisant le fait que pour tout $i = 1, \dots, n$ si U_1, \dots, U_n sont n variables aléatoires indépendantes de loi uniforme sur $[0,1]$ alors $\mathcal{G}_{\gamma,\sigma}(X_i) \stackrel{\mathcal{L}}{=} U_i$. Comme Les variables aléatoires sont indépendantes, nous pouvons dire que le résultat reste vrai pour le vecteur, c'est-à-dire

$$\{\mathcal{G}_{\gamma,\sigma}(X_i), i = 1, \dots, n\} \stackrel{\mathcal{L}}{=} \{U_i, i = 1, \dots, n\}$$

Une fonction de répartition étant croissante, on en déduit que

$$\{1 - \mathcal{G}_{\gamma,\sigma}(X_{i,n}), i = 1, \dots, n\} \stackrel{\mathcal{L}}{=} \{1 - U_{i,n}, i = 1, \dots, n\} \stackrel{\mathcal{L}}{=} \{U_{n-i+1,n}, i = 1, \dots, n\}$$

En utilisant la densité d'une statistique d'ordre pour une variable uniforme et son moment d'ordre s nous avons

$$\mathbb{E}(U_{n-i+1,n}^s) = \frac{(n-i+s)!}{(n-i)!} \frac{n!}{(n+s)!},$$

finalement nous estimons le moment pondéré μ_s par

$$\hat{\mu}_s = \frac{1}{n} \sum_{i=1}^n X_{i,n} \frac{(n-i+s)!}{(n-i)!} \frac{n!}{(n+s)!}.$$

4. Méthode des percentiles

Cette méthode a l'avantage, par rapport à la méthode des moments pondérés, d'être utilisable quelque soit la valeur de l'indice γ . Nous avons déjà vu précédemment que si nous disposons d'un échantillon X_1, \dots, X_n de variables aléatoires indépendantes de loi de Pareto Généralisée de paramètres $\gamma \in \mathbb{R}$ et $\sigma > 0$ alors

$$\{\bar{\mathcal{G}}_{\gamma,\sigma}(X_{i,n}), i = 1, \dots, n\} \stackrel{\mathcal{L}}{=} \{U_{n-i+1,n}, i = 1, \dots, n\}$$

où $\bar{\mathcal{G}}_{\gamma,\sigma} = 1 - \mathcal{G}_{\gamma,\sigma}$.

L'approximation suivante est la base de la méthode des percentiles

$$\bar{\mathcal{G}}_{\gamma,\sigma}(X_{i,n}) \approx \mathbb{E}[\bar{\mathcal{G}}_{\gamma,\sigma}(X_{i,n})] = \mathbb{E}[U_{n-i+1,n}] = 1 - \frac{i}{n+1} \quad (3.8)$$

En 1975, Pickands a fait la construction de son estimateur pour les paramètres d'une loi GPD $\mathcal{G}_{\gamma,\sigma}$ en procédant de la manière suivante, en premier lieu, il a formé un système en prenant dans l'approximation (3.8) : $i = n/2$ et $j = 3n/4$ où n est un multiple de 4.

$$\begin{cases} \bar{\mathcal{G}}_{\gamma,\sigma}(X_{n/2,n}) = (1 + \gamma X_{n/2,n}/\sigma)^{-1/\gamma} & \approx 1 - n/[2(n+1)] \approx 1/2 \\ \bar{\mathcal{G}}_{\gamma,\sigma}(X_{3n/4,n}) = (1 + \gamma X_{3n/4,n}/\sigma)^{-1/\gamma} & \approx 1 - 3k/[4(n+1)] \approx 1/4 \end{cases}$$

pour une valeur de n assez grande. Puis il a passé au logarithme ce qu'il lui a donné

$$\begin{cases} \ln(1 + \gamma X_{n/2,n}/\sigma) \approx \gamma \ln(2) \\ \ln(1 + \gamma X_{3n/4,n}/\sigma) \approx \gamma \ln(4) \end{cases} \quad (3.9)$$

A partir de ce système, il a fait la soustraction des deux équations, il a obtenu

$$\ln(1 + \gamma X_{3n/4,n}/\sigma) - \ln(1 + \gamma X_{n/2,n}/\sigma) \approx \gamma \ln(2)$$

puis

$$\gamma \approx \frac{1}{\ln(2)} [\ln(1 + \gamma X_{3n/4,n}/\sigma) - \ln(1 + \gamma X_{n/2,n}/\sigma)] \quad (3.10)$$

En divisant les deux équations du système (3.9) , il est arrivé à

$$\frac{\ln(1 + \gamma X_{3n/4,n}/\sigma)}{\ln(1 + \gamma X_{n/2,n}/\sigma)} \approx 2$$

l'approximation ci-dessus lui a permit d'écrire

$$(1 + \delta X_{n/2,n})^2 \approx (1 + \delta X_{3n/4,n})$$

avec $\delta = \gamma/\sigma$ qui est une équation du second degré en δ et dont les solutions sont 0 et

$$\delta \approx \frac{X_{3n/4,n} - 2X_{n/2,n}}{X_{n/2,n}^2} \quad (3.11)$$

En remplaçant (3.11) dans l'approximation (3.10), il a obtenu

$$\begin{aligned} \gamma &\approx \frac{1}{\ln(2)} \ln \frac{X_{n/2,n}^2 + X_{3n/4,n}^2 - 2X_{n/2,n}X_{3n/4,n}}{X_{n/2,n}^2 - 2X_{n/2,n}^2 + X_{n/2,n}^2 X_{3n/4,n}} \\ &\approx \frac{1}{\ln(2)} \ln \frac{(X_{3n/4,n} - X_{n/2,n})^2}{X_{n/2,n} [X_{3n/4,n} - X_{n/2,n}]} \\ &\approx \frac{1}{\ln(2)} \ln \frac{X_{3n/4,n} - X_{n/2,n}}{X_{n/2,n}} \end{aligned}$$

Finalement il est arrivé à estimer γ et σ par les statistiques

$$\hat{\gamma} = \frac{1}{\ln(2)} \ln \frac{X_{3n/4,n} - X_{n/2,n}}{X_{n/2,n}} \text{ et } \hat{\sigma} = \frac{\hat{\gamma} X_{n/2,n}^2}{X_{3n/4,n} - 2X_{n/2,n}}$$

Remarque 3.3

En pratique, d'estimateur de γ dont on vient de voir ce fait à partir des variables des excès Y_1, \dots, Y_{k_n} qui sont définies par

$$Y_i = Y_{i,k_n} = X_{n-k_n+i,n} - X_{n-k_n,n}$$

au lieu d'utiliser les variables aléatoires initiales X_1, \dots, X_n en négligeant le fait que les excès ne sont ni indépendants ni de loi GPD. En procédant de la sorte, l'estimateur de γ va s'écrire pour $1 \leq k_n < n$ comme suit

$$\hat{\gamma} = \frac{1}{\ln(2)} \ln \frac{X_{n-k_n+1,n} - X_{n-2k_n+1,n}}{X_{n-2k_n+1,n} - X_{n-4k_n,n}}.$$

Cet estimateur est dit Estimateur de Pickands. C'est un estimateur consistant de γ . Cette consistance a été démontrée par son auteur, par contre sa convergence forte ainsi que sa normalité asymptotique ont été établies par Dekkers et de Haan (1989) [31]. Nous avons le Théorème suivant qui donne ses propriétés.

Théorème 3.3

Soit $(k_n)_n \geq 1$ une suite d'entiers telle que $1 \leq k_n < n$, $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$ quand $n \rightarrow \infty$. Alors

- ✓ *Consistence faible* : $\hat{\gamma}_{k_n}^P$ converge en probabilité vers γ .
- ✓ *Consistence forte* : Si de plus $k_n/\ln \ln(n) \rightarrow \infty$ quand $n \rightarrow \infty$, $\hat{\gamma}_{k_n}^P$ converge presque sûrement vers γ .
- ✓ *Normalité asymptotique* : Sous des conditions additionnelles citées dans [31] sur la suite $(k_n)_n \geq 1$ et la fonction de répartition F ,

$$\sqrt{k_n} (\hat{\gamma}_{k_n}^P - \gamma) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\gamma^2(2^{2\gamma+1} + 1)}{4(\ln 2)^2(2^\gamma - 1)^2} \right).$$

Le paragraphe suivant traitera le cas très important quand trouve souvent dans l'étude des valeurs extrêmes et les phénomènes rares est celui des distributions à queues lourdes et dont leur domaine d'attraction et celui de Fréchet, c'est à dire celui où l'indice de queue γ est strictement positif.

Estimation pour le domaine d'attraction de Fréchet

Soit à considérer les variables aléatoires X_1, \dots, X_n dont elles sont indépendantes et de même fonction de répartition F et qui appartiennent au domaine d'attraction de Fréchet. Dans ce cas, le quantile extrême $q(\alpha_n)$ avec $\alpha_n \rightarrow 0$ est donné par

$$q(\alpha_n) \approx q(t\alpha_n) t^\gamma \tag{3.12}$$

pour $t > 1$. Pour utiliser cette approximation afin d'estimer le quantile extrême $q(\alpha_n)$, nous devons se ramener au cas classique. Pour y parvenir nous prenons un t assez grand pour que $t\alpha_n$ soit de l'ordre d'un quantile classique. Pratiquement, nous prenons $t = t_n = k_n/(n\alpha_n)$ où (k_n) est une suite intermédiaire. Nous obtenons

$$q(t\alpha_n) = q(k_n/n).$$

Comme nous sommes maintenant dans le cas d'un quantile classique, nous pouvons estimer le quantile en inversant la fonction de répartition empirique par la statistique d'ordre $X_{n-k_n, n}$. Nous donnons un estimateur de l'indice de queue γ que nous allons noter $\hat{\gamma}_n$ et en se servant de 3.12 nous pouvons estimer $q(\alpha_n)$ par

$$\hat{q}(\alpha_n) = X_{n-k_n, n} \left(\frac{k_n}{n\alpha_n} \right)^{\hat{\gamma}_n}$$

Cet estimateur est connu dans la littérature sous le nom d'estimateur de Weissman. Comme nous venons de mentionner, pour arriver à cet estimateur nous étions obligé

à donner un estimateur de l'indice de queue γ et ce qui nous pousse aussi à répondre à la question suivante : comment et quels sont les estimateurs que nous pouvons utiliser pour estimer γ ? Nous commençons pour répondre à cette question par l'estimateur incontestablement le plus connu et le plus utilisé pour estimer γ qui est l'estimateur de Hill.

Estimateur de Hill

Définition 3.3

Soient X_1, \dots, X_n des variables aléatoires indépendantes et de même fonction de répartition F appartenant au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes $\gamma > 0$. L'estimateur de Hill de γ est

$$\hat{\gamma}_n^{(H)}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln(X_{n-i+1,n}) - \ln(X_{n-k_n,n})$$

où $1 \leq k_n \leq n$ est une valeur à choisir par l'utilisateur.

Nous pouvons construire l'estimateur de Hill en utilisant plusieurs méthodes. Nous citons dans ce qui suit deux méthodes qui sont les plus utilisées.

Méthode du maximum de vraisemblance Sous l'hypothèse que les variables aléatoires X_1, \dots, X_n sont indépendantes et de même fonction de répartition F appartenant au domaine d'attraction de Fréchet d'indice $\gamma > 0$, la fonction de répartition s'écrit

$$F(x) = 1 - x^{-1/\gamma} L(x)$$

où L est une fonction à variations lentes. Nous utilisons le fait que pour x assez grand, la fonction à variations lentes comme son nom l'indique peut être considérée comme constante au voisinage de l'infini. En utilisant cette remarque, la fonction de répartition F peut être approximée par

$$F(x) \approx 1 - (x/c)^{-1/\gamma},$$

Ainsi, les valeurs les plus grandes de l'échantillon X_1, \dots, X_n obéissent approximativement à une loi de Pareto de paramètres c et γ à déterminer. Prenons les k_n plus grandes observations $X_{n-k_n+1,n}, \dots, X_{n,n}$ de l'échantillon X_1, \dots, X_n et calculons la fonction de vraisemblance associée.

$$\mathcal{L}(\gamma, c) = \prod_{i=1}^{k_n} \frac{1}{\gamma c} \left(\frac{X_{n-i+1,n}}{c} \right)^{-1/\gamma-1} \mathbb{1}_{\{X_{n-i+1,n} \geq c\}}$$

Nous commençons par la maximisation par rapport à γ de la vraisemblance que nous venons de donner par rapport à γ et pour simplifier les calculs nous utilisons à la place la Log-vraisemblance. Sous l'hypothèse les statistiques d'ordres sont supérieures à c nous obtenons

$$\ln \left(\prod_{i=1}^{k_n} \frac{1}{\gamma c} \left(\frac{X_{n-i+1,n}}{c} \right)^{-1/\gamma-1} \right) = k_n \ln(\gamma) - \frac{1}{\gamma} \sum_{i=1}^{k_n} \ln \frac{X_{n-i+1,n}}{c}.$$

En annulant la dérivée de cette Log-vraisemblance par rapport à γ nous arrivons à l'équation

$$-\frac{k_n}{\gamma} + \frac{1}{\gamma^2} \sum_{i=1}^{k_n} \ln \frac{X_{n-i+1,n}}{c} = 0$$

Ce qui nous donne

$$\gamma = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln (X_{n-i+1,n}) - \ln(c) \quad (3.13)$$

Il reste à donner une estimation de c pour compléter la construction de l'estimateur de γ . La condition que les statistiques d'ordres soient supérieures à c vient pour assurer que la vraisemblance ne s'annule pas et pour être sûr que cette condition soit vérifiée il suffit de prendre $c = X_{n-k_n,n} \leq \min_{1, \dots, k_n} (X_{n-i+1,n})$. Finalement, en remplaçant l'estimateur de c dans 3.13 nous trouvons

$$\hat{\gamma}_n^{(H)}(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln (X_{n-i+1,n}) - \ln (X_{n-k_n,n}).$$

Remarque 3.4

Nous constatons d'après l'expression de l'estimateur de Hill et entre autres sa construction qu'il dépend du choix du nombre k_n des plus grandes observations $X_{n-k_n+1,n}, \dots, X_{n,n}$ de l'échantillon X_1, \dots, X_n . Ce choix doit se faire d'une telle manière à trouver un compromis entre

1. *Un k_n trop grand \implies une mauvaise approximation par une loi de Pareto \implies un biais important de l'estimateur de Hill.*
2. *Un k_n est trop petit \implies très peu d'observations pour calculer l'estimateur de Hill \implies une variance importante.*

Donc le meilleur choix est de minimiser simultanément le biais ainsi que la variance de l'estimateur. Nous verrons plus loin dans ce document que théoriquement, la consistance de l'estimateur de Hill, est vérifié en prenant une suite intermédiaire k_n où : $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$. Pratiquement, le choix de cette suite k_n se fait souvent par une méthode graphique dite "Hill Horror Plot" qui consiste à tracer le graphique des points $\left\{ \left(k_n, \hat{\gamma}_n^{(H)}(k_n) \right), k_n = 1, \dots, n \right\}$ La figure suivante montre le Hill Horror plot pour une loi de Fréchet de paramètre $\gamma = 2/3$ et de taille $n = 500$ pour l'échantillon.

La figure suivante montre le Hill Horror plot pour une loi de Fréchet de paramètre $\gamma = 2/3$ et de taille $n = 500$ pour l'échantillon.

FIGURE 3.2 – Comprtement de l'estimateur de Hill en fonction de k_n 

Nous constatons sur la Figure 3.2 que le meilleur choix de k_n se situe dans la zone car les fluctuations de l'estimateur de Hill sont très grandes pour k_n petit (entre 0 et 250), ensuite il devient plus stable (entre 250 et 800) et enfin pour $k_n > 800$, l'estimateur de Hill subit des déviations qui représente le biais. Le meilleur choix de k_n se situe dans la zone de stabilité du "Hill Horror Plot".

Pareto Quantile Plot Une autre approche pour trouver l'expression de l'estimateur de Hill est d'utiliser le "Pareto Quantile Plot". C'est une approche qui utilise l'approximation suivante pour α proche de 0 :

$$\ln q(t\alpha) - \ln q(\alpha) \approx \gamma \ln(t^{-1}), \quad (3.14)$$

qui est déduite en utilisant le fait que pour $\alpha \in]0, 1[$, $q(\alpha) = \alpha^{-\gamma} \ell(\alpha^{-1})$ où ℓ est une fonction à variations lentes. Puis pour tout $t > 0$

$$\ln \frac{q(t\alpha)}{q(\alpha)} = \gamma \ln(t^{-1}) + \ln \frac{\ell(t^{-1}\alpha^{-1})}{\ell(\alpha^{-1})}.$$

Et finalement en utilisant la caractérisation d'une fonction à variations lentes. Le "Pareto Quantile Plot" est construit en suivant les étapes suivantes :

1. poser $i = 1, \dots, k_n$, $t_{i,n} = (2i - 1) / (2k_n)$ où (k_n) est une suite intermédiaire,
2. utiliser l'approximation 3.14 pour $t \in \{t_{i,n}, i = 1, \dots, k_n\}$ et $\alpha = k_n/n$, ce qui donne

$$\ln q\left(\frac{2i-1}{2n}\right) - \ln q\left(\frac{k_n}{n}\right) \approx \gamma \ln\left(\frac{2k_n}{2i-1}\right)$$

pour tout $i \in \{1, \dots, k_n\}$,

3. remplacer q par son estimateur empirique classique défini par

$$\hat{q}_n(\alpha) = X_{n-\lfloor n\alpha \rfloor, n},$$

pour tout $i \in \{1, \dots, k_n\}$, ce qui aboutit à

$$\ln X_{n-i+1, n} - \ln X_{n-k_n, n} \approx \gamma \ln \left(\frac{2k_n}{2i-1} \right) \quad (3.15)$$

Ce qui nous conduit à la définition du "Pareto Quantile Plot".

Définition 3.4

Le "Pareto Quantile Plot" est le graphique de l'ensemble des points

$$\left\{ \left(\ln \left(\frac{2k_n}{2i-1} \right), \ln X_{n-i+1, n} - \ln X_{n-k_n, n} \right), i = 1, \dots, k_n \right\}$$

où X_1, \dots, X_n sont des variables aléatoires indépendantes et de même loi et $k_n \in \{1, \dots, n\}$ est une suite intermédiaire.

Si la loi des variables aléatoires X_1, \dots, X_n est une loi appartenant au domaine d'attraction de Fréchet, le "Pareto Quantile Plot" formera approximativement une droite de pente $\gamma > 0$. Nous citons deux méthodes possible pour estimer cette pente :

Première méthode Prendre la moyenne dans chacun des membres de l'approximation 3.15 ce qui conduit à l'approximation

$$\gamma \approx \frac{1}{k_n} \sum_{i=1}^{k_n} [\ln X_{n-i+1, n} - \ln X_{n-k_n, n}] / \frac{1}{k_n} \sum_{i=1}^{k_n} \ln \left(\frac{2k_n}{2i-1} \right).$$

Ce qui nous ramène à l'estimateur de Hill en remarquant que

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{i=1}^k \ln \left(\frac{2k}{2i-1} \right) = 1.$$

Deuxième méthode Utiliser la méthode des moindres carrés c'est-à-dire estimer γ par

$$\arg \min_{\gamma} \sum_{i=1}^{k_n} \left\{ \ln [X_{n-i+1, n} - \ln X_{n-k_n, n}] - \gamma \ln \left(\frac{2k_n}{2i-1} \right) \right\}^2$$

ce qui nous donne l'estimateur dit de Zipf dont la définition est la suivante.

Définition 3.5 (Estimateur de Zipf)

Soit $(k_n)_n \geq 1$ une suite d'entiers avec $1 < k_n < n$, l'estimateur de Zipf est défini par

$$\hat{\gamma}_n^Z = \frac{\frac{1}{k_n} \sum_{i=1}^{k_n} \left(\ln \left(\frac{k_n+1}{i} \right) - \frac{1}{k_n} \sum_{i=1}^{k_n} \ln \left(\frac{k_n+1}{i} \right) \right) \ln(X_{n-i+1, n})}{\frac{1}{k_n} \sum_{i=1}^{k_n} \left(\ln \left(\frac{k_n+1}{i} \right) \right)^2 - \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \ln \left(\frac{k_n+1}{i} \right) \right)^2} \quad (3.16)$$

En 1996, cette estimateur de l'indice des valeurs extrêmes a été le fruit des deux travaux indépendants qui sont de Schultze et Steinebach [68] et de Kratz et Resnick [49]. C'est un estimateur qui a l'avantage d'avoir un biais moindre que celui de l'estimateur de Hill par contre il a une variance asymptotique plus importante.

Il est primordiale de donner les propriétés asymptotiques des précédents estimateurs de l'indice des valeurs extrêmes γ . Nous commençons par celui qui est le plus utilisé : L'estimateur de Hill.

Comportement asymptotique de l'estimateur de Hill L'étude du comportement asymptotique d'un l'estimateur vise surtout sa normalité asymptotique et c'est se que nous allons essayer de voir dans ce paragraphe pour l'estimateur de Hill. Afin d'arriver à montrer la normalité asymptotique de cet estimateur nous devons suivre plusieurs étapes. La première étape est de faire une décomposition de l'estimateur de Hill. Pour y parvenir, nous adopterons les notations suivantes : $\{E_i, i \geq 1\}$ pour des variables aléatoires indépendantes de loi exponentielle et $\{U_i, i \geq 1\}$ pour des variables aléatoires indépendantes de loi uniforme.

Proposition 3.2

Si X_1, \dots, X_n sont des variables aléatoires indépendantes et de même loi F qui appartient au domaine d'attraction de Fréchet vérifiant donc la relation $q(\alpha) = \alpha^{-\gamma} \ell(\alpha^{-1})$ où $q(\alpha)$ est le quantile d'ordre α alors

$$\hat{\gamma}_n^{(H)}(k_n) \stackrel{\mathcal{L}}{=} \gamma A(k_n) + R(k_n)$$

où

$$A(k_n) \stackrel{\mathcal{L}}{=} \frac{1}{k_n} \sum_{i=1}^{k_n} E_i \text{ et } R(k_n) = \frac{1}{k_n} \sum_{i=1}^{k_n} \ln \frac{\ell(U_{i,n}^{-1})}{\ell(U_{k_n+1,n}^{-1})}$$

En se servant de cette décomposition et en utilisant le Théorème Central Limite sur le cas d'une loi de Pareto où la fonction à variations lentes ℓ est constante et en prenant $k_n = n$.

$$\sqrt{n} (\hat{\gamma}_n^{(H)}(n) - \gamma) \stackrel{\mathcal{L}}{\rightarrow} \mathcal{N}(0, \gamma^2).$$

Si la fonction à variations lentes n'est pas constante, nous procédons comme suit : pour montrer que

$$\sqrt{k_n} (\hat{\gamma}_n^{(H)}(k_n) - \gamma) \stackrel{\mathcal{L}}{\rightarrow} \mathcal{N}(0, \gamma^2)$$

nous montrons $(k_n)^{1/2} R(k_n) \xrightarrow{\text{P}} 0$ lorsque $n \rightarrow \infty$ car

$$\begin{aligned} \sqrt{k_n} (\hat{\gamma}_n^{(H)}(k_n) - \gamma) &\stackrel{\mathcal{L}}{=} \gamma \sqrt{k_n} [A(k_n) - 1] + \sqrt{k_n} R(k_n) \\ &\stackrel{\mathcal{L}}{=} \gamma \sqrt{k_n} \left(\frac{1}{k_n} \sum_{i=1}^{k_n} E_i - 1 \right) + \sqrt{k_n} R(k_n) \\ &\stackrel{\mathcal{L}}{=} \gamma \sqrt{k_n} \left(\frac{1}{k_n} \sum_{i=1}^{k_n} E_i - \mathbb{E} \left(\frac{1}{k_n} \sum_{i=1}^{k_n} E_i \right) \right) + \sqrt{k_n} R(k_n) \end{aligned}$$

D'après le TCL, le premier terme converge en loi vers une loi normale. Il nous reste seulement à contrôler le terme $R(k_n)$ contenant la fonction à variations lentes ℓ). Pour cela nous introduisons les conditions supplémentaires suivantes.

H.1 La fonction à variations lentes ℓ est normalisée i.e. il existe x_0 tel que

$$\ell(x) = c \exp \left\{ \int_{x_0}^x \frac{\Delta(t)}{t} dt \right\}$$

avec $c > 0$ et $\Delta(t) \rightarrow 0$ lorsque $t \rightarrow \infty$.

H.2 La fonction $|\Delta|$ est à variations régulières d'indice $\rho < 0$.

Le paramètre ρ qui contrôle la vitesse de convergence de $\ell(tx)/\ell(x)$ vers 1 lorsque $x \rightarrow \infty$, est appelé paramètre du second ordre. Si ρ est petit, la convergence est rapide et l'estimateur de Hill aura un bon comportement. Par contre, si ρ est proche de 0, la convergence est lente et l'estimateur de Hill aura un biais important.

H.3 La fonction $|\Delta|$ est asymptotiquement décroissante.

Sous les hypothèses que nous venons de citer, nous avons le Théorème suivant.

Théorème 3.4

Si X_1, \dots, X_n sont des variables aléatoires indépendantes dont la loi commune appartient au domaine d'attraction de Fréchet, sous les hypothèses (H.1), (H.2), (H.3), si $k_n \rightarrow \infty, n/k_n \rightarrow \infty$ et $k_n^{1/2} \Delta(n/k_n) \rightarrow 0$ lorsque $n \rightarrow \infty$ alors

$$\sqrt{k_n} (\hat{\gamma}_n^{(H)}(k_n) - \gamma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2)$$

Remarque 3.5

Un choix possible pour la suite (k_n) est

$$k_n = cn^{-2(\rho+\eta)/(1-2\rho)}$$

où $c > 0, \eta \in]0, -\rho[$ aussi proche de 0.

Comportement asymptotique de l'estimateur de Weissman On rappelle que pour toute suite $\alpha_n \rightarrow 0$, l'estimateur de Weissman du quantile $q(\alpha_n)$ est donné par

$$\hat{q}_n(\alpha_n) = X_{n-k_n, n} \left(\frac{k_n}{n\alpha_n} \right)^{q_n^{(H)}(k_n)}$$

en utilisant l'estimateur de Hill comme estimateur de $\gamma > 0$. On a le résultat suivant.

Théorème 3.5

Soient X_1, \dots, X_n des variables aléatoires indépendantes dont la loi commune appartient au domaine d'attraction de Fréchet. Sous les hypothèses (H.1), (H.2), et (H.3), pour toutes suites k_n et α_n telles que $k_n \rightarrow \infty, n/k_n \rightarrow \infty, k_n^{1/2} \Delta(n/k_n) \rightarrow 0$ et $n\alpha_n/k_n \rightarrow 0$, on a

$$\frac{k_n^{1/2}}{\ln[k_n/(n\alpha_n)]} \ln \left(\frac{\hat{q}_n(\alpha_n)}{q(\alpha_n)} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma^2)$$

3.3 Etude pratique : maximas annuelles des températures

Nous étudions dans ce paragraphe les maximas annuelles des température en degré Fahrenheit ($^{\circ}F$) de la ville de New York de l'année 1938 à l'année 2017. Les données de cette étude peuvent être téléchargées du lien [Maximas Annuelles \[1\]](#).

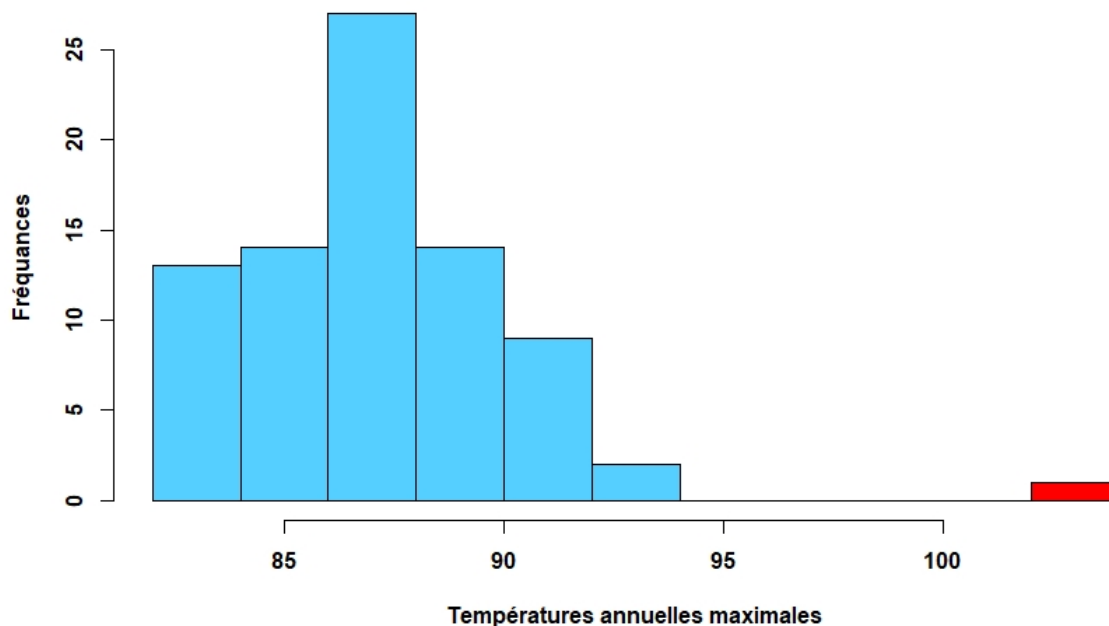


FIGURE 3.3 – les maximas annuelles des températures en $^{\circ}F$ de la ville de New York de l'année 1938 à l'année 2017.

Nous constatons de ce graphique qu'il y a une valeur qui est complètement différente et loin de toutes les autres (en couleur rouge). Ce phénomène est caractéristique des valeurs extrêmes. Nous optons donc à faire un ajustement par une loi \mathcal{GEV} en se servant du package **extRemes** et la fonction **fevd**. Les résultats sont illustrés par le graphique suivant qui montre que l'ajustement est de bonne qualité.

fevd(x = T_max_Ann, type = "GEV")

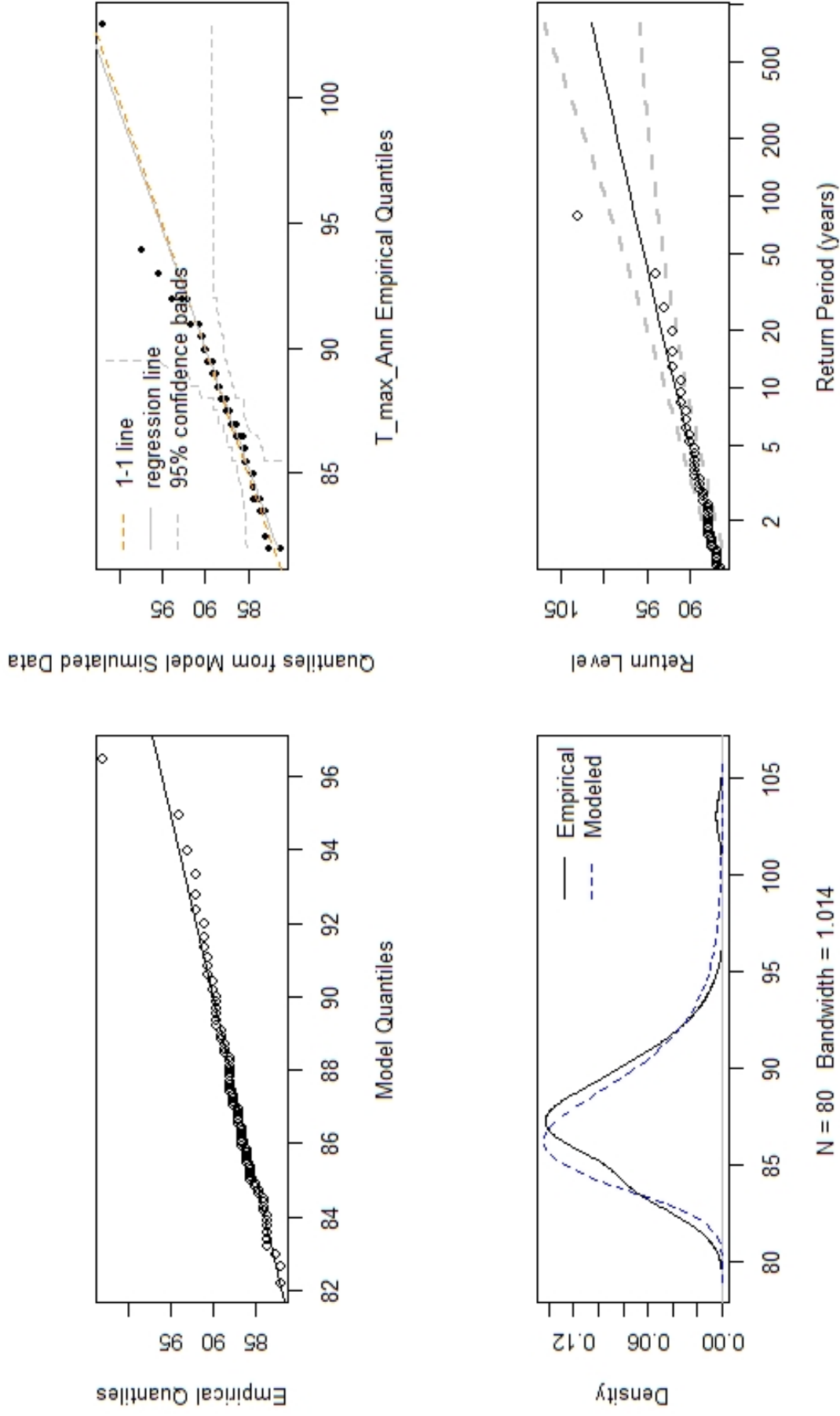


FIGURE 3.4 – Etude des valeurs extrêmes des températures maximales annuelles en degré Fahrenheit ($^{\circ}F$) de la ville de New York de l'année 1938 à l'année 2017 par approche \mathcal{GEV} .

Deuxième partie

Mesures de Risques

Chapitre 4

Mesures de risques

Sommaire

4.1	Première mesure de risque	76
4.2	Approche axiomatique des mesures de risque et mesures de risque usuelles	76
	L'équivalent certain	78
	La Value-at-Risk	79
	La Tail Value-at-Risk	80
	Transformée de Hardy-Littlewood	81
	La Conditional Tail Expectation	81
	La Conditional-VaR	81
	L'Expected Shortfall	82
4.2.1	Liens entre mesures de risque.	82
4.2.2	Mesures de risque de Wang	82
4.2.3	Mesures de risque par distorsion	83
4.3	Estimation des mesures de risque définies par distorsion	84
4.3.1	Estimation de la Value-at-Risk	85
	Estimation paramétrique	85
	Estimation non-paramétrique	86
	Estimation semi-paramétrique	87
4.4	Une nouvelle mesure de risque : le Conditional Tail Moment	87

Résumé

Le métier d'actuaire par exemple est un métier qui fait appel au calcul des probabilités et de la statistique aux questions d'assurances, de prévention, de finance et de prévoyance

sociale. Il analyse l'impact financier du risque (risques financiers, risques juridiques (la mauvaise volonté dans l'application d'une loi), risques industriels et risques environnementaux, etc...) et estime les conséquences futurs qui y sont associés. Pour faire cette estimation il doit quantifier le risque ou d'une autre manière donner et travailler avec une mesure de risque.

4.1 Première mesure de risque

La première approche pour mesurer le risque est d'utiliser la variance de la statistique étudiée. L'histoire de Paul Samuelson (voir [67]) aboutit à la conclusion qu'elle l'est une mais elle n'est pas adéquate pour tous les cas. Il raconte que si on considère un jeu de pile ou face où l'on décide à l'avance du nombre de lancers n que l'on fera et où l'on peut gagner 200 si la pièce tombe sur pile et on perd 100 si la pièce tombe sur face et que la pièce est bien équilibrée. Si on décide qu'on va prendre n comme nombre de lancers que l'on est prêt à faire, en calculant la variance on trouve

$$\text{Var}(X_n) = \frac{300^2}{4n} \rightarrow 0 \text{ lorsque } n \rightarrow \infty.$$

où X_n est le gain (ou la perte) au bout de n lancers. Cette formule nous donne théoriquement la possibilité de trouver n au delà duquel le risque pris devient aussi petit qu'on le souhaite. Mais en pratique, Paul Samuelson raconte qu'il a trouvé des personnes prêtes à jouer $n = 5$, voire $n = 10$ parties. Mais personne pour en jouer $n = 100$ malgré le fait que pour ce nombre de lancers on est presque certain de ne rien perdre car

$$P(X_{100} > 0) = P(\text{au moins 34 piles sur 100 lancers}) \sim 99,91\%$$

. Ce qui fait que la variance n'est pas une mesure "adéquate" pour quantifier le risque. Donc quelles sont les critères à vérifier pour dire qu'on est en présence de mesure de risque adéquate? Ce qui suit répond à cette question.

4.2 Approche axiomatique des mesures de risque et mesures de risque usuelles

Définition 4.1

On appelle mesure de risque toute application \mathcal{R} associant un risque X à un réel $\mathcal{R}(X) \in \mathbb{R}_+ \cup \{+\infty\}$.

On donne quelques propriétés naturelles et souhaitables de ces mesures :

- **invariance en loi** : si $X \stackrel{\mathcal{L}}{=} Y$, alors $\mathcal{R}(X) = \mathcal{R}(Y)$,
- **croissance** : si $X \geq Y$, alors $\mathcal{R}(X) \geq \mathcal{R}(Y)$,

- **invariance par translation** : si $k \in \mathbb{R}$, alors $\mathcal{R}(X + k) = \mathcal{R}(X) + k$,
- **homogénéité** : si $\lambda \in \mathbb{R}_+$, alors $\mathcal{R}(\lambda X) = \lambda \mathcal{R}(X)$,
- **sous additivité** : pour tous risques X et Y , $\mathcal{R}(X + Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$,
- **convexité** : si $\beta \in [0, 1]$, alors

$$\mathcal{R}(\beta X + (1 - \beta) Y) \leq \beta \mathcal{R}(X) + (1 - \beta) \mathcal{R}(Y)$$

Par la suite, nous ne nous intéresserons qu'aux mesures de risque invariantes en loi.

Remarque 4.1

Une conséquence de l'invariance par translation est :

$$\mathcal{R}(X - \mathcal{R}(X)) = 0.$$

Définition 4.2

Une mesure de risque est dite :

- *monétaire* si elle est monotone et invariante par translation,
- *convexe* si elle est monétaire et convexe,
- *cohérente* si elle est monétaire, homogène et sous-additive.

Corollaire 4.1

Si \mathcal{R} est une mesure de risque monétaire, homogène et normalisée à 0, alors la convexité et la sous-additivité sont des notions équivalentes.

Définition 4.3

Si \mathcal{R} est une mesure de risque, on définit la région de risques acceptables pour la mesure \mathcal{R} comme

$$\mathcal{A} = \{X, \mathcal{R}(X) \leq 0\}.$$

Réciproquement, si \mathcal{A} est une région de risques acceptables, la mesure de risque induite \mathcal{R} est

$$\mathcal{R}(X) = \inf \{m, X - m \in \mathcal{A}\}.$$

Proposition 4.1

Si \mathcal{R} est une mesure de risque monétaire alors

- \mathcal{R} est convexe si et seulement si \mathcal{A} est convexe,
- \mathcal{R} est positivement homogène si et seulement si \mathcal{A} est un cône.

On a le théorème de représentation suivant

Théorème 4.1

\mathcal{R} est une mesure de risque monétaire convexe si et seulement si pour tout X bornée ($X \in L^\infty$) :

$$\mathcal{R}(X) = \max_{\mathbb{Q} \in \mathcal{M}} \{\mathbb{E}_{\mathbb{Q}} - \alpha(\mathbb{Q})\},$$

où \mathcal{M} est l'ensemble des mesures additives et normalisées à 1, et

$$\alpha(\mathbb{Q}) = \sup_{X \in \mathcal{A}} \{\mathbb{E}_{\mathbb{Q}}(X)\},$$

où \mathcal{A} est l'ensemble d'acceptation associé à \mathcal{R} .

Proposition 4.2

Si R est une mesure de risque monétaire alors elle vérifie la propriété

- d'additivité pour les risques comonotones si, pour tout X et Y comonotones, $\mathcal{R}(X + Y) = \mathcal{R}(X) + \mathcal{R}(Y)$,
- de corrélation maximale (par rapport à une mesure μ) si pour tout X ,

$$R(X) = \sup \{\mathbb{E}(X \cdot U) \text{ où } U \sim X\},$$

- de cohérence forte si pour tout X et Y , $\sup \{R(\tilde{X} + \tilde{Y})\} = R(X) + R(Y)$, pour $\tilde{X} \stackrel{\mathcal{L}}{\cong} X$ et $\tilde{Y} \stackrel{\mathcal{L}}{\cong} Y$.

Proposition 4.3

Si \mathcal{R} est une mesure de risque convexe, les trois propriétés suivantes sont équivalentes

- \mathcal{R} est fortement cohérente,
- \mathcal{R} est additive pour des risques comonotones,
- \mathcal{R} est une mesure de corrélation maximale.

S. Kusuoka (2001) [50] a montré le résultat suivant dès lors que R est une mesure invariante en loi.

Proposition 4.4

Une mesure de risque cohérente R est additive pour des risques comonotones si et seulement s'il existe une fonction décroissante positive g sur $[0, 1]$ telle que

$$\mathcal{R}(X) = \int_0^1 g(t) F_X^{-1}(1-t) dt$$

où $F_X(x) = \mathbb{P}(X \leq x)$.

L'équivalent certain**Définition 4.4**

Soit u une fonction d'utilité concave, strictement croissante, alors l'équivalent certain $\mathcal{R}(X)$ associé à une perte X vérifie

$$u(\mathcal{R}(X)) = \mathbb{E}(u(X)) \text{ soit } \mathcal{R}(X) = u^{-1}(\mathbb{E}(u(X))).$$

La Value-at-Risk

La Value-at-Risk est apparue (sous ce nom) dans les années 90. C'est la mesure la plus utilisée et la plus étudiée dans la littérature.

Définition 4.5

On appelle Value-at-Risk de niveau $\alpha \in (0, 1)$ le quantile de niveau α ,

$$\mathcal{R}_\alpha(X) = VaR(X; \alpha) = x_\alpha \text{ où } \mathbb{P}(X \leq x_\alpha) = \alpha,$$

où encore

$$VaR(X; \alpha) = \inf \{x, \mathbb{P}(X \leq x) \geq \alpha\} = F_X^{-1}(\alpha) = Q_X(\alpha).$$

Remarque 4.2

Certains ouvrages notent $\mathcal{R}_\alpha(X)$ le quantile d'ordre $1 - \alpha$.

Lemme 4.1

Pour tout $\alpha \in (0, 1)$, si g est une fonction strictement croissante et continue à gauche :

$$VaR(g(X); \alpha) = F_{g(X)}^{-1}(\alpha) = g(F_X^{-1}(\alpha)) = g(VaR(X; \alpha)),$$

alors que si g est une fonction strictement décroissante, continue à droite, et si F_X est bijective :

$$VaR(g(X); \alpha) = F_{g(X)}^{-1}(\alpha) = g(F_X^{-1}(1 - \alpha)) = g(VaR(X; 1 - \alpha)),$$

Remarque 4.3

En prenant $g(x) = x + c$ et $g(x) = cx$, on déduit que la VaR est invariante par translation et homogène. Par contre elle n'est pas sous-additive et donc non cohérente.

Exemple 4.1

Soient X_1, X_2 deux v.a.i.i.d de fonction de répartition commune $F(x) = 1 - 1/\sqrt{x}$ pour $x > 1$. Nous avons alors

$$VaR(X_1; \alpha) = VaR(X_2; \alpha) = \frac{1}{(1 - \alpha)^2}.$$

Nous pouvons montrer

$$\begin{aligned} F_{X_1+X_2}(x) &= \mathbb{P}(X_1 + X_2 \leq x) \\ &= \int_1^{x-1} F_X(x-y) dF_X(y) \\ &= 1 - 2 \frac{\sqrt{x-1}}{x} \\ &< 1 - \sqrt{\frac{2}{x}} \\ &= F_{2X}(x), \end{aligned}$$

où $F_{2X}(t) = \mathbb{P}(2X_1 \leq t)$, pour $t \geq 2$.

Ainsi on a

$$\begin{aligned} \text{VaR}(X_1 + X_2; \alpha) &> \text{VaR}(2X_1; \alpha) \\ &\stackrel{\text{homogénéité}}{=} \text{VaR}(X_1; \alpha) + \text{VaR}(X_1; \alpha), \end{aligned}$$

et comme $\text{VaR}(X_1; \alpha) = \text{VaR}(X_2; \alpha)$, on aura

$$\text{VaR}(X_1 + X_2; \alpha) > \text{VaR}(X_1; \alpha) + \text{VaR}(X_2; \alpha),$$

ce qui fait que la VaR n'ai pas sous-additive.

La figure 4.1 illustre le résultat du dernier exemple.

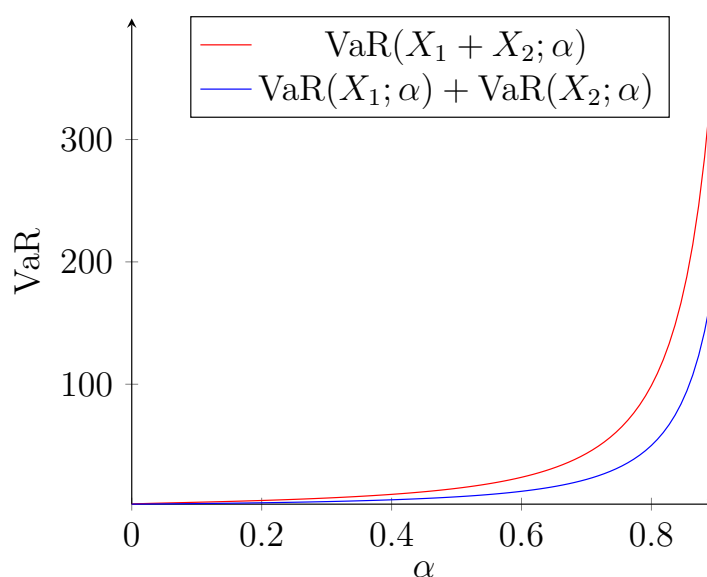


FIGURE 4.1 – Comparaison entre la VaR de la somme de deux variables aléatoires et la somme de chacune de leur VaR.

La Tail Value-at-Risk

La Tail-Value-at-Risk, s'intéresse à la moyenne des événements rares et extrêmes survénus. C'est une mesure qui vient remédier la faiblesse de la Value-at-Risk qui s'intéresse seulement aux probabilités de ces événements extrêmes.

Définition 4.6

La Tail Value-at-Risk au niveau α , notée $TVaR(X; \alpha)$ est définie par :

$$TVaR(X; \alpha) = \frac{1}{1 - \alpha} \int_{\alpha}^1 \text{VaR}(X; t) dt.$$

La $TVaR$ est la moyenne des VaR de niveau supérieur à α .

Transformée de Hardy-Littlewood Il existe une fonction de répartition F_X , appelée transformée de Hardy-Littlewood de F_X (voir G. H. Hardy and J. E. Littlewood. A maximal theorem with functiontheoretic applications. Acta Math., 54(1) :81–116, 1930), telle que pour tout α :

$$\tilde{F}_X^{-1}(\alpha) = TVaR(X; \alpha).$$

Si on note \tilde{X} une variable aléatoire de fonction de répartition \tilde{F}_X , on a :

$$TVaR(X; \alpha) = VaR(\tilde{X}; \alpha).$$

La $TVaR$ d'un risque X est donc la VaR de la transformée de Hardy-Littlewood de X .

Propriété 4.1

1. $TVaR(X; 0) = E(X)$.
2. $TVaR(X; \alpha) = \frac{1}{1-\alpha} \{E(X) - \int_0^\alpha VaR(X; t) dt\}$.
3. La Tail Value-at-Risk est une fonction croissante du niveau α .
4. $TVaR(X; \alpha) \geq TVaR(X; 0) = \mathbb{E}(X)$.

La Conditional Tail Expectation

Définition 4.7

La Conditional Tail Expectation au niveau de probabilité α , notée $CTE[X; \alpha]$, est l'espérance de la perte sachant que la VaR au niveau α est dépassée :

$$CTE(X; \alpha) = \mathbb{E}(X | X > VaR(X; \alpha)).$$

Cette formule peut être exprimée comme perte moyenne dans les pires $(1 - \alpha)\%$ des cas.

La Conditional-VaR

Définition 4.8

La Conditional-VaR au niveau de probabilité α , notée $CVaR(X; \alpha)$, est la valeur moyenne des pertes qui excèdent la VaR , correspondant à l'excédent moyen au-delà de la VaR ou encore :

$$\begin{aligned} CVaR(X; \alpha) &= \mathbb{E}(X - VaR(X; \alpha) | X > VaR(X; \alpha)) \\ &= e_X(VaR(X; \alpha)) \\ &= CTE(X; \alpha) - VaR(X; \alpha), \end{aligned}$$

où $e_X(s) = \mathbb{E}(X | X > s)$.

L'Expected Shortfall

Définition 4.9

L'Expected Shortfall au niveau de probabilité α , notée $ES(X; \alpha)$, n'est autre que la prime stop-loss dont la rétention est fixée à $VaR(X; \alpha)$:

$$ES(X; \alpha) = \mathbb{E}((X - VaR(X; \alpha))_+) = \mathbb{E}(\max(0, (X - VaR(X; \alpha))).$$

4.2.1 Liens entre mesures de risque.

Proposition 4.5

Quel que soit le niveau de probabilité $\alpha \in (0, 1)$, on a :

$$TVaR(X; \alpha) = VaR(X; \alpha) + \frac{1}{1 - \alpha} ES(X; \alpha) \quad (4.1)$$

$$CTE(X; \alpha) = VaR(X; \alpha) + \frac{1}{\overline{F}_X(VaR(X; \alpha))} ES(X; \alpha) \quad (4.2)$$

La Propriété 4.1 donne :

Proposition 4.6

Si la fonction de répartition des risques est continue alors la CTE et la TVaR sont égales :

$$CTE(X; \alpha) = TVaR(X; \alpha), \text{ pour } \alpha \in (0, 1). \quad (4.3)$$

Propriété 4.2

La TVaR est invariante par translation et homogène :

$$\begin{aligned} TVaR(X + c; \alpha) &= TVaR(X; \alpha) + c. \\ TVaR(cX; \alpha) &= cTVaR(X; \alpha) \end{aligned}$$

Proposition 4.7

La TVaR est cohérente et coïncide avec la CTE pour les risques continus.

4.2.2 Mesures de risque de Wang

Définition 4.10

Toute fonction croissante

$$g : [0, 1] \rightarrow [0, 1]$$

telle que $g(0) = 0$ et $g(1) = 1$ est dite fonction de distorsion.

Nous nous intéressons dans cette partie aux mesures de risque définies par distorsion pour les variables aléatoires strictement positives

Définition 4.11

La mesure de risque de Wang associée à la fonction de distorsion g , notée \mathcal{R}_g , est définie par :

$$\mathcal{R}_g(X) = \int_0^{+\infty} g(1 - F_X(x)) dx = \int_0^{+\infty} g(\bar{F}_X(x)) dx \quad (4.4)$$

Proposition 4.8

Quel que soit le risque X , la mesure de risque de Wang associée à la fonction de distorsion g peut s'écrire :

$$\mathcal{R}_g(X) = \int_0^1 VaR(X; 1 - \alpha) dg(\alpha) \quad (4.5)$$

Ainsi, les mesures de risque de Wang sont des moyennes pondérées de VaR .

Nous citons les propriétés des mesures de risque de Wang suivantes :

Proposition 4.9

Les mesures de risque de Wang sont

- homogènes,
- invariantes par translation,
- monotones.

Proposition 4.10

Lorsque la fonction de distorsion est concave, la mesure de risque correspondante est sous-additive.

Corollaire 4.2

Les mesures de risque de Wang correspondant à des fonctions de distorsion concaves sont cohérentes.

Le résultat suivant montre que la $TVaR$ est la plus petite mesure de risque cohérente excédant la VaR .

Proposition 4.11

Quel que soit le niveau de probabilité $0 < \alpha < 1$ et le risque X , nous avons :

$$TVaR(X; \alpha) = \min \{ \mathcal{R}_g(X) \mid g \text{ est concave et } \mathcal{R}_g(X) \geq VaR(X; \alpha) \}. \quad (4.6)$$

4.2.3 Mesures de risque par distorsion

Les mesures de risque par distorsion est une classe plus grande contenant les mesures de Wang. Elles sont définies même si les variables aléatoires ne sont pas positives.

Définition 4.12

On appelle mesure de risque par distorsion la quantité :

$$\mathcal{R}(X) = \int_0^1 F_X^{-1}(1 - u) dg(u).$$

où g est une fonction de répartition sur $[0, 1]$, appelée fonction de distorsion.

Cette expression peut s'écrire sous la forme

Proposition 4.12

$$\mathcal{R}(X) = \int_0^{+\infty} g(1 - F_X(x)) dx - \int_{-\infty}^0 (1 - g(1 - F_X(x))) dx.$$

Remarque 4.4

Dans le cas particulier où g est la fonction de répartition de la loi uniforme sur $[0, 1]$, on a :

$$\mathcal{R}(X) = \int_0^{+\infty} (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx = \mathbb{E}(X).$$

Principales mesures de risque par distorsion le tableau suivant donne les principales mesures de risque définies par distorsion :

TABLE 4.1 – Principales mesures de risque par distorsion (où α est une constante comprise entre 0 et 1).

Mesure de risque \mathcal{R}	fonction de distorsion g
VaR	$g(x) = \mathbf{1}(x \geq \alpha)$
$TVaR$	$g(x) = \min\{x/\alpha, 1\}$
PH	$g(x) = x^\alpha$
Dual Power	$g(x) = 1 - (1 - x)^{1/\alpha}$
Gini	$g(x) = (1 + \alpha)x - \alpha x^2$
Transformation exponentielle	$g(x) = (1 - \alpha^x) / (1 - \alpha)$

4.3 Estimation des mesures de risque définies par distorsion

Considérons une mesure de risque par distorsion, c'est-à-dire de la forme

$$\mathcal{R}(X) = \int_0^1 F_X^{-1}(1 - u) dg(u),$$

où F_X est supposée strictement croissante, et continue. L'estimateur naturel de cette mesure de risque est alors

$$\widehat{\mathcal{R}}(\{X_1, \dots, X_n\}) = \int_0^1 F_n^{-1}(1 - u) dg(u).$$

Etant données que F_n^{-1} est une fonction en escalier, alors

$$\widehat{\mathcal{R}}(\{X_1, \dots, X_n\}) = \sum_{i=1}^n \left(g\left(\frac{i}{n}\right) - g\left(\frac{i-1}{n}\right) \right) X_{n+i+1:n},$$

qui est simplement un L estimateur¹. On peut alors, tout naturellement, espérer retrouver facilement des propriétés comme celles obtenues pour l'estimation non-paramétrique de la VaR . Par exemple

Proposition 4.13

Si X_1, \dots, X_n est un échantillon i.i.d.,

$$\sqrt{n} \left(\mathcal{R}(X) - \widehat{\mathcal{R}}(\{X_1, \dots, X_n\}) \right) \rightarrow \int_0^1 \frac{B_{1-t}}{f(F^{-1}(1-t))} dg(t),$$

où $(B_t)_{t \in [0,1]}$ est un pont brownien. La loi limite est centrée, de variance asymptotique

$$\int_0^1 \int_0^1 \frac{\min\{t, u\} - ut}{f(F^{-1}(1-t))f(F^{-1}(1-u))} dg(t)dg(u).$$

Nous nous intéressons dans ce qui suit à l'estimation de la Value-at-Risk qui représente un cas particulier des mesures de risque par distorsion mais qui joue un rôle central dans la construction de ces mesures.

4.3.1 Estimation de la Value-at-Risk

Estimation paramétrique

Nous supposons que la loi de la variable étudiée, appartient à une famille paramétrique $F_X \in \mathcal{F} = \{F_\theta, \theta \in \Theta\}$ (que l'on supposera continue), dans ce cas

$$Q_X(\alpha) = F_\theta^{-1}(\alpha)$$

, et un estimateur naturel du quantile de niveau α est

$$\widehat{Q}_X(\alpha) = F_{\widehat{\theta}}^{-1}(\alpha) \tag{4.7}$$

où $\widehat{\theta}$ est un estimateur de θ (par maximum de vraisemblance, par la méthode des moments, etc).

Exemple 4.2

Loi de Pareto

$$F_X(x) = P(X \leq x) = 1 - \left(\frac{1}{x}\right)^\theta, \text{ pour } x \geq 1 \text{ et } \theta \geq 0.$$

► *Méthode du maximum de vraisemblance : La Log-vraisemblance est*

$$\ln \mathcal{L}(\theta) = \sum_{i=1}^n \ln \left(\frac{\theta}{X_i^{\theta+1}} \right) = n \log \theta - (\theta + 1) \sum_{i=1}^n \ln X_i,$$

1. Un L -estimateur est une combinaison linéaire de statistiques basées sur des statistiques d'ordre ou des quantiles d'échantillons

Son maximum est atteint pour

$$\hat{\theta}_1 = \left(\frac{1}{n} \sum_{i=1}^n \ln X_i \right)^{-1}.$$

► **Méthode des moments** : En utilisant le fait que $\mathbb{E}(X) = \frac{\theta}{\theta-1}$ pour $\theta > 1$, l'estimateur de la méthode des moments, $\hat{\theta}_2 = \frac{\hat{\mu}}{\hat{\mu}-1}$.

► **Méthode des moindres carrés** En prenant le logarithme de la fonction de répartition, on trouve la relation linéaire

$$\ln(1 - F_X(x)) = \ln(\bar{F}_X(x)) = 1 - \theta \ln(x),$$

dont l'estimateur par moindres carrés de la pente est donné par

$$\hat{\theta}_3 = \frac{-n \sum_{i=1}^n \ln(X_{i:n}) \ln\left(\frac{n-i}{n}\right) + \sum_{i=1}^n \ln(X_{i:n}) \sum_{i=1}^n \ln\left(\frac{n-i}{n}\right)}{n \sum_{i=1}^n (\ln(X_{i:n}))^2 - (\sum_{i=1}^n \ln(X_{i:n}))^2},$$

En résumé, on a obtenu trois estimateurs paramétriques qui donnent trois estimateurs de quantiles :

$$\hat{Q}_k(\alpha) = (1 - \alpha)^{-1/\hat{\theta}_k}, k = 1, 2, 3.$$

Estimation non-paramétrique

L'estimation non-paramétrique vient pour éviter les erreurs dans le choix de la famille paramétrique \mathcal{F} .

Définition 4.13

La fonction de répartition empirique F_n , construite à partir de $\{X_1, \dots, X_n\}$ est

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x).$$

La loi (forte) des grands nombres, montre que $F_n(x)$ converge presque sûrement vers $F(x)$ quand $n \rightarrow \infty$, et le théorème de Glivenko-Cantelli assure sa convergence uniforme,

$$\sup \{|F_n(x) - F(x)|\} \text{ tend presque sûrement vers } 0 \text{ lorsque } n \rightarrow \infty.$$

Le théorème central limite assure la normalité asymptotique,

$$\sqrt{n}(F_n(x) - F(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, F(x)(1 - F(x))) \text{ lorsque } n \rightarrow \infty.$$

Définition 4.14

La fonction quantile empirique, construite à partir de $\{X_1, \dots, X_n\}$ est

$$Q_n(\alpha) = \inf \{x \in \mathbb{R} : F_n(x) \geq \alpha\} = F_n^{-1}(\alpha).$$

On notera que

$$Q_n(\alpha) = X_{k:n} \text{ où } \frac{k-1}{n} < \alpha \leq \frac{k}{n}.$$

Le théorème de Glivenko Cantelli garantie la convergence (forte) de $Q_n(\alpha)$ vers $Q(\alpha)$.

Proposition 4.14

Si X_1, \dots, X_n est un échantillon de i.i.d. de loi absolument continue de densité f , et que $f(Q(\alpha)) > 0$, alors

$$\sqrt{n}(Q_n(\alpha) - Q(\alpha)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f^2(Q(\alpha))}\right) \text{ lorsque } n \rightarrow \infty.$$

Estimation semi-paramétrique

Cette approche utilise la théorie des valeurs extrêmes en se servant du théorème de Pickands-Balkema-de Haan qui dit que si u est suffisant grand, alors la loi conditionnelle de $Y-u$ sachant $Y > u$ suit une de Pareto généralisée, de paramètres ξ et β dont on peut estimer par la méthode du maximum de vraisemblance. En prenant pour seuil $u = Y_{n-k:n}$, avec k suffisamment grand, et en notant $\hat{\xi}_k$ et $\hat{\beta}_k$ les estimateurs du maximum de vraisemblance de la loi de Pareto généralisé sur l'échantillon $\{Y_{n-k+1:n} - Y_{n-k:n}, \dots, Y_{n:n} - Y_{n-k:n}\}$,

$$\hat{Q}(Y; \alpha) = Y_{n-k:n} + \frac{\hat{\beta}_k}{\hat{\xi}_k} \left(\left(\frac{n}{k} (1-\alpha) \right)^{-\hat{\xi}_k} - 1 \right). \quad (4.8)$$

Alternativement, si $\xi > 0$, on peut utiliser l'estimateur de Hill,

$$\hat{Q}(Y; \alpha) = Y_{n-k:n} \left(\frac{n}{k} (1-\alpha) \right)^{-\hat{\xi}_k}, \quad (4.9)$$

où

$$\hat{\xi}_k = \frac{1}{k} \sum_{i=1}^k \log Y_{n+1-i:n} - \log Y_{n-k:n}.$$

4.4 Une nouvelle mesure de risque : le Conditional Tail Moment

Cette nouvelle mesure de risque dite : le Conditional Tail Moment (CTM), est introduite par J.El Methni [33]. Elle vient afin d'unifier la formulation de toutes les mesures de risque qu'on vient de citer auparavant et bien d'autres. Cette unification, a l'avantage d'estimer d'un seul coup toutes les mesures de risque unifiées en estimant seulement le Conditional Tail Moment.

Définition 4.15

Le Conditional Tail Moment au niveau $\alpha \in]0, 1[$ est défini par

$$\text{CTM}_a(\alpha) = \mathbb{E}(X^a \mid X > \text{VaR}(\alpha))$$

où $a \geq 0$ est tel que le moment d'ordre a de Y existe.

Le tableau suivant donne la réécriture de quelques mesures de risque de niveau α en fonction du CTM.

Mesure de risque	Expression comme une combinaison de $\text{CTM}_a(\alpha)$ et $\text{VaR}(\alpha)$
$\text{CTE}(\alpha)$	$\text{CTM}_1(\alpha)$
$\text{CVaR}_\lambda(\alpha)$	$\lambda \text{VaR}(\alpha) + (1 - \lambda) \text{CTM}_1(\alpha)$ where $\lambda \in [0, 1]$
$\text{SP}(\alpha)$	$(1 - \alpha) (\text{CTM}_1(\alpha) - \text{VaR}(\alpha))$
$\text{CTV}(\alpha)$	$\text{CTM}_2(\alpha) - \text{CTM}_1^2(\alpha)$
$\text{CTS}(\alpha)$	$\text{CTM}_3(\alpha) / (\text{CTM}_2(\alpha) - \text{CTM}_1^2(\alpha))^{3/2}$

TABLE 4.2 – Lien entre le CTM et certaines mesures de risque lorsque la répartition de X est continue.

Où

- SP est la Stop-loss Premium reinsurance qui est définie comme suit

$$\text{SP}(\alpha) = \mathbb{E}((X - \text{VaR}(\alpha))_+) = \alpha(\text{CTE}(\alpha) - \text{VaR}(\alpha))$$

- CTV est la Conditional Tail Variance qui est donnée par

$$\text{CTV}(\alpha) = \mathbb{E}((Y - \text{CTE}(\alpha))^2 | Y > \text{VaR}(\alpha))$$

- $\text{CTS}(\alpha)$ est la Conditional Tail Skewness dont sa définition est

$$\text{CTS}(\alpha) = \mathbb{E}(Y^3 | X > q(\alpha)) / (\text{CTV}(\alpha))^{3/2}$$

En fait, le CTM peut s'écrire dans le cas d'une fonction de répartition continue à l'aide de la fonction de distorsion

$$g(x) = \min\{1, x/(1 - \alpha)\}$$

et on aura pour tout ordre a strictement positif

$$\text{CTM}_a(\alpha) = \mathcal{R}_g(X^a).$$

Nous avons essayé de travailler sur cette nouvelle mesure de risque en lui donnant une estimation empirique en présence de données censurées. Notre travail n'est pas encore achevé. Les résultats préliminaires de ce travail ont fait l'objet de notre participation à la conférence internationale qui a eu lieu en 2019 à Croatie, mais comme le travail est en cours nous n'allons pas les présenter.

Troisième partie
Travaux de recherche

Chapitre 5

Improved Estimator of the Conditional Tail Expectation in the case of heavy-tailed losses

In this paper, we investigate the extreme-value methodology, to propose an improved estimator of the conditional tail expectation (*CTE*) for a loss distribution with a finite mean but infinite variance. The present work introduces a new estimator of the *CTE* based on the bias-reduced estimators of high quantile for heavy-tailed distributions. The asymptotic normality of the proposed estimator is established and checked, in a simulation study. Moreover, we compare, in terms of bias and mean squared error, our estimator with the known old estimator.

5.1 Introduction and Motivation

Risk management is a subject of concern in finance and actuarial science. Protecting against financial and actuarial risks is essential in order to anticipate financial crises or major insurance claims. For this, there are many tools to quantify and predict risk (risk measures). They make it possible to evaluate a level of danger of a risk, but also to compare different risks between them. Quantification, prevention, comparison and prediction of risk are essential elements of our society.

One of the best known and used risk measure is the Value-at-Risk (or *VaR*), it is introduced in the 1990s by Morgan [57] (see Jorion [47] for more details). The idea of the *VaR* is as follows : we fixe a threshold α and calculate a value $VaR(\alpha)$ which will be such that the probability that the catastrophe will occur is smaller than α . $VaR(\alpha)$ can be seen as the amount of extra capital and business needs in order to reduce the probability of going

bankrupt to α .

In statistical terms, the VaR of level $\alpha \in]0, 1[$ corresponds to the level $(1 - \alpha)$ quantile of the distribution function of the losses. This risk measure has several flaws. It gives only one point information to the quantile $q(\alpha)$ and no information beyond this point; it does not take into account the importance of the disaster when it occurs but only its frequency. To respond to the need for theoretical practical principles, Artzner et al. 1999 [8] introduced the concept of coherent risk measure. A risk measure is said to be coherent if it satisfies the four properties : monotony, positive homogeneity, invariance by translation and subadditivity, (for more details see, [8]).

Unfortunately, the VaR risk measure is not subadditive which implies that it is not coherent. For this reason, many authors proposed to replace standard VaR with alternative risk measures such as Conditional Tail Expectation (CTE), also called in the literature on risk measure Tail Value-at-Risk ($TVaR$) or Expected Shortfall (ES) (see, for example, [48], [32], [2], [3], [4], [66], [65], [77] and [79]).

This risk measures gives information of the distribution beyond the $VaR(\alpha)$. Thus, unlike the VaR , it takes into account the information contained in the tail of the distributi which is essential in the finanacial context. The CTE has been studied by many authors such that : [8], [21], [22], [72], [17], [52].

The application of these alternative measures has gained interest growing in literature and industry. In particular, the CTE , due to its properties and its effectiveness in different fields such as finance and actuaries, its use and studies keep increasing (see, e.g., [32], [71] and references therein).

In the literature, the estimation of the CTE and its asymptotic normality under the assumption that the second moment of the loss variable is finite has been established but this assumption is very restrictive in practical problems. For this reason, Necir et al (2010) [58] has extended that in the case of infinite variance. In this paper, we propose an improvement of the estimator established by Necir et al. (2010) [58], our considerations are based on the bias-reduced estimators of high quantile for heavy-tailed distributions introduced by Li (2010) [54], also we show its efficiency and its asymptotic normality theoretically, finally we prove the performance of our estimator by some results of simulation study. For this aim, we proceed as follow :

Let X be a loss random variable with cumulative distribution function (cdf) F . F is assumed to be continuous throughout the present paper and defined on the entire real line, with negative loss interpreted as gain. The CTE of the risk X is then defined, for every $t \in (0, 1)$, by the function $\mathbb{C}(t)$ writed as follow

$$\mathbb{C}(t) = \mathbf{E}(X | X > \mathbf{Q}(t)), \quad (5.1)$$

where $\mathbf{Q}(t) = \inf \{x : F(x) \geq t\}$ is the quantile function corresponding to the cdf F . Since

F is continuous, we easily check that $\mathbb{C}(t)$ is equal to

$$\mathbb{C}(t) = \frac{1}{1-t} \int_t^1 \mathbf{Q}(s) ds. \quad (5.2)$$

In this paper we are interesting by the heavy-tailed distribution. A model F is said to be heavy-tailed, with a tail index α ($\alpha > 0$), if the tail function $\bar{F} = 1 - F \in \mathcal{RV}_{-\alpha}$, where denotes the class of regularly varying functions with index of regular variation equal to α , i.e., non-negative measurable functions g such that, for all $x > 0$, $g(tx)/g(t) \rightarrow x^\alpha$, as $t \rightarrow \infty$ (see, [38] for more details).

Suppose that X_1, X_2, \dots, X_n are independent and identically distributed (i.i.d.) random variables with common distribution function F which has regularly varying tails with index $\alpha > 1$, i.e.,

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha}, \text{ for every } x > 0. \quad (5.3)$$

This class includes a number of popular distributions such as Pareto, generalized Pareto, Burr, Fréchet, Student, etc., which are known to be appropriate models for fitting large insurance claims, fluctuations of prices, log-returns, etc. (see, e.g., [13]). In this paper, we restrict ourselves to this class of distributions. For more information on the topic and, generally, on extreme value models and their manifold applications, we refer to the monographs by [27], [60] and [64].

In particular, if the index $\alpha \in (1, 2)$, then, condition (5.3) is equivalent to the statement that F has infinite second moment. In this case, in order to estimate the *CTE*, let $X_{1:n} < \dots < X_{n:n}$ denote the order statistics of X_1, \dots, X_n , we set

$$\tilde{\mathbb{C}}_n(t) = \frac{1}{1-t} \left[\int_t^{1-k/n} \mathbf{Q}_n(s) ds + \int_0^{k/n} \mathbf{Q}_n^w(1-s) ds \right], \quad (5.4)$$

where $\mathbf{Q}_n(s)$ is the empirical estimation of the quantile function, which is equal to the i^{th} order statistic $X_{i:n}$ for all $s \in ((i-1)/n, i/n)$ and for all $i = 1, \dots, n$ and $\mathbf{Q}_n^w(1-s)$ is the estimation of the extreme quantile established by [75], this estimator is given by

$$\mathbf{Q}_n^w(1-s) = \left(\frac{k}{n} \right)^{1/\hat{\alpha}_n} X_{n-k,n} s^{-1/\hat{\alpha}_n}. \quad (5.5)$$

where $\hat{\alpha}_n$ is the Hill estimator [45]

$$\hat{\alpha}_n = \left(\frac{1}{k} \sum_{i=1}^k \log X_{n-i+1:n} - \log X_{n-k:n} \right)^{-1} \quad (5.6)$$

of the tail index $\alpha \in (1, 2)$ and $k = k(n)$ is an intermediate integer sequence satisfying the condition

$$k \rightarrow \infty \text{ and } k/n \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (5.7)$$

After integration, we obtain

$$\tilde{\mathbf{C}}_n(t) = \frac{1}{1-t} \left[\int_t^{1-k/n} \mathbf{Q}_n(s) ds + \left(\frac{k}{n} \right) \frac{\hat{\alpha}_n X_{n-k,n}}{(\hat{\alpha}_n - 1)} \right]. \quad (5.8)$$

the asymptotic normality of this estimator is found in [58] under some conditions.

We note that, the Hill's estimator $\hat{\alpha}_n$ plays a pivotal role in statistical inference on distribution tails. This estimator has been extensively studied, improved and even generalized to any real parameter α . Weak consistency of $\hat{\alpha}_n$ was established by [55] assuming only that the underlying cdf F satisfies condition (5.3). The asymptotic normality of $\hat{\alpha}_n$ has been established (see [29]) under the following stricter condition that characterizes Hall's model (see [43] and [44]), where there exist $c > 0$, $d \neq 0$ and $\beta > \alpha > 0$ such that

$$1 - F(x) = cx^{-\alpha} + dx^{-\beta} + o(x^{-\beta}), \text{ as } x \rightarrow \infty, \quad (5.9)$$

Note that (5.9), which is a special case of a more general second-order regular variation condition (see [28]), is equivalent to

$$\mathbf{Q}(1-s) = c^{1/\alpha} s^{-1/\alpha} (1 + \alpha^{-1} c^{-\beta/\alpha} d s^{\beta/\alpha-1} + o(1)), \text{ as } s \downarrow 0, \quad (5.10)$$

The constants α and β are called, respectively, first-order (tail index, shape parameter) and second-order parameters of cdf F . In this paper, we use the bias-reduced estimator of the high quantile $\mathbf{Q}(1-s)$, proposed by [54] who exploited the censored maximum likelihood (CML) based estimators $\hat{\alpha}$ and $\hat{\beta}$ of the couple of regular variation parameters (α, β) introduced by [61]. The CML estimators $\hat{\alpha}$, $\hat{\beta}$ are defined as the solution of the two equations (under the constraint $\beta > \alpha$):

$$\frac{1}{k} \sum_{i=1}^k \frac{1}{G_i(\alpha, \beta)} = 1 \text{ and } \frac{1}{k} \sum_{i=1}^k \frac{1}{G_i(\alpha, \beta)} \log \frac{X_{n-i+1,n}}{X_{n-k,n}} = \beta^{-1}, \quad (5.11)$$

where

$$G_i(\alpha, \beta) = \frac{\alpha}{\beta} \left(1 + \frac{\alpha\beta}{\alpha - \beta} H(\alpha) \right) \left(\frac{X_{n-i+1,n}}{X_{n-k,n}} \right)^{\beta-\alpha} - \frac{\alpha\beta}{\alpha - \beta} H(\alpha), \quad (5.12)$$

and

$$H(\alpha) = \frac{1}{\alpha} - \frac{1}{k} \sum_{i=1}^k \log \frac{X_{n-i+1,n}}{X_{n-k,n}}. \quad (5.13)$$

Li et al. [54] obtained their bias-reduced estimators $\mathbf{Q}_n^L(1-s)$, of the high quantiles $\mathbf{Q}(1-s)$, by substituting $(\hat{\alpha}, \hat{\beta})$ to (α, β) in (5.10). That is

$$\hat{\mathbf{Q}}_n^L(1-s) = \hat{c}^{1/\hat{\alpha}} s^{-1/\hat{\alpha}} (1 + \hat{\alpha}^{-1} \hat{c}^{-\hat{\beta}/\hat{\alpha}} \hat{d} s^{\hat{\beta}/\hat{\alpha}-1} + o(1)), \text{ as } s \downarrow 0, \quad (5.14)$$

where

$$\left\{ \begin{array}{l} \hat{c} = \frac{\hat{\alpha}\hat{\beta}}{\hat{\alpha}-\hat{\beta}} \frac{k}{n} X_{n-k,n}^{\hat{\alpha}} \left(\frac{1}{\hat{\beta}} - \frac{1}{k} \sum_{i=1}^k \log \frac{X_{n-i+1,n}}{X_{n-k,n}} \right) \\ \hat{d} = \frac{\hat{\alpha}\hat{\beta}}{\hat{\alpha}-\hat{\beta}} \frac{k}{n} X_{n-k,n}^{\hat{\beta}} \left(\frac{1}{\hat{\alpha}} - \frac{1}{k} \sum_{i=1}^k \log \frac{X_{n-i+1,n}}{X_{n-k,n}} \right) \end{array} \right\}. \quad (5.15)$$

The consistency and asymptotic normality of $\widehat{\mathbf{Q}}_n^L(1-s)$ are established by the same authors. By using \mathbf{Q}_n in formula (5.4), we get, after integration, the new estimator of the *CTE* as follows :

$$\bar{\mathbf{C}}_n(t) = \frac{1}{1-t} \left[\int_t^{1-k/n} \mathbf{Q}_n(s) ds + \left(\frac{k}{n}\right) \left(\frac{n\widehat{c}}{k}\right)^{1/\widehat{\alpha}} \left(\frac{\widehat{\alpha}}{\widehat{\alpha}-1} + \frac{\widehat{d}\widehat{c}^{-\widehat{\beta}/\widehat{\alpha}} (k/n)^{\widehat{\beta}/\widehat{\alpha}-1}}{\widehat{\beta}-1} \right) \right], \quad (5.16)$$

provided that $\widehat{\beta} > \widehat{\alpha} > 1$ so that $\bar{\mathbf{C}}_n$ is finite.

The rest of the paper is organized as follows. In Section 5.2 we study the asymptotic normality of the new *CTE* estimator. In section 5.3, we make a simulation study of the new estimator and we show the performance of our estimator and compare it with the old one. We finish our paper by a conclusion given in section 5.4. The proof of the main result, which is Theorem 5.1 in Section 5.2, is postponed to Section 5.5.

5.2 Main result and its practical implementation

In the field of the extreme values theory, a function denoted by \mathbf{U} and (sometimes) called tail quantile function, is defined by

$$\mathbf{U}(t) = F^{\leftarrow}(1-1/t) = \mathbf{Q}(1-1/t), \quad (5.17)$$

where F^{\leftarrow} represented the generalized inverse of the df F . Then, we say that F is heavy-tailed iff $\mathbf{U} \in RV_{1/\alpha}$ (de Haan, 1970 [26]), i.e.

$$\lim_{t \rightarrow \infty} \frac{\mathbf{U}(tx)}{\mathbf{U}(t)} = x^{1/\alpha}, \text{ for any } x > 0. \quad (5.18)$$

In terms of this function, Hall's conditions (5.9) and (5.10) are equivalent to

$$\mathbf{U}(t) = c^{1/\alpha} t^{1/\alpha} (1 + \alpha^{-1} c^{-\beta/\alpha} dt^{1-\beta/\alpha} + o(1)), \text{ as } t \rightarrow \infty. \quad (5.19)$$

This implies that, there is a function $A_1(t)$, which tends to zero as $t \rightarrow \infty$ (because $\beta > \alpha$), determines the rate of convergence of $\log(\mathbf{U}(tx)/\mathbf{U}(t))$ to its limit $\alpha^{-1} \log x$, such that :

$$\lim_{t \rightarrow \infty} \frac{\log(\mathbf{U}(tx)/\mathbf{U}(t)) - \alpha^{-1} \log x}{A_1(t)} = \frac{x^{1-\beta/\alpha} - 1}{1 - \beta/\alpha}; \text{ for any } x > 0, \quad (5.20)$$

where

$$A_1(t) = d\alpha^{-1} (1 - \beta/\alpha) c^{-\beta/\alpha} t^{1-\beta/\alpha}.$$

Relation (5.20) is known as the second-order condition of regular variation (see, e.g., [27]). Unfortunately, the second-order regular variation is not sufficient to find asymptotic distributions for the estimators defined by the systems (5.11) and (5.15). We strengthen

it with a condition, called third-order condition of regular variation and given by (5.21), that specifies the rate of (5.20) (see, e.g., [28] or [7]).

$$\lim_{t \rightarrow \infty} \frac{\frac{\log(\mathbf{U}(tx)/\mathbf{U}(t)) - \alpha^{-1} \log x}{A_1(t)} - \frac{x^{1-\beta/\alpha} - 1}{1-\beta/\alpha}}{A_2(t)} = D(\alpha, \beta, \rho) \frac{x^{1-\beta/\alpha} - 1}{1-\beta/\alpha}; \text{ for any } x > 0. \quad (5.21)$$

where $A_2(t) \rightarrow 0$ as $t \rightarrow \infty$, with constant sign near infinity and

$$D(\alpha, \beta, \rho) = \frac{1}{\rho} \left(\frac{x^{1-\beta/\alpha+\rho} - 1}{1-\beta/\alpha+\rho} - \frac{x^{1-\beta/\alpha} - 1}{1-\beta/\alpha} \right).$$

with ρ being a positive constant called third-order parameter. Peng and Qi, [61] established the asymptotic normality of $\hat{\alpha}$, $\hat{\beta}$ and \hat{c} under the following extract conditions on the sample fraction k , as $n \rightarrow \infty$:

$$(i) : k^{1/2} |A_1(n/k)| \rightarrow \infty, \quad (ii) : k^{1/2} A_1^2(n/k) \rightarrow 0 \text{ and } (iii) : k^{1/2} A_1(n/k) A_2(n/k) \rightarrow 0. \quad (5.22)$$

As for \hat{d} , it is asymptotically normal under the assumption $k^{1/2} |A_1(n/k)| / \log(n/k) \rightarrow \infty$ added to (ii) and (iii). Note that, from a theoretical point of view, assumptions (5.7) and (5.22) are realistic, as the following example shows, indeed, let us choose

$$k = [n^{1-\varepsilon}], 0 < \varepsilon < 1, \quad (5.23)$$

then it easy to verify that these assumptions hold for any $1/5 < \varepsilon < 1/3$. The notation $[\cdot]$ stands for the integer part of real numbers.

Our main result, namely the asymptotic normality of new estimator of the *CTE* is formulated in the following theorem.

Théorème 5.1

Assume that the cdf F satisfies condition (5.21) with $\alpha \in (1, 2)$ and $\beta/\alpha = \lambda > 1$. Then for any sequence of integers $k = k_n$ satisfying the conditions (5.7) and (5.22). Then

$$\frac{\sqrt{n}(\overline{\mathbb{C}}_n(t) - \mathbb{C}(t))(1-t)}{(k/n)^{1/2} (nc/k)^{1/\alpha}} \rightarrow_d \mathcal{N}(0, \sigma^2(\alpha, \beta)) \quad (5.24)$$

for any fixed $t \in (0, 1)$, where the asymptotic variance $\sigma^2(\alpha, \beta)$ is given by the formula

$$\sigma^2(\alpha, \beta) = \frac{\alpha^2 \beta^4}{(\alpha-1)^4 (\alpha-\beta)^4} + \frac{2}{2-\alpha} + \frac{2\alpha\beta^2}{(\alpha-1)^2 (\alpha-\beta)^2}.$$

5.3 Simulation study

In this section, the biased estimator $\tilde{\mathbb{C}}_n(t)$ and the reduced-bias one $\overline{\mathbb{C}}_n(t)$ are compared using simulation study. For this reason, 1000 samples of size $n \in \{250, 500, 1000, 2000\}$ are simulated from a two heavy-tailed distributions :

Fréchet model : defined as : $F(x) = \exp(-x^{-\alpha})$ we take $\alpha = 1,5$ and $\alpha = 1,75$, respectively, and according to the Hall's model, we find $\beta = 2\alpha, c = 1$, and $d = -1/2$.

Burr model : defined as : $F(x) = 1 - (1 + x^\tau)^{-\lambda}$ with $\lambda = 1,5$ and $\lambda = 1,75$, both with $\tau = 1$ and in this case we have $c = 1, \alpha = \tau\lambda, d = -\lambda$ and $\beta = \lambda\tau + \tau$ (For more flexibility, we can take $\tau = \alpha/\lambda$ in the expression of Burr distribution).

In all cases, we assume that α, β, c and d are unknowns, then by the resolving the equations (5.11) and (5.12) we calculate the values of $\hat{\alpha}$ and $\hat{\beta}$, and resolving the system (5.15) to calculate the estimators \hat{c} and \hat{d} . We fix two values of t , for example $t = 0.90$ and $t = 0.95$ to calculate the different values of the two estimators, the new estimator \bar{C}_n and the old estimator \tilde{C}_n . We note that, the mean, the bias and root mean squared error (RMSE) of these estimators are estimated over the 1000 replications. We illustrate and compare the bias and the root mean square error (RMSE) of $\tilde{C}_n(t)$ and $\bar{C}_n(t)$. Finally, we summarize the results of simulations in the tables (5.1, 5.2, 5.3, and 5.4). We remark, on one hand, that the bias and the RMSE of the new estimator increase when the sample size increase, and on other hands, the bias and RMSE of our estimator is smaller than the old estimator. In order to show the influence of the choice of the integer value k to the performance of our estimator $\bar{C}_n(t)$ and the old estimator $\tilde{C}_n(t)$, we generate 1000 samples of size 1000 of the parent model (Fréchet and Burr) with two values of index $\alpha = 1.5$ and 1.75 and we picture the two estimators when k varied from 50 to 850. The results are displayed in figures 5.1 and 5.2. We observe that the performance of our estimator is clearly better than the old estimator.

TABLE 5.1 – Comparison between the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in terms of bias and RMSE respecting to the variation of sample size n based on Fréchet distribution.

Fréchet distribution, $\alpha = 1.5$								
t	n	CTE	$\bar{C}_n(t)$	bias	RMSE	$\tilde{C}_n(t)$	bias	RMSE
0.9	250	13.793	14.457	-0.664	0.66415	11.048	2.745	2.7454
	500		14.409	-0.616	0.61578	11.418	2.375	2.3755
	1000		14.287	-0.494	0.49345	11.837	1.956	1.9559
	2000		14.14	-0.347	0.34733	11.852	1.941	1.941
0.95	250	21.984	22.78	-0.796	0.79685	16.115	-5.869	5.8683
	500		22.703	-0.719	0.71965	16.473	-5.511	5.5105
	1000		22.674	-0.69	0.69043	16.944	-5.04	5.0392
	2000		22.67	-0.686	0.68661	17.247	-4.737	4.7364

TABLE 5.2 – Comparison between the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in terms of bias and RMSE respecting to the variation of sample size n based on Fréchet distribution.

Fréchet distribution, $\alpha = 1.75$								
t	n	CTE	$\bar{C}_n(t)$	bias	RMSE	$\tilde{C}_n(t)$	bias	RMSE
0.9	250	8.6207	8.6977	-0.077	0.076993	8.2478	0.3729	0.37295
	500		8.6771	-0.0564	0.05637	8.3136	0.3071	0.30711
	1000		8.6586	-0.0379	0.037901	8.5085	0.1122	0.11218
	2000		8.6504	-0.0297	0.029695	8.5374	0.0833	0.083322
0.95	250	12.866	12.963	-0.097	0.096298	12.185	0.681	0.68126
	500		12.956	-0.09	0.089431	12.218	0.648	0.64858
	1000		12.937	-0.071	0.070963	12.624	0.242	0.24248
	2000		12.917	-0.051	0.050889	12.766	0.100	0.10069

TABLE 5.3 – Comparison between the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in terms of bias and RMSE respecting to the variation of sample size n based on Burr distribution.

Burr distribution, $\alpha = 1.5$								
t	n	CTE	$\bar{C}_n(t)$	bias	RMSE	$\tilde{C}_n(t)$	bias	RMSE
0.9	250	13.676	14.365	-0.689	0.68807	14.909	-1.233	1.2321
	500		14.226	-0.55	0.54908	14.658	-0.982	0.98132
	1000		14.124	-0.448	0.44795	14.506	-0.83	0.82922
	2000		14.026	-0.35	0.34997	14.491	-0.815	0.81428
0.95	250	21.891	23.239	-1.348	1.3474	24.05	-2.159	2.1587
	500		22.996	-1.105	1.1044	23.874	-1.983	1.9827
	1000		22.753	-0.862	0.86155	23.709	-1.818	1.8173
	2000		22.602	-0.711	0.71095	23.64	-1.749	1.7489

TABLE 5.4 – Comparison between the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in terms of bias and RMSE respecting to the variation of sample size n based on Burr distribution.

Burr distribution, $\alpha = 1.75$								
t	n	CTE	$\bar{C}_n(t)$	bias	RMSE	$\tilde{C}_n(t)$	bias	RMSE
0.9	250	8.5455	9.0332	-0.4877	0.48776	9.3583	-0.8128	0.81281
	500		8.9188	-0.3733	0.37334	9.0502	-0.5047	0.5047
	1000		8.8353	-0.2898	0.28979	8.9942	-0.4487	0.44873
	2000		8.7708	-0.2253	0.22533	8.9115	-0.366	0.36607
0.95	250	12.811	13.697	-0.886	0.88641	14.386	-1.575	1.5751
	500		13.556	-0.745	0.74521	14.082	-1.271	1.2709
	1000		13.381	-0.57	0.56966	13.788	-0.977	0.97676
	2000		13.246	-0.435	0.4355	13.596	-0.785	0.78538

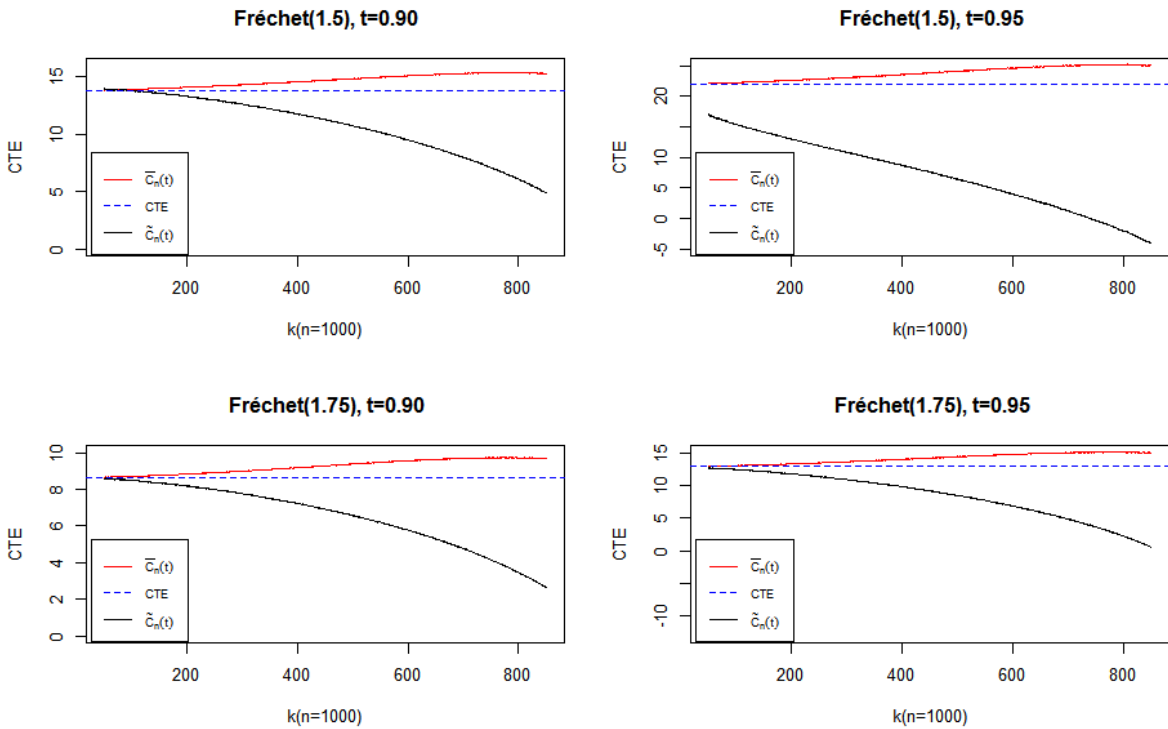


FIGURE 5.1 – Behavior of the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in respect to the variation of k for Fréchet distribution.

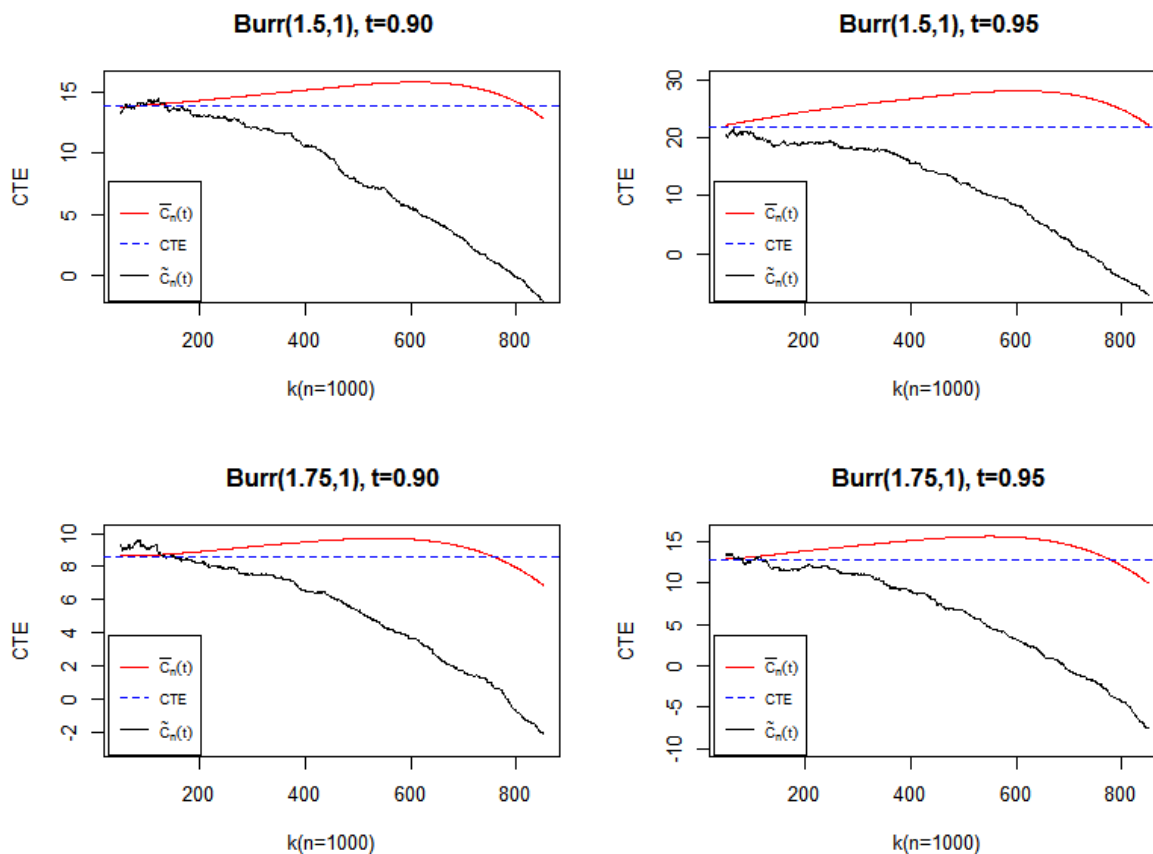


FIGURE 5.2 – Behavior of the new estimator \bar{C}_n and old estimator \tilde{C}_n of CTE in respect to the variation of k for Burr distribution.

The analysis and simulation as shown in tables 5.1, 5.2, 5.3, 5.4, and figures 5.1, 5.2 indicate that the new CTE estimator is more performant than the estimator constructed by Necir et al (2010) [58].

5.4 Conclusion

From the research that has undertaken, it is possible to conclude that the estimation of the CTE by the use of extreme quantile established by Li (2010) [54] gave us a more efficient estimate compared to that introduced by Necir et al (2010) [58]. The results of the simulations show significant improvement either in term of bias or in term of RMSE. The next stage of our research will be the construction of a confidence interval of our CTE estimator with the covering probability.

5.5 Proof of Theorem 5.1

We start the proof of Theorem 5.1 with the decomposition

$$(1-t)\mathbb{C}(t) = C_{1,n}(s) + C_{2,n}(s),$$

where

$$C_{1,n}(s) = \int_t^{1-k/n} \mathbf{Q}(s) ds \text{ and } C_{2,n}(s) = \int_0^{k/n} \mathbf{Q}(1-s) ds.$$

Also, we have

$$(1-t)\overline{\mathbb{C}}(t) = \overline{C}_{1,n}(s) + \overline{C}_{2,n}(s),$$

where

$$\overline{C}_{1,n}(s) = \int_t^{1-k/n} \mathbf{Q}_n(s) ds,$$

and

$$\overline{C}_{2,n}(s) = \left(\frac{k}{n}\right) \left(\frac{n\widehat{c}}{k}\right)^{1/\widehat{\alpha}} \left(\frac{\widehat{\alpha}}{\widehat{\alpha}-1} + \frac{\widehat{d}\widehat{c}^{-\widehat{\beta}/\widehat{\alpha}}(k/n)^{\widehat{\beta}/\widehat{\alpha}-1}}{\widehat{\beta}-1}\right).$$

First, it is easy to verify that, as $n \rightarrow \infty$

$$C_{2,n}(s) = (1 + o_p(1)) \left(\frac{k}{n}\right) \left(\frac{nc}{k}\right)^{1/\alpha} \left(\frac{\alpha}{\alpha-1}\right),$$

and, under the condition (5.6), we have

$$\overline{C}_{2,n}(s) = (1 + o_p(1)) \left(\frac{k}{n}\right) \left(\frac{n\widehat{c}}{k}\right)^{1/\widehat{\alpha}} \left(\frac{\widehat{\alpha}}{\widehat{\alpha}-1}\right).$$

It follows that

$$\overline{C}_{2,n}(s) - C_{2,n}(s) = (1 + o_p(1)) \left(\frac{k}{n}\right) \left[\frac{\widehat{\alpha}}{\widehat{\alpha}-1} \left(\frac{n\widehat{c}}{k}\right)^{1/\widehat{\alpha}} - \frac{\alpha}{\alpha-1} \left(\frac{nc}{k}\right)^{1/\alpha} \right], \quad (5.25)$$

let us write $\overline{C}_{2,n}(s) - C_{2,n}(s) = A_{n,1}(s) + A_{n,2}(s)$, where

$$A_{n,1}(t) = (1 + o_p(1)) \sqrt{k} \left(\frac{\widehat{\alpha}}{\widehat{\alpha}-1} - \frac{\alpha}{\alpha-1}\right), \quad A_{n,2} = (1 + o_p(1)) \sqrt{k} \left(\frac{(n\widehat{c}/k)^{1/\widehat{\alpha}}}{(nc/k)^{1/\alpha}} - 1\right).$$

We begin by showing that $A_{n,2} \rightarrow 0$, as $n \rightarrow \infty$. First observe that $A_{n,2}$ may be rewritten into

$$A_{n,2} = (1 + o_p(1)) \sqrt{k} \left((nc/k)^{1/\widehat{\alpha}-1/\alpha} (\widehat{c}/c)^{1/\widehat{\alpha}-1/\alpha} - 1 \right)$$

Assumptions (i) and (ii), imply that $\sqrt{k}/\log(n/k) \rightarrow \infty$. Also, from Theorem 1 of [61], the asymptotic normality of $\widehat{\alpha}$ gives $\widehat{\alpha} - \alpha = O(1/\sqrt{k})$. Therefore

$$(1/\widehat{\alpha} - 1/\alpha) \log(nc/k) \xrightarrow{P} 0,$$

this implies that

$$(nc/k)^{1/\hat{\alpha}-1/\alpha} \xrightarrow{P} 1 \text{ as } n \rightarrow \infty.$$

On the other hand, from equation (4.7) in [54], we have

$$(\hat{c}/c) - 1 = \alpha^{-1} (1 + o_p(1)) (\hat{\alpha} - \alpha) \log(n/k) + o_p\left(\frac{1}{\sqrt{k}} \log \frac{n}{k}\right).$$

Since \hat{c} is a consistent estimator of c , then Taylor's expansion gives

$$(\hat{c}/c)^{1/\hat{\alpha}-1/\alpha} - 1 = \alpha^{-1} (\hat{\alpha} - \alpha) \log(n/k) (\hat{c}/c - 1), \text{ as } n \rightarrow \infty.$$

It suffices now to show that $\sqrt{k} \left((\hat{c}/c)^{1/\hat{\alpha}-1/\alpha} - 1 \right)$ converges to 0 in probability. Indeed, again by using the fact that $\hat{\alpha} - \alpha = O_p\left(1/\sqrt{k}\right)$, yield

$$\sqrt{k} \left((\hat{c}/c)^{1/\hat{\alpha}-1/\alpha} - 1 \right) = O_p(1) \left(\frac{1}{\sqrt{k}} \log \frac{n}{k} + o_p\left(\frac{\log(n/k)}{\sqrt{k}}\right) \right),$$

which tends in probability to 0, because we already have $\sqrt{k}/\log(n/k) \rightarrow \infty$. Now, we consider the term $A_{n,1}$. Since $\hat{\alpha}$ is a consistent estimator of α , then it is easy to show that

$$A_{n,1} = - (1 + o_p(1)) \frac{\sqrt{k}}{(\alpha - 1)^2} (\hat{\alpha} - \alpha).$$

From theorem 2.1 in [16], we infer that

$$\frac{\sqrt{n}(1-t)}{(k/n)^{1/2} (nc/k)^{1/\alpha}} (\bar{C}_{2,n}(s) - C_{2,n}(s)) = -\frac{\alpha}{(\alpha-1)^2} [\eta_1 W_1 + \eta_2 W_2 + \eta_3 W_3] + o_{\mathbf{P}}(1),$$

where

$$\begin{aligned} W_1 &= \sqrt{n/k} B_n(1-k/n) - \sqrt{n/k} \int_0^1 s^{-1} B_n(1-ks/n) ds, \\ W_2 &= (\lambda^{-1} - 1) \sqrt{n/k} B_n(1-k/n) + (\lambda - 1) \sqrt{n/k} \int_0^1 s^{\lambda-2} B_n(1-ks/n) ds, \\ W_3 &= (1 - \lambda) \sqrt{n/k} \int_0^1 s^{\lambda-2} (\log s) B_n(1-ks/n) ds \\ &\quad + \lambda^{-2} \sqrt{n/k} B_n(1-k/n) - \sqrt{n/k} \int_0^1 s^{\lambda-2} B_n(1-ks/n) ds, \end{aligned}$$

and

$$\eta_1 = \frac{\lambda^4}{(\lambda-1)^4}, \eta_2 = \frac{\lambda^2(2\lambda-1)(3\lambda-1)}{(\lambda-1)^5}, \eta_3 = \frac{\lambda^3(2\lambda-1)^2}{(\lambda-1)^4}.$$

From the Proof of statement (4.2) in [58], we shall show below that there are Brownian bridges B_n such that

$$\frac{\sqrt{n}(1-t)}{(k/n)^{1/2} \mathbf{Q}(1-k/n)} (\bar{C}_{1,n}(s) - C_{1,n}(s)) = -\frac{\int_{k/n}^1 B_n(1-s) d\mathbf{Q}(1-s)}{(k/n)^{1/2} \mathbf{Q}(1-k/n)} + o_{\mathbf{P}}(1) \quad (5.26)$$

On the other hand, from (5.10), we have $\mathbf{Q}(1 - k/n) \sim (nc/k)^{1/\alpha}$, as $n \rightarrow \infty$, it follows that

$$\frac{\sqrt{n}(1-t)}{(k/n)^{1/2}(nc/k)^{1/\alpha}} (\bar{C}_{1,n}(s) - C_{1,n}(s)) = W_4 + o_{\mathbf{P}}(1),$$

where

$$W_4 = -\frac{\int_{k/n}^1 B_n(1-s) d\mathbf{Q}(1-s)}{(k/n)^{1/2} \mathbf{Q}(1-k/n)},$$

Finally, we have

$$\frac{\sqrt{n}(1-t)}{(k/n)^{1/2}(nc/k)^{1/\alpha}} (\bar{C}_n(t) - \mathbb{C}(t)) = -\frac{\alpha}{(\alpha-1)^2} [\eta_1 W_1 + \eta_2 W_2 + \eta_3 W_3] + W_4 + o_{\mathbf{P}}(1).$$

All done, after a standard calculation we obtain

$$\frac{\sqrt{n}(1-t)}{(k/n)^{1/2}(nc/k)^{1/\alpha}} (\bar{C}_n(t) - \mathbb{C}(t)) \rightarrow_d N(0, \sigma^2(\alpha, \beta)).$$

5.6 Confidence interval of improved Estimator

This section is a continuation of the last one of which we have shown the asymptotic normality of our improved estimator of the CTE.

In order to construct a confidence interval using the Theorem 5.1, we first fix a significance level $\zeta \in (0, 1)$ and use the classical notation $z_{\zeta/2}$ for the $(1 - \zeta/2)$ -quantile of the standard normal distribution $\mathcal{N}(0, 1)$.

Given a realization of the random variables X_1, \dots, X_n (e.g., claim amounts), which follow a cdf F satisfying the conditions of Theorem, we construct a level $1 - t$ confidence interval for $\mathbb{C}(t)$ as follows. First, we need to choose an appropriate number k^* of extreme values by choose for example $\varepsilon = 0.005$ an appropriate value k^* of k . Having computed Hill's estimator and consequently determined $X_{n-k^*:n}$, we then compute the corresponding values of $\bar{C}_n(t)$ and σ^2 , and denote them by $\bar{C}_n^*(t)$ and $\sigma^{2*}(\alpha^*, \beta^*)$, respectively. Finally, using Theorem ?? we arrive at the following $(1 - t)$ -confidence interval for $\mathbb{C}(t)$

$$\bar{C}_n^*(t) \pm z_{\zeta/2} \frac{(k^*/n)^{1/2} (nc/k^*)^{1/\alpha} \sigma^*(\alpha, \beta)}{(1-t)\sqrt{n}}$$

To use this formula, we used frechet distribution with tail index $\alpha = 1.5$ and a 95% confidence threshold by changing the sample sizes from 1 to 2000. The results of this simulation are shown in the following figure.

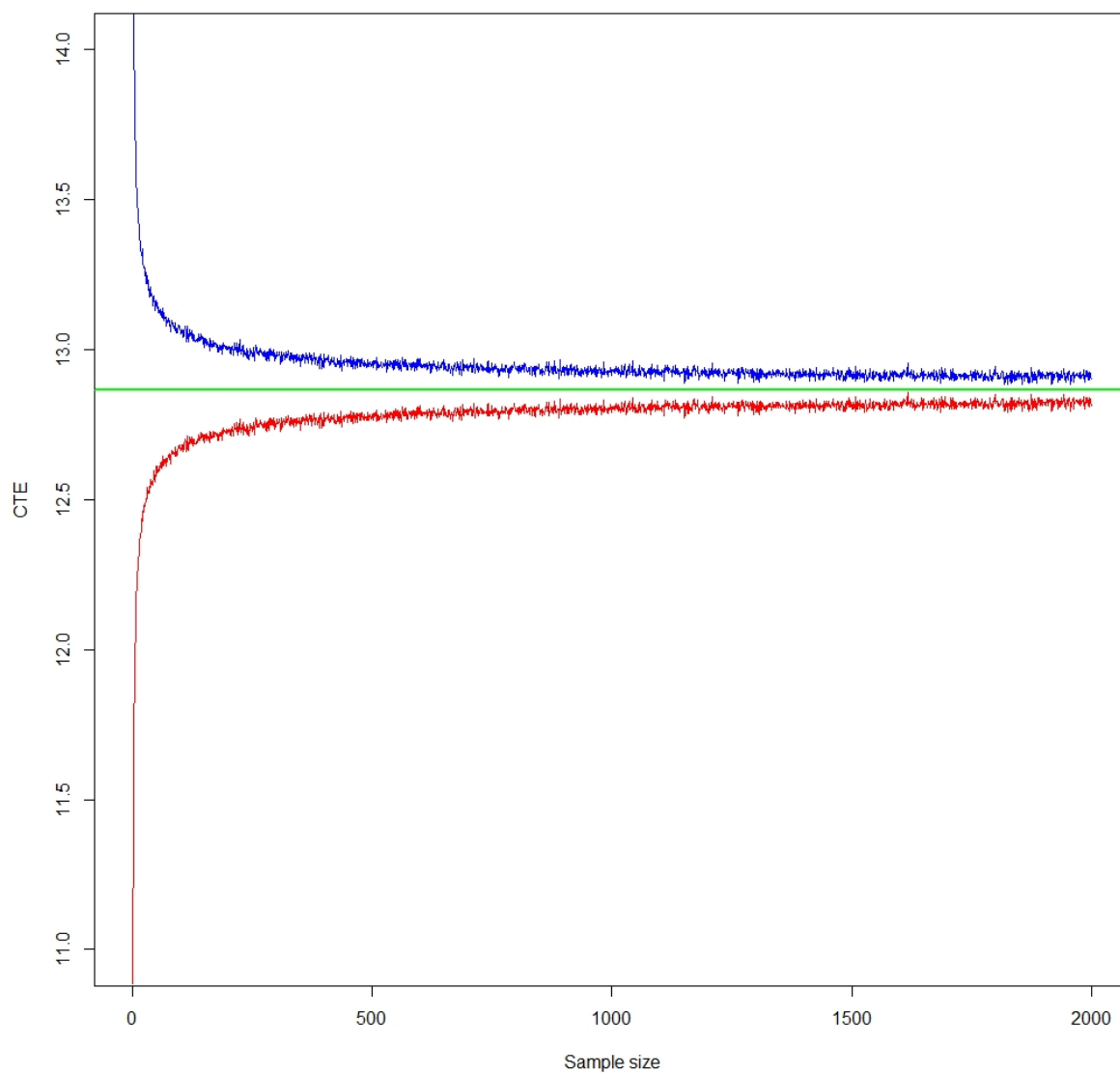


FIGURE 5.3 – 95%-Confidence interval of improved Estimator $\bar{C}_n(t)$ in respect of sample size for Fréchet distribution with $\alpha = 1.5$.

Conclusion et Perspectives

"The key message is that extreme value theory cannot do magic, but it can do a whole lot better than empirical curvefitting and guesswork. My answer to the sceptics is that people aren't given well-founded methods like extreme value theory, they will just use dubious ones instead."

Jonathan Tawn

Cette thèse nous a permis à découvrir et à étudier une théorie non conventionnelle de la statistique. Une théorie qui ne s'appuie pas sur la tendance centrale d'un échantillon tant utilisée dans le cas classique, mais du comportement de leurs valeurs les plus éloignées de son centre qu'on appelle les valeurs extrêmes et d'où le nom de cette théorie : la Théorie des Valeurs Extrêmes.

C'est une théorie qui est en étroite relation avec les risques car, en général, un risque est un événement rare et donc se trouve dans la queue de distribution du phénomène étudié ou d'une autre manière, il est une valeur extrême. Quand on parle d'un risque, on doit mesurer sa dangerosité, et donc on doit définir des mesures de risque pour y parvenir. Parmi ces mesures, il y a la CTE qui est une substituable de la VaR souvent utilisée dans la littérature. Nous avons contribué à améliorer son estimateur classique proposé par Lie et al. (2010) [54] par réduction de biais. Une amélioration qui a donné naissance à un nouveau estimateur nettement meilleur que le classique et donc plus efficace pour estimer la CTE. Une efficacité montrée théoriquement et par simulation.

Nous visons dans les prochaines travaux :

- Reprendre le même travail en introduisant la notion des données censurées dans la base de données.
- L'étude comparative des méthodes d'extrapolation des données de la queue et leur impacts sur la qualité des estimateurs des mesures de risque.
- L'étude d'une nouvelle mesure risque introduite par El Methni [56] dite conditional Tail Mean (CTM) qui est une généralisation de la CTE.

Quatrième partie

Annexes

Annexe A

Définitions et propriétés indispensables

A.1 Convergence de variables aléatoires

A.1.1 Modes de convergences

On considère l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Soient $X, X_n, n \geq 1$ des variables aléatoires définies sur cette espace.

Converge Presque Sûre

Définition A.1

On dit que X_n converge presque sûrement (p.s.) vers X et on écrit $X_n \xrightarrow{p.s.} X$, si

$$\lim_{n \rightarrow \infty} X_n(\omega) \rightarrow X(\omega) \text{ pour tout } \omega \in A \text{ où } \mathbb{P}(\bar{A}) = 0,$$

ou d'une façon équivalente, si pour tout $\varepsilon > 0$

$$\mathbb{P}\left(\sup_{m \geq n} |X_m - X| > \varepsilon\right) \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Converge en Probabilité

Définition A.2

On dit que X_n converge en probabilité vers X et on écrit $X_n \xrightarrow{\mathbb{P}} X$, si pour tout $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Condition suffisante de convergence en probabilité

$$\lim_{n \rightarrow \infty} \mathbb{E}(X_n) = c \text{ et } \lim_{n \rightarrow \infty} V(X_n) = 0 \Rightarrow X_n \xrightarrow{\mathbb{P}} c.$$

Propriété A.1

$$X_n \xrightarrow{\mathbb{P}} X \text{ et } Y_n \xrightarrow{\mathbb{P}} Y \Rightarrow X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$$

Convergence en moyenne d'ordre p **Définition A.3**

Soit $0 < p < \infty$. On dit que X_n convergence en moyenne d'ordre p ou en norme L^p vers X et on écrit $X_n \xrightarrow{L^p} X$, si

$$\mathbb{E}(|X_n - X|^p) \rightarrow 0 \text{ quand } n \rightarrow \infty \text{ pour } X_n, X \in L^p(\mathbb{P}).$$

Convergence en Loi**Définition A.4**

On dit que X_n convergence en Loi, ou en Distribution vers X et on écrit $X_n \xrightarrow{\mathcal{L}} X$, ou $X_n \xrightarrow{d} X$ si les distributions F_n et F de X_n et X respectivement satisfont

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \text{ pour tout point } x \text{ de continuité de } F.$$

On note $(\phi_n)_{n \in \mathbb{N}}$ la suite des fonctions caractéristiques de $(X_n)_{n \in \mathbb{N}}$ et ϕ celle de X : $\phi_n(t) = \mathbb{E}(e^{itX_n})$.

Théorème A.1 (Théorème de Paul Levy)

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \forall x \in \mathbb{R}, \lim_{n \rightarrow \infty} \phi_n(x) = \phi(x).$$

Lemme A.1 (Lemme de Portmanteau)

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \forall g \text{ continue et bornée, } \lim_{n \rightarrow \infty} \mathbb{E}(g(X_n)) = \mathbb{E}(g(X)).$$

Théorème A.2

$$\forall c \in \mathbb{R}, X_n \xrightarrow{\mathcal{L}} X \text{ et } Y_n \xrightarrow{\mathcal{L}} c \Rightarrow (X_n, Y_n) \xrightarrow{\mathcal{L}} (X, c).$$

Théorème A.3 (Théorème de Slutsky)

$$\forall c \in \mathbb{R}, X_n \xrightarrow{\mathcal{L}} X \text{ et } Y_n \xrightarrow{\mathcal{L}} c \Rightarrow$$

1. $X_n + Y_n \xrightarrow{\mathcal{L}} X + c$,
2. $X_n Y_n \xrightarrow{\mathcal{L}} Xc$,
3. $X_n/Y_n \xrightarrow{\mathcal{L}} X/c, c \neq 0$.

Attention ! $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y \not\Rightarrow X_n + Y_n \xrightarrow{\mathcal{L}} X + Y$

A.1.2 Liens entre modes de convergences

Définition A.5 (Intégrabilité uniforme)

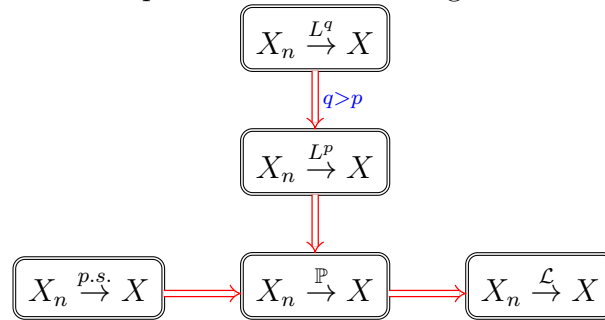
Une suite de variables aléatoires $(X_n)_{n \in \mathbb{N}}$ est dit uniformément intégrable si

$$\lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{E}(|X_n| \geq \alpha) = 0.$$

Théorème A.4 (Implications de la convergence)

1. Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.
2. Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_{n'} \xrightarrow{p.s.} X$ pour une sous-suite $(n')_{n' \in \mathbb{N}}$.
3. Si $X_n \xrightarrow{L^r} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.
4. Si $X_n \xrightarrow{\mathbb{P}} X$ et $|X_n^r|$ est uniformément intégrable, alors $X_n \xrightarrow{L^r} X$.
Si $X_n \xrightarrow{\mathbb{P}} X$ et $\limsup_{n \rightarrow \infty} \mathbb{E}(|X_n|^r) \leq \mathbb{E}(|X|^r)$, alors $X_n \xrightarrow{L^r} X$.
5. Si $X_n \xrightarrow{L^r} X$ alors $X_n \xrightarrow{L^{r'}} X$ pour tout $0 < r' \leq r$.
6. Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{d} X$.
7. $X_n \xrightarrow{\mathbb{P}} X$ si et seulement si chaque sous-suite $(n')_{n' \in \mathbb{N}}$ contient une autre sous-suite $(n'')_{n'' \in \mathbb{N}}$ pour laquelle $X_{n''} \xrightarrow{p.s.} X$.

Nous donnons l'essentiel des implications de la convergence dans le diagramme suivant



La convergence en Loi est la convergence la plus faible

A.1.3 Remplacement de la convergence en loi par la convergence presque sûre

Définition A.6

Soit F une fonction de distribution. Pour tout $0 < t < 1$

$$F^{-1}(t) := \inf\{x : F(x) \geq t\}.$$

Proposition A.1

F^{-1} est continue à droite.

Proposition A.2

Si X a une loi de distribution F qui est continue, alors $F(X) \sim \mathcal{U}(0, 1)$. Si la continuité de F n'est pas assurée, on a pour tout $t \in (0, 1)$,

$$\mathbb{P}(F(X) \leq t) \leq t,$$

avec une égalité si et seulement si t appartient à l'image de F .

Proposition A.3 (Théorème de Skorokhod)

On suppose que $X_n \xrightarrow{d} X$. Alors il existe des variables aléatoires $X_n^*, n \geq 0$ définies sur l'espace de probabilité commun $([0, 1], \mathcal{B}[0, 1], \lambda)$ pour lesquelles $X_n^* \stackrel{d}{=} X_n$ pour tout $n \geq 0$ et $X_n^* \xrightarrow{p.s.} X_0^*$.

Théorème A.5 (Théorème de Mann-Wald)

Soit g une fonction continue.

1. $X_n \xrightarrow{p.s.} X \Rightarrow g(X_n) \xrightarrow{p.s.} g(X)$,
2. $X_n \xrightarrow{\mathbb{P}} X \Rightarrow g(X_n) \xrightarrow{\mathbb{P}} g(X)$,
3. $X_n \xrightarrow{\mathcal{L}} X \Rightarrow g(X_n) \xrightarrow{\mathcal{L}} g(X)$.

A.1.4 Lois fondamentales de la statistique

Soit $(X_i)_{1 \leq i \leq n}$ une suite de variables aléatoires i.i.d. d'espérance μ et de variance σ^2 . Soit

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Loi des grands nombres

1. Loi faible des grands nombres

$$\bar{X} \xrightarrow{\mathbb{P}} \mu.$$

2. Loi forte des grands nombres

$$\bar{X} \xrightarrow{p.s.} \mu.$$

Théorème central limite (TCL)

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Annexe B

Statistique d'ordre

Définition B.1

Soit (X_1, \dots, X_n) n variables aléatoires indépendantes identiquement distribuées de distribution commune F et de densité f . On appelle statistique d'ordre (croissante) la suite des variables aléatoires (X_1, \dots, X_n) qui sont rangées par ordre croissant, soit :

$$X_{1:n} \leq \dots \leq X_{n:n}$$

Remarque B.1

Pour $1 \leq k \leq n$ la variable $X_{k:n}$ est connue sous le nom de la k^{eme} statistique d'ordre ou statistique d'ordre k . Deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des événements extrêmes. Ce sont les statistiques d'ordre extrêmes qui sont données par la définition suivante.

Définition B.2 (Statistiques d'ordre extrême)

Les statistiques d'ordre extrême sont définies comme termes du maximum et du minimum des n variables aléatoires (X_1, \dots, X_n) la variable $X_{1:n}$ est la plus petite statistique d'ordre (ou statistique du minimum) et $X_{n:n}$ est la plus grande statistique d'ordre (ou statistique du maximum) $M_n = \max(X_1, \dots, X_n)$ et $m_n = \min(X_1, \dots, X_n)$

B.1 Fonction de répartition empirique et statistiques d'ordre

- $F_n(t)$ est la proportion des n variables qui sont inférieurs ou égales à t .
- $\overline{F}_n(t)$ c'est la proportion d'observations qui dépasse à t .

Pour $1 \leq i \leq n$, les fonctions $F_n(t)$ et $\overline{F}_n(t)$ s'écrivent en termes des valeurs des statistiques d'ordre comme suit

$$F_n(t) = \begin{cases} 0 & \text{si } t \leq X_{1:n} \\ \frac{i}{n} & \text{si } X_{i:n} \leq t \leq X_{i+1:n} \\ 1 & \text{si } t \geq X_{n:n} \end{cases}$$

et

$$\overline{F}_n(t) = \begin{cases} 1 & \text{si } t \leq X_{1:n} \\ 1 - \frac{i}{n} & \text{si } X_{i:n} \leq t \leq X_{i+1:n} \\ 0 & \text{si } t \geq X_{n:n} \end{cases}$$

B.1.1 Fonction de répartition et densité du maximum

Proposition B.1 (Balakrishnan & Nagaraja)

— La fonction de distribution (de répartition) F_{M_n} de M_n est donnée par :

$$\forall x \in \mathbb{R}, \quad F_{M_n}(x) = \mathbb{P}(M_n \leq x) = F^n(x) \quad (\text{B.1})$$

— Si X est absolument continue de densité f , alors la fonction de densité f_{M_n} de M_n est donnée par :

$$\forall x \in \mathbb{R}, \quad f_{M_n}(x) = nF^{n-1}(x)f(x) \quad (\text{B.2})$$

Remarque B.2

Par analogie, on peut facilement trouver la répartition et la densité du minimum et donner par :

$$\forall x \in \mathbb{R}, \quad F_{m_n}(x) = \mathbb{P}(m_n \leq x) = 1 - \mathbb{P}(m_n > x) = 1 - (1 - F(x))^n \quad (\text{B.3})$$

Remarque B.3

D'après la formule (B.3), les propriétés statistiques de M_n , données par sa fonction de distribution, dépendant principalement des propriétés de F pour les grandes valeurs de x . Pour les autres valeurs de x , l'influence de F décroît rapidement avec la croissance de n . Donc, l'information la plus importante à propos des extrêmes est contenue dans la queue de la loi de probabilité de X . La loi de probabilité de M_n devait nous fournir des information au sujet des événements extrêmes.

M_n tend vers un nombre réel qui peut être infini, qu'on appelle point extrême ou point terminal de la fonction de distribution F , quand $n \rightarrow \infty$.

B.1.2 Point extrême

Définition B.3

On note par x_F (resp. x_F^*) le point extrême supérieur (resp. inférieur) de la distribution F (i.e. la plus grande valeur possible pour $X_{k:n}$ qui peut prendre la valeur $+\infty$ (resp. $-\infty$))

au sens où :

$$x_F = \sup\{x : F(x) < 1\} \leq \infty$$

respectivement

$$x_F^* = \inf\{x : F(x) > 0\}$$

Proposition B.2 (Embrechts & Mikosch)

La suite des maximums $\{M_n = \max(X_1, X_2, \dots, X_n), n \geq 1\}$ converge presque-sûrement vers x_F quand $n \rightarrow \infty$, i.e :

$$M_n \xrightarrow{p.s.} x_F \quad \text{quand} \quad n \rightarrow \infty \quad (\text{B.4})$$

Le résultat suivant découle automatiquement de (B.3).

Corollaire B.1

La suite des maximums $\{M_n, n \in \mathbb{N}^*\}$ converge en loi vers une variable aléatoire dégénérée concentrée en x_F , car :

$$\lim_{n \rightarrow \infty} F_{M_n}(x) = \lim_{n \rightarrow \infty} \mathbb{P}(M_n \leq x) = \lim_{n \rightarrow \infty} F^n(x) = \begin{cases} 0 & \text{si } x < x_F \\ 1 & \text{si } x \geq x_F \end{cases} \quad (\text{B.5})$$

Remarque B.4

Les formules exactes de la loi du terme maximum M_n donnée dans (B.3) et de la loi limite donnée dans (B.5) présentent toute fois un intérêt limité, et elles ne fournissent pas beaucoup d'informations sur les extrême. En pratique, la loi de la variable aléatoire parente X est rarement connue avec précision, et par conséquent, il en est de même pour la loi du terme maximum M_n .

De plus, même si la loi de X est connue avec exactitude, la loi de M_n n'est pas toujours facilement calculable. Par exemple, la fonction de distribution d'une variable aléatoire normale ne possède pas d'expression analytique, puisque c'est une intégrale incalculable. Donc, sa puissance n^{eme} est difficile à calculer et conduit à de sérieux problèmes numériques pour de grandes valeurs de n et pour de grandes valeurs de x . Pour ces raisons, il est intéressant de considérer le comportement asymptotique du terme maximum M_n normalisé.

B.1.3 Fonction quantile

Définition B.4 (Fonction Quantile)

Pour tout $0 < s < 1$, la fonction quantile est définie par :

$$Q(s) = \inf\{t : F(t) \geq s\} = F^{-1}(s)$$

Où F^{-1} représente la fonction inverse généralisée de F avec la convention que $\inf\{\emptyset\} = +\infty$ et $\mathbb{P}(X \leq Q(s)) = s$.

On l'exprime en termes de la fonction de survie par :

$$Q(s) = \inf\{t : \bar{F}(t) \leq 1 - s\}$$

Remarque B.5

Pour tout $0 < s < 1$ lorsque la fonction de répartition F est strictement croissante et continue alors :

$$Q(s) = F^{-1}(s) = \bar{F}^{-1}(1 - s)$$

Le quantile d'ordre s est défini par :

$$t_s = F^{-1}(s)$$

Définition B.5 (Fonction Quantile empirique)

La fonction de quantile empirique de l'échantillon (X_1, \dots, X_n) est définie par :

$$Q_n(s) = F_n^{-1}(s) = \inf\{t : \bar{F}_n(t) \geq s\}$$

où

$$Q_n(s) = \bar{F}_n^{-1}(1 - s) = \inf\{t : \bar{F}_n(t) \leq (1 - s)\}$$

Q_n peut être exprimée comme une fonction simple des statistiques d'ordre concernant l'échantillon (X_1, \dots, X_n) . Donc, on a :

$$Q_n(s) = X_{i:n} \quad \text{si} \quad \frac{i-1}{n} \leq s \leq \frac{i}{n}, \quad i = 1, \dots, n$$

On note que $X_{[ns]+1:n}$ est le quantile empirique d'ordre s où $[ns]$ désigne la partie entière de ns .

Remarque B.6

— Une fonction, notée par U et (parfois) appelée fonction de quantile de queue, elle est définie par :

$$U(t) = F^{-1}\left(1 - \frac{1}{t}\right) = \left(\frac{1}{\bar{F}}\right)^{-1}(t), \quad t \geq 1 \quad (\text{B.6})$$

— la fonction empirique correspondante est :

$$U_n(t) = Q_n\left(1 - \frac{1}{t}\right), \quad t \geq 1$$

B.1.4 Distributions exactes des statistiques d'ordre

Proposition B.3 (Distributions du maximum et du minimum)

Soient n variables aléatoires (X_1, \dots, X_n) indépendantes identiquement distribuées de fonction de répartition F , la distribution exacte du maximum $X_{n:n}$ est simplement donnée par la formule suivante :

$$\forall x \in \mathbb{R}, \quad F_{X_{n:n}}(x) = [F(x)]^n$$

La distribution exacte du minimum est donnée par :

$$\forall x \in \mathbb{R}, \quad F_{X_{1:n}}(x) = 1 - [1 - F(x)]^n$$

Preuve

$$\begin{aligned} F_{X_{n:n}}(x) &= \mathbb{P}(X_{n:n} \leq x) \\ &= \mathbb{P}(X_1 \leq x, \dots, X_n \leq x) \\ &= \prod_{i=1}^n \mathbb{P}(X_i \leq x) \\ &= [F(x)]^n \end{aligned}$$

et

$$\begin{aligned} F_{X_{1:n}}(x) &= \mathbb{P}(X_{1:n} \leq x) \\ &= 1 - \mathbb{P}(X_{1:n} > x) \\ &= 1 - \mathbb{P}(X_1 > x, \dots, X_n > x) \\ &= 1 - (1 - [F(x)])^n \end{aligned}$$

Ce sont des cas particuliers importants du résultat général de $F_{X_{k:n}}(x)$ qui est donné par :

$$\begin{aligned} \mathbb{P}(X_{k:n} \leq x) &= F_{X_{k:n}}(x) \\ &= \mathbb{P}(\text{au moins } k \text{ des } n \text{ } X \text{ sont } \leq x) \\ &= \sum_{j=k}^n \binom{n}{j} \mathbb{P}(X_1 \leq x)^j (1 - \mathbb{P}(X_1 \leq x))^{n-j} \\ &= \sum_{j=k}^n \binom{n}{j} F(x)^j (1 - F(x))^{n-j}. \end{aligned}$$

Annexe C

Propriétés statistiques des lois limites des valeurs extrêmes

TABLE C.1 – Propriétés statistiques des lois limites des valeurs extrêmes

Loi limite des valeurs extrêmes	Weibull	Gumbel	Fréchet
Répartition	$1 - \exp\left(-\left(\frac{x}{\lambda}\right)^k\right)$	$\exp\left(-e^{-\left(\frac{x-\mu}{\beta}\right)}\right)$	$\exp(-x^{-\alpha})$
Paramètres	$k > 0, \lambda > 0$	$\mu \in \mathbb{R}, \beta > 0$	$\alpha > 0$
Support	$x > 0$	$x \in \mathbb{R}$	$x > 0$
Densité	$\left(\frac{k}{\lambda}\right) \left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right)$	$\frac{1}{\beta} e^{-\left(\frac{x-\mu}{\beta}\right)} \exp\left(-e^{-\left(\frac{x-\mu}{\beta}\right)}\right)$	$\frac{\alpha}{x^{\alpha+1}} \exp(-x^{-\alpha})$
Espérance	$\mu = \lambda \Gamma\left(1 + \frac{1}{k}\right)$	$\mu + \beta \gamma, \gamma : \text{constante d'Euler}$	$\begin{cases} \Gamma\left(1 - \frac{1}{\alpha}\right) & \text{si } \alpha > 1 \\ \infty & \text{sinon} \end{cases}$
Médiane	$\lambda (\ln 2)^{\frac{1}{k}}$	$\mu - \beta \ln(\ln 2)$	$\frac{1}{(\ln 2)^{\frac{1}{\alpha}}}$
Mode	$\lambda \left(\frac{k-1}{k}\right)^{\frac{1}{k}}, k > 1$	μ	$\left(\frac{\alpha}{1+\alpha}\right)^{\frac{1}{\alpha}}$
Variance	$\sigma^2 = \lambda^2 \Gamma\left(1 + \frac{2}{k}\right) - \mu^2$	$\frac{\pi^2}{6} \beta^2$	$\begin{cases} \Gamma\left(1 - \frac{2}{\alpha}\right) - \left(\Gamma\left(1 - \frac{1}{\alpha}\right)\right)^2 & \text{si } \alpha > 2 \\ \infty & \text{sinon} \end{cases}$
Fonction génératrice des moments/moments	$\lambda^n \Gamma\left(1 + \frac{n}{k}\right)$	$\Gamma(1 - \beta t) \exp(\mu t)$	$\Gamma\left(1 + \frac{n}{\alpha}\right), \alpha > n$

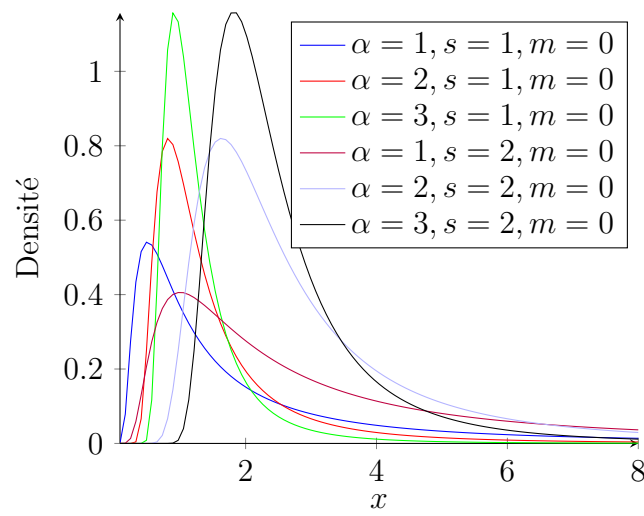
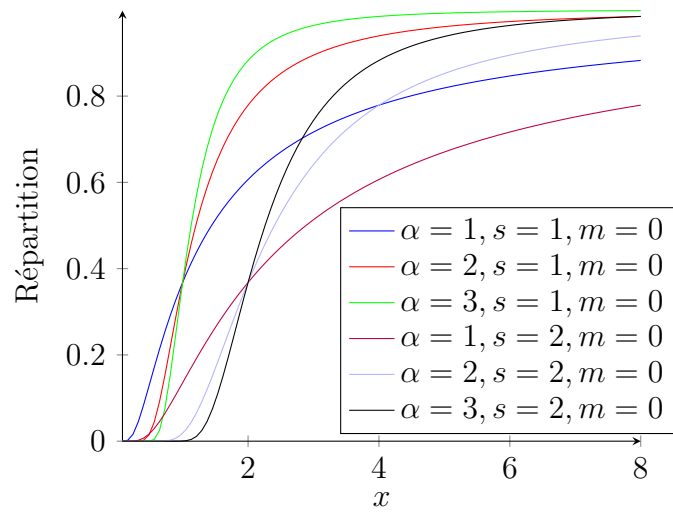
Selon le tableau C.1, la loi de Fréchet à la particularité d'avoir :

$$\mathbb{E}(X) = \infty \text{ si } 0 < \alpha \leq 1$$

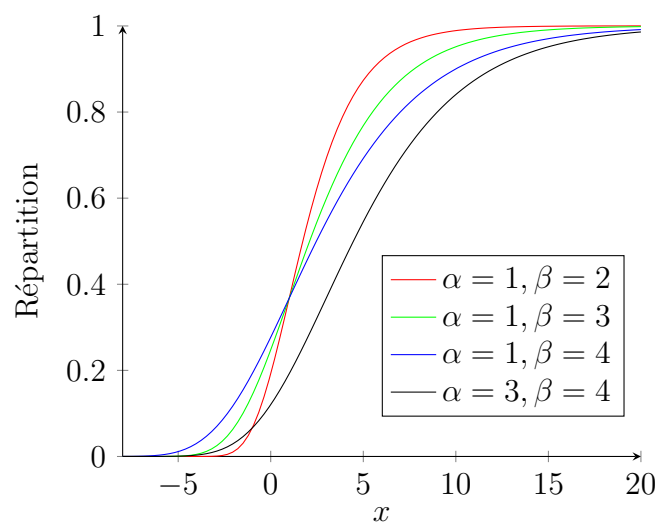
et

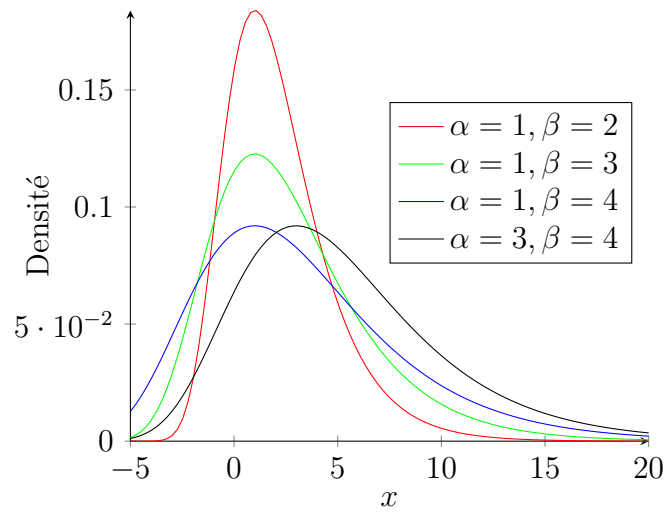
$$\text{Var}(X) = \infty \text{ si } 0 < \alpha \leq 2$$

Cette particularité pose problème à l'estimation classique de ses paramètres. Ce point est traité dans cette thèse.

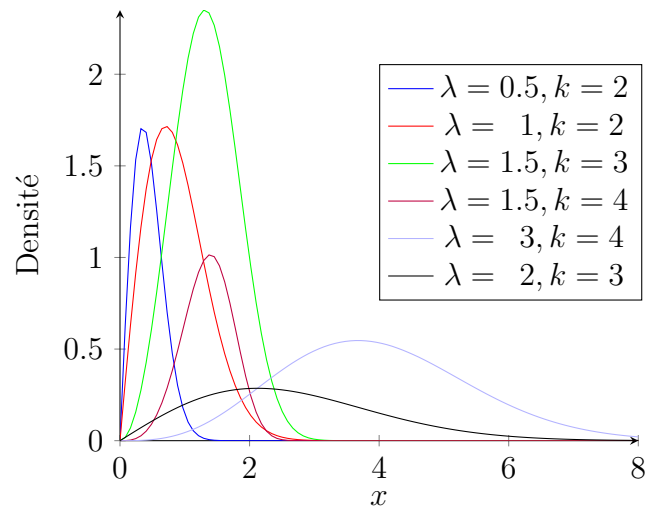


$$\frac{\alpha}{s} \left(\frac{x-m}{s}\right)^{-\alpha-1} \exp\left(-\left(\frac{x-m}{s}\right)^{-\alpha}\right)$$

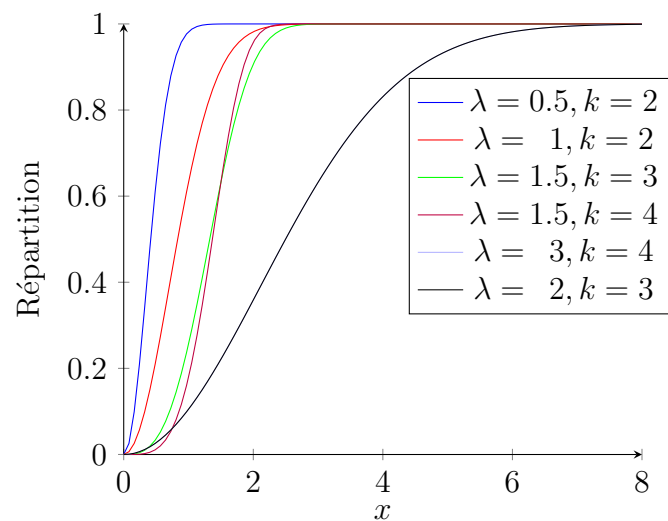




$$\exp(-\exp(-(\frac{x-\mu}{\beta})))$$



$$\frac{k}{\lambda} (\frac{x}{\lambda})^{k-1} \exp(-(\frac{x}{\lambda})^k)$$



Bibliographie

- [1] http://www.dataanalysisclassroom.com/wp-content/uploads/2018/03/cp_temperature.csv, 03 2018.
- [2] C. Acerbi. Coherent representations of subjective risk-aversion. *Risk measures for the 21st century*, pages 147–207, 2004.
- [3] C. Acerbi and D. Tasche. Expected shortfall : a natural coherent alternative to value at risk. *Economic notes*, 31(2) :379–388, 2002.
- [4] C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7) :1487–1503, 2002.
- [5] F. I. Alves, L. D. Haan, and T. Lin. Third order extended regular variation. *Publications de l'Institut Mathématique*, 80(94) :109–120, 2006.
- [6] M. F. Alves, M. I. Gomes, and L. de Haan. A new class of semi-parametric estimators of the second order parameter. *Portugaliae Mathematica*, 60(2) :193–214, 2003.
- [7] M. F. Alves, M. I. Gomes, L. de Haan, and C. Neves. A note on second order conditions in extreme value theory : linking general and heavy tail conditions. *REVSTAT Statistical Journal*, 5(3) :285–304, 2007.
- [8] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical finance*, 9(3) :203–228, 1999.
- [9] A. A. Balkema and L. De Haan. Residual life time at great age. *The Annals of probability*, pages 792–804, 1974.
- [10] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes : Theory and Applications*, volume 558. John Wiley & Sons, 2004.
- [11] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of extremes : theory and applications*. John Wiley & Sons, 2006.

- [12] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of extremes : theory and applications*. John Wiley & Sons, 2006.
- [13] J. Beirlant and G. Matthys. Extreme quantile estimation for heavy-tailed distributions. *Rapport technique, Department of Mathematics, KU Leuven*, page 23, 2001.
- [14] J. Beirlant, W. Schoutens, and J. Segers. Mandelbrot’s extremism. 2004.
- [15] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.
- [16] B. Brahim, D. Meraghni, A. Necir, and D. Yahia. A bias-reduced estimator for the mean of a heavy-tailed distribution with an infinite second moment. *Journal of Statistical Planning and Inference*, 143(6) :1064–1081, 2013.
- [17] V. Brazauskas, B. L. Jones, M. L. Puri, and R. Zitikis. Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, 138(11) :3590–3604, 2008.
- [18] L. Breiman, C. J. Stone, and C. Kooperberg. Robust confidence bounds for extreme upper quantiles. *Journal of Statistical Computation and Simulation*, 37(3-4) :127–149, 1990.
- [19] F. Caeiro and M. I. Gomes. A new class of estimators of a “scale” second order parameter. *Extremes*, 9(3-4) :193–211, 2006.
- [20] F. Caeiro and M. I. Gomes. Bias reduction in the estimation of a shape second-order parameter of a heavy-tailed model. *Journal of Statistical Computation and Simulation*, 85(17) :3405–3419, 2015.
- [21] J. Cai and K. S. Tan. Optimal retention for a stop-loss reinsurance under the var and cte risk measures. *ASTIN Bulletin : The Journal of the IAA*, 37(1) :93–112, 2007.
- [22] J.-J. Cai, J. H. Einmahl, L. Haan, and C. Zhou. Estimation of the marginal expected shortfall : the mean when a related variable is extreme. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 77(2) :417–442, 2015.
- [23] G. Ciuperca and C. Mercadier. Semi-parametric estimation for heavy tailed distributions. *Extremes*, 13(1) :55–87, 2010.
- [24] S. Coles, J. Bawa, L. Trenner, and P. Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

- [25] A. C. Davison and R. L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society : Series B (Methodological)*, 52(3) :393–425, 1990.
- [26] L. de Haan. De (1970) on regular variation and its application to the weak convergence of sample extremes. *Mathematical Centre Tracts*, 32, 1970.
- [27] L. de Haan and A. Ferreira. *Extreme value theory : An introduction*, 2006.
- [28] L. de Haan and U. Stadtmüller. Generalized regular variation of second order. *Journal of the Australian Mathematical Society (Series A)*, 61(03) :381–395, 1996.
- [29] L. d. De Haan and L. Peng. Comparison of tail index estimators. *Statistica Neerlandica*, 52(1) :60–70, 1998.
- [30] T. de Wet, Y. Goegebeur, and M. R. Munch. Asymptotically unbiased estimation of the second order tail parameter. *Statistics & Probability Letters*, 82(3) :565–573, 2012.
- [31] A. L. Dekkers, L. De Haan, et al. On the estimation of the extreme-value index and large quantile estimation. *The annals of statistics*, 17(4) :1795–1832, 1989.
- [32] J. Dhaene, L. Henrard, Z. Landsman, A. Vandendorpe, and S. Vanduffel. Some results on the cte-based capital allocation rule. *Insurance : Mathematics and Economics*, 42(2) :855–863, 2008.
- [33] J. El Methni. *Contributions à l'estimation de quantiles extrêmes. Applications à des données environnementales*. PhD thesis, 2013.
- [34] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events : for insurance and finance*, volume 33. Springer Science & Business Media, 2013.
- [35] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge University Press, 1928.
- [36] M. Fréchet. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Math. Polon.*, 6 :93–116, 1927.
- [37] J. Geluk, L. De Haan, S. Resnick, and C. Stărică. Second-order regular variation, convolution and the central limit theorem. *Stochastic processes and their applications*, 69(2) :139–159, 1997.

- [38] B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.
- [39] B. Gnedenko, A. Kolmogorov, B. Gnedenko, and A. Kolmogorov. Limit distributions for sums of independent. *Am. J. Math*, 105, 1954.
- [40] Y. Goegebeur, J. Beirlant, and T. de Wet. Kernel estimators for the second order parameter in extreme value statistics. *Journal of statistical Planning and Inference*, 140(9) :2632–2652, 2010.
- [41] C. M. Goldie and C. Klüppelberg. Subexponential distributions. *A practical guide to heavy tails : statistical techniques and applications*, pages 435–459, 1998.
- [42] M. I. Gomes, L. De Haan, and L. Peng. Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, 5(4) :387–414, 2002.
- [43] P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 37–42, 1982.
- [44] P. Hall, A. Welsh, et al. Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13(1) :331–341, 1985.
- [45] B. M. Hill et al. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5) :1163–1174, 1975.
- [46] J. R. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized pareto distribution. *Technometrics*, 29(3) :339–349, 1987.
- [47] P. Jorion. *Value at risk : the new benchmark for controlling market risk*. Irwin Professional Pub., 1997.
- [48] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern actuarial risk theory : using R*, volume 128. Springer Science & Business Media, 2008.
- [49] M. Kratz and S. I. Resnick. The qq-estimator and heavy tails. *Stochastic Models*, 12(4) :699–724, 1996.
- [50] S. Kusuoka. On law invariant coherent risk measures. In *Advances in mathematical economics*, pages 83–95. Springer, 2001.
- [51] M. Laidi, A. Rassoul, and H. O. Rouis. Improved estimator of the conditional tail expectation in the case of heavy-tailed losses. *Statistics, Optimization & Information Computing*, 8(1) :98–109, 2020.

- [52] Z. M. Landsman and E. A. Valdez. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4) :55–71, 2003.
- [53] P. Lévy. Calcul des probabilités (gauthier-villars), 1925.
- [54] D. Li, L. Peng, and J. Yang. Bias reduction for high quantiles. *Journal of Statistical Planning and Inference*, 140(9) :2433–2441, 2010.
- [55] D. M. Mason. Laws of large numbers for sums of extreme values. *The Annals of Probability*, pages 754–764, 1982.
- [56] J. E. Methni, L. Gardes, and S. Girard. Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics*, 41(4) :988–1012, 2014.
- [57] J. Morgan. Creditmetrics-technical document. *JP Morgan, New York*, 1997.
- [58] A. Necir, A. Rassoul, and R. Zitikis. Estimating the conditional tail expectation in the case of heavy-tailed losses. *Journal of Probability and Statistics*, 2010.
- [59] C. Neves. From extended regular variation to regular variation with application in extreme value statistics. *Journal of mathematical analysis and applications*, 355(1) :216–230, 2009.
- [60] L. Peng. Asymptotically unbiased estimators for the extreme-value index. *Statistics & Probability Letters*, 38(2) :107–115, 1998.
- [61] L. Peng and Y. Qi. Estimating the first-and second-order parameters of a heavy-tailed distribution. *Australian & New Zealand Journal of Statistics*, 46(2) :305–312, 2004.
- [62] J. Pickands III et al. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1) :119–131, 1975.
- [63] S. Resnick. Extreme values, regular variation, and point processes, appl. *Probab. Ser. Appl. Probab. Trust*, 4 :33, 1987.
- [64] S. I. Resnick. *Heavy-tail phenomena : probabilistic and statistical modeling*. Springer Science & Business Media, 2007.
- [65] R. T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7) :1443–1471, 2002.

- [66] R. T. Rockafellar, S. Uryasev, et al. Optimization of conditional value-at-risk. *Journal of risk*, 2 :21–42, 2000.
- [67] P. A. Samuelson. The “fallacy” of maximizing the geometric mean in long sequences of investing or gambling. *Proceedings of the National Academy of sciences*, 68(10) :2493–2496, 1971.
- [68] J. Schultze and J. Steinebach. On least squares estimates of an exponential tail coefficient. *Statistics & Risk Modeling*, 14(4) :353–372, 1996.
- [69] R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1) :67–90, 1985.
- [70] R. L. Smith. Estimating tails of probability distributions. *The annals of Statistics*, pages 1174–1207, 1987.
- [71] K. S. Tan, C. Weng, and Y. Zhang. Optimality of general reinsurance contracts under cte risk measure. *Insurance : Mathematics and Economics*, 49(2) :175–187, 2011.
- [72] D. Tasche. Expected shortfall and beyond. *Journal of Banking & Finance*, 26(7) :1519–1533, 2002.
- [73] R. Von Mises. La distribution de la plus grande de n valeurs. *Rev. math. Union interbalcanique*, 1 :141–160, 1936.
- [74] X. Q. Wang and S. H. Cheng. General regular variation of n -th order and the 2nd order edgeworth expansion of the extreme value distribution (i). *Acta Mathematica Sinica*, 21(5) :1121–1130, 2005.
- [75] I. Weissman. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364) :812–815, 1978.
- [76] T. Werner and C. Upper. Time variation in the tail behaviour of bunds futures returns. 2002.
- [77] J. L. Wirch and M. R. Hardy. A synthesis of risk measures for capital adequacy. *Insurance : mathematics and Economics*, 25(3) :337–347, 1999.
- [78] J. Worms and R. Worms. Estimation of second order parameters using probability weighted moments. *ESAIM : Probability and Statistics*, 16 :97–113, 2012.
- [79] Y. Yamai, T. Yoshida, et al. Comparative analyses of expected shortfall and value-at-risk : their estimation error, decomposition, and optimization. *Monetary and economic studies*, 20(1) :87–121, 2002.

Index alphabétique

C

Comportement
asymptotique 55, 70, 71

D

domaines
d'attraction 18

E

estimateur
de Hill 66, 67, 70
de Pickand 51
de Pickands 64
de Weissman 65
de Zipf 69
des moments 61
du maximum de vraisemblance 61
empirique 57, 61
paramétriques 86
du maximum de vraisemblance 87

I

indice
de queue 65, 66
de queue 23, 51, 54, 65

L

loi
de pareto 87

M

méthode
des moments 86

mesure
de Wang 83
invariante 78
par distorsion 83
de risque 77
de risque monétaire 77
par distorsion 84

mesures
additives 78
de risques 74
de risques usuelles 76
moindres carrés 86

Q

quantile 79
de queue 49, 113
empirique 86
queue
de la loi 111
de distribution 23
légère 24
lourde 24, 34, 65

R

risques
comonotones 78

V

valeurs
extrêmes 47, 65
extrêmes 18, 39, 59, 61, 66, 70

ملخص

يقدم هذا العمل دراسة لمقاييس المخاطر في حالة التوزيعات الثقيلة الذيل للقيم المتطرفة. يقترح تحسيناً لمقدر كلاسيكي لأحد مقاييس المخاطر المسماة الأمل الشرطي للذيل (CTE) وذلك عن طريق تقليل التحيز. تم من خلال هذه الأطروحة إجراء دراسة للسلوك المقارب، ولا سيما التقارب الطبيعي لهذا المقدّر الجديد، بالإضافة إلى المقارنة مع المقدّر الكلاسيكي، مما يثبت تحسناً واضحاً في جودة التقدير. كلمات مفتاحية: مقياس خطر، ذيل التوزيع، الأمل الشرطي، القيمة المتطرفة، المقدر، السلوك المقارب.

Résumé

Ce travail présente une étude des mesures du risque dans le cas des distributions à queues lourdes des valeurs extrêmes.

Il propose une amélioration à un estimateur classique de l'une des mesures de risque appelée Espérance Conditionnelle de queue (CTE) en réduisant le biais.

Dans cette thèse, une étude du comportement asymptotique, en particulier de la convergence normale de ce nouvel estimateur, a été menée, en plus d'une comparaison avec l'estimateur classique, ce qui prouve une nette amélioration de la qualité de l'estimation.

Mots clés : mesure du risque, queue de distribution, espérance conditionnelle, valeur extrême, estimateur, comportement asymptotique.

Abstract

This work presents a study of risk measures in the case of heavy tailed extremes values distributions. It proposes an improvement to a classical estimator of one of the risk measures called Conditional Tail Expectation (CTE) by reducing the bias. In this thesis, a study of the asymptotic behavior, in particular of the normal convergence of this new estimator, was carried out, in addition to a comparison with the classical estimator, which proves a clear improvement of the quality of the estimate.

Keswords : risk measure, distribution's tail, conditional expectation, extreme value, estimator, Asymptotic behavior.

