

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
UNIVERSITÉ KASDI MERBAH, OUARGLA
FACULTÉ DE MATHÉMATIQUES ET SCIENCES DE LA MATIÈRE
DÉPARTEMENT DE MATHÉMATIQUES



Mémoire Présenté En Vue De L'obtention Du

DIPLÔME DE MASTER

EN MATHÉMATIQUES

Option : Probabilité et Statistique

Par

Chaima NEGAIS

Titre

**ESTIMATION NON PARAMETRIQUE DE L'INDICE DE QUEUE DES
VALEURS EXTRÊMES**

Membres du Comité d'Examen

Abdelkader AMARA	Maître assistant classe "A"	UKMO	Président
Hadda SAIDANE	Maître assistant classe "A"	UKMO	Examineur
Hanane ARBIA	Maître assistant classe "A"	UKMO	Rapporteur

mai 2017

Table des matières

Remerciements	i
Table des figures	ii
Notations et abriviations	ii
Introduction	1
1 La Théorie des Valeurs extrêmes	2
1.1 Statistiques d'ordre	3
1.1.1 Distribution et fonction de quantile	6
1.1.2 Distribution et fonction de quantile empiriques	7
1.2 La loi du maximum	8
1.3 Lois limites des valeurs extrêmes	10
1.4 Loi Généralisée des Valeurs Extrêmes (GEV)	12
1.4.1 Représentation de <i>Jenkinson-Von Mises</i>	12
1.4.2 Loi max-stable	15
1.5 Domaine d'attraction	15
1.5.1 Fonction de même type	18
1.5.2 Caractérisations des domaines d'attraction	19
1.5.3 Condition nécessaire et suffisante de convergence	23
1.5.4 Condition de second ordre :	24
1.5.5 Constantes de normalisation	25
1.6 Loi Généralisée de Pareto (GPD)	25
1.6.1 Distribution des excès	26
1.6.2 Distribution de Pareto Généralisé (GPD)	27
1.6.3 Le choix du seuil	28
1.6.4 L'espérance au-delà d'un seuil	28
1.7 Théorème de Balkema-de Haan-Pickands	28

2 Estimation non paramétrique de l'indice de queue des valeurs extrêmes	29
2.1 L'estimateur de Pickands	29
2.2 L'estimateur de Hill	31
2.3 L'estimateur des moments de Dekkers - Einmahl-De Haan	32
2.4 Comparaison entre les trois estimateurs	34
2.5 Choix du seuil k	34
2.5.1 Estimation de la queue de la distribution	36
2.5.2 Détermination du seuil u	36
2.6 Construction de l'estimateur de Hill	37
2.6.1 Normalité asymptotique d'estimateur de Hill(1975)	41
2.7 Simulation	43
Conclusion générale	46
Bibliographie	47

Remerciements

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

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

J'adresse mes sincères remerciements au Pr SAIDANE HADDA pour l'intérêt qu'il a manifesté a ce mémoire en acceptant de l'examiner. J'exprime ma profonde gratitude au Pr AMARA ABDELKADER qui m'a fait l'honneur de présider le jury de ce mémoire. Je lui suis très reconnaissante pour l'attention qu'il a porté à ce travail. Merci sincèrement du temps et de l'énergie que vous avez consacré à la lecture de mon travail.

Mes vifs remerciements vont également à tous mes enseignants en graduation, et en post graduation.

J'adresse un amical remerciement le plus sincère à tous mes collègues de master pour leur sympathie et leur soutien.

Enfin j'exprime ma gratitude à ma petite famille : Sofiane, Allale, Nasreddine, Sabrina, Ritadj, mes Parents, Ammar et Farida qui m'a toujours soutenue et encouragée dans la voie que je m'étais fixée. Je remercie particulièrement mes parents qui m'ont stimulée et encouragé pendant mes études.

Merci à tous ceux que j'ai oubliés...Voilà, avec un tel entourage, on ne pouvait que faire au mieux! .

Table des figures

1.1	Maxima annuels du niveau de la mer au Port Pirie, Australie du sud	3
1.2	Fonctions de distribution du Gumbel (en noire), de Fréchet (en bleu) et de Weibull (en rouge).	14
1.3	Queue légère	16
1.4	Queue lourde	17
1.5	Queue finie	18
1.6	Représentation des excès Y issus des dépassements X au-delà d'un seuil u .	26
2.1	Estimateur de Pickands avec un intervalle de confiance 95% de γ basés sur 100 échantillons de taille 3000 pour la loi uniforme.	30
2.2	L'estimateur de Hill avec un intervalle de confiance 95% de γ basés sur 100 échantillons de taille 3000 pour la loi de Pareto standard.	32
2.3	Estimateur des moments avec un intervalle de confiance 95% de γ basés sur 100 échantillons de taille 3000 pour la loi de Gumbel standard.	34

Notations et abriviations

x_F	Le point terminal de F.
$\mathcal{D}()$	Domaine d'attraction.
σ	paramètre d'échelle
F	Fonction de distribution.
F^{-1}	Fonction quantile.
F^*	Distribution de probabilité distordue.
F_u	Distribution conditionnelle des excès.
F_i	Fonction de distribution de la $i^{\text{ème}}$ marginale.
γ	paramètre de forme
G	Fonction de répartition de la v.a Z .
GEV	Distribution généralisée de valeurs extrêmes.
GPD	Distribution généralisée de Pareto.
H_γ	Distribution des valeurs extrêmes standard.
$v.a$	variable aléatoire
POT	Peaks Over Threshold.
$\xrightarrow{p.s.}$	Converge p.s
\xrightarrow{p}	Converge en propabilité
TVE	Théorie des valeurs extrêmes.
\mathbb{R}	ensemble de nombre réels
(X_1, \dots, X_n)	Echantillon de taille n de X .
$(X_{1,n}, \dots, X_{n,n})$	Echantillon ordonné.
\mathbb{R}_+	ensemble de nombre réels positifs

$X_{i,n}$	La $i^{\text{ième}}$ statistique d'ordre.
ϕ	Distribution normale standard.
$\Lambda(x)$	Distribution des valeurs extrêmes de Gumbel.
$\Phi_\alpha(x)$	Distribution des valeurs extrêmes de Fréchet.
$\Psi_\alpha(x)$	Distribution des valeurs extrêmes de Weibull.

Introduction

La modélisation des événements extrêmes (ouragan, tremblement de terre ou inondation, crues, crises financières, krachs, chocs pétroliers) est aujourd'hui un champ de recherches particulièrement actif, notamment par l'importance de leurs impacts économiques et sociaux. En particulier, depuis quelques années, on note un intérêt croissant pour l'application de la Théorie des Valeurs Extrêmes (TVE) pour la modélisation de tels événements. La théorie des valeurs extrêmes permet d'évaluer les événements rares et les pertes associées à leur apparition. En d'autres termes lorsqu'une perte importante survient, cette théorie permet d'en évaluer l'ampleur. Les résultats théoriques sur le comportement stochastique des extrêmes d'échantillon qu'offre la théorie des valeurs extrêmes permettent de proposer un cadre mathématique rigoureux pour réaliser de telles extrapolations. De plus cette théorie acquiert une importance particulière du fait qu'elle s'intéresse directement à la queue de la loi.

Notre mémoire est organisé comme suit :

Chapitre 1 : Nous proposons dans ce chapitre la théorie des valeurs extrêmes. Cette théorie est basée sur la statistique d'ordre d'un échantillon de taille n , le comportement asymptotique du maximum. Pour cela on a introduit deux théorèmes essentielles, le premier concerne les lois des valeurs extrêmes (*GEV : Generalized Extreme Value*) et qui s'intéresse à la loi du maximum. Cette première approche est basé sur le théorème de *Fisher et Tippett* qui sert à modéliser les maximums prouvés en 1928, la preuve complète a été proposée par *Gnedenko (1943, [31])*, et le second concerne les lois des excès (*POT : Peaks Over threshold*). Il consiste à utiliser les observations qui dépassent un certain seuil déterministe. Les différences entre ces observations et le seuil sont appelées excès. Cette deuxième approche est fondée sur la convergence vers une loi de *Pareto Généralisée (GPD)* assurée par la théorème de *Belkema, De Han et Pickands* qui sert à modéliser les excès déterminé en 1975. Nous donnons ensuite les caractérisations des domaines d'attraction du maximum, la notion des fonctions à variations régulières.

Chapitre 2 : Dans ce chapitre, on s'intéresse à l'estimation de l'indice des valeurs extrêmes γ par des méthodes non paramétriques. On a d'abord définit l'estimateur de *Pickands* et sa propriétés. Ensuite l'estimateur des moments et l'estimateur de *Hill (1975, [34])* et ses propriétés et on fait comparaison entre les trois estimateurs. On définit aussi le seuil qui doit être choisi à partir de la méthode de *Cheng et Peng [7]* qui est proposé une valeur optimale qui est donne un meilleur modèle. On a présenté la construction de l'estimateur de *Hill(1975, [34])* et on étudie sa normalité asymptotique.

Chapitre 1

La Théorie des Valeurs extrêmes

La théorie des valeurs extrêmes permet d'évaluer les événements rares et les pertes associées à leur apparition. En d'autres termes lorsqu'une perte importante survient, cette théorie permet d'en évaluer l'ampleur. Les résultats théoriques sur le comportement stochastique des extrêmes d'échantillon qu'offre la théorie des valeurs extrêmes permettent de proposer un cadre mathématique rigoureux pour réaliser de telles extrapolations. De plus cette théorie acquiert une importance particulière du fait qu'elle s'intéresse directement à la queue de la loi. Par définition, les événements rares sont des événements ayant une faible probabilité d'apparition. Lorsque le comportement de ces événements est dû au hasard on peut étudier leur loi. Ils sont dits extrêmes quand il s'agit des valeurs beaucoup plus grandes ou plus petites que celles observées habituellement. On illustre l'utilisation de la TVE par un recueil d'exemples d'application, tiré du livre de *Cole (2001,[8])*. L'exemples ci-dessous, accompagné par des questions qui sont habituellement soulevées par les praticiens, réfèrent à plusieurs domaines d'applications.

Exemple 1 : La figure ci-dessous présente les maxima annuels du niveau de la mer au Port Pirieau nord d'Adélaïde, en Australie du Sud au cours de la période 1923-1987. A partir de ces données, nous pouvons estimer le niveau maximal susceptible de se reproduire dans la région étudiée au cours des 100 ou 1000 ans à venir. Cependant, comment pouvons nous estimer les niveaux qui peuvent se reproduire dans les 1000 prochaines années sans vérifier la validité de la stabilité des changements climatiques durant 1000 ans ?

Certes, les données présentées ne donnent aucun signe sur une non-stationnarité au cours de la période observée, mais cette stabilité peut ne pas persister dans l'avenir, surtout lorsqu'il s'agit d'une période très longue (1000 ans).

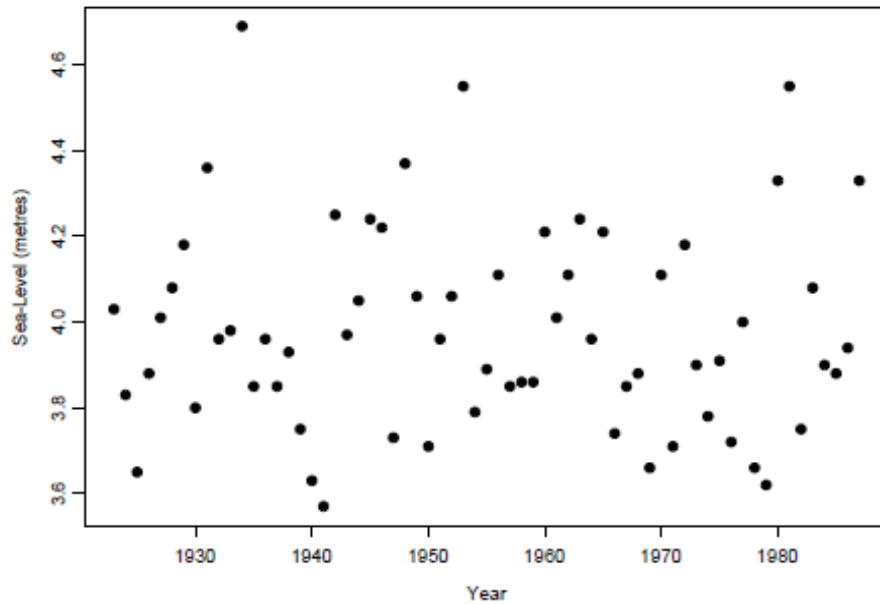


FIG. 1.1 – Maxima annuels du niveau de la mer au Port Pirie, Australie du sud

1.1 Statistiques d'ordre

Dans la théorie des valeurs extrêmes, la statistique d'ordre occupe une place très importante parce qu'elle donne des informations sur la queue de la distribution.

Dans la suite nous considérons (X_1, X_2, \dots, X_n) un échantillon de n variables aléatoires indépendantes et identiquement distribuées *i.i.d*

Définition 1.1.1 (Statistique d'ordre)

Les statistiques d'ordre d'un échantillon (X_1, \dots, X_n) sont le réangement croissant de (X_1, \dots, X_n) , elles sont dénotées par $X_{1,n}, \dots, X_{n,n}$ telles que

$$X_{1,n} \leq \dots \leq X_{n,n}$$

$X_{i,n}$ est la $i^{\text{ième}}$ statistique d'ordre.

La première statistique d'ordre, noté $X_{1,n}$ est le minimum de l'échantillon (X_1, \dots, X_n)

$$X_{1,n} = \min(X_1, \dots, X_n).$$

La dernière statistique d'ordre $X_{n,n}$ est le maximum de l'échantillon (X_1, \dots, X_n)

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

Dans la suite de notre exposé, nous présentons les résultats ne concernant que le maximum, puisque les résultats relatifs au minimum se déduisent directement des résultats du maximum en considérant la série opposée $-X_1, -X_2, \dots, -X_n$, d'après l'égalité suivante :

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

Proposition 1.1.1 (*La distribution et la densité du maximum et du minimum*)
Soit X_1, X_2, \dots, X_n *n* v.a i.i.d, de fonction de répartition commune F . La distribution de la plus petite statistique d'ordre est

$$F_{X_{1,n}}(x) = 1 - [1 - F(x)]^n, \quad x \in \mathbb{R},$$

et sa densité est donnée par

$$f_{X_{1,n}}(x) = n [1 - F(x)]^{n-1} f(x).$$

Pour la plus grande statistique d'ordre, on a

$$F_{X_{n,n}}(x) = [F(x)]^n, \quad x \in \mathbb{R},$$

et

$$f_{X_{n,n}}(x) = n [F(x)]^{n-1} f(x).$$

Grâce à la propriété d'indépendance des variables aléatoires X_1, \dots, X_n et qu'ils ont la même distribution F , nous obtenons les expressions de $F_{X_{1,n}}$ et $F_{X_{n,n}}$

$$\begin{aligned} F_{X_{1,n}}(x) &= P(X_{1,n} \leq x) \\ &= 1 - P(X_{1,n} > x) \\ &= 1 - P(X_1 > x, \dots, X_n > x) \\ &= 1 - P\left(\bigcap_{i=1}^n \{X_i > x\}\right) \\ &= 1 - \prod_{i=1}^n P(X_i > x) \\ &= 1 - [\bar{F}(x)]^n. \end{aligned}$$

et

$$\begin{aligned}
 F_{X_{n,n}}(x) &= P\{X_{n,n} \leq x\} \\
 &= P\{X_1 \leq x, \dots, X_n \leq x\} \\
 &= P\left(\bigcap_{i=1}^n \{X_i \leq x\}\right) \\
 &= \prod_{i=1}^n P(X_i \leq x) \\
 &= [F(x)]^n.
 \end{aligned}$$

Lemme 1.1.1 (La distribution et la densité de la $k^{\text{ième}}$ statistique d'ordre)
 La loi de la variable aléatoire $X_{k,n}$, pour $1 \leq k \leq n$ est donnée par :

$$\begin{aligned}
 F_{X_{k,n}}(x) &= P[X_{k,n} \leq x] \\
 &= \sum_{j=k}^n C_n^j [F(x)]^j [1 - F(x)]^{n-j}.
 \end{aligned}$$

Sa densité est

$$f_{X_{k,n}}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x).$$

Convergence presque sur des statistique d'ordre

Proposition 1.1.2 Soit $p = 1$ (respectivement $p = 0$). Soit $(k(n); n \geq 1)$ suite d'entiers telle que naturels non décroissante c'est à dire : $1 \leq k(n) \leq n$ telle que

$$\lim_{n \rightarrow \infty} \frac{k(n)}{n} = p.$$

Alors, la suite des quantiles empirique données par la statistique d'ordre $X_{(k(n),n)}$, $n \geq 1$ converge presque sûrement vers

$$x_F = \inf \{x : F(x) = 1\},$$

respectivement

$$\bar{x}_F = \sup \{x : F(x) = 0\}.$$

1.1.1 Distribution et fonction de quantile

Définition 1.1.2 (*Fonction de distribution et de survie*)

On appelle fonctions de distribution et de survie d'une variable aléatoire X les fonctions définies respectivement sur \mathbb{R} par

$$F(x) := p(X \leq x) \quad \text{et} \quad \bar{F}(x) := 1 - F(x).$$

Définition 1.1.3 (*Fonction de quantile*)

On appelle fonction de quantile ou l'inverse généralisé de F la fonction continue à gauche $Q :]0, 1[\rightarrow \mathbb{R}$ définie par

$$Q(s) = F^{\leftarrow}(s) := \inf \{x \in \mathbb{R}, F(x) \geq s\}, \quad 0 < s < 1$$

où par convention $\inf \emptyset = +\infty$ (\emptyset est l'ensemble vide)

Définition 1.1.4 (*Fonction de quantile de queue*)

Dans la TVE, nous utilisons souvent la fonction quantile de queue U définie par :

$$\begin{aligned} U(t) &= Q(1 - 1/t), \\ &= F^{\leftarrow}(1 - 1/t), \quad 1 < t < \infty \end{aligned}$$

Proposition 1.1.3 (*Propriétés de la fonction de distribution*)

La fonction de distribution F est croissante, continue à droite et vérifie

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{et} \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

Définition 1.1.5 (*point terminal*)

Le point terminal de F défini par

$$x_F = \sup \{x : F(x) < 1\} \leq \infty.$$

1.1.2 Distribution et fonction de quantile empiriques

Définition 1.1.6 (*Distribution empirique*)

On appelle distribution empirique associée à un échantion (X_1, X_2, \dots, X_n) , la fonction F_n définie par :

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}}, \quad x \in \mathbb{R} \\ &= \frac{1}{n} \text{card} \{i \in \{1, \dots, n\} : X_i \leq x\} \end{aligned}$$

où $1_{\{X_i \leq x\}}$: c'est la fonction indicatrice, elle est définie par :

$$1_{\{X_i \leq x\}} = \begin{cases} 1 & \text{si } X_i \leq x \\ 0 & \text{si } X_i > x \end{cases}$$

La relation entre les statistiques d'ordre et la distribution empirique est

$$F_n(x) = \begin{cases} 0 & \text{si } x < X_{1,n}, \\ 1/n & \text{si } X_{1,n} < x < X_{2,n}, \\ 2/n & \text{si } X_{2,n} < x < X_{3,n}, \\ \cdot & \cdot \\ \cdot & \cdot \\ i/n & \text{si } X_{i,n} < x < X_{i+1,n}, \text{ pour } i = 1, \dots, n-1, \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & \text{si } x \geq X_{n,n}. \end{cases}$$

On remarque que $F_n(x_{i,n}) = i/n$, pour $i = 1, \dots, n$.

La statistique F_n est l'estimateur naturel de F . En appliquant les lois des grands nombres sur F_n on obtient les résultats suivants :

Corollaire 1.1.1 *Pour tout $x \in \mathbb{R}$, on a*

1. *consistance faible* : $F_n(x) \xrightarrow{p} F(x)$ quand $n \rightarrow \infty$
2. *consistance forte* : $F_n(x) \xrightarrow{p.s} F(x)$ quand $n \rightarrow \infty$

D'après *Glivenco-Cantelli (1933)*, nous avons que $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$ converge presque sûrement vers 0 quand $n \rightarrow \infty$. Alors on dit que $F_n(x)$ est un estimateur de la fonction de répartition F :

Théorème 1.1.1 (*Glivenco-Cantelli (1933)*)

Soit $(X_n, n \in \mathbb{N})$ une suite de variables aléatoires réelles indépendantes et de même loi de F

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{p.s} 0 \text{ quand } n \rightarrow \infty$$

Définition 1.1.7 (*Fonction de quantile empirique*)

La fonction de quantile empirique de (X_1, \dots, X_n) est La fonction de quantile de sa distribution empirique qui est définie par :

$$Q_n(s) = F_n^{\leftarrow}(s) := \inf \{x \in \mathbb{R}, F_n(x) \geq s\}, \quad 0 < s < 1$$

Définition 1.1.8 (*Fonction de quantile empirique de queue*)

La fonction de quantile empirique de queue est définie par :

$$\begin{aligned} U_n(t) &= Q_n(1 - 1/t), \\ &= F_n^{\leftarrow}(1 - 1/t), \quad 1 < t < \infty \end{aligned}$$

1.2 La loi du maximum

Nous considérons n variables aléatoires X_1, \dots, X_n *i.i.d* de fonction de répartition F . Une manière simple d'étudier le comportement des événements extrêmes est de considérer la variable aléatoire :

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

La loi exacte de $X_{n,n}$ est

$$F_{X_{n,n}}(x) = [F(x)]^n. \quad (1.1)$$

La formule (1.1) présente un intérêt très limité. De plus, la loi d'une variable aléatoire parente X est rarement connue avec précision et, même si la loi de cette variable parente X est connue avec exactitude, la loi du terme maximum n'est

pas toujours facilement calculable. i.e : La difficulté provient du fait que l'on ne connaît pas en général la fonction de répartition F , en plus la loi du maximum quand n tend vers à l'infini est une loi dégénérée.

$$\lim_{n \rightarrow \infty} F_{X_{n,n}}(x) = \begin{cases} 0, & \text{si } x < x_F, \\ 1, & \text{si } x \geq x_F. \end{cases}$$

Pour cette raison, on essaie de trouver des constantes a_n et b_n de telle sorte que la loi du maximum normalisée, i.e., la loi de $\frac{X_{n,n} - b_n}{a_n}$ ne le soit plus.

Notons

$$Y_n = \frac{X_{n,n} - b_n}{a_n} \quad \text{avec } a_n > 0, b_n \in \mathbb{R}$$

$$\begin{aligned} F_{Y_n}(x) &= P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right), \\ &= P(X_{n,n} \leq a_n x + b_n), \\ &= (F(a_n x + b_n))^n. \end{aligned}$$

On étudie

$$\lim_{n \rightarrow +\infty} F_{Y_n}(x) = \lim_{n \rightarrow +\infty} (F(a_n x + b_n))^n$$

Définition 1.2.1 On appelle loi asymptotique des extrêmes, la loi de la v.a Y telle que :

$$Y_n = \frac{X_{n,n} - b_n}{a_n} \xrightarrow{L} Y$$

On se pose les questions suivantes :

1. Est-il possible de trouver deux suites $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ tels que

$$F_{X_{n,n}}(a_n x + b_n) \rightarrow H(x) \quad \text{quand } n \rightarrow \infty$$

en tout point x de continuité de H , où H est la fonction de répartition de Y .

2. H est de quel type ?
3. Quel est le lien entre F et H ?
4. Comment peut-on choisir les constantes de normalisation a_n et b_n afin que

$$a_n^{-1}(X_{n,n} - b_n) \xrightarrow{L} Y.$$

1.3 Lois limites des valeurs extrêmes

Le principal résultat de la théorie des valeurs extrêmes repose sur le théorème de Fisher-Tippett *Fisher et Tippett (1928, [29])* dont la première preuve rigoureuse est due à *Gnedenko (1943, [31])*. Le théorème de *Fisher et Tippett (1928, [29])* est un résultat très connu dans la théorie des valeurs extrêmes pour caractériser la loi de distribution des extrêmes. Ce théorème est une généralisation du théorème centrale limite (TCL) dans le cas des valeurs extrêmes.

Avant d'énoncer le théorème de Fisher-Tippett nous exposons le TCL.

Théorème 1.3.1 (*Théorème Central Limite (TCL)*)

Le Théorème Central Limite (TCL) donne, sous des conditions standards, la loi asymptotique de la moyenne

$$X_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

d'un échantillon (X_1, \dots, X_n) de variables indépendantes et identiquement distribuées :

$$\sqrt{n} \left(\frac{X_n - E(x)}{\sigma(x)} \right) \xrightarrow{\mathcal{L}} N(0, 1),$$

ou en termes de fonction de répartition

$$\lim_{n \rightarrow \infty} P \left(\sqrt{n} \left(\frac{X_n - E(x)}{\sigma(x)} \right) \leq x \right) = \Phi(x),$$

où Φ est la fonction de répartition de la loi $N(0, 1)$ (toute somme de variables aléatoires indépendantes et identiquement distribuées tend vers une variable aléatoire gaussienne).

Théorème des types extrêmes

Théorème 1.3.2 (*Théorème limite de Fisher-Tippett, 1928- Gnedenko, 1943*)

Soit $(X_i, i \geq 1)$ une suite des variables aléatoires indépendantes et de même loi et $X_{n,n} = \max(X_1, X_2, \dots, X_n)$. Supposons qu'il existe des constantes $a_n > 0$ et $b_n \in \mathbb{R}$ et une loi non-dégénérée de loi H telles que :

$$\lim_{n \rightarrow +\infty} F^n(a_n x + b_n) = H(x),$$

avec H une fonction de distribution non dégénérée.

Le théorème limite de Fisher-Tippett est vrai pour la majorité des lois usuelles.

Alors H appartient à l'un des trois types suivants (I, II, ou III)

Type I : Gumbel

$$\Lambda(x) = \exp\{-\exp(-x)\}, \quad x \in \mathbb{R}$$

Type II : Fréchet

$$\Phi_\alpha(x) = \begin{cases} \exp\{-x^{-\alpha}\}, & \text{si } x > 0 \\ 0, & \text{si } x \leq 0 \end{cases}$$

Type III : Weibull

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & \text{si } x \leq 0 \\ 1, & \text{si } x > 0 \end{cases}$$

Avec Φ_α , Λ et Ψ_α sont les lois limites possibles pour le maximum.

Supposons donner, n variables aléatoires X_1, \dots, X_n i.i.d de fonction de répartition commune F . Soit F suit une loi exponentielle de paramètre 1, c'est à dire

$$F(x) = 1 - \exp(-x)$$

Soit F_Y la fonction de répartition de la variable Y avec $Y = X_{n,n} - \log(n)$, alors

$$\begin{aligned}
F_Y(x) &= P[X_{n,n} - \log(n) \leq x] \\
&= P[X_{n,n} \leq x + \log(n)] \\
&= (P[X \leq x + \log(n)])^n \\
&= (1 - \exp(-(x + \log(n))))^n \\
&= (1 - \exp(-x - \log(n)))^n \\
&= \log\left(1 - \frac{1}{n} \exp(-x)\right)^n \\
&= n \log\left(1 - \frac{1}{n} \exp(-x)\right) \\
&= \exp\left(n \log\left(1 - \frac{1}{n} \exp(-x)\right)\right) \\
&\simeq \exp(-\exp(-x)) \quad \text{quand } n \rightarrow \infty \\
&= \Lambda(x).
\end{aligned}$$

Prenons $a_n = 1$, et $b_n = \log(n)$, alors $(X_{n,n} - b_n)/a_n$ tend asymptotiquement vers la loi de *Gumbel*.

Le théorème de Fisher-Tippett établit que la GEV est la distribution limite du maximum (adéquatement normalisé) d'une série de variables aléatoires *i.i.d.*

1.4 Loi Généralisée des Valeurs Extrêmes (GEV)

Grâce aux travaux de *von Mises* (1936, [53]) et de *Jenkinson* (1955, [35]), on a une forme unifiée de la fonction de répartition de la loi des valeurs extrêmes à un facteur d'échelle et de position près.

1.4.1 Représentation de *Jenkinson-Von Mises*

Il est possible de rassembler les trois familles de lois de valeur extrême en une seule famille paramétrique $(H_\gamma(x); \gamma \in \mathbb{R})$ dite famille des lois des valeurs extrêmes généralisées notée $(G.E.V.D)$ ou $(G.E.V)$. Elle est paramétrée par une seule variable $\gamma \in \mathbb{R}$, mais toujours à un facteur de changement d'échelle et de translation près. La fonction de répartition est pour $\gamma \in \mathbb{R}$; et pour tout x tel que $1 + \gamma x > 0$;

$$H_\gamma(x) = \begin{cases} \exp\left[-(1 + \gamma x)^{-\frac{1}{\gamma}}\right] & \text{si } \gamma \neq 0, 1 + \gamma x > 0 \\ \exp[-\exp(-x)] & \text{si } \gamma = 0, -\infty \leq x \leq \infty \end{cases}$$

L'ensemble $D = \left\{x \in \mathbb{R} : \left[1 + \gamma \frac{x-\mu}{\sigma}\right]^{-\frac{1}{\gamma}} > 0\right\}$.

La forme la plus générale de la distribution des valeurs extrêmes est :

$$H_{\gamma,\mu,\sigma} = \exp\left(-\left(1 + \gamma \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right), \quad \gamma \neq 0, 1 + \gamma \frac{x-\mu}{\sigma} > 0.$$

où H est une fonction de répartition non-dégénérée. μ est un paramètre de localisation, il est directement lié à la valeur la plus probable de la loi, il indique donc approximativement où se trouve le coeur de la distribution. σ est un paramètre de dispersion, il indique l'étalement des extrêmes. γ est l'indice de queue, déjà introduit.

Remarque 1.4.1 *Les trois lois des extrêmes peuvent être représentés en fonction de H_γ par*

$$H_\gamma = \begin{cases} \Lambda(x) & \gamma = 0, \\ \Phi_{1/\gamma}(x) & \gamma > 0, \\ \Psi_{-1/\gamma}(x) & \gamma < 0. \end{cases}$$

La densité correspondante à $H_{\gamma,\mu,\sigma}$ est

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

Cette densité est présentée dans la figure suivante

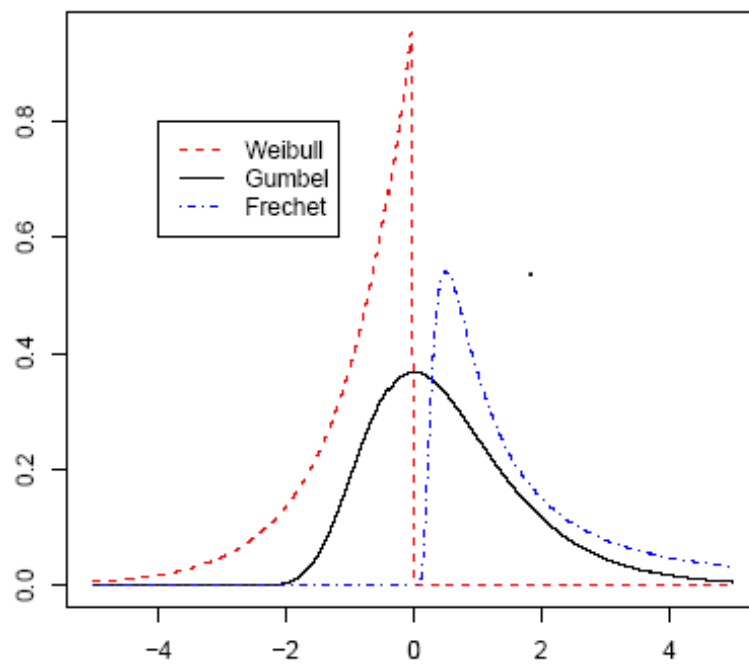


FIG. 1.2 – Fonctions de distribution du Gumbel (en noire), de Fréchet (en bleu) et de Weibull (en rouge).

1.4.2 Loi max-stable

Une loi H non dégénérée est dite max-stable s'il existe des constantes $a_n \in \mathbb{R}^{+*}$ et $b_n \in \mathbb{R}$ telles que

$$H^n(a_n x + b_n) = H(x); \quad \forall x \in \mathbb{R}$$

La loi de Weibull (Ψ_α) est une loi max-stable, en effet,

$$\begin{aligned} \Psi_\alpha^n\left(n^{-\frac{1}{\alpha}}x\right) &= \exp\left(-\left(-n^{-\frac{1}{\alpha}}x\right)^\alpha\right)^n \\ &= \exp\left(-\left(-n^{-1}x^\alpha\right)n\right) \\ &= \exp\left(-(-x^\alpha)\right) \\ &= \Psi_\alpha(x). \end{aligned}$$

1.5 Domaine d'attraction

L'analyse des valeurs extrêmes repose principalement sur les distributions limites des extrêmes et leurs domaines d'attraction. Ces distributions apparaissent comme les seules distributions limites possibles du maximum d'un échantillon de variables aléatoires indépendantes et identiquement distribuées (*i.i.d*) X_1, \dots, X_n de loi F .

La recherche du domaine d'attraction est équivalente à la réponse à la question suivante :

Etant donnée une distribution de valeurs extrêmes H_γ , sous quelles conditions la fonction de distribution F des maxima $X_{n,n}$ normalisés converge-t-elle faiblement vers H_γ ?

Définition 1.5.1 (*Domaine d'Attraction*)

On dit qu'une distribution F appartient au domaine d'attraction de la distribution des valeurs extrêmes H_γ ($F \in \mathcal{DA}(H_\gamma)$) s'il existe des suites réelles $a_n > 0$ et $b_n \in \mathbb{R}$ telles que :

$$P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) \xrightarrow{\mathcal{L}} H_\gamma(x) \quad \text{telle que } n \longrightarrow \infty.$$

c'est à dire

$$\lim_{n \rightarrow \infty} F_{X_{n,n}}(a_n x + b_n) = H_\gamma(x)$$

D'après cette définition, nous voyons que les trois distributions de valeurs extrêmes sont très différentes en terme de domaine d'attraction :

1. Si $\gamma = 0$, on dit que F appartient au domaine d'attraction de *Gumbel* et on notera $F \in \mathcal{DA}(\Lambda(x))$, nous trouvons des distributions dont la fonction de survie décroît vers zéro à un vitesse exponentielle (exponentielle, Normale, Gamma, Weibull, etc).

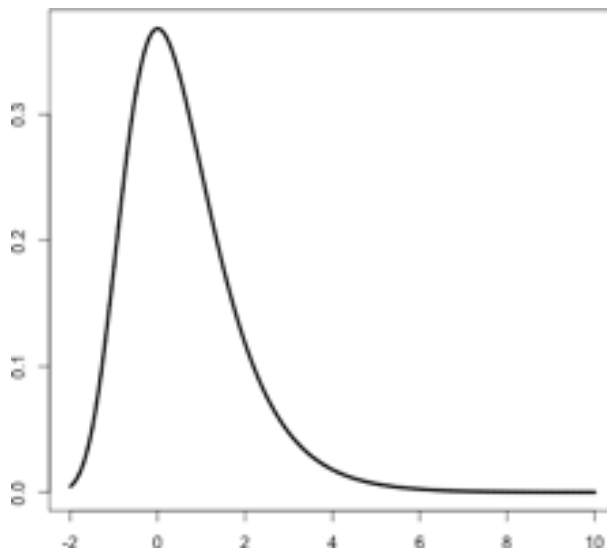


FIG. 1.3 – Queue légère

2. Si $\gamma > 0$, on dit que F appartient au domaine d'attraction de *Fréchet* et on notera $F \in \mathcal{DA}(\Phi_\alpha(x))$, nous trouvons des distributions à queue lourde (Cauchy, Pareto, Student, etc).

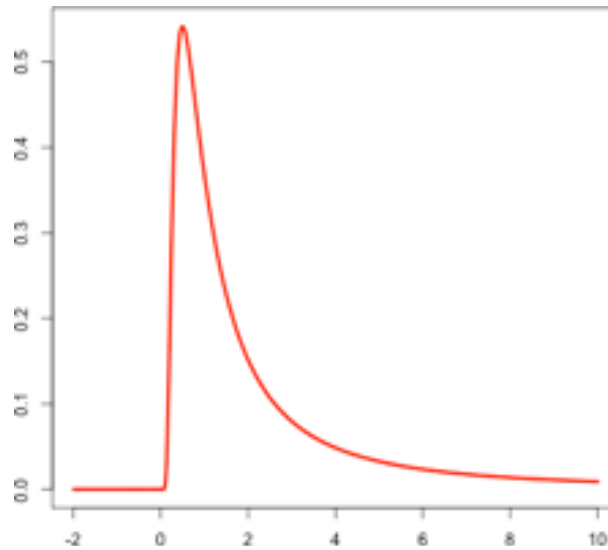


FIG. 1.4 – Queue lourde

3. Si $\gamma < 0$, on dit que F appartient au domaine d'attraction de *Weibull* et on notera $F \in \mathcal{DA}(\Psi_\alpha(x))$, nous trouvons des distributions dont le point terminal est fini (Uniforme, Beta, etc).

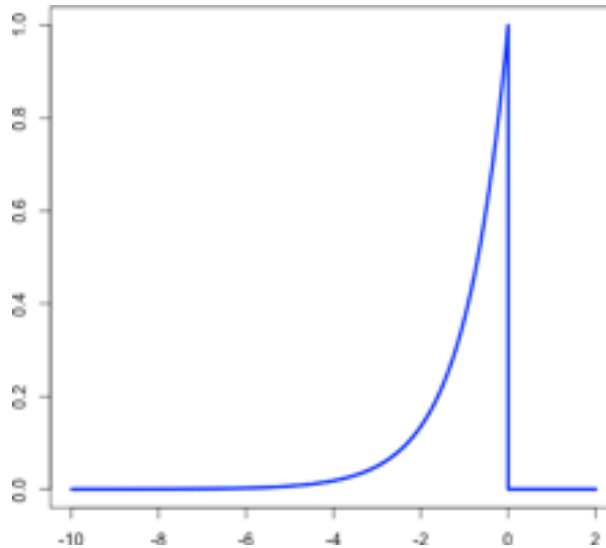


FIG. 1.5 – Queue finie

Sachant la distribution F , nous voudrions connaître à quel max-domaine d'attraction elle appartient et quelles sont les constantes de normalisation. La réponse est relativement complexe et n'est pas unique. Nous indiquons ici les critères les plus utilisés.

1.5.1 Fonction de même type

Définition 1.5.2 Deux fonctions de répartition F_1 et F_2 sont dites de même type si :

$$\exists a_0 > 0 \text{ et } b_0 \in \mathbb{R} \text{ tel que } F_2(x) = F_1(a_0x + b_0), \quad \forall x$$

Exemple 1.5.1 Soient X_1 et X_2 deux variables aléatoires des fonctions de répartition respectives F_1 et F_2 , On suppose $X_1 = a_0X_2 + b_0$

$$\text{Donc } X_2 = \frac{X_1 - b_0}{a_0}$$

$$\begin{aligned} F_2(x) &= P[X_2 \leq x] \\ &= P\left[\frac{X_1 - b_0}{a_0} \leq x\right] \\ &= P[X_1 - b_0 \leq a_0x] \\ &= P[X_1 \leq a_0x + b_0] \\ &= F_1(a_0x + b_0). \end{aligned}$$

F_1 et F_2 sont donc de même type.

1.5.2 Caractérisations des domaines d'attraction

Nous donnons dans cette partie la caractérisation des trois domaines d'attractions, *Fréchet*, *Weibull* et *Gumbel*.

D'abord, nous donnons la définition d'une fonction à variation régulière parce qu'elle est présente dans la caractérisation des domaines d'attraction.

Définition 1.5.3 (Fonction à variation régulière)

Une fonction mesurable f est dite à variation régulière (à l'infini) d'indice $\alpha \in \mathbb{R}$ si f est positive à l'infini

$$\lim_{x \rightarrow +\infty} \frac{f(tx)}{f(x)} = t^\alpha, \quad \forall t > 0.$$

Définition 1.5.4 (Fonction à variation lente)

f est dit à variation lente si :

$$\lim_{x \rightarrow +\infty} \frac{f(tx)}{f(x)} = 1$$

En remarquant que si f est à variation régulière d'indice α alors $\frac{f(x)}{x^\alpha}$ est à variation lente, il est facile de montrer qu'une fonction à variation régulière d'indice α peut toujours s'écrire sous la forme $x^\alpha l(x)$, où l est à variation lente.

Lemme 1.5.1 Soit l une fonction à variation lente. Alors on a : pour tout $\alpha > 0$; $l(x) = o(x^\alpha)$ en $+\infty$ et

$$\int_x^\infty l(t) t^{-\alpha-1} dt \sim \frac{1}{\alpha} x^{-\alpha} l(x) \quad \text{en } +\infty$$

Théorème 1.5.2 l est une fonction à variation lente si et seulement si : pour tout $x > 0$;

$$l(x) = c(x) \exp \left\{ \int_1^x t^{-1} \epsilon(t) \right\}$$

où c et ϵ sont des fonctions positives telles que

$$\lim_{x \rightarrow \infty} c(x) = c \in]0, +\infty[\text{ et } \lim_{t \rightarrow \infty} \epsilon(t) = 0$$

Exemple 1.5.3 $f(x) = \log(x)$

$$\lim_{t \rightarrow +\infty} \frac{\log(tx)}{\log(x)} = \lim_{t \rightarrow +\infty} \left(1 + \frac{\log t}{\log x} \right) = 1$$

La fonction \log est à variation lente.

Remarque 1.5.1 Si la fonction c est constante, on dit que l est normalisée.

Remarque 1.5.2 Si f est à variation régulière alors $f(x) = x^\alpha l(x)$, où $l(x)$ est à variation régulière lente.

Remarque 1.5.3 Soit f une fonction à variation régulière d'indice α . En utilisant le fait que $f(x) = x^\alpha l(x)$, on déduit facilement que pour tout $x > 0$

$$f(x) = c(x) \exp \left\{ \int_1^x t^{-1} \alpha(t) \right\}.$$

où c et α sont des fonctions positives telles que

$$\lim_{x \rightarrow \infty} c(x) = c \in]0, +\infty[\text{ et } \lim_{t \rightarrow \infty} \alpha(t) = \alpha.$$

Domaine d'attraction de Fréchet $DA(\Phi_\alpha(x))$

Théorème 1.5.4 (Caractérisation de $DA(\Phi_\alpha(x))$)

La fonction de répartition F appartient au domaine d'attraction de la loi de Fréchet $DA(\Phi_\alpha(x))$ de paramètre $\alpha > 0$ si et seulement si

$$x_F = +\infty \text{ et } \bar{F}(x) = x^{-\alpha} l(x), \quad \forall x > 0.$$

où la fonction l est à variation lente.

De plus

$$a_n = F^{-1}\left(1 - \frac{1}{n}\right) \quad \text{et} \quad b_n = 0.$$

La suite $(a_n^{-1}X_{n,n}, n \geq 1)$ converge en loi vers une variable aléatoire de fonction de répartition Φ_α

Les distributions du $DA(\Phi_\alpha(x))$ sont dites à queue épaisse.

Ce théorème formulé par *Gnedenko (1943)* permet de caractériser très simplement les distributions $F \in MDA(\Phi_\alpha(x))$. En effet, elles doivent vérifier

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

Prenons par exemple le cas de la distribution *Pareto*. Nous avons

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

Nous en déduisons que

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

Donc $\bar{F} \in RV_{-\frac{1}{\gamma}}$ et $F \in MDA(\Phi_{-\frac{1}{\gamma}}(x))$. Remarquons que $a_n = F^{-1}\left(1 - \frac{1}{n}\right) = n^\gamma$. Bien sur, nous vérifions

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

qui est la loi de Fréchet. La loi de Pareto appartient au domaine d'attraction maximal de Fréchet. Aussi, les lois dans le domaine d'attraction de la loi de Fréchet sont parfois appelées lois de type Pareto.

Le théorème précédent suggère une vérification graphique de la propriété $F \in MDA(\Phi_\alpha(x))$. Nous avons

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} \simeq x^{-\alpha}.$$

pour t suffisamment grand. Dans ce cas, nous vérifions que

$$\ln(1 - F(tx)) \simeq \ln(1 - F(t)) - \alpha \ln x$$

Par exemple, nous vérifions graphiquement que la distribution gaussienne n'appartient pas au max-domaine d'attraction de la distribution Fréchet.

Domaine d'attraction de Weibull $DA(\Psi_\alpha(\mathbf{x}))$

Théorème 1.5.5 (*Caractérisation de $DA(\Psi_\alpha(x))$*)

Une distribution F appartient au $DA(\Psi_\alpha(x))$ si

$$x_F < +\infty \quad \text{et} \quad 1 - F\left(x_F - \frac{1}{x}\right) = x^{-\alpha}l(x),$$

où l est une fonction à variation régulière lente.

Dans ce cas

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

Prenons par exemple le cas de la distribution uniforme. Nous avons $x_F = 1$

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1 - F(1 - t^{-1}x^{-1})}{1 - F(t^{-1})} &= \lim_{t \rightarrow \infty} \frac{t^{-1}x^{-1}}{t^{-1}} \\ &= x^{-1}. \end{aligned}$$

Donc $F \in MDA(\Psi_1)$. Remarquons $a_n = 1 - F^{-1}(1 - n^{-1}) = n^{-1}$. Bien sur, nous vérifions

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

Les distributions appartenant à ce domaine d'attraction sont sans queue.

Domaine d'attraction de Gumbel $DA(\Lambda(\mathbf{x}))$

Théorème 1.5.6 (*Caractérisation de $DA(\Lambda(x))$*)

Une distribution F appartient au $DA(\Lambda(x))$ si F est une fonction de von Mises

Définition 1.5.5 (*Fonction de von Mises*)

Soit F une fonction de distribution de point terminal x_F . S'il existe $z < x_F$ tel que

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

où g et c sont deux fonctions mesurables telles que $c(x) \rightarrow c > 0$ et $g(x) \rightarrow 1$ quand $x \rightarrow x_F$ et a est une fonction positive absolument continue de densité a' vérifiant $\lim_{x \rightarrow x_F} a'(x) = 0$, alors F est une fonction de von Mises et a sa fonction auxiliaire.

Un choix possible pour a_n et b_n est

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

Prenons par exemple le cas de la loi exponentielle. Pour la loi exponentielle de paramètre $\lambda = 1$, la fonction de répartition de cette loi est

$$F(x) = 1 - e^{-x} \quad \text{pour } x \geq 0$$

En posant $b_n = \log(n)$ et $a_n = 1$ alors,

$$\begin{aligned} \lim_{n \rightarrow \infty} F^n(a_n x + b_n) &= \lim_{n \rightarrow \infty} (1 - e^{-x - \log n})^n \\ &= \lim_{n \rightarrow \infty} \left(1 - \frac{e^{-x}}{n} \right)^n \\ &= \exp \{ -e^{-x} \} \\ &= \Lambda(x). \end{aligned}$$

c'est-à-dire que le maximum convenablement normalisé de la loi exponentielle converge vers la loi de Gumbel. Cette loi appartient au domaine d'attraction maximal de Gumbel.

Aussi, les lois dans le domaine d'attraction de la loi de Gumbel sont parfois appelées lois de type exponentiel.

Les distributions appartenant au $DA(\Lambda(x))$ sont des queues légères.

1.5.3 Condition nécessaire et suffisante de convergence

1. F appartient au domaine maximal d'attraction de Fréchet $DA(\Phi_\alpha(x))$ si et seulement si

$F^{-1}(1) = +\infty$ et $\exists \alpha > 0$ tel que

$$\lim_{x \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\alpha} \quad \forall x > 0$$

On peut écrire,

$$x^{-\alpha} = -\log \Phi_\alpha(x)$$

2. F appartient au domaine maximal d'attraction de de Weibull $DA(\Psi_\alpha(x))$ si et seulement si $F^{-1}(1) < +\infty$ et $\exists \alpha > 0$ et

$$\lim_{\gamma \rightarrow \infty} \frac{1 - F(F^{-1}(1) - \gamma x)}{1 - F(F^{-1}(1) - \gamma)} = x^{-\alpha} \quad \forall x > 0,$$

On peut écrire,

$$x^{-\alpha} = -\log \Psi_\alpha(x)$$

3. F appartient au domaine maximal d'attraction de de Gumbel $DA(\Lambda(x))$ si et seulement si

$$E[X/X > C] < +\infty \quad \text{et} \quad \forall C < F^{-1}(1) \quad \text{et}$$

$$\lim_{t \rightarrow F^{-1}(1)} \frac{1 - F(t + xE[X - t/X > t])}{1 - F(t)} = e^{-x} \quad \forall x \in \mathbb{R}$$

De même, $e^{-x} = -\log \Lambda_\alpha, \forall x \in \mathbb{R}$

1.5.4 Condition de second ordre :

On définit une fonction U par :

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma},$$

où γ est une valeur constant dans \mathbb{R} appelle l'indice des valeurs extrêmes.

Définition 1.5.6 On dit que une fonction U (ou est une lois de probabilité) satisfait la condition de second ordre si pour une fonction positive a et pour une fonction négative A avec

$$\lim_{t \rightarrow \infty} A(t) = 0,$$

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

Où la fonction A (resp la fonction a) est le second fonction auxiliaire(resp est le premier fonction auxiliaire). Telle que H n'est pas un multiple de $x^\gamma - 1/\gamma$ en particulier n'est pas identiquement zéro.

Théorème 1.5.7 (Théorème de Karamata)

Soit F une fonction à variation régulière. Alors il existe une fonction à variation lente L telle que :

$$F(x) = x^{-\alpha} L(x)$$

1.5.5 Constantes de normalisation

Théorème 1.5.8 *On peut choisir des constantes $a_n > 0$ et $b_n \in \mathbb{R}$ soit vérifiée de la manière suivante :*

1. $b_n = 0$, $a_n = F^{-1} \left(1 - \frac{1}{n} \right)$ si $H = \Phi$
2. $b_n = F^{-1} (1)$, $a_n = F^{-1}(1) - F^{-1} \left(1 - \frac{1}{n} \right)$ si $H = \Psi$
3. $b_n = F^{-1} \left(1 - \frac{1}{n} \right)$, $a_n = E \left(X - a_n / X > a_n \right)$ si $H = \Lambda$
ou bien $a_n = F^{-1} \left(\frac{1}{ne} \right) - F^{-1} \left(1 - \frac{1}{n} \right)$.

1.6 Loi Généralisée de Pareto (GPD)

La méthode des excès au-delà d'un seuil (ou Peak Over Threshold, POT) s'intéresse au comportement (statistique) des observations au-delà d'un seuil u fixé (a priori u sera choisi « *grand* »). En d'autres termes, elle consiste à observer non pas le maximum ou les plus grandes valeurs mais toutes les valeurs des réalisations qui excèdent un certain seuil élevé. L'idée de base de cette approche consiste à choisir un seuil suffisamment élevé et à étudier les excès au-delà de ce seuil. Cette méthode initialement développée par *Pickands (1975)* et abondamment étudiée par divers auteurs tels que de *Smith (1987)*, *Davison et Smith (1990)*, ou *Reiss et Thomas (2001)* pour de plus amples références.

Nous supposons une suite d'observations X_1, X_2, \dots, X_n *i.i.d.* de fonction de répartition F et x_F un point terminal.

Les excès au-delà du seuil u sont définis comme l'ensemble des variables aléatoires Y tel que : $y_j = x_j - u$, $x_j > u$.

Le principe de cette méthode est représenté dans la figure suivante

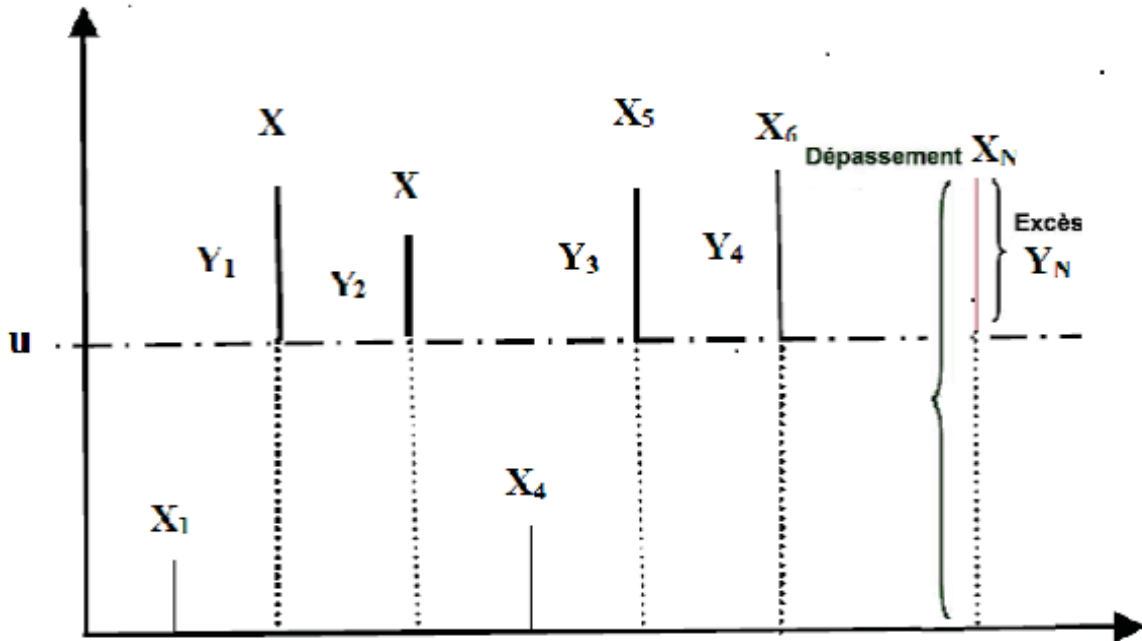


FIG. 1.6 – Représentation des excès Y issus des dépassements X au-delà d'un seuil u .

1.6.1 Distribution des excès

On cherche à partir de la loi F de X à définir une loi conditionnelle F_u par rapport au seuil u pour les variables aléatoires dépassant ce seuil.

On définit alors la loi conditionnelle des excès F_u par : pour un seuil $u < x_F$ fixé. On déduit la variable $X_i - u$, $i = 1, \dots, n$, l'excès au-dessus du seuil u .

Dans les modèles à seuil (*threshold models*), nous nous intéressons à la probabilité conditionnelle suivante

$$\Pr \{X - u \leq y / X > u\} = \frac{1 - F(u + y)}{1 - F(u)} = 1 - \frac{\bar{F}(u + y)}{\bar{F}(u)} \quad (1.2)$$

Pour cela, nous introduisons la fonction de distribution des excès par rapport à un seuil u :

$$F_u(y) = \Pr \{X - u \leq y / X > u\}$$

pour $0 \leq y \leq x_F - u$ avec $x_F = \sup \{x \in \mathbb{R} : F(x) < 1\}$. La fonction de distribution des excès représente la probabilité que la variable aléatoire X dépasse

le seuil u d'au plus une quantité y , sachant qu'elle dépasse u . Remarquons que l'on peut écrire cette fonction de distribution sous la forme

$$\begin{aligned} F_u(y) &= 1 - \left(\frac{1 - F(u+y)}{1 - F(u)} \right) \\ &= \frac{F(u+y) - F(u)}{1 - F(u)} \quad \text{pour } 0 \leq y \leq x_F - u \end{aligned}$$

1.6.2 Distribution de Pareto Généralisé (GPD)

Lorsqu'on s'intéresse aux valeurs extrêmes d'un phénomène, il est naturel de considérer la loi des observations « au-delà d'un seuil » alors on déduit que la distribution de Pareto généralisées, joue un rôle essentiel dans la modélisation des excès.

Définition 1.6.1 Une distribution $G_{\gamma, \beta(u)}$ est dite de Pareto généralisée de paramètres γ et $\beta(u)$ si elle s'écrit :

$$G_{\gamma, \beta(u)}(x) = \begin{cases} 1 - (1 + \gamma \frac{x}{\beta(u)})^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, \\ 1 - \exp(-\frac{x}{\beta(u)}) & \text{si } \gamma = 0. \end{cases}$$

où $x \geq 0$ si $\gamma \geq 0$ et $0 \leq x \leq -\frac{x}{\beta(u)}$ si $\gamma < 0$.

$\gamma, \beta(u)$ sont des paramètres de forme et d'échelle.

Cette distribution limite est la distribution de Pareto généralisée (GPD). La loi des valeurs au-delà d'un seuil admet ainsi une distribution limite. L'égalité ci-dessus peut être reformulée de la manière suivante : lorsque u est grand (c'est à dire proche du point terminal x_F), alors on peut trouver

$$F(u) \approx G_{\gamma, \beta(u)}(x).$$

L'objectif de la méthode des excès est de déterminer par quelle loi de probabilité on peut approcher la distribution des excès. Le théorème Pickands-Balkema-de Haan est le second théorème de la théorie des valeurs extrêmes, il permet d'approximer la distribution au-delà d'un certain seuil (assez grand) par une distribution de Pareto généralisée.

1.6.3 Le choix du seuil

Une bonne détermination de u est primordiale si nous voulons appliquer la théorie des valeurs extrêmes. En effet, un seuil trop faible conduit à une mauvaise adéquation entre la fonction de survie et la loi *GPD* et les estimateurs sont biaisés. Inversement, un seuil trop élevé entraîne une faible taille d'échantillon de dépassements et donc des estimateurs volatiles.

1.6.4 L'espérance au-delà d'un seuil

L'espérance au-delà du seuil u (ou « *moyenne en excédent* ») est définie pour une distribution F par

$$e(u) = E(X - u | X > u), \quad u \leq x_F$$

Il s'agit donc de l'espérance de la distribution conditionnelle.

1.7 Théorème de Balkema-de Haan-Pickands

Le théorème suivant fait le lien entre le comportement asymptotique de la distribution des excès et la loi de Pareto généralisé.

Théorème 1.7.1 (de Balkema-de Haan-Pickands)

Soit F_u la distribution des excès. Si $F \in \mathcal{DA}(H_\gamma)$, la *GPD* est la distribution limite de la distribution des excès lorsque le seuil tend vers x_F :

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x \leq x_F - u} |F_u(x) - G_{\gamma, \beta(u)}(x)| = 0.$$

où $\beta(u)$ une fonction positive mesurable.

Chapitre 2

Estimation non paramétrique de l'indice de queue des valeurs extrêmes

Une large catégorie d'estimateurs à été élabornée spécifiquement pour le paramètre γ . Les estimateurs qui font partie de cette catégorie, tel que l'estimateur de *Pickands (1975, [43])*, *Hill (1975, [34])*, et *Dekkers- Einmahl-De Haan (1989, [19])*, sont qualifiés d'estimateurs non paramétriques. Une bonne estimation de γ peut être obtenue en identifiant une zone de stabilité sur le graphique qui représente différentes valeurs des estimateurs non paramétriques, en fonction du rang k correspondant à une statistique d'ordre $X_{(k)}$.

2.1 L'estimateur de Pickands

Cet estimateur a été introduit par *Pickands (1975, [43])* pour $\gamma \in \mathbb{R}$. Il est le premier estimateur de γ .

Définition 2.1.1 *L'estimateur de Pickands (1975, [43]) combine 4 statistique d'ordre. Il est calculé pour un ensemble de rang k . Sa formule est donnée par :*

$$\hat{\gamma}_{n,k}^{(p)} = \frac{1}{\ln 2} \ln \left(\frac{X_{(n-k+1,n)} - X_{(n-2k+1,n)}}{X_{(n-2k+1,n)} - X_{(n-4k+1,n)}} \right) \quad (2.1)$$

où X_1, \dots, X_n un échantillon de variables aléatoires (*i.i.d*) et n est la taille de l'échantillon observé.

Pickands (1975, [43]) a démontré la consistance faible de son estimateur. La convergence forte ainsi que la normalité asymptotique ont été démontrées par *Dekkers et De Haan [21]* Des améliorations de cet estimateur ont été introduites notamment par *Drees [23]* et *Segers [48]*.

Théorème 2.1.1 (*propriétés de $\hat{\gamma}_{n,k}^{(p)}$*)

Supposons que $F \in DA(H_\gamma)$, $\gamma \in \mathbb{R}$, $k = k(n) \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$.

(a) Consistance faible :

$$\hat{\gamma}_{n,k}^{(P)} \xrightarrow{p} \gamma \text{ quand } n \rightarrow \infty$$

(b) Consistance forte : Si $k/\log\log n \rightarrow \infty$ quand $n \rightarrow \infty$, alors

$$\hat{\gamma}_{n,k}^{(P)} \xrightarrow{p.s.} \gamma \text{ quand } n \rightarrow \infty$$

(c) Normalité asymptotique : Sous des conditions additionnelles sur la suite $k(n)$ et la distribution F

$$\sqrt{k}(\hat{\gamma}_{n,k}^{(P)} - \gamma) \xrightarrow{\mathcal{L}} N\left(0, \frac{\gamma^2(2^{2\gamma+1} + 1)}{4(\log 2)^2(2^\gamma - 1)^2}\right) \text{ quand } n \rightarrow \infty$$

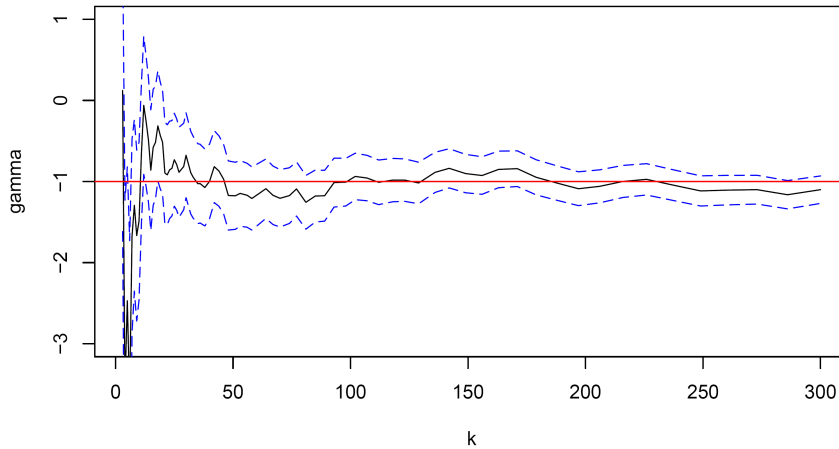


FIG. 2.1 – Estimateur de Pickands avec un intervalle de confiance 95% de γ basés sur 100 échantillons de taille 3000 pour la loi uniforme.

2.2 L'estimateur de Hill

L'estimateur de *Hill* (1975, [34]) est l'un des estimateurs les plus répandus dans les applications de la TVE. Introduit par *Hill* (1975, [34]), il est applicable seulement dans le cas où $\gamma > 0$.

Définition 2.2.1 *L'estimateur de Hill* (1975, [34]) *défini par :*

$$\hat{\gamma}_{k,n}^{(H)} = \frac{1}{k} \sum_{i=n-k+1}^n \ln X_{(i,n)} - \ln X_{(n-k+1,n)} \quad (2.2)$$

La consistance faible de cet estimateur a été démontrées par *Mason* [38] et *Deheuvels et al.* [18] et plus récemment par *Necir* en 2006 [40] La normalité asymptotique a été démontrées par *Davis et Resnick* [13], *Csörgő et Mason* [13], *Häusler et Teugels* [32], ...etc. Le théorème suivant présente les propriétés de l'estimateur de *Hill* :

Théorème 2.2.1 (Propriétés de $\hat{\gamma}_n^H$)

Supposons que $F \in DA(H_\gamma)$, $\gamma > 0$, $k = k(n) \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$.

(a) Consistance faible :

$$\hat{\gamma}_{n,k}^{(H)} \xrightarrow{p} \gamma \text{ quand } n \rightarrow \infty$$

(b) Consistance forte : Si $k/\log \log n \rightarrow \infty$ quand $n \rightarrow \infty$

$$\hat{\gamma}_{n,k}^{(H)} \xrightarrow{p.s.} \gamma \text{ quand } n \rightarrow \infty$$

(c) Normalité asymptotique : Soient $F \in \mathcal{D}(H_\gamma)$, $k(n)_{n \geq 1}$ une suite d'entiers telle que $1 \leq k(n) \leq n$ où $k(n) \rightarrow \infty$ et $\frac{k(n)}{n} \rightarrow 0$. Si F vérifier la condition de second ordre avec $\lim_{n \rightarrow \infty} \sqrt{k}A(n/k) = 0$; alors

$$\sqrt{k}(\hat{\gamma}_{n,k}^{(H)} - \gamma) \xrightarrow{\mathcal{L}} N(0, \gamma^2) \text{ quand } n \rightarrow \infty$$

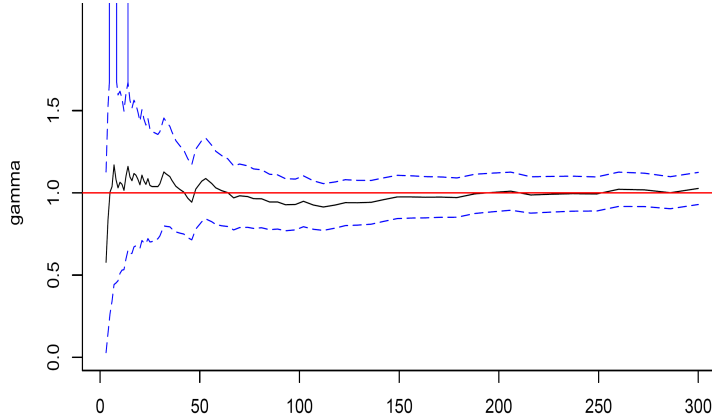


FIG. 2.2 – L'estimateur de Hill avec un intervalle de confiance 95% de γ basés sur 100 échantillons de taille 3000 pour la loi de Pareto standard.

2.3 L'estimateur des moments de Dekkers - Einmahl-De Haan

En 1989, *Dekkers-Einmahl-DeHaan* [19] ont proposé un estimateur de $\gamma \in \mathbb{R}$ qui présente une extension de l'estimateur de Hill i.e : l'estimateur de Hill est une généralisation de l'estimateur de Hill à l'ensemble des lois, appelé un estimateur de moment car il a été obtenu comme l'estimateur des moments pour $\gamma \in \mathbb{R}$.

Définition 2.3.1 *L'estimateur des moments est défini comme suit :*

$$\hat{\gamma}_{n,k}^{(M)} = 1 + M_n^{(1)} - \frac{1}{2} \left(\frac{(M_n^{(1)})}{(M_n^{(2)})} - 1 \right)^{-1} \quad (2.3)$$

$$M_n^{(r)} = M_n^{(r)}(k) = \frac{1}{k} \sum_{i=1}^{n-k} (\ln X_{(n-i+1,n)} - \ln X_{(n-k,n)})^r, \quad r = 1, 2. \quad (2.4)$$

Sous certaines conditions sur k , l'estimateur de *Dekkers-Einmahl-DeHaan* converge asymptotiquement vers la loi normale.

Dekkers-Einmahl-DeHaan [19] ont démontré la consistance faible, la consistance forte et la normalité asymptotique de leur estimateur.

Théorème 2.3.1 (Propriétés de $\hat{\gamma}_{n,k}^{(M)}$)

Supposons que $F \in DA(H_\gamma)$, $\gamma \in \mathbb{R}$, $k = k(n) \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$.

(a) Consistance faible :

$$\hat{\gamma}_{n,k}^{(M)} \xrightarrow{p} \gamma \text{ quand } n \rightarrow \infty$$

(b) Consistance forte : Si $k/\log \log n \rightarrow \infty$ quand $n \rightarrow \infty$, alors

$$\hat{\gamma}_{n,k}^{(M)} \xrightarrow{p.s.} \gamma \text{ quand } n \rightarrow \infty$$

(c) Normalité asymptotique : Sous des conditions de régularité convenables

$$\sqrt{k}(\hat{\gamma}_{n,k}^{(M)} - \gamma) \xrightarrow{\mathcal{L}} N(0, \eta^2) \text{ quand } n \rightarrow \infty,$$

ou

$$\eta^2 = \begin{cases} 1 + \gamma^2 & \text{si } \gamma \geq 0 \\ (1 - \gamma)^2(1 - 2\gamma)[4 - 8\frac{1 - 2\gamma}{1 - 3\gamma} + \frac{(5 - 11\gamma)(1 - 2\gamma)}{(1 - 3\gamma)(1 - 4\gamma)}] & \text{si } \gamma < 0 \end{cases}$$

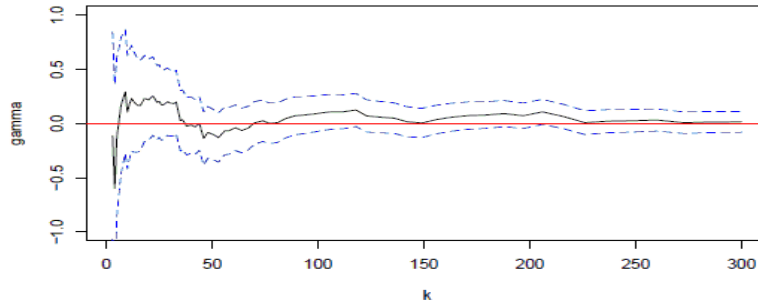


FIG. 2.3 – Estimateur des moments avec un intervalle de confiance 95% de γ basés sur 100 échantillons de taille 3000 pour la loi de Gumbel standard.

2.4 Comparaison entre les trois estimateurs

Les estimateurs non paramétriques de γ ont été comparés par *Emprechts et al. [26]*, la question du meilleur estimateur n'a pas de réponse tranchée. L'estimateur de Pickands est en particulier difficile à utiliser car il présente une forte instabilité. L'estimateur de Hill présente une variance asymptotique plus faible que les deux autres dans le cas où $\gamma > 0$. Il est généralement le plus utilisé.

2.5 Choix du seuil k

Les estimateurs de l'indice de queue γ sont basé sur l'utilisation du k plus grandes observations. La question que l'on peut posé est comment choisir le seuil k ?

Le seuil doit être choisi de façon à faire un compromis : plus le seuil est élevé, plus le biais est réduit et la variance est grande ce qui donne un meilleur modèle, par ailleurs, plus le seuil est bas, plus le biais est grand et la variance de l'estimateur est réduite car plus de données participent à l'estimation.

Plusieurs techniques ont été proposées pour trouver une fonction optimale $k_{opt}=k_{opt}(n)$, comme les méthodes de *Dekkers-Einmahl-De Haan [21]* et *Cheng et*

Peng [7], Drees et Kaufmann [24], Hall et Welsh [33] et Neves et al.[41] . Dans notre mémoire nous commençons présenter une méthode graphique universelle qui devrait être appliquée avant toute recherche numérique puis la méthode de *Cheng et Peng [7]*.

Méthodes Graphique

Cette méthode consiste à tracer

$$\{(k, \hat{\gamma}(k)) : k = 1, \dots, n\},$$

L'idée est de choisir le seuil k_{opt} dans la première région pour laquelle le graphe des $\hat{\gamma}(k)$ est approximativement horizontal. Dans *la figure 2.1*, il semble que k_{opt} est compris entre 90 et 102 pour la loi de Pareto standard.

Méthode de Cheng et Peng

En 2001, *Cheng et Peng* dans [7] ont proposé une valeur optimale k_{opt} définie comme suite

$$k_{opt} = \begin{cases} \eta^{-\hat{\rho}/(1-\hat{\rho})} \left(\frac{(1+2z_\alpha^2)}{3\hat{\delta}(1-2\hat{\rho})} \right)^{1/(1-\hat{\rho})} & \text{si } \hat{\delta} > 0 \\ \eta^{-\hat{\rho}/(1-\hat{\rho})} \left(\frac{(1+2z_\alpha^2)}{-3\hat{\delta}} \right)^{1/(1-\hat{\rho})} & \text{si } \hat{\delta} < 0 \end{cases}$$

ou z_α est le quantile d'ordre $1 - \alpha$ de $N(0, 1)$,

$$\hat{\rho} = -\frac{1}{\log 2} \left(\left| \frac{M_n^{(2)}(n/(2\sqrt{\log n})) - 2 \{M_n^{(1)}(n/(2\sqrt{\log n}))\}^2}{M_n^{(2)}(n/(\sqrt{\log n})) - 2 \{M_n^{(1)}(n/\sqrt{\log n})\}^2} \right| \right)$$

$$\hat{\delta} = (1 + 2\hat{\rho}) (\log n)^{-\hat{\rho}} \frac{M_n^{(2)}(n/(\sqrt{\log n})) - 2 \{M_n^{(1)}(n/\sqrt{\log n})\}^2}{-2\hat{\rho} \{M_n^{(1)}(n/\sqrt{\log n})\}^2}$$

avec $M_n^{(r)}(k)$, ($r = 1, 2$), définies dans (2.4).

2.5.1 Estimation de la queue de la distribution

L'étude des événements extrêmes se ramène à l'estimation des queues de la distribution, i.e : la probabilité de survenir un événement extrême. La GPD est employée pour modéliser la queue (supérieure ou inférieure) d'une distribution présentant des valeurs extrêmes.

En posant $x = u + y$ dans la formule (1.2), on peut écrire \bar{F} sous la forme

$$\bar{F}(x) = \bar{F}(u) \bar{F}(x - u), \quad u < x < x_F$$

Donc, l'estimateur de $\hat{\bar{F}}$ est

$$\hat{\bar{F}}(x) = \hat{\bar{F}}(u) \hat{\bar{F}}(x - u), \quad u < x < x_F$$

or,

$$\hat{\bar{F}}_u(x - u) = \bar{G}_{\hat{\gamma}_{N_u}, \hat{\sigma}_{N_u}}(x - u) = \left(1 + \hat{\gamma}_{N_u} \frac{x - u}{\hat{\sigma}_{N_u}}\right)^{-1/\hat{\gamma}_{N_u}}, \quad u < x < x_F$$

et

$$\hat{\bar{F}}(u) = \hat{\bar{F}}_u(u) = \frac{1}{n} \sum_{i=1}^n 1_{\{x_i > u\}} = \frac{N_u}{n}, \quad u < x_F$$

Donc

$$\hat{\bar{F}}(u) = \hat{p} := \frac{N_u}{n} = \left(1 + \hat{\gamma}_{N_u} \frac{x - u}{\hat{\sigma}_{N_u}}\right)^{-1/\hat{\gamma}_{N_u}}, \quad u < x < x_F$$

2.5.2 Détermination du seuil u

Il existe différentes méthodes pour choisir le seuil u au dessus duquel les valeurs sont modélisées par une GPD (ce qui parallèle le problème de déterminer k pour les estimateurs de GEVD).

Méthode des moyennes des excés

Davison et Smith [14] ont proposé une méthode graphique de l'estimation du seuil.

Cette méthode consiste à tracer la fonction des moyennes des excés pour différents seuils u

$$\{(u, e_n(u)) : X_{1,n} \leq u \leq X_{n,n}\},$$

ou, $e_n(u)$ est la moyenne des excés empirique qui donnée comme suit

$$e_n(u) = \frac{1}{\bar{F}_n(u)} \int_u^\infty \bar{F}_n(t) dt = \frac{1}{N_u} \sum_{i=1}^n (X_i - u) 1_{\{x_i > u\}}$$

L'idée est de retenir le seuil u pour lequel la fonction des moyennes des excès $e_n(x)$ est approximativement linéaire pour $x \geq u$ qui est justifiée par le fait que cette fonction est linéaire en u pour une GPD.

2.6 Construction de l'estimateur de Hill

Cet estimateur a été introduit par *Hill (1975, [34])* pour estimer d'une manière non paramétrique le paramètre de queue des lois appartenant au domaine d'attraction de Fréchet. Pour construire son estimateur, *Hill* utilise la méthode du maximum de vraisemblance sur l'ensemble des k plus grandes observations d'un échantillon, en supposons que $\gamma > 0$, et F vérifie la condition :

$$\frac{1 - F(x)}{1 - F(u)} := \left(\frac{x}{u}\right)^{-\frac{1}{\gamma}}, \quad x \geq u \text{ où } u \in \mathbb{R} \quad (2.5)$$

Alors on utilise seulement les statistiques d'ordre qui dépassent le seuil u : On a, les statistique d'ordre $X_{(n-k+1,n)}, \dots, X_{(n,n)}$. La fonction de log-vraisemblance est donnée par :

$$L(\gamma, X_{n-k+1,n}, \dots, X_{n,n}) = -k \log(\gamma u) + \log(1 - F(u)) - \frac{\gamma + 1}{\gamma} * \sum_{i=1}^k (\log(X_{n-i+1}) - \log(u)) \quad (2.6)$$

En résolvant cette équation et remplaçant le seuil u par $X_{n-k,n}$, on obtient l'estimateur de Hill. Pour caractériser l'estimateur de Hill, nous allons recourir à leur formule principal.

Il est défini de la façon suivante (pour tout $\gamma > 0$) :

$$\hat{\gamma}_{k,n}^H = \frac{1}{k} \sum_{i=n-k+1}^n \ln X_{(i,n)} - \ln X_{(n-k+1,n)} \quad (2.7)$$

Soit F une fonction de répartition, le paramètre $\alpha = \frac{1}{\gamma} > 0$ est appelé indice de queue de F . Soit F une fonction de répartition. on dit que $F \in D(H_{\frac{1}{\gamma}})$ ssi,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} &= x^{-\frac{1}{\gamma}}, \quad \gamma > 0 \\ \gamma > 0 : F(x) < 1, \quad \forall x, \quad \int_1^{+\infty} (1 - F(x)) \frac{dx}{x} &< \infty \\ \lim_{t \rightarrow \infty} \frac{\int_t^{+\infty} (1 - F(x)) \frac{dx}{x}}{1 - F(t)} &= \gamma \end{aligned} \quad (2.8)$$

La formule (2.8) équivalence à :

$$\lim_{t \rightarrow \infty} E(\log X - \log t / X > t) = \gamma$$

donc :

$$E[(\log X - \log t / X > t) = \frac{\int_t^{+\infty} (1 - F(x)) \frac{dx}{x}}{1 - F(t)}$$

Par intégration par partie on trouve que :

$$\int_t^{+\infty} (1 - F(s)) \frac{ds}{s} = \int_t^{+\infty} (\log(u) - \log(t)) dF(u)$$

Donc nous avons :

$$\lim_{t \rightarrow \infty} \frac{\int_t^{+\infty} (\log(u) - \log(t)) dF(u)}{1 - F(t)} = \gamma \quad (2.9)$$

Afin de développer un estimateur basé sur ce resultat asymptotique, on remplace dans (2.9) la valeur t par la statistique d'ordre $X_{n-k;n}$ et F par la distribution empirique F_n on trouve,

$$\hat{\gamma} = \frac{\int_{X_{n-k;n}}^{+\infty} (\log(u) - \log(X_{n-k;n})) dF_n(u)}{1 - F_n(X_{n-k;n})}$$

ce $\hat{\gamma}$ est appelé l'estimateur de Hill (1975, [34]) et on note $\hat{\gamma}_{n,k}^{(H)}$. Sa forme générale s'écrit par :

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

Lemme 2.6.1

Soit F une distribution appartient au domaine d'attraction de la loi de Fréchet de paramètre $\alpha > 0$. On a :

$$\frac{1}{\bar{F}(t)} E[(\log(X) - \log(t)) \mathbf{1}_{\{X > t\}}] \xrightarrow{t \rightarrow \infty} \frac{1}{\alpha} = \gamma$$

Démonstration :

On déduit de la définition des fonctions à variation lente et du théorème (1.5.4) que $F \in D(H\gamma)$; où $1/\alpha = \gamma$; si et seulement si $\lim_{t \rightarrow \infty} \frac{\bar{F}(x)}{\bar{F}(t)} = x^{-\alpha}$ pour tout $x > 0$. Supposons par simplicité que la loi de X possède la densité f .

Par intégration par partie, on a pour $t > 1$

$$\begin{aligned} E[(\log(X) - \log(t))\mathbf{1}_{\{X>t\}}] &= \int_t^{+\infty} (\log(x) - \log(t))f(x)dx \\ &= [-\bar{F}(x)(\log(X) - \log(t))]_t^{+\infty} + \int_t^{+\infty} \frac{\bar{F}(x)}{x} dx \end{aligned}$$

En fait le membre de gauche est égal au membre de droite en toute généralité. Grâce au lemme (1.5.1), on a $\bar{F}(x) = x^{-\zeta}\ell(x) = o(x^{-\alpha+a})$, $-\alpha + a < 0$.

Le membre de droite de l'équation ci-dessus se réduit donc à $\int_t^{+\infty} \frac{\bar{F}(x)}{x} dx$. On a d'après le deuxième partie du lemme (1.5.1) :

$$\int_t^{+\infty} \frac{\bar{F}(x)}{x} dx = \int_t^{+\infty} x^{-\alpha-1}\ell(x)dx \sim \frac{1}{\alpha}t^{-\alpha}\ell(t)dt = \frac{1}{\alpha}\bar{F}(t) \quad (2.10)$$

Il nous faut maintenant trouver un estimateur de $\bar{F}(t) = E[\mathbf{1}_{\{X_i>t\}}]$ et un estimateur de $E[(\log(X) - \log(t))\mathbf{1}_{\{X>t\}}]$.

La loi forte des grands nombres assure que $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i>t\}}$ converge presque sûrement vers $\bar{F}(t)$. Il reste à remplacer t par une quantité qui tende vers $+\infty$ avec n , il est naturel de remplacer t par $X_{n-k_n+1,n}$, où la suite $(k_n, n \geq 1)$ satisfait les hypothèses suivantes : $\lim_{n \rightarrow \infty} k_n = +\infty$, et $\lim_{n \rightarrow \infty} k_n/n = 0$.

Cette dernière condition assure d'après la proposition (1.1.2) et le théorème (1.5.4) que presque sûrement $X_{n-k_n+1,n}$ diverge vers l'infinie.

Pour alléger les notations, notons $k_n = k$. S'il on suppose que F est continue, la statistique d'ordre est strictement croissante presque sûrement, et on a pour estimation de $\bar{F}(X_{n-k+1,n})$:

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i>X_{n-k+1,n}\}} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_{i,n}>X_{n-k+1,n}\}} = \frac{k-1}{n} \quad (2.11)$$

La loi forte des grands nombres assure que $\frac{1}{n} \sum_{i=1}^n (\log(X_i) - \log(t))\mathbf{1}_{\{X_i>t\}}$ converge presque sûrement vers $g(t) = E[(\log(X) - \log(t))\mathbf{1}_{\{X>t\}}]$.

On remplace à nouveau t par $X_{n-k+1,n}$ et on obtient comme estimation de $g(X_{n-k+1,n})$:

$$\frac{1}{n} \sum_{i=1}^n (\log(X_i) - \log(X_{n-k+1,n})) \mathbf{1}_{\{X_i > X_{n-k+1,n}\}} = \frac{1}{n} \left(\sum_{i=n-k+2}^n \log(X_{i,n}) - (k-1) \log(X_{n-k+1,n}) \right)$$

On en déduit que :

$$\frac{1}{(k-1)} \sum_{i=n-k+2}^n \log(X_{i,n}) - \log(X_{n-k+1,n})$$

est un bon candidat pour l'estimation de γ .

Théorème 2.6.1 Soit $(X_n; n \geq 1)$ une suite des variables aléatoires indépendantes de loi $F \in D(H_\gamma)$, où $\gamma > 0$. Si $\lim_{n \rightarrow \infty} k_n = +\infty$, et $\lim_{n \rightarrow \infty} \frac{k_n}{n} = 0$ alors l'estimateur de Hill (1975, [34]) définit par :

$$\hat{\gamma}_{k,n}^{(H)} = \frac{1}{k_n} \sum_{i=n-k_n+1}^n \log(X_{i,n}) - \log(X_{n-k_n+1,n}) \quad (2.12)$$

Définition 2.6.1 Soit $(k_n)_{n \geq 1}$ une suite d'entiers avec $1 < k_n = k \leq n$, si $\lim_{n \rightarrow \infty} k_n = +\infty$, et $\lim_{n \rightarrow \infty} \frac{k_n}{n} = 0$ alors on peut définir l'estimateur de Hill (1975, [34]) par :

$$\hat{\gamma}_{k_n,n}^{(H)} = \frac{1}{k_n - 1} \sum_{i=1}^{k_n-1} \log(X_{n-i+1,n}) - \log(X_{n-k_n+1,n}) \quad (2.13)$$

Remarque 2.6.1 Il est d'usage avec l'estimateur de Hill (1975, [34]) de remplacer la somme de $i = 1, \dots, k_n - 1$; par la somme de $i = 0, \dots, k_n - 1$; et $k_n - 1$ par k_n sauf dans le dernier terme, ce qui ne change rien au résultat asymptotique. Nous admettrons le résultat suivant :

$$\hat{\gamma}_{k,n}^{(H)} = \frac{1}{k} \sum_{i=0}^{k-1} \log(X_{n-i,n}) - \log(X_{n-k,n}) \quad (2.14)$$

Pour alléger les notations, notons $k_n = k$.

2.6.1 Normalité asymptotique d'estimateur de Hill(1975)

La normalité asymptotique de l'estimateur de Hill (1975, [34]) est due entre autres à Davis et Resnick (1984, [13]), Csörgö et Mason (1985, [11]), Häusler et Teugels (1985, [32]) et Smith (1987, [49]).

Théorème 2.6.2 *Supposons que la fonction de distribution F a vérifiée la condition de second ordre de subsection (1.5.4), c'est-à-dire :*

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho} \quad (2.15)$$

ou il équivalence a :

$$\lim_{t \rightarrow \infty} \frac{\frac{1-F(tx)}{1-F(t)} - x^{-\frac{1}{\gamma}}}{A\left(\frac{1}{1-F(t)}\right)} = x^{-\frac{1}{\gamma}} \frac{x^\rho - 1}{\gamma\rho}$$

où $\gamma > 0, \rho \leq 0$; et A est une fonction positive ou négative avec $\lim_{t \rightarrow \infty} A(t) = 0$ alors :

$$\sqrt{k} (\hat{\gamma}_k^H - \gamma) \rightarrow^d N\left(\frac{\lambda}{1-\rho}, \gamma^2\right)$$

avec N est la loi normal centré réduite, à condition que $k = k_n \rightarrow \infty, \frac{k_n}{n} \rightarrow 0$ quand $n \rightarrow \infty$ et

$$\lim_{n \rightarrow \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda < \infty \quad (2.16)$$

Remarque 2.6.2 *Il est possible que la convergence de $U(tx) = U(t)$ vers x est très rapide pour tout puissance négative de t c'est-à-dire :*

$$\forall x > 0 \text{ et } \forall \alpha > 0 \quad \lim_{t \rightarrow \infty} t^\alpha \left(\frac{U(tx)}{U(t)} - x^\gamma \right) = 0$$

Dans ce cas le résultat de théorème (2.6.2) satisfaisant la biaisé de partie $\frac{\lambda}{(1-\rho)}$ remplace par 0. Alors on a :

$$\sqrt{k} \left(\hat{\gamma}_{k,n}^{(H)} - \gamma \right) \xrightarrow{d} N(0, \gamma^2)$$

Théorème 2.6.3 *Soit X_1, \dots, X_n une suite des variables aléatoires indépendantes et identiquement distribuées avec une fonction de répartition F . Avec $F \in D(\Phi_\gamma), \gamma > 0$. si $n \rightarrow \infty, k = k_n \rightarrow \infty, \frac{k}{n} \rightarrow 0$, alors*

$$\sqrt{k} \left(\hat{\gamma}_{k,n}^{(H)} - \gamma \right) \xrightarrow{d} N(0, \gamma^2)$$

Théorème 2.6.4 *Soit X_1, \dots, X_n une suite des variables aléatoires indépendantes et identiquement distribuées avec une fonction de répartition F . Avec $F \in D(\Phi_\gamma)$, $\gamma > 0$. Quand $n \rightarrow \infty, k = k_n \rightarrow \infty, \sqrt{k}b_{k,n} \rightarrow 0$, alors:*

$$\sqrt{k} \left(\frac{\hat{\gamma}_{k,n}^{(H)}}{\gamma} - 1 \right) \xrightarrow{d} N(0, 1)$$

Ces dernier théorèmes va nous permettre de calculer des intervalles de confiance pour $\hat{\gamma}_k^H$. On ne doit pas interpréter le théorème comme assurant que l'estimateur de *Hill (1975)* est toujours le meilleur. Le théorème (2.6.3) indique seulement que les propriétés statistiques usuelles que l'on demande aux estimateurs sont vérifiées, mais les conditions de régularité de deuxième ordre qu'il faut exiger à \bar{F} , ce qui n'est pas vérifiable en pratique, font qu'il soit nécessaire de faire attention au moment de l'utiliser. En fait, comme on le verra plus tard, parfois l'estimateur de *Pickands* donne des valeurs moins oscillantes. De plus, l'estimateur de *Hill* n'est valable que pour le bassin d'attraction de Fréchet ; dans les autres cas on est contraints d'utiliser les autres estimateurs.

Proposition 2.6.1 *En pratique, le choix du paramètre k_n pose des problèmes. Si l'on trace le diagramme de Hill, c'est-à-dire la fonction $k_n \rightarrow \hat{\gamma}_{k,n}^{(H)}$, on observe une extrême volatilité qui rend difficile l'utilisation de cet estimateur en pratique si l'on n'a aucune indication sur le choix de k_n . De plus, cet estimateur est biaisé.*

Par conséquent, si k_n est petit, on aura fortiori, compte tenu des remarques faites précédemment, une estimation avec un intervalle de confiance large et a contrario, si k_n est grand, on aura un intervalle de confiance plus étroit mais pas centrée sur la vraie valeur.

2.7 Simulation

Générer de N observations de Pareto

```
N<-5000
Gamma<-1
r<-100 #nombre de réplifications
lev<-0.05
nq<-qnorm(1-lev/2)
k1<-numeric(r)
for(j in 1 :r){
  U<-runif(N)
  X<-(1-U)^(-Gamma)
  Y<-sort(X)
  Y1<-numeric(N)
  for(i in 1 :N){
    Y1[i]<-max(Y[i],1)
  }
}
```

K optimal et l'estimateur de Hill

Kopt par Cheng et Peng

```
W<-N/sqrt(log(N))
V<-W/2
w<-floor(W)
v<-floor(V)
```

```

K11<-(1/w)*sum(log(Y1[(N-w+1) :N]))-log(Y1[N-w])
K12<-(1/v)*sum(log(Y1[(N-v+1) :N]))-log(Y1[N-v])
K21<-(1/w)*sum((log(Y1[N : (N-w+1)]))^2)-(2/w)*(log(Y1[N-w]))*sum(log(Y1[N : (N-w+1)])))+(log(Y1[N-w]))^2
K22<-(1/v)*sum((log(Y1[N : (N-v+1)]))^2)-(2/v)*(log(Y1[N-v]))*sum(log(Y1[N : (N-v+1)])))+(log(Y1[N-v]))^2
ro<-abs((1/log(2))*log(abs((K22-2*(K12^2))/(K21-2*(K11^2))))))
d<-(K21-2*(K11^2))*(1+ro)*((sqrt(log(N)))^ro)/(2*(K11^2)^ro)
K<-function(t){if (t>0)
(((1+2*(nq^2))/(3*t*(1+2*ro)))^(1/(1+ro))*(N^(ro/(1+ro))))
else
(((1+2*(nq^2))/(-3*t))^(1/(1+ro))*(N^(ro/(1+ro))))}
k1[j]<-ceiling(K(d))
}
k<-k1[k1>N/100&k1<N/10]
kopt<-floor(mean(k))

## L'estimateur de Hill ##
l<-length(k)
vectgamma<-numeric(l)
sqerr<-numeric(l)
sig<-numeric(l)
a<-numeric(l)
b<-numeric(l)
compt<-numeric(l)

```

```
for(j in 1 :l){
vectgamma[j]<-(1/k[j])*sum(log(Y1[(N-k[j]+1) :N]))-log(Y1[N-k[j]])
sqerr[j]<-(vectgamma[j]-Gamma)^2
a[j]<-vectgamma[j]-(nq*(vectgamma[j]/sqrt(k[j])))
b[j]<-vectgamma[j]+(nq*(vectgamma[j]/sqrt(k[j])))
}
gamma<-mean(vectgamma)

> gamma
[1] 0.9849614

rmse<-sqrt(mean(sqerr))

> rmse
[1] 0.0318162

sd<-sd(vectgamma)

> sd
[1] 0.02827837
```

Conclusion

Dans notre thèse nous avons abordé différents aspects de la théorie des valeurs extrêmes.

Après avoir dans le premier chapitre rappelé certaines des notions clés en théorie des valeurs extrêmes, nous nous sommes intéressés dans le deuxième chapitre aux techniques non paramétriques pour l'estimation de l'indice de queue, et aussi nous nous sommes intéressés à la famille des lois à queue de type Fréchet on peut citer la loi de Pareto. Ce type de lois interviennent dans de nombreuses applications qui utilisent la méthode des excès au-dessus d'un seuil pour estimer l'indice de queue γ . On a présenté trois estimateurs (*estimateur de Pickands*, *estimateur de Hill* et *estimateur de moment*). On étudie la construction de l'estimateur de *Hill (1975)* qui est valable pour $\gamma > 0$. Il est considéré comme le plus simple des estimateurs de l'indice de queue et nous établissons la normalité asymptotique de cet estimateur.

Pour conclure nous signalons que notre thèse n'est qu'un point de départ pour mieux connaître ce monde immense des valeurs extrêmes.

Bibliographie

- [1] Adlouni, S., Bobée, B. et Ouarda, T. B., (2007). Caractérisation des distributions à queue lourde pour l'analyse des crues. Rapport technique no r-929, INRS-ETE, Université du Québec.
- [2] Beirlant, J., Dierckx, G., Y., G. et Matthys, G., (1999). Tail index estimation and an exponential regression model. *Extremes*, 2(2) :177–200.
- [3] Beirlant, J., Dierckx, G. et Guillou, A., (2005). Estimation of the extreme value index and regression on generalized quantile plots. *Annals of Statistics*, 11(6) :949–970.
- [4] Beirlant, J., Vynckier, P. et Teugels, J., (1996). Excess functions and estimation of the extreme value index. *Bernoulli*, 2(4) :293–318.
- [5] Benlagha, N., Grun-Réhomme, M., Vasechko, O., (2009). Les sinistres graves en assurance automobile : Une nouvelle approche par la théorie des valeurs extrêmes. *Revue Modulad*, N°39.
- [6] Bezat, A., Nikeghbali, A., (2000). La théorie des extrêmes et la gestion des risques de marché, ENSAE.
- [7] Cheng, S. and Peng, L., (2001). Confidence Intervals for the Tails Index. *Bernoulli* 7, 751-760.
- [8] Cole S.G., (2001). *An Introduction to Statistical Modeling of Extreme Value*, Springer.
- [9] Cont, R., (2009). La statistique face aux événements rares = statistics dealing with rare events. *Pour la science*, 385 :116–123.
- [10] Csörgő, S., Deheuvels, P. et Mason, D., (1985). Kernel estimates of the tail index of a distribution. *Annals of Statistics*, 13(3) :1050–1077.
- [11] Csörgő, S. et Mason, D., (1985). Central limit theorems for sums of extreme values. *Mathematical Proceedings of the Cambridge Philosophical Society*, 98(3) :547–558.
- [12] Danielsson, J., Dennis, W. J. et de Vries, C. G., (1996). The method of moments ratio estimator for the tail shape parameter. *Communication in Statistics, Theory and Methods*, 4(25) :711–720

-
- [13] Davis, R. et Resnick, S., (1984). Tail estimates motivated by extreme value theory. *Annals of Statistics*, 12(4) :1467–1487.
- [14] Davison, A. C. et Smith, R. L., (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, B*, 52(3) :393–442.
- [15] De Haan, L., Ferreira, A., (2006). *Extreme Value Theory-An Introduction*. Springer, New York.
- [16] de Haan, L., (1990). Fighting the Arch-enemy with mathematics. *Statistica Neerlandica*, 44(2) :45–68.
- [17] de Haan, L. et Peng, L., (1998). Comparison of tail index estimators. *Statistica Neerlandica*, 52(1) :60–70.
- [18] Deheuvels, P., Häeuser, E. et Mason, D., (1988). Almost sure convergence of the Hill estimator. *Mathematical Proceedings of the Cambridge Philosophical Society*, 104(2) : 371–381.
- [19] Dekkers, A., Einmahl, J. H. J. et de Haan, L., (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4) :1833–1855.
- [20] Dekkers, A. et de Haan, L., (1989). On the estimation of the extreme value index and large quantile estimation. *Annals of Statistics*, 17(4) :1795–1832.
- [21] Dekkers, A.L.M., de Haan, L., (1989). On the estimation of the extreme-value index and large quantile estimation. *The annals of statistics* 4, 1833-1855.
- [22] Dekkers, A. et de Haan, L., (1993). Optimal Choice of Sample Fraction in Extreme Value Estimation. *Journal of Multivariate Analysis* 47, 173-195.
- [23] Drees, H., (1996). Refined Pickands estimator of the extreme value index. *Annals of Statistics*, 23(6) :2059–2080.
- [24] Drees, H., and Kaufmann, E., (1998). Selection of the Optimal Sample Fraction in Univariate Extreme Value Estimation. *Stochastic Processes and their applications* 75,149-195.
- [25] Einmahl, J., (1990). The empirical distribution function as a tail estimator. *Statistica Neerlandica*, 44(3) :79–82.
- [26] Embrechts, P., Klüppelberg, C. et Mikosch, T., (1997). *Modelling Extremal Events for Insurance and Finance*. Springer Berlin.
- [27] FALK, M. (1995). On testing the extreme value index via the Pot-method. *Annals of Statistics*, 23(6) :2013–2035.
- [28] Fréchet, M., (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique*, 6 :93–116.
- [29] Fisher, R.A., Tippett, L.H.C., (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, 24, 180-190.

-
- [30] Garrido, M., (2002). Modélisation des événements rares et estimation des quantiles extrêmes, méthodes de sélection de modèles pour les queues de distribution. Thèse de doctorat, Université Grenoble 1.
- [31] Gnedenko, B., (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3) :423–453.
- [32] Haeusler, E. et Teugels, J., (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation. *Annals of Statistics*, 13(2) :743–756.
- [33] Hall, P., (1990). Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Non Parametric Problems. *Journal of Multivariate Analysis* 32, 177-203.
- [34] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5) :1163–1174.
- [35] Jenkinson, A. F., (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *The Quarterly Journal of the Royal Meteorological Society*, 81(384) :158–171..
- [36] Lekina, A., (2010). Estimation non-paramétrique des quantiles extrêmes conditionnels. Université Joseph-Fourier - Grenoble I; Université de Grenoble.
- [37] Longin, F., (1998). Value at Risk : Une nouvelle approche fondée sur les valeurs extrêmes. *Annales d'économie et de statistique* N°52 .
- [38] Mason, D., (1982). Laws of large numbers for sums of extreme values. *Annals of Probability*, 10(3) :754–764.
- [39] Meraghni, D., (2008). Modelling distribution tails. A Thesis Presented For The Degree Of Doctor of Sciences, University Mohamed Kheider, Biskra, Algeria.
- [40] Nesir, A., (2006). A Functional law of the Iterated Logarithm for Kernel-type Estimators of the tail Index. *Journal of Statistical Planning and Inference* 136, 780-802.
- [41] Neves, C., and Fraga Alves, M.I., (2004). Reiss and Thomas' Automatic Selection of the Number of Extremes. *Computational Statistics and Data Analysis* 47, 689-704.
- [42] Peng, L., (1998). Asymptotically unbiased estimators for the extreme-value index. *Statistics and Probability Letters*, 38(2) :107–115.
- [43] Pickands, J., (1975). Statistical Inference Using Extreme Order Statistics. *Annals of statistics* 3, 119-131.
- [44] Resnick, S., (1987). *Extreme Values, regular Variation, and Point Process*. Springer *istical Planning and Inference* 98, 15-27.,New-York.

-
- [45] Rényi, A., (1953). On the theory of order statistics. *Acta Mathematica Hungarica*, 4(3-4) :191-231.
- [46] Rosen, O. et WEISSMAN, I., (1996). Comparison of estimation methods in extreme value theory. *Communication in Statistics-Theory and Methods*, 24(4) :759-773.
- [47] Schultze, J. et STEINEBACH, J., (1996). On least squares estimates of an exponential tail coefficient. *Statistics and Decisions*, 14(3) :353-372.
- [48] Segars, J., (2001). Residuel Estimators. *Journal of StSmith, R. L., (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika*, 72(1) :67-92.
- [49] Smith, R., (1987). Estimating tails of probability distributions. *Annals of Statistics*, 15(3) : 1174-1207
- [50] Smith, R., (1996). Exreme Value Analysis for Insurance Risk. *Statistical Science* 4, 367-393.
- [51] Viharos, L., (1993). Asymptotic distributions of linear combinations of extreme values. *Acta Scientiarum Mathematicarum*, 58 :211-231.
- [52] Viharos, L., (1995). Limit theorems for linear combinations of extreme values with applications to inference about the tail of a distribution. *Acta Scientiarum Mathematicarum*, 60 :761-777.
- [53] von Mise, R., (1936). La distribution de la plus grande de n valeurs. *Revue deMathématique Union Interbalcanique*, 1 :141-160.
- [54] von Mise, R., (1954). La distribution de la plus grande de n valeurs. *American Mathematical Society, RI, USA, vol. II, pp. 271-294.*
- [55] Weissman, I., (1978). Estimation of parameters and large quantiles based on the klargest observations. *Journal of the American Statistical Association*, 73(364) :812-815.
- [56] Weinstein, S. B., (1973). Theory and application of some classical and generalized asymptotic distributions of extreme values. *IEEE Transactions on Information Theory*, 19(2) :148-154.

Résumé

Depuis quelques années, la théorie des valeurs extrêmes a reçu beaucoup d'attention aussi bien sur le plan théorique que sur le plan pratique. Les domaines d'applications sont en effet très variés : hydrologie, météorologie, biologie, ingénierie, finance, assurance, etc. Cette thèse est une contribution à la modélisation statistique des valeurs extrêmes.

Nous avons cherché à estimer le paramètre de queue d'une distribution de type Fréchet par des méthodes non paramétriques .

Mots clés : queue, valeur extrême; indice de la valeur extrême, occurrences, seuil, excès, estimation des paramètres.

Abstract

For many years, Extreme value theory has received much attention on the theoretical level as well as on the practical one. The scope of its applications are indeed very varied and include hydrology, meteorology, biology, engineering, finance, insurance, etc.

We have sought to estimate the tail parameter of a Fréchet distribution by not parametric methods .

Keywords: Tail, extreme value; Index of extreme value, occurrences, threshold, excess, estimation of parameters .

ملخص

في السنوات الأخيرة، تلقت نظرية القيم القصوى إهتماما كبيرا سواء من حيث الجانب النظري كما الجانب التطبيقي. مجالات التطبيق متنوعة جدا في الواقع: الهيدرولوجيا و الأرصاد الجوية، وعلم الأحياء، والهندسة، التمويل، التأمين، والعلوم الاجتماعية، الخ...

في هذه الأطروحة سعينا لتقدير معلمة الذيل ذات التوزيع من نوع فريشت عن طريق أساليب غير وسائطية .

الكلمات المفتاحية: الذيل، القيمة القصوى، مؤشر القيمة القصوى، حوادث عتبة ، وتقدير المعلمة.