

UNIVERSITÉ ABOU BEKR BELKAID TLEMCEM



Faculté de Sciences

Département de Mathématiques

MÉMOIRE DE MASTER

Spécialité : STATISTIQUES ET PROBABILITÉS APPROFONDIES

présenté par

Abdallah Abdallah Hamisi

Soutenu le : 12/ 10/ 2025

Théorie des valeurs extrêmes, estimation et applications

Soutenu devant le jury composé de :

Mme Bensmain Nawel	Maître de Conférences A	Université de Tlemcen	Présidente
Mme Benyahia Ouahiba	Maître de Conférences A	Université de Tlemcen	Examinatrice
Mme Khettab Zahira	Maître de Conférences A	Université de Tlemcen	Examinatrice
Mme Boukhiar Souad	Maître de Conférences B	Université de Tlemcen	Encadrante

Année Universitaire : 2024-2025

Dédicaces

Je dédie ce travail :

À mes parents bien-aimés, pour leur soutien dans ma vie et mes études,
À mes chers frères et mes chères sœurs,
À toute ma famille,

À tous mes amis.

Remerciements

Je tiens tout d'abord à remercier Dieu le tout puissant, qui m'a donné la capacité et la patience d'effectuer ce travail.

Je tiens particulièrement à remercier sincèrement mon encadreur Madame Boukhiar Souad, d'avoir bien assuré la direction de ce mémoire. Merci pour votre gentillesse, votre patience, votre simplicité et vos précieux conseils. Vraiment merci beaucoup pour la deuxième chance que vous m'avez donné pour accomplir ce mémoire.

Je souhaite aussi à remercier également les membres de jury, pour avoir accepté d'examiner et évaluer ce mémoire.

Mes remerciements les plus sincères vont également à l'ensemble des enseignants et des professeurs du département de Mathématiques qui ont contribué à ma formation.

Enfin, j'adresse mes remerciements à mes parents et tous mes amis pour leur encouragement et support pendant la réalisation de ce mémoire.

Table des matières

Dédicaces	i
Remerciements	ii
Table des matières	iv
Table des figures	v
Liste des tableaux	vi
Abréviations et Notations	vii
Introduction	1
1 S.O et Concepts de Base	4
1.1 Statistique d'Ordre	4
1.1.1 Loi du vecteur SO	5
1.1.2 Lois des statistiques extrêmes	6
1.1.3 Loi de la $k^{\text{ième}}$ statistique d'ordre	8
1.2 Les Quantiles	9
1.2.1 Quantiles d'une loi	9
1.2.2 Quantiles empiriques	10
2 Théorie des Valeurs Extrêmes	15
2.1 Introduction	15
2.2 Propriétés probabilistes asymptotiques des valeurs extrêmes . . .	15
2.2.1 Théorème de Fisher-Tippett-Gnedenko	16
2.3 Distribution Généralisée des Valeurs Extrêmes (GEV)	22
2.4 Domaine d'attraction	22
2.4.1 Fonction à variations régulières	24
2.4.2 Caractérisation des domaines d'attraction	26
2.5 Distribution de Paréto Généralisée (GPD)	28
2.5.1 Théorème de Pickands-Balkema-de Haan	29
3 Estimation de l'indice des Valeurs Extrêmes	32
3.1 Introduction	32
3.2 Estimation Paramétrique	32
3.2.1 Méthode du maximum de vraisemblance	32
3.3 Estimation par les méthodes semi-paramétrique	35

3.3.1	Estimateur de Hill	35
3.3.2	Estimateur des Moments	39
3.3.3	Estimateur de Pickands	40
4	Étude Numérique	45
4.1	Généralités	45
4.2	Simulation par la methode du Maximum Vraisemblance (MV) . .	46
4.3	Simulation par la méthode des Moments (EM)	47
4.4	Comparaison et conclusion	48
	Conclusion Générale	49

Table des figures

1.1	Comparaison de l'ecdf F_n et la fdr F	11
1.2	Comparaison des quantiles et Q-Q plot.	12
1.3	Convergence uniforme - Théorème de Glivenko-Cantelli	13
2.1	Distribution de lois des valeurs extrêmes : Fréchet, Gumbel et Weibull.	21
2.2	Densités des trois types des distributions des valeurs extrêmes :Fréchet($\gamma = 1$), Gumbel($\gamma = 0$) et Weibull($\gamma = -1$)	23
2.3	Les dépassements	29
2.4	Comparaison des densités GPD pour différentes valeurs γ	31
3.1	Estimateur de Hill et l'intervalle de confiance de niveau 95% de l'indice des valeurs extrêmes pour la loi de Pareto Standard ($\gamma = 1$) basé sur 100 échantillons de 3000 observations.	39
3.2	Estimateur de Moment, en fonction du nombre des extrêmes (en trait plein) avec l'intervalle de confiance de niveau 95%, pour la distribution de Burr(1,1,1) basée sur 100 échantillons de 3000 observations.	41
3.3	Estimateur de Pickands, en fonction du nombre des extrêmes avec l'intervalle de confiance 95%, pour l'indice de valeurs extrêmes de la distribution uniforme avec $\gamma = 1$ basé sur 100 échantillons de 3000 observations.	44

Liste des tableaux

2.4.1 Domaines d'attraction et quelques lois associées	28
4.2.1 Estimations de μ et δ et leur biais absolu pour $\mu_0 = 0.25$ et $\delta_0 = 0.15$ par la méthode de MV.	46
4.2.2 Estimations de μ et δ et leur erreur quadratique moyenne pour $\mu_0 = 0.25$ et $\delta_0 = 0.15$ par la méthode de MV.	46
4.2.3 Estimations de μ et δ et leur biais absolu pour $\mu_0 = 0.5$ et $\delta_0 = 0.75$ par la méthode de MV.	46
4.2.4 Estimations de μ et δ et leur erreur quadratique moyenne pour $\mu_0 = 0.5$ et $\delta_0 = 0.75$ par la méthode de MV.	47
4.3.1 Estimations de μ et δ et leur biais absolu pour $\mu_0 = 0.25$ et $\delta_0 = 0.15$ par la méthode des moments.	47
4.3.2 Estimations de μ et δ et calcul du biais absolu pour $\mu_0 = 0.5$ et $\delta_0 = 0.75$ par la méthode des moments.	47
4.3.3 Estimations de μ et δ et le calcul des erreurs quadratiques moyennes pour $\mu_0 = 0.5$ et $\delta_0 = 0.75$ par la méthode des moments.	48
4.4.1 Erreur quadratique moyenne de μ , $MSE(\mu)$ par les deux estimateurs, MV et EM pour $\mu_0 = 0.12$ et $\delta_0 = 0.45$	48
4.4.2 Erreur quadratique moyenne de δ , $MSE(\delta)$ par les deux estimateurs, MV et EM pour $\mu_0 = 0.12$ et $\delta_0 = 0.45$	48

Abréviations et Notations

Les différentes abréviations et notations utilisées dans ce memoire sont expliquées ci-dessous.

TVE	Théorie des valeurs extrêmes.
EVT	Extreme Values Theory.
SO	Statistiques d'ordre.
iid	Indépendantes et identiquement distribuées.
ddp	Densité de probabilité.
fdr	Fonction de répartition.
v.a	Variable aléatoire.
p.s	Presque sûre.
ecdf	Empirical cumulative distribution function.
GEV	Distribution généralisée des valeurs extrêmes.
GPD	Distribution de Paréto généralisée.
POT	Excès au-delà d'un seuil (Peak over threshold).
MSE	Erreur quadratique moyenne (Mean squared error).
abs.bias	Absolute bias.
MV	Maximum de vraisemblance.
EM	Estimateur des moments.
al.	Autres.
$\mathbb{R}$	Ensemble des valeurs réelles.
$\mathbb{N}$	Ensemble des valeurs entières.
f	Densité de probabilité d'une variable aléatoire.
F	Fonction de répartition d'une variable aléatoire.
$X^{(\cdot)}$	Vecteur ordonné SO.
Q	Fonction de quantile.
$X_1, \dots, X_n$	Échantillon de taille n .
$X_{(1)}$	Minimum de $X_1, \dots, X_n$.
$X_{(n)}$	Maximum de $X_1, \dots, X_n$.
$X_{(k)}$	$k^{\text{ème}}$ statistique d'ordre.
$\mathbf{1}_A$	Fonction indicatrice de l'ensemble A .
$\inf A$	Infimum de l'ensemble A .
Q_p	Quantile d'ordre p .
F_n	Fonction de répartition empirique.
Q_n	Quantile empirique.
$Q_{n,p}$	Quantile empirique d'ordre p .
x_F	Point terminal de F
$F^{\leftarrow}$	Inverse généralisée de F .

U	Fonction quantile de queue.
$\lfloor a \rfloor$	Partie entière de a
$ \cdot $	Valeur absolue.
u	Seuil.
σ^2	Variance d'une variable aléatoire.
$\mathcal{RV}_\beta$	Ensemble des fonctions à variations régulières d'indice β .
L	Fonction à variations lente.
H_γ	Famille de la lois de valeurs extrêmes généralisée.
D	Domaine d'attraction.
γ	Indice des valeurs extrêmes.
$\overline{F}$	Fonction de survie.
Ψ	Distribution de Weibull.
Φ	Distribution de Fréchet.
Λ	Distribution de Gumbel.
$\hat{\gamma}_{k,n}^H$	Estimateur de Hill.
$\hat{\gamma}_{k,n}^P$	Estimateur de Pickands.
$\hat{\gamma}_n^M$	Estimateur des moments pour l'indice γ .

Introduction Générale

La théorie des valeurs extrêmes (TVE) ou "Extreme Values Theory (EVT)" en anglais est une approche statistique largement utilisée dans divers domaines d'applications : en finance pour évaluer le risque du marché, risque de crédit ou risque opérationnels là où les méthodes classiques échouent (Les extrêmes sont des événements rares conduisant à des pertes financières remarquables). En météorologie pour analyser les températures extrêmes, vitesse des vents intenses ainsi que les précipitations fortes (Les catastrophes naturelles conduisent à des pertes financières importantes). En épidémiologie pour modéliser les pics épidémiques pouvant contribuer à la planification des solutions face aux crises sévères.

La question clé qui se pose est la suivante : Admettant qu'un événement extrême va survenir, quelles sont les pertes financières maximales possibles ?

Il faudrait donc modéliser des événements rares et anticiper des situations imprévisibles. La TVE offre une base probabiliste nécessaire à l'élaboration des modèles statistiques pour prévoir la taille et la fréquence de ces phénomènes rares.

La TVE se focalise sur l'étude des événements rares (outliers) en se servant des lois de probabilités spécifiques. Elle permet d'estimer la probabilité qu'un événement rare (phénomènes dans les queues de distribution) se reproduise, ce qui est primordial dans la gestion des risques.

Sous des conditions appropriées, Fisher et Tippet (1928) [9], ont pu montrer que les seules distributions limites des extrêmes sont les lois de Fréchet, Gumbel et Weibull.

Dans la suite de ce mémoire, nous nous concentrerons sur l'étude de la loi du maximum. Cette loi étant difficile à déterminer, nous privilégierons l'analyse du comportement asymptotique du maximum afin de déterminer les lois limites non-dégénérées associées.

Fisher et Tippet [9] ont suggéré que, pour un seul processus le comportement des maxima peut être décrit par les trois distributions des valeurs extrêmes – Gumbel, Fréchet et Weibull. Gnedenko (1943) [10] a établi une extension des travaux de Fisher et Tippet s'appellant "le théorème de Fisher-Tippet-Gnedenko" connu sous le nom de "théorème de la valeur extrême".

L'étude et la caractérisation du comportement du maximum reposent essentiellement sur deux approches de modélisation : **Approche des maxima par**

blocs, qui consiste à modéliser l'échantillon des maxima en utilisant la Distribution Généralisée des Valeurs Extrêmes (GEV : Generalized Extreme Value), on extrait les valeurs maximales de blocs de données consécutifs et non superposés (par exemple, la pluviosité maximale quotidienne sur chaque année). La distribution GEV est ensuite ajustée à ces maxima par blocs.

La seconde est l'**Approche du dépassement du seuil** (POT : Peaks Over Threshold), quant à elle, se concentre sur la modélisation de la distribution des valeurs qui dépassent un certain seuil élevé. On ajuste dans ce cas la Distribution de Pareto Généralisée (GPD) aux "dépassements". Cette approche est souvent privilégiée lorsqu'il y a de nombreux événements extrêmes ou si l'on souhaite analyser des événements au-delà d'un niveau particulier, et non seulement les maxima par blocs.

Divers travaux ont été consacrés à l'estimation de l'indice des extrêmes dont l'objectif revient à construire des estimateurs et étudier leurs propriétés, voir Hill [11] dans le cas où l'indice est positif, et Pickands [13] dans le cas général. Dekkers et al.(1989) ont généralisé l'estimateur de Hill, dénommé l'estimateur des Moments. Beirlant et al. (1996) ont présenté à leur tour, l'estimateur de l'indice des valeurs extrêmes généré à partir de l'estimateur de Hill.

Nous pouvons citer d'autres références existant dans la littérature concernant la théorie des valeurs extrêmes, voir Embrechts et al. [8], De Haan et Ferreira [4], ?].

Ce mémoire est constitué de quatre chapitres :

Dans le premier chapitre, nous commençons par revoir quelques concepts fondamentaux sur la statistique d'ordre associée, essentiels pour développer la théorie des valeurs extrêmes. Nous explorons les distributions théoriques associées aux statistiques d'ordre notamment les distributions du maximum et du minimum. Nous rappelons ensuite les notions de base des quantiles, quantiles empiriques et quantiles extrêmes. Enfin, nous terminons par un rappel de quelques définitions utiles dans la suite de notre étude.

Dans le deuxième chapitre, nous commençons par introduire la TVE, nous étudierons ensuite le comportement asymptotique du maximum. Nous présentons le théorème de la valeur extrême, fondamental puisqu'il révèle la forme des lois limites non-dégénérées pour le maximum. Nous présentons la distribution Généralisée des Valeurs Extrêmes (GEV), et donnons les caractérisation des domaines d'attraction de Gumbel, Fréchet et Weibull. Nous concluons ce chapitre par une étude de la Distribution de Pareto généralisée (GDP).

Dans le troisième chapitre, nous commençons l'estimation de l'indice des valeurs extrêmes par l'approche paramétrique dans laquelle la méthode du maximum de vraisemblance est utilisée. Nous présentons l'approche d'estimation semi-paramétrique dans laquelle trois estimateurs de Hill, Pickands et Moments sont considérés. Nous explorons l'estimateur de Hill qui est limité dans le cas où l'indice est positif. Cependant l'estimateur des Moments constituant une extension de l'estimateur de Hill est utile quelque soit le signe de l'indice. Enfin, nous terminons ce chapitre en explorant l'estimateur de Pickands qui offre une

alternative importante aux deux estimateurs précédents.

Dans le dernier chapitre, nous présentons une étude de simulation pour différentes tailles finies (n) d'échantillons, afin d'étudier le comportement de nos estimateurs et d'illustrer et de comparer leurs performances.

Chapitre 1

Statistique d'Ordre et Concepts de Base

Les statistiques d'ordre (S.O) représentent un outil essentiel pour l'analyse des données. Elles constituent un fondement de la théorie des valeurs extrêmes (TVE : outliers en Anglais), développée par Fisher, Tippet et Gnedenko qui s'intéressaient au comportement asymptotique lorsque la taille de l'échantillon devient assez grande.

Les valeurs extrêmes représentent les observations qui s'écartent significativement de l'ensemble des données dans un échantillon donné. Ainsi, la théorie des valeurs extrêmes (TVE) devient pertinente, puisqu'elle offre la possibilité d'inférer le comportement probabiliste des queues de distribution en se reposant sur les valeurs les plus élevées observées.

Les applications de cette théorie sont divers, allant de la modélisation des phénomènes météorologiques (pluviosité, tempêtes) à l'analyse des risques financiers, en passant par la démographie, pour prévoir la distribution de probabilité de l'âge maximum que l'être humain pourra atteindre, en finance ou encore en hydrologie. Les statistiques d'ordre, quant à elles, permettent de construire des estimateurs efficaces et des intervalles de confiance non-paramétriques, illustrant ainsi la richesse et la diversité des applications de ces concepts en statistique moderne.

1.1 Statistique d'Ordre

Une réalisation d'une S.O est obtenue en classant dans l'ordre croissant les réalisations de l'échantillon de v.a indépendantes et identiquement distribuées. Nous commençons par les définitions fondamentales et quelques propriétés nécessaires sur les SO, avant d'aborder leurs différentes distributions théoriques.

Dans la suite, afin d'alléger les preuves, nous supposerons que la variable aléatoire (v.a) X est absolument continue. Tirons un n -échantillon $(X_1, X_2, \dots, X_n)$ issu de la v.a X de composantes aléatoires réelles $(X_i)_i$ indépendantes et identiquement distribuées (iid) de densité de probabilité (ddp) f et d'une fonction

de répartition (fdr) F .

Définition 1.1 (statistique d'ordre).

On appelle statistique d'ordre associée à l'échantillon $(X_1, X_2, \dots, X_n)$, la suite croissante notée $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ telle que :

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

où $X_{(i)}$ est la $i^{\text{ème}}$ statistique d'ordre, $(1 \leq i \leq n)$.

Théorème 1.1.

Soit $X = (X_1, X_2, \dots, X_n)$ un échantillon de loi F . Si

F est une loi continue alors $(X_i)_{1 \leq i \leq n}$ sont presque sûrement distinctes.

$$i.e; \quad \forall (i, j) \in \{\overline{1, n}\}^2 / i \neq j, \quad \mathbb{P}(X_i = X_j) = 0 \quad p.s$$

Démonstration.

Supposons que F est une loi continue de densité de probabilité f , pour $(i, j) \in \{\overline{1, n}\}^2 / i \neq j$, Nous avons

$$\mathbb{P}(X_i = X_j) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{1}_{\{x\}}(y) f(x) f(y) dx dy.$$

En utilisant le théorème de Fubini, nous obtenons

$$\mathbb{P}(X_i = X_j) = \int_{\mathbb{R}} f(x) \left\{ \int_{\{x\}} f(y) dy \right\} dx.$$

Il est à noter que la deuxième intégrale est nulle dans la mesure où le singleton $\{x\}$ est de mesure de Lebesgue nulle. Par suite

$$\mathbb{P}(\exists (i, j) \in \{\overline{1, n}\}^2 / i \neq j, X_i = X_j) = \mathbb{P}\left(\bigcup_{i \neq j, i, j=1}^n X_i = X_j\right) \leq \sum_{i \neq j} \mathbb{P}(X_i = X_j) = 0,$$

ce qui nous ramène au résultat.

Pour l'étude des événements extrêmes, nous nous intéressons aux deux statistiques d'ordres suivantes

- $X_{(1)} := \min(X_1, X_2, \dots, X_n), n \geq 1$: La plus petite statistique d'ordre ou encore statistique du minimum notée $\min_{1 \leq i \leq n} (X_i)$.
- $X_{(n)} := \max(X_1, X_2, \dots, X_n), n \geq n$: La plus grande statistique d'ordre, ou bien statistique du maximum notée $\max_{1 \leq i \leq n} (X_i)$.

Dans la suite de ce mémoire, nous nous focaliserons sur l'analyse du comportement du maximum.

1.1.1 Loi du vecteur SO

Le résultat suivant donne la loi du vecteur ordonné SO.

Lemme 1.1.

Soit $X_1, X_2, \dots, X_n$, n variables aléatoires indépendantes et identiquement distribuée (i.i.d) de densité f . Alors la densité jointe des n statistiques d'ordre $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ est donnée par :

$$f_{X_{(1)}, X_{(2)}, \dots, X_{(n)}}(t_1, \dots, t_n) = n! \prod_{i=1}^n f(t_i) \mathbf{1}_D(t_1, \dots, t_n) \quad (1.1)$$

où $D = ((t_1, \dots, t_n) \in \mathbb{R}^n \mid t_1 < t_2 < \dots < t_n)$.

Démonstration.

Posons $X^{(\cdot)} = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$, $t = (t_1, \dots, t_n)$ et $f^{(\cdot)}(t) = f_{X_{(1)}, X_{(2)}, \dots, X_{(n)}}(t_1, \dots, t_n)$. Notons que la densité du SO n'est que la dérivée de la fdr associée, autrement dit

$$f^{(\cdot)}(t) = \lim_{\Delta t \rightarrow 0} \frac{F^{(\cdot)}(t + \Delta t) - F^{(\cdot)}(t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq X^{(\cdot)} < t + \Delta t)}{\Delta t}.$$

Ainsi,

$$\begin{aligned} f^{(\cdot)}(t) &= \lim_{\Delta t_i \rightarrow 0, i \in \{1, \dots, n\}} \frac{\mathbb{P}(t_1 \leq X_{(1)} < t_1 + \Delta t_1, t_2 \leq X_{(2)} < t_2 + \Delta t_2, \dots, t_n \leq X_{(n)} < t_n + \Delta t_n)}{\Delta t_1 \Delta t_2 \dots \Delta t_n} \\ &= \lim_{\Delta t_i \rightarrow 0, i \in \{1, \dots, n\}} \frac{\mathbb{P}(t_1 \leq X_1 < t_1 + \Delta t_1, t_2 \leq X_2 < t_2 + \Delta t_2, \dots, t_n \leq X_n < t_n + \Delta t_n)}{\Delta t_1 \Delta t_2 \dots \Delta t_n}. \end{aligned}$$

Lorsque Δt_i tend vers 0, et pour tout $-\infty < t_1 < t_2 < \dots < t_n < +\infty$, et le fait que, pour n valeurs distinctes, il y a $n!$ possibilités pour les statistiques d'ordre de prendre les valeurs. Nous obtenons

$$f^{(\cdot)}(t) = n! f(t_1) \dots f(t_n) = n! \prod_{i=1}^n f(t_i) \quad \text{pour tout } -\infty < t_1 < t_2 < \dots < t_n < +\infty.$$

Exemple 1.1 (Exemple d'Application).

Considérons $X_1, X_2, \dots, X_n$ des v.a iid de fonction de loi uniforme $\mathbb{U}_{[0,1]}$. La densité du vecteur aléatoire $X^{(\cdot)} = (X_{(1)}, X_{(2)}, \dots, X_{(n)})$ est donnée par

$$f^{(\cdot)}(t_1, \dots, t_n) = n! \mathbf{1}_{\{0 < t_1 < t_2 < \dots < t_n < 1\}}, \quad (1.2)$$

qui est la ddp d'une loi uniforme spatiale sur le simplexe

$$S_n = \left\{ (t_1, \dots, t_n) \mid 0 \leq t_i \leq 1, \sum_{i=1}^n t_i = 1, t_1 < t_2 < \dots < t_n \right\},$$

appelée aussi la loi de Dirichlet que l'on note $\mathbb{D}(1, \dots, 1; 1)$.

1.1.2 Lois des statistiques extrêmes

Considérons un n -échantillon $(X_1, X_2, \dots, X_n)$, nous nous intéressons à la loi du minimum et du maximum de l'échantillon.

La fdr du minimum $X_{(1)}$, est donnée pour tout $u \in \mathbb{R}$, par :

$$\begin{aligned} F_{(1)}(u) &= P(X_{(1)} \leq u) \\ &= P(\min(X_1, X_2, \dots, X_n) \leq u) \\ &= 1 - P(\min(X_1, X_2, \dots, X_n) > u) \\ &= 1 - P(X_1 > u, X_2 > u, \dots, X_n > u). \end{aligned}$$

Comme les $X_1, X_2, \dots, X_n$ sont indépendantes et de même loi, nous aurons :

$$\begin{aligned} F_{(1)}(u) &= 1 - P\left(\bigcap_{i=1}^n (X_i > u)\right) \\ &= 1 - \prod_{i=1}^n P(X_i > u) \\ &= 1 - [P(X_i > u)]^n, \quad (1 \leq i \leq n) \\ &= 1 - [1 - P(X_i \leq u)]^n, \quad (1 \leq i \leq n). \end{aligned}$$

Par suite

$$F_{(1)}(u) = 1 - [1 - F(u)]^n, \quad (1.3)$$

où F étant la fdr des v.a X_i ($1 \leq i \leq n$).

Pour la ddp associée à $X_{(1)}$, elle sera donnée par :

$$\begin{aligned} f_{(1)}(u) &= \frac{d}{du} F_{X_{(1)}}(u) \\ &= n \frac{d}{du} F(u) [1 - F(u)]^{n-1}, \quad n \geq 1, \end{aligned}$$

Par conséquent

$$f_{(1)}(u) = n f(u) (1 - F(u))^{n-1}, \quad n \geq 1, u \in R. \quad (1.4)$$

Nous procédons de la même façon pour conclure la fdr du maximum $X_{(n)}$. Pour $u \in \mathbb{R}$, nous avons :

$$\begin{aligned} F_{(n)}(u) &= P(X_{(n)} \leq u) \\ &= P(\max(X_1, X_2, \dots, X_n) \leq u) \\ &= P(X_1 \leq u, X_2 \leq u, \dots, X_n \leq u), \end{aligned}$$

or les v.a $X_1, X_2, \dots, X_n$ sont iid, nous obtenons donc :

$$\begin{aligned} F_{(n)}(u) &= P\left(\bigcap_{i=1}^n (X_i \leq u)\right) \\ &= \prod_{i=1}^n P(X_i \leq u) \\ &= [P(X_i \leq u)]^n, \quad (1 \leq i \leq n), \end{aligned}$$

alors

$$F_{(n)}(u) = [F(u)]^n \quad (1.5)$$

La densité correspondante sera définie dans ce cas par :

$$\begin{aligned} f_{(n)}(u) &= \frac{d}{du} F_{(n)}(u) \\ &= n \frac{d}{du} F(u) [F(u)]^{n-1}, \quad n \geq 1, \end{aligned}$$

et donc

$$f_{(n)}(u) = n f(u) (F(u))^{n-1}. \quad (1.6)$$

Remarque 1.1.

1. *Contrairement aux v.a initiales $(X_i, 1 \leq i \leq n)$ qui sont iid, l'échantillon ordonné $(X_{(i)}, 1 \leq i \leq n)$, les composantes aléatoires $X_{(i)}$, $(1 \leq i \leq n)$ ne sont plus (i.i.d). En effet :
pour $i \neq j$, $X_{i,n}$ et $X_{j,n}$ sont dépendants l'une de l'autre, c'est-à-dire que si $i < j$ alors $X_{(i)} < X_{(j)}$, $(1 \leq i, j \leq n)$.
De plus, d'après ce qui a été établi plus haut, nous constatons que effectivement, (voir (1.3), (1.4), (1.5), et (1.6)), $F_{(1)} \neq F_{(n)}$ et que $f_{(1)} \neq f_{(n)}$ ce qui valide que $\mathcal{L}(X_{(1)}) \neq \mathcal{L}(X_{(n)})$.*
2. *Les résultats relatifs au minimum se déduisant à partir de ceux du maximum par la relation :*

$$\min(X_1, X_2, \dots, X_n) = -\max(-X_1, -X_2, \dots, -X_n) \quad (1.7)$$

1.1.3 Loi de la $k^{\text{ième}}$ statistique d'ordre

Proposition 1.1.

Soit $(1 \leq k \leq n)$; la fonction de répartition de la $k^{\text{ème}}$ statistique d'ordre $X_{k,n}$ est donnée par :

$$F_{(k)}(x) = \sum_{i=k}^n C_n^i [F(x)]^i [1 - F(x)]^{n-i}, \quad -\infty < x < +\infty \quad (1.8)$$

La densité associée est donnée par :

$$f_{(k)}(x) = \frac{n!}{(n-k)!(k-1)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x), \quad -\infty < x < +\infty \quad (1.9)$$

Démonstration.

Pour une preuve détaillée de la proposition, nous pouvons nous référer à Reiss et Thomas [14]. Nous donnons une esquisse de la preuve.

Soit $x \in \mathbb{R}$, la fonction de répartition de la $k^{\text{ème}}$ SO $(X_{(k)}, (1 \leq k \leq n))$ est donnée par

$$\begin{aligned} F_{X_{(k)}}(x) &= \mathbb{P}(X_{(k)} \leq x) \\ &= \mathbb{P}(\text{au moins } k \text{ des } X_i \text{ sont inférieurs à } x) \quad 1 \leq i, k \leq n \\ &= \mathbb{P}\left(\sum_{i=1}^n \mathbf{1}_{(X_i \leq x)}(x) \geq k\right) \quad 1 \leq i, k \leq n. \end{aligned}$$

Puisque les variables aléatoires $(X_i \ (1 \leq i \leq n))$ sont iid, la fonction indicatrice $\mathbf{1}_{(X_i \leq x)}(x)$ est une v.a de loi de Bernoulli de paramètre p avec $p = \mathbb{E}[\mathbf{1}_{\{X_i \leq x\}}(x)] = F(x)$, par conséquent, $\sum_{i=1}^n \mathbf{1}_{(X_i \leq x)}(x)$ est une v.a de loi binomiale de paramètres n et $F(x)$, alors

$$\begin{aligned} F_{X_{(k)}}(x) &= \mathbb{P}\left(\sum_{i=1}^n \mathbf{1}_{(X_i \leq x)}(x) \geq k\right) \quad 1 \leq k \leq n \\ &= \sum_{i=k}^n C_n^i [F(x)]^i [1 - F(x)]^{n-i}, \quad \forall x \in \mathbb{R}. \end{aligned}$$

La densité de la $k^{\text{ème}}$ SO $(X_{(k)}, \ (1 \leq k \leq n))$ est obtenue par dérivation de la fonction de répartition de $X_{(k)}$ par rapport à x , nous avons

$$\begin{aligned} f_{X_{(k)}}(x) &= \frac{d}{dx} (F_{X_{(k)}}(x)) \\ &= \frac{d}{dx} \left(\sum_{i=k}^n C_n^i [F(x)]^i [1 - F(x)]^{n-i} \right) \\ &= \sum_{i=k}^n C_n^i (iF'(x)[F(x)]^{i-1}[1 - F(x)]^{n-i} + [F(x)]^i(n-i)(-F'(x))[1 - F(x)]^{n-i-1}) \\ &= F'(x) \sum_{i=k}^n iC_n^i [F(x)]^{i-1}[1 - F(x)]^{n-i} - F'(x) \sum_{i=k}^n (n-i)C_n^i [F(x)]^i [1 - F(x)]^{n-i-1}. \end{aligned}$$

Rappelons que $iC_n^i = nC_{n-1}^{i-1}$ et $(n-i)C_n^i = nC_{n-1}^i$, nous arrivons à

$$f_{X_{(k)}}(x) = n f(x) \sum_{i=k}^n C_{n-1}^{i-1} [F(x)]^{i-1} [1 - F(x)]^{n-i} - n f(x) \sum_{i=k}^n C_{n-1}^i [F(x)]^i [1 - F(x)]^{n-i-1}.$$

En effectuant un changement de variable $j = i - 1$ dans la première somme et en arrangeant les calculs, nous avons

$$\begin{aligned} f_{X_{(k)}}(x) &= n f(x) C_{n-1}^{k-1} [F(x)]^{k-1} [1 - F(x)]^{n-k} \\ &= \frac{n!}{(n-k)!(k-1)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x). \end{aligned}$$

1.2 Les Quantiles

Les quantiles sont les valeurs qui divisent un jeu de données en intervalles de même probabilité. Dans notre cas, ils sont utilisés dans l'analyse des valeurs extrêmes.

1.2.1 Quantiles d'une loi

Définition 1.2 (Quantile d'ordre p).

Soit X une v.a de fdr F . Pour tout $p \in (0, 1)$, le quantile théorique d'ordre p

noté Q_p (appelé aussi p -quantile) est donnée par

$$Q_p = \inf\{x \in \mathbb{R} : F(x) > p\}.$$

Notons que $F(Q_p^-) \leq p \leq F(Q_p)$.

Si F est strictement croissante, alors Q_p est la solution unique de la fonction quantile : $F(Q_p) = p$, ou encore $Q_p = F^{-1}(p)$.

Définition 1.3 (Quatiles et Médiane).

La médiane (Me) est la deuxième quartile (Q_2) qui partage la population en deux sous-groupes de pourcentages égaux à 50%. Autrement dit, $\mathbb{P}(X > Me) = \mathbb{P}(X \leq Me) = 0.5$.

Si $p = \frac{1}{4}$, on a $Q_{\frac{1}{4}} = Q_1$ (Premier quartile) et pour $p = \frac{3}{4}$, on a $Q_{\frac{3}{4}} = Q_3$ (Troisième quartile). Les quartiles Q_1 et Q_3 partagent la population en deux sous-groupes de pourcentages égaux respectivement à 25% et 75%, ou encore $\mathbb{P}(X < Q_{\frac{1}{4}}) = 0.25$ et $\mathbb{P}(X \geq Q_{\frac{1}{4}}) = 0.75$.

En générale, les quartiles Q_1, Q_2 et Q_3 partagent la population en quatre sous-groupes de mêmes poids, i.e

$$\mathbb{P}(X < Q_1) = \mathbb{P}(Q_1 \leq X < Me) = \mathbb{P}(Me \leq X < Q_3) = 0.25.$$

1.2.2 Quantiles empiriques

Définition 1.4 (Fonction de répartition empirique).

Soit le n -($X_1, X_2, \dots, X_n$) de variables aléatoires iid de fonction de répartition F . La fonction de distribution empirique associée (empirical cumulative distribution function "ecdf"), notée F_n est définie, pour tout $x \in \mathbb{R}$, par :

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(X_i \leq x)}(x), \quad (1.10)$$

où $\mathbf{1}$ représente la fonction indicatrice qui vaut 1 si $X_i \leq x$ et 0 sinon.

Notons que F_n est un estimateur sans biais de F (esb). En effet :

$$\begin{aligned} \mathbb{E}[F_n(x)] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\mathbf{1}_{(X_i \leq x)}(x)) \\ &= P(X_i \leq x) \\ &= F(x). \end{aligned}$$

La fonction de répartition empirique prends ses valeurs dans l'ensemble $\{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}$

i.e :

$$F_n(x) = \begin{cases} 0 & \text{si } x < X_{1,n} \\ \frac{1}{n} & \text{si } X_{(1)} \leq x < X_{(2)} \\ \frac{2}{n} & \text{si } X_{(2)} \leq x < X_{(3)} \\ \cdot & \\ \cdot & \\ \cdot & \\ \frac{n-1}{n} & \text{si } X_{(n-1)} \leq x < X_{(n)} \\ 1 & \text{si } X_{(n)} \leq x \end{cases}$$

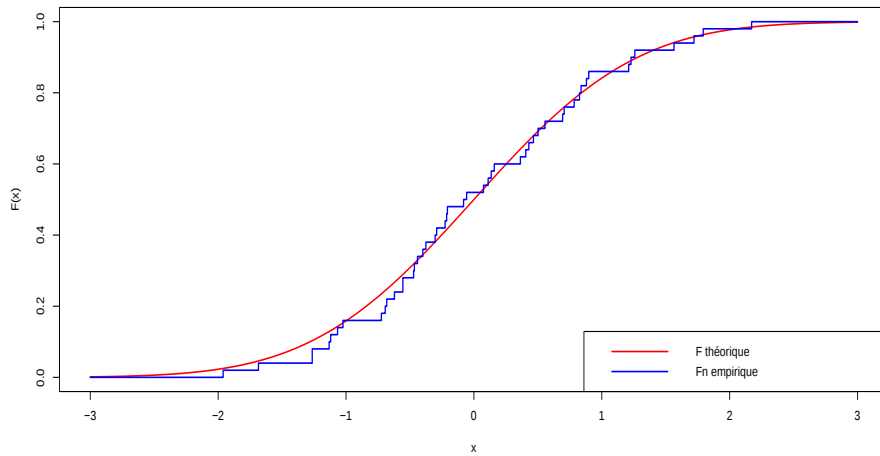


FIGURE 1.1 – Comparaison de l'ecdf F_n et la fdr F .

Définition 1.5 (Quantile Empirique).

Soient $(X_1, X_2, \dots, X_n)$ le n -échantillon de v.a de même loi F . Pour $p \in (0, 1)$, le quantile empirique d'ordre p noté $Q_{n,p}$ associé à la fonction de répartition empirique F_n est donné par :

$$Q_{n,p} = \inf\{x \in \mathbb{R} : F_n(x) \geq p\}. \quad (1.11)$$

En pratique, pour un vecteur SO $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$, le quantile d'ordre p peut être approximé par $X_{(\lfloor np \rfloor + 1)}$ où $\lfloor \cdot \rfloor$ représente la partie entière.

L'analyse graphique des quantiles empiriques constitue une étape cruciale dans l'évaluation de l'ajustement d'un ensemble de données (échantillon) à une distribution théorique. Cette approche, qui repose profondément sur la comparaison systématique entre les quantiles observés (empiriques) et leurs équivalents théoriques, fournit une perspective à la fois descriptive et inférentielle sur la structure probabiliste des données. Il convient à noter que des écarts significatifs mènent à un écart par rapport à la loi de distribution théorique.

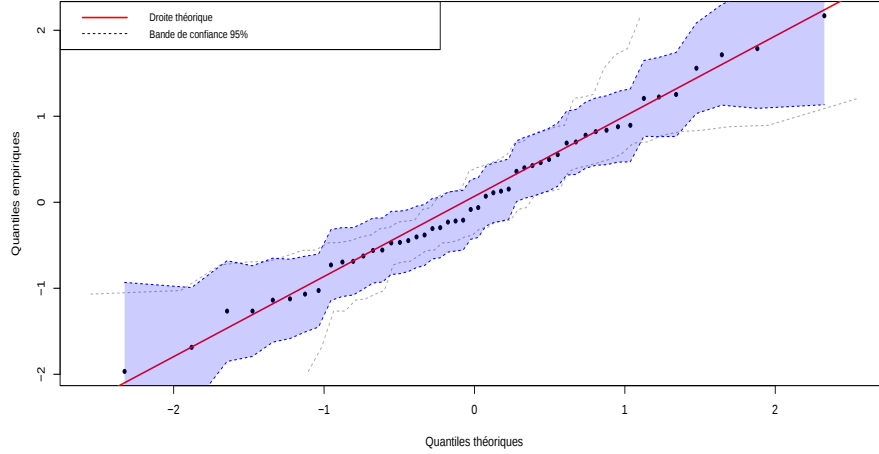


FIGURE 1.2 – Comparaison des quantiles et Q-Q plot.

Cette formulation met en lumière plusieurs propriétés notables : l'erreur standard atteint son minimum pour les quantiles centraux (médiane) et son maximum aux extrêmes, ce qui reflète la plus grande variabilité des statistiques d'ordre extrêmes. En outre, l'erreur standard est inversement proportionnelle à la densité de probabilité au point examiné, ce qui explique pourquoi les queues de distribution, qui se caractérisent par de faibles densités, affichent une plus grande incertitude d'estimation.

Théorème 1.2 (Convergence des quantiles empiriques).

Soit $(X_1, X_2, \dots, X_n)$ un échantillon i.i.d. de loi F . Soit $Q_{n,p}$ le quantile empirique d'ordre p et x_p le quantile théorique d'ordre p associé à F .

1. **Convergence presque sûre** : Pour tout $p \in (0, 1)$, si F possède un unique p -quantile, qui est en fait égal à x_p (c'est-à-dire que F^- est continue en p), alors

$$Q_{n,p} \xrightarrow[n \rightarrow \infty]{p.s.} x_p.$$

2. **Convergence en loi** : Pour tout $p \in (0, 1)$, si la loi admet une ddp f , alors

$$\sqrt{n}(Q_{n,p} - x_p) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_p^2)$$

où

$$\sigma_p^2 = \frac{p(1-p)}{[f(x_p)]^2}.$$

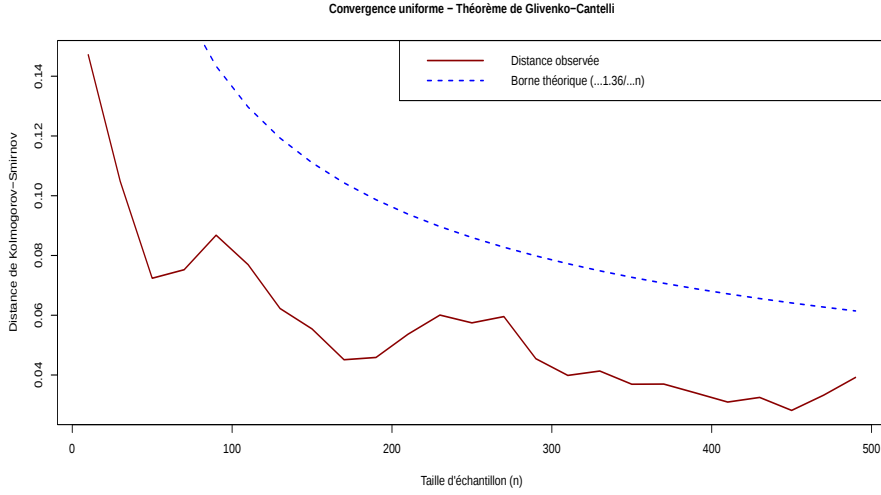


FIGURE 1.3 – Convergence uniforme - Théorème de Glivenko-Cantelli

Démonstration.

La première partie se déduit du théorème de Glivenko-Cantelli. En utilisant l'inégalité triangulaire, et le fait que $F_n(Q_{p,n}) = \frac{[np]+1}{n}$, le deuxième terme dans la première inégalité implique que

$$\begin{aligned} |F(Q_{n,p}) - F(x_p)| &\leq |F(Q_{n,p}) - F_n(Q_{n,p})| + |F_n(Q_{n,p}) - F(x_p)| \\ &\leq \|F_n - F\|_\infty + \left| \frac{[np]+1}{n} - p \right| \xrightarrow[p \rightarrow \infty]{p.s.} 0, \end{aligned}$$

où $[np]$ désigne la partie entière de np . Le terme $\|F_n - F\|_\infty \xrightarrow[p \rightarrow \infty]{p.s.} 0$ par le théorème de Glivenko-Cantelli, et $\left| \frac{[np]+1}{n} - p \right| \rightarrow 0$ puisque $\frac{[np]}{n} \rightarrow p$. Nous utilisons ensuite la continuité de F^- en p pour conclure que $F(Q_{p,n}) \rightarrow F(x_p) = p$ entraîne $Q_{p,n} \rightarrow x_p$ presque sûrement.

Pour la convergence en loi, nous pourrions se référer à Cortel et al. [3] pour un exemple détaillé et à Tassi [15] pour le résultat général. La preuve s'appuie sur un développement de Taylor de F au voisinage du point x_p , aussi sur le théorème de la limite centrale pour les processus empiriques. Quant à la variance limite $\sigma_p^2 = \frac{p(1-p)}{[f(x_p)]^2}$, elle découle de la méthode delta appliquée au quantile empirique.

Remarque 1.2.

1. Si la condition de continuité de F^- en p est remplie, l'unicité du p -quantile est assurée.
2. Pour la valeur $p = 1/2$ (médiane), la variance asymptotique σ_p^2 est minimale et croît vers les extrêmes.

Corollaire 1.1 (Intervalle de confiance asymptotique).

Sous les conditions du Théorème 1.2, pour $\alpha \in (0, 1)$, un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour x_p est de la forme

$$Q_{p,n} \pm q_{\alpha/2} \frac{\sqrt{p(1-p)}}{\sqrt{n}f(x_p)}$$

où $q_{\alpha/2}$ représente le quantile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite.

Nous terminons cette première partie par quelques définitions utiles dans la suite.

Définition 1.6 (Point terminal).

Le point terminal d'une fonction de répartition F est donné par :

$$x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} \leq \infty. \quad (1.12)$$

Définition 1.7 (Inverse Généralisée).

On appelle l'inverse généralisée (ou fonction des quantiles) de la fdr F , l'application notée $F^{\leftarrow}$ définie par :

$$Q(p) = F^{\leftarrow}(p) := \inf\{x \in \mathbb{R} : F(x) \geq p\}, \quad 0 < p < 1. \quad (1.13)$$

Remarque 1.3.

L'inverse généralisée $F^{\leftarrow}$ coïncide avec l'inverse F^{-1} lorsque F est continue et strictement croissante.

Définition 1.8 (Fonction des quantiles empiriques).

On appelle fonction des quantiles empiriques de la fonction F , l'application notée $F_n^{\leftarrow}$ définie pour tout $n \geq 1$ par :

$$Q_n(p) = F_n^{\leftarrow}(p) := \inf\{x \in \mathbb{R} : F_n(x) \geq p\}, \quad 0 < p < 1, \quad (1.14)$$

où F_n est la fonction de répartition empirique de F

Définition 1.9 (Fonction quantile de queue).

On appelle fonction quantile de queue de la distribution F , la fonction $U :]1, +\infty[\rightarrow \mathbb{R}$, définie par :

$$U(t) := Q\left(1 - \frac{1}{t}\right) = F^{\leftarrow}\left(1 - \frac{1}{t}\right), \quad \forall t > 1, \quad (1.15)$$

où Q est la fonction des quantiles associée à F .

Chapitre 2

Théorie des Valeurs Extrêmes

2.1 Introduction

Les valeurs extrêmes représentent les valeurs les plus faibles ou plus élevées dans les suites de variables aléatoires. Elles se définissent par des événements rares dont la probabilité d'apparition est considérablement faible.

Les cracks boursiers et les catastrophes naturelles comme le séisme et le tsunami sont des exemples d'événements extrêmes qui conduisent à des pertes financières importantes pour les assureurs et les réassureurs.

La TVE permet de comprendre le comportement asymptotique des maxima de valeurs prises par les valeurs de variables aléatoires indépendantes et identiquement distribuées. Elle fournit une base mathématique et probabiliste rigoureuse sur laquelle il est possible de développer des modèles statistiques pour prévoir la taille et la fréquence de ces phénomènes rares.

2.2 Propriétés probabilistes asymptotiques des valeurs extrêmes

D'après les calculs établis dans la partie précédente, nous rappelons que pour n v.a $X_1, X_2, \dots, X_n$ iid de loi commune F , la fonction de répartition de leur maximum $X_{(n)}$ est donnée, pour tout réel x , par $F_{(n)}(x) = [F(x)]^n$

D'une manière générale, la loi du maximum est reliée à la loi F . Toutefois, F n'est pas systématiquement connue, et même lorsqu'elle l'est, la détermination de la loi du maximum reste complexe. C'est la raison pour laquelle nous nous focalisons sur l'étude du comportement asymptotique du maximum en faisant tendre n vers l'infini. Ainsi, nous avons :

$$\lim_{n \rightarrow \infty} F_{(n)}(x) = \lim_{n \rightarrow \infty} [F(x)]^n = \begin{cases} 0 & \text{si } F(x) < 1 \\ 1 & \text{si } F(x) = 1 \end{cases} \quad (2.1)$$

Il est à noter que, pour n suffisamment grand, le comportement asymptotique du maximum mènent à une loi dégénérée qui ne prend que les valeurs 0 et 1. Par

suite, nous cherchons à obtenir une loi non dégénérée pour le maximum $X_{(n)}$, c'est-à-dire une loi qui ne se limite pas uniquement aux valeurs 0 et 1.

2.2.1 Théorème de Fisher-Tippett-Gnedenko

Nous présentons un résultat qui a été proposé par Gnedenko en 1943, et qui se base sur les travaux réalisés par Fisher et Tippett en 1928. Ce théorème est connu sous le nom de théorème de Fisher-Tippett-Gnedenko ou bien théorème de la valeur extrême qui donne la forme des lois limites non-dégénérée pour le maximum normalisé.

Le théorème ci-dessous présente une caractérisation de la loi limite du maximum $X_{(n)}$.

Définition 2.1 (Loi de même type).

Soient X et Y deux variables aléatoires réelles de lois respectives F et H . On dit que X et Y sont de même type, s'il existe des constantes $a > 0$ et $b \in \mathbb{R}$ tels que $aX + b \stackrel{L}{=} Y$, ou encore $F(ax + b) = H(x)$.

Théorème 2.1 (Fisher-Tippett-Gnedenko [14]).

Soient $X_1, X_2, \dots, X_n$ une suite de n variables aléatoires iid de loi F . S'il existe deux suites de constantes de normalisation $(a_n)_n > 0$ et $(b_n)_n \in \mathbb{R}$ et une loi non-dégénérée de distribution H telle que

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = H(x), \quad \forall x \in \mathbb{R}, \quad (2.2)$$

où H est la distribution des valeurs extrêmes (Extreme Values Distribution, EVD). Alors H appartient à une des trois distributions standard des valeurs extrêmes suivantes

1. Fréchet : $\Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{si } x > 0 \end{cases} \quad \alpha > 0,$
2. weibull : $\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \quad \alpha < 0,$
3. Gumbel : $\Lambda_0(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R}$

Le Théorème 2.1 repose sur le lemme ci-dessous.

Lemme 2.1.

Soient la suite $(f_n)_{n \in \mathbb{N}}$ de fonctions non-décroissantes et g est une fonction non-décroissante, pour $a < b$ et $x \in [a, b]$ (x est un point de continuité de g), tel que $\lim_{n \rightarrow \infty} f_n(x) = g(x)$. Posons $f_n^\leftarrow$ et $g^\leftarrow$ inverses généralisées de f_n et g respectivement. Alors, pour tout $x \in [g(a), g(b)]$ (point de continuité de $g^\leftarrow$), nous avons

$$\lim_{n \rightarrow \infty} f_n^\leftarrow(x) = g^\leftarrow(x). \quad (2.3)$$

Démonstration.

Pour x un point de continuité de $g^\leftarrow$, et $\varepsilon > 0$ fixé, nous montrons qu'il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$

$$f_n^\leftarrow(x) - \varepsilon \leq g^\leftarrow(x) \leq f_n^\leftarrow(x) + \varepsilon.$$

Traisons l'inégalité droite, la preuve de l'inégalité gauche est similaire. Choisissons $0 < \varepsilon_1 < \varepsilon$ tel que $g^\leftarrow(x) - \varepsilon_1 \in [a, b]$ est un point de continuité de g . Ceci est possible dans la mesure où les points de discontinuité de g forment un ensemble au plus dénombrable, par conséquent, les points de continuité de g sont denses dans $[a, b]$. Puisque $g^\leftarrow$ est continue en x , par définition de l'inverse généralisé, nous avons $g(g^\leftarrow(x)) \geq x$. Aussi, pour ε_1 assez petit, nous obtenons

$$g(g^\leftarrow(x) - \varepsilon_1) < x.$$

Choisissons $\delta < x - g(g^\leftarrow(x) - \varepsilon_1)$. Puisque $g^\leftarrow(x) - \varepsilon_1$ est un point de continuité de g appartenant à $[a, b]$, et que par hypothèse $f_n(g^\leftarrow(x) - \varepsilon_1) \rightarrow g(g^\leftarrow(x) - \varepsilon_1)$, alors $\exists n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$,

$$f_n(g^\leftarrow(x) - \varepsilon_1) < g(g^\leftarrow(x) - \varepsilon_1) + \delta < x \text{ pour } n > n_0.$$

D'après la définition de la fonction inverse généralisé $f_n^\leftarrow$, nous avons

$$g^\leftarrow(x) - \varepsilon_1 < f_n^\leftarrow(x),$$

et par conséquent

$$g^\leftarrow(x) < f_n^\leftarrow(x) + \varepsilon_1.$$

Démonstration (Preuve du Théorème 2.1).

Supposons qu'il existe deux suites de constantes de normalisation $(a_n)_{n \in \mathbb{N}} > 0$ et $(b_n)_{n \in \mathbb{N}} \in \mathbb{R}$ et une loi non-dégénérée de distribution H . Si les hypothèses du théorème sont remplies, pour tout point de continuité x de H , nous avons

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} [F(a_n x + b_n)]^n = H(x).$$

En prenant le logarithme, nous obtenons

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

Rappelons que le développement limité au voisinage de 1 de $\log u \sim u - 1$, dans ce cas

$$\lim_{n \rightarrow \infty} \frac{1}{n[1 - F(a_n x + b_n)]} = -\frac{1}{\log H(x)}. \quad (2.4)$$

Nous désignons par U l'inverse généralisée de $\frac{1}{(1-F)}$ et par D l'inverse généralisée de $\frac{-1}{\log H}$. Notons que puisque H est non-décroissante, D sera également non-décroissante. Posons $u = \frac{-1}{\log H(x)}$ alors $H(x) = e^{-\frac{1}{u}}$ et par conséquent $x = H^\leftarrow(e^{-\frac{1}{u}})$. L'inverse généralisée de $\frac{-1}{\log H}$ sera donc

$$D(x) = H^\leftarrow(e^{-\frac{1}{x}}).$$

Pour l'inverse généralisée de $\frac{1}{(1-F)}$, reprenons les mêmes techniques utilisées ci-dessus. Posons $u = \frac{1}{n[1-F(a_n x + b_n)]}$ alors $F(a_n x + b_n) = 1 - \frac{1}{nu}$ et par conséquent

$$a_n x + b_n = F^{\leftarrow}\left(1 - \frac{1}{nu}\right) := U(nu) \quad (2.5)$$

En substituant nu par x , nous arrivons à :

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

De la relation 2.5, $a_n x + b_n = U(nu)$, alors

$$x = \frac{U(nu) - b_n}{a_n}.$$

En appliquant le Lemme 2.1 à la relation (2.4), nous obtenons

$$\lim_{n \rightarrow \infty} \frac{U(nx) - b_n}{a_n} = D(x).$$

Plus généralement, nous trouverons des conditions sur D telles qu'il existe deux fonctions a et b telles que

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x), \quad t \in \mathbb{R}.$$

Si par exemple 1 est un point de continuité de D , alors nous pouvons nous débarrasser de b en écrivant

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} &= \lim_{t \rightarrow \infty} \left(\frac{U(tx) - b(t)}{a(t)} - \frac{U(t) - b(t)}{a(t)} \right) \\ &= D(x) - D(1) = E(x). \end{aligned}$$

Avec une nouvelle notation $E(x) = D(x) - D(1)$, et $E(1) = D(1) - D(1) = 0$. Pour simplifier, nous supposons que toutes les fonctions impliquées sont différentiables. Nous écrivons alors

$$\frac{U(txy) - U(t)}{a(t)} = \frac{U(txy) - U(ty)}{a(ty)} \cdot \frac{a(ty)}{a(t)} + \frac{U(ty) - U(t)}{a(t)}. \quad (2.6)$$

Lorsque t devient assez grand, nous déduisons que $A(y) = \lim_{t \rightarrow \infty} \frac{a(ty)}{a(t)}$ existe et l'équation (2.6) devient :

$$E(xy) = E(x)A(y) + E(y) \quad \text{et} \quad E(1) = 0. \quad (2.7)$$

Nous notons que

$$A(xy) = \lim_{t \rightarrow \infty} \frac{a(txy)}{a(t)} = \lim_{t \rightarrow \infty} \frac{a(txy)}{a(ty)} \cdot \frac{a(ty)}{a(t)} = A(x)A(y),$$

et que $A(1) = 1$, en dérivant par rapport à x , nous aurons

$$yA'(xy) = A'(x)A(y),$$

en prenant $x = 1$, nous en déduisons que $\frac{A'(y)}{A(y)} = \frac{A'(1)}{y}$ et en résolvant cette équation différentielle avec une condition initiale $A(1) = 1$, nous obtenons alors

$$A(y) = y^\xi \quad \text{pour } \xi \in \mathbb{R}, \quad (2.8)$$

En remplaçant l'équation (2.8) dans (2.7), nous arrivons à

$$E(xy) = E(x)y^\xi + E(y) \quad \text{et} \quad E(1) = 0. \quad (2.9)$$

(a) Si $\xi \neq 0$, en dérivant l'équation (2.9) ci-dessus par rapport à x , alors $yE'(xy) = E'(x)y^\xi$ et en prenant $x = 1$, cela produit

$$E'(y) = E'(1)y^{\xi-1} \quad (2.10)$$

avec $E'(1) > 0$. Sachant que D et par conséquent E sont des fonctions croissantes. En résolvant cette équation différentielle 2.10 avec la condition $E(1) = 0$, alors

$$E(y) = \frac{k}{\xi}(y^\xi - 1), \quad (2.11)$$

pour une certaine constante positive $k = E'(1)$. Ceci implique qu'il existe une constante β telle que pour tout $y \geq 0$,

$$H^\leftarrow(e^{-\frac{1}{y}}) = D(y) = D(1) + E(y) = \frac{ky^\xi}{\xi} + \beta,$$

ce qui implique $e^{-\frac{1}{y}} = H(\frac{ky^\xi}{\xi} + \beta)$. Ainsi, en posant $x = \frac{ky^\xi}{\xi} + \beta$, nous aurons $y = \left(\frac{x-\beta}{\frac{k}{\xi}}\right)^{\frac{1}{\xi}}$ ce qui nous ramène vers

$$H(x) = \exp \left[- \left(\frac{x-\beta}{\frac{k}{\xi}} \right)^{-\frac{1}{\xi}} \right]. \quad (2.12)$$

Ensuite, en faisant la translation et la multiplication par une constante positive, (2.12) devient

$$H_\xi(x) = e^{-(1+\xi x)^{-\frac{1}{\xi}}}, \quad \text{où } 1 + \xi x > 0. \quad (2.13)$$

De plus, en prenant $\xi = -\frac{1}{\alpha}$ with $\alpha > 0$, nous obtenons **la distribution de Weibull** suivante

$$\Psi_\alpha(x) = H_{-\frac{1}{\alpha}}(\alpha(x+1)), \quad x < 0, \quad (2.14)$$

et en prenant $\xi = \frac{1}{\alpha}$ avec $\alpha > 0$, nous aurons **la distribution de Fréchet**

$$\Phi_\alpha(x) = H_{\frac{1}{\alpha}}(\alpha(x-1)), \quad x > 0. \quad (2.15)$$

(b) Si $\xi = 0$, en dérivant l'équation 2.9 par rapport à x , alos $yE'(xy) = E'(x)$ et en prenant $x = 1$, cela produit

$$yE'(y) = E'(1), \quad (2.16)$$

avec $E'(1) > 0$ puisque D et par conséquent E sont des fonctions croissantes. En résolvant cette équation différentielle 2.16 avec la condition $E(1) = 0$, nous obtenons alors

$$E(y) = k \log y. \quad (2.17)$$

pour une certaine constante positive $k = E'(1)$. Enfin, en suivant les mêmes arguments utilisés plus haut, nous arrivons à conclure **la distribution de Gumbel**

$$H(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}. \quad (2.18)$$

Exemple 2.1 (La distribution exponentielle).

Considérons les n v.a $X_1, X_2, \dots, X_n$ iid suivant une loi exponentielle de paramètre λ où $\lambda > 0$, et de fonction de répartition commune $F : x \mapsto [1 - \exp(-\lambda x)] \mathbf{1}_{\mathbb{R}^+}(x)$. Posons $a_n = \frac{1}{\lambda}$ et $b_n = \frac{1}{\lambda} \log(n)$, alors $Y = \frac{X_{(n)} - b_n}{a_n}$ converge en loi vers la loi de Gumbel. En effet :

$$\begin{aligned} F_Y(x) &= P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) = P[\lambda X_{(n)} - \log(n) \leq x] \\ &= P\left[X_{(n)} \leq \frac{x + \log(n)}{\lambda}\right] = \left[P\left(X_1 \leq \frac{x + \log(n)}{\lambda}\right)\right]^n \\ &= \left[F\left(\frac{x + \log(n)}{\lambda}\right)\right]^n = [1 - e^{-(x + \log(n))}]^n, \quad x \in \mathbb{R}^+ \\ &= \left[1 - \frac{e^{-x}}{n}\right]^n \xrightarrow{n \rightarrow \infty} \exp(-\exp(-x)), \quad x \in \mathbb{R}^+. \\ F_Y(x) &\xrightarrow{n \rightarrow \infty} \Lambda(x), \quad x \in \mathbb{R}^+ \end{aligned}$$

Exemple 2.2 (La distribution de Cauchy).

Soient $X_1, X_2, \dots, X_n$ des v.a iid de loi de Cauchy de paramètre 1. La densité associée noté f est donnée par : $x \mapsto \frac{1}{\pi(1+x^2)}$. Nous choisissons $a_n = \frac{n}{\pi}$ et $b_n = 0$. Alors $Y = \frac{X_{(n)} - b_n}{a_n}$ converge en loi vers la loi de Fréchet. En effet :

$$\begin{aligned} F_Y(x) &= P\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) = P\left(\frac{\pi X_{n,n}}{n} \leq x\right) \\ &= P\left(X_{(n)} \leq \frac{nx}{\pi}\right) = \left[P\left(X_1 \leq \frac{nx}{\pi}\right)\right]^n \\ &= \left[1 - P\left(X_1 > \frac{nx}{\pi}\right)\right]^n = \left[1 - \int_{\frac{nx}{\pi}}^{+\infty} \frac{1}{\pi(1+y^2)} dy\right]^n. \end{aligned}$$

Pour $x > 0$,

$$\int_{\frac{nx}{\pi}}^{+\infty} \frac{1}{\pi(1+y^2)} dy = \int_{\frac{nx}{\pi}}^{+\infty} \frac{1}{\pi y^2} dy + \int_{\frac{nx}{\pi}}^{+\infty} \left[\frac{1}{\pi(1+y^2)} - \frac{1}{\pi y^2} \right] dy. = \frac{1}{nx} + o\left(\frac{1}{n}\right).$$

D'où

$$F_Y(x) = \left[1 - \frac{1}{nx} + o\left(\frac{1}{n}\right) \right]^n \xrightarrow{n \rightarrow +\infty} \exp(-x^{-1}).$$

En conclusion, $F_Y(x) \xrightarrow{n \rightarrow +\infty} \Phi_1(x)$ pour $x > 0$.

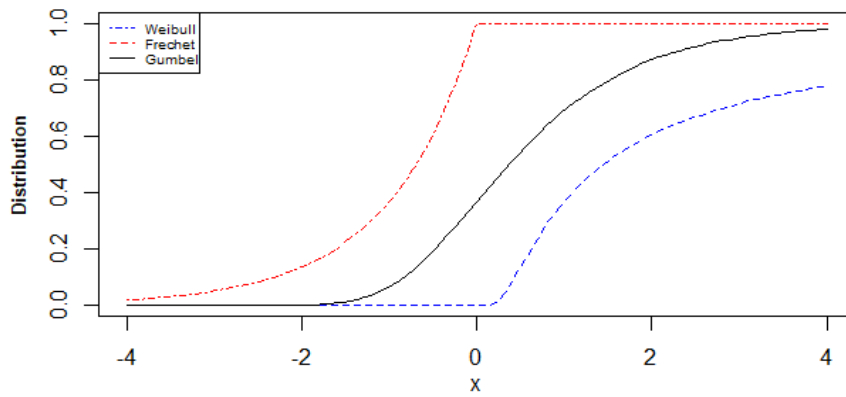


FIGURE 2.1 – Distribution de lois des valeurs extrêmes : Fréchet, Gumbel et Weibull.

2.3 Distribution Généralisée des Valeurs Extrêmes (GEV)

Afin d'alléger le travail avec les trois distributions limites, Jenkinson [12] et Von Mises [16] ont proposé une famille paramétrique combinant les trois lois (FRECHET, WEIBULL et GUMBEL). Cette nouvelle famille est appelée distribution généralisée des valeurs extrêmes (Generalized Extreme Values Distribution), que l'on note GEV, permettant d'unifier les trois lois extrêmes ci-dessous de probabilité qui modélise le comportement du maximum d'un échantillon. Nous définissons la fonction H par

$$H_\gamma(x) := \begin{cases} \exp(-(1 + \gamma x))^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, \ 1 + \gamma x > 0 \\ \exp(-\exp(-x)) & \text{si } \gamma = 0, \ x \in \mathbb{R}, \end{cases} \quad (2.19)$$

où H_γ est la distribution GEV et γ est un paramètre réel appelé **indice des valeurs extrêmes** ou **indice de queue** ("tail index" en anglais) qui contrôle la lourdeur de la queue de loi.

Remarque 2.1.

1. Pour $\gamma = 0$, $H_0(x) = \Lambda(x)$ correspond à la loi de Gumbel.
2. Pour $\gamma = \frac{1}{\alpha} > 0$, $H_{\frac{1}{\alpha}}(\alpha(x - 1)) = \Phi(x)$ correspond à la loi de Frechet de paramètre $\alpha > 0$.
3. Pour $\gamma < 0$, $H_\alpha(-\alpha(x + 1)) = \Psi_\alpha(x)$ correspond à la loi de Weibull de paramètre $\alpha < 0$.

La forme la plus générale de la distribution GEV pour les variables non-centrées et non-réduites est obtenu en remplaçant x par $\frac{x-\mu}{\sigma}$ dans l'équation. Nous obtenons dans ce cas

$$H_{\gamma,\mu,\sigma}(x) := \begin{cases} \exp\left[-\left(1 + \gamma \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right], & \text{si } \gamma \neq 0, \ 1 + \gamma \frac{x-\mu}{\sigma} > 0 \\ \exp(-\exp(-\frac{x-\mu}{\sigma})) & \text{si } \gamma = 0, \ x \in \mathbb{R}, \end{cases} \quad (2.20)$$

où $\mu \in \mathbb{R}$ est le paramètre de localisation et $\sigma > 0$ celui de l'échelle. Cette forme générale est la forme paramétrée de Von Mises [16]. La fonction de densité de la distribution GEV est de la forme

$$h_{\gamma,\mu,\sigma}(x) := \begin{cases} \frac{1}{\sigma} \left[1 + \gamma \frac{x-\mu}{\sigma}\right]^{-\left(\frac{1+\gamma}{\gamma}\right)} H_{\gamma,\mu,\sigma}(x) & \text{si } \gamma \neq 0 \\ \frac{1}{\sigma} \exp\left[-\left(\frac{x-\mu}{\sigma}\right) - \exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right] & \text{si } \gamma = 0 \end{cases} \quad (2.21)$$

2.4 Domaine d'attraction

Le domaine d'attraction est l'ensemble des distributions pour lesquelles la distribution limite du maximum converge vers la même limite GEV.

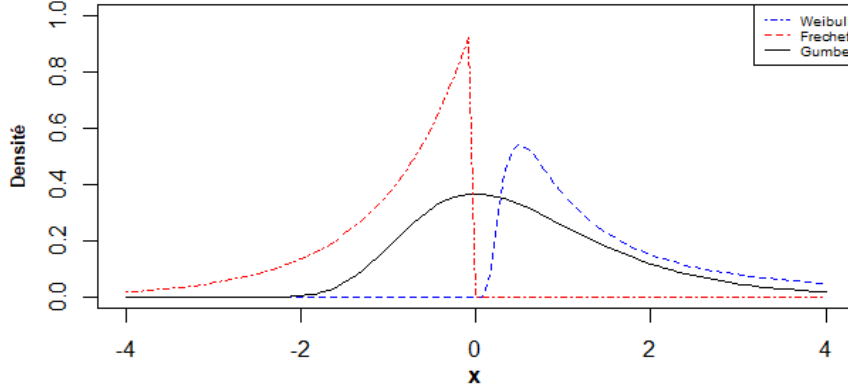


FIGURE 2.2 – Densités des trois types des distributions des valeurs extrêmes : Fréchet($\gamma = 1$), Gumbel($\gamma = 0$) et Weibull($\gamma = -1$)

En théorie des valeurs extrêmes, les informations les plus importantes résident dans les queues de la distribution et sont caractérisées par le paramètre de forme γ . Il faut bien noter que le paramètre γ joue un rôle crucial dans la détermination de la nature de la loi des valeurs extrêmes. En effet, selon le signe de γ , nous pouvons distinguer trois domaines d'attractions qui seront présentés dans la suite.

Définition 2.2.

Une distribution F appartient au domaine d'attraction de H_γ , que nous notons $F \in D(H_\gamma)$, s'il existe des suites de réels $(a_n > 0, n \in \mathbb{N}^*)$ et $(b_n \in \mathbb{R}, n \in \mathbb{N})$ telles que

$$\lim_{n \rightarrow +\infty} F^n(a_n x + b_n) = H_\gamma(x). \quad (2.22)$$

Proposition 2.1.

Soit le n -échantillon $X_1, X_2, \dots, X_n$ de variable aléatoires i.i.d ayant comme fdr F . Considérons les deux suites de normalisation $(a_n)_n$ et $(b_n)_n$ définies plus haut. Pour $\gamma \in \mathbb{R}$, nous avons

$$F \in D(H_\gamma) \iff \lim_{n \rightarrow +\infty} n\bar{F}(a_n x + b_n) = -\log H_\gamma(x) \quad (2.23)$$

Démonstration.

Supposons que : $F \in D(H_\gamma)$. Nous avons par définition

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

par suite, en remplaçant par la queue de distribution (ou fonction de survie), nous aurons

$$\lim_{n \rightarrow +\infty} [1 - \bar{F}(a_n x + b_n)]^n = H_\gamma(x).$$

En prenant le logarithme de l'expression ci-dessus, nous obtenons

$$\lim_{n \rightarrow +\infty} n \log [1 - \overline{F}(a_n x + b_n)] = \log H_\gamma(x).$$

Nous avons $\forall (1 + \gamma x) > 0$, $\overline{F}(a_n x + b_n)$ tend vers 0 et par le développement de Taylor pour la fonction logarithme, nous arrivons à

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

d'où

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

La réciproque est évident : si nous avons la convergence ci-dessus, alors $F \in D(H_\gamma)$.

2.4.1 Fonction à variations régulières

Dans la théorie des valeurs extrêmes, les fonctions à variations régulières constituent un outil essentiel pour la caractérisation des domaines d'attraction. Cependant, ces fonctions permettent de fournir une représentation unique pour chaque domaine d'attraction simplifiant ainsi l'analyse théorique.

Définition 2.3 (Fonction à variations régulières).

Une fonction $g : \mathbb{R}_+ \mapsto \mathbb{R}_+$ mesurable et positive, est dite à variations régulières d'indice $\beta \in \mathbb{R}$ si pour tout $t > 0$:

$$\lim_{x \rightarrow +\infty} \frac{g(tx)}{g(x)} = t^\beta. \quad (2.24)$$

L'ensemble des fonctions à variations régulières d'indice β est noté par $\mathcal{RV}_\beta$.

Définition 2.4 (Fonction à variations lente).

Une fonction L à variations régulières d'indice $\beta = 0$ est appelée fonction à variations lente ($L \in \mathcal{RV}_0$). Dans ce cas

$$\lim_{x \rightarrow +\infty} \frac{L(tx)}{L(x)} = 1, \quad t > 0. \quad (2.25)$$

Exemple 2.3.

Soit $x \in]0, +\infty[$, une fonction $f : x \mapsto \log x$ est une fonction à variation lente. En effet

$$\begin{aligned} \lim_{x \rightarrow +\infty} \frac{f(tx)}{f(x)} &= \lim_{x \rightarrow +\infty} \frac{\log(tx)}{\log(x)}, \quad t > 0 \\ &= 1 + \lim_{x \rightarrow +\infty} \frac{\log(t)}{\log(x)} = 1, \quad t > 0. \end{aligned}$$

Remarque 2.2.

Tout fonction à variation régulière d'indice $\beta \in \mathbb{R}$ ($g \in \mathcal{RV}_\beta$), peut s'écrire sous la forme : $g(x) = x^\beta L(x)$, où $L(x) \in \mathcal{RV}_0$.

Proposition 2.2 (Représentation de Karamata).

Toute fonction $L \in \mathcal{RV}_0$, s'écrit sous la forme :

$$L(x) = c(x) \exp \left(\int_{x_0}^x \frac{\delta(t)}{t} dt \right), \quad (2.26)$$

où $x_0 \in \mathbb{R}$, c et δ sont des fonctions mesurables avec $c(x) \rightarrow c_0 > 0$ et $\delta(x) \rightarrow 0$ lorsque x converge vers l'infini.

Démonstration.

Il suffit de montrer que

$$L(x) = c_0(x) \exp \left(\int_{x_0}^x \frac{\delta(t)}{t} dt \right) \implies \forall c > 0 : \frac{L(cx)}{L(x)} \xrightarrow{x \rightarrow \infty} 1,$$

où $c_0(x) \xrightarrow{x \rightarrow \infty} c_0 > 0$ et $\delta(t) \xrightarrow{t \rightarrow \infty} 0$. $\forall c > 0$,

$$\begin{aligned} \lim_{x \rightarrow +\infty} \frac{L(cx)}{L(x)} &= \lim_{n \rightarrow +\infty} \frac{c_0(cx) \exp \left(\int_{x_0}^{cx} \frac{\delta(t)}{t} dt \right)}{c_0(x) \exp \left(\int_{x_0}^x \frac{\delta(t)}{t} dt \right)} \\ &= \frac{c_0}{c_0} \cdot \lim_{x \rightarrow +\infty} \frac{\exp \left(\int_{x_0}^{cx} \frac{\delta(t)}{t} dt \right)}{\exp \left(\int_{x_0}^x \frac{\delta(t)}{t} dt \right)} \\ &= \lim_{x \rightarrow +\infty} \exp \left(\int_{x_0}^{cx} \frac{\delta(t)}{t} dt - \int_{x_0}^x \frac{\delta(t)}{t} dt \right). \end{aligned}$$

Supposons que $c > 1$ et par conséquent $cx > x$. Alors

$$\lim_{x \rightarrow +\infty} \frac{L(cx)}{L(x)} = \lim_{x \rightarrow +\infty} \exp \left(\int_x^{cx} \frac{\delta(t)}{t} dt \right).$$

Il suffit de montrer que $\int_x^{cx} \frac{\delta(t)}{t} dt \xrightarrow{x \rightarrow \infty} 0$. Puisque $\lim_{t \rightarrow +\infty} \delta(t) = 0$ alors $\forall \xi > 0, \exists t_0 > 0$ tel que $\forall t \geq t_0 : |\delta(t)| < \xi$. Dans ce cas, pour $x \geq t_0$

$$\begin{aligned} \left| \int_x^{cx} \frac{\delta(t)}{t} dt \right| &\leq \int_x^{cx} \frac{|\delta(t)|}{t} dt \\ &\leq \int_x^{cx} \frac{\xi}{t} dt = \xi \int_x^{cx} \frac{1}{t} dt = \xi \log c = k(\text{constante}). \end{aligned}$$

D'où $\left| \int_x^{cx} \frac{\delta(t)}{t} dt \right| \leq k$, ainsi $\int_x^{cx} \frac{\delta(t)}{t} dt \xrightarrow{x \rightarrow \infty} 0$ et, par conséquent :

$$\frac{L(cx)}{L(x)} \xrightarrow{x \rightarrow \infty} 1.$$

Remarque 2.3.

La fonction $L(\cdot)$ est dite normalisée si la fonction $c(\cdot)$ est constante.

2.4.2 Caractérisation des domaines d'attraction

Les différentes caractérisations des trois domaines d'attraction de Fréchet $D(\Phi_\gamma)$, de Weibull $D(\Psi_\gamma)$ et de Gumbel $D(\Lambda)$ ont été proposées dans Embrechts et al. [8], Reiss et Thomas [14] et De Haan et Ferreira [4].

Dans cette partie, nous allons donner les conditions nécessaires et suffisantes sur la fdr F pour qu'elle appartienne à l'un des domaines d'attraction de l'un des trois lois limites des valeurs extrêmes.

Théorème 2.2 (Caractérisation de $D(\Phi_\gamma)$).

La fonction de répartition F appartient au domaine d'attraction de la distribution de Fréchet de paramètre $\gamma > 0$ si et seulement si $x_F = +\infty$ et

$$\bar{F}(x) = x^{-\frac{1}{\gamma}} L(x). \quad (2.27)$$

où L est une fonction à variation lente.

De plus, si $F \in D(\Phi_\gamma)$, alors les suites de normalisation (a_n) et (b_n) sont données par : $a_n = U_n = F^{-1}(1 - \frac{1}{n})$ et $b_n = 0$. La suite $(a_n^{-1}X_{(n)})_{n \geq 1}$ converge en loi vers une variable aléatoire de fdr Φ_γ quand $n \rightarrow +\infty$.

Démonstration.

Supposons que $\bar{F}(x) = x^{-\frac{1}{\gamma}} L(x)$ avec L une fonction à variation lente. En utilisant la notation de la **Proposition 2.1**, nous avons $\bar{F}(x) \sim g(x)$ où $g(x) = x^{-\frac{1}{\gamma}} c_0 \exp\left(\int_a^x \frac{k(u)}{u} du\right)$ est une fonction continue. Posons $a_n = U_n$, en générale $\bar{F}(x) \leq \frac{1}{n} \leq \bar{F}(a_n^-)$ et a_n tend vers l'infini avec n . Pour simplifier, on suppose que F est continue en a_n , alors

$$\begin{aligned} a_n = U_n = F^{-1}\left(1 - \frac{1}{n}\right) &\implies F(a_n) = 1 - \frac{1}{n} \\ &\implies \bar{F}(a_n) = \frac{1}{n}. \end{aligned}$$

Pour $x > 0$ et $b_n = 0$, nous avons

$$\begin{aligned} \lim_{n \rightarrow +\infty} n\bar{F}(a_n x) &= \lim_{n \rightarrow +\infty} \frac{\bar{F}(a_n x)}{\frac{1}{n}} = \lim_{n \rightarrow +\infty} \frac{\bar{F}(a_n x)}{\bar{F}(a_n)} \\ &= \lim_{n \rightarrow +\infty} \frac{(a_n x)^{-\frac{1}{\gamma}} L(a_n x)}{(a_n)^{-\frac{1}{\gamma}} L(a_n)} = \lim_{n \rightarrow +\infty} x^{-\frac{1}{\gamma}} \frac{L(a_n x)}{L(a_n)} \\ &= x^{-\frac{1}{\gamma}}. \end{aligned}$$

En suivant la preuve de la Proposition 2.1,

$$n\bar{F}(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{} -\log \Phi_\gamma(x) = -\log e^{-x^{-\frac{1}{\gamma}}} = x^{-\frac{1}{\gamma}}.$$

Cela implique que $F \in D(\Phi_\gamma)$. L'implication inverse est omise.

Exemple 2.4 (loi de Paréto).

Considérons la suite de v.a $X_1, X_2, \dots, X_n$ iid issue d'une loi de Paréto de paramètre $\gamma > 0$, et de fdr $F(x) = 1 - kx^{-\gamma}$ avec $k > 0$. Prenons pour $n \geq$

1, $a_n = (kn)^{\frac{1}{\gamma}}$ et $b_n = 0$, on a :

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

D'où, le maximum normalisé de la loi de Paréto converge vers la loi de Fréchet et par conséquent $F(x) \in D(\Phi_\gamma(x))$.

Théorème 2.3 (Caractérisation de $D(\Psi_\gamma)$).

La fonction de répartition F appartient au domaine d'attraction de la distribution de Weibull de paramètre $\gamma < 0$ si et seulement si $x_F < +\infty$ et pour $x \leq x_F$

$$\overline{F}(x) = (x_F - x)^{-\frac{1}{\gamma}} L[(x_F - x)]. \quad (2.28)$$

où L est une fonction à variation lente.

De plus, si $F \in D(\Psi_\gamma)$ alors les suites de normalisation (a_n) et (b_n) sont données par : $a_n = x_F - U_n = x_F - F^{-1}(1 - \frac{1}{n})$ et $b_n = x_F$. La suite $(a_n^{-1}(X_{(n)} - x_F))_{n \geq 1}$ converge en loi vers une v.a de fdr Ψ_γ quand $n \rightarrow +\infty$.

Démonstration.

La preuve du théorème 2.3 est similaire à celle du théorème 2.2, voir Embrechts et al. [8].

Théorème 2.4 (Caractérisation de $D(\Lambda)$).

La fonction de répartition F d'une variable aléatoire X appartient au domaine d'attraction de la distribution de Gumbel si et seulement s'il existe $x_0 < x_F \leq +\infty$ tel que

$$\overline{F}(x) = c(x) \exp\left(-\int_{x_0}^x \frac{g(t)}{\sigma(t)} dt\right), \quad (2.29)$$

où $c(\cdot)$ et $g(\cdot)$ sont des fonctions mesurables satisfaisantes $c(x) \rightarrow c_0 > 0$ et $g(x) \rightarrow 1$ quand $x \rightarrow x_F$, et $\sigma(\cdot)$ est une fonction positive dérivable de dérivée σ' vérifiant $\lim_{x \rightarrow x_F} \sigma'(x) = 0$.

Les suites de normalisation (a_n) et (b_n) pouvant être choisies de telle façon où pour tout $n > 0$ $a_n = \sigma(b_n)$ et $b_n = U(x) = F^{\leftarrow}(1 - \frac{1}{n}) = \overline{F}^{\leftarrow}(\frac{1}{n})$, où U est la fonction quantile de queue de la variable X , avec $\sigma(\cdot)$ choisie comme

$$\sigma(x) = \int_x^{x_F} \frac{\overline{F}(t)}{\overline{F}(x)} dt, \quad x < x_F. \quad (2.30)$$

Démonstration.

Pour la preuve de ce résultat, nous pouvons consulter Beirlant et al. [2], Chapitre 2.

Exemple 2.5 (loi exponentielle).

Soit $X_1, X_2, \dots, X_n$ une suite de variables aléatoires indépendantes de loi exponentielle de paramètre $\lambda > 0$. Prenons $a_n = \frac{1}{\lambda}$ et $b_n = \frac{\log(n)}{\lambda}$, pour $x > 0$ nous obtenons

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

Le maximum normalisé de la loi exponentielle converge dans ce cas vers la loi de Gumbel et $F(x) \in D(\Lambda(x))$.

Le tableau (2.4.1) ci-après résume quelques lois et leurs domaines d'attraction.

Domaine d'attraction	Gumbel ($\gamma = 0$) $D(\Lambda)$	Fréchet ($\gamma > 0$) $D(\Phi_\gamma)$	Weibull ($\gamma < 0$) $D(\Psi_\gamma)$
Lois	Exponentielle Normale Weibull Log-normale Gamma	Cauchy Pareto Student Log-gamma	Uniforme Beta Reverse Burr

TABLE 2.4.1 – Domaines d'attraction et quelques lois associées

Remarque 2.4.

La relation suivante relie les variables aléatoires dans les trois domaines d'attraction de Fréchet, de Weibull et de Gumbel. (Voir Embrechts et al. [8] pour plus de détails)

$$X \in D(\Phi_\gamma) \iff \log(X^{-\frac{1}{\gamma}}) \in D(\Lambda) \iff -X^{-1} \in D(\Psi_\gamma) \quad (2.31)$$

2.5 Distribution de Paréto Généralisée (GPD)

L'approche par la GEV consiste à utiliser un seul maximum. Cependant, cette méthode a été critiquée dans la mesure où l'utilisation d'un seul maxima peut conduire à une perte d'information contenue dans les autres grandes valeurs de l'échantillon.

Afin de franchir ce problème, Pickands a introduit une autre approche dans la TVE connue par la méthode POT (Peaks Over Threshold) ou bien la méthode des excès au-delà d'un seuil élevé. L'idée consiste à observer pas seulement le maximum ou les plus grandes valeurs, mais toutes les valeurs des réalisations

qui excèdent un certain seuil élevé. Cela revient à choisir un seuil suffisamment élevé et étudier les excès au-delà de ce seuil.

Notons u un réel suffisamment élevé appelé seuil. Définissons les excès au-delà du seuil u comme l'ensemble des variables aléatoires conditionnelles $Y_i = X_j - u / X_j > u$.

À partir de la loi F de X , nous cherchons à définir une loi conditionnelle F_u par rapport au seuil u pour les v.a dépassant ce seuil. La loi conditionnelle des excès F_u sera donné par

$$F_u(y) = P(X - u \leq y / X > u).$$

De plus

$$\begin{aligned} F_u(y) &= \frac{P(X - u \leq y, X > u)}{P(X > u)} = \frac{P(u < X \leq y + u)}{1 - P(X \leq u)} \\ &= \frac{P(X \leq y + u) - P(X \leq u)}{1 - F(u)} \\ &= \frac{F(y + u) - F(u)}{\bar{F}(u)}, \quad y \geq 0 \end{aligned}$$

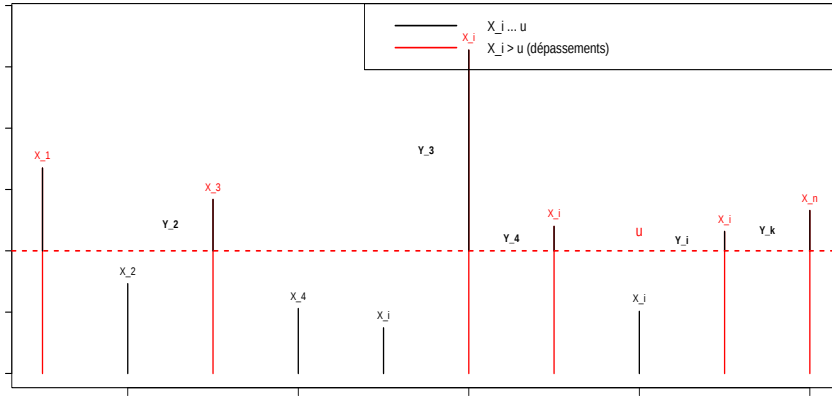


FIGURE 2.3 – Les dépassements

2.5.1 Théorème de Pickands-Balkema-de Haan

Le théorème de Pickands-Balkema-de Haan donne la forme de la loi limite pour les valeurs extrêmes. Sous certaines conditions de convergence, la loi limite est une loi de Paréto généralisée (Generalized Pareto Distribution que l'on note GPD).

Définition 2.5 (Distribution de Paréto Généralisée).

La loi de Paréto généralisée de paramètres $\sigma > 0$, $\gamma \in \mathbb{R}$ est définie par sa

fonction de répartition $G_{\gamma,\mu,\sigma}$ donnée par

$$G_{\gamma,\mu,\sigma}(x) := \begin{cases} 1 - \left(1 + \gamma \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{si } \gamma = 0 \end{cases} \quad (2.32)$$

où

$$\begin{cases} x \geq 0 & \text{si } \gamma \geq 0 \\ \mu \leq x \leq \mu - \frac{\sigma}{\gamma} & \text{si } \gamma < 0. \end{cases}$$

La GPD standard correspond au cas où $\mu = 0$ et $\sigma = 1$, telle que

$$G_{\gamma}(x) := \begin{cases} 1 - (1 + \gamma x)^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, \\ 1 - \exp(-x) & \text{si } \gamma = 0, \end{cases} \quad (2.33)$$

avec

$$\begin{cases} x \geq 0 & \text{si } \gamma \geq 0 \\ 0 \leq x \leq -\frac{1}{\gamma} & \text{si } \gamma < 0 \end{cases}$$

Nous donnons un résultat concernant la distribution conditionnelle des excès lorsque le seuil déterministe tend vers le point terminal x_F .

Théorème 2.5 (Pickands-Balkema-de Haan).

Soit $X_1, X_2, \dots, X_n$ un échantillon de variables aléatoires i.i.d de fonction de distribution F . Si cette loi appartient au domaine d'attraction maximale de la distribution GEV alors il existe une fonction positive $\sigma(u)$ et un réel γ tels que

$$\lim_{u \rightarrow x_F} \sup_{0 < y < x_F - u} |F_u(y) - G_{\gamma,\sigma(u)}(y)| = 0. \quad (2.34)$$

Où F_u est une fonction de répartition des excès au-delà du seuil u , x_F est le point terminal de F et $G_{\gamma,\sigma(u)}$ est la fonction de répartition de la loi de Pareto généralisée donnée par :

$$G_{\gamma,\sigma(u)}(x) := \begin{cases} 1 - \left(1 + \gamma \frac{x}{\sigma(u)}\right)^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, \\ 1 - \exp\left(-\frac{x}{\sigma(u)}\right) & \text{si } \gamma = 0, \end{cases} \quad (2.35)$$

où

$$\begin{cases} x \geq 0 & \text{si } \gamma \geq 0 \\ 0 \leq x \leq -\frac{\sigma(u)}{\gamma} & \text{si } \gamma < 0 \end{cases}$$

On peut conclure de ce théorème que, si la fonction de répartition F appartient au domaine d'attraction de la loi H_{γ} , alors il existe une fonction positive $\sigma(u)$ et un réel γ tels que la loi des excès F_u peut-être uniformément approchée par une distribution de Paréto généralisée GPD notée $G_{\gamma,\sigma(u)}$.

Si H_{γ} est écrite sous la forme d'une GEV de l'équation (2.19), alors l'indice de queue γ est le même que celui de la GPD de l'équation (2.33).

Remarque 2.5.

Le Théorème 2.5 montre l'existence d'une relation étroite entre la GPD standard $G_{\gamma}(x)$ et la GEV standard $H_{\gamma}(x)$, tels que :

$$G_{\gamma}(x) = 1 + \log(H_{\gamma}(x)), \quad \text{si } \log(H_{\gamma}(x)) > -1$$

Exemple 2.6.

Prenons une loi exponentielle de fdr F donnée pour $x > 0$ par, $F(x) = 1 - e^{-x}$. Alors

$$\begin{aligned} F_u(y) &= \frac{F(y+u) - F(u)}{1 - F(u)}, \quad y \geq 0 = \frac{(1 - e^{-(u+y)}) - (1 - e^{-u})}{1 - (1 - e^{-u})} \\ &= \frac{e^{-u} - e^{-(u+y)}}{e^{-u}} = 1 - e^{-y}. \end{aligned}$$

D'où, la loi limite des excès est la loi GPD standard de paramètre $\gamma = 0$.

Exemple 2.7.

Considérons cette fois la loi de Paréto de paramètre $\gamma > 0$ de fdr donnée par : $F(x) = 1 - kx^{-\frac{1}{\gamma}}$, $k > 0$. Alors

$$\begin{aligned} F_u(y) &= \frac{F(y+u) - F(u)}{1 - F(u)}, \quad y \geq 0 \\ &= \frac{(1 - k(u+y)^{-\frac{1}{\gamma}}) - (1 - ku^{-\frac{1}{\gamma}})}{1 - (1 - ku^{-\frac{1}{\gamma}})} \\ &= 1 - \left(1 + \frac{1}{u}y\right)^{-\frac{1}{\gamma}} \end{aligned}$$

Ceci correspond à la loi GPD de paramètre $\gamma > 0$ et $\sigma(u) = \gamma u$.

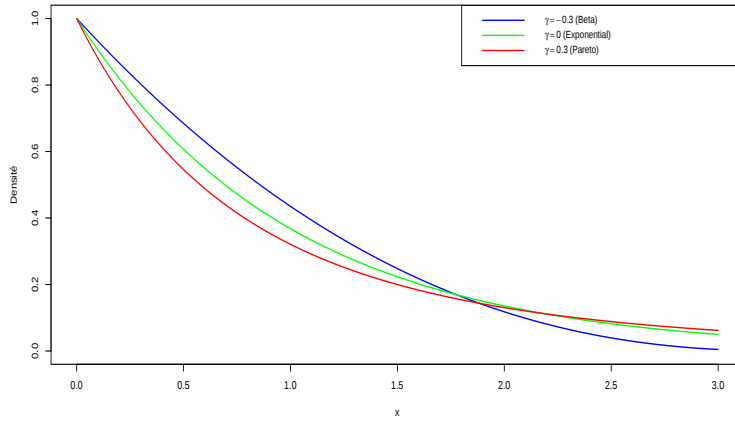


FIGURE 2.4 – Comparaison des densités GPD pour différentes valeurs γ .

Chapitre 3

Estimation de l'indice des Valeurs Extrêmes

3.1 Introduction

La distribution asymptotique du maximum suit une loi des valeurs extrêmes caractérisée par le paramètre de queue γ . Ce paramètre fournit une information cruciale sur le comportement de la queue de la distribution. Comme établi dans la partie précédente, le signe du paramètre détermine l'appartenance à l'un des trois domaines d'attraction : Fréchet, Weibull ou Gumbel. Dans ce cas, une estimation de γ revêt une importance fondamentale dans l'identification de la loi limite appropriée.

Pour estimer ce paramètre, deux approches sont utilisées : paramétrique et semi-paramétrique. Dans le cadre paramétrique, la méthode du maximum de vraisemblance est largement utilisée, tandis que le second fait appel à plusieurs méthodes d'estimation, notamment l'estimateur de Hill, l'estimateur des moments et l'estimateur de Pickands.

Dans cette partie nous présentons les différentes méthodes d'estimation des paramètres intervenant dans les distributions extrêmes présentées précédemment.

3.2 Estimation Paramétrique

3.2.1 Méthode du maximum de vraisemblance

Les résultats asymptotiques obtenus par cette méthode sont efficaces, et les estimateurs obtenus convergent sous certaines conditions vers les vraies valeurs des paramètres.

Nous construisons l'estimateur du maximum de vraisemblance à partir des observations des maxima. Nous estimons l'indice des valeurs extrêmes γ ainsi que les deux paramètres, de position μ et d'échelle σ .

Pour la loi GEV

Considérons un n -échantillon $X_1, X_2, \dots, X_n$ de variables aléatoires i.i.d de densité h_θ et de fonction de répartition égale à H_θ où $\theta = (\gamma, \mu, \sigma) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$, γ

est l'indice de queue, μ est le paramètre de position et σ est celui d'échelle. Nous nous intéressons à l'estimation des paramètres de la loi des valeurs extrêmes de la distribution GEV suivante

$$H_\theta(x) := \begin{cases} \exp \left[- \left(1 + \gamma \frac{x-\mu}{\sigma} \right)^{-\frac{1}{\gamma}} \right], & \text{si } \gamma \neq 0, \quad 1 + \gamma \frac{x-\mu}{\sigma} > 0 \\ \exp(-\exp(-\frac{x-\mu}{\sigma})) & \text{si } \gamma = 0, \quad x \in \mathbb{R}. \end{cases}$$

La fonction de vraisemblance (likelihood function) est donnée par

$$\mathcal{L}((X_1, X_2, \dots, X_n); \theta) = \prod_{i=1}^n h_\theta(x_i); \quad \theta = (\gamma, \mu, \sigma);$$

et pour $\gamma \neq 0$, la densité h_θ est donnée par

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

pour $\gamma \neq 0$, on a la fonction de log-vraisemblance suivante

$$\log \mathcal{L}((X_1, X_2, \dots, X_n); (\gamma, \mu, \sigma)) = -n \log \sigma - \left(\frac{1+\gamma}{\gamma} \right) \sum_{i=1}^n \log \left[1 + \gamma \frac{x_i - \mu}{\sigma} \right] - \sum_{i=1}^n \left[1 + \gamma \frac{x_i - \mu}{\sigma} \right]^{-\frac{1}{\gamma}}.$$

En dérivant cette fonction par rapport aux trois paramètres μ , σ et γ , nous obtenons le système d'équations suivant

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \mu} = 0 &\iff (1+\gamma) \sum_{i=1}^n \frac{1}{1 + \gamma \frac{x_i - \mu}{\sigma}} - \frac{1}{\sigma} \sum_{i=1}^n \left(1 + \gamma \frac{x_i - \mu}{\sigma} \right)^{-\left(\frac{1+\gamma}{\gamma}\right)} = 0 \\ \frac{\partial \log \mathcal{L}}{\partial \sigma} = 0 &\iff -\frac{n}{\sigma} + \left(\frac{1+\gamma}{\sigma} \right) \sum_{i=1}^n \frac{x_i - \mu}{\sigma + \gamma(x_i - \mu)} - \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \left[1 + \gamma \frac{x_i - \mu}{\sigma} \right]^{-\left(\frac{1+\gamma}{\gamma}\right)} = 0 \\ \frac{\partial \log \mathcal{L}}{\partial \gamma} = 0 &\iff -\left(1 - \frac{1}{\gamma^2} \right) \sum_{i=1}^n \log \left(1 + \gamma \frac{x_i - \mu}{\sigma} \right) - \left(\frac{\gamma-1}{\sigma} \right) \sum_{i=1}^n \frac{x_i - \mu}{1 + \gamma(x_i - \mu)} + \\ &\sum_{i=1}^n \left(1 + \gamma \frac{x_i - \mu}{\sigma} \right)^{-\frac{1}{\gamma}} \left[\frac{1}{\sigma} \sum_{i=1}^n \frac{x_i - \mu}{1 + \gamma^2(x_i - \mu)} - \frac{1}{\gamma^2} \log \left(1 + \gamma \frac{x_i - \mu}{\sigma} \right) \right] = 0 \end{aligned}$$

Le système d'équations ci-dessus est un système non-linéaire pour lesquels aucune solution explicite existe. Pour $\gamma = 0$, la densité h_θ devient

$$h_\theta(x) = \frac{1}{\sigma} \exp \left[\frac{x_i - \mu}{\sigma} - \exp \left(\frac{x_i - \mu}{\sigma} \right) \right].$$

La fonction de log-vraisemblance est donnée par

$$\log \mathcal{L}((X_1, X_2, \dots, X_n); (\mu, \sigma)) = -n \log \sigma - \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^n \exp \left(-\frac{x_i - \mu}{\sigma} \right).$$

En prenant la dérivée de cette fonction par rapport aux deux paramètres μ et σ , nous obtenons le système d'équations ci-dessous

$$\frac{\partial \log \mathcal{L}}{\partial \mu} = 0 \iff n - \sum_{i=1}^n \exp \left(-\frac{x_i - \mu}{\sigma} \right) = 0.$$

$$\frac{\partial \log \mathcal{L}}{\partial \sigma} = 0 \iff n + \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right) \left[\exp \left(-\frac{x_i - \mu}{\sigma} \right) - 1 \right] = 0.$$

Vu que la résolution de ce système est relativement difficile et n'admet pas généralement de solutions explicites, nous faisons dans ce cas appel à des méthodes d'optimisation numérique.

Pour la loi GPD

Soit $X_1, X_2, \dots, X_n$ un échantillon de n variables aléatoires i.i.d de densité g_θ et de fonction de répartition G_θ ou $\theta = (\gamma, \sigma) \in \mathbb{R} \times \mathbb{R}_+$. Nous nous intéressons à l'estimation des deux paramètres σ et γ de la loi des valeurs extrêmes de la distribution GPD $G_\theta(x) := \begin{cases} 1 - (1 + \gamma \frac{x}{\sigma})^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, \\ 1 - \exp(-\frac{x}{\sigma}) & \text{si } \gamma = 0 \end{cases}$

La fonction de vraisemblance associée est écrite sous la forme

$$\mathcal{L}((X_1, X_2, \dots, X_n); \theta) = \prod_{i=1}^n g_\theta(x_i); \quad \theta = (\gamma, \sigma).$$

Pour $\gamma \neq 0$, la densité est

$$g_\theta(x) = \frac{1}{\sigma} \exp \left[- \left(\frac{x - \mu}{\sigma} \right) - \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \right],$$

par conséquent, la fonction log-vraisemblance pour $\gamma \neq 0$ devient

$$\log \mathcal{L}((X_1, X_2, \dots, X_n); (\gamma, \sigma)) = -n \log \sigma - \left(\frac{1 + \gamma}{\gamma} \right) \sum_{i=1}^n \log \left[1 + \frac{\gamma}{\sigma} x_i \right].$$

En prenant la dérivée de cette fonction par rapport à σ et γ , nous arrivons au système d'équations

$$\begin{cases} \frac{\partial \log \mathcal{L}}{\partial \sigma} = 0 \iff -n + (1 + \gamma) \sum_{i=1}^n \frac{x_i}{\sigma + \gamma x_i} = 0 \\ \frac{\partial \log \mathcal{L}}{\partial \gamma} = 0 \iff \left(\frac{1 + \gamma}{\gamma} \right) \sum_{i=1}^n \frac{x_i}{\sigma + \gamma x_i} - \frac{1}{\gamma^2} \sum_{i=1}^n \log \left(1 + \frac{\gamma}{\sigma} x_i \right) = 0. \end{cases}$$

Dans le cas où $\gamma = 0$, la densité g_θ est donnée par

$$g_\theta(x) = \frac{1}{\sigma} \exp \left(-\frac{x}{\sigma} \right),$$

la fonction log-vraisemblance associée est de la forme

$$\begin{aligned} \log \mathcal{L}((X_1, X_2, \dots, X_n); (0, \sigma)) &= \log \prod_{i=1}^n g_\theta(x_i); \quad \theta = (0, \sigma) \\ &= -n \log \sigma - \frac{1}{\sigma} \sum_{i=1}^n x_i. \end{aligned}$$

En dérivant par rapport à σ , nous obtenons

$$\begin{aligned}\frac{\partial \log \mathcal{L}}{\partial \sigma} = 0 &\iff -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n x_i = 0 \\ &\iff n = \frac{1}{\sigma} \sum_{i=1}^n x_i \\ &\iff \hat{\sigma} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}_n.\end{aligned}$$

3.3 Estimation par les méthodes semi-paramétrique

3.3.1 Estimateur de Hill

L'idée de base de la construction des estimateurs consiste à trouver des conditions équivalentes d'appartenance à un domaine d'attraction et qui s'expriment à partir du paramètre de forme γ .

En 1975, Hill a introduit un estimateur de forme semi-paramétrique pour l'indice des valeurs extrêmes en se basant fortement sur les plus grandes statistiques d'ordre.

L'utilisation de cet estimateur est limitée aux distributions du domaine d'attraction de Fréchet, ce qui correspond au scénario où l'indice des valeurs extrêmes est positif.

Définition 3.1 (Estimateur de Hill).

Soit $X_1, X_2, \dots, X_n$ un échantillon de variables aléatoires i.i.d de fonction de répartition F appartenant au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes $\gamma > 0$. Soient $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ les statistiques d'ordre associées. Notons par $X_{(n-k)}$, $1 \leq k \leq n$, la $k^{\text{ième}}$ plus grande valeur de l'échantillon. Soit $(k_n)_{n \geq 1}$ une suite d'entiers avec $1 < k_n < n$, l'estimateur de Hill est défini par

$$\hat{\gamma}_{k,n}^H := \frac{1}{k} \sum_{i=1}^k \log(X_{(n-i+1)}) - \log(X_{(n-k)}), \quad 1 \leq k \leq n. \quad (3.1)$$

La construction de cet estimateur est basée sur l'estimation par la méthode du maximum de vraisemblance appliquée à la queue de la distribution F .

Nous supposons que l'indice des valeurs extrêmes γ est positif, alors la distribution F doit vérifier la condition suivante :

$$\frac{\bar{F}(x)}{\bar{F}(u)} = \left(\frac{x}{u}\right)^{-\frac{1}{\gamma}}, \quad x \geq u, \quad u \in \mathbb{R}, \quad (3.2)$$

avec $\bar{F}(z) = 1 - F(z)$.

Il serait dans ce cas plus raisonnable d'utiliser les statistiques d'ordre dépassant le seuil u . Il faut donc utiliser les statistiques d'ordre $X_{(n-k+1)}, \dots, X_{(n)}$.

Comme $\frac{\bar{F}(x)}{\bar{F}(u)} = \left(\frac{x}{u}\right)^{-\frac{1}{\gamma}}$, alors

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

$$\Rightarrow -F'(x) = -\frac{\bar{F}(u)}{\gamma u} \cdot \left(\frac{x}{u}\right)^{-\frac{\gamma+1}{\gamma}}, \gamma > 0.$$

La densité f des variables aléatoires $X_1, \dots, X_n$ est la suivante

$$f(x) = \frac{\bar{F}(u)}{\gamma u} \cdot \left(\frac{x}{u}\right)^{-\frac{\gamma+1}{\gamma}},$$

par conséquent, la fonction de vraisemblance pour les statistiques d'ordre $X_{(n-k+1)}, \dots, X_{(n)}$.

$$\mathcal{L}(X_{(n-k+1)}, \dots, X_{(n)}; \gamma) = \left[\frac{\bar{F}(u)}{\gamma u} \right]^k \cdot \prod_{i=1}^k \left(\frac{x_{(n-i+1)}}{u} \right)^{-\frac{\gamma+1}{\gamma}}.$$

La fonction log-vraisemblance sera alors

$$\log \mathcal{L}(X_{(n-k+1)}, \dots, X_{(n)}; \gamma) = \log \bar{F}(u) - k \log(\gamma u) - \frac{\gamma+1}{\gamma} \sum_{i=1}^k (\log x_{n-i+1,n} - \log u).$$

En dérivant par rapport à γ , nous arrivons à :

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \gamma} = 0 &\iff -\frac{k}{\gamma} + \frac{1}{\gamma^2} \sum_{i=1}^k (\log x_{(n-i+1)} - \log u) = 0 \\ &\iff k = \frac{1}{\gamma} \sum_{i=1}^k (\log x_{(n-i+1)} - \log u) \\ &\iff \hat{\gamma} = \frac{1}{k} \sum_{i=1}^k (\log x_{(n-i+1)} - \log u). \end{aligned}$$

En remplaçant le seuil u par la statistique d'ordre $X_{(n-k)}$, nous aboutissons à l'estimateur de Hill.

Propriétés asymptotiques de l'estimateur de Hill

Théorème 3.1.

Soit le n -échantillon $X_1, X_2, \dots, X_n$ de fonction de répartition F appartenant au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes $\gamma > 0$. Pour $k = (k_n)_{n \geq 1}$ une suite d'entiers telle que $1 < k_n \leq n$, nous avons

(a) Consistance faible

Si $k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$ quand $n \rightarrow \infty$, alors l'estimateur de Hill converge en probabilité vers γ . Nous avons

$$\hat{\gamma}_{k,n}^H \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \gamma.$$

(b) Consistance forte

Si $\frac{k_n}{n} \rightarrow 0$ et $\frac{k_n}{\log(\log n)} \rightarrow \infty$ quand $n \rightarrow \infty$, alors l'estimateur de Hill converge presque sûrement vers γ . Nous avons

$$\hat{\gamma}_{k,n}^H \xrightarrow[n \rightarrow +\infty]{p.s.} \gamma$$

Démonstration.

Soit $X_1, \dots, X_n$ un échantillon de variables aléatoires i.i.d de fonction de répartition F appartenant au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes $\gamma > 0$. Soit U la fonction quantile de queue de la distribution F définie par

$$U(t) := F^{\leftarrow} \left(1 - \frac{1}{t} \right), \quad \forall t > 1.$$

Admettons que U est une fonction à variations régulières d'indice γ , ($U \in \mathcal{RV}_\gamma$). Alors

$$\lim_{t \rightarrow +\infty} \frac{U(zt)}{U(t)} = z^\gamma, \quad \forall z > 0. \quad (3.3)$$

Considérons $X_{(n-j+1)}$, la j -ième plus grande valeur d'échantillon, $1 \leq j \leq n$. Nous supposons qu'il existe une suite d'entiers $k = (k_n)_{n \geq 1}$ avec $(1 < k \leq n)$ telle que $k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$ quand $n \rightarrow \infty$, dans ce cas, les approximations suivantes sont obtenues comme suit

$$X_{(n-j+1)} \approx U \left(\frac{n}{j} \right) \quad \text{avec} \quad \frac{X_{(n-j+1)}}{U \left(\frac{n}{j} \right)} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 1,$$

et

$$X_{(n-k)} \approx U \left(\frac{n}{k+1} \right) \approx U \left(\frac{n}{k} \right).$$

L'expression ci-dessus signifie que la j -ième plus grande observation de l'échantillon se comporte asymptotiquement comme le quantile $U(\frac{n}{j})$. Si F est une fonction de répartition, alors le rang attendu de $X_{(n-j+1)}$ correspond approximativement au niveau de probabilité $1 - \frac{j}{n}$ et son quantile est $F^{\leftarrow} \left(1 - \frac{j}{n} \right)$. Or, puisque $U(t) := F^{\leftarrow} \left(1 - \frac{1}{t} \right)$, nous remplaçons donc $t = \frac{n}{j}$, ce qui nous conduit à $U(\frac{n}{j}) := F^{\leftarrow} \left(1 - \frac{j}{n} \right)$. Pour k assez grand, $\frac{n}{k+1}$ et $\frac{n}{k}$ sont asymptotiquement

les mêmes. Alors $\begin{cases} \log X_{(n-j+1)} & \approx & \log U \left(\frac{n}{j} \right) \\ \log X_{(n-k)} & \approx & \log U \left(\frac{n}{k} \right) \end{cases}$ d'où

$$\log U \left(\frac{n}{j} \right) - \log U \left(\frac{n}{k} \right) = \log \frac{U \left(\frac{n}{j} \right)}{U \left(\frac{n}{k} \right)},$$

$$\log U(n/j) - \log U(n/k) = \log \frac{U(n/j)}{U(n/k)}.$$

Cependant, d'après la relation 3.3, quand n tend vers ∞ ,

$$\frac{U(n/j)}{U(n/k)} = \frac{U \left(\frac{n}{k} \cdot \frac{k}{j} \right)}{U \left(\frac{n}{k} \right)} \longrightarrow \left(\frac{k}{j} \right)^\gamma,$$

Nous arrivons à

$$\log \frac{U(n/j)}{U(n/k)} \longrightarrow \gamma \log \frac{k}{j}.$$

L'estimateur de Hill devient

$$\hat{\gamma}_{k,n}^H \approx \frac{1}{k} \sum_{j=1}^k \log \frac{U(n/j)}{U(n/k)} \xrightarrow[n \rightarrow +\infty]{} \frac{1}{k} \sum_{j=1}^k \gamma \log \frac{k}{j}.$$

D'après la limite de la somme de Riemann, nous allons obtenir

$$\frac{1}{k} \sum_{j=1}^k \log \frac{k}{j} = \frac{1}{k} \sum_{j=1}^k \log \left(1 + \frac{k-j}{j} \right) \xrightarrow{k \rightarrow +\infty} \int_0^1 \log \frac{1}{t} dt = 1.$$

Ce qui nous permet de conclure le résultat.

Concernant la consistance forte, nous renvoyons le lecteur à Deheuvels et al. [5].

Remarque 3.1.

Notons que la condition $k_n \rightarrow \infty$ assure que la taille de la statistiques d'ordre k est assez grande dans le but de construire des estimateurs stables. En revanche, la condition $\frac{k_n}{n} \rightarrow 0$ permet de rester dans la queue de distribution.

Normalité asymptotique

Afin d'établir la normalité asymptotique de l'estimateur de Hill $\hat{\gamma}_{k_n, n}^H$, nous avons besoin d'introduire une hypothèse sur la fonction de répartition F , nous supposons qu'elle est à variation régulière du second ordre.

Définition 3.2 (Condition du second ordre).

Une fonction de queue de quantile U est à variation régulière du second ordre avec le paramètre du premier ordre $\gamma > 0$ et le paramètre du second ordre $\rho < 0$ (on écrit $U \in 2\mathcal{RV}_{\gamma, \rho}$), s'il existe une fonction A satisfaisant $\lim_{t \rightarrow \infty} A(t) = 0$ ne changeant pas son signe près de ∞ , telles que pour tout $x > 0$

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho} \quad (3.4)$$

Où $A \in \mathcal{RV}_\rho$ appelée la fonction auxiliaire de U .

Considérons deux fonctions A et A^* à variations régulières avec $A^*(t) = A\left(\frac{1}{F(t)}\right)$ dont rôle est de contrôler la vitesse de convergence vers la loi limite.

Théorème 3.2 (Normalité asymptotique).

Soit $X_1, \dots, X_n$ un échantillon de n variables aléatoires i.i.d de fonction de répartition F appartenant au domaine d'attraction de Fréchet. Soit $(k_n)_{n \geq 1}$ une suite d'entiers telle que $1 < k_n \leq n$. Si $k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$ quand $n \rightarrow \infty$ et si la condition du second ordre (3.4) est satisfaite avec $\lim_{n \rightarrow \infty} \sqrt{k} A^*\left(\frac{n}{k}\right) = \lambda$, alors

$$\sqrt{k} (\hat{\gamma}_{k, n}^H - \gamma) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(\frac{\lambda}{1 - \rho}, \gamma^2\right) \quad (3.5)$$

La preuve détaillée du Théorème 3.2 est détaillée notamment dans le livre de De Haan et Ferreira [4] et récemment dans Deme [7].

Le choix du nombre k_n

Le choix de k_n (C'est-à-dire du nombre d'observations que l'on considère) est crucial. Si k_n est trop grand, alors l'approximation par une loi de Paréto sera mauvaise et l'estimateur de Hill aura un biais important. À l'inverse si k_n est trop petit, on disposera de très peu d'observations pour le calcul de l'estimateur

qui aura alors une variance importante. Le bon choix de k_n est donc celui de meilleur compromis biais/variance.

En théorie, on a vu que pour obtenir la consistance de l'estimateur de Hill, il faut prendre pour k_n une suite intermédiaire : $k_n \rightarrow \infty$ (k_n assez grand) et $\frac{k_n}{n} \rightarrow 0$ (k_n pas trop grand).

Pour choisir cette suite k_n en pratique, une méthode graphique est souvent utilisée. Elle consiste à tracer le "Hill Horror Plot", c'est-à-dire le graphe des points $\{(k_n, \hat{\gamma}_n^H(k_n)), k_n = 1, \dots, n\}$

L'estimateur de Hill est biaisé et le résultat sur sa normalité asymptotique permet de donner un intervalle de confiance pour l'indice γ . On peut alors associer un intervalle de confiance $I_n(\alpha)$ de niveau α à l'estimateur de Hill.

$$I_n(\alpha) = \left[\hat{\gamma}_{k_n, n}^H \mp \frac{z_{\frac{\alpha}{2}}}{\sqrt{k_n}} \right] \quad (3.6)$$

Où $z_{\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ d'une loi normale centrée réduite.

La FIGURE 2.9 ci-après représente l'estimateur de Hill $\hat{\gamma}_{k_n, n}^H$ et l'intervalle de confiance associé en fonction de k_n , pour un échantillon de taille $n = 3000$ issu de la loi de Paréto standard avec 100 réplifications. La figure montre que le

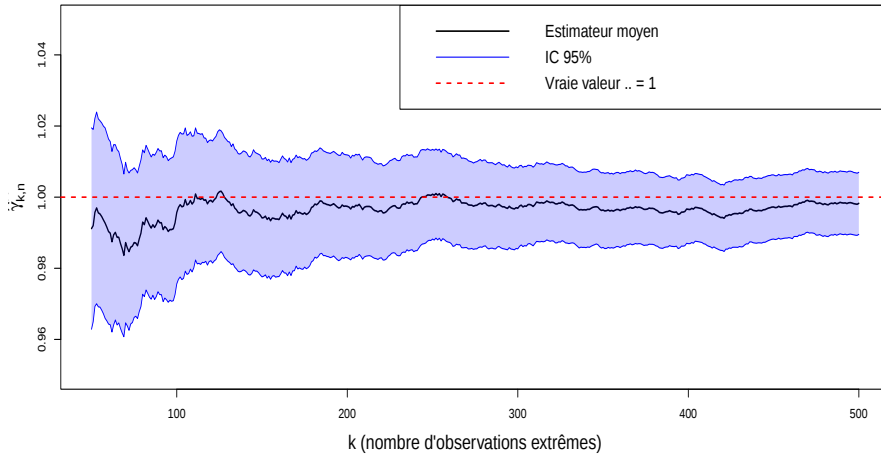


FIGURE 3.1 – Estimateur de Hill et l'intervalle de confiance de niveau 95% de l'indice des valeurs extrêmes pour la loi de Pareto Standard ($\gamma = 1$) basé sur 100 échantillons de 3000 observations.

graphe est très chaotique en particulier pour k_n petit (entre 0 et 20), autrement dit, l'estimateur fluctue davantage. Pour k élevé entre 300 et 500, à cause d'une variance faible avec un intervalle de confiance qui se resserre, l'estimateur de Hill devient de plus en plus performant montrant plus de précision.

3.3.2 Estimateur des Moments

Une méthode plus robuste constituant une extension de l'approche précédente a été développée par Dekkers et Haan [6], il s'agit des estimateurs des moments d'ordre élevés, utilisés quelque soit le signe du paramètre γ .

Définition 3.3 (Estimateur des moments).

Pour tout $\gamma \in \mathbb{R}$, l'estimateur des moments est défini par

$$\hat{\gamma}_n^M = M_n^{(1)} + 1 - \frac{1}{2} \left[1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right]^{-1}, \quad (3.7)$$

où

$$M_n^{(r)} = \frac{1}{k} \sum_{i=1}^k [\log X_{(n-i+1)} - \log X_{(n-k)}]^r, \quad \text{pour } r = 1, 2. \quad (3.8)$$

Théorème 3.3 (Propriétés asymptotiques de $\hat{\gamma}_n^M$).

[6] Soit $X_1, \dots, X_n$ un échantillon de n variables aléatoires i.i.d de fonction de répartition F appartenant au domaine d'attraction de la loi généralisée GEV (i.e $F \in D(H_\gamma)$), $\gamma \in \mathbb{R}$. Pour $k \rightarrow \infty$ et $\frac{k}{n} \rightarrow 0$, nous avons

(i) **Consistance faible**

$$\hat{\gamma}_n^M \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \gamma.$$

(ii) **Consistance forte** : Si $k/(\log n)^{\delta} \rightarrow \infty$ pour un certain $\delta > 0$, alors

$$\hat{\gamma}_n^M \xrightarrow[n \rightarrow +\infty]{p.s} \gamma.$$

(iii) **Normalité asymptotique**, pour $\gamma \geq 0$:

$$\sqrt{k} (\hat{\gamma}_n^M - \gamma) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \alpha^2),$$

$$\text{où } \alpha^2 = \begin{cases} 1 + \gamma^2, & \gamma \geq 0, \\ (1 - \gamma^2)(1 - 2\gamma) \left(4 - 8 \frac{1 - 2\gamma}{1 - 3\gamma} + \frac{(5 - 11\gamma)(1 - 2\gamma)}{(1 - 3\gamma)(1 - 4\gamma)} \right), & \gamma < 0. \end{cases}$$

La FIGURE 3.2 représente l'estimateur des moments $\hat{\gamma}_n^M$. Nous constatons que la convergence est stable vers la vraie valeur $\gamma = 0$ pour k entre 150 et 400. Contrairement à l'estimateur de Hill qui ne peut traiter que le cas $\gamma > 0$, le nouvel estimateur démontre sa robustesse dans l'estimation de l'indice de queue. En effet, pour une variance asymptotique faible, nous obtenons un resserrement progressif de l'intervalle de confiance avec l'augmentation de k .

3.3.3 Estimateur de Pickands

L'estimateur de Pickands, introduit par James Pickands, constitue une alternative importante aux estimateurs de Hill et des moments pour l'estimation de l'indice de queue γ . Contrairement aux deux derniers qui utilisent l'ensemble des statistiques d'ordre extrêmes, cet estimateur a l'avantage d'être applicable quelque soit le domaine d'attraction de la distribution H_γ , pour tout $\gamma \in \mathbb{R}$.

En finance, l'estimateur de Pickands trouve est utilisé dans l'estimation des queues de distributions de rendements. Aussi, on peut le trouver dans la modélisation des précipitations extrêmes, niveaux de crues, ou encore la vitesse des vents maximales.

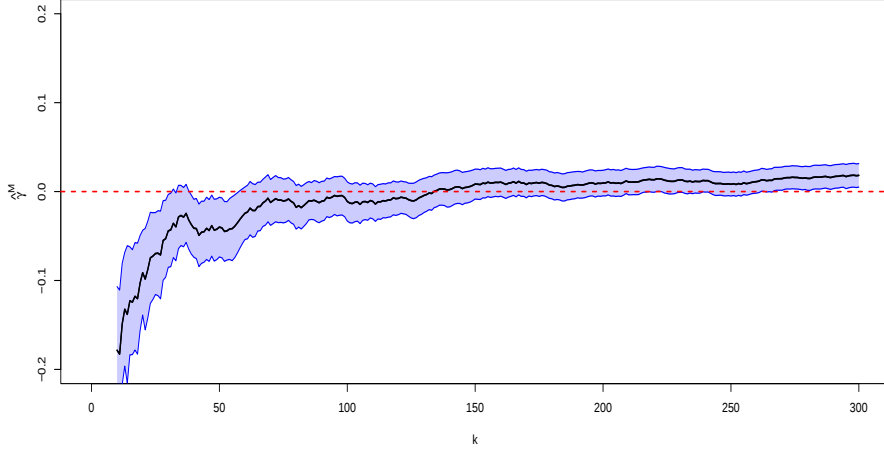


FIGURE 3.2 – Estimateur de Moment, en fonction du nombre des extrêmes (en trait plein) avec l'intervalle de confiance de niveau 95%, pour la distribution de Burr(1,1,1) basée sur 100 échantillons de 3000 observations.

Définition 3.4 (Estimateur de Pickands).

Pour un n -échantillon $X_1, \dots, X_n$ dev.a iid de loi F appartenant au domaine d'attraction de la loi généralisée GEV , i.e $F \in D(H_\gamma)$ avec $\gamma \in \mathbb{R}$. Soit le paramètre de troncature $k = (k_n)_{n \geq 1}$ une suite d'entiers telle que $1 < k < n$. L'estimateur de Pickands est défini par la statistique suivante

$$\hat{\gamma}_{k,n}^P := \frac{1}{\log 2} \log \left(\frac{X_{(n-k+1)} - X_{(n-2k+1)}}{X_{(n-2k+1)} - X_{(n-4k+1)}} \right) \quad (3.9)$$

Construction de l'estimateur de Pickands

Considérons le n -échantillon $X_1, \dots, X_n$ de variables aléatoires i.i.d de fdr $F \in D(H_\gamma)$ où $\gamma \in \mathbb{R}$. Soit U une fonction quantile de queue de la distribution F donnée par

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

Supposons que U est une fonction à variations régulières d'indice γ , ($U \in \mathcal{RV}_\gamma$). Alors

$$\lim_{t \rightarrow +\infty} \frac{U(zt)}{U(t)} = z^\gamma, \quad \forall z > 0. \quad (3.10)$$

Soit $X_{(n-k+1)}$ la k -ième plus grande valeur d'échantillon. Supposons l'existence d'une suite d'entiers $k = (k_n)_{n \geq 1}$ avec $1 < k < n$ telle que $k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$

quand $n \rightarrow \infty$, alors

$$\begin{aligned} X_{(n-k+1)} &\sim U\left(\frac{n}{k}\right) \\ X_{(n-2k+1)} &\sim U\left(\frac{n}{2k}\right) \\ X_{(n-4k+1)} &\sim U\left(\frac{n}{4k}\right). \end{aligned}$$

Remarquons que l'estimateur de Pickands utilise les niveaux k , $2k$ et $4k$, donc nous devons avoir $4k \leq n$. Posons $t = \frac{n}{4k}$ et par conséquent $t \rightarrow \infty$ puisque $\frac{k}{n} \rightarrow 0$. Les trois quantiles ci-dessus deviennent : $U\left(\frac{n}{k}\right) = U(4t)$, $U\left(\frac{n}{2k}\right) = U(2t)$ et $U\left(\frac{n}{4k}\right) = U(t)$. Ainsi

$$\frac{X_{(n-k+1)} - X_{(n-2k+1)}}{X_{(n-2k+1)} - X_{(n-4k+1)}} \sim \frac{U(4t) - U(2t)}{U(2t) - U(t)}.$$

Où encore

$$\lim_{t \rightarrow \infty} \frac{U(4t) - U(2t)}{U(2t) - U(t)} = \frac{X_{(n-k+1)} - X_{(n-2k+1)}}{X_{(n-2k+1)} - X_{(n-4k+1)}}. \quad (3.11)$$

D'autre part, d'après la relation 3.10, pour t assez grand

$$\begin{aligned} U(t) &\sim U(t), \\ U(2t) &\sim 2^\gamma U(t), \\ U(4t) &\sim 4^\gamma U(t), \end{aligned}$$

d'où :

$$\begin{aligned} \frac{U(4t) - U(2t)}{U(2t) - U(t)} &\sim \frac{4^\gamma U(t) - 2^\gamma U(t)}{2^\gamma U(t) - U(t)} \\ &\sim \frac{4^\gamma - 2^\gamma}{2^\gamma - 1} \\ &\sim \frac{2^\gamma (2^\gamma - 1)}{2^\gamma - 1} \\ &\sim 2^\gamma. \end{aligned}$$

Par suite

$$\lim_{t \rightarrow \infty} \frac{U(4t) - U(2t)}{U(2t) - U(t)} = 2^\gamma = \frac{X_{(n-k+1)} - X_{(n-2k+1)}}{X_{(n-2k+1)} - X_{(n-4k+1)}}.$$

En conclusion, l'estimateur de Pickands prendra la forme

$$\hat{\gamma}_{k,n} := \frac{1}{\log 2} \log \left(\frac{X_{(n-k+1)} - X_{(n-2k+1)}}{X_{(n-2k+1)} - X_{(n-4k+1)}} \right).$$

Le choix de l'entier k est crucial pour la performance des estimateurs en question. Pour des valeurs de k faibles, nous obtenons une variance importante et conduisant à une instabilité évidente. En revanche, pour des valeurs de k élevées en tenant compte de la condition $\frac{k}{n} \rightarrow 0$, nous obtenons un biais potentiel

avec une réduction de la variance. Cette configuration entraîne éventuellement une stabilité qui cache un biais important. Ceci est effectivement interprété par l'inclusion d'observations qui s'éloigne de la queue de distribution réelle. Plus le nombre de points impliqués est grand plus la sensibilité aux valeurs aberrantes modérées augmente. En pratique, une sélection de l'ordre de $k = n^{2/3}$ permettra un compromis satisfaisant minimisant les erreurs de l'estimation.

Propriétés asymptotiques de l'estimateur de Pickands

Théorème 3.4.

Soit $X_1, X_2, \dots, X_n$ un échantillon de variables aléatoires i.i.d de fonction de répartition F appartenant au domaine d'attraction de la loi généralisée des valeurs extrêmes GEV (i.e $F \in D(H_\gamma)$) et $\gamma \in \mathbb{R}$. Soit $(k_n)_{n \geq 1}$ une suite d'entiers telle que $1 < k_n \leq n$

(a) Consistance faible

Si $k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$ quand $n \rightarrow \infty$, alors l'estimateur de Pickands converge en probabilité vers γ . i.e.

$$\hat{\gamma}_{k_n, n}^P \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \gamma.$$

(b) Consistance forte

Si $\frac{k_n}{n} \rightarrow 0$ et $\frac{k_n}{\log(\log n)} \rightarrow \infty$ quand $n \rightarrow \infty$, alors l'estimateur de Pickands converge presque sûrement vers γ . i.e.

$$\hat{\gamma}_{k_n, n}^P \xrightarrow[n \rightarrow +\infty]{p.s.} \gamma.$$

(c) Normalité asymptotique

Sous certaines conditions sur k_n et F , nous avons

$$\sqrt{k} (\hat{\gamma}_{k_n, n}^P - \gamma) \xrightarrow[n \rightarrow +\infty]{en\ loi} \mathcal{N}\left(0, \frac{\gamma^2(2^{2\gamma+1} + 1)}{[2(2^\gamma - 1) \log 2]^2}\right).$$

Pour la preuve du Théorème 3.4 et pour plus de détails, nous pouvons consulter Pickands [13] pour la consistance faible, Dekkers et Haan [6] pour le résultat de la normalité asymptotique.

La FIGURE 3.3 montre qu'en moyenne, l'estimateur converge vers la vraie valeur $\gamma = -1$ pour un nombre de points extrêmes k élevé. Pour les petites valeurs de k , l'estimateur apparaît moins stable donnant des résultats moins fiables. Pour des valeurs de k trop élevées, nous pouvons obtenir un risque de biais, puisque les observations utilisées ne sont pas considérées comme des extrêmes. Par conséquent, le choix d'un k optimal est un compromis entre une variance élevée pour les petits k et un biais potentiel pour les grandes valeurs de k .

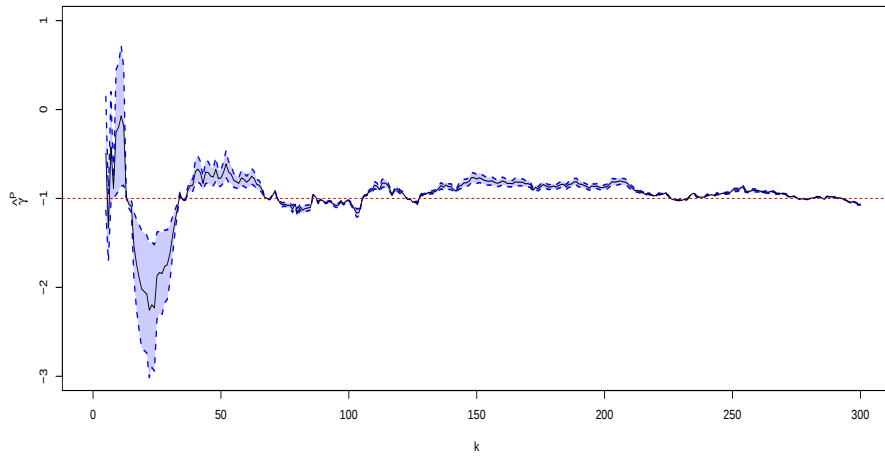


FIGURE 3.3 – Estimateur de Pickands, en fonction du nombre des extrêmes avec l'intervalle de confiance 95%, pour l'indice de valeurs extrêmes de la distribution uniforme avec $\gamma = 1$ basé sur 100 échantillons de 3000 observations.

Chapitre 4

Étude Numérique

4.1 Généralités

Dans le but d'étudier le comportement asymptotiques des estimateurs présentés dans la section précédente, et de comparer leurs performances, nous réalisons une étude numérique pour différentes tailles finies n d'échantillons.

Nous utilisons la loi de Fréchet des paramètres γ , δ et μ , de fonction de répartition :

$$F(x) = \exp\left(-\left(\frac{x-\mu}{\delta}\right)^{-\frac{1}{\gamma}}\right).$$

En prenant $\gamma = 0.25$, nous choisissons les valeurs initiales des paramètres comme suit $\delta_0 = 0.15$ et 0.75 , et $\mu_0 = 0.25$ et 0.5 . Prenons $n = 50, 100, 500, 1000$ et 1500 .

Nous générons $N = 4000$ échantillons indépendants de taille n , issus d'une loi $\mathbb{U}_{[0,1]}$, de sorte que si $Y \sim \mathbb{U}_{[0,1]}$ alors $X = \mu + \delta(-\log Y)^{-\gamma} \sim \text{Fréchet}(\gamma, \delta, \mu)$. En utilisant le logiciel R, nous estimons les deux paramètres δ et μ pour chacun des $N = 4000$ échantillons par la méthode du maximum de vraisemblance(MV) et ensuite par la méthode des moments(EM). La qualité de l'estimation des paramètres est mesurée par les quantités biais absolu ("Absolute Bias, (abs.bias)") et l'erreur quadratique moyenne ("Mean Squared Error, (MSE)") des estimateurs obtenu pour les différents scénarios.

Pour $\theta \in \{\mu, \delta\}$, nous avons,

$$abs.bias(\theta) = \frac{1}{N} \sum_{i=1}^N \left| \hat{\theta}_i - \theta_0 \right| \quad (4.1)$$

et

$$MSE(\theta) = \frac{1}{N} \sum_{i=1}^N \left(\hat{\theta}_i - \theta_0 \right)^2 \quad (4.2)$$

4.2 Simulation par la methode du Maximum Vraisemblance (MV)

Dans un premier temps, nous commençons par les simulations en utilisant la méthode du Maximum de Vraisemblance (MV). Nous fixons la valeur de $\gamma = 0.25$, et nous considérons les deux scénarios pour $\delta_0 = 0.15$ et 0.75 , ensuite $\mu_0 = 0.25$ et 0.5 . L'étude est réalisée pour les différentes tailles d'échantillons $n = 50, 100, 500, 1000$ et 1500 . Nous donnons les estimations de μ et δ ainsi que leur biais absolu (abs.bias) et leur erreur quadratique moyenne (MSE) pour chacune $N = 4000$ répliques. Les résultats des estimations et simulations sont présentés dans les tableaux ci-dessous.

N	$\hat{\mu}$	$\hat{\delta}$	abs.bias(μ)	abs.bias(δ)
$\mu_0 = 0.25, \delta_0 = 0.15$				
n = 50	0.2524972	0.1478822	2.497232e-03	2.117762e-03
n = 100	0.2510161	0.1491507	1.016139e-03	8.492740e-04
n = 500	0.2501638	0.1498522	1.638329e-04	1.478218e-04
n = 1000	0.2499623	0.1500694	3.771310e-05	6.942136e-05
n = 1500	0.2500176	0.1500133	1.760037e-05	1.325526e-05

TABLE 4.2.1 – Estimations de μ et δ et leur biais absolu pour $\mu_0 = 0.25$ et $\delta_0 = 0.15$ par la méthode de MV.

n	$\hat{\mu}$	$\hat{\delta}$	MSE(μ)	MSE(δ)
$(\mu_0 = 0.25, \delta_0 = 0.15)$				
50	0.2524972	0.1478822	2.984350e-04	4.065793e-04
100	0.2510161	0.1491507	2.233034e-04	3.082372e-04
500	0.2501638	0.1498522	2.776420e-05	3.903846e-05
1000	0.2499623	0.1500694	1.442748e-05	2.022479e-05
1500	0.2500176	0.1500133	9.293555e-06	1.296370e-05

TABLE 4.2.2 – Estimations de μ et δ et leur erreur quadratique moyenne pour $\mu_0 = 0.25$ et $\delta_0 = 0.15$ par la méthode de MV.

n	$\hat{\mu}$	$\hat{\delta}$	abs.bias(μ)	abs.bias(δ)
$(\mu_0 = 0.5, \delta_0 = 0.75)$				
50	0.5131837	0.7384529	0.0131836721	0.0115470901
100	0.5062027	0.7446293	0.0062027439	0.0053707058
500	0.5012548	0.7488027	0.0012548238	0.0011973452
1000	0.5004338	0.7495982	0.0004338180	0.0004018378
1500	0.5005852	0.7494793	0.0005851958	0.0005207188

TABLE 4.2.3 – Estimations de μ et δ et leur biais absolu pour $\mu_0 = 0.5$ et $\delta_0 = 0.75$ par la méthode de MV.

n	$\hat{\mu}$	$\hat{\delta}$	MSE(μ)	MSE(δ)
$(\mu_0 = 0.5, \delta_0 = 0.75)$				
50	0.5131837	0.7384529	0.0074598814	0.0101848072
100	0.5062027	0.7446293	0.0034890643	0.0048371895
500	0.5012548	0.7488027	0.0006879553	0.0009456080
1000	0.5004338	0.7495982	0.0003604897	0.0005041604
1500	0.5005852	0.7494793	0.0002322813	0.0003235809

TABLE 4.2.4 – Estimations de μ et δ et leur erreur quadratique moyenne pour $\mu_0 = 0.5$ et $\delta_0 = 0.75$ par la méthode de MV.

4.3 Simulation par la méthode des Moments (EM)

Nous réalisons cette fois-ci les simulations pour nos échantillons par la méthode des moments. Nous fixons toujours $\gamma = 0.25$, et nous utilisons les mêmes valeurs mentionnées plus haut des paramètres $\delta_0 = 0.15$ et 0.75 , et $\mu_0 = 0.25$ et 0.5 , pour différentes tailles d'échantillons $n = 50, 100, 500, 1000$ et 1500 . Nous présentons les estimations de μ et δ ainsi que leur biais absolu et leur erreur quadratique moyenne, chaque scénarios est répété $N = 4000$.

Les résultats des simulations sont regroupés dans les tableaux suivantes :

n	$\hat{\mu}$	$\hat{\delta}$	abs.bias(μ)	abs.bias(δ)
$(\mu_0 = 0.25, \delta_0 = 0.15)$				
50	0.2608777	0.1412415	0.010877733	0.0087585289
100	0.2565896	0.1447165	0.006589558	0.0052834931
500	0.2519875	0.1483803	0.001987472	0.0016197437
1000	0.2513587	0.1489142	0.001358702	0.0010857560
1500	0.2508394	0.1493278	0.000839397	0.0006721963

TABLE 4.3.1 – Estimations de μ et δ et leur biais absolu pour $\mu_0 = 0.25$ et $\delta_0 = 0.15$ par la méthode des moments.

n	$\hat{\mu}$	$\hat{\delta}$	abs.bias(μ)	abs.bias(δ)
$(\mu_0 = 0.5, \delta_0 = 0.75)$				
50	0.5543887	0.7062074	0.054388667	0.043792645
100	0.5329478	0.7235825	0.032947789	0.026417466
500	0.5099374	0.7419013	0.009937361	0.008098719
1000	0.5067935	0.7445712	0.006793508	0.005428780
1500	0.5041970	0.7466390	0.004196985	0.003360982

TABLE 4.3.2 – Estimations de μ et δ et calcul du biais absolu pour $\mu_0 = 0.5$ et $\delta_0 = 0.75$ par la méthode des moments.

n	$\hat{\mu}$	$\hat{\delta}$	MSE(μ)	MSE(δ)
$(\mu_0 = 0.5, \delta_0 = 0.75)$				
50	0.5543887	0.7062074	0.086461637	0.073288723
100	0.5329478	0.7235825	0.059561066	0.048327414
500	0.5099374	0.7419013	0.021804763	0.016340811
1000	0.5067935	0.7445712	0.012413498	0.009381866
1500	0.5041970	0.7466390	0.009561265	0.007126712

TABLE 4.3.3 – Estimations de μ et δ et le calcul des erreurs quadratiques moyennes pour $\mu_0 = 0.5$ et $\delta_0 = 0.75$ par la méthode des moments.

4.4 Comparaison et conclusion

Nous avons vu dans tous les tableaux ci-dessus que les biais absolus et les erreurs quadratiques moyennes pour chaque parametre diminuent vers zero quand n tend vers l'infini.

Afin de comparer les performances des deux estimateurs, nous allons comparer leurs erreurs quadratiques moyennes (MSE), pour chaque estimateur. Celui avec les plus petites valeurs de MSE est le meilleur estimateur.

N	$MSE(\mu)_{MV}$ $(\mu_0 = 0.12, \delta_0 = 0.45)$	$MSE(\mu)_{EM}$ $(\mu_0 = 0.12, \delta_0 = 0.45)$
n = 50	2.700929e-03	0.031126189
n = 100	1.261285e-03	0.021441984
n = 500	2.466597e-04	0.007849715
n = 1000	1.381230e-04	0.004468859
n = 1500	8.343801e-05	0.003442055

TABLE 4.4.1 – Erreur quadratique moyenne de μ , MSE(μ) par les deux estimateurs, MV et EM pour $\mu_0 = 0.12$ et $\delta_0 = 0.45$

N	$MSE(\delta)_{MV}$ $(\mu_0 = 0.85, \delta_0 = 0.05)$	$MSE(\delta)_{EM}$ $(\mu_0 = 0.85, \delta_0 = 0.05)$
n = 50	4.452490e-05	3.257277e-04
n = 100	2.096701e-05	2.147885e-04
n = 500	4.139318e-06	7.262583e-05
n = 1000	2.285686e-06	4.169718e-05
n = 1500	1.502546e-06	3.167428e-05

TABLE 4.4.2 – Erreur quadratique moyenne de δ , MSE(δ) par les deux estimateurs, MV et EM pour $\mu_0 = 0.12$ et $\delta_0 = 0.45$

À partir des deux tableaux ci-dessus, nous voyons clairement que pour tout taille $n \in \{50, 100, 500, 1000, 1500\}$;

$$MSE(\mu)_{MV} < MSE(\mu)_{EM} \quad \text{et} \quad MSE(\delta)_{MV} < MSE(\delta)_{EM}$$

Nous concluons alors que la méthode du Maximum Vraisemblance, (MV) est meilleure que la méthode des Moments(EM).

Conclusion Générale

L'étude des trois approches d'estimation de l'indice de queue γ met en évidence l'apport complémentaire des méthodes disponibles en théorie des valeurs extrêmes. Chaque estimateur présente des propriétés distinctes qui en font des outils appropriés à des contextes particuliers.

Pour les distributions à queues lourdes, l'estimateur de Hill reste crucial. Sa structure simple, son efficacité optimale et sa variance réduite nous permet de privilégier cet outil, notamment dans la modélisation des événements extrêmes en finance et des sciences environnementales. Néanmoins, cette perfection est limitée au où $\gamma > 0$ représentant ainsi une restriction importante.

L'estimateur des moments est plus flexible dans la mesure où l'approche couvre les valeurs $\gamma \in \mathbb{R}$. Cette qualité combinée à ces propriétés asymptotiques dans tous les domaines d'attraction, le rend un complément essentiel à l'estimateur de Hill. Avec une variance asymptotique importante (pour des $\gamma > 0$, cela n'empêche pas à gérer le cas où $\gamma < 0$ qui compense largement cette diminution de performance.

L'estimateur de Pickands se caractérise par sa robustesse inhérente, reposant seulement sur trois points de la queue empirique, permettant une efficacité face aux valeurs extrêmes, mais au coût d'une variance asymptotique considérablement importante. Sa capacité à agir sur le domaine $\gamma \in \mathbb{R}$ en fait un outil crucial pour l'analyse des résultats issus d'autres méthodes.

Bien que ces estimateurs aient des qualités remarquables, restent limités. La sensibilité au choix de k présente un défi pratique, nécessitant une analyse graphique dans la détermination des régions de stabilité.

Bibliographie

- [1] Arnold, Barry C., Balakrishnan, N. et Nagaraja, H.N., (1992). *A Örst course in order statistics*. Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York.
- [2] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., (2006). Statistics of Extremes- Theory and Applications. *Wiley*.
- [3] Cotrell, M., Genon-Catalot, V., Duhamel, C. et Meyre, T. (1999) *Exercices de probabilités*, Cassini.
- [4] De Haan, L. and Ferreira, A. (2006). Extreme Value Theory : An Introduction. *SpringerVerlag, New York*.
- [5] Deheuvels, P., Häeusler, E., and Mason, D. M. (1988). *Almost sure convergence of the Hill estimator*. Math. Proc. Cambridge Philos. Soc., 104(02) : 371–381.
- [6] Dekkers.A.L.M, and Haan.L, D. (1989). *On the estimation of the extreme-value index and largequantile estimation*. The Annals of Statistics. 17(4). 1795–1832.
- [7] Deme, E.H. (2013). Quelques contributions à la Théorie univariée des valeurs extrêmes et estimation des mesures de risque actuariel pourdes pertes à queues lourdes, Université Gaston Berger.
- [8] Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). Modelling Extremal Event for Insurance and Finance. *Springer, Berlin*.
- [9] Fisher, R., Tippet, L.(1928). *Limitting forms the frequency distribution of the largestor smallest member of a sample*. Proceedings of the Cambridge Philosophical Society,24 : 180–19.
- [10] Gnedenko, B. V. (1943). *Sur la distribution du terme maximum d’une serie aléatoire*. Ann. Math, 423-453.
- [11] Hill, B. M. (1975). *A simple general approach to inference about the tail of a distribution*. Ann. Statist., 3(5), 1163–1174.
- [12] Jenkinson, A. F. (1955). *The frequency distribution of the annual maximum (or minimum) values of meteorological elements*. Quarterly J. R. Methodol.Soc., 81(348), 158.

-
- [13] Pickands III, J. (1975). *Statistical inference using extreme order statistics*. Ann. Statist., 119–131.
 - [14] Reiss, R. D., and Thomas, M. (2007). Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields. *Birkhäuser, Basel*.
 - [15] Tassi, P. (1985). *Méthodes statistiques*, Economia.
 - [16] Von Mises R. (1954). *La distribution de la plus grande de n valeurs*. Amer. Math.Soc., 2 : 271–294.

ملخص

تشرح هذه الرسالة كيف تسمح نظرية القيم القصوى (TVE) بدراسة والتنبؤ بالأحداث نادرة الحدوث، مثل الأمطار الاستثنائية، العواصف الشديدة، أو الأزمات المالية. تبدأ بعرض مفاهيم بسيطة: كيف نصنف البيانات، كيف نحدد القيم الأكبر، ولماذا هي مهمة لفهم المخاطر. ثم يوضح العمل أنه من خلال تحليل القيم الأكثر تطرفًا، يمكن تجميعها في ثلاث عائلات رئيسية من السلوكيات الإحصائية. تسمح هذه الفكرة باستخدام نموذج يُسمى GEV، الذي يساعد على تقدير احتمال حدوث أحداث نادرة جدًا. كما تقدم الرسالة طريقة أخرى تُعرف باسم POT، التي تهتم بالزيادة عن حد معين وتستخدم التوزيع الباريتو المعمم.

جزء كبير من العمل يركز على كيفية تقدير مستوى النهاية لتوزيع بطريقة صحيحة من خلال عدة تقنيات، بعضها بسيط جدًا والبعض الآخر أكثر صلابة. وأخيرًا، وبفضل المحاكاة الحاسوبية، يقارن البحث هذه الأساليب ويظهر أن تقنية الاحتمالية القصوى تعطي عادة أدق النتائج. يساعد العمل بالكامل على فهم أفضل لكيفية تحليل المخاطر والتنبؤ بها المتعلقة بالظواهر القصوى في مجالات مختلفة.

RÉSUMÉ

Ce mémoire explique comment la Théorie des Valeurs Extrêmes (TVE) permet d'étudier et de prévoir des événements très rares, comme des pluies exceptionnelles, des tempêtes fortes ou des crises financières. Il commence par présenter des notions simples : comment on classe des données, comment on repère les valeurs les plus grandes, et pourquoi elles sont importantes pour comprendre les risques. Le travail montre ensuite qu'en analysant les valeurs les plus extrêmes, on peut les regrouper en trois grandes familles de comportements statistiques. Cette idée permet d'utiliser un modèle appelé GEV, qui aide à estimer la probabilité d'événements très rares. Le mémoire présente aussi une autre méthode, dite POT, qui s'intéresse aux dépassements au-delà d'un certain seuil et utilise la distribution de Pareto généralisée.

Une grande partie du travail porte sur la manière d'estimer correctement le niveau d'extrémité d'une distribution grâce à plusieurs techniques, dont certaines très simples et d'autres plus robustes. Enfin, grâce à des simulations informatiques, le mémoire compare ces méthodes et montre que la technique du maximum de vraisemblance donne en général les résultats les plus précis. L'ensemble du travail aide à mieux comprendre comment analyser et prédire les risques liés aux phénomènes extrêmes dans différents domaines.

ABSTRACT

This thesis explains how the Extreme Values Theory (EVT) allows studying and predicting very rare events, such as exceptional rainfall, severe storms, or financial crises. It begins by presenting simple concepts: how data is classified, how the largest values are identified, and why they are important for understanding risks. The work then shows that by analyzing the most extreme values, they can be grouped into three major families of statistical behavior. This idea allows the use of a model called GEV, which helps estimate the probability of very rare events. The thesis also presents another method, called POT, which focuses on exceedances beyond a certain threshold and uses the generalized Pareto distribution.

A large part of the work focuses on how to correctly estimate the tail level of a distribution using several techniques, some very simple and others more robust. Finally, through computer simulations, the thesis compares these methods and shows that the maximum likelihood technique generally provides the most accurate results. The entire work helps to better understand how to analyze and predict risks related to extreme phenomena in different fields.