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

Bacha Dounia

Titre

La théorie des valeurs extrêmes

Membres du Comité d'Examen :

Dr.	**SAYAH ABDALLAH**	UMKB	Encadreur
Dr.	**BENATIA FATEH**	UMKB	Président
Dr.	**TOUBA SONIA**	UMKB	Examinateur

Juin 2022

Dédicace

Á mes trés chères parents, source de vie, d'amour et d'affection.

Á mes chers frères et mes chères soeur

Á mes chères amise.

Á toute ma chère famille.

Á mon cher fils.

Je n'oublie jamais tous mes professeurs dans tous mes voyages d'études.

REMERCIEMENTS

Tout d'abord, je remercie dieu "**Allah**" tout-puissant de m'avoir donné la force et l'énergie pour atteindre ce que je suis.

Je tiens à remercier sincérement mon superviseur "**Sayah Abdallah**" de m'avoir suivi et guidé tout au long de la production de ce message.

Je tiens également à remercier les membres du jury "**Benatia Fateh**" et "**Touba Sonia**" qui m'ont fait l'honneur d'accepter l'étude et l'évaluation de ce travail.

Je voudrais remercier ma chère famille et mes chers parents, en particulier ma mère qui m'ont soutenu durant mon parcours universitaire et m'a donné le courage de faire ce travaille.

Table des matières

Liste des tableaux

Introduction

La théorie des valeurs extrêmes (**TVE)** est une branche de la théorie des probabilités et de statistiques mathématiques qui se concentre sur l'analyse et l'inférence d'évènements extrêmes, il a une faible probabilité d'occurence. Les évènements extrêmes peuvent devenir catastrophiques et sont donc d'une grande importance dans prèsque tous les domaines de la science et de la technologie et sont donc trés significatifs.

Quand à la littérature liée à la théorie des valeurs extrêmes, nous référons aux travaux de Embrechts et al. (1997), Beirlant et al. (2005), Reiss et Thomas (2007) et de Haan et Ferreira (2006).

La théorie des valeurs extrêmes est appliquée en hydrologie pour prévoir les crues, en océanographie dans l'étude des vagues scélérates, en épidémiologie pour identifier rapidement les maladies émergenres, en démographie pour prévoir la distribution de probabilité de l'âge maximum que l'être humain pourra sinisters, en finance ou encore en météorologie.

Ainsi, notre objectif dans cette thèse est de choisir le nombre optimal de statistique de classements extrêmes qui sont essentielles à l'estimation de (**TVE**) et permettent d'améliorer les performances des estimateurs.

L'organisation de ce mémoire est la suivante :

chapitre 01 : (Préliminaire sur la Théorie de valeurs extrêmes) : Dans ce chapitre, nous rappelons quelques éléments de base de la théorie des valeurs extrêmes. Nous commencons par rappeler la définition des statistiques d'ordre, qui est trés utile en TVE et nous étudions le comportement des valeurs extrême d'un échantillon de variables aléatoires, nous définissons rapidement les caractérisations des domaines d'atractions. Nous donnons des critères pour que la limite en loi du maximum suive une loi des valeurs extrême. Et nous présentons les estimateurs de l'indice des valeurs extrêmes les plus utilisés avec une attention particulière à l'estimation paramétrique et semi-paramétrique (Hill, Moment, pickands).

chapitre 02 : (Etimation de quantiles extrêmes) : Dans ce chapitre, nous présentons l'éstimation de quantiles extrêmes (Quelques résultats sur les statistiques d'ordre, approche des quantiles extrêmes par la loi des valeurs extrêmes, approche des quantiles extrêmes par la méthode des excées, approche des quantiles extrêmes par l'approche semi-paramétrique).

Chapitre 1

Préliminaire sur la théorie des valeurs extrêmes

1.1 Théorie des Valeurs Extrêmes

1.1.1 Théorèmes Limites

Définition 1.1.1 *(Fonction de distribution empirique)*

Soit un échantillon $X = (X_1, X_2, ..., X_n)$, de variables aléatoires indépendantes et identiquement distribuées définies sur le même espace de probabilité (Ω, F, P) avec une distribution de probabilité F. La fonction de répartition empirique notée F_n et définie par :

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

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

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

Définition 1.1.2 *(Somme et moyenne empirique)*

Soit $X_1, ..., X_n$ une suite de variables aléatoires i.i.d, et F la fonction de répartition commune. Pour un entier $n \geq 1$, on définit la somme partielle et la moyenne arithmétique correspondonée respectivement par :

$$S_n = \sum_{i=1}^{n} x_i \text{ et } \bar{X}_n = \frac{S_n}{n},$$

ou $\bar{X}_n$ est la moyenne empirique.

1.1.2 Lois des grands nombres

Ces lois décrivent le comportement asymptotique de la moyenne de l'échantillon. Il y a deux types de lois :
La loi faible mettant en jeu la convergence en probabilité et la loi forte relative à la convergence présque sûre.

Théorème 1.1.1 *(Lois des grands nombres)*

Si $(X_i, ..., X_n)$ une suite de v.a, tel que $E|X| < \infty$ alors :

· la loi forte $\quad \bar{X}_n \overset{ps}{\to} \mu \quad$ quand $n \to \infty$,

· la loi faible $\quad \bar{X}_n \overset{p}{\to} \mu \quad$ quand $n \to \infty$,

où

$$\bar{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \text{ et } \mu = E(X).$$

1.1.3 Théorème centrale limite

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

Théorème 1.1.2 ***(TCL)***

Si $(X_1, ..., X_n)$ est une suite des v.a définie sur le même espace de probabilité de variance $\sigma^2 < \infty$ et de moyenne μ alors

$$\frac{1}{\sqrt{n}}(\frac{S_n - n\mu}{\sigma}) \xrightarrow{L} N(0,1) \text{ quand } n \to \infty.$$

1.2 Les statistiques d'ordre

Définition 1.2.1 *Soint $X_1, X_2, ..., X_n$, n variables aléatoire i.i.d. Rangeons ces variables aléatoires par "ordre croissant de grandeur". Pour cela, nous introduisons la notation $X_{i,n}$ avec*

$$X_{1,n} \leq X_{2,n} \leq ... \leq X_{n-1,n} \leq X_{n,n}$$

$X_{i,n}$ est donc la i-éme statistique d'ordre (ou statistique d'ordre i) dans un échantillon de taille n.
Deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des évènements extrêmes ce sont :

La statistique du minimun qui est la plus petite statistique :

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

La statistique du maximun qui est la plus grande statistique :

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

Dans la suite de notre travail, on s'interesse à la statistique du maximun puisque la statistique du minimun se déduit par relation suivante :

$$\min(X_1, ..., X_n) = -\max(-X_1, ..., -X_n).$$

Remarque 1.2.1 *Même si les variables aléatoires $X_1, ..., X_n$ sont indépendantes, les statistique d'ordre ne le sont pas.*

Définition 1.2.2 *(Fonction des quantiles)*
La fonction des quantiles associée à F est l'inverse généralisée de F notée par Q etdéfinie par

$$Q(t) := F^{-}(t) = \inf \{s : F(s) \geq t\}, \quad 0 < t < 1.$$

Définition 1.2.3 *(Fonction des quantiles empirique).*
La fonction des quantiles empirique d'un échantillon $(X_1, X_2, ..., X_n)$ est définie par

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

où F^{-} est l'inverse généralisée de la fonction de distribution F.

Définition 1.2.4 *(Fonction des quantiles de queues).*
Dans la théorie des extrême, une fonction notée U et appelée fonction quantile de queue, est définie par :

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

Définition 1.2.5 *(Fonction des quantiles de queues empirique).*

La fonction des quantiles de queues empirique notée U_n est donnée par :

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

1.3 Lois des statistique d'orde

Lemme 1.3.1 *Arnold Barry et al.* (1992) : *La densité conjointe de la statistique d'ordre $X_{1,n} \leq ... \leq X_{n,n}$ est donnée par*

$$f_{(X_{1,n} \leq \cdots \leq X_{n,n})}(x_1, x_2, ..., x_n) = n! \prod_{i=1}^{n} f(x_i),$$

avec $x_1 \leq x_2, ..., \leq x_n$.

Lemme 1.3.2 *Arnold Barry et al.* (1992) : *La densité conjointe de $(X_{r,n} \leq X_{s,n})$ avec $r \leq s$ est donnée par :*

$$f_{(X_{r,n} \leq X_{s,n})}(x, y) = n! \frac{[F(x)]^{r-1} f(x) [F(y) - F(x)]^{s-r-1} f(y) [1 - F(y)]^{n-s}}{(r-1)!(s-r-1)(n-s)!},$$

avec $-\infty < x < y < +\infty$.

La fonction de répartition $F_{(X_{r,n} \leq X_{s,n})}$ est

$$F_{(X_{r,n} \leq X_{s,n})}(x, y) = \begin{cases} F_{X_{s,n}}(y), & \text{pour } x \geq y, \\ \displaystyle\sum_{j=s}^{n} \sum_{i=r}^{j} n! \frac{[F(x)]^{i-1} f(x) [F(y) - F(x)]^{j-i-1} f(y) [1 - F(y)]^{n-j}}{(i-1)!(j-i-1)(n-j)!}, & \text{pour } x < y. \end{cases}.$$

1.4 Lois des valeurs extrêmes

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

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

Cette convergence est équivalente à la convergence

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

Ainsi, en posant $y = a_n x + b_n$, on a pour n assez grand (ou de manière équivalente pour y proche du point terminal de F puisque $a_n x + b_n \longrightarrow x_F$ lorsque $n \longrightarrow \infty$)
L'approximation

$$\bar{F}(y) \approx -\frac{1}{k} \log H_\gamma(\frac{y - b_n}{a_n}) = \frac{1}{k}(1 + \gamma \frac{y - b_n}{a_n})^{-1/\gamma}.$$

Noter que lorsque $\gamma = 0$, l'expression ci-dessus a un sens par passage à la limite.
Pour obtenir une approximation de $q(\alpha)$ avec α proche de 0, il suffit simplement d'inverser l'expression ci-dessus ce qui donne

$$q(\alpha) \approx b_n + \frac{a_n}{\gamma}\left[(n\alpha)^{-\gamma} - 1\right].$$

Théorème 1.4.1 *(Fisher et Tippet (1928) ; Gnedenko(1943)). Sous certaines conditions de régularité sur la fonction de répartition F, il existe un paramètre réel γ et deux suites*

$(a_n)_{n\geq 1} > 0$ *et* $(b_n)_{n\geq 1} \in \mathbb{R}$ *tels que pour tout* $x \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left|\frac{X_{n,n}-b_n}{a_n} \leq x\right| = H_\gamma(x),$$

avec,

loi de Fréchet $\quad H_\gamma(x) = \begin{cases} 0 & \text{si } x < 0 \\ \exp\left[-(x)^{-1/\gamma}\right] & \text{si } x \geq 0 \end{cases}$ et $\gamma > 0$,

loi de Weibull $\quad H_\gamma(x) = \begin{cases} \exp\left[-(x)^{-1/\gamma}\right] & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$ et $\gamma < 0$,

loi de Gumbel $\quad H_0(x) = \exp\left[-\exp(-x)\right]$ si $x \in \mathbb{R}$ et $\gamma = 0$,

et H_γ la fonction de répartition de la **loi des valeurs extrêmes (EVD).**

1.5 Domaines d'attraction et coefficients de normalisation

Définition 1.5.1 *On appelle domaine d'attraction O l'ensemble des lois F pour les quelles le maximum normalisé suit asymptotoquement la loi de O.*

Définition 1.5.2 *On dit qu'une distribution F appartient au max-domaine d'attraction d'une distribution des valeurs extrêmes O et on note $F \in MDA(O)$, si la distribution du maximum normalisé converge vers O i.e si F est la distribution commune de v.a $X_1, ..., X_n$ (iid) de maximum M_n, alors il existe des constantes $a_n > 0$, $b_n \in \mathbb{R}$ telles que :*

$\forall x \in \mathbb{R}$

$$\lim_{n\to\infty} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = O.$$

Exemple 1.5.1 *La loi exponentielle du paramètre 1*

Soit X suit la loi $\exp(1)$; $F(x) = 1 - e^{-x}$.

On pose $a_n = 1$, $b_n = \ln n$

$$\lim_{n\to\infty} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \Lambda(x).$$

Alors la distribution exponentielle appartient au max-domaine d'attraction de Gumbel.

Exemple 1.5.2 *La loi Uniforme*

Soit X suit la loi $U([0,1])$;

$\forall x \in \mathbb{R}$

$$F(x) = \begin{cases} 0 & \text{si} \quad x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si} \quad x > 0 \end{cases}.$$

On pose $a_n = \frac{1}{n}$ et $b_n = 1$ alors :

$\forall x \in \mathbb{R}$, $\forall n \in \mathbb{N}^*$

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(1 + \frac{x}{n}).$$

Donc $\forall x \in \mathbb{R}$

$$\lim_{n\to\infty} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \begin{cases} e^x & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases}$$
$$= \Psi_1(x)$$

Alors la distribution uniforme appartient au max-domaine d'attraction de Weibull.

1.6 Caractérisation des domaines d'attraction

Nous allons donner des conditions sur la fonction de répartition F pour qu'elle appartienne à l'un des trois domaines d'attraction. Dans la suite, on note $x_F = \sup\{x \backslash F(x) < 1\}$ le poit terminal de F et $F^{\leftarrow}(y) = \inf\{x \in \mathbb{R} \backslash F(x) \geq y\}$ l'inverse généralisée de F (avec la convention $\inf\{\emptyset\} = \infty$).

Avant de donner les expressions de la fonction de répartition dans chaque domaine, il convient d'introduires des définitions utiles.

Définition 1.6.1 *On appelle point terminale de la fonction F, le réel x_F défini par*

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

avec la convention $\sup\{\varnothing\} = \infty$.

Définition 1.6.2 *Une fonction ℓ mesurable et positive sur* $]0, \infty[$ *est à variations lentes, si, pour tout $x > 0$,*

$$\lim_{t \to \infty} \frac{\ell(tx)}{\ell(t)} = 1.$$

Théorème 1.6.1 *Tout fonction à variation lentes ℓ s'écrit sous la forme*

$$\ell(x) = c(x) \exp \int_1^x \frac{\varepsilon(t)}{t} dt,$$

où $c > 0$ et ε sont deux fonctions mesurables telles que

$$\lim_{x \to \infty} c(x) = c_0 \in]0, \infty[\text{ et } \lim_{x \to \infty} \varepsilon(x) = 0.$$

Définition 1.6.3 *On dit qu'une fonction G est à variations régulières d'indice $p \in \mathbb{R}$ à l'infini si G est positive à l'infini (i.e s'il existe A telle que pour tout $x \geq A$, $G(x) > 0$) et si pour tout $t > 0$,*

$$\lim_{x \to \infty} \frac{G(tx)}{G(x)} = t^p.$$

Dans le cas particulier où $p = 0$, G est une fonction à variations lentes à l'infini. une fonction à variations régulières d'indice p peut toujours s'écrire sous la forme $x^p \ell(x)$ où ℓ est une fonction à variations lentes à l'infini.

1.6.1 Domaine d'attraction de Fréchet

Théorème 1.6.2 *$F \in D$(Frèchet) (la distribution au domaine d'attraction de la loi de Frèchet) avec un indice des valeurs extrêmes $\gamma > 0$ si et seulement si $x_F = +\infty$ et la fonction F est à variantions régulières d'indice*
$(-1/\gamma = -\alpha)$. Dans ce cas, un choix possible pour les suites a_n et b_n est

$$a_n = F^{\leftarrow}(1 - n^{-1}) \text{ et } b_n = 0,$$

où

$$F^{\leftarrow}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}.$$

Dans ce théorème, on en déduit que $F \in D$(Frèchet) si et seulement si le point terminal x_F est infini et
$\bar{F}(x) = x^{-\alpha}\ell(x)$, où ℓ est une fonction à variations lentes à l'infini et α un réel strictement positif.

Frèchet	$\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}, \quad x > 0, \quad \alpha > 0.$
MDA(Φ_α)	$x_F = \infty, \;\; \bar{F}(x) = x^{-\alpha}\ell(x), \quad \ell \in R_{0.}$
Constantes normalisation	$a_n = F^{\leftarrow}(1-n^{-1}) = n^{1/\alpha}\ell_1(n), \quad \ell_1 \in R_0, \quad b_n = 0.$
Résultat limite	$a_n^{-1}M_n \xrightarrow{d} \Phi_\gamma$

Exemples

Cauchy	$f(x) = (\pi(1+x^2))^{-1}, \quad x \in \mathbb{R}.$ $a_n = n/\pi$
Pareto Bavure stable avec index $\alpha < 2$	$\bar{F}(x) \sim Kx^{-\alpha}, \quad K, \alpha > 0.$ $a_n = (K_n)^{1/\alpha}$
Loggamma	$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)}(\ln x)^{\beta-1}x^{-\alpha-1}, \quad x > 1, \quad \alpha, \beta > 0.$ $a_n = ((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n)^{1/\alpha}$

TAB. 1.1 – domaine d'attraction maximal de la distribution de Fréchet

Exemple 1.6.1 *Distribution de Cauchy, distribution de Pareto, distribution Loggamma,...*

domaine d'attraction maximal de la distribution de Fréchet

1.6.2 Domaine d'atraction de weibull

Théorème 1.6.3 *$F \in D$(Weibull) (domaine d'attraction de Weibull) avec un indice des valeurs extrême $\gamma < 0$ si et seulement $x_F < +\infty$ et la fonction $(1-\bar{F})$ est à variations régulières d'indice $(\alpha = 1/\gamma)$ avec*

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

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

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

De ce théorème, on en déduit que $F \in D(\text{Weibull})$ si et seulement si le point terminal x_F

est fini et $\bar{F}(x) = (x_F - x)^{-\alpha} \ell \left[(x_F - x)^{-1} \right]$ avec ℓ est une fonction à variations lentes à l'infini et α un réel strictement négatif.

Weibull	$\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}, \quad x < 0, \quad \alpha > 0.$
MDA(Ψ_α)	$x_F < \infty, \quad \bar{F}(x_F - x^{-1}) = x^{-\alpha}\ell(x), \quad \ell \in R_0.$
Constantes de normalisation	$a_n = x_F - F^{\leftarrow}(1 - n^{-1}) = n^{-1/\alpha}, \quad \ell_1 \in R_0, \quad b_n = x_F.$
Résultat Limite	$a_n^{-1}(M_n - x_F) \xrightarrow{d} \Psi_\alpha$

Exemples

Uniforme	$f(x) = 1, \quad x \in (0,1).$ $a_n = n^{-1}, \quad b_n = 1$
Comportement en loi de puissance à x_F	$\bar{F}(x) = K(x_F - x)^\alpha, \quad x_F - K^{-1/\alpha} \leq x \leq x_F, \; K, \alpha > 0$ $a_n = (Kn)^{-1/\alpha}, \quad b_n = x_F$
Béta	$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad 0 < x < 1, \; a, b > 0$ $a_n = (n\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)})^{-1/b}, \quad b_n = 1$

TAB. 1.2 – domaine d'attraction maximal de la distribution de Weibull

Exemple 1.6.2 *Distribution Beta, distribution uniforme ...*

1.6.3 Domaine d'attraction de Gumbel

Théorème 1.6.4 *$F \in D$(Gumbel) (domaine d'atraction de Gumbel) si et seulement si'il existe*

une fonction positive g telle que :

$$\lim_{x \to x_F} \frac{\bar{F}(x + tg(x))}{\bar{F}(x)} = \exp(-t), \qquad t \in \mathbb{R},$$

avec

$$\bar{F}(x) = c(x) \exp\left\{ -\int_z^x \frac{1}{a(t)} dt \right\}, \text{où } \lim_{x \to x_F} c(x) = c.$$

Un choix possible pour la fonction g est

$$g(x) = \frac{1}{\bar{F}(a_n)} \int_{a_n}^{x_F} \bar{F}(t) dt = E(X - x \backslash X > x).$$

Les suites a_n et b_n ont alors la forme suivantes :

$$b_n = F^{\leftarrow}(1 - n^{-1})$$

et

$$a_n = g(b_n) = \frac{1}{\bar{F}(b_n)} \int_{b_n}^{x_F} \bar{F}(t)dt = F^{\leftarrow}(1 - ne^{-1}) - F^{\leftarrow}(1 - n^{-1}).$$

Gumbel	$\Lambda(x)=\exp\{-e^{-x}\}, \quad x\in\mathbb{R}.$
MDA(Λ)	$x_F\leq\infty, \ \ \bar{F}(x)=c(x)\exp\{-\int_{x_0}^{x}\frac{g(t)}{a(t)}dt\}, \quad x_0<x<x_F,$
	où $c(x)\to c>0, \quad g(x)\to 1, \quad g'(x)\to 0 \quad$ comme $x\uparrow x_F.$
Constantes normalisation	$b_n=F^{\longleftarrow}(1-n^{-1}), \qquad a_n=g(d_n).$
Résultat limite	$a_n^{-1}(M_n-d_n)\xrightarrow{d}\Lambda$

Exemples

Type exponentiel	$\bar{F}(x)\sim Ke^{-\lambda x}, \quad K,\lambda>0$ $a_n=\lambda^{-1}$ $b_n=\lambda^{-1}\ln(Kn)$
Ressemblant à un weibull	$\bar{F}(x)\sim Kx^{\alpha}\exp\{-cx^{\tau}\}, \quad K,c,\tau>0,\alpha\in\mathbb{R}.$ $a_n=(c\tau)^{-1}(a^{-1}\ln n)^{1/\tau-1}$ $b_n=(c^{-1}\ln n)^{1/\tau}+\frac{1}{\tau}(c^{-1}\ln n)^{1/\tau-1}\{\frac{\alpha}{c\tau}\ln(c^{-1}\ln n)+\frac{\ln K}{c}\}$

Tab. 1.3 – Domaine d'attraction maximal de la distribution de Gumbel

Gamma	$f(x)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, \quad x>0, \quad \alpha,\beta>0.$ $a_n=\beta^{-1}$ $b_n=\beta^{-1}(\ln n+(\alpha-1)\ln\ln n-\ln\Gamma(\alpha))$
Normal	$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \quad x\in\mathbb{R}.$ $a_n=(2\ln n)^{-1/2}$ $b_n=\sqrt{2\ln n}-\frac{\ln(4\pi)+\ln\ln n}{2(2\ln n)^{1/2}}$
Log-normal	$f(x)=\frac{1}{\sqrt{2\pi}\sigma x}e^{-(\ln x-\mu)^2/2\sigma^2}, \quad x>0, \quad \mu\in\mathbb{R},\sigma>0.$ $a_n=\sigma(2\ln)^{-1/2}d_n$ $b_n=\exp\{\mu+\sigma(\sqrt{2\ln n}-\frac{\ln(4\pi)+\ln\ln n}{2(2\ln n)^{1/2}})\}$

Tab. 1.4 – Distribution de Gamma,distibution de Normale, distribution de Log-normal

Exemple 1.6.3 *Distribution Normale, distribution Exponentielle...*

1.7 La distribution GEV

Grâce aux travaux de Von Mises (1936) et de Jenkinson (1955), on a une forme unifiée de la fonction de répartition de la loi des valeurs extrêmes à un facteur d'échelle et de position prés. Ce théorème montre que la loi limite des extrêmes a toujours la même forme. Les trois formules précédentes peuvent être combinées en une seule paramétrisatiion :

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

où H est une fonction de répartition non-dégénérée et γ est le paramètre de forme que l'on appelle indice des valeurs extrême (IVE) ou indice de queuse. Cette loi est appelée loi de valeurs extrême généralisée (Generalized Extrême Value) que l'on note GEV. La forme la plus générale de la GEV est :

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

où $\mu \in \mathbb{R}$ et $\sigma > 0$ sont respectivement les paramètres de localisation et de dispertion. La fonction de densité standard correspondante $h_{\zeta,\mu,\sigma}(x)$

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

Suivant le signe du paramètre de forme, on définit trois types de GEV :

1. $\gamma = 0$, lois à queue légère (distribution de Gumbel).

2. $\gamma > 0$, lois à queue lourde (distribution de Fréchet).

3. $\gamma < 0$, lois à queue bornée (distribution de Weibul).

Proposition 1.7.1 *On a, les relations suivantes entre* Λ , Φ_γ *et* Ψ_γ *en terme de* $H_{\gamma,\mu,\sigma}$

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

1.8 Estimation de l'indice des queues

1.8.1 Estimation du maximun de vraisemblance

Soit un échantillon de k maxima $Z_1, ..., Z_k$ i.i.d. suivant une loi GEV. Pour $\xi \neq 0$, la fonction de log-vraisemblance est donnée par :

$$\mathcal{L}(Z, \xi, b_n, a_n) = -k \log a_n - (\frac{1}{\xi} + 1) \sum_{i=1}^{k} \log(1 + \xi \frac{Z_i - b_n}{a_n}) - \sum_{i=1}^{k} (1 + \xi \frac{Z_i - b_n}{a_n})^{-1/\xi}.$$

Pour le cas $\xi = 0$ on obtient :

$$\mathcal{L}(Z, 0, b_n, a_n) = -k \log a_n - \sum_{i=1}^{k} \exp(\frac{Z_i - b_n}{a_n}) - \sum_{i=1}^{k} (\frac{Z_i - b_n}{a_n}).$$

Smith (1985) a démontré les consistance et la normalité asymptotique de cet estimateur lorsque $\xi > 1/2$ et $m \longrightarrow \infty$:

$$\sqrt{m}((\hat{a}_n, \hat{\xi} \ , \hat{b}_n) - (a_n, \xi, b_n)) \longrightarrow N(0, I^{-1}),$$

où I est la matrice d'information de Fisher estimée par sa version empirique

$$I(\Theta) = -E(\frac{\partial^2 \mathcal{L}(Z; \Theta)}{\partial \Theta^2}).$$

$\mathcal{L}(Z;\Theta)$ est la fonction de log-vraisemblance associée à la loi de la variable aléatoire Z, paramétrèe par un ensemble de paramètres Θ.

1.8.2 Estimation de Hill

Définition 1.8.1 *Soient $X_1,...,X_n$ des variables aléatoire indépendantes et de même fonction de répartition F appartenant au domaine d'atraction de Fréchet avec un indice des valeurs extrêmes $\xi > 0$. L'estimateur de Hill de ξ est*

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

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

1.8.3 Estimateur de Pickands

Théorème 1.8.1 *Soit $(X_n, n \geq 1)$ une suite de variable aleatoire indépendante de même fonction de répartition $F \in D(H(\xi))$, ou $\xi \in \mathbb{R}$.*
Si $\lim\limits_{n\longrightarrow\infty} k(n) = \infty$ et $\lim\limits_{n\longrightarrow\infty} \frac{k(n)}{n} = 0$, alors l'estimateur de Pickand

$$\xi^p_{(k(n),n)} = \frac{1}{\log 2}\log\frac{X_{(n-k(n)+1,n)} - X_{(n-2k(n)+1,n)}}{X_{(n-2k(n)+1,n)} - X_{(n-4k(n)+1,n)}},$$

converge en probabilité vers ξ.

Proof. Pour $\xi \in \mathbb{R}$, on a avec le choix $t = 2s$ et $y = 1/2$

$$\lim_{t\longrightarrow\infty}\frac{U(t) - U(t/2)}{U(t/2) - U(t/4)} = 2^{\xi}.$$

■

En fait en utilisant la croissance de U qui se déduit de la croissance de F, on obtient

$$\lim_{t\longrightarrow\infty}\frac{U(t)-U(t_{c1}(t))}{U(t_{c1}(t))-U(t_{c2}(t))}=2^{\xi},$$

dés que

$$\lim_{t\longrightarrow\infty}c_1(t)=1/2 \text{ et } \lim_{t\longrightarrow\infty}c_2(t)=1/4,$$

il reste donc à trouver des estimateurs pour $U(t)$ soit $(k(n),\ n\geq 1)$ une suite d'entiers telle que,

$$1(n)\geq n/4 \quad , \quad \lim_{t\longrightarrow\infty}k(n)=\infty \text{ et } \lim_{t\longrightarrow\infty}\frac{k(n)}{n}=0.$$

Nous ecrivons k pour $k(n)$. Soit $(V_{(1,n)},...,V_{(n,n)})$ la statistique d'ordre d'un echantillon de variables aléatoires indépendantes de loi de pareto. On note $F_v(x)=1-\frac{1}{x}$, pour $x\geq 1$, la fonction de repartition de la loi de pareto. On a les convergences en probabilités suivantes :

$$V_{(n-k+1,n)}\longrightarrow\infty,$$

$$\frac{V_{(n-2k+1,n)}}{V_{(n-k+1,n)}}\longrightarrow 1/2,$$

$$\frac{V_{(n-4k+1,n)}}{V_{(n-k+1,n)}}\longrightarrow 1/4.$$

On en deduit donc que la concergence suivante a lieu en probabilité

$$\frac{U\left(V_{(n-k+1,n)}\right)-U\left(V_{(n-2k+1)}\right)}{U\left(V_{(n-2k+1)}\right)-U\left(V_{(n-4k+1)}\right)}\longrightarrow 2^{\xi}.$$

Il reste maintenant à determiner la loi de $U(V_{(1,n)}),...,U(V_{(n,n)}))$. Remarquons que si $x\geq 1$ alors,

$$U(x)=F^{-1}(F_V(x)),$$

on a donc

$$U(V_{(1,n)}), ..., U(V_{(n,n)})) = (F^{-1}(F_V(V_{(1,n)})), ..., F^{-1}(F_V(V_{(n,n)}))),$$

où F_V est la fonction de répartition de la loi de pareto. On deduit du lemme que le vecteur aleatoire

$$(F^{-1}(F_V(V_{(1,n)})), ..., F^{-1}(F_V(V_{(n,n)}))),$$

à la même loi que $(X_{(1,n)}, ..., X_{(n,n)})$ la statistique d'ordre d'un echantillon de n variables aleatoires indépendantes dont la loi à pour fonction de repartition F. Donc la variable aleatoire

$$\frac{U\left(V_{(n-k+1,n)}\right) - U\left(V_{(n-2k+1)}\right)}{U\left(V_{(n-2k+1)}\right) - U\left(V_{(n-4k+1)}\right)}$$

à la même loi que

$$\frac{X_{(n-k(n)+1,n)} - X_{(n-2k(n)+1,n)}}{X_{(n-2k(n)+1,n)} - X_{(n-4k(n)+1,n)}},$$

ainsi cette quantité converge en loi vers 2^{ξ} quand n tend vers l'infini. Comme la fonction logarithme est continue sur $\mathbb{R}_+^*$. On en déduit que l'estimateur de pickand converge en loi vers ξ. Mais comme ξ est une constante, ona également la convergence en probabilité.

1.8.4 Estimateur des moments

Cet estimateur est par Dekkers et de Haan (1989). C'est une généralisation de l'estimateur de Hill, valable pour tout $\xi \in \mathbb{R}$:

$$\hat{\xi}^{(M)}_{X,k,n} = \hat{\xi}^{(H)}_{X,k,n} + 1 - \frac{1}{2}(1 - \frac{\hat{\xi}^{(H)}_{X,k,n}}{M^{(2)}_n})^{-1}.$$

On a la normalité asymptotique suivante lorsque n et k tendent vers l'infini, pour $\hat{\xi} \geq 0$:

$$\sqrt{k}(\hat{\xi}^{(M)}_{X,k,n} - \xi) \longrightarrow N(0, 1 + \xi^2).$$

Chapitre 2

Estimation de quantile extrême

2.1 Estimation de quantile extrême

Dans tout ce qui suit, on fait l'hypothèse que F appartient à l'un des domaines d'attractions définis précédemment. Afin de résumer le problème d'estimation investigué dans ces travaux, on introduit le résultat suivant appelé approximation de Poisson.

2.1.1 Quelques résultats sur les stastistiques d'ordre

Pour des raisons de commodités, commençons par rappeler un résultat sur la transformation des quantiles.

Lemme 2.1.1 *(**Transformation des quantiles**). Soient $\{X_i,\ i=1,...,n\}$ des variables aléatoires indépendantes et de fonction de répartition F. Soient $\{V_{1,n},...,V_{n,n}\}$ les statistiques d'ordre associées aux variables aléatoires indépendantes de loi uniforme standard $\{V_i, i=1,...,n\}$. Alors,*

$\cdot F^{\longleftarrow}(V_1) \overset{\mathcal{L}}{=} X_1$.

$\cdot \{F^{\longleftarrow}(V_{1,n}),...,F^{\longleftarrow}(V_{n,n})\} \overset{\mathcal{L}}{=} \{X_{1,n},...,X_{n,n}\}$, pour tout $n \in \mathbb{N}$.

$\cdot F(X_1)$ suit une loi uniforme standard si et seulement si F est continue.

Ce lemme nous assure que l'étude des quantiles empiriques associés à une loi quelconque peut se déduire de l'étude des statistiques d'ordre associées à la loi uniforme sur $[0, 1]$. donnons à présent un résultat sur les statistiques d'ordre de loi uniforme et exponentielle. Soit $\{E_i, i = 1, ..., n\}$ une suite de variables aléatoires de loi exponentielle standard. On note $T_n = \sum_{i=1}^{n} E_i$. Soit $\{V_i, i = 1, ..., n\}$ une suite de variables aléatoires indépendantes de loi uniforme standard.

Lemme 2.1.2 ***(Réprésentation de Rényi (1953))****. La suite de variables aléatoires* $\{V_{1,n}, V_{n,n}\}$ *la même loi que la suite de variables aléatoires* $\{\frac{T_1}{T_{n+1}}, ..., \frac{T_n}{Tn+1}\}$*,ie* $\{V_{j,n} \overset{\mathcal{L}}{=} \frac{T_j}{T_{n+1}}\}_{(j=1,...,n)}$.

Ce lemme nous permet d'établir la convergence faible d'une suite de statistiques d'ordres de loi exponentielle.

Proposition 2.1.1 *Soit* $(K_n)_{n \geq 1}$ *une suite d'entiers telle que* $1 \leq K_n \leq n, K_n \longrightarrow \infty$ *et* $K_n/n \longrightarrow 0$ *quand* $n \longrightarrow \infty$*. Alors,*

$$\sqrt{K_n} \left(E_{n-K_n+1,n} - \log(n/K_n)\right) \xrightarrow{\mathcal{L}} N(0, 1).$$

Définition 2.1.1 *Pour tout* $t > 0$*, la fonction queue (en anglais* $\ll$*tail quantile function* $\gg$*) est donnée par*

$$U(t) \overset{def}{=} \inf\{x : F(x) \geq 1 - 1/t\} = q_{1/t}.$$

L'utilité de cette définition est essentiellement d'ordre pratique. En effet, il peut être plus commode de travailler non pas sur la fonction de survie $\bar{F}$ elle-même, mais plutôt sur la fonction queue U.

Théorème 2.1.1 *(De Haan et Ferreira (2006)). Soit* $(K_n)_{n \geq 1}$ *une suite d'entiers telle que* $1 \leq K_n \leq n, K_n \longrightarrow \infty$ *et* $K_n/n \longrightarrow 0$ *quand* $n \longrightarrow \infty$*. Si*

$$\lim_{x \longrightarrow x_F^-} \frac{\bar{F}(x) F''(x)}{(F'(x))^2} = -\gamma - 1,$$

alors

$$\sqrt{K_n}\left(\frac{X_{n-k_n+1,n}-U(n/k_n)}{(n/k_n)U'(n/k_n)}\right)\xrightarrow{\mathcal{L}} N(0,1).$$

Ce théorème montre que l'estimateur de quantile extrême $X_{n-K_n+1,n}$ avec $K_n \overset{def}{=} \lfloor na \rfloor$ est asymptotiquement gaussien. Aussi, remarquons que la proposition (2.1.1) est une cas particulier du théorème (2.1.1) avec $U = \log$.

Proposition 2.1.2 *Dans le cas particulier où $\gamma > 0$, si la condition $\lim\limits_{x\longrightarrow\infty} \dfrac{xU^{'}(x)}{U(x)} = \gamma$ est vérifiée alors, on a*

$$\sqrt{K_n}\left(\frac{X_{n-K_n+1,n}}{(qk_n/n}-1\right)\xrightarrow{\mathcal{L}} N(0,\gamma^2).$$

La proposition (2.1.2) est encore un cas particulier du Théorème (2.1.1). Il apparait que pour des lois dont la fonction de répartition $F \in D$(Fréchet), l'estimateur de quantile extrême $X_{n-k_n+1,n}$ est asymptotiquement normal avec variance asymptotiquement proportionnelle à γ^2/k_n.

2.1.2 Approche des quantiles extrêmes par la loi des valeurs extrêmes

Pour estimer le quantile q_{a_n}, Guida et Longo (1988) a utilisé l'approximation $P(X_{n,n} \leq a_nx+b_n) = F^n(a_nx+b_n) \simeq H_\gamma(x)$. En effet d'après le Théorème (1.5.1), on peut écrire

$$\lim_{n\longrightarrow\infty} n\log F(a_nx+b_n) = \log H_\gamma(x),$$

soit encore

$$\lim_{n\longrightarrow\infty} n\log[1-\bar{F}(a_nx+b_n)] = \log H_\gamma(x).$$

comme $n \longrightarrow \infty$, on peut montrer que $a_nx+b_n \longrightarrow x_F$ et par conséquent $\bar{F}(a_nx+b_n)$

converge vers 0. Un développement limité au premier ordre de $\log(1+u)$ donne donc

$$\bar{F}(a_n x + b_n) \sim -\frac{1}{n} \log H_\gamma(x).$$

Pour tout γ, on peut alors approcher le quantile q_{a_n} par :

$$q_{a_n} \simeq \; a_n x_{a_n} + b_n \text{ où } x_{a_n} \text{ vérifier } - \log H_\gamma(x) = n a_n.$$

On a alors un estimateur de quantile extrême de type

$$\begin{aligned} \hat{q}_{a_n} &= \hat{a}_n x_{a_n} + \hat{b}_n \\ &= \begin{cases} \hat{a}_n (n a_n)^{-\hat{\gamma}} + \hat{b}_n & \text{si } F \in D(\text{Fréchet}) \\ -\hat{a}_n (n a_n)^{-\hat{\gamma}} + \hat{b}_n & \text{si } F \in D(\text{Weibull}) \\ -\hat{a}_n \log(n a_n) + \hat{b}_n & \text{si } F \in D(\text{Gumbel}) \end{cases}, \end{aligned}$$

où $(\hat{a}_n, \hat{b}_n)$ et $\hat{\gamma}$ sont respectivement des estimateurs des suites (a_n, b_n) et de l'indice de queue γ.

Dans le cas particulier où $\gamma = 0$, les autres proposent d'utiliser l'approche basée sur la loi GEV dont le résultat s'énonce ainsi.

Théorème 2.1.2 *(Weinstein (1973)). Soit $F \in D(Gumbel)$, il existe deux suites $(a_n)_{n\geq 1} > 0$ et $(b_n)_{n\geq 1} \in \mathbb{R}$ telles que pour tout $x \in \mathbb{R}$ et $v > 0$,*

$$\lim_{n\to\infty} n\bar{F}\left[(b_n^v + c_n x)^{1/v}\right] = \exp(-x),$$

où $c_n = a_n v b_n^{v-1}$.

Dans une telle situation, on approche le quantile par

$$q_{\alpha_n} \simeq (b_n^v + c_n x_{\alpha_n})^{1/v} \text{ où } x_{\alpha_n} \text{ vérifie } \exp(-x_{\alpha_n}) = n\alpha_n,$$

et un estimateur du quantile extrême est obtenu en remplaçant les suites b_n et c_n respectivement par leurs estimateurs $\hat{b}_n$ et $\hat{c}_n$, i.e

$$\hat{q}_{\alpha_n} = (\hat{b}_n^v - \hat{c}_n \log(n\alpha_n))^{1/v}.$$

L'avantage d'utiliser le résultat (2.1) provient du fait qu'il existe des valeurs de v pour lesquelles la convergence dans (2.1) est plus rapide que dans le cas $v = 1$. Dans ce cas, sur simulation, l'approximation du quantile q_{α_n} est de meilleure qualité que l'approximation basée sur l'approche EVD, i.e avec $v = 1$. Les auteurs fournissent la valeur optimale du paramètre v.

Les paramètres γ, a_n, b_n et c_n de ces lois peuvent être estimès par la méthode du maximun de vraisemblance (Prescott et Walden, 1980, 1983) ou la méthode des moments pondérés (Hosking et al., 1985). Dans le cas de l'approche EVD, Smith (1985) fait une étude détaillée du comportement asymptotique des estimateurs des paramètres γ, a_n, b_n obtenus par la méthode du maximun de vraisemblance. Toutefois, il est conseillé d'utiliser les estimateurs des moments pondérés car ceux-ci sont non seulement explicites et faciles à calculer mais aussi parce qu'ils donnent de meilleurs résultats que les estimateurs du maximun de vraisemblance quand on a des échantillons de petite ou de moyenne taille.
La principale difficulté de l'estimation des paramètres γ , a_n, b_n et c_n est due au fait qu'il faut un échantillon de maxima, laquel est parfois difficile à extraire des données initiales.

2.1.3 Approche des quantiles extrêmes par la méthode des excès

Avant de présenter cette approche, il convient de commencer par une définition .

Définition 2.1.2 ***(loi de Pareto Généralisée (GPD)).*** *La loi de Pareto Généralisée est la loi dont la fonction de répartition est donnée par*

$$\zeta_{\gamma,\beta}(y) = \begin{cases} 1-(1+\gamma\frac{y}{\beta})^{1/\gamma} & si\ \gamma \neq 0\ \ et\ \beta > 0 \\ 1-\exp(-\frac{y}{\beta}) & si\ \gamma = 0\ \ et\ \beta > 0 \end{cases},$$

avec $y \in \mathbb{R}_+$ si $\gamma \geq 0$ ou $[0, -\beta/\gamma[$ si $\gamma < 0$.

Dans l'expression précédente, β représente le paramètre d'échelle et γ le paramètre de forme : il s'agit du même paramètre de forme introduit dans la partie (1.5) et que l'on appelle indice des valeurs extrêmes.

La loi GPD présente quelques particularités. En voici une liste non exhaustive :

·Si $\beta = 1$,on parle la loi GPD standard.

·Si $\gamma = 0$, la GPD correspond à une loi exponentielle d'espérance β.

·Si $\gamma = -1$, elle correspond à une loi uniforme sur $[0, \beta]$.

·Si $\gamma > 0$, on retrouve la loi de pareto décentrée.

Dans cette approche d'estimation des quantiles extrêmes, on ne retient que les observations dépassant un seuil fixé $u < x_F$. On définit alors l'excés Y de la variable X au dessus du seuil u par $X - u$ sachant $X > u$. Si l'on note par F_u la fonction de répartition d'un excés au dessus du seuil u, on a pour tout $y > 0$

$$\begin{aligned} 1 - F_u(y) &= \mathbf{P}(Y > y) \\ &= \mathbf{P}(X - u > y \backslash X > u) \\ &= \frac{\mathbf{P}(X > u + y, X > u)}{\mathbf{P}(X > u)} \\ &= \frac{1 - F(u + y)}{1 - F(u)}. \end{aligned}$$

Lorsque le seuil u est grand, on peut approcher cette quantité par la fonction de suivie d'une loi GPD. Afin d'approcher le quantile, il suffit alors d'utiliser le résultat de Balkema et de Haan (1974) et Pickands(1975) qui établit l'équivalence entre la convergence en loi du maximum vers une loi des valeurs extrêmes H_γ et la convergence en loi d'un excés vers une GPD. Ce résultat s'énonce comme suit.

Théorème 2.1.3 *(Balkema et de Haan (1974) et Pickands(1975)). Si F appartient au domaine d'attraction de H_γ , alors*

$$\lim_{u\to x_F} \sup_{y\in]0,x_F-u[} |F_u(y) - \zeta_{\gamma,\beta}(y)| = 0.$$

D'aprés ce résultat, si pour une fonction de répartition F inconnue, l'échantillon des maxima normalisés converge en loi vers une distribution non dégénérée, alors il s'en déduit que la distribution des excés au-dessus d'un seuil élevé converge vers une GPD lorsque le seuil tend vers la limite supérieure du sipport de F. Cette caractérisation est à la base des méthodes d'estimation de type Peaks Over Yhreshold (POT).
Comme $1 - F(u+y) = [1 - F(u)][1 - F_u(y)]$, si pour tout $y \geq 0$ on pose $q_{\alpha_n} = u + y$,alors

$$\begin{aligned} \alpha_n = 1 - F(q_{\alpha_n}) &= [1 - F(u)][1 - F_u(q_{\alpha_n} - u)] \\ &\simeq [1 - F(u)][1 - \zeta_{\gamma,\beta}(q_{\alpha_n} - u)]. \end{aligned}$$

Pour k_n excès au-dessus du seuil u, l'approximation $1 - F(u) \simeq k_n/n$ conduit à

$$\frac{k_n}{n}(1 - \zeta_{\gamma,\beta}(q_{\alpha_n} - u)) \simeq \alpha_n,$$

et si $y \neq 0$, alors on approche le quantile par

$$q_{\alpha_n} \simeq u + \frac{\beta}{\gamma}((\frac{k_n}{n\alpha_n})^{\gamma} - 1).$$

On a alors un estimateur de type

$$\hat{q}_{\alpha_n} = \frac{(\frac{k_n}{n\alpha_n})^{\gamma} - 1}{\hat{\gamma}}\hat{\beta} + u,$$

où $\hat{\gamma}$ et $\hat{\beta}$ sont respectivement des estimateurs des paramètres de forme et d'échelle. On peut noter la similitude entre l'estimateur de quantileet l'expression du quantile avec $\hat{\beta} = \hat{\alpha}_n$ et $u = \hat{b}_n$.

Les paramètres γ et β de la GPD peuvent être estimés par la méthode des moments, la méthode des moments pondérées (Hosking et Wallis, 1987) ou la méthode du maximum de vraisemblance (Smith, 1987 ; Davison et Smith, 1990).

Cette méthode présente un avantage par rapport à la précédente en ce sens qu'il est plus facile d'avoir un échantillon d'éxcès que de maxima. Dans la pratique, on remplace u par $X_{n-K_n+1,n}$ c'est-à-dire la K_n plus grande observation de l'échantillon $\{X_i, i = 1, ..., n\}$.

Deux variantes de cette méthode ont été présentées par Breiman et al. (1990) sous les apppellations Exponential tail (**ET**) et qadratique tail (**QT**).

2.1.4 Approche des quantiles extrêmes par l'approche semi-paramétrique

On se retreint aux fonctions $F \in D(\text{Fréchet})$ pour lesquelles on a la caractérisation suivante

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

avec ℓ une fonction à variations lentes à l'infini et $\alpha > 0$. Conformément au Lemme de (Inverse d'une fonction à variation régulières), cette caractérisation implique que

$$q_{\alpha_n} := \bar{F}^{\longleftarrow}(\alpha_n) = \alpha_n^{-\gamma} L(1/\alpha_n) \text{ avec } \alpha_n \leq 1/n,$$

$$q_{\beta_n} := \bar{F}^{\longleftarrow}(\beta_n) = \beta_n^{-\gamma} L(1/\beta_n) \text{ avec } \beta_n \geq 1/n,$$

où L est une fonction à variations lentes à l'infini. En ce qui concerne les fonctions L et ℓ, il apparait important de signaler ici qu'il ne s'agit pas de la même fonction à variations lentes.

Vu la définition d'une fonction à variations lentes, pour β_n suffisamment petit, on a

$$\bar{F}^{\longleftarrow}(\alpha_n) \simeq \bar{F}^{\longleftarrow}(\beta_n)(\frac{\beta_n}{\alpha_n})^{\gamma}.$$

En remplaçant $\bar{F}^{\longleftarrow}(\beta_n)$ et γ par des estimateurs, on obtient l'estimateur de Weissman (1978) défini par

$$\hat{q}_{\alpha_n}^{W} = X_{n-\lfloor n\beta_n \rfloor+1,n}(\frac{\beta_n}{\alpha_n})^{\hat{\gamma}}.$$

Pour les propriétés de l'estimateur de Weissman, on peut se référer à l'ouvrage de Embrechts et al. (1997).

Comme autre estimateur des quantiles extrêmes, on peut citer celui obtenu par l'approximation

$$\bar{F}^{\longleftarrow}(\alpha_n) \simeq \frac{(\beta_n/\alpha_n)^{\gamma} - 1}{1 - 2^{-\gamma}}(\bar{F}^{\longleftarrow}(\beta_n) - \bar{F}^{\longleftarrow}(2\beta_n)) + \bar{F}^{\longleftarrow}(\beta_n),$$

et valable quel que soit le domaine d'attraction de la fonction F. La normalité asymptotique de l'estimateur de quantile extrême qui en découle, i.e

$$\hat{q}_{\alpha_n}^{DH} = \frac{(\beta_n/\alpha_n)^\gamma - 1}{1 - 2^{-\gamma}}(X_{n-\lfloor n\beta_n\rfloor+1,n} - X_{n-\lfloor 2n\beta_n\rfloor+1,n}) + X_{n-\lfloor n\beta_n\rfloor+1,n},$$

a été établie par Dekkers et de Haan (1989). Il apparait clairement que cet estimateur de quantile extrême peut se mettre sous la forme (Estimation du paramètre de loi des valeurs extrêmes) et donc (Caractérisation des domaines d'attractions) avec

$$\hat{a}_n = \frac{\hat{\gamma}}{1 - 2^{-\hat{\gamma}}}(X_{n-\lfloor n\beta_n\rfloor+1,n} - X_{n-\lfloor 2n\beta_n\rfloor+1,n}) \text{ et } \hat{b}_n = X_{n-\lfloor n\beta_n\rfloor+1,n}.$$

Conclusion

Pour conclure, la théorie des valeurs extrêmes comme dans plusieurs applications, montre comment faire face à des sinistres graves (évènements extrêmes). Dans notre cas de tels évènements sont rares, et leurs prévisions ou estimations doivent souvent être établies avec une grande prudence et en marge des données disponibles. Les modèles doivent être utilisé de féçon souple, en restant paragmatique.
Nous avons vu que contrairement à la théorie des grands nombres, la loi asymtotique du maximum (resp. monimum) normalisé tend vers trois types de lois appelées lois des valeurs extrêmes et distingués selon le signe du paramètre de forme ξ (Frèchet ($\xi > 0$), Gumbel ($\xi = 0$) et Weibull ($\xi < 0$)).

Bibliographie

[1] Arnold, Barry C., Balakrishnan, N. et Nagaraja, H.N., (1992). A first course in order statistique. Wiley Series in probability and Mathematical Statistique : Probability and Mathematical Statistique. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1992.

[2] BALKEMA, A. et de HAAN.L. (1974). Residual life time at a great age. Annzls Of Probability, $792-804$.

[3] BRAHIMI, B. (2018). Sur l'estimateur des moments pour la queue de distribution. Mémoire de magister, Université Mohamed Khider, Algeria.

[4] BREIMAN, L., STONE, C. J. et KOOPERBERG, C. (1990). Robust confidence bounds for extreme upper quantiles. Journal Of Statistical Computation and Simulation, $127-149$.

[5] DEKKERS, A, L., EINMAHL, J. H. , &De HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. The Annals of Statistics, $1833-1855$.

[6] De HAAN, L. et FERREIRA, A. (2006). Extreme Value Theory : An Introduction. Springer Series in Operations Research and Financial Engineering, New York Inc.

[7] DAVISON, A. C. et SMITH, R. L. (1990). Models forexcudances over high thresholds. Journal Of The Royal Statistical Society, B. $393-442$.

[8] EMBRECHTS, P., KLÜPPELLERG, C. et MIKOXH, T. (1997). Modelling Extremal Events for Insurance and Finance. Springer verlag.

[9] GUIDA, M. et LONGO, M. (1988). Estimation of probability tails based on generalized extreme value distributions. Reliability Engineering and System Safety, $219-242$.

[10] HOSKING, J. R. M. et WALLIS, J. R.(1987). Parameter and quantile estimation for the generalized Pareto distribution. Technometrics, $339-1349$.

[11] HOSKING, J. R. M., WALLIS, J. R. et WOOD, E .F.(1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted comments. Technometrics, $251-261$.

[12] JENKINSON, A. F. (1955). The frequency distribution of the annual maximum(or minimum) values of meteorological elements. The Quarterly Journal of The Royal Meteorological Society, $158-171$.

[13] LEKINA, A. (2010). Estimation non-paramétrique des quantiles extrêmes conditionnels. Thèse Docteur, Université de Grenoble.

[14] NELSEN, R. B. (2006). An introduction to copules, second ed. Springer, New York.

[15] PIKANDS, J. (1975). Statical inference using extreme order statics. Annals Of Statistics, $119-131$.

[16] PRESCOTT, P. et WALDEN, A. T. (1980). Maximum likelihood estimation of the parametres of generalized extreme-value distribution.Biometrika, $723-724$.

[17] PRESCOTT, P. et WALDEN, A. T. (1983). Maximum likelihood estimation of the parameters of the three-parameter generalized extreme-value distribution for censored samples. Journal Of Statistical Computation and Simulation, $241-250$.

[18] Reiss, R.D. et Thomas, M. (2007). Statistical Analysis of Extreme Value with Applications to Insurance, Finance, Hidrology andOther Fields. Birkhäuser, Basel.

[19] RéNYI, A. (1953). On the theory of order statistics. Acta Mathimatica Hungarica, $191-231$.

[20] SMITH, R. (1987). Estimating tails of probability distributions. Annals Of Statistiques, $1174-1207$.

[21] SMITH, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika, $67-92$.

[22] WEISSMAN, L. (1978). Estimation of parameters and large quantiles based on the klargest observations. Journal Of The American statistical Association, $812-815$.

Annexe B : Abréviations et Notations

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

$\mathbb{E}[X]$	Espérance mathématique ou moyenne du v.a. X.
F	Fonction de répartition
F_n	Fonction de répartition empirique
$(X_1, ..., X_n)$	Échantillon de taille n de v.a's
$i.i.d$	Indépendantes et identiquement distribuées
$\xrightarrow{P}$	Convergence en probabilité
$\xrightarrow{L}$	Convergence en loi
F_n	Fonction de distribution empirique
$\bar{F}$	Fonction de survie
$F^{\leftarrow}$	L'inverce généralisé de F
(Ω, F, P)	Espace de probabilité
$N(0,1)$	Loi normale standard
$(X_{1,n}, ..., X_{n,n})$	Statistiques d'ordre associé à $(X_1, ..., X_n)$
$X_{1,n}$	Minimum de $X_1, ..., X_n$
$X_{n,n}$	Maximum de $X_1, ..., X_n$
$X_{i,n}$	I-ème statistique d'ordre
Q	Fonction des quantiles
Q_n	Fonction des quantiles empirique
$U(t)$	Fonction des quantiles de queue
$U_n(t)$	Fonction des quantiles de queue empirique
Φ_α	Loi de Frèchet
Ψ_α	Loi de Weibull
Λ	Loi de Gumbel
GPD	Généralised Pareto distribution
$P.S$	Presque sûre
H_ξ	Famille des lois des extrême généralisées
$D(H_\xi)$	Domaine d'attraction
$\hat{\xi}^M$	Estimateur des moments
$\hat{\gamma}^H$	Estimateur de Hill
$\hat{\xi}^P$	Estimateur de Pickands

$\mathbb{I}$	La matrice d'information de Fisher estimé.
x_F	Point terminale.
x_p	Quantile d'ordre p.
$\stackrel{d}{=}$	Égalité en distribution.
TCL	Théorème centrale limite.
EVD	Extrem Value Yheory
u	Seuil
GEV	Lois de valeurs extrêmes généralisées
POT	Type over yhreshold

Résume

Dans ce mémoire nous avons la théorie des valeurs extrêmes , et caractérisation des domaines d’attraction puis la distribution GEV.

Il existe trois type de domaine d’attraction sont (Frèchet,Weibull,Gumbel).

Abstract

In This Memory we have the theory of extreme values, and characterization some domains of attraction then the distribution GEV.

There are three type of area of attraction are (Frèchet,Weibull,Gumbel)

ملخص

في هذه المذكرة درسنا نظرية القيم المتطرفة و خصائص مناطق الجذب ثم التوزيع لقانون القيم المتطرفة.

هناك ثلاثة انواع من مناطق الجذب (فريشيت ، ويبل، قيمبل)