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 **:**

MASTER en Mathématiques

Option : **Statistique**

Par

SAKER SARA

Titre :

L'application de l'indice de queue

Membres du Comité d'Examen :

Dr.	**TOUBA SOUNIA**	UMKB	Encadreur
Dr.	**DJABER IBTISEM**	UMKB	Président
Dr.	**ABDELLI JIHANE**	UMKB	Examinateur

2020

Dédicace

JE dédie ce travail :

A ma mére

A mes frérее et soeurs

Qui m'ont toujours soutenu et encouragé

A mes amis et collégues et tous ceux qui m'a aide

Et à tous ceux qui m'ont soutenu

REMERCIEMENTS

Tout d'abord, je remercie Dieu le tout-puissant qui m'a donné la force et le savoir afin d'accomplir ce travail.

Un grand merci pour mon encadreur Dr.**TOUBA SOUNIA** pour son encouragement et son suivi attentif pour la réalisation de ce travail.

Je tiens aussi à remercier également tous les membres de jiry : Dr.**DJEBER IBTISEM** et Dr.**ABDELLI JIHANE** pour avoir accepter d'évaluer mon travail.

Je remercie tous les enseignants qui ont contribuée à ma formation, ainsi que tous les employée du département de Mathématiques

A mes chéres amis qui ont toujours été présents et fidéles. Enfin, pour toute personne qui acontribué, de prés ou de loin, à l'élaboration de ce mémoire. Veuilez bien trouver ici l'expression de mes sincéres remerciements.

Table des matières

Table des figures

Liste des tableaux

Introduction

La théorie des valeurs extrêmes (TVE) ou "ExtremeValueTheory" en anglais, est une branche de la théorie des probabilités et des statistiques mathématiques, est apparue grace à Frésher et Tippett,Gumbel et Gnedenko. Lorsque l'on caractérise les domaines d'attractions de la plus grande observation. On cherche alors à décrire le compoortement des valeurs extrêmes d'un échantillon. C'est-à-dire que l'on veut approcher la loi que suit le maximum ou le minimum des observations lorsque celles ci suivent une loi inconnue. Les domaines d'applications sont en effet trés variés : hydrologie, méthorologie, biologie, finance, assurance, etc , en effet la gestion des risques est devenue aujourd'hui fondamentale dans tous ces domaines. Durant ces derniéres années, plusieurs hauteurs comme Embrechts et al, Danielsson et de vries ont noté que la TVE est appropriée à la modélisation des observations en hautes fréquences en finance. Deux théorème sont essentiels à la théorie des valeurs extrêmes celui de Fisher-Tippet et celui de Balkema de Hann, Pickands. En effet, deux approches sont possibles à la modélisatio des évenements rares la méthode blok maximum qui modélise la distribution des extrêmes par la distribution (GEV) et la méthode POT(Pics au delà d'un seuil) qui modélise la distribution des excés au-dessus d'un seuil élevé par la Distribution de Pareto Généralisée(GPD).

La théorie des valeurs extrêmes fournit des outils pour estimer la loi des observations au delà d'un seuil donné et pour calculer les quantiles extrêmes à l'aide de cette loi estimée.

En effet, la loi des valeurs « au-delà d'un seuil » pour un échantillon iid, est sous des conditions assez générales, une loi universelle (la distribution de Pareto généralisée), et ilexiste des méthodes statistiques pour estimer les paramètres de cette « loi de queue ».

Chapitre1 : Dans ce chapitre, nous rappelons quelques éléments théoriques essentiels de la théorie des valeurs extrêmes (TVE). Il contient des rappels sur la statistique d'ordre, qui est très utile en théorie des valeurs extrêmes et on fait une introduction sur l'étude du comportement asymptotique du maximum d'un échantillon. Cette étude faisant appel à la notion de fonctions à variations régulières, on rappelle préalablement la définition de telles fonctions et on en donne quelques propriétés. On donne ensuite des résultats décrivant les limites possibles de la loi du maximum d'un échantillon. Deux théorèmes sont essentiels à la compréhension de la Théorie des Valeurs Extrêmes : celui de Fisher-Tippett et celui de Balkema- de Haan-Pickands.

Chapitre2 : Dans ce chapitre, on passe en revue les différentes méthodes d'estimation de l'indice des valeurs extrêmes. La vaste collection des estimateurs de l'indice des valeurs extrêmes γ qui caractérise les queue de distributions, à été la question centrale dans ce chapitre. Nous avons présenté les principales approches pour l'estimation de γ, Le plus utilisé est les estimateurs du maximum de vraisemblance. Avec une attention particuliéreà l'estimation semi-paramétrique(l'estimateur de Hill ; l'estimateur de Pickands et l'estimateur de Moment). Le plus connu est l'estimateur de Hill(1975) où nous basons notre travai.(voir [2] , [3] et[10])

Chapitre 1

La théorie des valeurs extrêmes

La théorie des valeurs extrêmes est une branche des statistiqies qui s'intéresse aux valeurs extrêmes des distributions de probabilité.

1.1 Définitions et caractéristiques de bases

Définition 1.1 *(**Fonction de Distribution empirique**) Soit $X_1, \cdots, X_n$ une suit de va's indépendantes et identiquements distribuées (i.i.d) définies sur le même espace de probabilité (Ω, F, P) d'une fonction de répartition commune F telle que :*

$$F(x) := P\{w \in \Omega / X(w) \leq x\} = P(X \leq x), \qquad x \in \mathbb{R}.$$

La fonction de répartition empirique notée F_n est définie par :

$$\begin{aligned} F_n(x) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leq x\}}, \qquad x \in \mathbb{R}. \\ &= \begin{cases} 0 & si \quad x < X_{1.n} \\ \frac{k}{n} & si \quad X_{k,n} \leq x < X_{k+1,n} \ , \quad pour \quad 1 \leq k \leq n \\ 1 & si \quad x \geq X_{n,n} \end{cases} \end{aligned}$$

En plus notons par $\bar{F}$ la fonction de survie (ou la fonction des queues) est définie par :

$$\bar{F}(x) = P\{w \in \Omega \diagup X(w) > x\} = 1 - F(x)$$

Théorème 1.1 ***(Glivenko- Cantelli, 1933)***

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

Définition 1.2 *(**Fonction des quantiles empiriques**) La fonction des quntiles ou l'inverse généralisé de la fonction de distribution F notée par Q telle que :*

$$Q(t) := F^{\longleftarrow}(t) = \inf\{x \in \mathbb{R}, F(x) \geq t\}, \qquad 0 < t < 1$$

La fonction des quantiles empiriques notée Q_n est définie par :

$$\begin{aligned} Q_n(t) &= F_n^{\longleftarrow}(t) = \inf\{x \in \mathbb{R},\ F_n(x) \geq t\},\ 0 < t < 1 \\ &= \begin{cases} X_{i,n} & \text{si} \quad \frac{i-1}{n} < t \leq \frac{i}{n} \\ X_{n,n} & \text{si} \quad 0 < t \leq 1 \end{cases} \end{aligned}$$

où $F^{\longleftarrow}$ est l'inverse généralisée de la fonction de distribution F.

Définition 1.3 *(**Fonction des quantiles de queue**) La fonction des quantiles de queue notée U est définie par :*

$$U(t) = Q(1 - 1/t) = F^{\longleftarrow}\left(1 - \frac{1}{t}\right), \qquad 1 < t < \infty$$

Et la fonction des quantiles de queue empirique notée U_n est donnée par :

$$U_n(t) := Q_n(1 - 1/t), \qquad 1 < t < \infty$$

Définition 1.4 ***(Point terminal)*** *Le point terminal de la fonction de répartition F est donné par :*

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

1.1.1 Lois des grands nombres

Ces lois décrivent le comportement asymptotique de la moyenne de l'échantillon. Elles sont de deux types : loi faible mettant en jeu la convergence en probabilité et loi forte relative à la convergence presque sûre.

Théorème 1.2 ***(Lois des grands nombres)*** *Si$(X_1, ..., X_n)$ une suite d'une va X tel que $\mathbb{E}[X] < \infty$ alors :*

$$\text{La loi faible :} \qquad \bar{X}_n \xrightarrow{p} \mu \qquad \text{quand} \quad n \longrightarrow \infty$$

$$\text{La loi forte :} \qquad \bar{X}_n \xrightarrow{p.s} \mu \qquad \text{quand} \quad n \longrightarrow \infty$$

où : $\mu := \mathbb{E}[X]$.

1.1.2 Théorème central limite

L'étude de somme de variables indépendantes et de même loi joue un rôle capitale en statistique. Le théorème suivant connu sous le non de théorème centrale limite(TCL) établit la convergence vers la loi de Guass.

Théorème 1.3 ***(TCL)*** *Soit $(X_;)_{i \geq 1}$une suite de va's réelles i.i.d, telle que $\mathbb{E}(X_i^2) < \infty$, et $\bar{X}_n$ sa moyenne empirique alors :*

$$\frac{\sqrt{n}\left(\bar{X}_n - \mathbb{E}(X_1)\right)}{\sqrt{var(X_1)}} \xrightarrow{d} N(0,1), \quad \text{quand} \quad n \longrightarrow \infty.$$

On peut aussi en déduire que la loi de $\sum_{i=1}^{n} X_i = n\ \bar{X}_n$ est proche de $N\left(n\mathbb{E}(X_1), n\ \sigma^2\right)$, où $\sigma^2 = var(X_1)$.

1.2 Statistiques d'ordre

1.2.1 Définition de statistiques d'ordre

Soit $X_1, \cdots, X_n$, n va i.i.d de densité de probabilité f et de fonction de distribution F. On appelle statistique d'ordre notées $X_{1,n}, \cdots, X_{n,n}$ les va's ordonnées comme suit : $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$ telle que :

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

et :

$$X_{n,n} = \max(X_1, \cdots, X_{n)}$$

1.2.2 Distribution du maximum et du minimum

La fonction de répartition de la statistique d'ordre du maximum $X_{n,n}$ est :

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

sa densité est :

$$f_{X_{n,n}}\left(x\right) = n \ f\left(x\right) F\left(x\right)^{n-1}, \quad pour \quad x \in \mathbb{R}$$

La fonction de répartition de la statistique d'ordre du minimum $X_{1,n}$ est :

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

et sa densité est :

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

1.2.3 Distribution de la K$^{\text{ième}}$statistique d'ordre

La fonction de répartition de la K$^{\text{ième}}$ statistique d'ordre est :

$$\begin{aligned} F_{X_{k,n}}(x) &= P(X_{k,n} \leq x) \\ &= P\{\text{au moins } k \text{ des } X_r \text{ sont inférieur à } x\} \\ &= \sum_{r=k}^{n} P\{\text{exactement } r \text{ de } X_1, \cdots, X_n \text{ sont inférieur à } x\} \\ &= \sum_{r=k}^{n} C_n^r [F(x)]^r [1 - F(x)]^{n-r}, \quad x \in \mathbb{R} \end{aligned}$$

où :

$$C_n^r = \frac{n!}{r!(n-r)!}$$

Sa fonction de densité est :

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

1.2.4 Densité jointe de n statistique d'ordre

Soit $X_1, \cdots, X_n$, n va i.i.d de densité de probabilité f, alors la densité jointe de la statistique d'ordre $(X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n})$ est :

$$f_{(X_{1,n}, \cdots, X_{n,n})}(x_1, \cdots, x_2) = n! \prod_{i=1}^{n} f(x_i), \quad -\infty < x_1 < \cdots < x_n < +\infty$$

1.2.5 Densité jointe d'un couple de statistique d'ordre

La densité jointe d'un couple de statistique d'ordre $(X_{j,n}, X_{k,n})$ avec $i \neq j$ est :

$$f_{X_{i,n}X_{j,n}}(x, y) = \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!} f(x)\, f(y)\, [F(x)]^{i-1}\, [F(y) - F(x)]^{j-i-1}\, [1 - F(y)]^{n-j}, \quad -$$

1.3 Distribution des valeurs extrêmes

Théorème 1.4 ***(Fisher et Tippet(1928),Gnedenko(1943))***

Soit $(X_1, \cdots, X_n)$ est une suite des va's i.i.d s'il existe un réel γ et deux suites (a_n) et (b_n), $n \in \mathbb{N}$ avec $b_n \in \mathbb{R}$ et $a_n > 0$ telles que :

$$\lim_{n\to\infty} P\left[\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right)\right] = \lim_{n\to\infty} F^n(a_n x - b_n) \longrightarrow H_\gamma(x).$$

pour tout x, où H est une fonction de distribution non dégénérée. Alors H est de même type que l'une des fonction suivantes :

– **Fréchet :**

$$H_\gamma(x) = \Phi_\gamma(x) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ \exp\left(-(x)^{-\frac{1}{\gamma}}\right) & \text{si} \quad x > 0 \end{cases} \quad \text{avec} \quad \gamma > 0$$

– **Weibull :**

$$H_\gamma(x) = \Psi_\gamma(x) = \begin{cases} 1 & \text{si} \quad x > 0 \\ \exp\left(-(-x)^{-\frac{1}{\gamma}}\right) & \text{si} \quad x \leq 0 \end{cases}, \text{avec} \quad \gamma < 0$$

– **Loi de Gumbel :**

$$H_0(x) = \Lambda(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R}$$

Définition 1.5 ***(distribution standard des valeurs extrêmes)***

Les trois fonctions de distribution du théorème **1.4** s'appellent les distribution standard des valeurs extrêmes.

– La fonction de répartition H_γ est appelée loi des valeurs extrêmes(que l'on note (EVD)"Extreme Value Distribution"). Le paramètre γ est un paramètre de forme encore appelé indice des valeurs extrêmes ou indice de queue, a_n est un paramètre de position et b_n est un paramètre déechelle.
– Les suites de normalisation $\{a_n\}$ et $\{b_n\}$ ne sont pas uniques.

Exemple 1.1 *Supposons que X suit une loi de probabilité exponentielle standard* $\exp(1)$. *Si nous posons $a_n = 1$ et $b_n = \ln n$, nous aurons :*

$$P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \begin{cases} \left[1 - e^{-(x+\ln n)}\right]^n & si \quad x + \ln n > 0 \\ 0 & si \quad x + \ln n \leq 0 \end{cases}$$

$$\lim_{n \longrightarrow \infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n \longrightarrow \infty} \left[1 + \frac{-e^{-x}}{n}\right]^n = \exp\left(-e^{-x}\right) = \Lambda(x).$$

Exemple 1.2 *Suposons que X suit une loi de probabilité uniforme $U([0,1])$. Si nous posons $a_n = n^{-1}$ et $b_n = 1$, nous aurons :*

$$P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \begin{cases} 0 & si \quad 1 + \frac{x}{n} < 0 \\ \left[1 + \frac{x}{n}\right]^n & si \quad 0 \leq 1 + \frac{x}{n} \leq 1 \\ 1 & si \quad 1 + \frac{x}{n} > 1 \end{cases}$$

$$\lim_{n\longrightarrow\infty} P\left(\frac{X_{n,n}-b_n}{a_n}\leq x\right)=\begin{cases}\exp(x) & si \quad x\leq 0\\ 1 & si \quad x>0\end{cases}=\Psi_1(x)$$

Proposition 1.1 ***(Relation entre*** Λ, Φ_γ ***et*** Ψ_γ***)***

Soit Y une va positive ($Y>0$) alors les affirmation suivantes sont équivalentes :

1. $Y \sim \Phi_\gamma$.

2. $\ln Y^\gamma \sim \Lambda$.

3. $-Y^{-1} \sim \Psi_\gamma$.

$$Y \sim \Phi\gamma \Longleftrightarrow -Y^{-1} \sim \Psi\gamma \Longleftrightarrow \ln Y^\gamma \sim \Lambda$$

1.3.1 Distribution des valeurs extrêmes généralisé (GEVD)

Pour faciliter le travail avec les trois distributions limites **Jenkinson-VonMises** a donnée une représentation qui a obtenu en introduisant les paramètres localisation μ et de dispersion σ dans la paramétrisation des distributions extrêmes,notée GEVD (Generalized Extreme Value Distribution) est donnée par :

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

Avec $\mu \in \mathbb{R}$ et $\sigma > 0$.

En remplaçant $\left(\frac{x-\mu}{\sigma}\right)$ par x on obtient la forme standard de la GEVD

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

$\gamma \in \mathbb{R}$ c'est l'indice de queue ou l'indice des valeurs extrêmes.

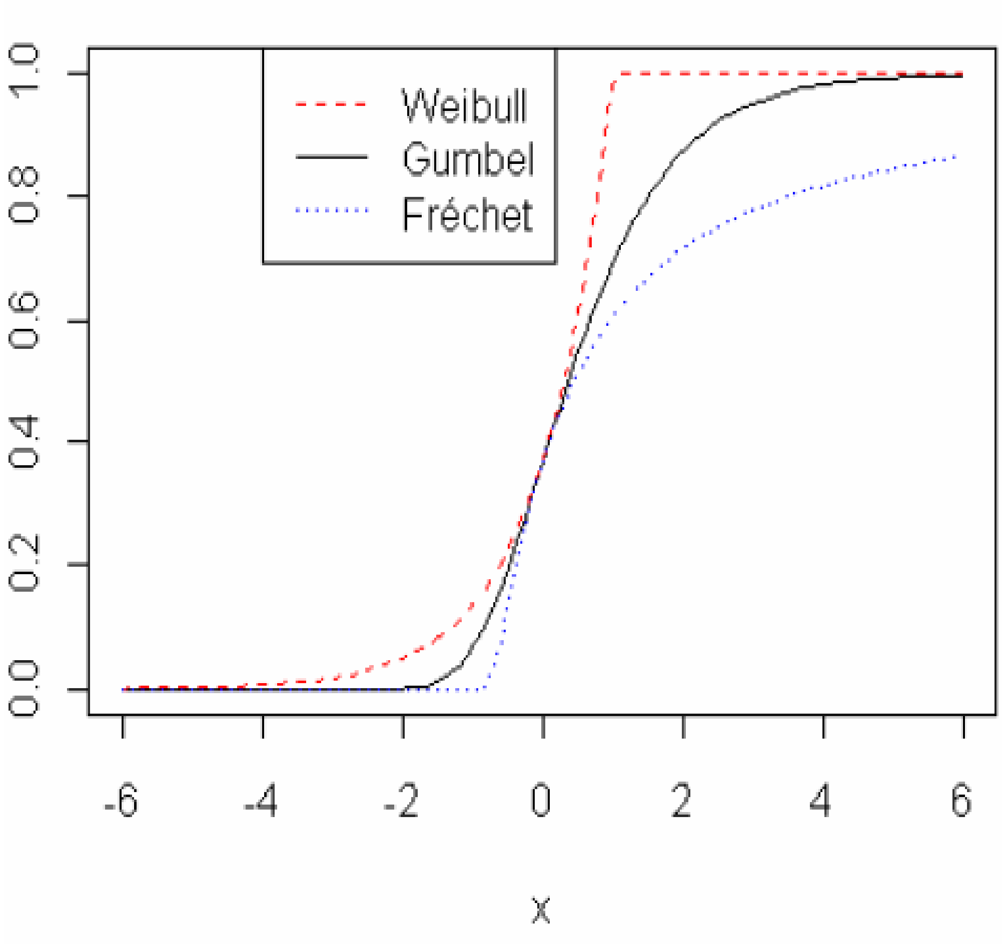

Fig. 1.1 – Distribution standard des valeurs extrêmes

La densité de la loi GEV séecrit pour $\gamma \neq 0$ comme suit :

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

La fonction de densité standard correspondante $h_{\mu,\sigma,\gamma}$ est :

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

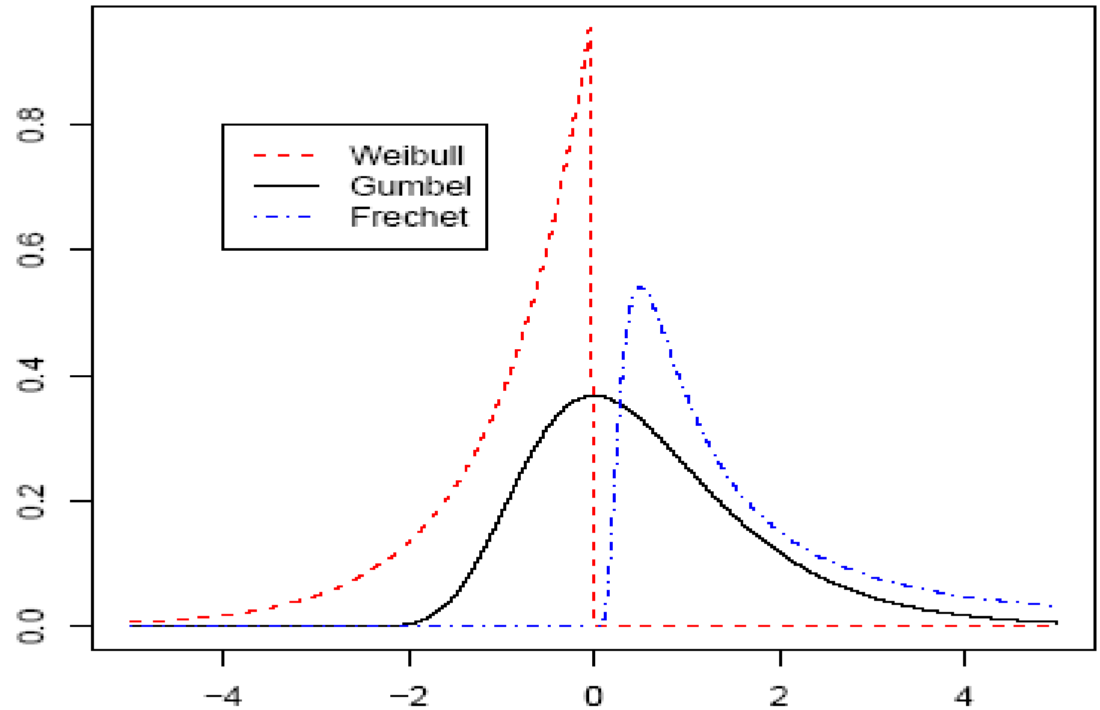

FIG. 1.2 – Densites standard des valeurs extrêmes

1.3.2 Domaines d'attraction

Définition 1.6 ***(Domaine d'attraction)***

Si F vérifie le théorème **1.4**, on dit que F appartient au domaine d'attraction de H_γ, notée par $F \in D(H_\gamma)$.

Définition 1.7 ***(Fonction à variation régulières)*** *On dit qu'une fonction G à variation réguliére d'indice $\alpha \in \mathbb{R}$ à l'infini et on note $G \in RV_\alpha$ si G est positive à l'infinie (i.e.s'il existe A tel que pour tout $x \geq A$, $G(x) > A$) et pour tout $t > 0$,*

$$\lim_{x \longrightarrow +\infty} \frac{G(tx)}{G(x)} = t^\alpha$$

Si $\alpha = 0$, on dit que G est à variation lente.à l'infini. Dans la suite, les fonction à variations lentes sont notées l. En remarquent que si G est à variations régulieres d'indice α alors $G(x)/x^\alpha$ est à variation lentes , il est facile de montrer qu'une fonction régulière d'indice α peut toujours s'ecrire sous la forme $x^\alpha l(x)$. Comme exemple de fonctions à variations lentes, citons les fonctions $\ln(1+x)$, $\ln[1+1/\ln(1+x)]$, etc...

Théorème 1.5 ***(Représentation de Karamata)*** *l est une fonction à variation lente si et seulement si pour tout $x>0$,*

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

où c et ε sont des fonctions positives telles que :

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

1. Si la fonction c est costante, on dit que l est normalisée.

2. Le théorème **1.5** implique que si l est normalisée alors l est dérivable le dérivé λ avec pour $x>0$:

$$\lambda(x)=\frac{\varepsilon(x)\,l(x)}{x}$$

3. Soit G une fonction à variation régulière d'indice α. En utilisant le fait que $G(x)=x^{\alpha}l(x)$, on déduit facilement du théorème **1.5** pour tout $x>0$:

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

où c et ε sont des fonctions positives telles que :

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

Caractérisation des domaines d'attraction

Nous allons donner des cnditions sur la fonction de répartition F pour qu'elle appartienne à l'un des trois domaines d'attraction.

Domaine d'attraction de Fréchet : Ce domaine d'attraction regroupe la majorité des distributions à queue lourde comme par exemple la loi de Cauchy, la loi de Pareto, Log-Gamma, et Student, etc...

Théorème 1.6 *$F \in D(\Phi_\gamma)$ avec $\gamma > 0$ ssi $x_F = +\infty$ et $1 - F$ est une fonction à variation régulière d'indice $-1/\gamma$ (i.e. $1 - F(x) = x^{-1/\gamma} l(x)$ où l est une fonction à variation lente).Dans ce cas, un choix possible pour les suites a_n et b_n est :*

$$a_n = F^{\longleftarrow}\left(1 - \frac{1}{n}\right) \qquad et \qquad b_n = 0$$

Exemple 1.3 *Soit $X_1, \cdots, X_n$ une suite de v.a's i.i.d de loi de Paréto de paramètre $\gamma > 0$, de fonction de répartition $F(x) := 1 - cx^{-\gamma}$, pour $a_n = (cn)^{\frac{-1}{\gamma}}$ et $b_n = 0$, on'a :*

$$\begin{aligned}
\lim_{n \longrightarrow \infty} F^n(a_n x + b_n) &= \lim_{n \longrightarrow \infty} P\left(\frac{X_{nn} - b_n}{a_n}\right) \leq x \\
&= \lim_{n \longrightarrow \infty} \left(1 - \frac{x^{-\gamma}}{n}\right)^n \\
&= \exp\left(-x^{-\gamma}\right)
\end{aligned}$$

Donc la loi limite est une loi de Fréchet $F \in D(\Phi_\gamma)$

Domaine d'attraction de Weibull : Ce domaine d'attraction contient la majorité des fonctions de répartition dont le point terminal est fini (loi Uniforme, Bêta, etc...)

Théorème 1.7 *$F \in D(\Psi_\gamma)$ avec $\gamma < 0$ ssi $x_F < +\infty$ et $1-$ F^* est une fonction à variation régulière d'indice $1/\gamma$ (i.e.$1 - F(x) = (x_F - x)^{-1/\gamma}\left[l(x_F - x)^{-1}\right]$). Avec :*

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

Dans ce cas un choix possible pour les suites a_n et b_n est :

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

Exemple 1.4 *Soit* $X_1, ..., X_n$ *une suite de v.a's i.i.d de loi uniforme sur* $[0,1]$, *de fonction de répartition* $F(x) = x$, *pour* $a_n = \frac{1}{n}$ *et* $b_n = 1$, *on'a :*

$$\begin{aligned}
\lim_{n \longrightarrow \infty} F^n(a_n x + b_n) &= \lim_{n \longrightarrow \infty} P\left(\frac{X_{nn} - b_n}{a_n}\right) \leq x \\
&= \lim_{n \longrightarrow \infty} \left(1 + \frac{x}{n}\right)^n \\
&= \exp(x) \\
&= \exp(-(-x))
\end{aligned}$$

Donc la loi limite est une loi de Weibull $F \in D(\Psi_\gamma)$

Domaine d'attraction de Gumbell : Ce domaine d'attraction regroupe la majorité des distributions à queue fine par exemple loi Normale, Exponentielle, Gamma, Lognormale, etc..

Définition 1.8 ***(Fonction de Von-Mises)*** *Soit* F *une fonction de répartition de point terminale* x_F *fini ou infini. S'il existe* $z < x_F$ *tel que*

$$1 - F(x) = c \exp\left\{-\int_z^x \frac{1}{a(t)} dt\right\}, \qquad z < x < x_F \leq \infty.$$

où $c > 0$ et a est une fonction positive absolument continue de densité à vérifiant $\lim\limits_{x \longrightarrow x_F^+} a'(x) = 0$, alors F est une fonction de Von-Mises et a est sa fonction auxiliaire.

Théorème 1.8 $F \in D(\Lambda)$ *avec* $\gamma = 0$ *ssi il existe une fonction de Von-Mises* F^* *telle que pour* $z < x < x_F$ *on ait :*

$$1 - F(x) = c(x)\left[1 - F^*(x)\right] = c(x) \exp\left\{-\int_z^x \frac{g(t)}{a(t)} d(t)\right\}$$

où g et c sont des fonctions mesurables tels que $g(x) \longrightarrow 1$ et $c(x) \longrightarrow c > 0$ quand $x \longrightarrow x_F$, et a est une fonction positive et absolument continue avec une densité a' satisfaire $\lim\limits_{x \longrightarrow x_f} a'(x) = 0.$

Dans ce cas, on peut choisir $b_n = Q\left(1 - \frac{1}{n}\right)$ et $a_n = a\left(b_n\right)$ comme des constantes de normalisation. Un choix possible pour a est :

$$a\left(x\right) = \int\limits_{z}^{x_F} \frac{F\left(t\right)}{F\left(x\right)} dt, \qquad x < x_F$$

1.3.3 Conditions de Von Mises

Von Mises (1936) a proposé quelques conditions simples et utiles pour les fonctions de répartition ayant une densité, ces conditions permettant de déterminer à quel des trois domaines d'attraction appartient la fonction de distribution correspondante F.

Soit F une fonction de distribution absolument continue dans l'intervalle $]x_1, x_F[$, de densité f. Alors les conditions suffisantes pour que F appartient à l'un des trois domaines d'attraction sont données dans le théorème suivant :

Théorème 1.9 ***(Conditions de Von Mises)***

Soit F une fonction de distribution absolument continue dans l'intervalle $]x_1, x_F[$, de densité f. Alors les conditions suffisantes pour que F appartient à l'un des trois domaines d'attraction sont :

1. Si f admet une dérivée $f^{'}$ négative pour tout $x \in \left]x_1, x_F\right[$, $f\left(x\right) = 0$ pour $x \geq x_F$ et

$$\lim_{t \longrightarrow x_F} \frac{f^{'}\left(t\right)\left(1 - F\left(t\right)\right)}{\left(f\left(t\right)\right)^2} = -1,$$

alors $F \in D\left(\Lambda\right).$

2. Si $f\left(x\right) > 0$ pour tout $x \in \left]x_1, \infty\right[$ et pour $\gamma > 0$

$$\lim_{t \longrightarrow \infty} \frac{tf\left(t\right)}{1 - F\left(t\right)} = \gamma,$$

alors $F \in D\left(\Phi_\gamma\right).$

3. Si $f(x) > 0$ pour tout $x \in]x_1, x_F[$, $f(x) = 0$ pour tout $x > x_F$ et pour $\gamma > 0$

$$\lim_{t \longrightarrow x_F^+} \frac{(x_F - t) f(t)}{1 - Ft} = \gamma,$$

alors $F \in D(\Psi_\gamma)$.

1.3.4 Distribution conditionnelle des excés

Plutôt que de considére le maximum d'un échantillon $X_1, \cdots, X_n$,on étudie les valeur dépassent un seuil donnée, l'excés Y de la variable X au-dessus du seuil u est définit par : $X - u$ quand $X \geq u$

Définition 1.9 ***(La fonction distribution des excés)*** *Soit X un variable aléatoire de fonction de répartition F et de points terminal x_F, pour tout $u < x_F$ la fonction de distribution des excés définit par :*

$$F_u(y) = P(X - u \leq y / X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}, \ y > 0$$

Définition 1.10 ***(Embrech et al (1997) , Coles)*** *On appelle fonction des excés en moyen la fonction $e(u)$ définit par :*

$$e(u) = \mathbb{E}[X - u / X > u], \ pour \ u \geq 0$$

1.3.5 Distribution de Pareto généralisé (GPD)

La loi de Pareto généralisé noté GPD :

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

$$\text{avec :}\quad x \geq 0 \qquad \text{si} \quad \gamma \geq 0$$
$$0 \leq x \leq -\frac{1}{\gamma} \qquad \text{si} \quad \gamma < 0$$

$H_\gamma$est appelé loi de Pareto généralisé standard

La fonction général de **GPD** noté $H_{\gamma,\mu,\beta}(x) = H_\gamma\left(\frac{x-\mu}{\beta}\right)$, telle que $\mu \in \mathbb{R}$ appelé le paramètre de position et $\beta > 0$ appelé paramètre d'chelle.

Le cas où le paramètre de position est nul $(\mu = 0)$ et $(\beta > 0)$ est important dans l'analyse statistique des événements extrêmes, cette famille noté par $H_{\gamma,\beta}(x)$ telle que :

$$H_{\gamma,\beta}(x) = \begin{cases} 1 - \left(1 + \gamma \frac{x}{\beta}\right)^{\frac{-1}{\gamma}} & \text{si} \quad \gamma \neq 0 \\ 1 - e^{-\frac{x}{\beta}} & \text{si} \quad \gamma = 0 \end{cases}$$

avec :

$$x \geq 0 \qquad \text{si} \quad \gamma \geq 0$$
$$0 \leq x \leq -\frac{\beta}{\gamma} \qquad \text{si} \quad \gamma < 0$$

Sa densité est $h_{\gamma,\beta}$ où :

$$h_{\gamma,\beta}(x) = \begin{cases} \frac{1}{\beta}\left(1 + \gamma \frac{x}{\beta}\right)^{\frac{-1}{\gamma}-1} & \text{si} \quad \gamma \neq 0 \\ \frac{1}{\beta} e^{-\frac{x}{\beta}} & \text{si} \quad \gamma = 0 \end{cases}$$

Théorème 1.10 ***(Pickands-Balkema-de Haan)***

Une fonction de répartition F appartient au domaine d'attraction de G_γ s'il existe une fonction positive $\beta(u)$ tell que :

$$\lim_{u \longrightarrow x_F} \sup_{0 < y < x_F - u} \left| F_u(y) - H_{\gamma,\beta(u)}(y) \right| = 0$$

où $F_u(y)$: la fonction des excés

et $H_{\gamma,\beta(u)}$: le **GPD** et x_F le point terminal.

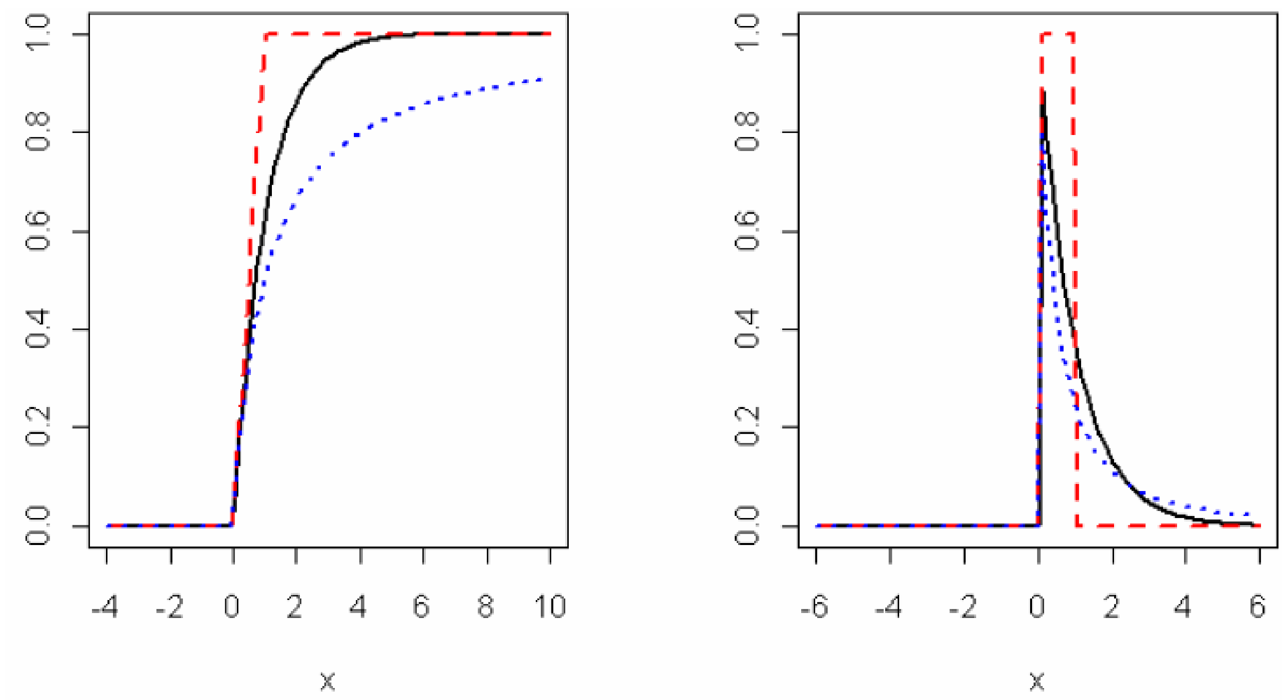

FIG. 1.3 – Densité (à gauche) et la distribution (à droite) de GPD standard

Chapitre 2

Estimation des paramétres de la loi des valeurs extrêmes

Dans ce chapitre on s'intéresse à l'estimation des paramètres de la GEV, et plus précisément on va s'intéresser à l'estimation de l'indice de queue γ qui joue un rôle essentiel dans le comportement de la loi des extrêmes.

2.1 Analyse exploratoire des données

L'information préliminaire utile sur un ensemble proposé de données à analyser peut être obtenu par plusieurs résultats graphiques et analytiques. On commence habituellement l'analyse statistique des données par la détermination des statistiques de base(la moyenne,la variance,...) et dessiner le nuage des points, l'histogramme,... En outre, dans l'analyse des extrêmes, la premiére tâche consiste à étudier le poids de la queue de distribution des données.

2.1.1 Probabilité et quantile plots

Supposons que $\hat{F}$ l'estimateur de la fonction de répartition F a été obtenu. La probabilité probabilité plot (PP-plot) et le quantile quantile plot (QQ-plot) peuvent fournir une estimation graphique à la fonction de distribution ajustée $\hat{F}$

Définition 2.1 ***(PP-plot)***

Le graphique PP-plot est l'ensemble des points suivantes :

$$\left\{\left(\hat{F}\left(X_{i,n}\right), \frac{i}{n+1}\right) : i = 1, ..., n\right\}.$$

Définition 2.2 ***(QQ-plot)***

Le graphique QQ-plot formé de l'ensemble des points suivantes :

$$\left\{X_{i,n}, \hat{F}^{-1}\left(\frac{i}{n+1}\right) : i = 1, ..., n\right\},$$

avec $X_{i,n}$ est le quantile empirique d'ordre$(i/\left(n+1\right))$ de la fonction de distribution F tandis que et $\hat{F}^{-1}\left(i/\left(n+1\right)\right)$ est son estimateur.

Le QQ-plot est un graphique qui oppose les quantiles de la distribution empirique aux quantiles de la distribution théorique envisagée. Si l'échantillon provient bien de cette distribution théorique, alors le QQ-plot sera linéare.

Dans la théorie des valeurs extrêmes, le QQ-plot se base sur la distribution exponentielle.

Dans ce cas, le QQ-plot est la représentation des quantiles de la distribution empirique sur l'axe des X contre les quantilles de la fonction de distribution exponentielle sur l'axe des Y.

Le graphique est donc l'ensemble des points :

$$\left\{x_{k,n}, G^{-1}\left(\frac{n-k+1}{n+1}\right) : k = 1, ..., n\right\},$$

avec $X_{k,n}$ représente le $k^{iéme}$ statistique d'ordre et G^{-1} est la fonction inverse de la distribution exponentielle.

L'intérêt du graphique QQ-plot est de nous permettre d'otenir la forme de la queue de la distribution. Trois cas de figures sont possibles :

- Les données suivent la loi exponentielle : la distribution présente une queue trés légère, les points du graphique QQ-plot présentent une forme linéaire.
- Les données suivent une distribution à queue lourde : le graphique QQ-plot est concave.

– Les données suivent une distribution à queue légère : le graphique QQ-plot a une forme convexe.

2.1.2 Paréto quantille plot

Les distributions de type Paréto sont caractérisées par l'équation :

$$U(x) := x^{\gamma} L_U(x), x > 0, \tag{2.1}$$

avec U est la fonction du queue et L_U est une fonction à variation lente à l'infini.

Définition 2.3 ***(Pareto quantile plot)***

Le "Paréto quntile plot" ou le "Zipf plot" correspondant au graphe :

$$\left\{\log\left(\frac{n+1}{i}\right), \log X_{n-i+1,n} : i = 1, ..., n\right\}.$$

Le"Paréto quantile plot" est une représentation trés utile pour visualisé graphiquement si les données sont distribuées selon une loi du domaine de fréchet ou non. En effet, de (2.1), il vient :

$$\log U(x) = \gamma \log x + \log L_U(x) = \gamma \log x \left(1 + \frac{\log L_U(x)}{\gamma \log x}\right).$$

En utilisant les propriétés des fonctions à variation lentes, on obtient que :

$$\log L_U(x) / \log x \longrightarrow 0 \text{ quand } x \longrightarrow \infty.$$

Ce qui implique que $\log U(x) \sim \gamma \log x$ quand $x \longrightarrow \infty$. En remplaçant la fonction du queue U par sa version empirique $\hat{U}_n$ et en remarquant que $\hat{U}_n\left(\frac{n+1}{i}\right) = X_{n-i+1,n}$, nous obtenons finalement l'équivalence suivante :

$$\log X_{n-i+1,n} \sim \gamma \log\left(\frac{n+1}{i}\right) \text{ quand } \left(\frac{n+1}{i} \longrightarrow \infty\right).$$

En d'autres termes, le" Paréto quantile plot" sera approximativement linéaire, avec une pente γ, pour les petites valeurs de i,i.e les points extrêmes.

2.1.3 Quantile plot généralisé

Définition 2.4 ***(Quantile plot généralisé)'***

Une approche permettant d'éviter le choix à priori du domaine d'attraction a été proposée par Beirlant et al en 1996. Elle consiste à utiliser un "quantile plot généralisé", défini par le graphe :

$$\left\{\left(\log\left(\frac{n+1}{j}\right), \log UH_{j,n}\right) : j = 1, ..., n\right\},$$

avec $UH_{j,n}$ est de la forme suivante :

$$UH_{j,n} = X_{n-j,n}\left(j^{-1}\sum_{i=1}^{j}\log X_{n-i+1,n} - \log X_{n-j,n}\right).$$

Suivant la courbure de ce graphe, on peut déduire dans quel domaine d'attraction on se situe.Si pour les points extrêmes on voit une droite de pente positive, on est alors dans le domaine de Fréchet. Si par contre il est plutôt constant, on est alors dans le dpmaine de Gumbel. Le cas d'une décroissance linéaire signifie que l'on appartient au domaine de Weibull

2.2 Méthodes d'estimation

2.2.1 La méthode du maximum de vraisemblance(EMV)

La méthode la plus classique est celle du maximum de vraisemblance. L'estimation par le maximum de vraisemblance donne des résultats asymptotiques efficaces, et les estimateurs obtenus convergent sous certaines conditions vers les vraies valeurs des paramètres.

Soit $X_1, X_2, \cdots, X_n$ un échantillon de n variables aléatoires supposéees indépendantes identiquement distribuées i.i.d, de densité h_θ, où $\theta = (\gamma, \sigma, \mu)$.

L'expression de la fonction de vraisemblance est donnée par :

$$l\left(\left(X_1,X_2,\cdots,X_n\right);\left(\gamma,\sigma,\mu\right)\right)=\prod_{i=1}^{n}h_\theta\left(x_i\right)\quad\text{avec}\quad\theta=\left(\gamma,\sigma,\mu\right).$$

L'estimateur $\hat{\theta}$ est donné par la résolution du système suivant :

$$\begin{cases}\frac{\partial\log l(\theta)}{\partial\theta}=0,\\ \frac{\partial^2\log l(\theta)}{\partial^2(\theta)}<0:\end{cases}$$

Dans le cas où $\gamma=0$ (loi de Gumbel), la fonction logarithme de vraisemblance est égale à :

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

En dérivant cette fonction relativement aux deux paramètres, nous obtenons le système d'équations à résoudre suivant :

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

c'est pas facile de résoudre ce système. En pratique, il est plus aisé d'utiliser des méthodes numériques.

Quand $\gamma\neq0$, la situation est encore plus compliquée.

2.2.2 Méthode des moments

Soit $X_1,...,X_n$ un n échantillon i.i.d de loi P_θ. La méthode des moments consiste à trouver le paramétre $\theta\in\Theta$ tel que les moments empiriques coincident avec les moments théoriques :

$$\hat{m}_j=\frac{1}{n}\sum_{i=1}^{n}X_i^j\approx m_j\left(\theta_0\right)=\mathbb{E}_{\theta_0}\left(X_i^j\right),\ j=1,...,d.$$

Si $d = 1$ la méthode revient à résoudre l'équation en θ :

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}_\theta \left(X_1\right).$$

L'estimateur des moments est defini comme la solution du systéme à d'équations :

$$\begin{cases} m_1\left(\theta\right) = \hat{m}_1. \\ \\ \\ m_d\left(\theta\right) = \hat{m}_d. \end{cases}$$

2.3 Estimation de l'indice des valeurs extrêmes γ

Les deux estimateurs les plus populaires sont les estimateurs de Hill (1975) et de pickands (1975).Ces estimateurs sont basés sur les plus grands statistiques d'ordre $X_{n-k,n} \leq ... \leq X_{n,n}$, où k est une suite intermédiaire d'entiers liés à la taille de l'échantillon n de la façon suivante :

$$k = k_n \longrightarrow \infty \qquad \text{et} \qquad k/n \longrightarrow 0 \qquad \text{quand } n \longrightarrow \infty$$

2.3.1 Estimateur de pickands

L'estimateur de pickands a été proposé par pickands en (1975) pour toute $\gamma \in \mathbb{R}$

Définition 2.5 ***(estimateur de pickands)***

Soit $X_1, X_2, \cdots, X_n$ une suite de v.a's i.i.d de fonction de répartition $F \in D\left(H_\gamma\right)$, où $\gamma \in \mathbb{R}$ et $k = k\left(n\right) \longrightarrow \infty$ quand $n \longrightarrow \infty$ une suite d'entrier .L'estimateur de pickands est donné par la statistique suivante :

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

Théorème 2.1 ***(Propriétés asymptotiques de l'estimateur de Pickands)***

Soit $X_1, X_2, \cdots, X_n$ une suite de v.a's i.i.d de fonction de répartition $F \in D(H_\gamma)$, où $\gamma \in \mathbb{R}$. si $k = k(n) \longrightarrow \infty$ et $\frac{k(n)}{n} \longrightarrow 0$ quand $n \longrightarrow \infty$, alors :

– **Consistance faible :**

$$\hat{\gamma}_n^p \xrightarrow{p} \gamma, \qquad \text{quand} \qquad n \longrightarrow \infty$$

– **Consistance forte :**

$$\text{Si} \underset{n \longrightarrow \infty}{\frac{k}{\log \log n}} \longrightarrow \infty \qquad \text{alors} \qquad \hat{\gamma}_n^p \xrightarrow{p.s} \gamma \quad \text{quand} \quad n \longrightarrow \infty.$$

– **Normalité asymptotique :**Sous certaines conditions pour k et F on a :

$$\sqrt{k}\left(\hat{\gamma}_n^p - \gamma\right) \xrightarrow{d} N\left(0, \eta^2\right), \quad \text{quand} \quad n \longrightarrow \infty.$$

où :

$$\eta^2 := \frac{\gamma^2 \left(2^{2\gamma+1} + 1\right)}{\left(2\left(2^\gamma - 1\right) \log 2\right)^2}$$

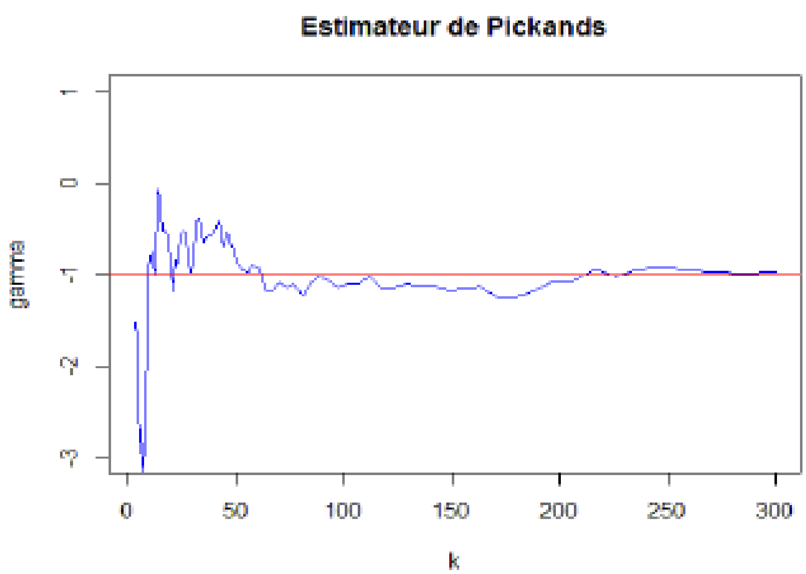

FIG. 2.1 – Représentation graphique de l'estimateur de Pickands

2.3.2 Estimateur de Hill

L'estimateur de l'GEV le plus populaire est l'estimateur de Hill, qui a été introduit en 1975 par B.Hill, il est applicable seulement dans le cas où $\gamma > 0$, il corresponds au cas de Fréchet.

Définition 2.6 *Soit $X_1, X_2, \cdot \cdot \cdot, X_n$ une suite de v.a's i.i.d de fonction de répartition F appartient au domaine d'attraction de Fréchet et la suit d'entier $k = k(n) \longrightarrow \infty$ et $\frac{k}{n} \longrightarrow 0$ quand $n \longrightarrow \infty$ L'estimateur de Hill est définie par la statistique suivante :*

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

ou bien :

$$\hat{\gamma}_n^H = \hat{\gamma}_n^H(k) := \frac{1}{k-1}\sum_{i=n-k+2}^{n} \log X_{i,n} - \log X_{n-k+1,n.}$$

Proposition 2.1 ***(Condition du premier ordre de Hann et Ferreira (2006))***

Les assertions suivantes sont équivalentes :

- F à queue lourde

$$F \in D\left(\Phi_{1/\gamma}\right), \quad \gamma > 0.$$

- $\bar{F}$ est à variation régulière à ∞ d'indice $-1/\gamma$

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

- $Q(1-s)$ est à variation régulière à 0 d'indice $-\gamma$

$$\lim_{s \longrightarrow \infty} \frac{Q(1-sx)}{Q(1-s)} = x^{-\gamma}, \quad x > 0.$$

Définition 2.7 ***(Fonction à variation régulière du second ordre)***

On dit que la queue de $F \in D(\Phi_\alpha)$, avec $\alpha = 1/\gamma$, est à variation régulière du sécond ordre, d'indice (γ, ρ) avec le paramétre du second ordre $\rho \leq 0$ (on note $\bar{F} \in 2RV_{(\gamma,\rho)}$, à l'infinie si l'une des conditions équivalentes est satisfaite :

a) Il existe un paramétre $\rho \leq 0$ et une fonction A^* satisfaisant $\lim_{t \longrightarrow \infty} A^*(t) = 0$ ne change pas son signe prés de ∞, telle que pour tout $x > 0$

$$\lim_{t \longrightarrow \infty} \frac{(1 - F(tx)/1 - F(t)) - x^{-\alpha}}{A^*(t)} = x^{-\alpha} \frac{x^\rho - 1}{\rho}.$$

b) Il existe un paramétre $\rho \leq 0$ et une fonction A^{**} satisfaisant $\lim_{t \longrightarrow \infty} A^{**}(t) = 0$ et ne change pas son signe prés de ∞, telle que pour tout $x > 0$

$$\lim_{t \longrightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}. \tag{2.2}$$

Si $\rho = 0$, $x^\rho - 1/\rho$ s'interprete comme $\log x$.

A, A^* et A^{**} sont des fonction à variation régulière avec $A^*(t) = A\left(1/\bar{F}(t)\right)$ et $A^{**}(s) = A(1/s)$. Leure rôle est de controler la vitesse de convergence.

Théorème 2.2 ***(Propriétés asymptotiques de l'estimateur de Hill)***

Soit $X_1, X_2, \cdots, X_n$ une suite de v.a's i.i.d de fonction de répartition $F \in D\left(\Phi_{\frac{1}{\gamma}}\right)$.Supposons que pour une suite intermédiaire $k = k(n) \longrightarrow \infty$, $\frac{k(n)}{n} \longrightarrow 0$, et $\frac{k(n+1)}{k(n)} \longrightarrow 1$,quand $n \longrightarrow \infty$ alors :

– **Consistance faible :**

$$\hat{\gamma}_n^H \xrightarrow{p} \gamma, \quad \text{quand} \quad n \longrightarrow \infty$$

– **Consistance forte :**

$$\text{Si} \quad \frac{k}{\log \log n} \longrightarrow \infty \quad \text{alors} \quad \hat{\gamma}_n^H \xrightarrow{p.s} \gamma, \quad \text{quand} \quad n \longrightarrow \infty.$$

– **Normalité asymptotique :** Si la condition (2.2) est satisfaite avec $\lim_{n \longrightarrow \infty} \sqrt{k} A\left(\frac{n}{k}\right) = \lambda$, alors

$$\sqrt{k}\left(\hat{\gamma}_n^H - \gamma\right) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right), \quad \text{quand} \quad n \longrightarrow \infty.$$

l'estimateur de Hill est plus asymptotiquement normal :

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

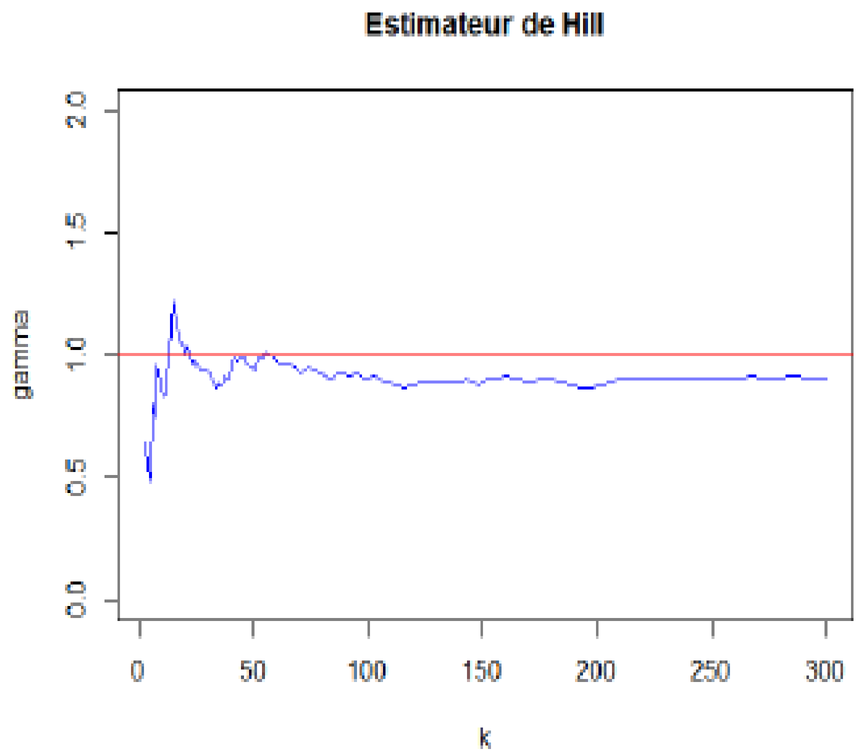

FIG. 2.2 – Représentation graphique de l'estimateur de Hill

2.3.3 Estimateur de Moment

Définition 2.8 *Un autre estimateur a été proposé par Dekkers, Einmahl et de Haan(1989), qui est valable quelque soit le singe de l'indice γ, c'est l'estimateur des moments, il est défini par :*

$$\hat{\gamma}_n^M = \hat{\gamma}_n^M\left(k\right) := M_n^{(1)} + 1 - \frac{1}{2}\left(1 - \frac{\left(M_n^{(1)}\right)^2}{M_n^{(2)}}\right)^{-1},$$

où

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

Théorème 2.3 *(Propriétés asymptotique de $\hat{\gamma}_n^M$)*

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

– **Consistance faible :**

$$\hat{\gamma}_n^M \xrightarrow{p} \gamma \text{ quand } n \longrightarrow \infty.$$

– **Consistance forte :**

Si $k/(\log n)^\delta \longrightarrow \infty$ quand $n \longrightarrow \infty$ pour certaine $\delta > 0$, alors :

$$\hat{\gamma}_n^M \xrightarrow{p.s} \gamma \text{ quand } n \longrightarrow \infty$$

– **Normalité asymptotoque :**

$$\sqrt{k}\left(\hat{\gamma}_n^M - \gamma\right) \xrightarrow{d} N\left(0, \eta^2\right) \text{ quand } n \longrightarrow \infty.$$

où

$$\eta^2 := \begin{cases} 1+\gamma^2 & \text{si } \gamma \geq 0; \\ \left(1-\gamma^2\right)(1-2\gamma)\left(4 - \frac{8-16\gamma}{1-3\gamma} + \frac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\right) & \text{si } \gamma < 0. \end{cases}$$

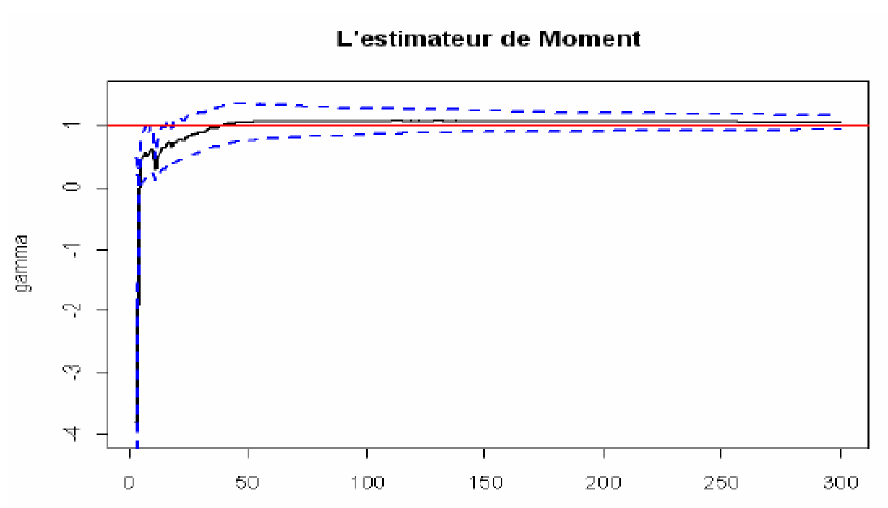

FIG. 2.3 – Représentation graphique de l'estimateur de Moment

2.3.4 Comparaison des différents estimateurs

Nous montre l'erreur de l'estimation en fonction de la méthode.

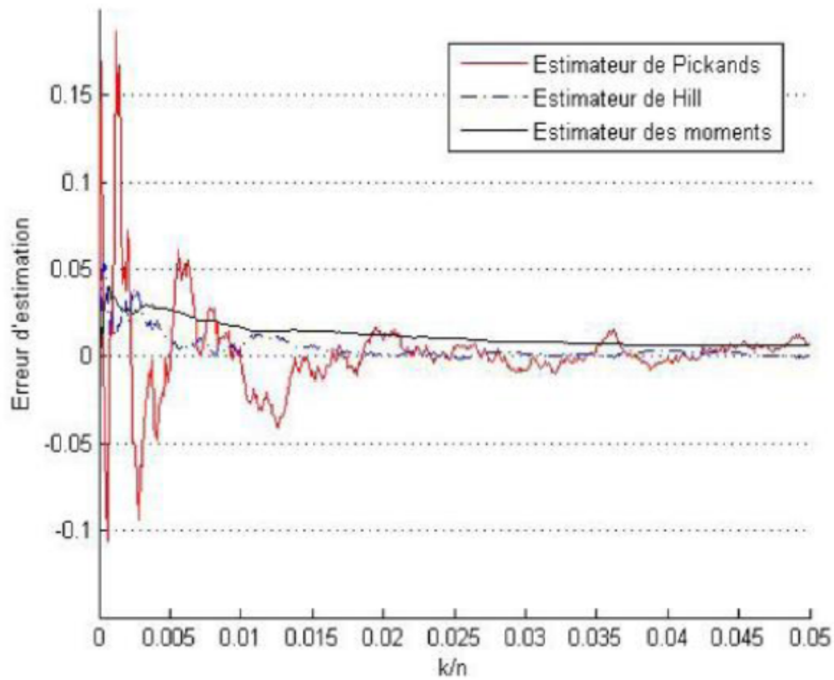

FIG. 2.4 – Comparaison de l'erreur d'estimation en fonction de la méthode

On ramarque que l'estimateur de Pickands est moins efficace dans l'estimation de l'indice de queue. Par contre, on observe une efficacité et une suprématie de l'estimateur de Hill sur ceux de Pickands et des moments. Par ailleurs, pour $k < 0.02 \times n$ léestimateur de Hill est relativement volatile. On serait donc amené à utiliser de l'ordre de $2,5\%$ des données les plus extrêmes pour estimer l'épaisseur de la queue de distribution.

Conclusion

Dans ce travail, nous avons introduit la notion des valeurs extrêmes et nous avons énoncé les principaux résultats les concernant, tout en mentionnant les deux principaux outils servant à modéliser le comportement des valeurs extrêmes : la loi des valeurs extrêmes et la loi des exés.

L'importance de l'estimateur de l'indice queue pour en savoir d'événements extrêmes quand les lois. Le TCL ne convient pas avec eux, donc pour obtenir les bornes de confiance en cherche autre méthode. Le théorème de Fisher-Tippett fournit, en quelque sorte, la contre partie du TCL dans le cas d'événements extrêmes. Cependant, contrairement au TCL, où la loi normale est la seule loi limite possible, dans le cas des extrêmes, trois types de loi limite sont possibles.

Ensuite, nous avons présenté des méthodes pour estimer les paramétres de la loi des valeurs extrêmes et les déférents estimateurs de l'indice des valeurs extrêmes(l'indice de queue). L'estimateur le plus populaire est l'estimateur de Hill, qui a été introduit en 1975 par B.Hill.

Bibliographie

[1] Beirlant ,J , Goegebeur, Y , Segers, J and Teugels J (2004), Statistics of Extremes-Theory and Application, Wiley.

[2] Beteka, S (2010). Déterminatio du nombre de statistiques d'ordre extrêmes.Memoire de magister, Université de Biskra.

[3] Beteka, S. Les valeurs extrêmes bivariees.THéses de doctorat, Université de Biskra.

[4] Benameure, S (2010). Sur l'stimation de l'indice des valeurs extrêmes. Memoire de magister, Université de Biskra.

[5] Coles, S. An Introduction to statistical Modeling of estreme values Springer

[6] De Haan, L et Ferreira, A (2006). Extremes Value Theory : au introduction Springer.

[7] Delmas J F (2013). Introduction au calcul des probabilités et à la statistique. Les presses de L'ENSTA.

[8] Embrechts, P , KlÜpperlberg,C and Mikosch, T (1997). Modeling Extreme Event for Insurance and Finance Springer, Berlin

[9] Touba, S. Sur l'estimation des parametres des lois stables. THéses de doctorat, Université de Biskra

Annexe A : Logiciel *R*

R est un système, communément appelé langage et logiciel, qui permet de réaliser des analyses statistiques. Plus particulièrement, il comporte des moyens qui rendent possible la manipulation des données, les calculs et les représentations graphiques. *R* a aussi la possibilité d'exécuter des programmes stockés dans des fichiers textes et comporte un grand nombre de procédures statistiques appelées paquets. Ces derniers permettent de traiter assez rapidement des sujets aussi variés que les modèles linéaires (simples et généralisés), la régression (linéaire et non linéaire), les séries chronologiques, les tests paramétriques et non paramétriques classiques, les différentes méthodes d'analyse des données,... Plusieurs paquets, tels *ade4, FactoMineR, MASS, multivariate, scatterplot3d et rgl* entre autres sont destinés à l'analyse des données statistiques multidimensionnelles.

Il a été initialement créé, en 1996, par *Robert Gentleman* et *Ross Ihaka* du département de statistique de l'Université d'Auckland en Nouvelle Zélande. Depuis 1997, il s'est formé une équipe "*R Core Team*" qui développe *R*. Il est conçu pour pouvoir être utilisé avec les systèmes d'exploitation *Unix*, *Linux*, *Windows* et *MacOS*.

Un élément clé dans la mission de développement de *R* est le *Comprehensive R Archive Network* (CRAN) qui est un ensemble de sites qui fournit tout ce qui est nécessaire à la distribution de R, ses extensions, sa documentation, ses fichiers sources et ses fichiers binaires. Le site maître du CRAN est situé en Autriche à Vienne, on peut y accéder par l'URL : "*http ://cran.r-project.org/*". Les autres sites du CRAN, appelés sites miroirs, sont répandus partout dans le monde.

R est un logiciel libre distribué sous les termes de la "GNU Public Licence". Il fait partie intégrante du projet GNU et possède un site officiel à l'adresse "*http ://www.R-project.org*". Il est souvent présenté comme un clone de *S* qui est un langage de haut niveau développé par les *AT&T Bell Laboratories* et plus particulièrement par *Rick Becker, John Chambers* et *Allan Wilks*. *S* est utilisable à travers le logiciel *S-Plus* qui est commercialisé par la société Insightful (*http : //www.splus.com/*).

Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous.

$\mathbb{E}[X]$	l'espérance de.X.
$X_{n,n}$	maximum de $X_1, ..., X_n$
F	fonction de répartition
$\bar{F}$	fonction de survie
$F^{\longleftarrow}$	l'inverse généralisé de F
f	densité de probabilité d'un va
$f_{X_{k,n}}$	fonction de densité de probabilité de $X_{k,n}$
F_n	fonction de répartition empirique
Q	fonction inverse généralisée de F
Q_n	fonction des quantiles empiriques
l	fonction à variation lente
i.i.d	indépendantes et identiquement distribuées
x_F	point terminale de F
TVE	théorie des valeurs extrêmes

GEV	valeurs extrêmes généralisé
H_γ	distribution des valeurs extrêmes (EVD)
GEVD	distribution des valeurs extrêmes généralisée
GPD	distribution de paréto généralisée
$D(H_\gamma)$	domaine d'attraction de H_γ
$\xrightarrow{d}$	convergence en distribution
$\xrightarrow{p}$	convergence en probabilité
$\xrightarrow{p.s}$	convergence presque sûre
(Ω, F, P)	espace de probabilité
TCL	théorème centrale limite
$:=$	égalité par définition
Φ_γ	loi de Fréchet
Ψ_γ	loi de Weibull
Λ	loi de Gumbel
$\hat{\gamma}_n^H$	estimateur de Hill
$\hat{\gamma}_{k,n}^p$	estimateur de Pickands