République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université Mohammed Seddik Ben Yahai - Jijel

Faculté des Sciences Exactes et Informatique
Département de Mathématiques

№ d'ordre : .....

№ de séries : .....

Mémoire de fin de cycle

Présenté pour l'obtention du diplôme de

Master

Filière : Mathématiques.

Option : Probabilités et statistique.

Thème

Distribution des valeurs extrêmes généralisées - Application en hydrologie

Présenté par :

- Tilbi Khaira.

Devant le jury :

Présidente	: Laoudj Farida	M.C.A Université de Jijel
Encadreur	: Gherda Mebrouk	M.A.A Université de Jijel
Examinateur	: Chraitia Hassen	M.C.B Université de Jijel

Promotion **2017/2018**

Remerciements

Je glorifie Allah le tout puissant de m'avoir donnée courage et patience qui m'ont permis d'accomplir ce travail.

Un travail, quel qu'il soit, n'est jamais individuel, de nombreuses personnes m'ont aidé et ont contribué, chacun à sa manière, à la réalisation de ce travail :
Mon encadreur, **M.Guerda Mebrouk**, merci de m'avoir laissé une grande liberté dans mon travail, et je vous remercie encore pour votre disponibilité et la confiance dont vous m'avez a fait preuve.
Mon enseignante **Laoudj**, je vous remercie pour tout le temps que vous m'avez offert et des précieux conseils pour m'aider à avancer, et j'ai l'honneur que vous allez participer au jury de mon mémoire de fin d'étude comme présidente et examinatrice.
M.Guerda et **M.Laoudj** merci pour vos hautes qualités humaines. Je garderai toujours avec moi le souvenir de notre premier contact et de notre discussion....
Je remercie **Monsieur Chraitia**, pour son aide dans ce travail ainsi d'avoir accepté de jury mon mémoire.
Dans le cadre d'application, je remercie **M.Dahoui Kamel** directeur du barrage Beni-Haroun (Mila) pour son accueil et son aide de nous avoir fourni les données hydrauliques pour la réalisation de l'application qui ont été très utiles pour notre travail et ses orientations ainsi que ses conseils.
Je n'oublie pas notre doyen de la faculté **M.bounames** pour sa disponibilité et son aide concernant les sorties et le stage, et aussi un remerciement pour le directeur de laboratoire Mathématiques et applications des mathématiques **M.Kerada** pour sa compréhensibilité et sa disponibilité.
Je remercie une personne qui était toujours avec moi, elle m'a encouragé, m'aider, me conseiller **Izza Sabrina**.
Un remerciement pour mes amies **Saliha**, **Zineb**, **Fahima** et **Soulef**.

Dédicaces

À mes parents

À la mémoire de ma grand mère

...

Table des matières

Table des figures

Liste des tableaux

Pourquoi les valeurs extrêmes ?

Pour un profane, la statistique est associée à la notion de moyenne ou l'écart-type. En effet dans de nombreuses applications, notamment en sciences sociales et en sciences de la physique, les statistiques se résument généralement au calcul des moyennes et à l'évaluation de dispersion d'une série de valeurs autour de leur moyenne.
Par définition, les événements rares sont des événements ayant une faible probabilité d'apparition. Lorsque le comportement de ces événements est dù au hasard on peut étudier leur loi. Ils sont dite extrêmes quand il s'agit de valeurs beaucoup plus élevées que les autres valeurs ou plus faibles que celles observées habituellement.
Les événements extrêmes peuvent être catastrophiques lorsque il s'agit (tremblements de terre , inondations, accidents nucléaires,...) dominent l'actualité quotidienne par leur caractère imprévisible compte tenu de l'importance des enjeux sociaux et scientifiques. Aucun débat sérieux sur le hasard ne serait être mené sans une réflexion sur les événements rares et extrêmes.
"La loi des grands nombres et la distribution gaussienne, fondements de l'étude statistique des grandeurs moyennes, échouent à rendre compte des événements rares ou extrêmes. Pour ce faire, des outils statistiques plus adaptés existent ... mais ne sont pas toujours utilisés !"(Rama Cont-Papers-Pour La Science -Décembre 2009).
Dés lors, la question que l'on pourrait se poser est de savoir ce que peuvent les outils statistiques face aux événements extrêmes ? Autrement dit, peut-on réellement prévoir ou quantifier le risque des événements extrêmes ?

La théorie des valeurs extrêmes (TVE) fournit une base mathématique et probabiliste rigoureuse sur laquelle il est possible de construire des modèles statistiques pour prévoir la taille et la fréquence de ces phénomènes dans les queues de distribution.
Le comportement extrême des lois appartenant aux différents domaines d'attraction maximal (DAM) est significativement différent. Les lois appartenant aux DAM de Weibull sont bornées à droite, celles appartenant aux DAM de Gumbel et Fréchet ont un support infini

à droite. Mais les premières ont des queues finies alors les secondes ont des queues épaisses et exposent donc à des situations extrêmes beaucoup plus dangereuses.
Les domaines d'application sont en effet trés variés : hydrologie, météorologie, biologie, ingénierie, gestion de l'environnement, finance, assurance, sciences sociales,...etc.
Les catastrophes naturelles sont des exemples d'évènements extrêmes qui conduisent à des pertes financières importantes. Les cracks boursiers sont d'autres exemples qui conduisent à des pertes financières très importantes.

Il existe essentiellement deux approches dans la modélisation des extrêmes : la méthode des blocs de maxima et celle de dépassement des excès. L'approche des maxima uni-variés établit qu'une famille paramétrique généralisée résume le comportement asymptotique de la loi du maximum convenablement normalisé. Dans la seconde approche, il est établi que seule la distribution généralisée de Pareto qui modélise la loi de la variable excédentaire au sel d'un certain seuil fixé et assez élevé.

Ce mémoire s'organise en deux parties.
La première partie s'intitule « Théorie des valeurs extrêmes » dont l'objectif est d'exposer la théorie probabiliste des valeurs extrêmes dans le cas uni varié. Elle se compose de trois chapitres :
• Au cours de premier chapitre, on présente tout d'abord quelques définitions de certains outils nécessaires dans la théorie des valeurs extrêmes.
• Dans le deuxième chapitre, on expose la théorie probabiliste des valeurs extrêmes dans le cas uni varie. D'abord, on a commencé par le théorème fondamental de la théorie des valeurs extrêmes (théorème de Fisher-Tippet) qui assure que la loi limite maximum est surement une des trois lois (Gumbel- Weibull-Fréchet). Puis on a unifié les trois lois dans une seule représentation qui est la représentation de Von-Mise (GEV) et qui sert à estimer l'indice de queue. Ensuite, on a défini les domaines d'attraction et les conditions pour que chaque distribution appartienne au max-domaine d'attraction associé. Après, on a passé à l'estimation des paramètres des valeurs extrêmes avec deux méthodes différentes. Puis, on a estimé les quantiles extrêmes afin d'arriver à le but principale qui est l'estimation de la période de retour et le niveau de retour.
• Dans ce chapitre, nous définissons une autre approche des valeurs extrêmes qui est basée sur la distribution de Pareto généralisée , On constate que le deuxième théorème fondamental des V.E (théorème de Balkema- Pickands-De Haan) est considéré comme le deuxième théorème fondamental des valeurs extrêmes. Puis, on a estimé les paramètres de

cette distribution mais après la détermination du seuil , et nous estimons les quantiles et la période de retour. Finalement, on a exposé une approximation de la période de retour et le niveau de retour.

La deuxième partie est une partie d'application. Malgré les difficultés rencontrés sur le terrain, nous avons réussi a réalisé une application dans laquelle on a fait une étude de cas réel pour modéliser la distributions des valeurs extrêmes (une fois dans la GEV et une autre dans GPD) et on a conclut par une comparaison entre les résultats obtenus dans les deux approches (GEV et GPD).

Première partie :

Partie théorique

Chapitre 1

Préliminaires

1.1 Loi indéfiniment divisible

L'idée principale des lois indéfiniment divisible vient du problème suivant : comment peut-on déterminer toutes les lois de probabilité qui s'expriment comme limites (en loi) d'une suite de sommes de n variables aléatoires indépendantes et identiquement distribuées à valeurs dans $\mathbb{R}$?

Définition 1.1.

Une variable aléatoire réelle X est dite textbfindéfiniment divisible si :

$$\forall n \in \mathbb{N}^* \ , \ \exists X_1, X_2, ..., X_n \text{ telles que } X \overset{D}{=} X_1 + X_2 + ... + X_n$$

avec : $X_1, X_2, ..., X_n$ sont des variables aléatoires indépendante et identiquement distribuées à valeurs dans $\mathbb{R}$.

Définition 1.2.

Soient X une variable aléatoire à valeurs dans $\mathbb{R}$ et $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$ la fonction caractéristique de sa loi de probabilité. On a :

$$X \text{ est indéfiniment divisible } \iff \forall n \in \mathbb{N}^*, \exists \varphi : \mathbb{R} \longrightarrow \mathbb{C} \text{ tel que } \forall t \in \mathbb{R}, \varphi(t) = [\varphi_n(t)]^n$$

avec : φ_n est une fonction caractéristique d'une certaine loi de probabilité.

1.1.1 Propriétés des lois indéfiniment divisibles

Théorème 1.3.

1- Si $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$ est la fonction caractéristique d'une loi de probabilité indéfiniment divisible, alors :
$\forall t \in \mathbb{R}$

$$\varphi(t) \neq 0$$

2- Si $(X_n)_{n\in\mathbb{N}^*}$ est une suite de variables aléatoires réelles indéfiniment divisibles telle que :

$$X_n \xrightarrow{D} X$$

alors : X est une variable aléatoire indéfiniment divisible.
3- Si $X_1, ..., X_n$ sont des variables aléatoires réelles indéfiniment divisibles alors : $\sum_{i=1}^{n} X_i$ est une variable aléatoire indéfiniment divisible.

1.2 Lois max-stables

La famille des lois max-stables est une famille introduite a partir de la famille des lois indéfiniment divisibles.

Définition 1.4.

La variable aléatoire non dégénérée X ou, la fonction de distribution F de X est dite ***max-stable****, s'il existe des constantes $a_n \in \mathbb{R}_+^*$ et $b_n \in \mathbb{R}$ telles que :*
Pour tout $n \in \mathbb{N}$

$$M_n \overset{\mathcal{D}}{=} a_n X + b_n$$

où $M_n = \max_{i\leq n} X_i$.

Définition 1.5.

Soit X une variable aléatoire non dégénérée et $(X_i)_i$ une suite de variable aléatoire réel de même loi que X. On dit que X est ***max-stable*** *s'il existe des constantes $a_n \in \mathbb{R}_+^*$ et $b_n \in \mathbb{R}$ telles que :*
$\forall n \in \mathbb{N}^*$

$$\max_{i\leq n} \frac{X_i - b_n}{a_n} \overset{\mathcal{D}}{=} X$$

Proposition 1.6.

La variable aléatoire non dégénérer X, la fonction de distribution F de X est dite max-stable si et seulement si

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

$$F^n(a_n x + b_n) = F(x)$$

Proposition 1.7.

Les trois lois ***Gumbel, Fréchet*** *et* ***Weibull*** *sont des lois max-stable.*

En effet :

1- Considérons la loi de Fréchet du paramètre $\alpha = 1$

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

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

Choisissons $a_n = n$ et $b_n = 0$

$$F^n(a_n x + b_n) \stackrel{?}{=} F(x)$$

$$F^n(a_n x + b_n) = F^n(nx) = \begin{cases} \exp(-(nx)^{-1})^n & \text{si } nx > 0 \\ 0 & \text{si } nx \leq 0 \end{cases}$$

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

$$= F(x)$$

Donc : $\forall x \in \mathbb{R}$, $\forall n \in \mathbb{N}^*$ $\exists a_n = n$ et $b_n = 0$ tel que $F^n(a_n x + b_n) = F(x)$

Alors la distribution de Fréchet est max-stable.

2- Considérons la loi de Weibull

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

Choisissons $a_n = n^{-1/\alpha}$ et $b_n = 0$

$$H^n(a_n x + b_n) \stackrel{?}{=} H(x)$$

$$\begin{aligned} H^n(a_n x + b_n) &= H^n(n^{-1/\alpha} x) \\ &= \exp(-(-n^{-1/\alpha} x)^\alpha)^n \\ &= \exp(-(-n^{-1} x^\alpha) n) \\ &= \exp(-(-x^\alpha)) \\ &= H(x) \end{aligned}$$

Donc : $\forall x \in \mathbb{R}$, $\forall n \in \mathbb{N}^*$ $\exists a_n = n^{-1/\alpha}$ et $b_n = 0$ tel que $H^n(a_n x + b_n) = H(x)$
Alors la distribution de Weibull est max-stable.

1.3 Statistique d'ordre

Soit $X_1, X_2, ..., X_n$ n variables aléatoires indépendantes et identiquement distribuées **iid** de densité commune f, et de fonction de répartition F.

1.3.1 Définitions et notations

Définition 1.8.

On appelle ***statistique d'ordre*** *notées $X_{(1,n)}, X_{(2,n)}, ..., X_{(n,n)}$, les variables aléatoires ordonnées :*

$$\min(X_1, X_2, ..., X_n) = X_{(1,n)} \leq X_{(2,n)} \leq ... \leq X_{(n,n)} = \max(X_1, X_2, ..., X_n)$$

* La statistique d'ordre de l'échantillon $(X_1, ..., X_n)$ est le réarrangement croissant (aléatoire) de $(X_1, ..., X_n)$ telle que $X_1 < X_2 < ... < X_n$ que l'on note $(X_{(1,n)}, ..., X_{(n,n)})$. En particulier, on note :

$$X_{(1,n)} = \min_{1 \leq i \leq n} (X_i)$$

et

$$M_n = X_{(n,n)} = \max_{1 \leq i \leq n} (X_i)$$

On s'intéresse à la variable aléatoire M_n.

Proposition 1.9.

Soit M_n une variable aléatoire qui représente la plus grande valeur observée sur les n observées. Comme les variables aléatoires sont indépendantes et identiquement distribuées, on obtient :

1- La distribution du maximum $F_{X_{(n,n)}}$ de la statistique d'ordre extrême M_n est donnée par :

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

$$\begin{cases} F_{M_n}(x) = \Pr\left(X_{(n,n)} \leq x\right) = [F(x)]^n \\ f_{X_{(n,n)}}(x) = n[F(x)]^{n-1} f(x) \end{cases} \tag{1.1}$$

2- La distribution du minimum $F_{X_{(1,n)}}$ de la statistique d'ordre extrême $X_{(1,n)}$ est donnée par :

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

$$\begin{cases} \Pr\left(X_{(1,n)} \leq x\right) = 1 - [1 - F(x)]^n \\ f_{X_{(1,n)}}(x) = n[1 - F(x)]^{n-1} f(x) \end{cases} \tag{1.2}$$

où $F(x)$ est la fonction de distribution des X_i et f est la densité des X_i.

1.3.2 Densité de la k-ème statistique d'ordre

Soit $x \in \mathbb{R}$ fixé. Les variables aléatoires $(\mathbf{1}_{(X_i \leq x)}, i \geq 1)$ sont des variables aléatoires indépendantes et de même loi de Bernoulli de paramètre $\Pr(X_i \leq x) = p$.
La variable aléatoire $S_n = \sum_{i=1}^n \mathbf{1}_{X_i \leq x}$ suit donc la loi binomial de paramètre (n, p), donc $S_n(x) \geq k$ si et seulement si $X_{(k,n)} \leq x$. Ainsi il vient :

$$\{X_{(k,n)} \leq x\} = \{S_n(x) \geq k\}$$

et on a le résultat suivant

Lemme 1.10.

La loi de la variable aléatoire $X_{(k,n)}$, pour $1 \leq k \leq n$ est donnée par :

$$F_{X_{(k,n)}}(x) = \Pr[X_{(k,n)} \leq x] = \sum_{j=k}^{n} C_n^j [F(x)]^j [1 - F(x)]^{n-j}$$

Sa densité est :

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

Corollaire 1.11.

Soit $X_1, ..., X_n$ n variables aléatoires (iid) de fonction de répartition F continue, alors la densité de $X_{(1,n)}, X_{(2,n)}, ..., X_{(n,n)}$ est donnée par :

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

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

1.3.3 Convergence de la fonction de répartition empirique

Théorème 1.12.

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

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|=0 \quad p.s$$

1.4 Fonction à variations régulière

Définition 1.13.

*On dit qu'une fonction l est **à variations lentes** à l'infini si $l(x)>0$ pour x assez grand et si pour tout $\lambda>0$ on a :*

$$\lim_{x\to\infty}\frac{l(\lambda x)}{l(x)}=1$$

1.4.1 Théorème de Karamata

Théorème 1.14.

Toute fonction à variation lente l s'écrit sous la forme

$$l(x)=c(x)\exp\int_1^x\frac{\varepsilon(t)}{t}dt \tag{1.3}$$

où $c>0$ et ε sont deux fonctions mesurables telles que :
$\lim_{x\to\infty}c(x)=0\in\,]0,\infty[$ *et* $\lim_{x\to\infty}\varepsilon(x)=0$
Si la fonction c est constante, alors onn dit que l est normalisée. L'équation (1.3) implique que si l est normalisée alors l est dérivable de dérivée l' avec :
$\forall x>0$

$$l'(x)=\frac{\varepsilon(x)l(x)}{x}$$

En particulier on a :

$$\lim_{x\to\infty}x\frac{l'(x)}{l(x)}=0$$

Définition 1.15.

*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 c-à-d s'il existe A tel que pour tout $x\geq A$, $G(x)\geq 0$ et si $\forall\lambda>0$*

$$\lim_{x\to\infty}\frac{G(\lambda x)}{G(x)}=\lambda^p$$

Si $p=0$, G est une fonction à variation lentes à l'infini.

Remarque 1.

Une fonction à variation régulières d'indice p peut toujours s'écrire sous la forme $x^p l(x)$ où l est une fonction à variations lentes à l'infini.

Lemme 1.16.

Ce lemme donne un résultat sur l'inverse d'une fonction à variation régulières.
** Si G est une fonction à variations régulières d'indice $p > 0$ alors G^{-1} est une fonction à variations régulières d'indice $1/p$.*
** Si G est une fonction à variations régulières d'indice $p < 0$ alors G^{-1} est une fonction à variations régulières d'indice $-1/p$.*

1.5 Autres définitions

1. **Inverse généralisé**
   Soit $F : \mathbb{R} \longrightarrow]a, b[$ $(-\infty \leq a < b \leq +\infty)$ une fonction croissante.
   On appelle inverse généralisé de F, l'application notée $F^{\leftarrow}$ définie par :
   $\forall p \in]a, b[$
   $$F^{\leftarrow}(p) = \inf\{x : F(x) \geq p\}$$
   avec la convention $\inf\{\emptyset\} = \infty$ avec $p \in [0, 1]$.
   L'inverse généralisé $F^{\leftarrow}$ coïncide avec l'inverse F^{-1} lorsque la fonction F est strictement croissante et continue.
2. **Fonction survie ou fonction de queue**
   On appelle fonction de survie ou fonction de queue de la variable aléatoire X, la fonction :$\bar{F} : \mathbb{R} \longrightarrow [0, 1]$ définie par :
   $\forall x \in \mathbb{R}$
   $$\bar{F}(x) = 1 - F(x)$$
3. **Point terminal**
   Le point terminal d'une distribution F est défini par :
   $$x_F = \sup\{x \in \mathbb{R}; F(x) < 1\} \tag{1.4}$$
4. **Fonction de hasard**
   Si la variable aléatoire X est absolument continue de densité f, alors on appelle fonction de hasard de X la fonction
   $\forall x \in \mathbb{R}$
   $$h(x) = \frac{f(x)}{\bar{F}(x)} = \frac{f(x)}{1 - F(x)}$$

5. **Quantile d'ordre P**

   On appelle quantile d'ordre p, le nombre x_p défini par :

   $$x_p = \inf\{x \in \mathbb{R}; F(x) \geq p\}$$

   avec $p \in [0, 1]$.

6. **Fonction quantile**

   On appelle fonction quantile de la variable aléatoire X la fonction Q définie par :
   $\forall p \in]0; 1[$

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

7. **Fonction quantile de queue**

   On appelle fonction quantile de queue de la variable aléatoire X la fonction U : $]1, \infty[\longrightarrow \mathbb{R}$ définie par :
   $\forall x \in]1, \infty[$

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

   où Q est la fonction quantile de X.

8. **Fonction de Von-Mises**

   La fonction de répartition F de la variable aléatoire X qui a un point terminal x_F est appelée fonction de Von-Mises s'il existe $a < x_F$ tel que
   $\forall x \in]a, x_F[$

   $$\bar{F}(x) = 1 - F(x) = c \exp\left(-\int_a^x \frac{1}{\delta(u)}\right)$$

   avec $c > 0$ et δ est une fonction absolument continue sur $]a, x_F[$ de densité δ' telle que
   $\forall x \in]a, x_F[$, $\delta(x) > 0$ et $\lim\limits_{x \to x_F} \delta'(x) = 0$.

9. **Quantile extrême**

   On appelle quantile extrême le quantile d'ordre $(1 - p)$, définie par :

   $$\begin{aligned} x_p &= \inf\{x \in \mathbb{R} : F(x) \geq p\} \\ &= F^{-1}(1-p) \end{aligned}$$

   où p proche de zéro.

10. **Lois à queue lourde**

    Une fonction F est dite à queue lourde si pour tout $a > 0$, $b > 0$

    $$\bar{F}(x) > ae^{-b}$$

Chapitre 2

Distribution des valeurs extrêmes généralisées

Motivation : La distribution des valeurs extrêmes généralisée est très utile en application de la théorie des valeurs extrême, car c'est la seule et unique loi de probabilité qui modélise le comportement du maximum d'un échantillon. Pour estimer ces paramètres, on utilise la méthode du "maximum par blocs" qui consiste à construire un échantillon de maximum à partir d'un échantillon de données en formant des blocs de même dimension.

2.1 Introduction

La théorie des valeurs extrêmes a le but d'étudier la loi du maximum d'une suite des variables aléatoires, cette théorie s'énonce de la façon suivante :
soit X_i une suite de variables aléatoires indépendantes et identiquement distribuées (iid) de fonction de répartition F qui est définie par

$$F(X) = \Pr(X \leq x)$$

L'étude des extrêmes passe par l'analyse du maximum d'un échantillon de taille n(n assez grand), ordonné, noté $M_n = Max(X_1, X_2, ..., X_n)$. Cette analyse est appelée analyse des maxima par bloc. Grâce a la caractérisation (iid) on aura :

$$\begin{aligned}\Pr(M_n \leq x) &= \Pr(X_1, X_2, ..., X_n \leq x) \\ &= \Pr(X_1 \leq x) \Pr(X_2 \leq x) ... \Pr(X_n \leq x) \\ &= \prod_{i=1}^{n} \Pr(X_i \leq x) \\ &= [F(x)]^n \end{aligned} \tag{2.1}$$

Généralement, Fest inconnue et la relation (2.1) n'est pas utilisable directement ; la théorie des valeurs extrêmes donne le comportement asymptotique de la variable M_n A partir de l'idée du théorème centrale limite (TCL), la théorie des valeurs extrêmes montre que s'il existe des suites normalisatrices $(a)_{n>0}n$ et $(b_n)_{n>0}$ sont des suites de normalisation avec $a_n > 0$ et $b_n \in \mathbb{R}$ telle que :

$$\lim_{n \to \propto} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n \to \infty} F(a_n x + b_n)^n = G \tag{2.2}$$

où G est une variable aléatoire non dégenerée.

Question : Quelle est la loi de G ?

Réponse : Le théorème suivant répond à cette question.

2.2 Théorème Fisher-Tippet(1928)

Théorème 2.1.

Sous certaines conditions de régularité sur la fonction de répartition F, s'il existe un paramètre réel α et deux suites $(a_n)_{n>0}$ et $(b_n)_{n>0}$ tels que $a_n > 0$ et $b_n \in \mathbb{R}$ et pour tout $x \in \mathbb{R}$

$$\lim_{n\to\infty} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F(a_n x + b_n)^n = G_\alpha(x)$$

alors $G_\alpha(x)$ est l'une des trois distributions suivantes :

***Gumbel** :*

pour tout $x \in \mathbb{R}$

$$\Lambda(x) = \exp(-\exp(-x))$$

***Fréchet** :*

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

***Weibull** :*

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

G_α est appelée fonction de répartition de la loi des valeurs extrêmes.

Démonstration.

Pour une démonstration de ce théorème, on pourra se référer aux ouvrages : Hann and Frreira ou Resnick (voir [7] ou [11]). ■

Commentaires sur le théorème :

1. On a une seule limite dans le TCL (la loi normale) par contre dans le cas des extrêmes trois limites sont possibles (*Gumbel, Fréchet et Weibull*).
2. $(a_n)_{n\geq 1}$ est le paramètre d'échelle,représente $\dfrac{\sigma(x)}{\sqrt{n}}$.
3. $(b_n)_{n\geq 1}$ est un paramètre de position,il représente $\mathbb{E}(x)$ dans le TCL.

Le paramètre $\alpha > 0$ qui apparaisse dans les types de distribution de **Fréchet** et de **Weibull** est appelé **paramètre de forme** (indice de queue).

Plus α est petit, plus la décroissance est lente, plus des valeurs extrêmes sont susceptibles de ce produire.
Plus α est grand,plus la décroissance est forte, moins des valeurs extrêmes vont se présenter.

Résultat important du théorème

Le théorème donne un résultat très intéressant :quelle que soit la loi limite des extrêmes a toujours la même forme ; bien que le comportement différent, elle peuvent être combinées en une seul paramétrisation contenant un unique paramètre ξ qui contrôle la lourdeur de la queue de loi appelé indice des valeurs extrêmes ou indice de queue.

2.2.1 Propriété de queue des lois extrêmes Gumbel, Fréchet et Weibull

A. Distribution de Gumbel

Proposition 2.2.

$\bar{\Lambda}(x) = 1 - \Lambda(x) = 1 - \exp(-e^{-x}) \sim e^{-x}$ *quand* $x \longrightarrow \infty$. *c-à-d : pour les grandes valeurs, la queue de la distribution décroit très rapidement (de façon exponentielle).*
La queue de ***Gumbel*** *est une fonction à variation rapide à l'infini d'indice* $-\infty$. *Donc c'est une lois à queue* ***légère****.*

B. Distribution de Fréchet

Proposition 2.3.

$\forall \alpha > 0$, $\bar{\Phi}_\alpha(x) = 1 - \Phi_\alpha(x) = 1 - \exp(-x^{-\alpha}) \sim x^{-\alpha}$ *quand* $x \longrightarrow \infty$. *Pour les grandes valeurs, la queue de cette distribution décroit d'une façon polynomiale .*
La queue de Φ_α *est une fonction à variation régulière à l'infini d'indice* $-\alpha$. *Donc on déduit que la queue de Fréchet est* ***lourde*** *à droite.*

C. Distribution de Weibull

La queue de la distribution de **Weibull** est finie, c-à-d c'est une distribution bornée à droite, donc elle a peu d'intérêt dans l'étude des événements extrêmes.

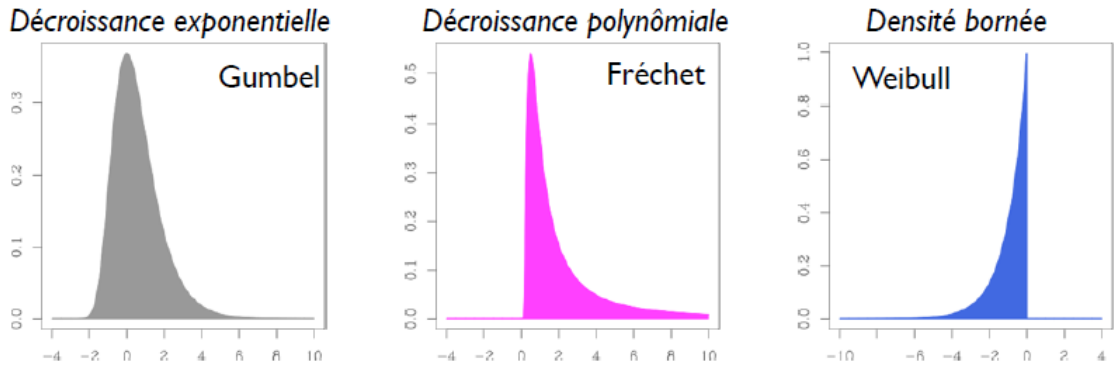

FIGURE 2.1 – queues des 3 lois extrêmes

*** Relation entre les trois lois**

Proposition 2.4.

La relation entre les trois distributions des valeurs extrêmes est donnée comme suit :

$X \sim$ ***Fréchet*** $\Longrightarrow Y = \ln X \sim$ ***Gumbel***.

$X \sim$ ***Weibull*** $\Longrightarrow Y = \dfrac{1}{X} \sim$ ***Fréchet***.

$X \sim$ ***Weibull*** $\Longrightarrow Y = \ln\left(\dfrac{1}{X}\right) \sim$ ***Gumbel***.

* Grâce au travaux de **Von Mises (1936)** et **Jenkinson(1955)** on a une forme unifiée de la fonction de répartition de la loi **EVD** à un facteur d'échelle et de position.

2.2.2 Représentation de Jenkinson-Von Mises(1954)

Proposition 2.5.

La représentation de ***Jenkinson-Von Mises*** *de la distribution des valeurs extrêmes****EVD*** *que l'on appelle loi des valeurs extrêmes généralisée notée* ***GEV*** *à pour fonction de répartition :*

$$G_\xi = \begin{cases} \exp\left(-\left(1+\xi x\right)\right)^{-1/\xi} & si\ \xi \neq 0, 1+\xi x > 0 \\ \exp\left(-\exp(-x)\right) & si\ \xi = 0, x \in \mathbb{R} \end{cases} \tag{2.3}$$

Pour obtenir une forme plus générale de la loi **GEV** en introduisent les paramètres de localisation μ et d'échelle σ on obtiendra :

$$G_{\xi,\mu,\sigma}(x) = \begin{cases} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right)\right) & \text{si } \xi \neq 0, 1+\xi(\frac{x-\mu}{\sigma}) > 0 \\ \exp\left(-\exp-\left(\frac{x-\mu}{\sigma}\right)\right) & \text{si } \xi = 0 \end{cases} \tag{2.4}$$

Remarque 2.

À partir de cette écriture, on peut distinguer 3 cas :

$\xi = 0$ *correspond à la loi de* ***Gumbel****.*

$\xi > 0$ *correspond à la loi de* ***Fréchet****.*

$\xi < 0$ *correspond à la loi de* ***Weibull****.*

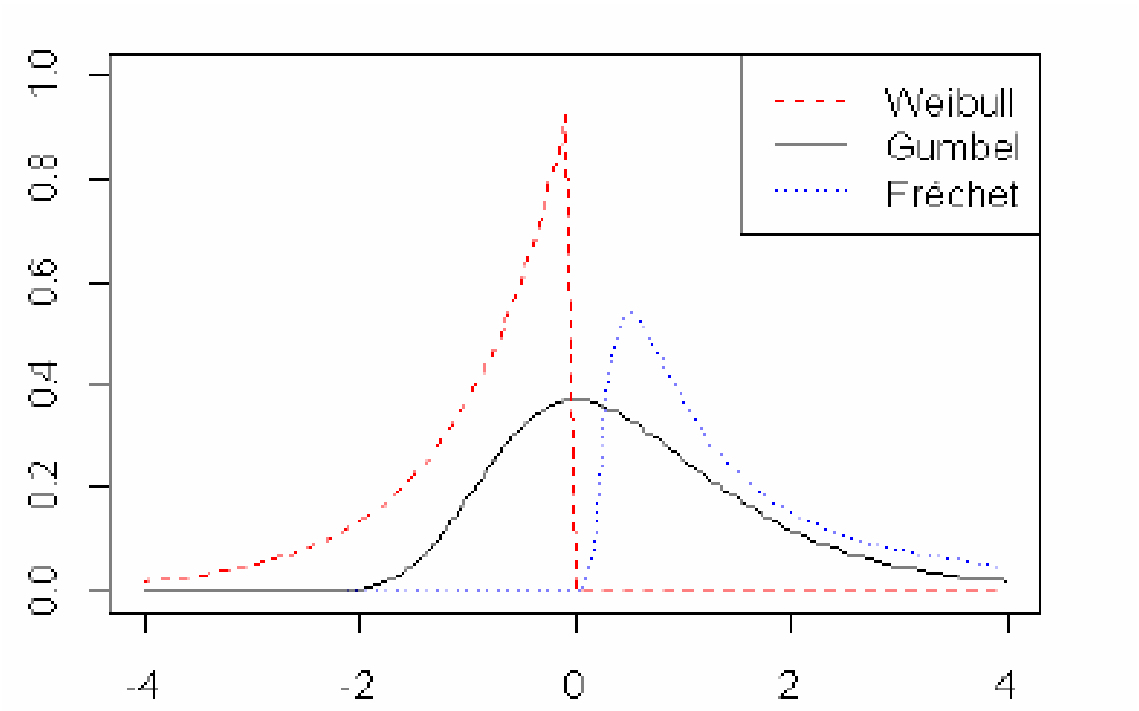

FIGURE 2.2 – Densité standard de GEV

Autres question :

1. Quelle est la relation entre F et G ?
2. Comment on peut choisir les coefficients de normalisation a_n et b_n ?

Réponse : La réponse de ces deux questions est dans la partie suivante.

2.3 Domaines d'attraction et coefficients de normalisation

La classe des fonctions de distribution F satisfaisant (2.2) appelé max-domaine d'attraction.

La recherche du domaine d'attraction peut être considérée comme l'étude réciproque de la recherche de la distribution des valeurs extrêmes associée à une distribution,étant donné G ,quelles sont les conditions nécessaires et/ou suffisante sur la variable aléatoire X pour que la limite $\lim_{n\to\infty} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F(a_n x + b_n)^n = G$ soit réalisée ?

Définition 2.6.

On appelle domaine d'attraction G l'ensemble des lois F pour les quelles le maximum normalisé suit asymptotiquement la loi de G.

Définition 2.7.

On dit qu'une distribution F appartient au max-domaine d'attraction d'une distribution des valeurs extrêmes G et on note $F \in MDA(G)$, si la distribution du maximum normalisé converge vers G 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 :

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

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

Exemple 2.8.

La loi exponentielle du paramètre 1

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

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

$$\begin{aligned}
\lim_{n\to\infty} \Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) &= \lim_{n\to\infty} \Pr\left(\frac{M_n - \ln n}{1} \leq x\right) \\
&= \lim_{n\to\infty} F\big(x + \ln n\big)^n \\
&= \lim_{n\to\infty} \left(1 - e^{-x+\ln n}\right)^n \\
&= \lim_{n\to\infty} \left(1 - \frac{e^{-x}}{n}\right)^n \\
&= e^{-e^{-x}} \\
&= \Lambda(x)
\end{aligned}$$

car $\lim_{n\to\infty} \big(1 - \frac{x}{n}\big)^n = e^{-x}$

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

Exemple 2.9.

On suppose que X *suit la loi uniforme* $\mathcal{U}([0,1])$

c-à-d :

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

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

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

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

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

$$= \begin{cases} 0 & si\ 1 + \frac{x}{n} < 0 \\ \left(1 + \frac{x}{n}\right)^n & si\ 0 \leq 1 + \frac{x}{n} \leq 1 \\ 1 & si\ 1 + \frac{x}{n} > 1 \end{cases}$$

$$= \begin{cases} 0 & si\ x < -n \\ \left(1 + \frac{x}{n}\right)^n & si\ -n \leq x \leq 0 \\ 1 & si\ x > 0 \end{cases}$$

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

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

$$= \Psi_1(x)$$

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

⋆ Les définitions précédentes concernant le domaine d'attraction basent sur le choix des facteurs de normalisation a_n et b_n. Mais souvent le choix n'est pas facile ; on propose d'autre théorèmes et des propositions sur le domaine d'attraction.

2.3.1 Domaine d'attraction maxima

Théorème 2.10.

Une condition nécessaire et suffisante pour q'une fonction F appartienne au domaine d'attraction maximal de $G_\alpha(x)$ est :

$$\lim_{\epsilon\to 0}\frac{F^{-1}(1-\epsilon)-F^{-1}(1-2\epsilon)}{F^{-1}(1-2\epsilon)-F^{-1}(1-4\epsilon)}=2^c \tag{2.5}$$

Cela implique que

1. *Si $c<0$ alors $F(x)\in MDA(\Psi_\alpha(x))$*
2. *Si $c=0$ alors $F(x)\in MDA(\Lambda_\alpha(x))$*
3. *Si $c>0$ alors $F(x)\in MDA(\Phi_\alpha(x))$*

Application du théorème

1. La loi uniforme
   On a $F^{-1}(\alpha)=\alpha$

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{F^{-1}(1-\epsilon)-F^{-1}(1-2\epsilon)}{F^{-1}(1-2\epsilon)-F^{-1}(1-4\epsilon)}&=\lim_{\epsilon\to 0}\frac{(1-\epsilon)-(1-2\epsilon)}{(1-2\epsilon)-(1-4\epsilon)}\\&=\frac{1}{2}\\&=2^{-1}\end{aligned}$$

   Donc $c=-1<0\Longrightarrow$ la loi uniforme $\in MDA(\Psi_\alpha(x))$

2. La loi de Cauchy
   On a $F^{-1}(p)=tg\big((p-\frac{1}{2})\Pi\big)$

$$\begin{aligned}\lim_{\epsilon\to 0}\frac{F^{-1}\big(1-\epsilon\big)-F^{-1}\big(1-2\epsilon\big)}{F^{-1}\big(1-2\epsilon\big)-F^{-1}\big(1-4\epsilon\big)}&=\lim_{\epsilon\to 0}\frac{tg\Big(\big(\frac{1}{2}-\epsilon\big)\Pi\Big)-tg\Big(\big(\frac{1}{2}-2\epsilon\big)\Pi\Big)}{tg\Big(\big(\frac{1}{2}-2\epsilon\big)\Pi\Big)-tg\Big(\big(\frac{1}{2}-4\epsilon\big)\Pi\Big)}\\&=2^1\end{aligned}$$

   Donc $c=1>0\Longrightarrow F(x)\in MDA(\Phi_\alpha(x))$

2.3.2 Domaine d'attraction de Fréchet, Weibull et Gumbel

A.Domaine d'attraction de Fréchet

Théorème 2.11.

La fonction de répartition F appartient au domaine d'attraction de la distribution de ***Fréchet*** *si et seulement si :*

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

où la fonction L est à variation lente.
En particulier $x_F = +\infty$; de plus si $F(x) \in DA\ \Phi_\alpha(x)$ avec a_n=$U(n)$=$F^{-1}(1-\frac{1}{n})$ alors :
La suite $(a_n^{-1}M_n)_{n>0}$ converge en loi vers une variable aléatoire de fonction de répartition ϕ.

Démonstration.

On suppose que

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

et L est à *variation lente* $\Longrightarrow$ on peut l'écrire sous la forme :

$$L(x) = c(x)\exp\int_a^x \frac{k(u)}{u}du$$

$\Longrightarrow$

$$\bar{F}(x) \approx x^{-\alpha}c\exp\int_a^x \frac{k(u)}{u}du$$

On pose a_n=$U(n)$

$$\bar{F}(x) = U(n)^\alpha c\exp\int_a^{U_n} \frac{k(u)}{u}du = (1-\frac{1}{n})c\exp\int_a^{1-\frac{1}{n}} \frac{k(u)}{u}du = (1-\frac{1}{n})c\exp$$

$$\bar{F}(x) \leq \frac{1}{n} \leq \bar{F}(\bar{a}_n)$$

Si F est continue en a_n,alors $\bar{F}\ (a_n) = \dfrac{1}{n}$
sinon,Comme $\bar{F}$ est équivalente en $+\infty$ à une fonction continue, on déduit que

$$\lim_{n\to\infty} \bar{F}(a_n) = \frac{1}{n}$$

Pour $x > 0$ on a donc :

$$\lim_{n\to\infty} \bar{F}(a_n x) = \lim_{n\to\infty} \frac{\bar{F}(a_n x)}{\bar{F}(a_n)} = x^{-\alpha}$$

■

Proposition 2.12.

Si $F(x) \in MDA\ (\Phi_\alpha(x))$ alors les constantes de normalisations $a_n > 0$ et $b_n \in \mathbb{R}$ telles que :

$$\forall x \in \mathbb{R}\,;\ \lim_{n\to\infty} F(a_n x + b_n)^n = \Phi_\alpha(x)$$

peuvent être choisies de la manière suivante :

$$a_n = U(n) \quad et \quad b_n = 0$$

où U est la fonction quantile de queue de la variable aléatoire X.

Exemple 2.13. *Dans ce tableau on présente quelques exemples de lois de probabilité qui appartient au max-domaine d'attraction de Fréchet*

Distribution de X	*coefficients de normalisation*
$X \sim Cauchy(0,1)$	$a_n = n/\pi$ *et* $b_n = 0$
$X \sim Pareto(\alpha)$	$a_n = n^{1/\alpha}$ *et* $b_n = 0$
$X \sim loggaamma(\alpha,\beta)(\beta > 0)$	$a_n = ((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n)^{1/\alpha}$ *et* $b_n = 0$

B.Domaine d'attraction de Weibull

Théorème 2.14.

La fonction de répartition F appartient au domaine d'attraction de la distribution de ***Weibull*** *si et seulement si :*

$x_F < +\infty$ *et* $\bar{F}\big(x_F - \frac{1}{x}\big) = x^{-\alpha}L(x)$ *où la fonction L est à variation lente.*
De plus ; si $F(x) \in DA\ (\Psi_\alpha(x))$ avec a_n=$x_F - U(n)$=$x_F - F^{-1}(1-\frac{1}{n})$ alors :
La suite $(a_n^{-1}M_n)_{n>0}$ converge en loi vers une variable aléatoire de fonction de répartition ψ.

Proposition 2.15.

Si $F(x) \in (MDA\ \Psi_\alpha(x))$ alors les constantes de normalisations $a_n > 0$ et $b_n \in \mathbb{R}$ telles que :

$$\forall x \in \mathbb{R}\,;\ \lim_{n\to\infty} F(a_n x + b_n)^n = \Psi_\alpha(x)$$

peuvent être choisies de la manière suivante :

$$a_n = x_F - U(n) \quad et \quad b_n = x_F$$

où U est la fonction quantile de queue de la variable aléatoire X et x_F est le point terminal.

Exemple 2.16. *Dans ce tableau on présente quelques exemples de lois de probabilité qui appartient au max-domaine d'attraction de Weibull*

Distribution de X	*coefficients de normalisation*
$X \sim \mathcal{U}[0;1]$	$a_n = 1/n$ *et* $b_n = 1$
$X \sim \beta(\alpha,\theta), (\theta > 0)$	$a_n = \left(\dfrac{\Gamma(\alpha+\theta)}{\Gamma(\alpha+1)\Gamma(\theta)} n\right)^{-1/\alpha}$ *et* $b_n = 1$

C.Domaine d'attraction de Gumbel

Le domaine d'attraction de **Gumbel** est le plus délicat car il est difficile à énoncer du fait qu'il n'existe pas une condition nécessaire et suffisante relativement simple.

Théorème 2.17.

*La fonction de répartition F de la variable aléatoire X avec le point terminal $x_F < +\infty$ appartient au domaine d'attraction de la distribution de **Gumbel** si et seulement s'il existe $a < x_F$ tels que :*

$$\bar{F}(x) = 1 - F(x) = c(x) \exp\left(- \int_a^x \frac{\theta(u)}{\delta(u)} du \right)$$

où c et θ sont des fonctions mesurables sur $]a; x_F[$ vérifiant :

$$\lim_{x\to x_F} c(x) = c > 0 \ et \ \lim_{x\to x_F} \theta(x) = 1$$

et δ est une fonction absolument continue sur $]a; x_F[$ de densité δ' telle que :

$$\forall x \in]a, x_F[\ : \delta > 0 \ et \lim_{x\to x_F} \delta'(x) = 0.$$

Proposition 2.18.

Si $F(x) \in MDA\ (\Lambda_\alpha(x))$ alors :
les constantes de normalisations $a_n > 0$ et $b_n \in \mathbb{R}$ telles que :
$\forall x \in \mathbb{R}$

$$\lim_{n\to\infty} F(a_n x + b_n)^n = \Lambda_\alpha(x)$$

peuvent être choisies de la manière suivante :

$$a_n = \delta(b_n) \quad et \quad b_n = U(n)$$

où U est la fonction quantile de queue de la variable aléatoire X et δ est la fonction définie dans le théorème précédent.

Proposition 2.19.

*Si la fonction de répartition F de la variable aléatoire X est une fonction de **Von-Mises**, alors F appartient au max-domaine d'attraction de **Gumbel**.*

Exemple 2.20.

La distribution exponentielle du paramètre $\lambda > 0$, i.e :

$$H(x) = \begin{cases} 1 - e^{-\lambda x} & si\ x > 0 \\ 0 & si\ x \leq 0 \end{cases}, \lambda > 0$$

*H est une fonction de **Von-Mises**.*
en effet : on a $x_H = +\infty$ et
$\forall x \in]0, +\infty[$

$$\bar{H}(x) = 1 - H(x) = e^{-\lambda x} = \exp\left(- \int_0^x \frac{1}{\lambda^{-1}} du \right)$$

*On déduit que H est une fonction de **Von-Mises** de fonction auxiliaire $\delta(x) = \lambda^{-1}$.*
Alors $H \in MDA\ (\Lambda)$

Exemple 2.21. *Dans ce tableau on présente quelques exemples de lois de probabilité qui appartient au max-domaine d'attraction de Gumbel*

Distribution de X	*coefficients de normalisation*
$X \sim Rayleigh$	$a_n = a\sqrt{\log n}$ *et* $b_n = a/2\sqrt{\log n}$
$X \sim \mathcal{N}(0,1)$	$a_n = (2\ln n)^{-1/2}$ *et* $b_n = \sqrt{2\ln n} - \dfrac{\ln(4\pi) + \ln\ln n}{2\sqrt{2\ln n}}$
$X \sim \Gamma(\alpha, \beta)$	$a_n = \beta^{-1}$ *et* $b_n = \dfrac{\ln n + (\alpha - 1)\ln\ln n - \ln\Gamma(\alpha)}{\beta}$

Les conditions nécessaires et suffisantes pour q'une fonction de répartition F appartienne à un domaine d'attraction d'une distribution extrême sont pas facile à vérifier.
La partie suivante on va présenter une condition suffisante **condition de Von-Mises**, elle est simple à vérifier mais dans le cas où les fonctions de répartitions ayant une densité car cette condition est basée sur *la fonction de Hasard.*

Condition suffisante de Von-Mises(1936)

Théorème 2.22.

Soit F une fonction de répartition absolument continue de densité f et soit la fonction de hasard

$$h(x) = \frac{f(x)}{1 - F(x)}$$

1. *Si $h(x) > 0$, x assez grand et s'il existe $\alpha > 0$ telle que :*
   $\lim_{x\to\infty} xh(x) = \alpha$ *alors :*
   $$F(x) \in MDA(\Phi_\alpha(x))$$
2. *Si $F^{-1}(1) < 1$ et s'il existe $\alpha > 0$ telles que :*
   $\lim_{x\to F^{-1}(1)} (F^{-1}(1) - x)h(x) = \alpha$
   alors :
   $$F(x) \in MDA(\Psi_\alpha(x))$$
3. *Si $h(x) \neq 0$ et h dérivable au voisinage de F^{-1} (où bien $F^{-1} = +\infty$) et si de plus ;*
   $\lim_{x\to F^{-1}(1)} \frac{d}{dx}\left(\frac{1}{h(x)}\right) = 0$ *alors :*
   $$F(x) \in MDA(\Lambda(x))$$

2.3.3 Domaine d'attraction de GEV

Théorème 2.23.

*La fonction de distribution de F de la variable aléatoire X appartient au max-domaine d'attraction de la distribution des valeurs extrêmes généralisés **GEV** G_ξ si et seulement si : il existe une fonction mesurable δ telle que :*

$$\lim_{t\to x_{\bar{F}}} \frac{\bar{F}(t + x\delta(t))}{\bar{F}(t)} = \begin{cases} (1 + \xi x)^{-1/\xi} & si\ \xi \neq 0 \\ e^{-x} & si\ \xi = 0 \end{cases} \tag{2.6}$$

avec $1 + \xi x > 0$ et $x \in \mathbb{R}$.

Proposition 2.24.

$F \in DAG_\xi$ si et seulement si :

$$\lim_{n\to\infty} \bar{F}(a_n x + b_n) = -\log G_\xi(x)$$

Pour une certaine suite $(a_n, b_n)_{n>0}$ où $a_n > 0$ et $b_n \in \mathbb{R}$; on a alors ***la convergence en loi*** *de $a_n^{-1}(M_n - b_n)_n$ vers une variable aléatoire de la fonction de répartition G_ξ.*

Démonstration.

$\Longrightarrow$ On suppose que $F \in DAG_\xi$ alors :

$$\lim_{n\to\infty} (1 - \bar{F}(a_n x + b_n))^n = G_\xi(x) \tag{2.7}$$

En prend le log de (2.7)

$$\lim_{n\to\infty} n \log(1 - \bar{F}(a_n x + b_n)) = \log G_\xi(x)$$

On a : $\forall (1 + \xi x) > 0$

$$\lim_{n\to\infty} \bar{F}(a_n x + b_n)) = 0$$

$\Longrightarrow$

$$\lim_{n\to\infty} n\bar{F}(a_n x + b_n) = -\log G_\xi(x)$$

$\Longleftarrow$ Si on a :

$$\lim_{n\to\infty} \bar{F}(a_n x + b_n) = -\log G_\xi(x)$$

alors : $F \in DA(G_\xi)$(évident)

■

Exemple 2.25.

Soit X ***v.a*** *suit la loi de Pareto standard*

$$F(x) = 1 - x^{-\theta} \text{ avec } \theta \in \mathbb{R}_+^*$$

Ce qui implique que :

$$1 - F(x) = \begin{cases} x^{-\theta} & si\ x > 1 \\ 1 & si\ x \geq 0 \end{cases}$$

On pose : $a_n = n^{1/\theta}$ *;* $b_n = 0$

alors :

$$n[1 - F(a_n x + b_n)] = n[1 - F(n^{\frac{1}{\theta}} x)] = \begin{cases} x^{-\theta} & si\ x > n^{-1/\theta} \\ n & si\ x < n^{-1/\theta} \end{cases}$$

passant à la limite

$$\begin{aligned} \lim_{n \to \infty} n[1 - F(a_n x + b_n)] &= \begin{cases} x^{-\theta} & si\ x > 0 \\ \infty & si\ x \leq 0 \end{cases} \\ &= \begin{cases} -\ln \exp(-x^{\theta}) & si\ x > 0 \\ -\ln \theta & si\ x \leq 0 \end{cases} \\ &= -\ln \Phi_{\alpha}(x) \end{aligned}$$

Donc $F \in MDA\ (\Phi_{\alpha}(x))$.

2.3.4 Résultats obtenus

Considérons les 2 lois suivantes :

1- La loi x-exponentielle

$\forall \lambda, \alpha > 0$

$$F(x) = (1 - (1 - \lambda x)e^{-\lambda x})^{\alpha}$$

,

2- La loi exponentielle généralisée

$\forall \alpha > 0, \forall \lambda > 0$

$$G(x) = (1 - e^{-\lambda x})^{\alpha}$$

La loi x-exponentielle

On a

$$F(x) = (1 - (1 - \lambda x)e^{-\lambda x})^{\alpha}$$

Pour trouver le domaine d'attraction maxima, on va utilisé le théorème (2.10)

D'abord :

$$x = F^{-1}(y) = -\frac{1}{\lambda}(1 + \omega((y^{1/\alpha})e^{-1}))$$

où ω est la fonction de Lambert.

$$\lim_{\epsilon \to 0} \frac{G^{-1}(1 - \epsilon) - G^{-1}(1 - 2\epsilon)}{G^{-1}(1 - 2\epsilon) - G^{-1}(1 - 4\epsilon)} = 1/2 = 2^{-1}$$

$$2^c = 1/2 \Rightarrow c = -1$$

Donc la distribution maximum de la loi exponentielle généralisée appartient au max domaine d'attraction de Weibull (c<0).
On peut calculer les coefficients de normalisation à l'aide de la proposition(2.15)

$$a_n = -\frac{1}{\lambda}\left(1 - \left(1 + \omega\big(\frac{n-1}{n}\big)^{1/\alpha} - 1\big)e^{-1}\right)$$
$$b_n = -\frac{1}{\lambda}$$

La loi exponentielle généralisée

On a

$$y = ((1 - e^{-\lambda x})^{\alpha})$$

Pour trouver le domaine d'attraction maxima, on va utilisé le théorème (2.10)
D'abord :

$$x = -\frac{1}{\lambda}\log(1 - y^{1/\alpha})$$

On applique le théorème (2.10)

$$\lim_{\epsilon \to 0} \frac{F^{-1}(1-\epsilon) - F^{-1}(1-2\epsilon)}{F^{-1}(1-2\epsilon) - F^{-1}(1-4\epsilon)} = 1$$

$$2^c = 1 \Rightarrow c = 0$$

Donc la distribution maximum de la loi exponentielle généralisée appartient au max domaine d'attraction de Gumbel (c=0).
On peut calculer les coefficients de normalisation à l'aide de la proposition (2.18)

$$a_n = -\frac{1}{\lambda}\log\big(1 - \big(\frac{n-1}{n}\big)^{1/\alpha}\big)$$
$$b_n = \frac{1}{\lambda}\big(\log\big(1 - (\frac{n-1}{n})^{1/\alpha}\big)\big) - \log\big(1 - \big(\frac{e^{-1}}{n}\big)^{1/\alpha}\big)$$

Remarque 3. *La fonction de Lambert, c'est la fonction inverse de la fonction f définie par* $f(x) = xe^x$.

Quelques propriétés de la fonction :

$y = xe^x \Rightarrow x = \omega(y)$

$\omega(xe^x) = x$

Remarque importante

La loi de Bernoulli de paramètre $p \in]0,1[$ n'admet pas une convergence du maximum normalisé.

En effet :

Soit X_i une variable aléatoire de loi de Bernoulli de paramètre $p \in]0,1[$ alors :

$$\Pr(X_i = 1) = p = 1 - \Pr(X_i = 0)$$

On pose

$$T = \inf\{k \geq 1; X_k = 1\}$$

Si $n > T$ on aura $M_n = 1$.La loi de T est une loi géométrique de paramètre p.Donc T est finis $p.s$ et M_n est constante égal à 1 $p.s$ alors,à partir d'un certain rang il n'existe pas de suite $(a_n, b_n)_{n\geq 1}$, avec $a_n > 0$ telle que la suite $a_n^{-1}(M_n - b_n)$ converge en loi vers une limite non dégénéré.

Donc il n'existe pas une limite non dégenerée pour la limite de maximum normalisé de la loi de Bernoulli.

De façon analogue, on peut démontrer aussi la même chose pour la loi géométrique et la loi de poisson.

2.4 Estimation des paramètres de la loi des valeurs extrêmes

Dans cette partie , on va donner quelques méthodes d'estimation pour estimer les paramètres des valeurs extrêmes et plus précisément pour estimer l'indice des valeurs extrêmes noté ξ qui joue un rôle très important et essentiel dans le comportement de la loi des valeurs extrêmes **V.E**.
Deux méthodes d'estimation seront discutés :

1. **Estimation par des méthodes paramétriques.**

2. **Estimation par des méthodes semi-paramétriques.**

2.4.1 Estimation par des méthodes paramétriques

Il y a deux méthodes plus fréquentes dans l'estimation paramétrique sont :
La méthode du maximum de vraisemblance et *la méthode des moments.*
Et aussi une autre méthode récente et donne des résultats très important c'est l'estimation par *la méthode des L-moments.*

A.Estimateur de Maximum de vraisemblance

Cet estimateur est le plus classique des estimateurs, donne des résultats asymptotiques efficaces,et les estimateurs obtenus convergent vers les vraie valeurs des paramètres.
Sous l'hypothèse que les X_i sont (iid) $i = \{1, ..., n\}$ et ayant la même loi de la distribution **GEV** avec une densité $g(x)$(g a 3 paramètres sont μ, σ, ξ).
La fonction de vraisemblance s'écrit comme suit :

$$L(\mu, \sigma, \xi) = \prod_{i=1}^{n} g(x_i, \mu, \sigma, \xi)$$

Pour faciliter les calculs, on travaille avec le logarithme des vraisemblances

$$\log L(\mu, \sigma, \xi) = \log \prod_{i=1}^{n} g(x_i, \mu, \sigma, \xi) = \sum_{i=1}^{n} \log g(x_i, \mu, \sigma, \xi)$$

Pour la loi GEV

$$\log L(\mu, \sigma, \xi) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log \left(1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right) - \sum_{i=1}^{n} \left(1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right)^{-1/\xi}$$

Le vecteur $\hat{\theta}=(\hat{\mu},\hat{\sigma},\hat{\xi})$ est solution du système suivant :

$$\hat{\mu} = \begin{cases} \dfrac{\delta \log L(\mu,\sigma,\xi)}{\delta\mu} = 0 \\ \dfrac{\delta^2 \log L(\mu,\sigma,\xi)}{\delta\mu^2} \leq 0 \end{cases}$$

$$\hat{\sigma} = \begin{cases} \dfrac{\delta \log L(\mu,\sigma,\xi)}{\delta\sigma} = 0 \\ \dfrac{\delta^2 \log L(\mu,\sigma,\xi)}{\delta\sigma^2} \leq 0 \end{cases}$$

et

$$\hat{\xi} = \begin{cases} \dfrac{\delta \log L(\mu,\sigma,\xi)}{\delta\xi} = 0 \\ \dfrac{\delta^2 \log L(\mu,\sigma,\xi)}{\delta\xi^2} \leq 0 \end{cases}$$

Il n'existe aucune solution explicite pour résoudre ces systèmes, le problème dans ce cas devient un problème d'optimalité numérique.
Mais à l'aide du langage $\mathcal{R}$ on peut estimer les paramètres de différentes lois avec le maximum de vraisemblance.

Lemme 2.26 (**Comportement des estimateurs MV dans le cas GEV**)**.**

D'après la théorie du ***MV*** *le support de la loi ne dépend pas des paramètres. Mais dans le cas des lois* ***EVD*** *le support dépend des paramètres sauf dans le cas particulier où* $\xi = 0$.
L'estimateur de ***MV*** *dans le cas* ***GEV*** *est* ***consistant, asymptotiquement efficace et asymptotiquement normal*** *pour tout* $\xi > \frac{1}{2}$. *Ce résultat était démontrer par* ***Smith(1985)***.

B.Estimation des paramètres avec la méthode des moments(PWM)

Cette méthode consiste à égaler les moments empiriques et les moments théoriques. L'inconvénient de cette méthode c'est que à partir du 2^{me} ordre les moments empiriques d'un échantillon sont biaisés.

Pour cela on propose d'égaler les les moments empiriques pondérés d'un échantillon aux moments pondérés.

Soit $X_1, X_2, ..., X_n$ un échantillon de fonction de répartition $G_{\mu,\sigma,\xi}$.

On définit :

les moments pondérés d'ordre r par :

$$\beta_r = \mathbb{E}(X {G^r}_{\mu,\sigma,\xi}(x))$$

avec $r \in \mathbb{N}$

Et ***les moments empiriques pondérés d'ordre r*** par :

$$W_r = \frac{1}{n}\sum_{i=1}^{n} X_{(i,n)} G^r_{\mu,\sigma,\xi}(x_{(i,n)})$$

avec $r = 0, 1, 2$. En utilisant la distribution des extrêmes généralisés $G_{\mu,\sigma,\xi}$ on obtient :

$$\beta_r = \frac{1}{r+1}\left[\mu - \frac{\sigma}{\xi}\big[1-(r-1)^{\xi}\Gamma(1-\xi)\big]\right]$$

telle que $\Gamma(t) = \int_0^{\infty} x^{t-1}\exp(-x)dx$ et $\xi < 1$. L'estimateur des moments pondérés $\hat{\theta} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})$ est la solution du système d'équation suivant :

$$\beta_0 = \mu - \frac{\sigma}{\xi}(1-\Gamma(1-\xi)) \tag{2.8}$$

$$2\beta_1 - \beta_0 = -\frac{\sigma}{\xi}(1-2^{\xi})\Gamma(1-\xi) \tag{2.9}$$

$$\frac{3\beta_2 - \beta_0}{2\beta_1 - \beta_0} = \frac{1-3^{\xi}}{1-2^{\xi}} \tag{2.10}$$

En inversant les formules, on obtient (μ, σ, ξ) en fonction de β_0,β_1 et β_2 ;puis en utilise les moments empiriques pondérés on obtiendra :

$$\begin{aligned} W_r &= \frac{1}{n}\sum_{i=1}^{n} X_{(i,n)} F_n^r\big(x_{(i,n)}\big) \\ &= \frac{1}{n}\sum_{i=1}^{n} X_{(i,n)}\left(\frac{i-1}{n}\right)^r \end{aligned}$$

De l'équation(2.10) on obtiendra $\hat{\xi}$.

De l'équation (2.9)

$$\hat{\sigma} = \frac{\hat{\xi}\big(2\hat{W}_1 - \hat{W}_0\big)}{\Gamma\big(1 - \hat{\xi}(2^{\hat{\xi}} - 1)\big)}$$

De l'équation (2.8)

$$\hat{\mu} = \hat{W}_0 + \frac{\hat{\sigma}}{\hat{\xi}}\Big(1 - \Gamma\big(1 - \hat{\xi}\big)\Big)$$

$\star$ Cet estimateur est ***consistant***.

C.Estimation des paramètres avec la méthode des L-moments

La méthode des **L-moments** est équivalente à la méthode des moments de probabilités.

$*$ Les estimateurs par **L-moments** des paramètres du modèle **GEV** lorsque $-1/2 < \xi < 1/2$ sont donnés par :

$$\hat{\xi} \approx 7.8590c + 2.9554c^2$$

$$\hat{\sigma} = \frac{l_2\hat{\xi}}{(1 - 2^{-\hat{\xi}})\Gamma(1 + \hat{\xi})}$$

$$\hat{\mu} = l_1 - \frac{\hat{\sigma}(1 - \Gamma(1 + \hat{\xi})}{\hat{\xi}}$$

où :

$$c = \frac{2}{3 + \tau_3} - \frac{\log 2}{\log 3}$$

et l_1,l_2 et l_3 sont respectivement les estimateurs des **L-moments** d'ordre 1 et 2 et du rapport des **L-moments** .

Ces estimateurs peuvent être définis à partir ***des moments de probabilité pondérés***

$$\beta_r = \mathbb{E}(XG^r(x))$$

Un estimateurs *sans biais* de β_r $(r > 0)$ est :

$$b_0 = \frac{1}{n}\sum_{i=1}^{n} X_j$$

$$b_r = \frac{1}{n}\sum_{i=1}^{n} \frac{(j-1)(j-2)...(j-r)}{(n-1)(n-2)...(n-3)} X_j$$

Et les estimateurs des quatre premiers **L-moments** sont :

$$\begin{aligned} l_1 &= b_0 \\ l_2 &= 2b_1 - b_0 \\ l_3 &= 6b_2 - 6b_1 + b_0 \\ l_4 &= 20b_3 - 30b_2 + 12b_1 - b_0 \end{aligned}$$

En divisant les **L-moments** d'ordre supérieure à 2 par la mesure de dispersion l_2 ; on obtient les estimateurs des rapports des **L-moments** $\tau_r = l_r/l_2$ avec $r = 3, 4$.

En pratique :

Le coefficient de l'asymétrie τ_3 peut s'écrire d'une manière directe de la façon suivante

$$\tau_3 = \sum_{i=1}^{n} \frac{c_i x_i + \dfrac{\bar{X}}{n}}{l_2}$$

où

$$c_i = 6\frac{(i-1)(i-2)}{n(n-1)(n-2)} - 6\frac{i-1}{n(n-1)}$$

et $\bar{X}$ est la moyenne empirique de l'échantillon.

Remarque 4.

1. *Les nombres l_1 et l_2 correspondent respectivement à **la moyenne** et à l'échelle de la distribution.*
2. *Le coefficient de **variation** est défini par : $\tau = l_2/l_1$.*
3. *Le coefficient τ_3 appelé le coefficient **d'asymétrie(Skewness)**, il mesure le degré d'asymétrie.*
4. *Le coefficient τ_4 appelé le coefficient **d'aplatissement (Kurtosis)**, il mesure le degré d'écrasement de la distribution.*

2.4.2 Estimation par des méthodes semi-paramétrique

A.Estimateur de Hill

C'est le plus simple des estimateurs de queue, il a été introduit par **Hill (1975)**, pour estimer d'une manière non paramétrique le paramètre de queue.

Définition 2.27.

Soit $(k_n)_{n>0}$ *une suite d'entier avec* $1 < k_n \leq n$ *et* $n > 0$ *l'estimateur de* ***Hill*** *est défini par :*

$$\hat{\xi}^H_{K_n} = \frac{1}{K_n - 1} \sum_{i=1}^{k_n - 1} \log X_{n-i+1,n} - \log X_{n-k_n+1,n}$$

pour $\xi > 0$.

Construction de l'estimateur de Hill

Il y a plusieurs approches pour construire l'estimateur de **Hill**, on va utiliser l'approche **EMV**.

• On considère une suite de variable aléatoire (iid) de loi de $Pareto$ de paramètre $\alpha > 0$, de fonction de répartition est donnée par :$F(x) = 1 - x^{-\alpha}$; $x \leq 1$.

Donc la densité est :$f(x) = \alpha x^{-\alpha-1}$.

On veut trouver l'estimateur du maximum de vraisemblance **(EMV)** de α, pour cela on donne L la fonction de vraisemblance et les dérivées du logarithme de la fonction vraisemblance par rapport à α.

$$\begin{aligned} L(x_1, x_2, ..., x_n, \alpha) &= \prod_{i=1}^{n} f(x_i) \\ &= \alpha^n \prod_{i=1}^{n} x_i^{-\alpha-1} \end{aligned}$$

Donc :

$$\frac{\delta \log L}{\delta \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \log x_i$$

$$\frac{\delta^2 \log L}{\delta \alpha^2} = -\frac{n}{\alpha^2} < 0$$

L'estimateur du **MV** de $1/\alpha$ est alors la statistique :

$$\hat{T} = \frac{1}{n} \sum_{i=1}^{n} \log X_i = \frac{1}{n} \sum_{i=1}^{n} \log X_{i,n}$$

Une généralisation concerne $F(x) = cx^{-\alpha}\,;\, x > 0$.
Si on pose $c = u^{\alpha}$ avec $x \geq u > 0$,on obtient donc l'**EMV** de α.

$$\begin{aligned}\hat{\alpha} &= \left(\frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{X_{(i,n)}}{u}\right)\right)^{-1} \\ &= \left(\frac{1}{n}\sum_{i=1}^{n}\log X_{(i,n)} - \log u\right)^{-1}\end{aligned}$$

Souvent,on n'a pas l'expression de la fonction de répartition mais dans le domaine d'attraction maximal de **Fréchet**. On sait que F suit la loi de **Pareto** à partir d'un certain seuil connu u.
Soit $K = card\{i = 1, ..., n; X_{i,n} > u\}$ conditionnellement, à l'événement $\{k = n\}$, les estimateurs du maximum de vraisemblance des paramètres α et c associes à un échantillon de la loi définie par :

$$f(X_{(1,n)}, ..., X_{(k,n)}) = \frac{n}{(n-k)}(1 - cx_k^{\alpha})^{n-k}c^k\alpha^k\prod_{i=1}^{k}x_i^{-(\alpha+1)}$$

Théorème 2.28.

Soit $(k_n)_{n>0}$ *une suite d'entier telle que*$1 < k_n \leq n$, $k_n \longrightarrow +\infty$ *et* $\lim\limits_{n\to\infty}\frac{k_n}{n} = 0$ *alors :*
$\hat{\xi}_{K_n}^{H}$ ***converge en probabilité*** *vers* ξ.
Si de plus ; si $\lim\limits_{n\to\infty}\frac{k_n}{\log\log n} = +\infty$ *alors :*
$\hat{\xi}_{K_n}^{H}$ ***converge presque surement*** *vers* ξ.
De plus aussi ; si autres conditions de variation sont vérifier avec $\sqrt{k_n}\varepsilon\big(\frac{n}{k_n}\big)\longrightarrow 0$ *alors :*
$\sqrt{k_n}\big(\hat{\xi}_{K_n}^{H} - \xi\big)$ ***converge en loi*** *vers* $\mathcal{N}(0, \xi^2)$ *(Normalité asymptotique).*

Quelques explications sur le théorème

1. Cet estimateur est biaisé, ce biais est de lo'ordre de $\varepsilon\big(\frac{n}{k_n}\big)$.La condition $\sqrt{k_n}\varepsilon\big(\frac{n}{k_n}\big)\longrightarrow 0$ impose au biais d'être négligeable devant l'écart type de l'estimateur qui est quand à lui égale à $\sqrt{k_n}$.
   Une minimisation de l'erreur en moyenne quadratique peut être utilisée comme critère. Mais cette méthode reste néanmoins inutilisable en pratique puisque l'erreur en moyenne quadratique reste inconnue.
2. Pour établir la normalité asymptotique de l'estimateur $\hat{\xi}_{K_n}^{H}$, on a besoin d'une hypothèse sur la fonction à *variations lentes* $\boldsymbol{l}$.

Il est en effet nécessaire d'imposer une condition qui spécifie la vitesse de convergence du rapport des fonctions à variations lentes vers $\boldsymbol{l}$ (définie à la partie de *fonction à variations lentes*).

Le résultat sur la normalité asymptotique de l'estimateur de **Hill** permet de donner un intervalle de confiance pour cet estimateur.

B.Estimateur de Hill négatif

Cet estimateur est un complément à l'estimateur de **Hill** dans le cas où $\xi < -1/2$
Il est défini comme suit :

Définition 2.29.

*Soit $(k_n)_{n>0}$ une suite d'entier avec $1 < k_n \leq n$ l'estimateur de **Hill négatif** est donné par :*

$$\hat{\xi}_F^H = \frac{1}{k}\sum_{i=1}^{k-1} \log(X_{n,n} - X_{n-i,n}) - \log(X_{n,n} - X_{n-k,n})$$

*Cet estimateur est **consistant** si $\xi < -1/2$ et **asymptotiquement normal** si $-1 < \xi < -1/2$.*

C.Estimateur de Hill adapté

Définition 2.30.

Cet estimateur est applicable pour tout $\xi \in \mathbb{R}$.
Il est défini comme suit :

$$\hat{\xi}_{ad}^H = \frac{1}{k}\sum_{i=1}^{k} \log(UH_i) - \log(UH_{k+1})$$

avec :
$UH_i = X_{(i+1,n)}\left(\frac{1}{i}\sum_{j=1}^{i} \log X_{j,n} - \log X_{(i+1,n)}\right)$ *est la fonction empirique deUH au point* $x = \frac{n}{i}$.

Théorème 2.31.

Soit UH une fonction à variation régulière,on suppose que $\delta(x)$ une fonction telle que $\delta(x) \longrightarrow 0$ et $\sqrt{k}\int_0^1 \delta\left(\frac{n}{kt}\right)dt \longrightarrow 0$ lorsque $n \longrightarrow +\infty$ alors :

$$\sqrt{k}(\hat{\xi}_{ad}^H - \xi) \textbf{ converge en loi } \textit{vers} \begin{cases} \mathcal{N}\big(0,(1+\xi)\big) & \textit{si } \xi \geq 0 \\ \mathcal{N}\left(0, \dfrac{1-\xi(1+\xi+2\xi^2)}{1-2\xi}\right) & \textit{sinon} \end{cases}$$

Remarque

L'estimateur de **Hill** estime le paramètre de queue des fonctions à queue ***lourd*** c-à-d les fonctions appartenant au domaine d'attraction de **Fréchet**.

D.Estimateur de Pickands

L'estimateur de **Pickands** est construit en utilisent trois statistiques d'ordre. Cet estimateur à l'avantage d'être valable quel que soit le domaine de définition de l'indice de queue.

Définition 2.32.

*On suppose que X_i $i = \{1, ..., n\}$ est une suite de variable aléatoire (iid) de loi F appartenant à l'un des domaines d'attractions.Soit $(K_n)_{n\geq 1}$ une suite d'entiers avec $1 < k_n \leq n$. L'estimateur de **Pickands** est défini par :*

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

Théorème 2.33.

Soit $(k_n)_n$ une suite d'entier telle que $1 < k_n \leq n$,$k_n \longrightarrow +\infty$ et $\lim_{n\to\infty} \frac{k_n}{n} = 0$ alors :

$\hat{\xi}^P_{K_n}$ ***converge en probabilité*** *vers ξ.*

Si de plus ; si $\lim_{n\to\infty} \frac{k_n}{\log\log n} = +\infty$ alors :

$\hat{\xi}^P_{K_n}$ ***converge presque surement*** *vers ξ.*

Sous des conditions additionnelles sur la suite k_n et la fonction de répartition F on aura :

$\sqrt{k_n}\big(\hat{\xi}^P_{K_n} - \xi\big)$ ***converge en loi*** *vers* $\mathcal{N}\left(0, \frac{\xi^2(2^{2\xi+1}+1)}{4(\log 2)^2(2^\xi - 1)^2}\right)$

Remarque 5. *L'estimateur de **Pickands** est biaisé et la normalité asymptotique permet de donner un intervalle de confiance pour l'estimateur.*

E.Estimateur de Zipf

Dans le but d'améliorer le biais asymptotique des des estimateurs,***Kratz, Resnick, Steinbach et Schultze*** ont indépendamment proposé d'estimer l'indice de queue par la méthode des moindres carrés classique. Leur estimateur connu sous le nom de **Zipf**.

Définition 2.34.

L'estimateur de ***Zipf*** *est défini par :*

$$\hat{\xi}^{Z}_{K_n} = \frac{\frac{1}{k_n}\sum_{j=1}^{k_n}\log\frac{k_n+1}{j}\log X_{n-j+1,n} - \frac{1}{k_n}\sum_{j=1}^{k_n}\log\frac{k_n+1}{j}\left(\frac{1}{k_n}\sum_{j=1}^{k_n}\log X_{n-i+1,n}\right)}{\frac{1}{k_n}\sum_{j=1}^{k_n}\left(\log\frac{k_n+1}{j}\right)^2 - \left(\frac{1}{k_n}\sum_{j=1}^{k_n}\log\frac{k_n+1}{j}\right)^2}$$

Cet estimateur est asymptotiquement gaussien.

2.5 Estimation de quantiles extrêmes et de période de retour

Le choix de la loi pour les valeurs extrêmes et l'estimation de la fonction de répartition en déterminant l'indice de queue de distribution ne sont souvent qu'un objectif intermédiaire, l'objectif principal étant plutôt l'estimation d'un quantile extrême ou d'un niveau de retour ainsi que l'estimation d'une période de retour.

2.5.1 Estimation des quantiles extrêmes

Nous estimerons le quantile d'ordre $p_t = 1 - \frac{1}{t}$ de F c-à-d le nombre x_{p_t} définie par $1 - F(x_{p_t}) = \frac{1}{t}$. Si t est grand, un tel quantile dit *quantile extrême.*

Pour l'estimation de ce quantile on va utiliser la méthode d'estimation des quantiles par la méthode des blocs car cette méthode est basée sur le théorème de ***Fisher-Tippet***, et en suppose que l'échantillon de maximum suit exactement une loi **GEV**.

Construction des blocs :

On divise l'échantillon $X_1, X_2, ..., X_n$ en k blocs de même taille n . Soit $(N = nk)$ le j^{me} bloc est défini par :

$$M_{n,j} = MaxX_{(j-1)n+1}, ..., X_{(j-1)n+n}$$

avec : $i = \{1...k\}$ et $j = \{1...k\}$

Proposition 2.35.

L'estimateur de quantile extrême obtenu par la méthode des blocs s'écrit sous la forme :

$$\hat{x}_{pt} = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}[1 - (-\log(1-p_t))^{-\hat{\xi}}] & \xi \neq 0 \\ \hat{\mu} - \hat{\sigma}\log(\log(1-p_t)) & \xi = 0 \end{cases} \tag{2.11}$$

où $\hat{\mu}$,$\hat{\sigma}$ et $\hat{\xi}$ sont estimateurs des paramètres de la loi ***VE****.*

En utilisant l'estimateur de Hill

On considère les fonctions de répartition appartenant au domaine d'attraction maximal de Fréchet ($\xi > 0$), alors la fonction de survie peut se mettre sous la forme $\bar{F} = x^{-1/\xi}L(x)$ où L est une fonction à variation lente.
Donc on peut écrire :

$$\frac{\bar{F}(x)}{\bar{F}(X_{n-k})} = \frac{L(x)}{L(X_{n-k})}\left(\frac{x}{X_{n-k}}\right)^{-1/\xi}$$

et on considèrent que le rapport des fonctions à variation lente est proche de 1 on trouve :

$$\bar{F}(x) \approx \bar{F}\big(X_{n-k}\big)\left(\frac{x}{X_{n-k}}\right)^{-1/\xi}$$

On déduit de cette expression un estimateur $F(x)$ pour $x > X_{n-k}$

$$\bar{F}(x) = 1 - \frac{k}{n}\bar{F}\big(X_{n-k}\big)\left(\frac{x}{X_{n-k}}\right)^{-1/\xi^H_{n,k}}$$

d'où

$$\hat{x}^H_p = X_{n-k}\left(\frac{n}{k}\big(1-p\big)\right)^{-\xi^H_{n,k}} \tag{2.12}$$

2.5.2 Estimation de période de retour

Nous définissons le niveau de retour comme la valeur x_t telle que nous espérons détecter en moyenne un seul dépassement de cette quantité au bout de **t** périodes c'est à dire :

$$\begin{aligned}
&\mathbb{E}\left(\textstyle\sum_{i=1}^{t} \mathbf{1}_{x_i > x_y}\right) = 1 \\
\iff &\Pr\big(X_i > x_t\big) = \frac{1}{t} \\
\iff &1 - F(x_t) = \frac{1}{t}
\end{aligned}$$

L'estimateur d'un niveau de retour d'ordre **t** revient à l'estimation d'un quantile extrême d'ordre $p_t = 1 - \dfrac{1}{t}$.

Proposition 2.36.

*Un estimateur d'ordre $\hat{p}^{GEV}_T$ du quantile x_pt pour la loi des **GEV** est donné par :*

$$\hat{p}^{GEV}_T = \begin{cases} \exp\left(-\left(1+\hat{\xi}\big(\frac{x_{pt}-\hat{\mu}}{\hat{\sigma}}\big)\right)^{-1/\xi}\right) & si\ \hat{\xi} \neq 0 \\ \exp\left(-\exp\big(-\frac{x_{pt}-\hat{\mu}}{\hat{\sigma}}\big)\right) & si\ \hat{\xi} = 0 \end{cases}$$

d'où :

$$\hat{T}_{GEV} = \begin{cases} \left(1 - \exp(-(1+\hat{\xi}(\frac{x_{pt}-\hat{\mu}}{\hat{\sigma}}))^{-1/\xi})\right)^{-1} & si\ \hat{\xi} \neq 0 \\ \left(1 - \exp\big(-\exp\big(-\frac{x_{pt}-\hat{\mu}}{\hat{\sigma}}\big)\big)\right)^{-1} & si\ \hat{\xi} = 0 \end{cases}$$

alors :

$$\hat{T}_{GEV} = \frac{1}{G_{\hat{\xi}}(x_{pt})} \quad \forall \hat{\xi} \in \mathbb{R}$$

2.5.3 Conclusion

Dans ce chapitre, on donne les fondements de la théorie probabiliste des valeurs extrêmes. Au début, on a présenté le théorème fondamental de la théorie des valeurs extrêmes (théorème de Fisher-Tippet) et les différents domaines d'attraction ainsi que le coefficient de normalisation avec des résultats obtenus dans ce cadre. Après on a passé à l'estimation des paramètres des valeurs extrêmes avec des méthodes paramétriques et des méthodes semi paramétriques. Enfin on a estimé les quantiles extrêmes ce qui conduit à l'estimation de la période de retour et le niveau de retour.

Chapitre 3

Distribution de Pareto généralisée

Motivation : L'approche basé sur la GEV a été critiqué dans la mesure où l'utilisation d'un seul maxima conduit à une perte d'information contenue dans les autres grandes valeurs de l'échantillon.
Pour pallier ce problème,la méthode POT (Peaks-Over-Threshold) où méthode des excès au-delà d'un seuil élevé a été introduit par Pickands(1975),basé sur la distribution de pareto généralisées GPD.

3.1 Introduction

La méthode des excès au-delà d'un seuil repose sur le comportement des données qui dépassent un seuil donnée.Autrement dit,elle consiste à étudier le comportement non pas du maximum des données qu'on a en main,mais toutes les données qui dépassent un seuil élevé u.Plus précisément,les différences entre ces données et le seuil u appelées *excès*.

Définition 3.1.

On appelle ***excès*** *de la variable aléatoire X au-delà d'un seuil $u < x_F$ la variable aléatoire Y qui prend ses valeurs sur $]0, x_F - u[$ définie par :*

$$Y = X - u | X > u \tag{3.1}$$

avec $u < x_F$.

Définition 3.2.

On appelle ***distribution des excès*** *de la variable aléatoire X par rapport à un seuil $u < x_F$ la loi de probabilité de la variable alétoire Y excès de X au-delà du seuil $u < x_F$ donnée par sa fonction de distribution de répartition F_u,qu'on appelle fonction de*

distribution des excès suivante :

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

$$F_u(y) = \begin{cases} 0 & y \leq 0 \\ 1 - \dfrac{1 - F(u+y)}{1 - F(u)} & 0 < y < x_F - u \\ 1 & y \geq x_F - u \end{cases} \tag{3.2}$$

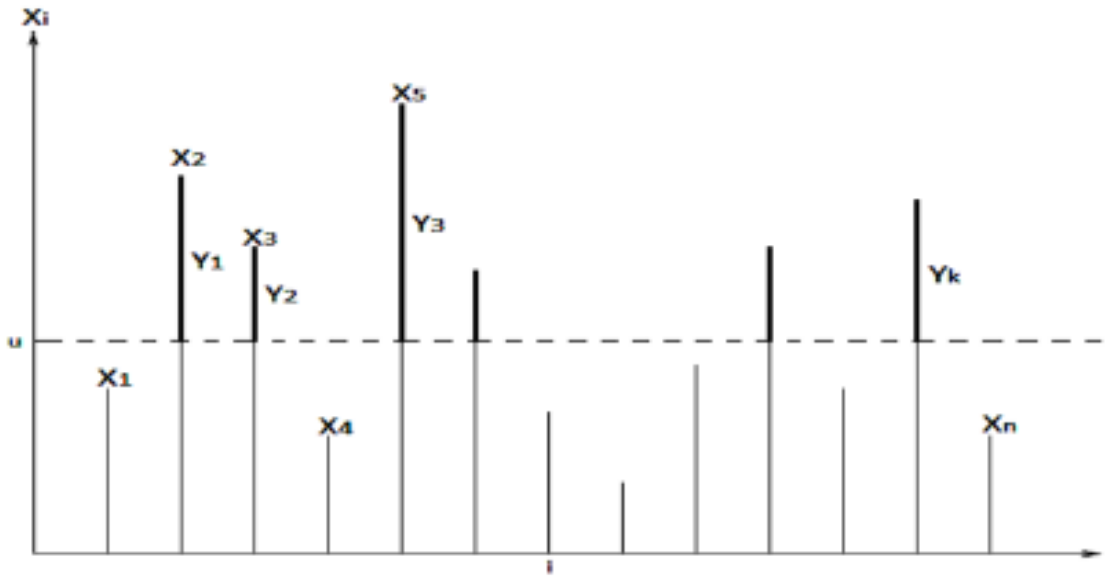

FIGURE 3.1 – excès

3.1.1 Distribution de Pareto généralisé(GPD)

La distribution de **Pareto généralisé** joue un rôle très important à la modélisation des excès.

Définition 3.3.

*Soit $\xi \in \mathbb{R}$,On appelle **distribution de Pareto généralisé standard** toute fonction de répartition H_ξ où toute loi de probabilité qui a H_ξ comme fonction de répartition telle que : $\forall x \in \mathbb{R}$, $1 + \xi x > 0$*

$$H_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi} & \xi \neq 0 \\ 1 - e^{-x} & \xi = 0 \end{cases} \tag{3.3}$$

Définition 3.4.

Une distribution $H_{\xi,\beta}$ est dite de ***Pareto généralisée*** *de paramètre $\xi \in \mathbb{R}$ et $\beta > 0$ si :*

$$H_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \xi = 0 \end{cases} \tag{3.4}$$

Cette distribution est définie pour :

$$\begin{cases} x \geq 0 & \xi \geq 0 \\ 0 < x \leq -\beta/\xi & \xi < 0 \end{cases}$$

β représente le paramètre d'échelle et ξ le paramètre de forme.

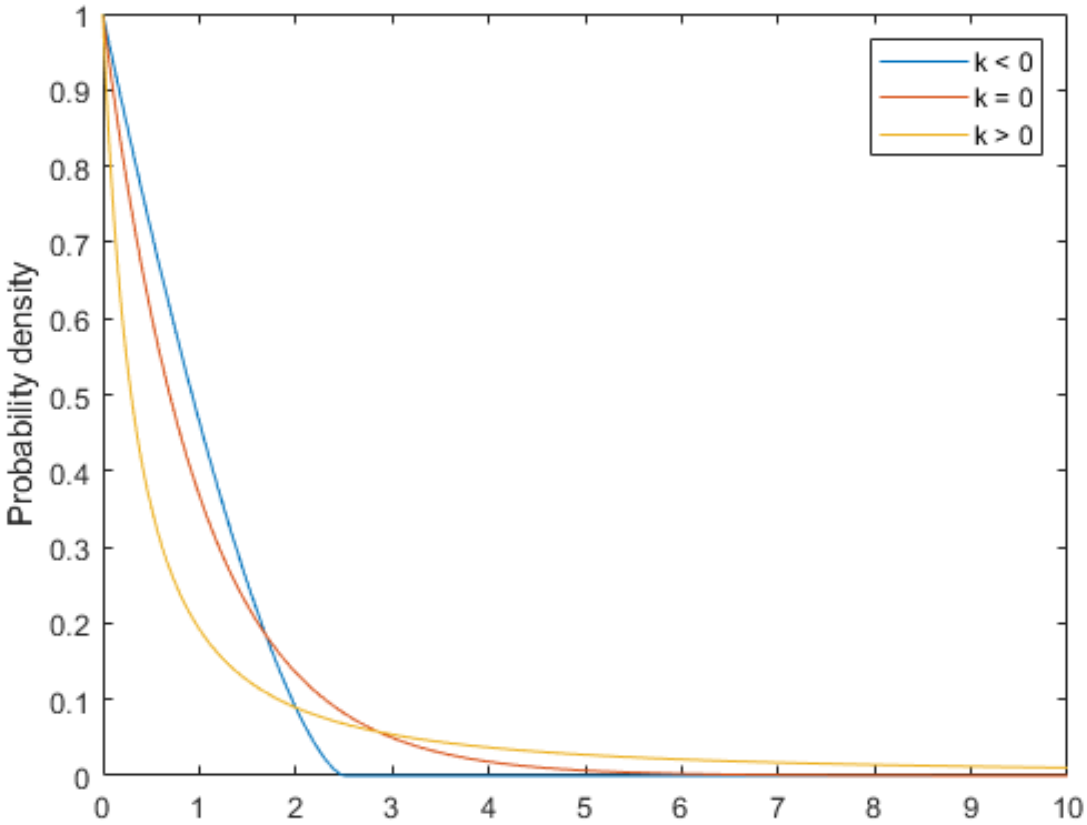

FIGURE 3.2 – Densité de Pareto

3.1.2 Particularités de la GPD

1. Si $\xi > 0$,la distribution $H_{\xi,\beta}$ est la loi de Pareto usuelle avec $\alpha = \frac{1}{\xi}$ et $K = \frac{\beta}{\xi}$.
   En effet :

$$\begin{aligned} H_{\xi,\beta}(x) &= 1 - \left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi} \\ &= 1 - \left(\frac{1}{1 + \frac{\xi}{\beta}x}\right)^{1/\xi} \\ &= 1 - \left(\frac{\beta}{\beta + \xi x}\right)^{1/\xi} \\ &= 1 - \left(\frac{\frac{\beta}{\xi}}{\frac{\beta}{\xi} + y}\right)^{1/\xi} \end{aligned}$$

2. Si $\xi = 0$,la distribution $H_{0,\beta}(x)$ est une distribution exponentielle d'espérance β.

$$\begin{aligned} \lim_{\xi\to 0} H_{\xi,\beta}(x) &= \lim_{\xi\to 0} 1 - \left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi} \\ &= 1 - \exp\left(- \lim_{\xi\to 0}\left(\frac{\frac{x}{\beta}}{1 + \frac{\xi x}{\beta}}\right)\right) \\ &= 1 - \lim_{\xi\to 0}\exp\left(- \frac{1}{\xi}\right)\log\left(1 + \frac{\xi}{\beta}x\right) \\ &= 1 - \exp\left(- \lim_{\xi\to 0}\left(\frac{\log\left(1 + \frac{\xi}{\beta}x\right)}{\xi}\right)\right) \\ &= 1 - \exp\left(\frac{-x}{\beta}\right) \\ &= H_{0,\beta}(x) \end{aligned}$$

3. $H_{\xi,\beta}(x) \in MDA(G_\xi)\,;\, \forall \xi \in \mathbb{R}$.
   En effet :
   Soit $\xi > 0$, on sait que si $F \in MDA(G_\xi)$ alors : $\bar{F}(x) = x^{-1/\xi}L(x)$ où $L(x)$ est une fonction à variation régulière (*théorème de Karamata*).

On montre que : $\bar{H}_{\xi,\beta}(x) = x^{-1/\xi}L(x)$.

$$\begin{aligned}
\bar{H}_{\xi,\beta}(x) &= 1 - H_{\xi,\beta}(x) \\
&= 1 - \left(1 - \left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi}\right) \\
&= \left(1 + \frac{\xi}{\beta}x\right)^{-1/\xi} \\
&= x^{-1/\xi}\left(\frac{x^{-1}\beta + \xi}{\beta}\right)^{-1/\xi} \\
&= x^{-1/\xi}\left(\frac{\xi}{\beta} + \frac{1}{x}\right) \\
&= x^{-1/\xi}L(x)
\end{aligned}$$

où : $L(x) = \dfrac{\xi}{\beta} + \dfrac{1}{x}$

On montre que $L(x)$ est à variation régulière.

L est à variation régulière $\iff \lim\limits_{x\to\infty} \dfrac{L(tx)}{L(x)} = 1$.

$$\begin{aligned}
\lim_{x\to\infty} \frac{L(tx)}{L(x)} &= \lim_{x\to\infty} \frac{\left(\dfrac{(tx)^{-1}\beta + \xi}{\beta}\right)^{-1/\xi}}{\left(\dfrac{x^{-1}\beta + \xi}{\beta}\right)^{-1/\xi}} \\
&= \lim_{x\to\infty} \left(\frac{\dfrac{(tx)^{-1}\beta + \xi}{\beta}}{\dfrac{x^{-1}\beta + \xi}{\beta}}\right)^{-1/\xi} \\
&= \lim_{x\to\infty} \left(\frac{(tx)^{-1}\beta + \xi}{x^{-1}\beta + \xi}\right)^{-1/\xi} \\
&= 1
\end{aligned}$$

Donc $L(x)$ est à variation régulière ; alors $H_{\xi,\beta}(x) \in MDA(G_\xi)$.

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

5. Si $\xi > 0$, on retrouve la loi de Pareto décentrée.

Question : Quel est le lien entre la loi de Pareto généralisée et la distribution des excès ?

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

3.2 Théorème de Pickands-Balkema-De Haan

Théorème 3.5.

La fonction de distribution F de la variable aléatoire X appartient au max-domaine d'attraction de la distribution des valeurs extrêmes généralisée standard G_ξ ($\xi \in \mathbb{R}$) si et seulement s'il existe une fonction strictement positive β telle que la fonction de distribution des excès F_u de X par rapport au seuil $u < x_F$ converge uniformément vers une distribution de Pareto généralisée $H_{\xi,\beta(u)}$ lorsque u tend vers x_F.

$$F \in MDA(G_\xi) \Longleftrightarrow \lim_{u \to x_F} \sup_{0<y<x_F-u} |F_u(y) - H_{\xi,\beta(u)}| = 0 \tag{3.5}$$

Démonstration.

Pour une démonstration de ce théorème, on pourra se référer au [6]. ■

Remarque 6.

⋆ *Le théorème de* ***Pickands-Balkema-De Haan*** *est considéré aussi comme le deuxième théorème fondamental de la théorie des valeurs extrêmes,ce qui a donné une importance à la distribution de Pareto généralisée dans cette théorie.*

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

$$\begin{aligned} 1 - F_u(y) &= \Pr(Y > y) \\ &= \Pr(X - u|X > u) \\ &= \frac{\Pr(X > u + y, X > u)}{\Pr(X > u)} \\ &= \frac{1 - F(u + y)}{1 - F(u)} \end{aligned} \tag{3.6}$$

Lorsque le seuil u est grand,on peut approcher cette quantité par la fonction de survie d'une loi **GPD**.Afin d'approcher le quantile,il suffit d'utiliser le résultat de **Pickands-Balkema-De Haan** qui établit l'équivalence entre la convergence en loi du maximum vers une loi des valeurs extrêmes G_ξ et la convergence en loi d'un excès vers une **GPD**.

Exemple 3.6.

1. *La loi exponentielle du paramètre 1 :* $F(x) = 1 - e^{-x}$
   on prend $\beta_u = 1$

$$\begin{aligned} F_u(y) = \Pr(X - u \leq x | X > u) &= \frac{F(u+y) - F(u)}{1 - F(u)} \\ &= \frac{e^{-u} - e^{-u-y}}{e^{-u}} \\ &= 1 - e^{-y} \end{aligned}$$

   Pour tout $y > 0$,*la loi limite est la loi* ***GPD*** *de paramètre* $\xi = 0$ *et* $\beta_u = 1$.
   Donc dans ce cas, la loi ***GPD*** *n'est pas simplement la loi limite, mais il s'agit de la loi exacte pour tout* u.

2. *La loi de Pareto :*$F(x) = 1 - cx^{-\alpha}$, $(c > 0, \alpha > 0)$
   On pose :$\beta_u = ub$, $b > 0$
   $\forall y > 0$

$$\begin{aligned} F_u(y) &= \frac{F(u + uby) - F(u)}{1 - F(u)} \\ &= \frac{cu^{-\alpha} - c(u + uby)}{cu^{-\alpha}} \\ &= 1 - (1 + by)^{-\alpha} \end{aligned}$$

   Si on pose $\xi = \frac{1}{\alpha}$ *et* $b = \xi$, *la limite est alors la loi* ***GPD*** *de paramètre* ξ.

Petite synthèse :(Relation entre **GEV** et **GPD**)

$\star$ Si pour une variable aléatoire X de fonction de répartition F inconnue,l'échantillon des maximums normalisés *converge en distribution* vers une loi de probabilité non dégénérer,alors il est équivalent de dire que F est dans le max-domaine d'attraction d'une distribution **GEV** d'indice de queue $\xi \in \mathbb{R}$ ***(Théorème de Fisher-Tippet)***.
Dans ce cas, il s'en déduit que la distribution des excès de X au-delà d'un seuil u *converge uniformément* vers une distribution **GPD**,de même indice de queue que celui de la distribution **GEV**,lorsque ce seuil tend vers le point terminal x_F de la fonction de répartition F ***(Théorème Pickands-Balkema De Haan)***.

3.3 Estimation des paramètres de la GPD

L'estimation des paramètres d'une distribution **GPD** pose le problème de la détermination du seuil u,car il doit être suffisamment grand pour que l'on puisse appliquer le théorème précédent,mais ne doit pas être trop grand afin d'avoir suffisamment de données pour obtenir des estimateurs de bonne qualité.
Donc,tout d'abord avant d'estimer les paramètre de cette distribution on doit choisir le bon seuil.

3.3.1 Choix du seuil

Le choix du seuil reste toujours délicat mais il y'a une méthode graphique nous aide à déterminer le bon seuil u,dans cette méthode en utilise une fonction qui est ***la fonction moyenne des excès***.

Définition 3.7.

On appelle ***fonction moyenne des excès*** *de la variable alétoire X par rapport au seuil $u < x_F$, et on l'a note $e(u)$,la fonction espérance de la variable aléatoire Y excès de X au-delà du seuil $u < x_F$ définie par :*
$\forall u < x_F$

$$e(u) = \mathbb{E}(X - u|X > u) = \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(t)dt \tag{3.7}$$

Définition 3.8.

Soient $(X_1, X_2, ..., X_n)$ l'échantillon de taille $n \in \mathbb{N}^*$ de la variable aléatoire X et F_n sa fonction de répartition empirique.
On appelle ***fonction moyenne des excès empirique*** *de la variable alétoire X par rapport au seuil $u < x_F$ la fonction $e_n(u)$ définie par :*
$\forall u < x_F$

$$\begin{aligned} e_n(u) &= \frac{1}{\bar{F}_n(u)} \int_u^{x_F} \bar{F}_n(t)dt \\ &= \frac{1}{card\{\Delta_n(u)\}} \sum_{i \in \Delta(u)} \big(X_i - u\big) \end{aligned} \tag{3.8}$$

avec : $\Delta_n(u)$=$\{i = 1, ..., n\}$ tel que $X_i > u$ et $\frac{0}{0} = 0$ (conventionnellement).

Proposition 3.9.

Si W est une variable aléatoire qui a comme fonction de répartition une distribution de Pareto généralisée $H_{\xi,\beta}$ avec $(\xi < 1, \beta > 0)$,alors sa fonction moyenne des excès $e(u)$

au-delà d'un seuil $u < w_0$ (w_0 est le point terminal de $H_{\xi,\beta}$) est donnée par :
$\forall u < w_0$

$$e(u) = \mathbb{E}(W - u|W > u) = \frac{\beta + \xi u}{1 - \xi} \tag{3.9}$$

avec : $\beta + \xi u > 0$.

• On peut résumé l'idées principale pour choisir le bon seuil on utilisant la méthode ***graphique*** comme suit :

1. **Cadre théorique :**

   D'après le théorème de **Pickands-Belkema-De Haan**, si à partir d'un certain seuil $u_0 < x_F$ l'excès de la variable aléatoire X au-delà du seuil u_0 suit une loi de Pareto généralisée, c-à-d : $\forall u_0 < x_F$

   $$\begin{aligned} e(u_0) &= \mathbb{E}(X - u_0|X > u_0) \\ &= \frac{\beta(u_0)}{1 - \xi} \end{aligned}$$

   • Si cette approximation est vraie pour le seuil u_0,elle sera vraie pour un autre seuil u tel que : $\forall u_0 < u < x_F$

   $$\begin{aligned} e(u) &= \mathbb{E}(X - u_0|X > u_0) \\ &= \frac{\beta(u_0) + \xi u}{1 - \xi} \end{aligned} \tag{3.10}$$

   avec : $\beta(u_0) + \xi u > 0$

   • Et pour déterminer le seuil u,on exploite la linéarité en u de la fonction moyenne des excès $e(u)$.

2. **Cadre empirique :**

   Si on suppose que nos données $x_1, ..., x_n$ ($n \in \mathbb{N}$) sont une réalisation de l'échantillon $(X_1, ..., X_n)$ de la variable aléatoire X,alors on procède de la manière suivante pour déterminer le seuil $u < x_F$:

   On calcule l'estimateur $\hat{e}_n(u)$ de la fonction moyenne des excès $e(u)$ de X au-delà du seuil u,en utilisant sa version empirique $e_n(u)$ par :

   $$\hat{e}_n(u) = \frac{1}{card\{i; x_i > u\}} \sum_{i;x_i>u} \big(x_i - u\big) \tag{3.11}$$

   avec :

   $$\min_{1\leq i\leq n} x_i \leq u < \max_{1\leq i\leq n} x_i$$

• On trace le graphe

$$\kappa_u = [u, \hat{e}_n(u)]; \min_{1\leq i\leq n} x_i \leq u < \max_{1\leq i\leq n} x_i \tag{3.12}$$

• Une fois le graphe est tracé,on exploite la linèarité en u de la fonction moyenne des excès d'une distribution de Pareto-généralisée $\kappa_{\xi,\beta}$ au-delà du seuil choisissant $u \leq x$ où x est le point à partir duquel le graphe $\kappa_{\xi,\beta}$ est approximativement une droite.

3.3.2 Estimation des paramètres de la GPD

Une fois le seuil optimal choisi,on construire une nouvelle série d'observations au dessus de ce seuil et la distribution de ces données suit approximativement une distribution **GPD**. Dans cette partie,on va estimer les paramètres de la **GPD** on utilisons trois méthodes différentes :

1. **Méthode du maximum de vraisemblance**.
2. **Méthode des moments pondérés**.
3. **Méthode basé sur les graphes de la moyenne des excès(Mean exces plot(MEP))**.

A. Méthode du maximum de vraisemblance

La densité de la distribution **GPD** est :

$$g_{\xi,\beta}(x) = \begin{cases} \beta^{-1/\xi}\big(\beta + \xi x\big)^{-\frac{1}{\xi}-1} & \text{si } \xi \neq 0 \\ \beta^{-1}\exp\big(-\dfrac{x}{\beta}\big) & \text{si } \xi = 0 \end{cases}$$

La fonction de vraisemblance est donnée par :

$$l(\xi, \beta, x_1, ..., x_n) = \prod_{i=1}^{n} g_{\xi,\beta}(x_i)$$

Ce qui implique

$$\begin{aligned} \log l(\xi, \beta, x_1, ..., x_n) &= \sum_{i=1}^{n} g_{\xi,\beta}(x_i) \\ &= -n\log\beta - \left(1 - \frac{1}{\xi}\right)\sum_{i=1}^{n}\log\left(1 + \frac{\xi}{\beta}x_i\right) \end{aligned}$$

On pose que $\tau = \dfrac{\xi}{\beta}$, l'annulation des dérivées partielles des logarithmes de la fonction de vraisemblance conduit au système :

$$\begin{cases} \hat{\xi} & = \dfrac{1}{n}\sum_{i=1}^{n}\log\big(1+\tau X_i\big) = \hat{\tau} \\ \dfrac{1}{\tau} & = \dfrac{1}{n}\big(1+\dfrac{1}{\xi}\big)\sum_{i=1}^{n}\dfrac{X_i}{1+\tau X_i} \end{cases}$$

L'estimateur de **MV** (ξ, τ) est $(\hat{\xi} = \hat{\xi}(\hat{\tau}), \tau)$ où τ est la solution de :

$$\frac{1}{\tau} = \frac{1}{n}\big(1+\frac{1}{\xi}\big)\sum_{i=1}^{n}\frac{X_i}{1+\tau X_i}$$

Cette équation se résous numériquement de manière itérative.

$\star$ Cet estimateur est *asymptotiquement normale*

$$\sqrt{n}\left(\hat{\xi}-\xi; \frac{\hat{\beta}}{\beta}-1\right) \longrightarrow \mathcal{N}(0, M^{-1})$$

où

$$M^{-1} = (1+\xi)\begin{pmatrix} 1+\xi & -1 \\ -1 & 2 \end{pmatrix}$$

Ce résultat a été démontrer par ***Hosking*** et ***Walis***.

B.Méthode des moments pondérés

Rappelons que le moment pondéré d'ordre r est définie par :

$$\mu_r = \mathbb{E}\big(Z(\bar{G}_{\xi,\beta}(Z))^r\big)$$

avec $r \in \mathbb{N}$.

On prend la variable aléatoire Z suit la loi de Pareto généralisée de paramètre (ξ, β). Le moment pondéré μ_r est égal à :

$$\mu_r = \frac{\beta}{(r+1)(r+1-\xi)} \tag{3.13}$$

avec $r = 0; 1$

Pour $r = 0\,;\, r = 1$

$$\hat{\beta} = \frac{2\mu_0\mu_1}{\mu_0 - 2\mu_1}$$

et

$$\hat{\xi} = 2 - \frac{\mu_0}{\mu_0 - 2\mu_1}$$

On remplace μ_0 et μ_1 par les estimateurs des moments empiriques pondérés :

$$\begin{aligned} M_r &= \frac{1}{n}\sum_{j=1}^{n}\left(\prod_{l=1}^{r}\frac{n-j-l+1}{n-l}\right)X_{(j,n)} \\ &= \frac{1}{n}\sum_{j=1}^{n}\left(1-\frac{j}{n+1}\right)^{r}X_{(j,n)} \end{aligned}$$

où $X_{(j,n)}$ est la j^{me} statistique d'ordre.
On trouve alors un estimateur de ξ par cette méthode qui est donné par :

$$\hat{\xi} = \frac{\frac{1}{n}\sum_{j=1}^{n}\left(4\frac{j}{n+1}-3\right)X_{(j,n)}}{\frac{1}{n}\sum_{j=1}^{n}\left(2\frac{j}{n+1}-1\right)X_{(j,n)}} \tag{3.14}$$

C.Méthode basé sur les graphes de la moyenne des excès(MEP)

Soit X une variable aléatoire de fonction de répartition F et d'espérance finie,on rappelle que la fonction de moyenne des excès est définie par :
$\forall u > 0$

$$e_X(u) = \mathbb{E}(X-u|X>u) = \frac{1}{\bar{F}(u)}\int_u^{x_F}\bar{F}(y)dy$$

Et le **MEP** correspond au graphe associé aux point $(u; e_X(u))$.
Si $F \in MAD(G_\xi)$ alors on aura :

$$e_{\log X}(\log u) = \mathbb{E}\big(\log X - \log u|X>u\big) \longrightarrow \xi$$

En remplace u par $X_{(k+1,n)}$

$$e_{\log X}(\log X_{(k+1,n)}) = \mathbb{E}\big(\log X - \log X_{(k+1,n)}|X > X_{(k+1,n)}\big)$$

$e_{\log X}(\log X_{(k+1,n)})$ est un estimateur consistant,donc on déduit que :

$$\hat{\xi} = e_{\log X}(\log X_{(k+1,n)}) \tag{3.15}$$

3.3.3 Estimation des quantiles extrêmes et période de retour

A.Estimation des quantiles extrêmes

Dans cette approche,on va utiliser une méthode nommée **Méthode POT (Peak-Over-Threshold)** pour estimer les quantiles extrêmes.
Cette méthode est basée sur l'approximation de la distribution des excès pour la loi de Pareto généralisée. De plus elle présente un avantage par rapport à la méthode des blocs, en ce sens qu'il est plus facile d'avoir un échantillon d'excès que de max.Dans la pratique, on remplace u par $X_{(n-k+1,n)}$ qui représente la K plus grande observation de l'échantillon.
Pour réaliser cette méthode,on va suivre les étapes suivantes :

1. Soit $X_1, ..., X_n$ un échantillon,à partir d'un certain seuil u, on note N_u le nombre d'observation qui dépassent ce seuil.
2. Soit $Y_1, ..., Y_{N_u}$ un échantillon des excès au dessus du seuil u de distribution conditionnelle
$$F_u(x) = \Pr(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}$$
D'aprés le théorème de **Belkema-Pickands De Haan** une approximation pour cette distribution on peut approximer cette distribution par :
$x > u$ et $u \longrightarrow \infty$
$$F_u(x) \approx H_{\xi,\beta}(x) \tag{3.16}$$
Et de plus :
$$\bar{F}_u(x) \approx \bar{H}_{\xi,\beta}(x) \tag{3.17}$$
3. La décomposition de F peut s'écrire de la manière suivante :
Pour $x > u$
$$F(x) = \Pr(X > u) F_u(x - u) + \Pr(X \leq u)$$
Cette distribution est estimée par :
$$\hat{F}(x) = (1 - F_n(u)) H_{\hat{\xi},\hat{\beta}}(x) + F_n(u) \tag{3.18}$$
Où $F_n(u)$ est la fonction de répartition empirique au point u, tel que
$$F_n(u) = 1 - \frac{N_u}{n}$$
Donc $\hat{F}$ est la loi **GPD** de paramètre $\xi = \hat{\xi}$, $\beta = \hat{\beta}\Big(1 - F_n(u)\Big)^{\hat{\xi}}$
et $\mu = u - \frac{\beta}{\xi}\Big(\big(1 - F_n(u)\big)^{-\hat{\xi}} - 1\Big)$.

L'estimateur pour la queue $\bar{F}(u+x)$ prend la forme :

$$\hat{\bar{F}}(u+x) = \frac{N_u}{n}\left(1+\hat{\xi}\frac{x}{\hat{\beta}}\right)^{-1/\hat{\xi}}$$

et

$$\hat{F}(u+x) = 1 - \frac{N_u}{n}\left(1+\hat{\xi}\frac{x}{\hat{\beta}}\right)^{-1/\hat{\xi}}$$

d'où

$$\hat{F}(x) = \frac{N_u}{n}\left(1+\hat{\xi}\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}$$

4. On inverse la fonction $\hat{F}$ définie dans (3.17), on trouve l'estimateur du quantile extrême qui est donnée par :

$$\begin{aligned}\hat{x}_p &= H^{-1}_{\hat{\xi},\hat{\beta}}\left(\frac{1-p-F_n(u)}{1-F_n(u)}\right)\\ &= u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\xi}} - 1\right)\end{aligned}$$

Donc :

$$\hat{x}_p^{POT} = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1-p)\right)^{-\hat{\xi}} - 1\right) \tag{3.19}$$

B.Estimation de période de retour

On sait que l'estimation de la période de retour T d'un quantile x_{POT} revient à estimer l'ordre de ce quantile $\hat{p}_T = \frac{1}{\hat{T}}$.
Selon (3.19) nous estimons la période de retour pour la loi des valeurs extrêmes GPD par :
$\forall \hat{\xi} \in \mathbb{R}$ et $\beta > 0$

$$\hat{T}_p^{POT} = \frac{n}{N_u}\frac{1}{1-H_{\hat{\xi},\hat{\beta}}(x_p)} \tag{3.20}$$

3.4 Approximation du niveau et du temps de retour

La théorie des valeurs extrêmes suppose la continuité de la loi de distribution pour estimer le niveau et la période de retour. Elle ne semble donc pas judicieuse dans le cas de données discrètes.
Comme nous nous intéressons à des données discrètes, nous allons proposer une estimation de bornes du niveau de retour et une estimation d'un temps de retour au lieu de les calculer exactement.
Soit un n échantillon $(X_1, X_2, ..., X_n)$ de n variable aléatoire discrètes positives(iid) de

variable aléatoire générique X de fonction de répartition F.

Problématique : Quelle valeur de niveau de retour $\hat{x}_t$ choisir pour qu'au bout de t-unités de temps,nous espérions un dépassement de $\hat{x}_t$ i.e $\hat{x}_t$? tel que :

$$\mathbb{E}\Big(\sum_{i=1}^{t}\mathbf{1}_{(X_i>\hat{x}_t)}\Big)=1\Longleftrightarrow \Pr\big(X_i>\hat{x}_t\big)=\frac{1}{t} \tag{3.21}$$

$$\Longleftrightarrow 1-F(\hat{x}_t)=\frac{1}{t} \tag{3.22}$$

$\forall\ i=1,...,t$

A défaut de calculer explicitement $\hat{x}_t$ nous proposons d'après **Guillou-al(2007)** une borne supérieur b_t du niveau de retour tel que pour tout t donné :

$\forall\ i=1,...,t$

$$\Pr\big(X_i>b_t\big)=\frac{1}{t}\leq \Pr\big(X_i>l_r\big)$$

3.4.1 Borne du niveau de retour

A.Borne supérieure du niveau de retour

La méthode développée par **Guillou-al(2007)** pour obtenir la borne supérieure repose sur *l'inégalité de Markov*

$$1-F(x)\leq\frac{\mathbb{E}(h(x))}{h(x)} \tag{3.23}$$

Pour $x>0$ et avec le choix de fonction h :$h(x)=u(x)v(F(x))$.

telle que u et v sont deux fonctions positives et croissantes définies respectivement sur $[0,+\infty]$ et $[0,1]$.

Pour un quantile $\hat{x}_t$ d'ordre $p_t=1-\frac{1}{t}$, la fonction de survie $\bar{F}(\hat{x}_t)=\frac{1}{t}$. En remplaçant x par $\hat{x}_t$ et $F(\hat{x}_t)$ par p_t dans *l'inégalité de Markov*, on obtient :

$u(\hat{x}_t)\leq\dfrac{t\theta(u,v)}{v(p_t)}$

avec $\theta(u,v)=\mathbb{E}\big(u(X)v(F(X))\big)$

d'où :

$$\hat{x}_t\leq u^{\longleftarrow}\left(\frac{t\theta(u,v)}{v(p_t)}\right)=b_t(u,v)$$

$u^{\longleftarrow}$ désignent la fonction inverse de u.

Ainsi :

$\hat{x}_t\leq \inf b_t(u,v)$; u et v sont deux fonctions croissantes positives.

Et pour un échantillon ordonné $(X_{1,n},...,X_{n,n})$ nous considérons l'estimateur naturel de

$\theta(u,v)$:

$$\hat{\theta}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} u(X_{i,n})v\left(\frac{i}{n}\right)$$

Proposition 3.10.

Soit X la variable aléatoire positive de loi F.

Si $v(.)$ est une fonction Lipshitzienne d'ordre 1 sur $[0,1]$ et $u(.)$ une fonction telle que les intégrales $\int |u(x)|dF(x)$ et $\int v^2(F(x))u^2(x)dF(x)$ sont finies alors :

1. *La variable aléatoire $\sqrt{n}\big(\hat{\theta}_n(u,v) - \theta(u,v)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma^2)$.*
   $\sigma^2 = \mathbb{E}\Big(- v(U)v(F^{\leftarrow}(U)) + \theta(u;v) - \int_0^1 \big(\mathbf{1}_{U\leq t} - t\big)v'(t)u(F^{\leftarrow}(t)dt)\Big)^2$
   où U est la fonction de répartition d'une loiuniforme sur$[0,1]$.
2. *La variable aléatoire $\sqrt{n}\big(\hat{b}_t(u,v) - b_t(u,v)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\psi^2)$.*
   $\psi^2 = \dfrac{t^2\sigma^2}{v^2 p_t}\left((u^{\leftarrow})'\dfrac{t\theta(u,v)}{vp_t}\right)^2$

Proposition 3.11.

Soit X une variable aléatoire positive de loi F telle que $F(0) = 0$.

• *Si $v(.)$ est une fonction positive,strictement croissante et Lipshitzienne d'ordre 1 sur $[0,1]$ et $u(.)$ est une fonction positive,continue et strictement croissante telle que $\mathbb{E}(u^3(X)) < +\infty$ alors :*

• *Si σ^2 est nul,*

$$\lim_{n\to\infty} \Pr\left(\frac{\hat{\theta}_n(u,v)}{v(F(x))} < 1 - F(x)\right) = 0$$

Sous les conditions des deux propositions,on peut remplacer $\theta(u,v)$ par $\hat{\theta}(u,v)$ dans l'inégalité de Markov et estimer ba borne supérieure du niveau de retour par :

$$\hat{b}_t(u,v) = u^{\leftarrow}\left(\frac{t\hat{\theta}(u,v)}{vp_t}\right) \tag{3.24}$$

Et un intervalle de confiance :

$$IC_{b_t} = \left[\hat{b}_t(u,v) \pm q_{1-\frac{1}{2}}\frac{t\hat{\sigma}}{\sqrt{n}vp_t}(u^{\leftarrow})'\left(\frac{t\hat{\theta}_n(u,v)}{vp_t}\right)\right] \tag{3.25}$$

où :

$\hat{\sigma}$ est l'estimateur de σ.

$q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centré réduite.

B.Borne inférieure du niveau de retour

Une borne inférieure du niveau de retour de $\hat{x}_t$ a été estimée par $\hat{l}$ selon :

$$\hat{x}_t \geq \hat{l}_t(u,w,q) = u^{-1}\left(\frac{\hat{\theta}^*(u,v) - t^{\frac{1}{q-1}}(\hat{\theta}^*(u,v))^{\frac{1}{q}}}{w(\frac{1}{t})p_t}\right) \tag{3.26}$$

avec :
$\hat{\theta}^*(u,v) = \frac{1}{n}\sum_{i=1}^{n} u(X_{i,n})w\left(1-\frac{1}{n}\right)$. Et w est une fonction positive et décroissante définie sur $[0,1]$ et $q > 1$.

3.4.2 Approximation du période de retour

A défaut de calculer explicitement la période de retour d'un quantile théorique $\hat{x}_t$. On propose une démarche pour estimer ce temps de retour T en deux étapes :

Étape 01 :

• Utiliser les estimateurs des deux bornes(sup et inf) pour tracer le graphe $(t, \hat{l}_t, \hat{b}_t)$ avec $\hat{l}_t \leq \hat{x}_t \leq \hat{b}_t$.

• La valeur de $\hat{l}_t$ étant constante pour tout t,nous utiliserons la borne supérieure du niveau de retour pour estimer approximativement le temps de retour de $\hat{x}_t$.

Étape 02 :

• Soit une nouvelle observation x_{t_0} à t_0.

• On pose $x_{t_0} = \hat{b}_T$.

• Nous lisons sur le graphe précédent la valeur T correspondante associe au niveau de retour où au quantile(inconnu) $\hat{x}_T$ avec $\hat{x}_t \leq \hat{b}_t$.

D'où :

$x_{t_0} \geq \hat{x}_T \Longleftrightarrow \exists \delta_t \geq 0$ tel que : $x_{t_0} = \hat{x}_{T+\delta_t}$

Donc le temps de retour théorique $(T + \delta_t)$ de l'observation x_{t_0} est supérieur au temps de retour T de $\hat{b}_t$.

Alors il serait naturel de procéder de même avec la borne inférieur $\hat{l}_t$ puisque $\hat{l}_t \leq \hat{x}_t \leq \hat{b}_t$.

$$T_l = \frac{1}{1-F(\hat{l}_t)} \leq T = \frac{1}{1-F(\hat{x}_t)} \leq T_b = \frac{1}{1-F(\hat{b}_t)}$$

où : T_l, T et T_b désignent les temps de retour associées à $\hat{l}_t$, $\hat{x}_t$ et $\hat{b}_t$.

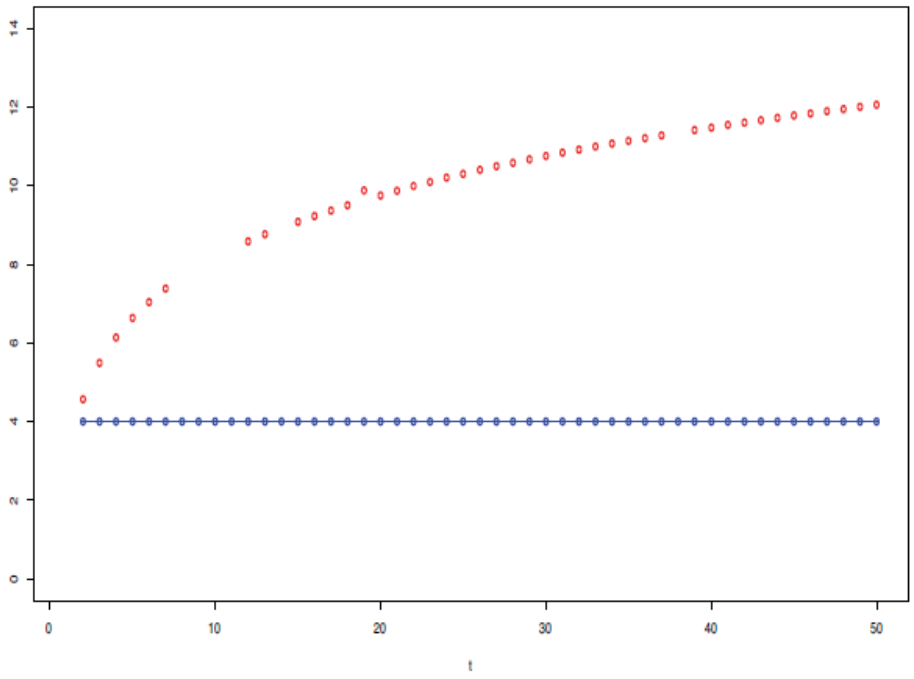

FIGURE 3.3 – Bornes du niveau de retour

L'axe des abscisses représente les valeurs de t et l'axe des ordonnées correspond aux bornes du niveau de retour ou du quantile. Les points rouges correspondent a la borne supérieure du niveau de retour b , les points bleus correspondent a la borne inférieure du niveau de retour .

3.5 Conclusion

Dans ce chapitre on a exposé une autre méthode pour étudier les valeurs extrêmes, la méthode **POT** (***Peaks-Over-Tershold***) où ***méthode des excès au-delà d'un seuil***. Cette dernière est basée sur la distribution de Paréto généralisée, elle a l'avantage de diminuer la perte d'information par rapport à la méthode précédente.
Ce chapitre souligne l'essentiel du fondement théorique de cette méthode et le théorème de ***Pickands-Belkama-De Haan*** qui est considéré comme le deuxième théorème fondamental de la théorie des valeurs extrêmes. On a estimé les paramètres avec trois méthodes différentes mais après la détermination du seuil afin d'arriver au le but essentiel des valeurs extrêmes, qui est l'estimation de niveau de retour et de la période de retour et en a clôturé ce chapitre par une approximation de la période de retour et le niveau de retour.

Deuxième partie :

Partie d'application

Chapitre 4

Application en hydrologie

Dans ce chapitre, nous réalisons une étude de cas réelle pour modéliser la distributions des valeurs extrêmes généralisées, les données sont obtenues suite à une collecte des données de barrage de Beni-Haroun (Mila).
Les objectifs principaux de ce chapitre sont :

1. Déterminer la loi.
2. Estimer les paramètres.
3. Prédire des futurs observations ainsi que la période de retour et le niveau de retour.

4.1 Problématique

Les crues d'origines naturelle peuvent causent le plus de dégâts en termes de vie humaine,matériels.
La connaissance des débits de crue reste un axe de recherche important en hydrologie pour la conception des aménagements des cours d'eau, le dimensionnement des ouvrages de franchissement et la protection des zones urbaines.
Les prévisions des valeurs extrêmes et des crues permettent d'élaborer un plan de prévention des risques inondation.
Les hydrologues sont exposés à deux problématiques majeures :

1. Comment estimer les débits de fréquence d'apparition rares (période de retour de plus de 100 ans) sur les sites où on dispose d'information pluvio-hydrométrique uniquement d'une dizaine d'années d'observation ?
2. Même question que précédemment mais sur les bassins où l'on ne dispose d'aucune information ?

La législation impose aux concepteurs des ouvrages des fréquences d'apparition d'environ 100 ans pour les digues, 1000 ans pour les barrages en béton et 10000 ans pour les barrages en remblais.

4.2 Application

D'abord on fait une simple analyse statistique descriptive, les résultats selon le langage $\mathcal{R}$ sont :

Min	1 st Qu	Median	Mean	3rd Qu	Max
-902.465	0.33	0.724	2.630	2.904	901.934

TABLE 4.1 – Stat-des des données

On remarque que la plage des variations des données est entre -902.465 (Hm^3)et 901.934 (Hm^3) mais la moyenne est 2.630 (Hm^3). Dans ce cas la moyenne n'est pas représentative. Donc on parle des extrêmes.

4.2.1 L'approche GEV

La distribution des valeurs extrêmes généralisée modélise le comportement du maximum d'un échantillon.
Pour déterminer la loi et estimer ces paramètres, on utilise la méthode du "maximum par blocs" qui consiste à construire un échantillon de maximum à partir d'un échantillon de données en formant des blocs de même dimension.

A partir 5234 observation on a former 172 blocs chaque bloc contient un seul maximum. La statistique descriptive du nouveau tableau est :

Min	1 st Qu	Median	Mean	3rd Qu	Max
0.1690	0.9272	3.3415	14.7211	9.7897	901.3400

TABLE 4.2 – stat-desc des blocs

Intervalle de confiance

Moy	$IC_{95\%}$
14.7211	$[4.3095, 25.1328]$

TABLE 4.3 – IC de la moyenne des blocs

Estimation des paramètres on utilisons la méthode MLE

Les résultats selon le langage $\mathcal{R}$ sont comme suit :

location	scale	shape
1.746263	2.253963	1.278713

TABLE 4.4 – Est-MLE pour GEV

Estimation des paramètres on utilisons la méthode LM

Les résultats selon le langage $\mathcal{R}$ sont comme suit :

location	scale	shape
2.413073	3.734732	0.736260

TABLE 4.5 – Est-LM pour GEV

Intervalles de confiance

	95% lower CI	Estimate	95% upper CI
location	1.367772	1.746263	2.124754
scale	1.654443	2.253963	2.853482
shape	1.063660	1.278713	1.493767

TABLE 4.6 – IC des paramètres de GEV

Interprétation des résultats

On a estimer les paramètres du modèle par 2 méthodes : MLE et L-Moments.

La moyenne actuelle des affluents est 14.721 (Hm^3), dans un intervalle de confiance $[4.3095; 25.1328]$, cette moyenne est statistiquement significatif au seuil $\alpha = 5\%$.

D'après les résultats obtenus les 3 paramètres sont statistiquement significatifs au seuil $\alpha = 5\%$.

1-L'estimateur du paramètre de localisation est inclue dans l'intervalle de confiance pour

un seuil $\alpha = 5\%$.

La largeur de cet intervalle est $L = 0.76$, donc il a une petite largeur, plus l'intervalle est petit plus qu'il est précis c-à-d il s'approche de la vraie valeur.

2-Le paramètre de dispersion (d'échelle) est inclue dans un intervalle de confiance $[4.3095; 25.1328]$.

3- Le paramètre le plus important dans les distributions extrêmes est l'indice de queue, l'estimateur de ce paramètre $\hat{\xi}$ qu'on a trouvé est inclue dans un intervalle de confiance a une largeur très petite (0.43) (tend vers zéro). Et on remarque qu'il est le plus petit intervalle par rapport aux autres estimateurs.

Choix de la distribution :

On a $\hat{\xi} = 1.27 > 0$ et $0 \notin IC_{\hat{\xi}}$ donc notre modèle suit la distribution de **Fréchet**.

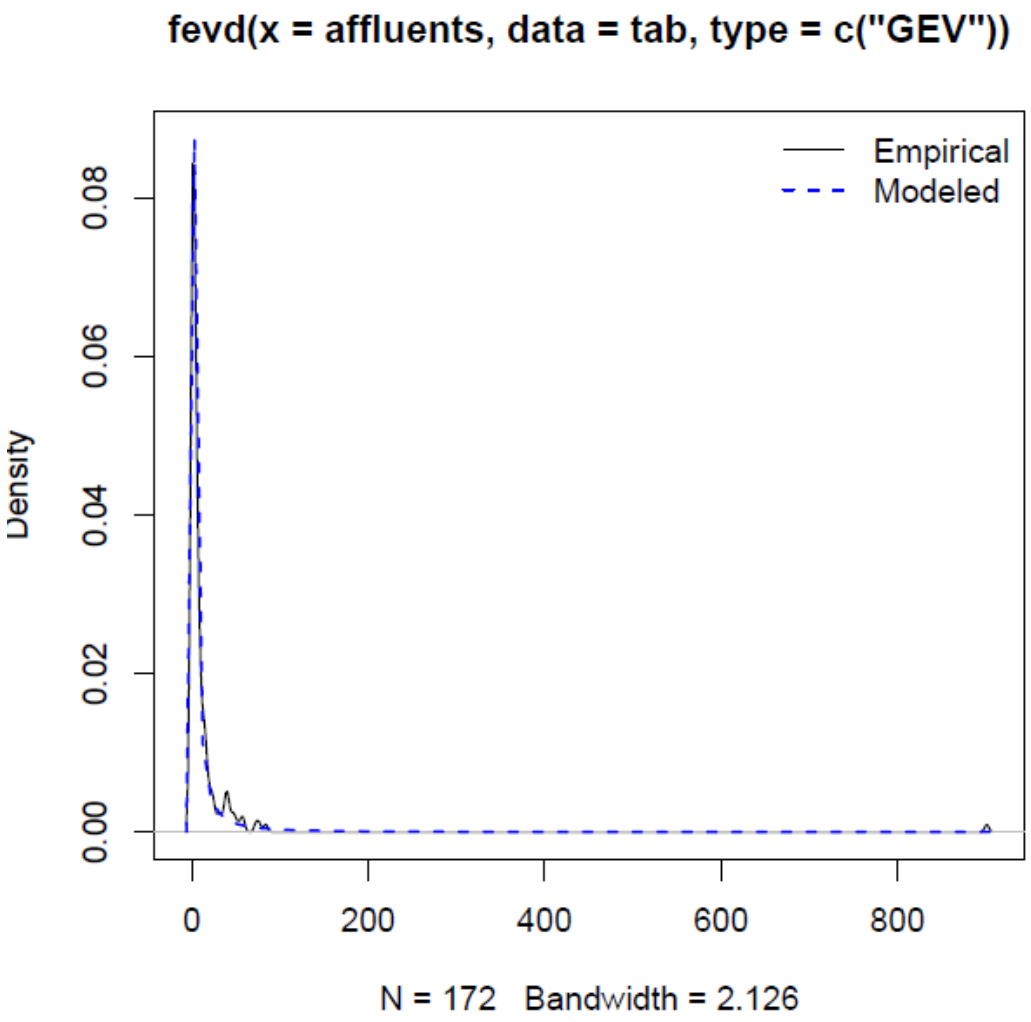

FIGURE 4.1 – Densité du modèle

La représentation graphique de la densité de notre modèle est adéquate avec la densité de Fréchet. Comme notre modèle admet Fréchet comme densité alors on conclue qu'il est à queue lourde c-à-d la queue d'une façon polynomiale.

Remarque 7.

Si $\xi = 0$ est dans l'intervalle de confiance ! Modèle de Gumbel adapté ?

Après la détermination de la distribution du modèle et estimer ses paramètres on arrive à notre but principal de l'étude du comportement des extrêmes qui est : l'estimation de la période de retour et le niveau de retour.

Période de retour et niveau de retour

Dans cette partie on va donner des prévisions à court terme et à long terme pour la période de retour et le niveau de retour associé. Les résultats obtenus sont les suivant :

	level	IC
3 ans	5.575	$[4.0313, 7.1178]$
5 ans	11.983	$[7.7109, 16.2544]$
10 ans	31.308	$[16.1571, 46.4596]$
15 ans	53.812	$[25.5008, 84.1228]$

TABLE 4.7 – Niveau de retour pour GEV

Interprétation des résultats

1. À court terme (3 ans - 5 ans)

   • Pour 3 ans :

   On a une probabilité $1/T = 1/3$ pour avoir un niveau de crue de moyenne 5.575 (Hm^3).

   Ce résultat est statistiquement significatif au seuil $\alpha = 5\%$.

   • Pour 5 ans : On a une probabilité $1/T = 1/5$ pour avoir un niveau de crue de moyenne 11.983 (Hm^3). Ce résultat est statistiquement significatif au seuil $\alpha = 5\%$.

2. À long terme (10 ans- 15 ans)

   • Pour 10 ans :

   On a une probabilité $1/T = 1/10$ pour avoir un niveau de crue de moyenne31.308 (Hm^3).

   Ce résultat est statistiquement significatif au seuil $\alpha = 5\%$.

   • Pour 15 ans :

On a une probabilité $1/T = 1/15$ pour avoir un niveau de crue de moyenne 53.812 (Hm^3).

En général, ce niveau est noté dans la période (Décembre-Janvier). La crue survient souvent après de fortes pluies en amont dans le bassin versant, plus rarement lors de la fonte des neiges. Ce résultat est statistiquement significatif au seuil $\alpha = 5\%$.

⋆ On remarque plus que la période s'éloigne l'intervalle de confiance s'élargit. La même remarque qu'on peut remarqué sur le graphe qui exprime la période de retour en fonction du niveau de retour.

A partir de 20 ans les intervalles de confiance s'élargissent. Donc notre étude est valable pour une durée compris entre 5 ans et 10 ans. **Mais** avec les même conditions actuelles.

Critiques :

Le départ était avec 5234 observations, après la construction des blocs on a gardé que 172 observations, ce qui provoque une perte d'information, c'est l'inconvénient de la **GEV**.

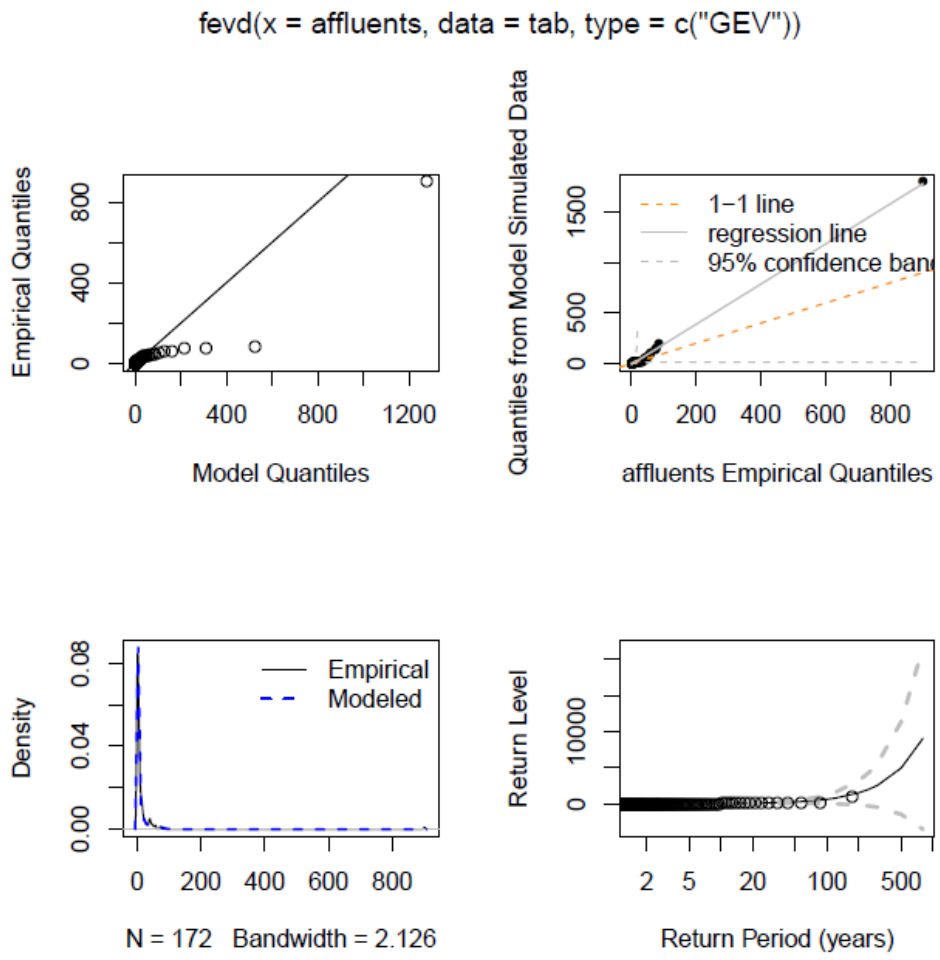

FIGURE 4.2 – Rep.G-Modèles dans le cas GEV

Probabilité et probabilité empirique du modèle

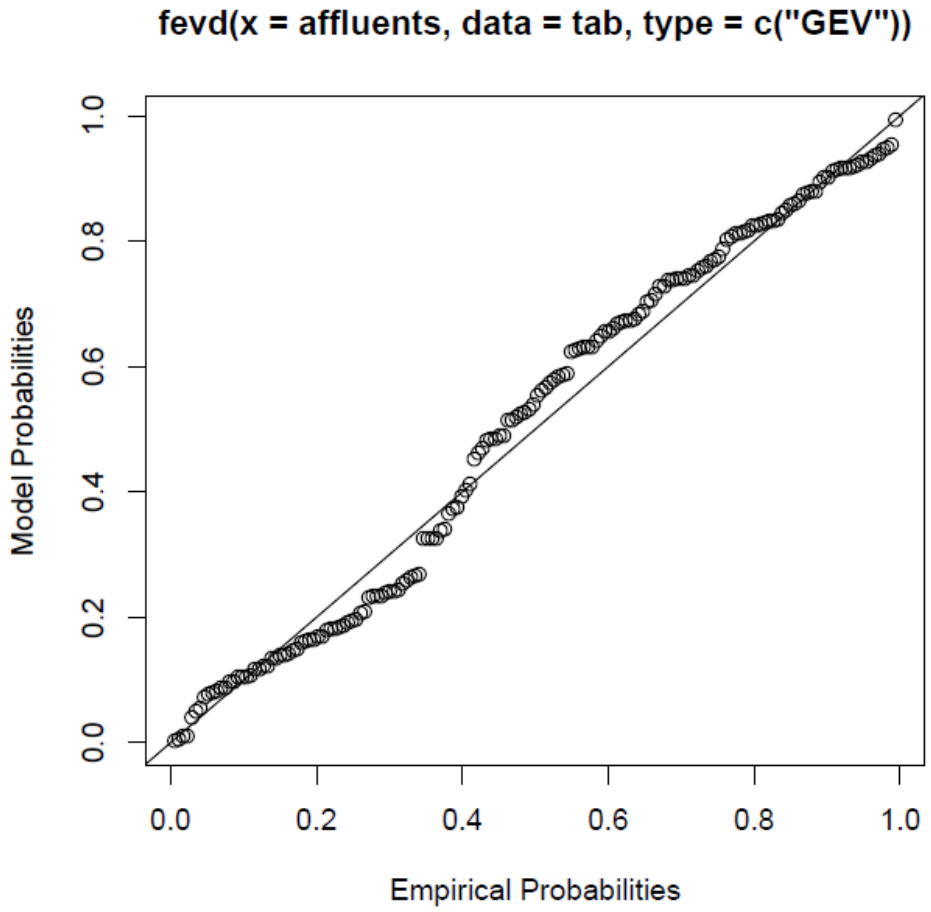

FIGURE 4.3 – Probabilité et probabilité empirique (GEV)

Les points sont proches de la diagonale de l'unité.

4.2.2 L'approche GPD

Dans cette partie de l'application, on va travailler avec le tableau initial nommé "tab", mais après la détermination d'un seuil, car cette méthode consiste à étudier les données qui dépassent ce dernier.

Choix du seuil

Selon le langage $\mathcal{R}$ le résultat est le suivant :

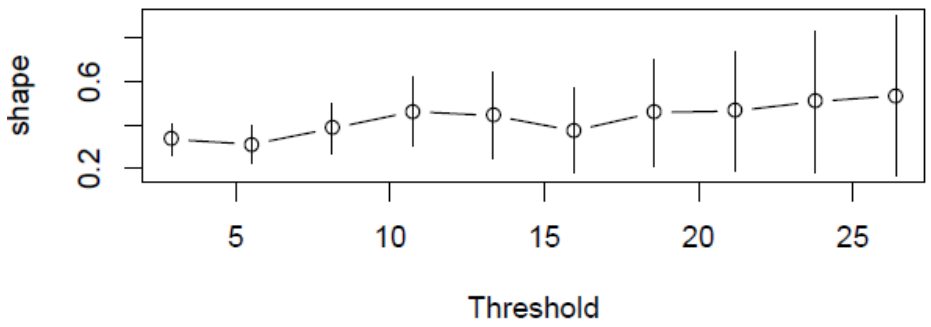

FIGURE 4.4 – Stabilité de la queue

Pour déterminer le seuil on a utilisé une méthode graphique basée sur la stabilité de paramètre indice de queue ξ. c-à-d a partir de quel niveau ξ devient stable ?
D'après le résultat on remarque que ξ est stable entre le niveau 10 et 15 puis il se perturbe légèrement puis il devient stable à partir du niveau 18.
Nous choisissons 10.65 (Hm^3) comme un seuil.

Distribution du modèle et estimation des paramètres

scale	shape
5.5691751	0.4527368

TABLE 4.8 – Es.Paramètres pour la méthode POT

Intervalles de confiance

	95% lower CI	Estimate	95% upper CI
scale	4.5374483	5.5691751	6.6009019
shape	0.2953281	0.4527368	0.6101455

TABLE 4.9 – IC pour les estimateurs (POT)

D'après ces résultats, l'estimateur d'indice de queue $\hat{\xi} = 0.4527 > 0$.
De plus on a $IC_{\hat{\xi}} = [0.3; 0.61]$ donc $0 \notin IC_{\hat{\xi}}$, alors notre modèle suit la distribution de Paréto .

Remarque 8.

Si $\xi = 0$ est dans l'intervalle de confiance ! Modèle de exponentiel adapté ?

Période de retour et niveau de retour

Dans cette partie on va donner des prévisions à court terme et à long terme pour la période de retour et le niveau de retour associé. Les résultats qu'on obtenus sont les suivants :

	level	IC($\alpha = 5\%$)
3 ans	80.887	$[55.8531, 105.9211]$
5 ans	102.363	$[64.2729, 140.4539]$
10 ans	140.707	$[75.6392, 205.7743]$
15 ans	169.392	$[81.7293, 257.0538]$

TABLE 4.10 – Retour-POT

D'après les résultats obtenu de cette étude laissent penser à la survenue des crues d'amplitudes assez importantes.
Ce résultat peut se justifier par l'échauffement climatique et ses conséquences sur le changement de climat et la perturbation des pluies.

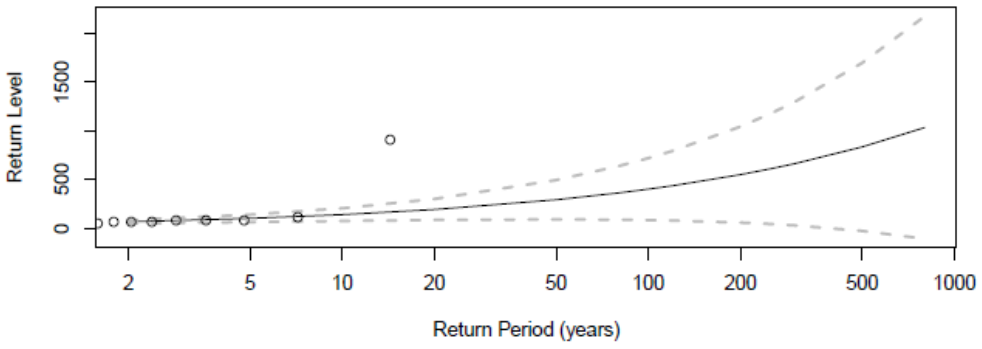

FIGURE 4.5 – Niveau de retour (POT)

La représentation graphique de la période de retour en fonction du niveau de retour exprime que les estimateurs de la période de retour et le niveau de retour sont inclus dans des intervalles de confiance large, plus la période est s'éloigne plus que l'intervalle est s'élargir.

4.2.3 Comparaison entre GEV et GPD

• Dans la première partie le nombre d'observations après la construction des blocs était 172 observations par contre dans la deuxième partie le nombre d'observation augmente jusqu'à 320 observations ce qui diminue la perte d'information.

Entre GEV et **POT**

Dans la **GEV** on ne prend en compte que les max $\sim$ Dans la méthode POT on ne prend en compte que les dépassement de seuils.

Choix de la taille du bloc $\sim$ Choix du seuil

compromis :

Biais (données pas assez extrêmes) $\sim$ Variance(pas assez de données).

• Le point commun qu'on a constaté c'est que l'indice de queue n'est pas changé ($\xi > 0$).

• Mais les résultats concernons la période de retour et le niveau de retour sont complètement différents peut être dù à la richesse de la GPD par rapport au GEV.

Conclusion générale

Au cour de ce travail, On a partagé le travail en deux partie, partie théorique et partie d'application. La partie théorique dans laquelle on a procédé à la présentation des fondements théoriques des valeurs extrêmes des distributions uni-variées qui modélisent les phénomènes extrêmes. On a exposé deux approches différentes (GEV et GPD). On a estimé les paramètres et les quantiles et la période de retour et le niveau de retour de chaque approche, et on a exposé des résultats obtenus dans la première approche. Et on a clôturé cet partie théorique par une approximation concernant la période de retour et le niveau de retour. La seconde partie est une réalisation de la première partie c-à-d on a appliquer les deux approches précédentes, pour analyser la variable "crue" aux données issues de barrage de Beni Haroun. Dans ce cadre on a arrivé à l'objectif de notre application qui est l'estimation la période de retour et le niveau de retour. On a déterminé la période de la survenue des crus avec le niveau associé.

Bibliographie

[1] **Anis Borchani**, *Statistiques des valeurs extrêmes dans le cas discrètes*, Research center. ESSEC working paper 10009, Décembre 2009.

[2] **Alexendre Leikima**, *Estimation non paramétrique des quantiles extrêmes conditionnels.*, thèse de doctorat, université de Grenoble 2006.

[3] **Bachir Reggad**, *Fondements de la théorie des valeurs extrêmes, ses principales application et son apport à la gestion des risques du marché pétrolier*, Math. Sci. Hum, Mathematics and social sciences, 47^{me} Année, N 186. (2009)2. 29-63.

[4] **Christan. Y.Robert**, *Cour Théorie des valeurs extrêmes* , ISFA-Université de Lyon 1, 2016.

[5] **Enrique Castillo**, *Extreme value theory in Engineering*, Spain (1988).

[6] **J.Berger**, *Lecture Notes in statistics.*

[7] **Laurens de Haan, Ana Ferreira**, *Extreme value theory . An introduction*, (2006)

[8] **Meraghni Djamel**, *Modelling Distribution Tails.*, doctorat thesis ,Biskra's university 2008.

[9] **Michael Falk**, *Best attrainable rate of joint convergence of extremes.*, Département ok mathematics, university of Siegen Holderlinstr. 3, 5900 Siegen 21, Fedral Republic of Germany.

[10] **M. Ait Yala Nabil**, *Modélisation des extrêmes spatiaux et application.*, Mémoire de magistère, Université Mouloud Mammeri de Tizi Ouzou , Année universitaire 2012-2013.

[11] **Resnick S.I**, *Extreme values, Regular variation, and Point Processes.*, Applied Probability Trust, Springer-Verlag. New York, (1987).

[12] **Temame Naima**, *Estimation du quantile extrême et VaR* , Mémoire de magister - École de doctorat Tizi-Ouzou, 2011.

Résumé

Les événements extrêmes ont été étudiés par de nombreux chercheurs durant les dernières décennies. Les statisticiens ont également développé des outils statistiques capables de traiter ce type de données statistiques. Dans ce mémoire, on a réalisé une synthèse concernant les outils fondamentaux utilisés dans l'analyse des valeurs extrêmes.
D'abord, on a procédé à la présentation des résultats fondamentaux des distributions univariées qui modélisent les phénomènes extrêmes, leurs différents domaines d'attraction et le coefficient de normalisation ainsi que l'estimation des paramètres par deux méthodes différentes. Puis on a exposé la distribution généralisée de Pareto comme deuxième approche pour la modélisation des valeurs extrêmes.
Enfin, on propose d'appliquer les deux approches précédentes, pour analyser la structure de dépendance pour le couple de variables "cote-crue" aux données issues de barrage de Beni Haroun. Les données proviennent de l'archive de la direction du barrage de Beni Haroun dans la wilaya de Mila entre 2004 et début 2018.

Abstract

Extreme events have been studied by many researchers in recent decades. Statisticians have also developed statistical tools capable of handling this type of statistical data. In this thesis, we synthesized the fundamental tools used in the analysis of extreme values. First, we presented the fundamental results of univariate distributions that model extreme events, their different domains of attraction and the coefficient of normalization and the estimation of parameters by two different methods. Then we exposed the generalized distribution of Pareto as a second approach for modeling extreme values. Finally, it is proposed to apply the two previous approaches, to analyze the dependency structure for the pair of "flood-side" variables to data from the Beni Haroun dam. The data come from the archive of the direction of the Beni Haroun dam in the wilaya of Mila between 2004 and early 2018.