
République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider, Biskra



Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département de Mathématiques

Mémoire présenté en vue de l'obtention du
DIPLÔME DE MASTER EN MATHÉMATIQUES

Option : **Statistique**

Par
Mokrani Ahmed

Thème

Estimation de l'Indice des Valeurs Extrêmes sous Troncature

Membres du Comité d'Examen

Pr.	Yahia Djabrane	UMKB	Président
Dr.	Soltane Louiza	UMKB	Encadreur
MAB.	Roubi Afaf	UMKB	Examineur

Septembre 2020

À ma mère pour sa générosité, et son soutien permanent,

À mon père bien-aimé,

À mes frères et mes soeurs.

REMERCIEMENTS

Je tiens à remercier ALLAH le tout-puissant qui m'a aidée et donnée la santé, la patience et le courage durant ces longues années d'études.

Mes plus vifs remerciements vont à mon encadreur Dr. Louiza Soltane pour sa disponibilité, ses encouragements, ses précieux conseils, son sens de l'écoute et pour toute l'aide qu'elles m'ont apporté durant toute la période du travail.

Je remercie également tous les enseignants et employés du département de Mathématiques présidé par Monsieur Mokhtar Hafayedh, avec une mention spéciale à Messieurs les membres du jury "le Pr. Djabrane Yahia et le Dr. Afaf Roubi" aient accepté la si difficile tâche d'examineur. Merci sincèrement du temps et de l'énergie que vous avez consacrés à la lecture de mon travail.

Je tiens aussi à remercier ma famille mes collègues et amis, grâce à qui ma vie universitaire a été très agréable et joyeuse.

Enfin, j'adresse mes remerciements à tous ceux qui ont contribué de près ou de loin à la réalisation de ce mémoire.

TABLE DES MATIÈRES

Dédicaces	2
Remerciements	3
Table des Matières	4
Liste des Figures	6
Liste des Tableaux	7
Introduction	1
1 Quelques Rappels et Définitions sur la Théorie des Valeurs Extrêmes	4
1.1 Définitions de base	5
1.1.1 Fonctions de répartition, de queue et du quantile	5
1.1.2 Fonctions empiriques de répartition, de queue et du quintile	6
1.1.3 Convergences des suites de variables aléatoires	7
1.2 Statistique d'ordre	8
1.3 Comportement asymptotique des extrêmes	10
1.3.1 Distributions à queues lourdes	11
1.3.2 Domaines d'attraction	12
2 Estimation de l'Indice des Valeurs Extrêmes	15
2.1 Données incomplètes	15

2.1.1	Données censurées	16
2.1.2	Données troncatures	18
2.1.3	Estimateur de Lynden-Bell	20
2.2	Estimation de l'IVE sous données complètes	21
2.2.1	Estimateur de Pickands	21
2.2.2	Estimateur de Hill	22
2.2.3	Estimateur des Moments	23
2.3	Estimation de l'IVE sous données incomplètes	25
2.3.1	Estimation sous censure aléatoire à droite	25
2.3.2	Estimation sous troncature aléatoire à droite	25
2.3.3	Application	28
	Conclusion	31
	Bibliographie	32
	Annexe B : Abréviations et Notations	35

TABLE DES FIGURES

1.1	Fonction Empirique de Répartition (Droite) et de Survie (Gauche) d'un Echantillon Gaussien Standard de Taille 200.	7
1.2	Comparaison du Comportement de la Queue	12
1.3	Densités de Lois des Valeurs Extrêmes avec $1/ \gamma $, $\alpha > 0$	14
2.1	Estimateur de Pickands, avec un intervalle de confiance de niveau 95%, pour l'EVI de la distribution de Pareto standard ($\gamma = -1$) basé sur 1000 échantillons de 5000 observations.	22
2.2	Estimateur de Hill, avec un intervalle de confiance de niveau 95%, pour l'EVI de la distribution de Pareto standard ($\gamma = 1$) basé sur 1000 échantillons de 5000 observations.	23
2.3	Estimateur de Moments, avec un intervalle de confiance de niveau 95%, pour l'EVI de la distribution de Gumbel standard ($\gamma = 0$) basé sur 1000 échantillons de 5000 observations.	24

LISTE DES TABLEAUX

1.1	Quelques distributions associées à un indice positif	13
1.2	Quelques distributions associées à un indice négatif	13
1.3	Quelques distributions associées à un indice nul	14
2.1	Biases and RMSE's de l'estimateur de Benchaira et al. (2015) (panneau de gauche) et celui de Gardes et Stupfler (2014) (panneau de droite) de l'IVE $\gamma_1 = 0.6$, basée sur 1000, échantillons de modèles de Burr	29
2.2	Biases and RMSE's de l'estimateur de Benchaira et al. (2015) (panneau de gauche) et celui de Gardes et Stupfler (2014) (panneau de droite) de l'IVE $\gamma_1 = 0.8$, basée sur 1000, échantillons de modèles de Burr	30

INTRODUCTION

La théorie des valeurs extrêmes (TVE) représente un outil approprié permettant d'extrapoler le comportement des queues de distributions à partir des plus grandes (ou plus petites) valeurs observées. Sur le plan statistique, l'étude permet de fournir des outils probabilistes et statistiques qui permettent de modéliser pour prévoir l'occurrence des événements rares. Ces événements rares sont des phénomènes aléatoires qui ont une faible probabilité. On peut citer entre autres :

- La météorologie : pour l'étude de vitesse de vent, extrême pluviométrique, des températures, ...
- En hydrologie : pour l'étude la probabilité que la hauteur d'eau d'un fleuve dépasse un certain seuil (Voir [Guillou et al \[18\]](#)).
- La finance (Marchés financiers) : Pour l'étude quantitative des boums et Krachs boursiers, qui se traduisent par de fortes variations des cours financiers
- L'assurance et la Réassurance : Pour l'étude des risques graves ou rares afin d'avoir une certaine stabilité des indicateurs qui traduit une bonne adéquation entre la sinistralité et la tarification (voir [Embrechts et al \[13\]](#)).

La théorie des valeurs extrêmes, fondée sur des résultats de la théorie des probabilités, offre un cadre mathématique rigoureux pour l'estimation des probabilités d'événements rares. Il s'agit fondamentalement de modéliser un phénomène aléatoire, en s'intéressant principalement aux quantiles extrêmes et à la queue de distribution souvent modélisée par un indice appelé indice des valeurs extrêmes.

L'utilisation des lois des valeurs extrêmes repose sur des propriétés des statistiques d'ordre et sur des méthodes d'extrapolation. Plus précisément, elle repose sur la convergence des maxima ou des minima des variables aléatoires indépendantes et identiquement distribuées, convenablement renormalisées. C'est dire donc qu'on étudie le comportement asymptotique des lois des extrêmes. Ces lois, appelées lois des valeurs extrêmes sont bien connues et elles sont de trois types : Fréchet, Gumbel, et Weibull.

Les lois des valeurs extrêmes, lorsqu'elle existe, sont indexées par l'indice des valeurs extrêmes (IVE). La connaissance de cet indice est un élément important car il contrôle la "lourdeur" de la queue de distribution. Ainsi de nombreux estimateurs de l'indice des valeurs extrêmes ont été proposé dans la littérature (Estimateur d'Hill, Pikands, ...). L'estimation de cet indice dépend largement de nombre de statistiques d'ordre extrêmes observées. Ce nombre détermine les valeurs qui sont réellement extrêmes.

L'analyse de survie est un domaine des statistiques qui trouve sa place dans tous les champs d'applications où l'on étudie la survenue d'un évènement. L'objectif de cette analyse réside dans l'analyse du délai de survie d'un évènement dans un ou plusieurs groupes d'individus. Dans le domaine biomédical, par exemple, plusieurs évènements sont intéressants à étudier :

- Le développement d'une maladie.
- La réponse à un traitement donné.
- La rechute d'une maladie ou décès.

Une des caractéristiques des données de survie est l'existence d'observation incomplète. La censure et la troncature font partie du processus générant de ce type de données.

Notamment au cours de la dernière décennie, de nombreux auteurs ont commencé à prêter attention à l'estimation de l'indice de queue sous données incomplètes : sous censure aléatoire à droite on trouve [Beirlant et al \[2\]](#), [Einmahl et al. \[12\]](#), [Brahimi et al. \[4\]](#) ...etc et sous troncature aléatoire à droite on trouve, telle que [Gardes et Stupfler \[16\]](#), [Benchaira et al. \[3\]](#) ... etc.

Ce mémoire est composé de deux chapitres :

Le premier Chapitre se regroupe en trois Sections, on débute avec quelques rappels et définitions sur les fonctions de répartition, de queue et du quantile et empiriques de répartition et convergences des suites de variables aléatoires. Après cela on parle sur les statistiques d'ordre. La dernière Section qui se compose de deux parties, la première partie on donne le résultat fondamental de la théorie des valeurs extrêmes, qui est le théorème de Fisher-Tippett (1928)

et Gnedenko (1943), et la deuxième partie on discute la notion des domaines d'attraction d'une distribution selon le paramètre de l'indice de queue.

Le deuxième Chapitre se compose de trois Sections. La première Section on présente deux cas de données incomplètes : censurées et tronquées et ses types (droite, gauche et par intervalle). Après cela dans la deuxième Section qui fournit les estimateurs classiques de l'IVE tels l'estimateur de Hill, de Pickands et de Moments dans le cadre de données complètes. La dernière Section on parle sur l'estimation de l'IVE sous données incomplètes et on s'intéresse sur l'estimation de l'IVE en présence de données tronquées aléatoirement à droite et fait une comparaison entre l'estimateur de Gardes et Stupfler et l'estimateur de Benchaira et al. dans la partie de l'application à l'aide du logiciel d'analyse statistique R ([The R Project for Statistical Computing](#)).

CHAPITRE 1

Quelques Rappels et Définitions sur la Théorie des Valeurs Extrêmes

La théorie des valeurs extrêmes communément appelée « Extreme Value Theory » (EVT) en anglais, est une vaste théorie dont le but est d'étudier les événements rares c'est-à-dire les événements dont la probabilité d'apparition est faible. Cette théorie est appliquée en :

- Hydrologie : pour prévoir les crues.
- Démographie : pour prévoir la distribution de probabilité de l'âge maximum que l'être humain pourra atteindre.
- Assurance, finance ou en météorologie : pour prévoir les grands sinistres.

En d'autres termes, on veut estimer des petites probabilités ou des quantités dont la probabilité d'observation est très faible c'est-à-dire proche de zéro. Ces quantités sont appelées quantiles extrêmes car l'ordre de ces quantiles tend vers zéro lorsque la taille de l'échantillon n tend vers l'infini. Pour plus de détails sur ce thème voir [Beirlant et al. \[2\]](#), [Reiss et Thomas \[27\]](#), [Novak \[24\]](#) et [Cláudia Neves \[5\]](#) comme des bonnes références classiques sur la théorie et les applications des valeurs extrêmes.

L'objectif essentiel de ce chapitre est de présenter les définitions et les résultats principaux sur la TVE dans le cas unidimensionnel.

1.1 Définitions de base

On commence dans un premier temps par donner les notions de base et les propriétés des statistiques d'ordre dans lesquelles repose l'utilisation des lois des valeurs extrêmes.

1.1.1 Fonctions de répartition, de queue et du quantile

Définition 1.1.1 (Variable aléatoire)

Une variable aléatoire (va) réelle X est une application mesurable de (Ω, \mathcal{F}, P) dans \mathbb{R} muni de sa tribu borélienne $(\mathbb{R}, \mathcal{B})$.

Définition 1.1.2 (Loi de probabilité)

On appelle loi de probabilité de X la mesure image de P par X et on la note P_X .

Définition 1.1.3 (Fonction de répartition)

La fonction de répartition (fdr) d'une va X est l'application F de \mathbb{R} dans $[0, 1]$ définie par :

$$F(x) := P(X \leq x). \quad (1.1)$$

F est aussi appelé fonction de distribution (fd).

Propriété 1.1.1 (Sans démonstration)

La fdr F d'une va X est monotone croissante continue à droite et vérifie

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

Définition 1.1.4 (Fonction de queue)

La fonction de queue, qu'on note par $\bar{F}(x)$ (ou $S(x)$) est définie sur \mathbb{R} par :

$$\bar{F}(x) = 1 - F(x) := P(X > x). \quad (1.2)$$

$\bar{F}(x)$ est aussi appelé fonction de survie.

Propriété 1.1.2 (Sans démonstration)

La fonction de survie d'une va X est décroissante monotone continue à gauche et vérifie

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

On parle aussi souvent du support de X , c'est l'ensemble de toutes les valeurs de x pour lesquelles la fdr est strictement positif. Le point terminal inférieur (ou le point terminal gauche) d'une distribution F est désigné par x_F c'est-à-dire,

$$x_F := \inf \{x \in \mathbb{R} : F(x) > 0\}.$$

De manière équivalente, on définit le point terminal supérieur (ou droit) de F par

$$x^F := \sup \{x \in \mathbb{R} : F(x) < 1\}. \quad (1.3)$$

Voir Embrechts et al. (1997), Exemple 3.3.22, p.139).

Définition 1.1.5 (Fonction du quantile)

Soit X une va et F sa fdr, la fonction du quantile est définie pour toute $0 < s < 1$ par :

$$Q(s) := \inf \{x : F(x) \geq s\} = F^{-1}(s), \quad (1.4)$$

où F^{-1} désignant l'inverse généralisé de F . On l'exprime en termes de la fonction de queue par :

$$Q(s) := \inf \{x : \bar{F}(x) \leq 1 - s\} = F^{-1}(s), \quad 0 < s < 1. \quad (1.5)$$

Remarque 1.1.1

1. Le quintile d'ordre q , où $q \in]0, 1[$, est défini par $x_q = F^{-1}(q)$.
2. Une fonction parfois appelée fonction du quantile de queue, notée par U , est définie par :

$$U(s) := F^{-1}(1 - 1/s) = (1/\bar{F})^{-1}(s), \quad s \geq 1, \quad (1.6)$$

On note que le point terminal supérieur peut donc être considéré comme le quantile ultime dans le sens où $x^F = U(\infty) := \lim_{s \rightarrow \infty} U(s) \leq \infty$.

1.1.2 Fonctions empiriques de répartition, de queue et du quintile

Définition 1.1.6 (Fonctions empiriques de répartition et de survie)

Soit X_1, \dots, X_n un échantillon de taille $n \geq 1$ d'une va positive X de fdr F et de fonction de survie S . Les fonctions empiriques de répartition et de survie, F_n et \bar{F}_n sont respectivement définies par

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i \leq x\}, \quad \forall x \in \mathbb{R}, \quad (1.7)$$

et

$$\bar{F}_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{I}\{X_i > x\}, \quad \forall x \in \mathbb{R}, \quad (1.8)$$

où $\mathbb{I}\{A\}$ est la fonction indicatrice de l'ensemble A .

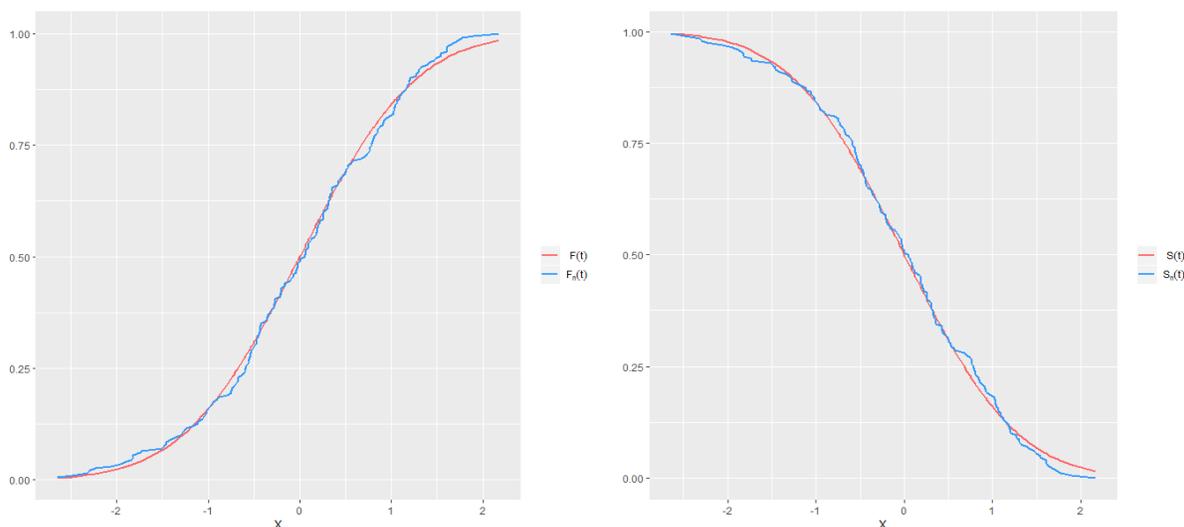


FIG. 1.1 – Fonction Empirique de Répartition (Droite) et de Survie (Gauche) d'un Echantillon Gaussien Standard de Taille 200.

Définition 1.1.7 (Fonction quantile empirique)

La fonction du quantil empirique de l'échantillon X_1, \dots, X_n est définie par :

$$Q_n(s) = F_n^{-1}(s) := \inf \{x : F_n(x) \geq s\}, \quad 0 < s < 1. \quad (1.9)$$

ou

$$Q_n(s) = \bar{F}_n^{-1}(1 - s) = \inf \{x : \bar{F}_n(x) \leq (1 - s)\}, \quad 0 < s < 1.$$

Remarque 1.1.2

La fonction empirique du quantile de queue est définie par :

$$U_n(s) = Q_n(1 - 1/s), \quad s \geq 1.$$

1.1.3 Convergences des suites de variables aléatoires

Une suite X_1, \dots, X_n de va's étant une suite de fonctions de Ω dans \mathbb{R} , il existe diverses façons de définir la convergence de X_1, \dots, X_n dont certaines jouent un grand rôle en calcul

des probabilités. Avant de passer à la théorie asymptotique (ou limite) des extrêmes, on doit nous familiariser avec les concepts suivants de convergence :

- On dit qu'une suite de va's X_1, \dots, X_n converge en distribution vers une variable aléatoire X , on la note $X_n \xrightarrow{d} X$, si pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

Ceci est également connu sous le nom de convergence faible.

- La suite X_1, \dots, X_n converge en probabilité vers X , on la note $X_n \xrightarrow{P} X$, si pour tout $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} (|X_n - X| > \epsilon) = 0.$$

- La suite X_1, \dots, X_n converge presque sûrement vers X , on la note $X_n \xrightarrow{p.s.} X$, si

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

La convergence presque sûre est également appelée convergence forte.

La convergence de F_n vers F est presque sûrement uniforme, c'est-à-dire que :

$$\sup_x |F_n(x) - F(x)| \xrightarrow{p.s.} 0 \text{ quand } n \rightarrow \infty. \quad (1.10)$$

La convergence (1.10) est connue sous le nom de théorème de Glivenko-Cantelli. Il est l'un des résultats fondamentaux en statistiques non-paramétriques. La preuve du résultat (1.10) peut être trouvée dans tout manuel standard de la théorie des probabilités comme (Pierre [26], page 229).

1.2 Statistique d'ordre

Soit X_1, \dots, X_n une suite de va's indépendantes identiquement distribuées (iid) et de même fdr F , on définit les statistiques d'ordres dans les suivant

Définition 1.2.1 (Statistique d'ordre)

Les statistiques d'ordres sont les observations ordonnées par ordre croissante, on écrit :

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

$X_{i:n}$ représente la $i^{\text{ème}}$ statistique d'ordre (statistique d'ordre i) dans un échantillon de

taille n . et les extrêmes sont comme termes du minimum et du maximum de l'échantillon X_1, \dots, X_n , tel que la variable $X_{1:n}$ est la plus petite valeur observée (où statistique du minimum) et $X_{n:n}$ est la plus grande statistique d'ordre (où statistique du maximum) qui sont données par :

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

et

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

On donne dans la suite l'expression de la distribution du maximum $X_{n:n}$ et du minimum $X_{1:n}$ par les deux formules suivantes :

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

En effet,

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

et

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

David [8] et Balakrishnan et Cohn [1] montre que l'expression de la distribution de $X_{i:n}$ est

$$F_{X_{i:n}} = P\{X_{i:n} \leq x\} = \sum_{r=i}^n \binom{n}{r} (F(x))^r (1 - F(x))^{n-r}.$$

Remarque 1.2.1

Les résultats correspondant pour les minima sont facilement accessibles en utilisant l'égalité suivante :

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

1.3 Comportement asymptotique des extrêmes

La théorie des valeurs extrêmes de suites de variables iid dans \mathbb{R} commence par la connaissance de la loi asymptotique de la suite des maxima $X_{n:n}$ quand n tend vers l'infini sous l'hypothèse que les observations sont iid. Il faut donc une information sur la forme de la distribution en étudiant la loi du maximum. Remarque d'abord

$$\lim_{n \rightarrow \infty} F_{X_{n:n}}(x) = \lim_{n \rightarrow \infty} [F(x)]^n = \begin{cases} 1 & x \geq x^F \\ 0 & x < x^F. \end{cases} \tag{1.11}$$

On constate que la distribution asymptotique du maximum donne une loi dégénérée, une masse de Dirac en x^F , puisque pour certaines valeurs de x , la probabilité peut être égale à 1 dans le cas où x^F est fini et donc $X_{n:n}$ tend vers x^F presque sûrement, avec $x^F \leq \infty$. Ce fait ne fournit pas assez d'informations, d'où l'idée d'utiliser une transformation afin d'obtenir des résultats plus exploitables pour les lois limites des maxima. On s'intéresse par conséquent à une loi non dégénérée pour le maximum, la TVE permet de donner une réponse à cette problématique. Les premiers résultats sur la caractérisation du comportement asymptotique des maxima $X_{n:n}$ convenablement normalisés ont été obtenus par [Fisher et Tippet \[15\]](#), [Gnedenko \[17\]](#) Ce résultat est analogue au Théorème Central Limite (TCL)¹ mais pour les phénomènes extrêmes. Voici le théorème qui donne la loi du maximum.

Théoreme 1.3.1 (Fisher et Tippet (1928), Gnedenko (1943))

Soit $(X_i)_{i \geq 1}$ une suite de va's iid de fonction de répartition F . S'il existe deux constantes de normalisation $a_n > 0$ et $b_n \in \mathbb{R}$, telle que :

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{n:n} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x - b_n) = g(x), \quad \forall x \in \mathbb{R}, \tag{1.12}$$

la limite converge en distribution vers une loi non dégénérée $g(x)$. Cette loi de la limite que l'un des trois types de lois suivantes :

$$\begin{aligned} \text{Fréchet} : \quad \Phi_\alpha(x) &= \exp\{-x^{-\alpha}\}, & x > 0 \\ \text{Gumbel} : \quad \Lambda(x) &= \exp\{-e^{-x}\}, & x \in \mathbb{R} \\ \text{Weibull} : \quad \Psi_\alpha(x) &= \exp\{-(-x)^{-\alpha}\}, & x \leq 0. \end{aligned} \tag{1.13}$$

Ces trois distributions limites sont appelées les distributions de valeurs extrêmes standard.

¹TCL : établit la convergence en loi vers la loi normale d'une somme de va. iid. sous des hypothèses très peu contraignantes.

Une démonstration détaillée de ce théorème est donnée dans [Resnick \[28\]](#) et avec des développements dans [Embrechts et al., \[13\]](#) page 152.

Il est possible de rassembler les trois familles de lois en une seule famille paramétrique (g_γ , $\gamma \in R$) de [Von Mises-Jenkinson \[\[29\], \[20\]\]](#), elle est dite famille des lois des valeurs extrêmes généralisées (GEV, Generalized Extreme Value distribution). Notamment, la distribution des valeurs extrêmes généralisées donnée par :

$$g_\gamma(x) = \begin{cases} \exp(-(1 - \gamma x)^{-1/\gamma}) & \gamma \neq 0, \quad 1 + \gamma x > 0 \\ \exp(-e^{-x}) & \gamma = 0, \end{cases} \quad (1.14)$$

Le paramètre $\alpha = 1/\gamma$ est appelé indice de queue ou indice de valeur extrême (IVE) si $\gamma > 0$. Dans la littérature, l'IVE est également appelé paramètre de forme du GEV.

1.3.1 Distributions à queues lourdes

Les distributions à queues lourdes jouent un rôle important dans la théorie des valeurs extrêmes. Elles ont été acceptées comme des modèles appropriés de divers phénomènes on peut citer dans le cas de montage de grand sinistre en assurance les fluctuations des prix en finance, ... etc. Mathématiquement, ce type des distributions est défini ainsi :

Définition 1.3.1 (Distribution à queue lourde)

Soit X une va de fdr F , donc cette dernière elle est dite distribution à queue lourde, s'il existe un constant positif γ qui représente l'indice de queue et prend la formule suivante :

$$\bar{F}(x) \sim x^{-1/\gamma} l(x), \text{ pour } x \rightarrow \infty, \quad (1.15)$$

où $l(x)$ la fonction à variation lente au voisinage de l'infini,

Ce type de distribution satisfaite pour tout $x > 0$, la condition suivante :

Définition 1.3.2 (Condition du 1^{er} ordre)

\bar{F} est dite variation régulière à l'infini d'indice $-1/\gamma < 0$, on a

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-1/\gamma}, \quad x > 0. \quad (1.16)$$

mais, en général cette condition n'est pas suffisante pour étudier les propriétés des estimateurs, en particulier la normaliser asymptotique. Dans ce cas, une condition du second ordre

des fonctions à variations régulières est nécessaire en spécifiant le taux de convergence dans l'Equation 1.16. La définition suivante de cette condition vient de de Haan et Ferreira [11], page 48.

Définition 1.3.3 (Condition du 2^{ème} ordre)

$\exists \rho \leq 0$ et une fonction $A \rightarrow 0$ et ne change pas le signe au voisinage de l'infini, tel que

$$\lim_{t \rightarrow \infty} \frac{\overline{F}(tx)/\overline{F}(t) - x^{-1/\gamma}}{A(t)} = x^{-1/\gamma} \frac{x^\rho - 1}{\gamma\rho}. \tag{1.17}$$

La famille de distribution de queue lourde, elle se caractérise par une décroissance lente vers zéro par rapport à la distribution exponentielle, comme le montre la figure suivante :

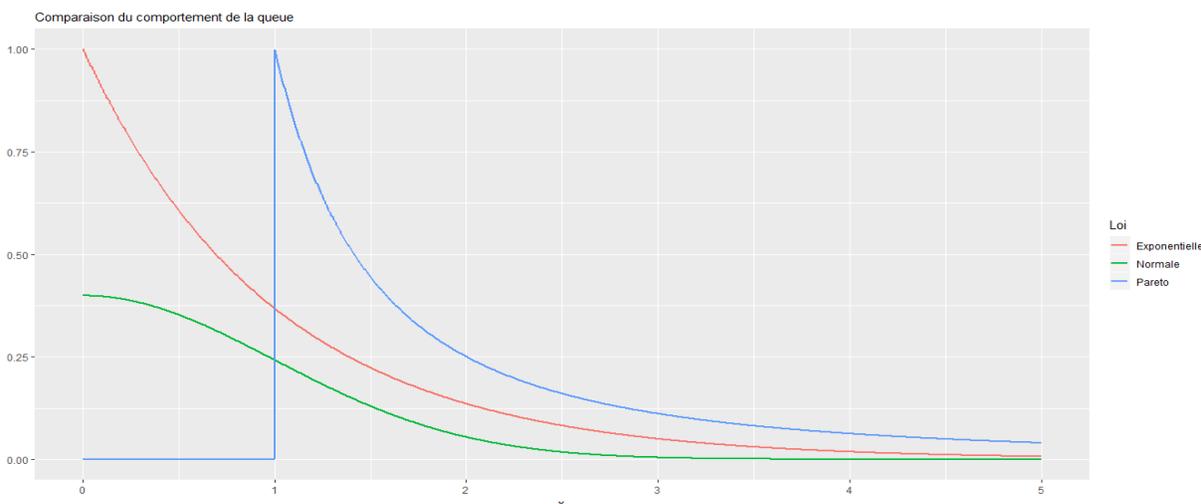


FIG. 1.2 – Comparaison du Comportement de la Queue

On a ici les distributions à queue lourdes, elles sont représentées par la courbe verte elle est décroît lentement vers zéro par rapport la distribution exponentielle et la distribution exponentielle que représentée par la courbe rouge et la distribution à queue légère représentée par la courbe bleue.

1.3.2 Domaines d'attraction

Définition 1.3.4 (domaines d'attraction)

On dit qu'une distribution F appartient au domaine d'attraction de g_γ , et on note $F \in \mathcal{D}(g_\gamma)$, s'il existe deux suites normalisantes $a_n > 0$ et b_n fourni dans l'Equation 1.12 de telle manière

que

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = g_\gamma(x), \quad (1.18)$$

avec g_γ donnée par l'Equation 1.14

Selon le signe de γ , on distingue trois domaines d'attraction :

- Si $\gamma > 0$, on dit que $F \in \mathcal{D}(\Phi_\alpha)$, et F a un point terminal à droite infinie ($x_F = +\infty$). Ce domaine d'attraction est celui des distributions à queues lourdes, c'est-à-dire qui ont une fonction de survie à décroissance polynomiale. Le tableau suivant présente quelques lois qui appartiennent au domaine d'attraction de Fréchet

Distributions	$\bar{F}(x)$	γ
<i>Pareto</i> (α) $\alpha > 0$	$x^{-\alpha}, x > 1$	$\frac{1}{\alpha}$
<i>Burr</i> (β, τ, λ), $\beta > 0, \tau > 0, \lambda > 0$	$\left(\frac{\beta}{\beta + x^\tau}\right)^\lambda$	$\frac{1}{\lambda\tau}$
Fréchet $\left(\frac{1}{\alpha}\right)$, $\alpha > 0$	$1 - \exp(-x^{-\alpha})$	$\frac{1}{\alpha}$
<i>Loggamma</i> (m, λ), $m > 0, \lambda > 0$	$\frac{\lambda^m}{\Gamma(m)} \int_x^\infty (\log u)^{m-1} u^{-\lambda-1} du$	$\frac{1}{\lambda}$

TAB. 1.1 – Quelques distributions associées à un indice positif

- Si $\gamma < 0$, on dit que $F \in \mathcal{D}(\Psi_\alpha)$, et F a un point terminal à droite finie ($x_F < +\infty$). Ce domaine d'attraction est celui des fonctions de survie dont le support est borné supérieurement. Le tableau suivant présente quelques lois qui appartiennent au domaine d'attraction de Weibull.

Distributions	$\bar{F}(x)$	γ
<i>Uniform</i> (0, 1)	$1 - x$	-1
<i>Reverse Burr</i> ($\beta, \tau, \lambda, x_\tau$), $\beta, \tau, \lambda > 0$	$\left(\frac{\beta}{\beta + (x_F + x)^{-\tau}}\right)^\lambda$	$-\frac{1}{\lambda\tau}$
<i>Weibull</i>	$\exp(-x^{-\alpha})$	$-\frac{1}{\alpha}$

TAB. 1.2 – Quelques distributions associées à un indice négatif

- Si $\gamma = 0$, on dit que $F \in \mathcal{D}(\Lambda)$, le point terminal x_F peut alors être fini ou non. Ce domaine d'attraction est celui des distributions à queues légères, c'est-à-dire qui ont une fonction de survie à décroissance exponentielle. Le tableau suivant fournit quelques exemples de lois qui appartiennent au domaine d'attraction de Gumbel.

Distributions	$\overline{F}(x)$	γ
$Gamma(m, \lambda), m \in \mathbb{N}, \lambda > 0$	$\frac{\lambda^m}{\Gamma(m)} \int_x^\infty u^{m-1} \exp(-\lambda u) du$	0
$Gumbel(\mu, \beta), \mu \in \mathbb{R}, \beta > 0$	$\exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right)$	0
logistic	$\frac{2}{1 + \exp(x)}$	0
$Lognormale(\mu, \sigma), \mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sqrt{2\pi}} \int_1^\infty \frac{1}{u} \exp\left(-\frac{1}{2\sigma^2} (\log u - \mu)^2\right) du$	0

TAB. 1.3 – Quelques distributions associées à un indice nul

Ebrechts et al. [13], dans les tableaux 3.4.2 – 3.4.4, ils ont donné différents exemples de distributions standard des trois domaines d’attraction précédente.

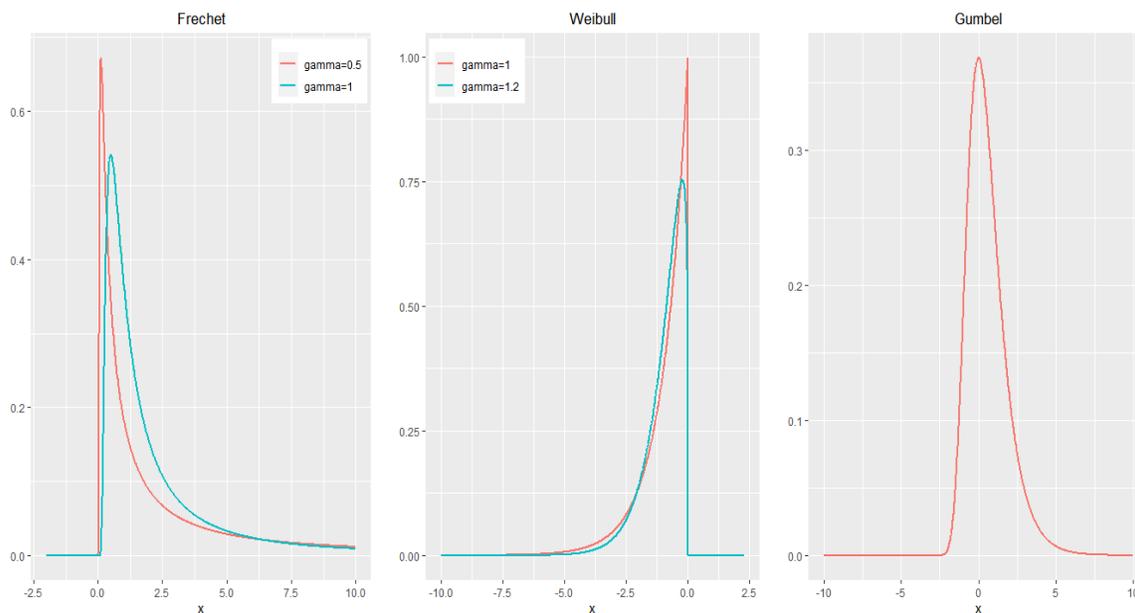


FIG. 1.3 – Densités de Lois des Valeurs Extrêmes avec $1/|\gamma|, \alpha > 0$.

La ?? montre les distributions prototypiques à chaque domaine d’attraction, pour les valeurs sélectionnées de α . On souligne les queues longues, les queues en décomposition polynomiales présentées par les distributions choisies de Fréchet, contrastant avec les queues courtes délimitées en haut, attribuées au domaine de Weibull.

CHAPITRE 2

Estimation de l'Indice des Valeurs Extrêmes

On a vu que pour la majorité des fdr F la loi asymptotique du maximum est une loi des valeurs extrêmes qui étant indexée par le paramètre γ (IVE). Dans la littérature de la théorie des valeurs extrêmes on trouve plusieurs techniques semi-paramétriques pour l'estimation de cet indice γ . En général, il n'y a pas une meilleure méthode dans toutes les situations. Mais, les méthodes les plus utilisées sont celles de Hill, de Pickands et des moments. Ceci est de probablement au fait qu'elles sont les plus anciennes.

Cependant, dans la réalité, quel que soit le phénomène étudié, on fait face toujours aux données incomplètes que ça soit tronqué ou censuré. Dans ce chapitre, on cite l'estimateur de Pickands, l'estimateur de Hill et l'estimateur des moments et on va s'intéresser à l'estimateur de [Gardes et Stupfler \[16\]](#) et l'estimateur de [Benchaira et al. \[3\]](#) en présence de troncature.

2.1 Données incomplètes

Une des caractéristiques des données de survie est l'existence d'observations incomplètes. En effet, les données sont souvent recueillies partiellement, notamment, à cause des processus de censure et de troncature. Les données censurées ou tronquées proviennent du fait qu'on n'a pas accès à toute l'information : au lieu d'observer des réalisations indépendantes et identiquement distribuées de durées X , on observe la réalisation de la variable X soumise à diverses perturbations, indépendantes ou non du phénomène étudié.

2.1.1 Données censurées

La censure est le phénomène le plus couramment rencontré lors du recueil de données de survie. Il y a plusieurs modes de censure : censure à droite, censure à gauche, censure par intervalle. on désigne par Y le temps de censure et par Z la durée réellement observée. On dispose donc 3 variables aléatoires X , Y et Z à partir desquelles on extrait des échantillons, de taille $n > 1$, (X_1, \dots, X_n) , (Y_1, \dots, Y_n) et (Z_1, \dots, Z_n) respectivement. Pour tout individu i , on a donc :

- son temps de survie X_i , de fonction de répartition F .
- son temps de censure Y_i , de fonction de répartition G
- la durée réellement observée Z_i , de fonction de répartition E .

Censure à droite

Une durée de survie est dite censurée à droite si l'individu concerné n'a aucune information sur sa dernière observation. Ainsi, en présence de censure à droite les variable d'intérêt ne sont pas toutes observées.

Censure à gauche

Il y a censure à gauche lorsque l'individu déjà subi l'événement avant qu'il soit observé. On sait uniquement que la variable d'intérêt est inférieure ou égale à une variable connue.

Censure par intervalle

Dans ce cas, comme son nom l'indique, on observe à la fois une borne inférieure et une borne supérieure de la variable d'intérêt .On retrouve ce modèle en général dans des études suivi médical ou les patients sont contrôlés et se présente ensuite après que l'événement d'intérêt se soit produit . Nous avons aussi ce genre de données qui sont censurées à droite ou, plus rarement, à gauche. Un avantage de ce type est qu'il permet de présenter les données censurées à droite au à gauche par des intervalles du type $[y, +\infty[$ et $[0, y]$ respectivement.

Censure double

La censure double (ou mixte) ce type de censure c'est un mélange entre les deux censure, la censure à droite et la censure à gauche, dans le même échantillon.

Ces quatre catégories de censure décrites ci-dessus peuvent se présenter en fonction du mode ou mécanisme de censure fixée, attendue, aléatoire. Mais, le type le plus intéressant dans les études est la censure aléatoire.

Censure aléatoire à droite

Définition 2.1.1 (Censure aléatoire à droite)

Soit X_1, \dots, X_n un échantillon d'une va positive X , on dit qu'il y a censure aléatoire de cet échantillon s'il existe une autre va positive elle aussi Y d'échantillon Y_1, \dots, Y_n dans ce cas au lieu d'observer les X_i 's, on observe un couple de va's (Z_i, δ_i) avec

$$Z_i := \min(X_i, Y_i) \text{ et } \delta_i := \mathbb{1}\{X_i < Y_i\} \text{ pour } i = 1, \dots, n. \quad (2.1)$$

où δ_i l'indicateur de censure, qui détermine si X a été censuré ou non :

- $\delta_i = 1$ si l'événement est observé ($Z_i = X_i$). On observe les "vraies" durées ou les durées complètes.
- $\delta_i = 0$ si l'individu est censuré ($Z_i = Y_i$). On observe des durées incomplètes.

Exemple 2.1.1

Lors d'un test de fiabilité, on observe n système sur une durée X : On définit la variable aléatoire X comme la durée de vie jusqu'à la première panne de chaque système. Cette panne ne sera pas forcément observée pour chaque élément de cet échantillon. Nous observerons alors, pendant ce temps Y , ($k < n$) durées de vie (pour $Y < C$) et $n - k$ fois C , (source : [CHAIB Yacine \[7\]](#), thèse de Doctorat de Probabilités et Statistiques).

Censure aléatoire à gauche

Définition 2.1.2 (Censure aléatoire à gauche)

La censure à gauche correspond au cas où l'individu a déjà subi l'événement avant que l'individu sont observe. La durée X est dite censurée aléatoirement à gauche si au lieu d'observer X_1, \dots, X_n on observe (Z_i, δ_i) où

$$Z_i := \max(X_i, Y_i) \text{ et } \delta_i := \mathbb{1}\{X_i \geq Y_i\}, \text{ pour } i = 1, \dots, n. \quad (2.2)$$

et Y_i est une censure aléatoire.

Exemple 2.1.2

Une ethnologie étudie la durée d'apprentissage d'une tâche. Cette durée est une va X et Y est l'âge de l'enfant. Pour les enfants qui savent déjà accomplir la tâche, X censure Y à gauche car pour eux X est inconnu mais inférieur a $Y : X < Y$.

Censure aléatoire par intervalle**Définition 2.1.3 (Cencure aléatoire par intervalle)**

Une date est censurée par intervalle si au lieu d'observer avec certitude le temps de l'événement, la seule information disponible est qu'il a eu lieu entre deux dates connue. La durée X est dite censurée par intervalle si au lieu d'observer X_1, \dots, X_n on observe aléatoirement (Z_i, δ_i) tel que

$$Z_i := \max [\min (X_i, Y_i^R), Y_i^L], \quad \text{pour } i = 1, \dots, n, \quad (2.3)$$

où Y_i^R et Y_i^L sont des censures aléatoires,

$$\delta_i := \begin{cases} 1 & \text{si } X_i \text{ observé,} \\ 0 & \text{si } X_i \text{ censure à droite,} \\ -1 & \text{si } X_i \text{ censure à gauche.} \end{cases}$$

Exemple 2.1.3

X durée entre l'infection par le **VIH** et le décès par le sida.

Pour des détails complets sur les types de censure avec des exemples, on réfère aux livres de [Elisa et Wang \[14\]](#) et [Klein et Moeschberger \[21\]](#).

2.1.2 Données troncatures

Une autre situation dans laquelle les données incomplètes apparaissent est celle des données tronquées. Les troncatures différent des censures au sens où elles concernent l'échantillonnage lui même. Une observation est dite tronquée si elle est conditionnelle à un autre événement. On dit que la variable Y de durée de vie est tronquée si Y n'est observable que sous une certaine condition dépendante de la valeur de Y . Il existe plusieurs types de troncature : troncature à droite, à gauche et par intervalle :

Troncature droite**Définition 2.1.4 (Troncature droite)**

On dit qu'il y a troncature droite lorsque X n'est observable que si elle est inférieure à Y . Y est alors la variable aléatoire de troncature droite :

$$X \text{ n'est observée que si } X \leq Y.$$

Exemple 2.1.4

Klein et Moeschberger [21] présentent des données sur les temps d'infection et l'induction pour 258 adultes et 37 enfants qui ont été infectés par le virus de SIDA. Ici, le nombre de personnes infectées est inconnu et l'information est disponible seulement pour ceux qui ont été infectés et développés le SIDA dans un certain laps de temps. Ainsi, les personnes qui n'ont pas encore développé le SIDA ne sont pas connues à l'enquêteur et ne sont pas incluses dans l'échantillon. C'est le cas de troncature à droite.

Troncature gauche

Définition 2.1.5 (Troncature gauche)

On dit qu'il y a troncature à gauche lorsque la variable d'intérêt X n'est observable que si elle est supérieure à Y . Y est alors la variable aléatoire de troncature gauche :

$$X \text{ n'est observée que si } X \geq Y.$$

Exemple 2.1.5

Durée de vie après la retraite : on étudie la durée de vie après la retraite de sujets qui entrent dans l'enquête à la suite d'un tirage au sort dans une caisse de retraite. Un sujet n'est donc observé que si sa durée de vie après la retraite excède le délai entre sa prise de retraite et l'instant de l'enquête. La durée de vie après la retraite est donc tronquée à gauche par ce délai. Elle peut aussi être censurée à droite si la fin de l'enquête a lieu alors que le sujet est toujours vivant (source : *CHAIB Yacine [7]*).

Troncature par intervalle

Définition 2.1.6 (Troncature par intervalle)

Quand une durée est tronquée à droite et à gauche, on dit qu'elle est tronquée par intervalle.

Exemple 2.1.6

On rencontre ce type de troncature lors de l'étude des patients d'un registre : les patients diagnostiqués avant la mise en place du registre ou répertoriés après la consultation du registre

ne seront pas inclus dans l'étude.

Remarque 2.1.1

Plus généralement, il y a troncature si l'observation de la variable d'intérêt X n'a lieu que conditionnellement à un événement B .

2.1.3 Estimateur de Lynden-Bell

Soit (Ω, A, P) un espace probabilisé, soient X_1, \dots, X_n une suite de va's réelles d'intérêt iid, de fdr commune F et Y_1, \dots, Y_n une suite de va's de troncature iid de fdr continue G . On suppose aussi que ces variables sont indépendantes des X_i . La fdr jointe de X_i et Y_i est :

$$\begin{aligned} H(x, y) &:= P(X \leq x, Y \leq y) \\ &= P(X \leq x, Y \leq y | X \leq Y) = p^{-1} \int_0^y F(\min(x, z)) dG(z). \end{aligned} \quad (2.4)$$

La taille de l'échantillon N est déterministe mais inconnue.

Soit $\{(X_i, Y_i), i = 1, \dots, n\}$ ($n < N$) l'échantillon observé (i.e $X_i \leq Y_i$). Une conséquence de la troncature, la taille de l'échantillon vraiment observé n est une va distribuée selon la loi Binomiale de paramètre N et p ou $p = P(X \leq Y) > 0$. Par la loi forte des grands nombres, on a

$$\frac{n}{N} \xrightarrow[n \rightarrow \infty]{p.s.} p.$$

Définition 2.1.7 (Estimateur de Lynden-Bell)

Lynden-Bell [22] en 1971 introduit les estimateurs de maximum de vraisemblance non-paramétriques de F et G donnés par les estimateurs produit-limite suivants :

$$\bar{F}_n^{LB}(x) := 1 - \prod_{\{i: X_i > x\}} \left(1 - \frac{1}{nC_n(X_i)}\right) \quad \text{et} \quad G_n^{LB}(y) := \prod_{\{i: X_i > t\}} \left(1 - \frac{1}{nC_n(Y_i)}\right), \quad (2.5)$$

où

$$C_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \leq y \leq Y_i\}.$$

Woodroffe [30] en (1985) étudie les propriétés asymptotiques des estimateurs de Lynden-Bell : les lois fortes et la normalité asymptotique.

2.2 Estimation de l'IVE sous données complètes

Le approche semi-paramétrique n'utilise que la caractérisation du domaine d'attraction de la distribution GEV. On présente ici trois estimateurs différents, tous basés sur les statistiques d'ordre $X_{1:n}, \dots, X_{n:n}$, obtenue à partir de la série initiale en considérant les k plus grande valeurs, l'idée est d'avoir $k \rightarrow \infty$ lorsque $n \rightarrow \infty$, mais sans prendre trop de valeurs de l'échantillon, ce qui conduit à imposer $k/n \rightarrow 0$ quand $n \rightarrow \infty$.

Incidemment, cela implique que se posera la question du choix optimal de k . En effet, il est indispensable de calculer ces estimateurs sur les queues de distribution. Choisir un k trop élevé engendre le risque de prendre en compte des valeurs qui ne sont pas extrêmes, à l'inverse, un sous échantillon trop petit ne permet pas aux estimateurs d'atteindre leur niveau de stabilité. Ce point délicat est abordé dans le livre [Reiss et Thomas \[27\]](#) et dans l'article de [Claudia et Fraga \[6\]](#).

2.2.1 Estimateur de Pickands

L'estimateur de Pickands a été introduit en 1975 par [J. Pickands](#) dans [25]. Il est défini par la statistique suivante :

$$\hat{\gamma}_n^P := \frac{1}{\log 2} \log \left(\frac{X_{n-k+1:n} - X_{n-2k_n+1:n}}{X_{n-2k_n+1:n} - X_{n-4k_n+1:n}} \right). \quad (2.6)$$

L'avantage de cet estimateur d'être valable quel que soit, le domaine de définition de l'IVE.

Théoreme 2.2.1 (Propriétés asymptotiques de $\hat{\gamma}_n^p$)

Pour $\gamma \in \mathbb{R}$, $F \in D(g_\gamma)$, et k que vérifie : $1 \leq k_n < n$, $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$ quand $n \rightarrow \infty$, on a :

- **Convergence en probabilité :**

$$\hat{\gamma}_n^p \xrightarrow{P} \gamma.$$

- **Convergence forte (presque sûr) :** $k/\log \log(n) \rightarrow \infty$, pour $n \rightarrow \infty$, alors :

$$\hat{\gamma}_n^p \xrightarrow{p.s.} \gamma.$$

- **Normalité asymptotiq :** sous certaines conditions sur k et F , on a :

$$\sqrt{k_n} (\hat{\gamma}_n^p - \gamma) \xrightarrow{\mathcal{D}} N \left(0, \frac{\gamma^2 (2^{2\gamma+1} + 1)}{4 (\log 2)^2 (2^\gamma - 1)^2} \right).$$

La consistance faible a été démontrée par [Pickands \[25\]](#) en 1975 et la consistance forte et ainsi la normalité asymptotique ont été démontrées par [Dekkers et de Haan \[10\]](#) en 1986.

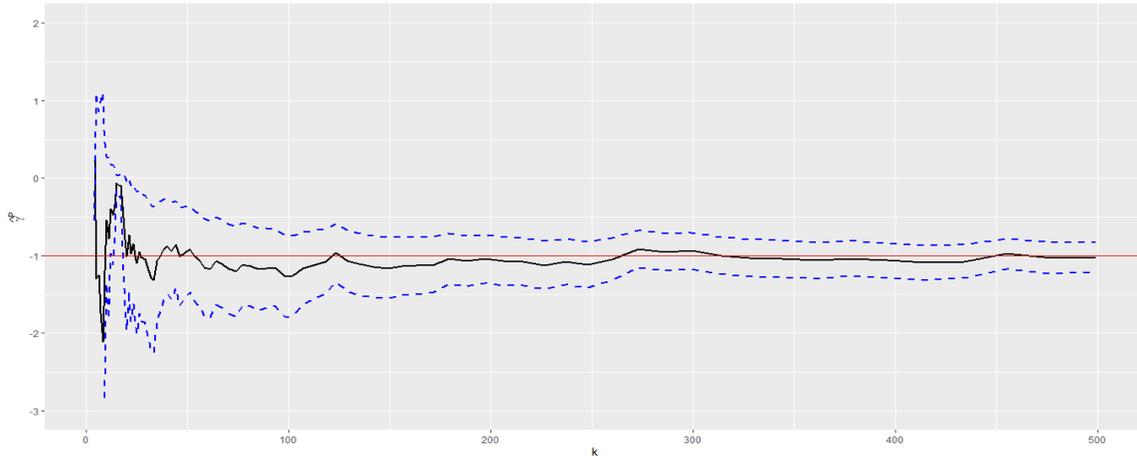


FIG. 2.1 – Estimateur de Pickands, avec un intervalle de confiance de niveau 95%, pour l'EVI de la distribution de Pareto standard ($\gamma = -1$) basé sur 1000 échantillons de 5000 observations.

2.2.2 Estimateur de Hill

Cet estimateur qui a été présenté par [Hill \[19\]](#) en 1975 et l'estimateur le plus célèbre parmi tous les estimateurs de l'indice de queue. Il s'applique seulement dans le cas où l'indice de queue est positif ($\gamma > 0$), qui correspond aux distributions appartenant au domaine d'attraction de Fréchet. Cet estimateur donné par la statistique suivante :

$$\hat{\gamma}_n^H = \frac{1}{k} \sum_{i=1}^k \log X_{n-i+1:n} - \log X_{n-k:n}. \quad (2.7)$$

Un grand nombre de travaux théoriques ont été consacrés à l'étude des propriétés de l'estimateur de Hill. [Mason \[23\]](#) en 1982 a démontré la consistance faible et [Deheuvels et al. \[9\]](#) en 1988 ont établi la consistance forte.

Théorème 2.2.2 (Propriétés asymptotiques de $\hat{\gamma}_n^H$)

Pour $\gamma > 0$, $F \in \mathcal{DA}(\Phi_{1/\gamma})$ et k que vérifie : $1 \leq k < n$, $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$, on a :

– **Convergence en probabilité :**

$$\widehat{\gamma}_n^H \xrightarrow{P} \gamma.$$

– **Convergence forte :**

$$\widehat{\gamma}_n^H \xrightarrow{p.s} \gamma.$$

– **Normalité asymptotique :** Si $\lim_{t \rightarrow \infty} (U(tx)/U(t) - x^{-\gamma}/A(t)) = x^\gamma (x^p - 1/p)$ est satisfaite avec $\sqrt{k_n}A(k/n) \rightarrow \lambda \in \mathbb{R}$, quand $n \rightarrow 1$, alors :

$$\sqrt{k_n} (\widehat{\gamma}_n^H - \gamma) \xrightarrow{\mathcal{D}} N\left(\frac{1}{1-p}, \gamma^2\right) \quad \text{quand } n \rightarrow \infty.$$

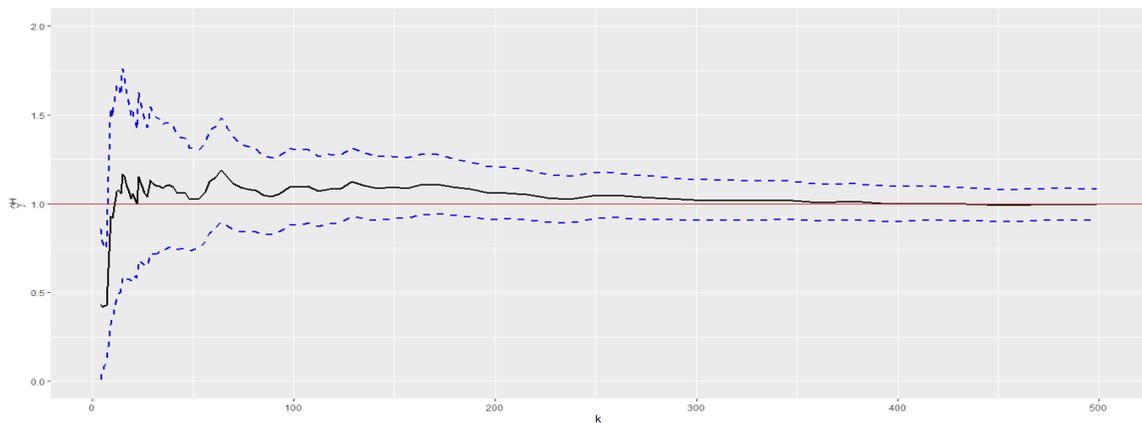


FIG. 2.2 – Estimateur de Hill, avec un intervalle de confiance de niveau 95%, pour l'EVI de la distribution de Pareto standard ($\gamma = 1$) basé sur 1000 échantillons de 5000 observations.

2.2.3 Estimateur des Moments

Un autre estimateur célèbre pour l'IVE d'une distribution, a été introduit par [Dekkers et al. \[9\]](#), c'est l'estimateur de moment. Ce dernier, il est défini par la statistique suivante :

$$\widehat{\gamma}_n^M := M_n^{(1)} + 1 - \frac{1}{2} \left(1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1}, \quad \text{Pour } \gamma \in \mathbb{R}. \quad (2.8)$$

où

$$M_n^{(j)} := \frac{1}{k} \sum_{i=0}^k (\log X_{n-i+1:n} - \log X_{n-k:n})^j, \quad j = 1, 2. \quad (2.9)$$

Théoreme 2.2.3 (Propriétés asymptotiques de $\widehat{\gamma}_n^M$)

Pour $\gamma \in \mathbb{R}$, $F \in \mathcal{DA}(\Phi_{1/\gamma})$ et k que vérifie : $1 \leq k < n$, $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$, on a :

– **Convergence en probabilité :**

$$\widehat{\gamma}_n^M \xrightarrow{P} \gamma \quad \text{quand } n \rightarrow \infty.$$

– **Convergence forte :** Si $k/(\log(n))^\delta \rightarrow \infty$, quand $n \rightarrow \infty$, pour $\delta > 0$ alors :

$$\widehat{\gamma}_n^M \xrightarrow{p.s.} \gamma.$$

– **Normalité asymptotique :**

$$\sqrt{k_n} (\widehat{\gamma}_n^M - \gamma) \xrightarrow{\mathcal{D}} N(0, \eta^2),$$

où

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

Les propriétés asymptotiques de cet estimateur ont été étudiées dans [9].

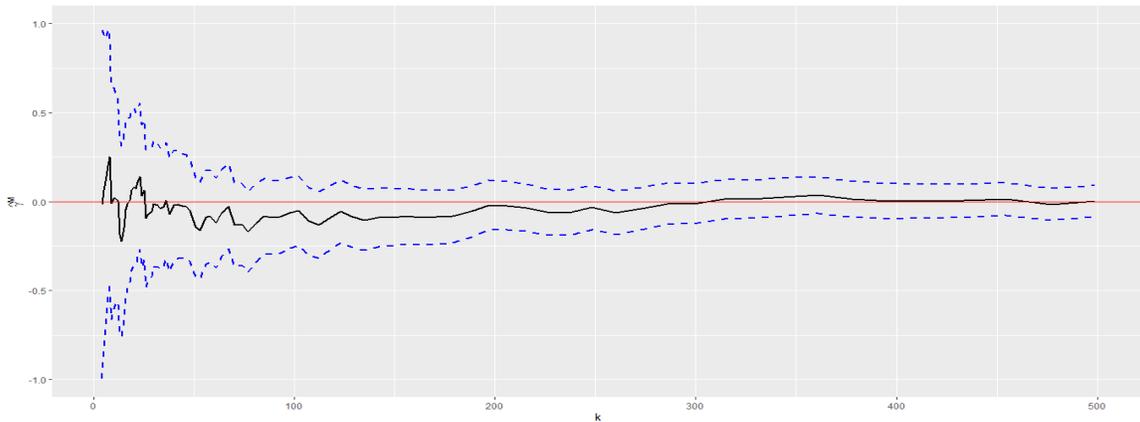


FIG. 2.3 – Estimateur de Moments, avec un intervalle de confiance de niveau 95%, pour l'EVI de la distribution de Gumbel standard ($\gamma = 0$) basé sur 1000 échantillons de 5000 observations.

2.3 Estimation de l'IVE sous données incomplètes

Dans cette partie on va intéresser au problème de l'estimation de l'IVE mais cette fois c'est en présence de données incomplètes.

2.3.1 Estimation sous censure aléatoire à droite

Les premiers qui ont mentionné ce sujet sont [Reiss et Thomas \[27\]](#) en (2007) mais sans résultat asymptotique. En 2007, [Beirlant et al \[2\]](#) ils ont introduit une méthode pour les estimateurs de Hill et de moment,... de plus, ils ont proposé les estimateurs des quantiles extrêmes et ont discuté leurs propriétés asymptotiques lorsque les données sont censurées pour un seuil déterministe et l'année suivante [Einmahl et al.\[12\]](#) ils ont adapté différents estimateurs de l'IVE au cas où les données sont censurées par un seuil aléatoire et ils ont proposé une méthode unifiée pour établir leur normalité asymptotique.

Soient \bar{F} et \bar{G} sont à queues lourdes avec les indices $-1/\gamma_1$ et $-1/\gamma_2$ où $\gamma_1 > 0$, $\gamma_2 > 0$ de l'échantillon X_1, \dots, X_n et Y_1, \dots, Y_n . Soit $\{(Z_i, \delta_i), 1 \leq i \leq n\}$ l'échantillon réellement observé défini par (2.1). Il est clair que les Z_i 's sont des variables indépendantes de loi E liée à F et G par la relation : $\bar{E}(t) = \bar{F}(t)\bar{G}(t)$. L'IVE de la fdr $E(t)$ de Z , existe et il est notée par γ où $\gamma := \frac{\gamma_1\gamma_2}{\gamma_1 + \gamma_2}$. Alors l'estimateur de Hill adapté de γ_1 aux données censurées est défini par [Einmahl et al.\[12\]](#), comme suit :

$$\hat{\gamma}_1^{(c,\cdot)}(k) = \hat{\gamma}^{(\cdot)}(k)/\hat{p},$$

où $\hat{\gamma}^{(\cdot)}$ peut être n'importe quel estimateur pas adapté à la censure, en particulier l'estimateur de Hill, moment, (basé sur les observations complètes) et $\hat{p}(k) = 1/k \sum_{i=1}^n \delta_{[n-i+1:n]}$ est un estimateur consistant de $p := \gamma/\gamma_1$ (voir [12]) qui représente la proportion de données observées dans la queue de distribution. Plus récemment, [Brahimi et al.\[4\]](#) ont proposé un nouvel estimateur du type Hill (pondéré) pour l'IVE positif. Ils ont aussi proposé sa consistance et sa normalité asymptotique sont prouvées au moyen du processus mentionné dans leur travail dans le cadre des conditions de second ordre de variation régulière. Ils ont également donné une étude de simulation comparative.

2.3.2 Estimation sous troncature aléatoire à droite

Dans les années récentes, le problème de l'étude des phénomènes extrêmes et de l'estimation de l'IVE pour des données tronquées a attiré l'attention d'un nombre croissant des chercheurs,

en raison des nombreuses applications qui appellent des solutions concrètes. Récemment, [Gardes et Stupfler \[16\]](#) ont introduit un estimateur de l'IVE et des quantiles extrêmes sous troncature aléatoire basée sur deux fractions d'échantillon distinctes des extrêmes ($k \neq k'$) à partir des données tronquées. Plus récemment, [Benchaira et al. \[3\]](#) ont proposé un nouvel estimateur sous troncature aléatoire à droite pour l'IVE, ils ont considéré la situation où $k = k'$ (au lieu de $k/k' \rightarrow 1$). Ils ont aussi proposé sa consistance et sa normalité asymptotique sont prouvées au moyen du processus de Wiener.

Soit $(\mathbf{X}_i, \mathbf{Y}_i)$, $1 \leq i \leq N$, soit $N \geq 1$ copies indépendantes d'un couple (\mathbf{X}, \mathbf{Y}) de va positives indépendantes définies sur un espace de probabilité (Ω, A, P) , avec les fonctions de distribution marginale continue \mathbf{F} et \mathbf{G} respectivement. On Suppose que \mathbf{X} est tronqué à droite par \mathbf{Y} , (on a déjà cité dans le paragraphe 2.1.2). On suppose que les deux fonctions de survie $\overline{\mathbf{F}}$ et $\overline{\mathbf{G}}$ vérifient la conduction (1.16) avec les indices négatifs respectifs $-1/\gamma_1$ et $-1/\gamma_2$ où $0 < \gamma_1 < \gamma_2$. Autrement on dit que, pour tout $s > 0$

$$\lim_{x \rightarrow \infty} \frac{\overline{\mathbf{F}}(sx)}{\overline{\mathbf{F}}(x)} = s^{-1/\gamma_1} \quad \text{et} \quad \lim_{y \rightarrow \infty} \frac{\overline{\mathbf{G}}(sy)}{\overline{\mathbf{G}}(y)} = s^{-1/\gamma_2}. \quad (2.10)$$

Ainsi, en notant (X_i, Y_i) , $i = 1, \dots, n$, peut être les données observées, sous forme de copies d'un couple de va's (X, Y) avec la fdr joint H défini par (2.4), correspondant à l'échantillon tronqué $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \dots, N$, où $n = n_N$ est une séquence de va's discrets on a déjà cité dans [Subsection 2.1.3](#).

Les distributions marginales de F et G sont donc respectivement définies par

$$F(x) := p^{-1} \int_0^x \overline{\mathbf{G}}(z) d\mathbf{F}(z) \quad \text{et} \quad G(y) := p^{-1} \int_0^y \mathbf{F}(z) d\mathbf{G}(z),$$

ce implique que les queues de distribution correspondants sont

$$\overline{F}(x) := -p^{-1} \int_x^\infty \overline{\mathbf{G}}(z) d\overline{\mathbf{F}}(z) \quad \text{et} \quad \overline{G}(y) := -p^{-1} \int_y^\infty \mathbf{F}(z) d\overline{\mathbf{G}}(z).$$

D'après [3], on a \overline{F} et \overline{G} sont des fonctions à variation régulière à l'infini avec les indices $\gamma := \gamma_1\gamma_2/(\gamma_1 + \gamma_2)$ et γ_2 respectivement.

Estimateur de Gardes et Stupfler

Gardes et Stupfler [16] en 2014 ont proposé un estimateur pour l'IVE en présence de troncature. Ils ont utilisé la définition de γ pour construire la statistique suivante :

$$\widehat{\gamma}_1^{GS}(k, k') := \frac{\widehat{\gamma}(k) \widehat{\gamma}_2(k')}{\widehat{\gamma}_2(k') - \widehat{\gamma}(k)}, \quad (2.11)$$

où $\widehat{\gamma}$ et $\widehat{\gamma}_2$ sont les estimateurs de Hill de γ et γ_2 (2.7) basés sur les observations (X_1, \dots, X_n) et (Y_1, \dots, Y_1) respectivement.

Estimateur de Benchaira et al.

Cependant, d'un point de vue pratique, il est plutôt inhabituel dans l'analyse de valeur extrême de traiter deux fractions d'échantillon distinctes simultanément ($k \neq k'$), ce qui est mentionné par [16] dans leur conclusion. On trouve Benchaira et al. [3] plus récemment ils ont considéré la situation où $k = k'$ (au lieu de $k/k' \rightarrow 1$), pour cela, leur estimateur est donné par la statistique suivant :

$$\widehat{\gamma}_1^{BMN} := \widehat{\gamma}_1^{BMN}(k) = \frac{1}{k} \frac{\sum_{i=1}^k \log \frac{X_{n-i+1:n}}{X_{n-k:n}} \sum_{i=1}^k \log \frac{Y_{n-i+1:n}}{Y_{n-k:n}}}{\sum_{i=1}^k \log \frac{X_{n-k:n} Y_{n-i+1:n}}{Y_{n-k:n} X_{n-i+1:n}}} \quad (2.12)$$

La normalité asymptotique de cet estimateur est donnée par le théorème suivant :

Théoreme 2.3.1 (Normalité asymptotique $\widehat{\gamma}_1^{BMN}$)

On suppose que la condition du second ordre de variation régulière (1.17) et $\lim_{t \rightarrow 0} t^{-1} \mathbf{P}(\overline{F}(X) \leq tx \text{ ou } \overline{G}(Y) \leq ty) =: l(x, y)$ sont satisfaisant. Soit k une suite d'entiers telle que vérifie : $1 \leq k < n$, $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$. Donc, il existe deux processus de Wiener standard $\{W_i(t), t \geq 0\}$, $i = 1, 2$, définis sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$ avec la fonction de covariance $R(\cdot, \cdot)$, et deux fonctions $A^ \sim A$ et $A_2^* \sim A_2$ avec $\sqrt{k}A^*(n/k) = O(1) = \sqrt{k}A_2^*(n/k)$, tel que*

$$\sqrt{k}(\widehat{\gamma}_1^{BMN} - \gamma_1) - \mu(k) = \int_0^1 t^{-1} (cW_1(t) - c_2W_2(t)) dt - cW_1(t) + c_2W_2(t) + o_{\mathbf{P}}(1),$$

où $c := \gamma_1^2/\gamma$, $c_2 := \gamma_1^2/\gamma_2$ et

$$\mu(k) := \frac{c\sqrt{k}A^*(n/k)}{\gamma(1-\tau)} + \frac{c_2\sqrt{k}A_2^*(n/k)}{\gamma_2(1-\tau_2)}.$$

Si en plus on a $\sqrt{k}A^*(n/k) \rightarrow \lambda$ et $\sqrt{k}A_2^*(n/k) \rightarrow \lambda_2$, alors $\sqrt{k}(\hat{\gamma}_1^{BMN} - \gamma_1) \xrightarrow{\mathcal{D}} N(\mu, \sigma^2)$, quand $n \rightarrow \infty$, où

$$\mu := \frac{c\lambda}{\gamma(1-\tau)} + \frac{c_2\lambda_2}{\gamma_2(1-\tau_2)} \quad \text{et} \quad \sigma^2 := 2c^2 + 2c_2^2 - 2cc_2\delta,$$

avec

$$\delta = \delta(R) := \int_0^1 \int_0^1 \frac{R(s,t)}{st} dsdt - \int_0^1 (R(s,1) - R(1,s)) ds + R(1,1).$$

Preuve. Voir [Benchaira et al. \[3\]](#). ■

2.3.3 Application

Pour voir la performance de l'estimateur de $\hat{\gamma}_1^{BMN}$ à celui présenté par la statistique (2.12) et de comparer avec l'estimateur de $\hat{\gamma}_1^{GS}$ à celle introduite par (2.11), en ce qui concerne le *biais* et la racine de l'erreur moyenne quadratique (root of the mean squared error : *rmse*) respectivement définis par

$$\text{biais} := \frac{1}{r} \sum_{k=1}^r (\hat{\gamma}_1^{BMN} - \gamma) \quad \text{et} \quad \text{rmse} := \sqrt{\frac{1}{r} \sum_{k=1}^r (\hat{\gamma}_1^{BMN} - \gamma)^2}$$

On a réalisé une étude de simulation pour les données simulées, on prend $\{200, 300, 500, 1000\}$ observations de la loi Burr de paramètre γ_1 tronquées par une autre variable de Burr de paramètre γ_2 , tel que le modèle de Burr défini par :

$$F(x) = 1 - (1 - x^{1/\eta})^{-\eta/\gamma_1} \quad \text{et} \quad G(x) = F(x) = 1 - (1 - x^{1/\eta})^{-\eta/\gamma_2}, \quad x \geq 0,$$

avec $\eta > 0$ and $0 < \gamma_1 < \gamma_2$. Les paramètres du second ordre de (1.17) sont $\tau = -2\gamma/\eta$ et $\tau_2 = -2\gamma_2/\eta$, on prend $\eta = 1/4$. La probabilité de troncature est égale à $1-p$ avec $p = \gamma_2/(\gamma_1 + \gamma_2)$. On fixe $p = 0.6, 0.8, 0.9$ et $\gamma_1 = 0.6, 0.9$. Les valeurs γ_2 correspondantes sont obtenues en résolvant cette dernière équation. On fait varier la taille commune N des deux échantillons et pour chaque taille, on génère 200 répétitions indépendantes. Les résultats globaux sont alors pris comme moyennes empiriques des valeurs obtenues dans les 200 répétitions. Pour

déterminer le nombre optimal d'observations extrêmes utilisées dans le calcul de $\hat{\gamma}_1^{BMN}$ et $\hat{\gamma}_1^{GS}$ on applique l'algorithme de [Reiss et Thomas \(2007\)](#) (voir la page 137). Les résultats obtenus de $\hat{\gamma}_1^{BMN}$, $\hat{\gamma}_1^{GS}$, *biais*, *rmse* sont résumés dans le [Table 2.1](#) et le [Table 2.2](#). Pour faciliter la lecture de ces tableaux, tout d'abord, on remarque que l'estimation de $\hat{\gamma}_1^{BMN}$ et $\hat{\gamma}_1^{GS}$ est meilleure pour un grand valeur de p et un petit valeur de γ_1 . Deuxièmement, on remarque que dans les deux tableaux que le estimateur de $\hat{\gamma}_1^{BMN}$ fonctionne mieux de l'estimateur de $\hat{\gamma}_1^{GS}$.

$\gamma_1 = 0.6$							
N	n	$\hat{\gamma}_1^{BMN}$	<i>biais</i>	<i>rmse</i>	$\hat{\gamma}_1^{GS}$	<i>biais</i>	<i>rmse</i>
$p = 0.6$							
200	120	0.325	-0.275	0.428	0.362	-0.298	3.617
300	179	0.362	-0.238	0.423	0.242	-0.358	2.342
500	299	0.469	-0.131	0.342	0.423	-0.177	0.854
1000	601	0.499	-0.101	0.264	0.462	-0.138	0.675
$p = 0.8$							
200	160	0.472	-0.128	0.407	0.496	-0.154	4.614
300	239	0.509	-0.091	0.347	0.578	-0.842	0.395
500	399	0.554	-0.079	0.265	0.532	-0.098	0.280
1000	800	0.569	-0.039	0.203	0.561	-0.039	0.251
$p = 0.9$							
200	180	0.537	-0.063	0.527	0.586	-0.094	1.187
300	270	0.546	-0.054	0.365	0.519	-0.084	0.388
500	450	0.534	-0.066	0.213	0.543	-0.077	0.224
1000	899	0.569	-0.031	0.160	0.562	-0.031	0.218

TAB. 2.1 – Biases and RMSE's de l'estimateur de [Benchaira et al. \(2015\)](#) (panneau de gauche) et celui de [Gardes et Stupfler \(2014\)](#) (panneau de droite) de l'IVE $\gamma_1 = 0.6$, basée sur 1000, échantillons de modèles de Burr

$\gamma_1 = 0.8$							
N	n	$\widehat{\gamma}_1^{BMN}$	<i>biais</i>	<i>rmse</i>	$\widehat{\gamma}_1^{GS}$	<i>biais</i>	<i>rmse</i>
$p = 0.6$							
200	120	0.454	-0.346	0.687	0.354	-0.446	4.530
300	179	0.482	-0.318	0.596	0.486	-0.414	2.581
500	299	0.555	-0.245	0.451	0.531	-0.369	0.650
1000	599	0.675	-0.125	0.414	0.732	-0.168	0.497
$p = 0.8$							
200	160	0.646	-0.154	0.559	0.348	-0.442	5.327
300	239	0.627	-0.173	0.418	0.563	-0.237	1.300
500	399	0.718	-0.082	0.348	0.721	-0.129	0.283
1000	800	0.767	-0.067	0.245	0.785	-0.075	0.209
$p = 0.9$							
200	180	0.644	-0.150	0.542	0.572	-0.228	2.199
300	269	0.748	-0.052	0.409	0.734	-0.066	0.374
500	450	0.721	-0.061	0.273	0.732	-0.068	0.252
1000	900	0.773	-0.033	0.223	0.742	-0.058	0.197

TAB. 2.2 – Biases and RMSE's de l'estimateur de Benchaira et al. (2015) (panneau de gauche) et celui de Gardes et Stupfler (2014) (panneau de droite) de l'IVE $\gamma_1 = 0.8$, basée sur 1000, échantillons de modèles de Burr

Conclusion

Il y a lieu de préciser que le travail entrepris dans le cadre de ce mémoire est la comparaison entre l'estimateur de Gardes et Stupfler et l'estimateur de Benchaira et al de l'IVE pour le type de données tronquées. Il arrive assez souvent qu'une série statistique, ayant des données tronquées, présente des données censurées. Il serait intéressant de faire le même travail, par ce que ce cas mérite d'être considéré attentivement vu. Le cas d'observations incomplètes issues de distributions à queues lourdes présente un intérêt particulier et ouvre une direction de recherche très intéressante qui mérite d'être approfondie.

BIBLIOGRAPHIE

- [1] Balakrishnan, N and Cohen, A. C.(1991) Order statistics and Inference. : estimation methods. statist. model. decis. sci. Academic Press.
- [2] Beirlant, J, Goegebeur, Y, Segers, J, and Teugels, J. (2006). Statistics of extrêmes : theory and applications. John Wiley.
- [3] Benchaira, S., Meraghni, D., and Necir, A. (2015). On the asymptotic normality of the extreme value index for right-truncated data. *Statist. Probab. Lett*, 378 – 384.
- [4] Brahimi, D., Meraghni, A. Necir and L. Soltane. (2018). Tail empirical process and a weighted extreme value index estimator for randomly right-censored data. Unpublished manuscript, available on the arxiv archive : <https://arxiv.org/abs/1801-00572>.
- [5] Cláudia, N , and Maria, J. (2019) . Forecasting and assessing risk of individual electricity peaks : mathematics of planet earth, Springer, International Publishing.
- [6] Cláudia, N, and Fraga A. M.I. (2004). Reiss and thomas’ automatic selection of the number of extremes. *Computational Statistics & Data Analysis*, 47, 689 – 704.
- [7] CHAIB, Y (2013) Estimation de la fonction mode pour des données tronquées et censurées. Thèse de doctorat de probabilités et statistiques. Université Badji Mokhtar, Annaba, Algérie.
- [8] David, H.A. (1970). Order Statistics. New York-London Sydney.
- [9] 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.
- [10] Dekkers, A. L. M., Einmahl, J. H. J., and de Haan, L. (1989). On the estimation of the extreme-value index and large quantile estimation. *Ann. Statist*, 17, 1795 – 1832.

-
- [11] de Haan, L., and Ferreira, A. (2006). *Extreme value theory : an Introduction*. Springer.
- [12] Einmahl, J. H., Fils-Villetard, A., and Guillou, A. (2008). Statistics of extremes under random censoring. *Bernoulli*, 14(1), 207 – 227.
- [13] Embrechts P., Klüppelberg C., and Mikosh T. (1997). *Modelling extremal events for Insurance and finance*. Springer, Verlag Berlin Heidelberg.
- [14] Elisa T. Lee, and John W. (2003). *Statistical methods for survival data analysis*, J. Wiley.
- [15] Fisher, R.T., and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math. Proc. Cambridge Philos. Soc*, 24(02), 180 – 190.
- [16] Gardes, L., and Stupfler, G. (2015). Estimating extreme quantiles under random truncation. *TEST*, 24, 207 – 227.
- [17] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d’une série aléatoire. *Ann. Math*, 44, 423 – 453.
- [18] Guillou , A and Beirlant, J. (2001). Pareto index estimation under moderate right censoring. *Scandinavian Actuarial Journal*, 2, 111 – 125.
- [19] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5), 1163 – 1174.
- [20] Jenkinson, A.F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quart. J. Roy. Meteo. Soc*, 81, 158 – 171.
- [21] Klein, J P, and Moeschberger, M L. (2003). *Survival analysis : techniques for censored and truncated data*. Springer.
- [22] Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices Roy. Astronom. Soc*, 155, 95 – 118.
- [23] Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *The Annals of Probability*, 754 – 764.
- [24] Novak, and Serguei, Y (2011) *Extreme value methods with applications to finance*. CRC Press.
- [25] Pickands, J. (1975). Statistical inference using Extreme order statistics. *The Annals of Statistics*, 3, 119 – 131.
- [26] Pierre, B (1988) *Introduction aux probailités modélisation des phénomènes aléatoires*. Springer.

-
- [27] Reiss, R.D, and Thomas, M (2007) Statistical analysis of extreme values with applications to Insurance, finance, hydrology and other fields. Birkhäuser, Basel.
- [28] Resnick, S.I. (1987). Extreme Values, Regular Variation, and point Processes. Springer, New York.
- [29] Von Mises, R. (1954). La distribution de la plus grande des n valeurs. Selected Papers, American Mathematical Society, 271 – 294.
- [30] Woodroffe, M.(1985). Estimating a distribution function with truncated data. Ann. Statist, 13, 163 – 177.

Annexe B : Abréviations et Notations

Les différentes abréviations et notation utilisées tout au long de cette thèse sont expliquées ci-dessous.

fdr ou fd	: La fonction de répartition.
GEV	: Distribution des valeurs extrêmes généralisées.
iid	: Indépendantes et identiquement distribué.
IVE	: Indice des valeurs extrêmes.
TCL	: Théorème Central Limite.
TEV	: Theorie des valeurs extremes.
va	: Variable aléatoire.
F	: Fonction de répartition.
F_n	: Fonction de répartition empirique.
F^{-1}	: Inverse généralisé de F .
\mathcal{G}_γ	: Famille de la loi de valeurs extrêmes généralisée.
$\mathbb{1}_A$: Fonction indicatrice de l'ensemble A .
\bar{F} ou S	: Fonction de queue ou Fonction de survie.
x_F	: Point terminal inférieur.
x^F	: Point terminal supérieur.
Q	: Fonction du quantile.
Q_n	: Fonction quantile empirique.
\xrightarrow{d}	: Converge en distribution.
\xrightarrow{P}	: Converge en probabilité.

$\xrightarrow{p.s}$: Converge presque sûrement.
X_1, \dots, X_n	: Une suite de n variable aléatoire.
Φ	: Loi de Fréchet.
Λ	: Loi de Gumbel.
Ψ	: Loi de Weibull.
g_γ	: La distribution des valeurs extrêmes généralisées
$D(\cdot)$: Domaine d'attraction.
\mathbb{R}	: Ensemble des valeurs réelles.
$N(0, 1)$: Loi normale standard.
(Ω, F, P)	: Espace probabilisé.
δ_i	: Indicateur de censure.
$\hat{\gamma}^P$: Estimateur de Pickands.
$\hat{\gamma}^H$: Estimateur de Hill.
$\hat{\gamma}^M$: Estimateur des Moments.
$:=$: Egalité par définition.
\overline{F}^{LP}	: Estimateur de Lynden-Bell.

مُلخَص

إن أهمية تقدير مؤشر القيم المتطرفة لذيل التوزيع لمتغير عشوائي تنبع من حقيقة أن العديد من الكميات الإحصائية، مثل المتوسط وقياسات خطر التشويه، يعبر عنها من حيث هذا المؤشر. وقد تم استخدامه تاريخياً في البحوث في الاقتصاد، والتمويل، والتأمين، والطب، وعلم الأحياء، وعلم الأوبئة، وعلم الفلك، والموثوقية، ... الهدف النهائي لهذه الأطروحة هو تقدير مؤشر القيم المتطرفة في حالة البيانات المقطوعة عشوائياً على اليمين ونقترح مثلاً تطبيقاً لمقارنة مقدر (2014) Gardes et Stupfler و مقدر (2015) Benchaira et al .

الكلمات المفتاحية: نظرية القيم المتطرفة، مؤشر القيم المتطرفة ، الاقتران الأيمن العشوائي، مقدر قاردز وستبفلر، مقدر بن شعيرة والآخرين.

Résumé

L'importance de l'estimation de l'IVE de la queue de distribution d'une variable aléatoire provient du fait que beaucoup de quantités statistiques, telles la moyenne et les mesures de risque de distorsion, sont exprimées en termes de cette l'indice. Elle a historiquement été utilisée dans la recherche en économie, finance, assurance, médecine, biologie, épidémiologie, astronomie, fiabilité, ... L'objectif final étant de ce mémoire est l'estimation de l'IVE sous données tronquées aléatoirement à droite et propose un exemple d'application pour comparer l'estimateur de Gardes et Stupfler (2014) et l'estimateur de Benchaira et al (2015).

Les mots clés : : Théorie des valeur extrême, Indice des valeurs extrêmes, Troncature aléatoire à droite, Estimateur de Gardes et Stupfler, Estimateur de Benchaira et al.

Abstract

The importance of estimating the EVI of the tail of a random variable distribution comes from the fact that many statistical quantities, such as the mean and the risk of distortion measures, are expressed in terms of this index. It has historically been used in research in economics, finance, insurance, medicine, biology, epidemiology, astronomy, reliability, ... The final objective of this memory is the estimation of the EVI under randomly truncated data on the right and proposes an example of application to compare the Gardes and Stupfler estimator (2014) and the Benchaira et al (2015) estimator.

Key words: Extreme value theory, Extreme value index, Random right truncation, Guards and Stupfler Estimator, Estimator Benchaira et al.