République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**UNIVERSITÉ MOHAMED KHIDER, BISKRA**

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

**DÉPARTEMENT DE MATHÉMATIQUES**

Mémoire présenté en vue de l'obtention du Diplôme **:**

**MASTER en Mathématiques**

Option : **Statistique**

Par

**TOUMI SAMIA**

Titre :

# Distribution des Excès et Selection du Seuil

Membres du Comité d'Examen :

| | | | |
|---|---|---|---|
| Pr. | NECIR ABDELHAKIM | UMKB | Président |
| Pr. | BENATIA FATEH | UMKB | Encadreur |
| Dr. | SOLTANE LOUIZA | UMKB | Examinateur |

Septembere 2020

## Dédicace

Je dédie ce humble travail

A mes *parents*.

A mes *frères*.

A mes *soeurs*.

A mes *famille*.

A mes *amis*.

A mes collegues de mathématiques $2019 - 2020$.

## REMERCIEMENTS

Tout d'abord, je remercie "**Allah**" le tout Puissant de m'avoir aidé et donné la santé et volonté pour arriver à ce stade.

Mes vifs remerciements, sont adressés à mon encadreur **Pr. Benatia Fateh** pour ses précieux conseils, ses orientations pertinentes et sa patience tout au long de la réalisation de ce mémoire.

Je tiens à remercier : **Pr. NECIR ABDELHAKIM** et **Dr. SOLTANE LOUIZA** qui m'ont fait l'honner de faire partie du jury de soutenance.

Je remercie tout les enseignants qui ont contribué à ma formation, ainsi que tous les employés du département de mathématiques.

Je remercie tout particulièrement mes parents, pour leur encouragement et soutien sur tout les aspects, ainsi que toute ma famille.

Je n'oublie pas l'ensemble de mes amis mes proches et aussi mes collégues d'études.

A ceux qui ont contribué, de près ou de loin, à la réalisation de ce modeste travail : un grand merci à vous tous.

# Table des matières

# Table des figures

# Liste des tableaux

# Introduction

La théorie des probabilités est une branche des mathématiques qui traite des propriétés de certaines structures modélisant des phénomènes survenant.

La théorie des probabilités permet de modéliser efficacement certains phénomènes aléatoires et d'en faire l'étude théorique.

Les statistique d'ordre sont très utiles en théorique des valeurs extrêmes.

La théorie des valeurs extrêmes est une branche des statistiques qui s'intéresse aussi bien aux valeurs extrêmes et leurs caractéristiques que leurs distributions de probabilité.

Lorsque l'on étudie un phénomène aléatoire, on s'intéresse principalement à la partie dite centrale de la loi modélisant aux mieux le phénomène considéré (calcul de l'espérance, la médiane, la variance,...) .

Le domaine d'application sont en effet très variés : hydrologie, météorologie, finance, assurance, ...

La modélisation par la loi $GPD$ passe par l'estimation des ces paramètres à savoir les paramètres de position échelle et l'indice de queue. Ces derniers peuvent être estimer par différentes méthode.

La méthode des excès au -delà d'un seuil repose sur le comportement des données qui dépassent certain un seuil donnée (fixé).

Le mémoire et organisé comme suit :

Nous commençons donc par rappeler dans le premier chapitre quelques éléments théorique sur espace des probabilités, la variable aléatoire et la statistique d'ordre.

Dans le deuxième chapitre, nous donnons un aperçu sur la théorie des valeurs extrême. Nous avons

présenter les caractérisations générales, lois des valeurs extrêmes et le domaine d'attraction ...

Le troisième chapitre, nous présentons un aperçu sur la loi des excès et quelques méthodes d'estimation du seuil.

# Chapitre 1

# Statistique d'ordre

Avant d'aborder les statistiques d'ordre, nous allons d'abords commencé par énoncer quelques notions élémentaires de probabilité et de statistiques nécessaire pour toute la suite, en le premier lieu l'espace probabilisé, la variable aléatoire et sa loi de probabilité.

## 1.1 Espace des probabilités

A toute expérience on associe l'ensemble de tous les résultats possibles noté $\Omega$ et appelé ensemble fondamentale ; et soit $\mathcal{F}$ une tribu définie sur cet ensemble.

Le couple $(\Omega, \mathcal{F})$ s'appelle espace probabilisable.

Tout élément de $\mathcal{F}$ s'appelle événement. Soit $\omega \in \Omega$ le résultat observé de l'expérience, $A$ est un événement si $\omega \in A$, on dit que $A$ est réalisé.

Soit $P$ une application de $\mathcal{F}$ à valeur dans $[0, 1]$, vérifiant les propriétés suivantes :

1. $P(\Omega) = 1$.
2. $\forall A \in \mathcal{F}$; $P\left(\overline{A}\right) = 1 - P(A)$.

3. $\forall A, B \in \mathcal{F};\ P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

$$\begin{array}{rccl} P & : & (\Omega, \mathcal{F}) & \rightarrow & [0,1] \\ & & A & \rightarrow & P(A). \end{array}$$

$P$ est dite probabilité. L'espace $(\Omega, \mathcal{F}, P)$ est dite espace de probabilité (ou espace probabilisé).

## 1.2 Variable aléatoire

On peut aussi bien s'intéresse à l'éventualité $\omega$, qu'à une mesure lié à $\omega$ notée $X(\omega)$ est appelé variable aléatoire.

**Définition 1.2.1**

*Une variable aléatoire réelle (v.a.r) $X$ est une fonction définie sur un espace probabilisable $(\Omega, \mathcal{F})$ à valeurs dans $\mathbb{R}$, et mesurable par rapport à la tribu $B_{\mathbb{R}}$ (tribu borélienne de $\mathbb{R}$).*

$$\begin{array}{rccl} X & : & \Omega & \rightarrow & \mathbb{R} \\ & & \omega & \rightarrow & X(\omega). \end{array}$$

*Selon l'ensemble d'arrivé, il ya deux types de variables aléatoires : discrète et continue.*

*-On dit qu'une v.a est discrète si elle ne prend que des valeurs dans un ensemble fini ou au plus dénombrable. En règle générale, toutes les variables qui sont le résultat d'un dénombrement ou d'une numération sont de type discrète.*

*-On dit qu'une v.a est continue si elle prend ses valeurs dans un intervalle de $\mathbb{R}$.*

## 1.3 Loi de probabilité

Soit $X$ une $v.a$ sur un espace probabilisé $(\Omega, \mathcal{F}, P)$. On appelle loi de probabilité de la variable $X$, la probabilité noté $P_X$ définie sur l'espace $(\mathbb{R}, B_{\mathbb{R}})$ vers $[0,1]$ par :

$$\begin{aligned} \forall B \in B_{\mathbb{R}} : P_X(B) &= P\{X^{-1}(B)\} \\ &= P\{w \in \Omega / X(w) \in B\} \\ &= P\{X \in B\} \end{aligned}$$

### 1.3.1 Probabilité conditionnelle

Pour tout évènement $B$ de probabilité non nulle i.e $P(B) \neq 0$, on appelle probabilité conditionnelle à $B$, la probabilité sur $(\Omega, \mathcal{F})$

$$\begin{array}{rccl} P & : & \mathcal{F} & \rightarrow & [0,1] \\ & & A & \rightarrow & P_B(A) := \frac{P(A \cap B)}{P(B)}, \end{array}$$

$P_B(A)$ s'appelle probabilité conditionnelle à $B$ de $A$ (ou encore probabilité de $A$ sachant $B$). On note aussi $P_B(A) = P(A/B)$.

### 1.3.2 Caractéristiques d'une loi de probabilité

1. Une fonction de répartition est une fonction croissante au sens large.
2. La limite de $F(x)$ quand $x$ tend vers $-\infty$ est égale à 0, et on note $F(-\infty) = 0$.
3. La limite de $F(x)$ quand $x$ tend vers $+\infty$ est égale à 1, et on note $F(+\infty) = 1$.
4. Une fonction de répartition est continue à droite admet une limite à gauche.

### 1.3.3 Densité de probabilité

On appelle variable aléatoire à densité toute fonction $X : \Omega \longrightarrow \mathbb{R}$ telle qu'il existe une fonction $f : \mathbb{R} \longrightarrow \mathbb{R}$ continue par morceaux vérifiant la propriété suivante :

$\forall a < b,\ P\left(X \in [a,b]\right) = \int_a^b f\left(x\right)\, dx.$

Si une telle fonction $f$ existe, elle est appelée densité de la $v.a\ X$.

### 1.3.4 Caractéristiques d'une densité de probabilité

1. Si $f$ est la densité d'une $v.a\ X$, alors nécessairement $f$ est positive, $f$ est intégrable sur $\mathbb{R}$ et vérifie $\int_{-\infty}^{+\infty} f\left(x\right)\, dx = 1$.
2. La fonction de densité ou densité de probabilité, égale à la dérivée de la fonction de répartition.

**Proposition 1.3.1** *:*

$$E[X] = \begin{cases} \int_{\mathbb{R}} x f(x) dx & \textit{dans le cas d'une v.a continue.} \\ \sum_{i\in\Omega} x_i P(X = x_i) & \textit{dans le cas discrète.} \end{cases}$$

*et,*

$$V\left[X\right] = \begin{cases} \int_{\mathbb{R}} (x - E(x))^2 f(x) dx & \textit{dans le cas d'une v.a continue.} \\ \sum_{i\in\Omega} (x_i - E(x_i))^2 P(X = x_i) & \textit{dans le cas discrète.} \end{cases}$$

$\sigma = \sqrt{V\left[X\right]}$ : *l'écart type.*

### 1.3.5 Fonction de répartition

**Définition 1.3.1**

*La fonction de répartition de la v.a.r $X$ est l'application $F$ de $\mathbb{R}$ dans $[0,1]$ définie par :*

*Si la v.a est continue alors :*

$$\forall x \in \mathbb{R}, \quad F_X(x) := P\left(\{-\infty, x[\right) = P\left(\omega \in \Omega / X(\omega) \leqslant x\right)$$

*ou encore :* $F_X(x) = P\left(X \leqslant x\right) = \int_{-\infty}^{x} f\left(t\right) dt$. *Où* $f_X\left(x\right)$ *est la densité de probabilité de* $X$.

*Si la v.a est discrète on a :*

$$F_X(x) := P\left(X = x\right) = P\left(\{\omega \in \Omega / X(\omega) = x\}\right).$$

## 1.4 Statistique d'ordre

Soient $X_1, X_2, ..., X_n$ $n$ *v.a* indépendantes et identiquement distribuées (*iid*) de densité $f$ et de fonction de répartition $F$.

**Définition 1.4.1 (Statistique d'ordre)**

*On appelle statistique d'ordre la suite de v.a notées :* $X_{1,n}, X_{2,n}, ..., X_{n,n}$, *et ordonnées comme suit :* $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$.*et pour* $i = 1, ..., n$ *, la v.a* $X_{i,n}$ *est appelée la* $i - iéme$ *statistique d'ordre* (*ou statistique d'ordre de rang* $i$).

*On note par* $X_{1,n}$ *la plus petite statistique d'ordre et* $X_{n,n}$ *la plus grande statistique d'ordre et qui sont définie par :*

$$X_{1,n} := \min\left(X_1, X_2, ..., X_n\right) \quad et \quad X_{n,n} := \max\left(X_1, X_2, ..., X_n\right)$$

**Exemple 1.4.1**

$X_1 = 5,\ X_2 = 2,\ X_3 = 1,\ X_4 = 7,\ X_5 = 4.$

$X_{1.5} = X_3,\ X_{2.5} = X_2,\ X_{3.5} = X_5,\ X_{4.5} = X_1,\ X_{5.5} = X_4.$

**Proposition 1.4.1**

*La relation entre les statistiques d'ordre extrêmes est la suivante :*

$$\min(X_1, X_2, ..., X_n) = -\max(-X_1, -X_2, ..., -X_n).$$

*La distance entre les statistique d'ordre extrêmes, est appelée déviation extrême ou l'étendu de l'échantillon, notée $D_n$ donnée par :*

$$D_n := X_{n,n} - X_{1,n}.$$

**Corollaire 1.4.1 (Distribution de statistique d'ordre)**

*Soit $X_1, X_2, ..., X_n$ n v.a iid de densité commune $f$ et de fonction de répartition $F$, alors on a :*

**a)** La fonction de densité de $X_{1,n}$ est donnée par

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

**b)** La fonction de répartition de $X_{1,n}$ est donnée par

$$F_{X_{1,n}}(x) = 1 - \left(\bar{F}(x)\right)^{n} = 1 - (1 - F(x))^{n}, \quad x \in \mathbb{R}.$$

**c)** La fonction de densité de $X_{n,n}$est donnée par

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

**d)** La fonction de répartition de $X_{n,n}$ est donnée par

$$F_{X_{n,n}}(x) = (F(x))^{n}, \quad x \in \mathbb{R}. \tag{1.1}$$

**e)** La fonction de densité conjointe de $X_{1,n}$ et $X_{n,n}$ est donnée par

$$f_{X_{1,n},X_{n,n}}(x,y) = n(n-1) f(x) f(y) (F(y) - F(x))^{n-2}, \quad -\infty < x < y < +\infty.$$

**f)** La densité jointe de $X_{1,n}, X_{2,n}, ..., X_{n,n}$ est donnée par

$$f_{X_{1,n},X_{2,n},...,X_{n,n}}(x_1, x_2, ..., x_n) = n! \prod_{i=1}^{i=n} f(x_i), \quad x \in \mathbb{R}.$$

**g)** La densité de $X_{i,n}$ pour $i = 1, ..., n.$ est donnée par

$$f_{X_{i,n}}(x) = n \begin{pmatrix} n-1 \\ i-1 \end{pmatrix} (F(x))^{i-1} \left(\bar{F}(x)\right)^{n-i} f(x), \quad x \in \mathbb{R}.$$

**h)** La fonction de répartition de $X_{i,n}$ pour $i = 1, ..., n.$est donnée par

$$F_{X_{i,n}}(x) = \sum_{i=1}^{i=m} \begin{pmatrix} n \\ m \end{pmatrix} (F(x))^{m} \left(\bar{F}(x)\right)^{n-m}, \quad x \in \mathbb{R}.$$

Où $\begin{pmatrix} n \\ m \end{pmatrix} = \frac{n!}{m!(n-m)!}$ est la combinaison de $m$ objets parmi $n$ objets sans remise.

**Remarque 1.4.1 (1)**

*Pour plus de détails sur démonstration en peut cité* ***Bateka*** *[3]*

**Remarque 1.4.2 (2)**

*Les v.a's de la statistique d'ordre $X_{1,n}, X_{2,n}, ..., X_{n,n}$ ne sont pas indépendantes et de loi différents.*

***Exemple 1.4.2***

*Soit $U_1, U_2, ..., U_n$ une suite de v.a iid selon une loi $Uniforme$ sur $[0,1]$.*

*Déterminer la loi de* $U_{1,n}$ *et* $U_{n,n}$.

*On a* $F_U(x) = \begin{cases} 0 & Si \quad x < 0 \\ x & Si \quad 0 \leqslant x < 1 \\ 1 & Si \quad x \geqslant 1 \end{cases}$

*et* $f_U(x) = \frac{\partial F_U(x)}{\partial x} = I_{[0,1]}$ *Où* $I_{[0,1]}$ *est la fonction indicatrice.de* $[0,1]$.

**a)** $f_{U_{1,n}}(x) = \begin{cases} n & \text{Si} \quad x < 0 \\ n(1-x)^{n-1} & \text{Si} \quad 0 \leqslant x < 1 \\ 0 & \text{Si} \quad x \geqslant 1 \end{cases}$

**b)** $F_{U_{1,n}}(x) = \begin{cases} 0 & \text{Si} \quad x < 0 \\ 1-(1-x)^n & \text{Si} \quad 0 \leqslant x < 1 \\ 1 & \text{Si} \quad x \geqslant 1 \end{cases}$

**c)** $f_{X_{n,n}}(x) = \begin{cases} 0 & \text{Si} \quad x < 0 \\ nx^{n-1} & \text{Si} \quad 0 \leqslant x < 1 \\ n & \text{Si} \quad x \geqslant 1 \end{cases}$

**d)** $F_{U_{n,n}}(x) = \begin{cases} 0 & \text{Si} \quad x < 0 \\ x^n & \text{Si} \quad 0 \leqslant x < 1 \\ 1 & \text{Si} \quad x \geqslant 1 \end{cases}$

# Chapitre 2

# Valeurs extrêmes

La formule 1.1 montre que la loi du maximum est relié d'une manière principale à $F(x)$, mais cette dernière n'est pas toujours connue, même si elle est connue, la loi du maximum n'est pas facile à calculer. Donc on s'intéresse à étudier les comportements asymptotiques du maximum en faisant tendre $n$ vers l'infini. De plus amples détails pour la théorie des valeurs extrêmes ($EVT$, englais.Extreme value theory) peuvent être trouves dans les références : **de Haan, Ferreria** [9], **Reiss-Thomas**(1997) [19] et, **Galambos** [10].

## 2.1 Caractérisations générales

Dans cette partie, nous commençons d'abords par présenter des généralité de probabilité et statistique avant de passer aux distributions des valeurs extrêmes.

**Définition 2.1.1 (Fonction de répartition empirique)**

*Soit $X_1, X_2, ..., X_n$ échantillon de $n$ $v.a's$ $iid$ de fonction de répartition $F$, alors la fonction de*

*répartition empirique est donnée par :*

$$\begin{aligned} F_n(x) &:= \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \preceq x\}}, \quad -\infty < x < +\infty \\ &= \begin{cases} 0 & si \quad x < X_{1,n} \\ i/n \quad si \quad X_{i,n} \leq x < X_{i+1,n}, & 2 \leqslant i < n \\ 1 & si \quad x \geqslant X_{n,n} \end{cases} \end{aligned}$$

*Où* $I_A(x)$ *est la fonction indicatrice définie par :* $I_A(x) = \begin{cases} 1 \quad si \quad x \in A \\ 0 \quad si \quad x \notin A \end{cases}$

**Définition 2.1.2 (Point terminal)**

*Le point terminal d'une distribution de loi* $F$ *notée* $x_F$ *est définie par :*

$$x_F := \sup_{x \in \mathbb{R}} \{x : F(x) < 1\}.$$

### 2.1.1 Convergence de la fonction de répartition empirique

**Théorème 2.1.1 (Glivenko-Canteli)**

*Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de v.a.r iid et de même loi* $F$*, et de fonction de répartition empirique* $F_n$*, alors :*

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

**Définition 2.1.3 (Fonction quantile et quantile de queue)**

*La fonction quantile lié à la fonction de répartition* $F$ *est la fonction inverse généralisée de* $F$ *notée* $F^{\leftarrow}$ *définie par :*

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

*Si* $F$ *est continue et strictement croissante alors :* $Q(p) = F^{-1}(p)$.

*Les fonctions $F^{\leftarrow}$ et $U$ sont liées par la relation suivante :*

$$U(p) := F^{\leftarrow}(1 - 1/p).$$

*et $U$ est appelée la fonction quantile de queue.*

*La fonction quantile empirique est donnée par :*

$$Q_n(p) := F_n^{\leftarrow}(p) = \inf_{x \in \mathbb{R}} \{x : F_n(x) \geqslant p\}, \quad 0 < p < 1.$$

*La fonction empirique du quantile de queue est :*

$$U_n(p) := F_n^{\leftarrow}(1 - 1/p).$$

### 2.1.2 Théorème central limite

Le théorème central limite ($T.C.L$) établit, sous des hypothèses très peu contraignantes, la convergence en distribution d'une moyenne de *v.a iid* vers la loi normale centrée réduite .

**Théorème 2.1.2 (TCL)**

*Soit $X_1, X_2, ..., X_n$ est une suite des v.aɪs iid de moyenne $\mu$ et de variance $\sigma^2$ finie alors :*

$$\sqrt{n}\left(\frac{S_n - \mu}{\sigma}\right) \xrightarrow{D} N(0,1) \quad \textit{quand } n \to \infty$$

*où $S_n := \sum_{i=1}^{n} X_i$ somme arithmétique.*

### 2.1.3 lois des grandes nombres

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

**Théorème 2.1.3** ($L.G.N$)

*Soit $X_1, X_2, ..., X_n$ est une suite des v.a's iid de moyenne $\mu = E[X] < \infty$, alors :*

$$La\ loi\ faible\ : \overline{X} \xrightarrow{P} \mu\ quand\ n \to \infty$$
$$La\ loi\ forte\ : \overline{X} \xrightarrow{Ps} \mu\ quand\ n \to \infty,$$

*où $\overline{X} := \sum_{i=1}^{n} X_i$ moyenne empirique.*

**Remarque 2.1.1**

*Pour plus de détails sur le théorème T.C.L et L.G.N en peut cité le livre* ***Saporta*** *[22].*

## 2.2 Loi des valeurs extrêmes

Le théorème $Fisher - Tippet\,(1928)$, $Gnedenko\,(1943)$ est fondamental en théorie des valeurs extrêmes.

### 2.2.1 Comportement asymptotique

Les distribution asymptotique du maximum quand $n$ tendre vers l'infini est :

$$F_{X_{n,n}}(x) = (F(x))^n.$$

$$\lim_{n\to\infty} F_{X_{n,n}}(x) = \lim_{n\to\infty} (F(x))^n = \begin{cases} 0 & si \quad x < x_F \\ 1 & si \quad x \geqslant x_F \end{cases}$$

La loi asymptotique de $X_{n,n}$ est dégénérée. Pour cette raison, on essaie de trouver des constantes $a_n$ et $b_n$ de telle sorte que la loi du maximum normalisée soit non dégénérée.

Notons $Z_n = \frac{X_{n,n}-b_n}{a_n}$ avec $a_n \in \mathbb{R}$ et $b_n > 0$.

$$F_{Z_n}(x) = P\left(\frac{X_{n,n}-b_n}{a_n} \leq x\right) = [F(a_n x + b_n)]^n$$

On étudie

$$\lim_{n\to\infty} F_{Z_n}(x) = \lim_{n\to\infty} [F(a_n x + b_n)]^n .$$

**Définition 2.2.1**

*On appelle loi asymptotique des extrêmes, la loi de la v.a $Z$ telle que :*

$$Z_n = \frac{X_{n,n}-b_n}{a_n} \xrightarrow{L} Z, \ \ quand\ n \to \infty$$

*Soient deux suites $(a_n)_{n\geqslant 1}$, $a_n > 0$ et $(b_n)_{n\geqslant 1}$, $b_n \in \mathbb{R}$ telles que :*

$$F_{X_{n,n}}(a_n x + b_n) \longrightarrow H(x) \ \ quand\ n \to \infty,$$

*en tout point $x$ de continuité de $H$, où $H$ est la fonction de répartition de $Z$.*

**Théorème 2.2.1 (*Fisher − Tippet* (1928), *Gnedenko* (1943))**

*S'il existe deux suites $(a_n)_{n\geq 1} \geq 0$, $(b_n)_{n\geq 1} \in \mathbb{R}$ tels que :*

$$\lim_{n\to\infty} F_{Z_n}(x) = H(x), \quad \forall x \in \mathbb{R}.$$

*Alors, $H$ est non dégénérée et elle est l'une des trois types suivants :*

*Type* 1 *: Weibull*

$$\Psi_\gamma(x) = \begin{cases} \exp\left(-\left(-\frac{x-\mu}{\sigma}\right)^{\frac{-1}{\gamma}}\right) & si \quad x \leqslant 0 \\ & \gamma < 0 \\ 1 & si \quad x > 0 \end{cases}$$

*Type* 2 *: Gumbel*

$$\Lambda_0(x) = \left\{ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \quad x \in \mathbb{R} \right.$$

*Type* 3 : *Fréchet*

$$\phi_\gamma(x) = \begin{cases} 0 & si \quad x \leqslant 0 \\ & \gamma > 0 \\ \exp\left(-\left(\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right) & si \quad x > 0 \end{cases}$$

*avec $\Psi_\gamma$, $\Lambda_0$, $\phi_\gamma$ sont des lois limites de $X_{n,n}$.*

*Si on prend $\mu = 0$ et $\gamma = 1$ on obtient les expressions standard de ces lois.*

1. $\Psi_\gamma(x) = \begin{cases} \exp\left(-(-x)^{\frac{-1}{\gamma}}\right) & si \quad x \leqslant 0 \\ & \gamma < 0 \\ 1 & si \quad x > 0 \end{cases}$
2. $\Lambda_0(x) = \left\{ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \quad x \in \mathbb{R} \right.$
3. $\phi_\gamma(x) = \begin{cases} 0 & si \quad x \leqslant 0 \\ & \gamma > 0 \\ \exp\left(-x^{-\frac{1}{\gamma}}\right) & si \quad x > 0 \end{cases}$

**Proposition 2.2.1 (Relation entre $\Psi_\gamma, \Lambda_0, \phi_\gamma$)**

*Soit $X$ une v.a positive alors les affirmations suivantes sont équivalentes :*

1. $X \rightsquigarrow \phi_\gamma$.

2. $\ln X^{\gamma} \rightsquigarrow \Lambda_0$.

3. $-X^{-1} \rightsquigarrow \Psi_{\gamma}$.

## 2.3 Le domaine maximal d'attraction et les constantes de normalisation

### 2.3.1 Caractérisation des domaines d'attraction

Si $F$ vérifie le théorème 2.2.1, alors on dit que $F$ appartient au domaine d'attraction de $H\gamma$ et selon le signe de $\gamma$, on distingue trois domaines d'attraction :

- Si $\gamma < 0$, on dit que $F$ appartient au domaine d'attraction de $Weibull$, la queue de cette distribution supérieure est finie, donc supérieurement bornée. Utiliser particulièrement en Fiabilité, l'étude de durée de vie des machines. Ce domaine d'attraction, comprends entre autre les lois Bêta, Uniforme...
- Si $\gamma = 0$, on dit que $F$ appartient au domaine d'attraction de $Gumbel$, la queue de cette distribution décroit d'une façon exponentielle, c'est une distribution non bornée, dont les moments existes. Utiliser particuliérement en Hydrologie(Etude de Grues) . Ce domaine d'attraction, comprends entre autre les lois de Gamma, Normale, Exponentielle, etc...
- Si $\gamma > 0$, on dit que $F$ appartient au domaine d'attraction de $Fréchet$, la queue de cette distribution décroit avec une vitesse polynomiale. C'est une distribution non bornée et dont les moment ne sont pas finis. Ce domaine d'attraction, comprends entre autre les lois de Paréto, de Student, de Cauchy, etc...

**Définition 2.3.1 (Fonction à variation lente)**

*On dit que L est une fonction à variation lente pour tout x du domaine de définition et on a :*

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

### 2.3.2 Condition nécessaire et sufisante de convergence

- Domaine d'attraction maximale de *Fréchet*

Soit $X_1, X_2, ..., X_n$ une suite de $n$ *v.a iid* ayant une loi de probabilité $F$, $F \in D(\phi_\gamma(x))$ si et seulement si : $F^{-1}(1) = +\infty, \exists\ \gamma > 0$ et vérifiant :

$$\lim_{t\to\infty}\frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\frac{1}{\gamma}} = -\log(\phi_\gamma(x)) \quad \forall x \succeq 0.$$

- Domaine d'attraction maximale de *Gumbel*

Soit $X_1, X_2, ..., X_n$ une suite de n *v.a iid* ayant une loi de probabilité $F$, $F \in D(\Lambda(x))$ si et seulement si : $E(X/X > c) < +\infty$,

$\forall c < F^{-1}(1)$ et $F$ vérifie :

$$\lim_{t\to\infty}\frac{\bar{F}(t + xE(X/X \succ c))}{\bar{F}(t)} = \exp(-x) = -\log(\Lambda_\gamma(x))$$

- Domaine d'attraction maximale de *Weibull*

Soit $X_1, X_2, ..., X_n$ une suite de n *v.a iid* ayant une loi de probabilité $F$, $F \in D(\Psi_\gamma(x))$ si et seulement si : $F^{-1}(1) < +\infty$, $\exists\gamma < 0$ et vérifiant :

$$\lim_{t\to\infty}\frac{\bar{F}(F^{-1}(1) - tx)}{\bar{F}(F^{-1}(1) - t)} = -x^{-\frac{1}{\gamma}} = -\log(\Psi_\gamma(x))$$

Constante de normalisation

**Théorème 2.3.1**

*On peut choisir des constantes* $a_n \geq 0$ *et* $b_n \in \mathbb{R}$ *pour* $n \in \mathbb{N}$ *telles que* (2.3.2) *soit vérifiée de la manière suivante :*

1) $b_n = 0$, $a_n = F^{-1}\left(1 - \frac{1}{n}\right)$ pour $H_\gamma = \phi_\gamma$

2) $b_n = F^{-1}\left(1 - \frac{1}{n}\right)$, $a_n = F^{-1}\left(1 - \frac{1}{n\exp(1)}\right) - F^{-1}\left(1 - \frac{1}{n}\right)$ pour $H_0 = \Lambda$

3) $b_n = F^{-1}(1)$, $a_n = F^{-1}(1) - F^{-1}\left(1 - \frac{1}{n}\right)$ pour $H_\gamma = \Psi_\gamma$.

**Remarque 2.3.1**

1. La fonction de répartition $H_\gamma$ est appelée loi des valeurs extrêmes. Le paramètre $\gamma$ est un paramètre de forme et encore appelé indice des valeurs extrêmes ou indice de queue, $a_n$ est un paramètre de position et $b_n$ est un paramètre d'échelle.
2. Les suites de normalisation $(a_n)_{n\geqslant 1}$ et $(b_n)_{n\geqslant 1}$ ne sont pas uniques.

**Exemple 2.3.1**

*Soit* $E_1, E_2, ..., E_n$ *une suite de v.a iid de loi exponentielle de paramètre* 1*, de fonction de répartition*

$F(x) = 1 - \exp(-x)$, $x \in \mathbb{R}_+$.

*Puisque* $F \in D(\Lambda(x))$ *alors les constantes de normalisation sont :*

$F(x) = y \Rightarrow F^{-1}(y) := x = -\ln(1-y)$

$b_n = F^{-1}\left(1 - \frac{1}{n}\right) = \ln(n)$, $a_n = F^{-1}\left(1 - \frac{1}{n\exp(1)}\right) - F^{-1}\left(1 - \frac{1}{n}\right) = 1$

*d'où*

$\lim_{n\to\infty} F\left(\frac{X_{n,n}-b_n}{a_n}\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty}\left(1 - \frac{\exp(-x)}{n}\right)^n = \exp(-\exp(-x))$.

**Exemple 2.3.2**

*Soit* $C_1, C_2, ..., C_n$ *une suite de v.a iid de loi Cauchy de paramètre* 0 *et* 1 *c-à-d* $C(0,1)$, *de fonction de densité* $f(x) = \frac{1}{\Pi}\frac{1}{1+x^2}$.

*Puisque* $F \in D(\phi_1(x))$ *alors la fonction de répartition et les constantes de normalisation sont :*

$y >> \infty$, $y^2 + 1 \backsimeq y^2$

$F(x) = \int_{-\infty}^{x} f(t)\, dt = 1 - \int_{x}^{+\infty} \frac{1}{\Pi}\frac{1}{1+t^2} dt \backsimeq 1 - \frac{1}{\Pi}\int_{x}^{+\infty} \frac{1}{t^2} dt \backsimeq 1 - \frac{1}{\Pi x}$

$F(x) = y \Rightarrow F^{-1}(y) := x = \frac{1}{\Pi(1-y)}$

$a_n = F^{-1}\left(1 - \frac{1}{n}\right) = \frac{n}{\Pi}$, $b_n = 0$

*d'où*

$$\lim_{n\to\infty} F\left(\frac{X_{n,n} - b_n}{a_n}\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty} F^n\left(\frac{n}{\Pi}x\right) \backsimeq \lim_{n\to\infty}\left(1 - \frac{1}{nx}\right)^n = \exp(-x^{-1}).$$

**Exemple 2.3.3**

*Soit* $U_1, U_2, ..., U_n$ *une suite de v.a iid de loi Uniforme sur* $[0,1]$, *de fonction de répartition*

$F(x) = x\ I_{[0,1]}$.

*Puisque* $F \in D(\Psi_{-1}(x))$ *alors les constantes de normalisation sont :*

$F(x) = y \Rightarrow F^{-1}(y) := x = y$

$b_n = F^{-1}(1) = 1$, $a_n = F^{-1}(1) - F^{-1}\left(1 - \frac{1}{n}\right) = 1 - \left(1 - \frac{1}{n}\right) = \frac{1}{n}$

*d'où*

$$\lim_{n\to\infty} F\left(\frac{X_{n,n} - b_n}{a_n}\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = \exp(x) = \exp(-(-x)).$$

## 2.4 Représentation de Jenkinson(1954)Von-Mises(1955)

Etant donné qu'il est difficile de travailler avec trois distributions des valeurs extrêmes à la fois. Grâce aux travaux de Von-Mises et de Jenkinson, Ils ont regroupés ces 3 types de fonction sous une même formalisation.

**Définition 2.4.1**

*La représentation de $Jenkinson-Von-Mises$ de la loi des valeurs extrêmes que l'on appelle "loi des valeurs extrêmes généralisées" notée $GEV$ a pour fonction de répartition :*

$$H_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right) & Si\ \gamma \neq 0\ et\ pour\ tout\ x\ :\quad 1+\gamma x > 0 \\ \exp\left(-\exp\left(-x\right)\right) & Si\ \gamma = 0, \quad x \in \mathbb{R}. \end{cases}$$

*Pour les variables centrées et réduites, on peut écrire $H_\gamma$ sous une forme plus générale (appelée forme paramétrée de $Von\ Mises$) :*

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

*Où $\mu$ est un paramètre de position, $\sigma$ est celui de dispersion, $\gamma$ l'indice.des queues.*

*Nous exprimons les trois distributions des valeur extrêmes $\phi_\gamma$, $\Lambda_0$, $\Psi_\gamma$ en termes de $GEV$. La Figure 2.1 ci-dessous illustre le comportement de $GEV$ standard.*

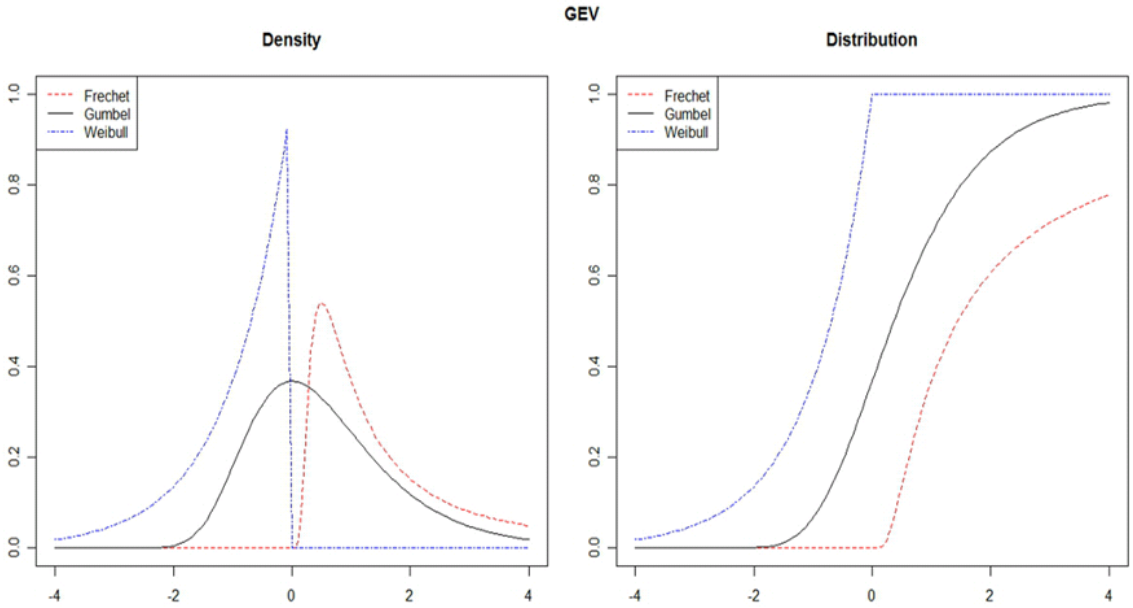


FIG. 2.1 – Densités standard des valeurs extrêmes

**Propriété 2.4.1**

1. $\boldsymbol{E}[X] = \frac{1}{\gamma}\Gamma(1-\gamma) - \frac{1}{\gamma}$, $\gamma \neq 0$
2. $\boldsymbol{E}[X^2] = \frac{1}{\gamma^2}\Gamma(1-2\gamma) - \frac{2}{\gamma^2}\Gamma(1-\gamma) + \frac{1}{\gamma^2}$, $\gamma \neq 0$
3. $\mathbf{V}[X] = \frac{1}{\gamma^2}\left(\Gamma(1-2\gamma) - (\Gamma(1-\gamma))^2\right)$, $\gamma \neq 0$
4. $Q(_p) = \begin{cases} -\log(-\log(p)) & \gamma = 0 \\ \frac{(-\log p)^{-\gamma}-1}{\gamma} & \gamma \neq 0. \end{cases}$

# Chapitre 3

# Loi des excès et selection du seuil

Le point faible de la distribution des valeurs extrêmes généralisés est qu'elle ne prend en considération qu'une seule valeur $(X_{n,n})$, pour éviter cette contrainte il est préferable de travailler avec plusieurs valeurs extrêmes le choix de celles-ci sera détaillé dans ce qui suit.

## 3.1 Distribution des excès

**Définition 3.1.1 (Fonction de répartition des excès)**

*Soit $X_1, X_2, ..., X_n$ une suite d'observations iid, de fonction de répartition $F$ et $x_F$ le point terminal. Alors, pour un seuil $u < x_F$ fixé, on étudie la distribution des valeurs dépassant $u$ appelées excès au- dessus du seuil $u$ la variable et définie par $Y_j = X_i - u$ pour $0 \leq j \leq N_u$ où $N_u$ est le nombre de dépassements du seuil $u$ par les $X_{i \leq n}$ et les $Y_{j \leq N_u}$ sont les excès correspondants Figure* 3.1

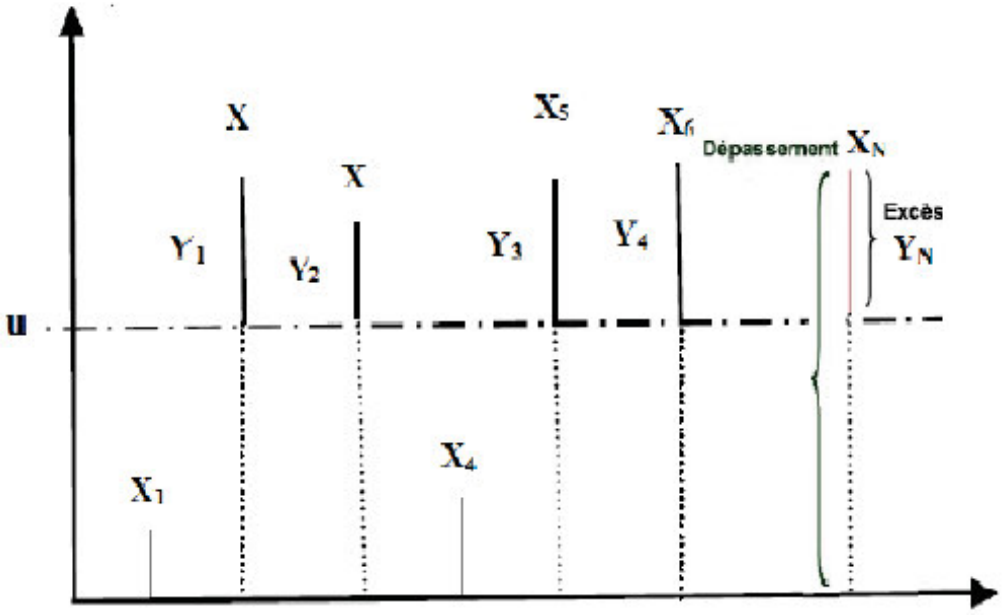


FIG. 3.1 – Dépassement de seuil ($POT$)

*La fonction de distribution des excès de $X$ au- dessus du seuil $u$ est :*

$$\begin{aligned} F_u(x) &= P(X-u \leqslant x \mid X>u) \\ &= \tfrac{P(X-u\leqslant x\ ,\ X>u)}{P(X>u)} \\ &= \tfrac{P(u<X<x+u)}{P(X>u)} \\ &= \tfrac{F(x+u)-F(u)}{1-F(u)} \quad pour\ 0 \leq x \leq x_F - u. \end{aligned}$$

## 3.2 Distribution de Pareto Généralisée$(GPD)$

La fonction de distribution de Pareto généralisée, joue un rôle essentiel dans la modélisation des excès.

**Définition 3.2.1**

*La fonction de distribution de Pareto généralisée standard* $(GPD,$ *en anglais generalized Pareto distribution*$)$

*définie pour* $\gamma \in \mathbb{R}$ *et* $\beta > 0$ *s'écrit sous la forme :*

$$G_{\gamma,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma}{\beta}x\right)^{-1/\gamma} & si \quad \gamma \neq 0 \\ 1 - \exp\left(-x/\beta\right) & si \quad \gamma = 0 \end{cases}, \; x \in G_{\gamma,\beta}.$$

*où :*

$$S(\gamma,\beta) = \begin{cases} x \geqslant 0 & si \quad \gamma \geqslant 0 \\ 0 \leqslant x \leqslant -\beta/\gamma & si \quad \gamma < 0 \end{cases}$$

*est le support de* $G_{\gamma,\beta}$.

*et sa densité s'écrit alors :*

$$g_{\gamma,\beta}(x) = \begin{cases} \frac{1}{\beta}\left(1 + \frac{\gamma}{\beta}x\right)^{-(1/\gamma)-1} & si \quad \gamma \neq 0 \\ (1/\beta)\exp\left(-x/\beta\right) & si \quad \gamma = 0. \end{cases}$$

*La Figure 3.2 illustre la distribution et la densité de la loi GPD.*

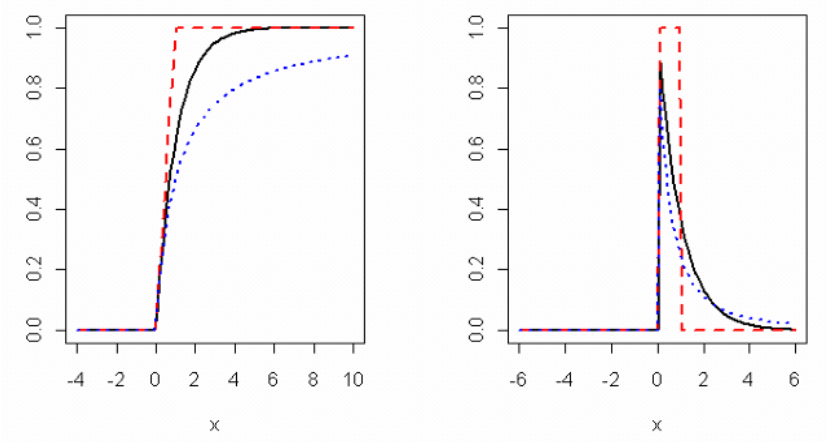


FIG. 3.2 – Densité (à droite) et la distribution (à gauche) de *GPD* standard.

*La GPD avec deux paramètre regroupe trois distributions selon les valeurs du paramètre de forme. Lorsque* $\gamma > 0$*, c'est la loi de Pareto ; lorsque* $\gamma < 0$*, nous avons la loi de Bêta et* $\gamma = 0$ *donne la loi Exponentielle.*

*Cas particulier, lorsque* $\begin{cases} \gamma = 0 & c'est\ la\ loi\ \exp(\lambda = 1/\beta) \\ \gamma = -1 & c'est\ la\ loi\ U_{[0,\beta]}. \end{cases}$

*Le paramètre* $\gamma$ *est le même d'une GEV et* $\beta$ *: paramètre d'échelle.*

## 3.3 Théorème de Balkema-de Haan-Pickands

Le théorème suivant fait le lien entre le comportement asymptotique de la distribution des excès et la distribution de Pareto Généralisée ($GPD$).

**Théorème 3.3.1**

*Si* $F_u$ *est la fonction de répartition des excès au-déla d'un seuil* $u$ *tel que* $u$ *tend vers le point terminal* $x_F$*, on a :*

$$\lim_{u \to x_F} \sup_{0 \leq x \leq x_F \ -u} | F_u(x) - G_{\gamma,\beta}(x) | = 0$$

*ou* $\beta(u)$ *une fonction positive mesurable.*

**Exemple 3.3.1**

*"La loi Exponentielle"*

*Soit* $F(x) = 1 - \exp(-x)$, $x > 0$

*et* $F_u(y) = \frac{F(u+y)-F(u)}{1-F(u)} = \frac{(1-\exp(-(u+y)))-(1-\exp(-u))}{1-(1-\exp(-u))} = 1 - \exp(-y)$, $y > 0$.

*On trouve la loi Exponentielle qui est également la loi GPD de paramètre* $\gamma = 0$ *et* $\beta = 1$.

### 3.3.1 Propriétés de la GPD

1. Si $\gamma > 0$, la distribution $G_{\gamma,\beta}(x)$ est la Pareto usuelle avec $\alpha = 1/\gamma$ et $k = \beta/\gamma$

$G_{\gamma,\beta}(x) = 1 - \left(1 + \frac{\gamma}{\beta}x\right)^{-1/\gamma} = 1 - \left(\frac{1}{1+(\gamma/\beta)x}\right)^{1/\gamma} = 1 - \left(\frac{\beta}{\beta+\gamma x}\right)^{1/\gamma} = 1 - \left(\frac{\beta/\gamma}{\beta/\gamma+x}\right)^{1/\gamma}$.

2. Si $\gamma = 0$, la distribution $G_{\gamma,\beta}(x)$ est une distribution exponentielle.

   $\lim_{\gamma\to 0} G_{\gamma,\beta}(x) = \lim_{\gamma\to 0} 1 - \left(1 + \frac{\gamma}{\beta}x\right)^{-1/\gamma} = 1 - \lim_{\gamma\to 0} \exp\left(\frac{-1}{\gamma}\ln\left(1 + \frac{\gamma}{\beta}x\right)\right) = 1 - \lim_{\gamma\to 0} \exp\left(\frac{-x}{\beta}\frac{\ln\left(1+\frac{\gamma}{\beta}x\right)}{\frac{\gamma x}{\beta}}\right) =$
   $1 - \exp(-x/\beta)$.

   On a $\lim_{u\to 0}\frac{\ln(1+u)}{u} = 1$.

3. Si $\gamma < 0$, c'est la loi de Pareto de type II.

# 3.4 Estimation des paramètres de la GPD

## 3.4.1 Méthode du maximum de vraisemblance

L'estimation par la méthode du maximum de vraisemblance donne des résultats asymptotique efficaces.

Soit $(X_1, X_2, ..., X_n)$ un n-échantillon, les $X_i$ supposé $iid$ de densité $g_\theta$ où $\theta = (\gamma, \beta)$.

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

$L_{(\gamma,\beta,X)} = \prod_{i=1}^{n} g_{(\gamma,\beta,X)}(x_i)$.

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

$$\begin{cases} \frac{\partial l(x_1,x_2,...,x_n,\theta)}{\partial\theta} = 0 \\ \frac{\partial^2 l(x_1,x_2,...,x_n,\theta)}{\partial^2\theta} < 0 \end{cases}$$

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

$l(x_1, x_2, ..., x_n, \theta) = \log L_{(\gamma,\beta,X)} = -n\log\beta - \frac{1}{\beta}\sum_{i=1}^{n} x_i$.

Dans le cas $\gamma \neq 0$, la fonction log de la vraisemblance égale à :

$logL_{(\gamma,\beta,X)} = -n\log\beta - \left(\frac{1}{\gamma} + 1\right)\sum_{i=1}^{n}\log\left(1 + \frac{\gamma}{\beta}x_i\right)$.

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

$$\begin{cases} \frac{\partial logL_{(\gamma,\beta,\mathbf{X})}}{\partial\beta} = 0 \iff -n + (\gamma+1)\sum_{i=1}^{n}\frac{x_i}{\beta+\gamma x_i} = 0. \\ \frac{\partial logL_{(\gamma,\beta,\mathbf{X})}}{\partial\gamma} = 0 \iff \frac{1}{\gamma}\sum_{i=1}^{n}\log\left(1+\frac{\gamma}{\beta}x_i\right) - (\gamma+1)\sum_{i=1}^{n}\frac{x_i}{\beta+\gamma x_i} = 0. \end{cases}$$

La résolution de ce système est relativement difficile et n'admet pas en général de solution explicites Dans ce cas, en utilisant les méthodes d'optimisation numériques c'est ce qui est fait par exemple par la fonction $f_{gev}$ du package $R$

### 3.4.2 Estimateur de Hill

**Définition 3.4.1**

*Soit* $(X_1, X_2, ..., X_n)$ *un échantillon de v.a iid de fonction de répartition* $F$ *avec* $F \in D(\phi_\gamma(x))$ *et* $\gamma > 0$*, on définit l'estimateur du paramètre de queue* $\gamma$ *par :*

$$H_{k,n} = \gamma_{k,n}^{Hill} = \frac{1}{k}\sum_{i=1}^{n}\log(X_{n-i+1,n}) - \log(X_{n-k,n}) .pour \quad k \in \{1, ..., n-1\}.$$

## 3.5 Sélection du seuil

Le choix du seuil doit établir un compromis entre le biais et la variance, le seuil doit être suffisamment grand pour pouvoir utiliser les résultats asymptotiques, mais pas trop élevé pour obtenir des estimations précises. Par contre le choix d'un seuil faible risque de déclarer abusivement des observations extrêmes, introduire un biais dans l'estimation et par conséquent, mal approximer la loi asymptotique.

L'estimation des paramètres $(\gamma, \beta)$ de la distribution $GPD$ pose le problème de la détermination du seuil $u$, le seuil ne doit pas être trop grand car il faut suffisamment de données pour que

l'approximation $GPD$ soit valide, généralement la technique est la plus utilisé pour estimer $u$. Il ya plusieurs méthodes d'estimation du seuil :

### 3.5.1 Méthodes graphiques

Le Peak Over Threshold ($POT$) est une technique qui a été développée par des chercheurs, et elle est fréquemment utilisée dans la structure des valeurs extrêmes. L'approche $POT$ consiste à ajuster un modèle paramétrique pour que ses excès au-dessus d'un seuil $\mu$ soient assez élevés. En d'autre termes, cette technique permet d'évaluer si le choix du seuil $u$ est adéquat pour être représenté par un modèle asymptotique. Afin de comprendre cette notion, la section suivante fera une brève introduction de quelques méthodes utilisant la technique $POT$

**Mean Residual life**($MRL - plot$)

Le graphe de la durée de vie moyenne résiduelle($MRL - plot$) introduit par **Division et Smith** [6] utilise la méthode d'espérance des excès de $GPD$, $\mathbb{E}\left(X - u \mid x > u\right) = \beta/\left(1 - \gamma\right)$, comme diagnostique, définie pour $\gamma < 1$ pour s'assurer de l'existence de la moyenne [5]. Pour tout $u > u^*$ assez grand, l'attente devient

$$e\left(u\right) = \mathbb{E}\left(X_i - \mu \ / \ X_i > u\right) = \frac{\beta_{u^*}}{1-\gamma} + \frac{\gamma}{1-\gamma}u, \quad \gamma < 1$$

qui est linéaire en $u^*$ avec $\gamma/\left(1-\gamma\right)$ et $\beta_{u^*}/1-\gamma$.

oú $u$ : indique le seuil

$\gamma$ : indique le paramètre de forme

$\beta_u$ : indique le paramètre d'échelle correspondant au seuil $u$

$e\left(u\right)$ est la moyenne des excès au delà du seuil

Le graphe de la durée de vie moyenne résiduelle est le graphe des points $\left\{\left(\mu, e_n\left(u\right)\right), X_{1,n} < u < X_{n,n}\right\}$.

Un estimateur empirique de cette fonction est donné par :

$$e_n(u) = \frac{1}{N_u}\sum_{i=1}^{N_\mu}(X_i - u), \quad X_i > u, \quad 0 < u < +\infty$$

où $N_u$ : indique le nombre de dépassement par rapport à $u$.

Supposons données les observations $X_1, X_2, ..., X_n$ on trace graphiquement $\hat{e}_n(u)$ en fonction de $u$ et on choisit le plus petit $u$ de manière à ce que $\hat{e}_n(u)$ soit approximativement linéaire pour tout $x \geqslant u$.

$$e_n(\mu) = \frac{\hat{\beta} + \hat{\gamma}u}{1-\gamma}, \ \hat{\beta} + \hat{\gamma}u > 0.$$

La fonction moyenne des excès empiriques sous la transformation affine s'écrit pour $x \leqslant x_F$, comme suit :

Trois cas peuvent alors se présenter :

1. Si à certain seuil, la fonction moyenne des excès empirique est marquée par une pente positive, alors les données suivent une distribution de Preto généralisée avec un paramètre $\gamma$ positif.
2. Si la fonction moyenne des excès empirique présente une pente horizontale, les données suivent une distribution *Exponentielle*.
3. Si le graphe mean excess plot est marquée par une pente négative. Alors les données suivent une distribution à queue légère.

Le Figure 3.3 présente le graphe de la durée de vie moyenne résiduelle de taille 3000.

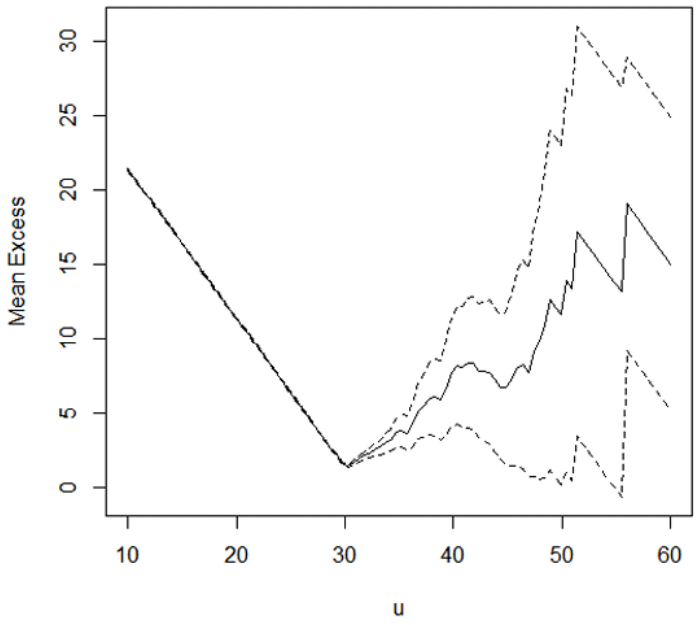


FIG. 3.3 – $MRL - plot$ pour $GPD$ considérés $n = 3000$ et $\gamma = 0.3$

Sur la Figure 3.3, on constate une linéarité entre 29 et 33. De ce fait, on peut affirmer que le seuil est compris entre 29 et 33.

**Estimateur de Hill**

Soit $X_1, X_2, ..., X_n$ suite d'observations $iid$, de fonction de répartition $F$. L'estimateur de $Hill$ utilise plus pour les distributions appartenant au domaine d'attraction de $Fréchet$, il est reformulé par l'expression suivante :

$$\gamma_{k,n}^{Hill} = \frac{1}{k}\sum_{i=1}^{k} \log\left(X_{n-i+1,n}\right) - \log\left(X_{n-k,n}\right), \text{ pour } k < n,$$

avec $k$, l'ordre statistique le plus élevé (le nombre des excès) et $\alpha = \frac{1}{\gamma}$ est l'indice de la queue de distribution.

Cet estimateur intervient dans la construction du graphique.

$Hill - plot$ : Représentation de $\alpha$ en fonction de la statistique d'ordre $X_{1,n}$. Le $Hill - plot$ nous permet de choisir un seuil élevé pour la construction d'un modèle $GPD$.

Le $Hill - plot$.est donc un outil à double utilité :

L'estimation de l'indice de la queue de la distribution, et l'estimation du seuil.

Le graphique $Hill - plot$, nous permet d'avoir des estimations du paramètre en fonction de l'ordre statistique le plus élevé (nombre des excès) , nous choisissons ainsi l'indice le plus stable.

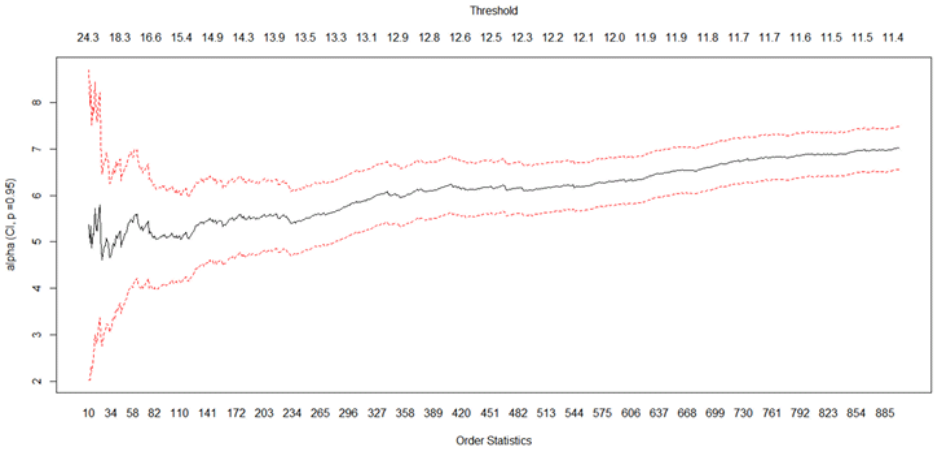


FIG. 3.4 – $Hill - plot$ pour $GPD$ $(n = 3000, u = 10, \beta = 1, \gamma = 0.3)$

Nous remarquons une zone de stabilité entre 141 et 234 excès. Au delà de 234 excès, l'estimateur n'est plus du tout stable.

Nous considèrerons donc que l'adéquation à une $GPD$ débute au niveau du 234 ème excès, soit un seuil à $[13.9 - 15.0]$

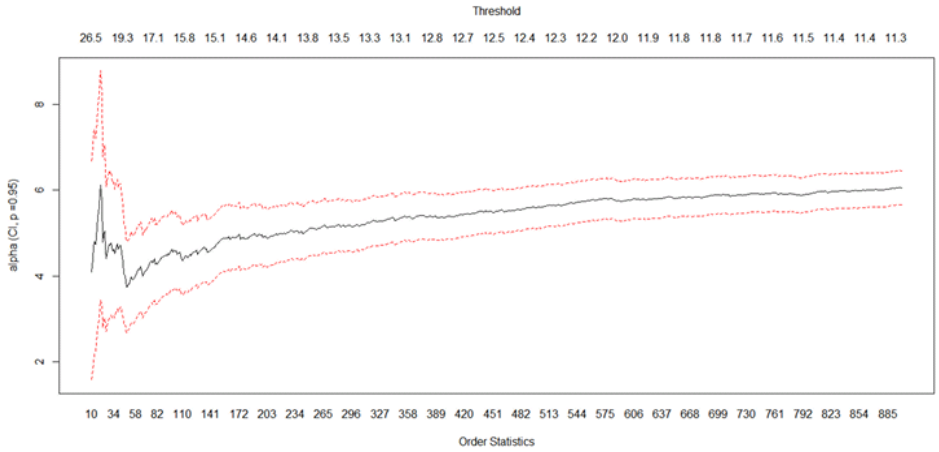


FIG. 3.5 – $Hill - plot$ pour $GEV$ $(n = 3000, u = 10, \beta = 1, \gamma = 0.3)$

Nous remarquons une zone de stabilité entre 172 et 220 excès. Au delà de 220 excès, l'estimateur n'est plus du tout stable. Nous considèrerons donc que l'adéquation à une $GEV$ débute au niveau du 220 ème excès, soit un seuil à $[13.8 - 15.1]$.

# Conclusion

Nous avons dans notre travail commencés par présenter les valeurs extrêmes et leurs distributions, ainsi que les différents domaines d'attractions, ($Fréchet, Weibull$ et de $Gumbel$) ; leurs caractéristique et propriétés particulières.

Un intérêt particulier pour les distributions des excès, et enfin les différentes façons de sélectionner le seuil $u$. Nous avons comme conclusion particulière fait le constat sur l'approche $POT$ ($Peaks\ Over\ Trecholds$.) qui est largement utilisée, mais avec un défaut principal : sa sensibilité au choix du seuil pour obtenir une bonne approximation des excès au delà d'un seuil considéré.

# Bibliographie


[1] Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (2008). A first course in order statistics. Society for Industrial and Applied Mathematics.

[2] Baldwin, J., Eddleman, C. D., Giblin-Davis, R. M., Williams, D., Vida, J., & Thomas, W. (1997). The buccal capsule of Aduncospiculum halicti (Nemata : Diplogasterina) : an ultrastructural and molecular phylogenetic study. Canadian Journal of Zoology, 75(3), 407-423..

[3] Bateka, S (2010) memoire de magister.Determination du nombre de statistiques D'ordre Extremes.

[4] Beirlant, J., Vynckier, P., & Teugels, J. L. (1996). Tail index estimation, Pareto quantile plots regression diagnostics. Journal of the AmeGnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. Annals of mathematics, 423-453.rican statistical Association, 91(436), 1659-1667.

[5] Cabanal-Duvillard, T., & Ionescu, V. (1997). Un théoreme central limite pour des variables aléatoires non-commutatives. Comptes Rendus de l'Académie des Sciences-Series I-Mathematics, 325(10), 1117-1120.

[6] Caeiro, F., et Gomes, MI (2016). Sélection de seuil dans l'analyse des valeurs extrêmes. Modélisation des valeurs extrêmes et analyse des risques : méthodes et applications , 1 , 69-86.

[7] Caers, J., Beirlant, J., et Maes, MA (1999). Statistiques pour la modélisation des distributions à queue lourde en géologie : Partie I. Méthodologie. Géologie mathématique , 31 (4), 391-410.

[8] Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. Journal of the Royal Statistical Society : Series B (Methodological), 52(3), 393-425

[9] De Haan, L., & Ferreira, A. (2007). Extreme value theory : an introduction. Springer Science & Business Media.

[10] Galambos, J. (1978). The asymptotic theory of extreme order statistics (No. 04 ; QA274, G3.).

[11] Garrido, M. (2002). Modélisation des événements rares et estimation des quantiles extrêmes, méthodes de sélection de modeles pour les queues de distribution (Doctoral dissertation, Université Joseph-Fourier-Grenoble I).

[12] Garrido, Myriam, and Zaher Khraibani. "Modélisation statistique des évènements rares et en particulier des valeurs extrêmes". Colloque émergences (2006). INRA editions, (2006).

[13] H. D. David, (1970), ordre statistics.John Wiley & Sons, Inc. ,New York-London-Sydney.

[14] Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics, 27(3), 251-261.

[15] Kotz, S. ans Nadarajah S.(2000) Extreme value distributions. Theory and applications.

[16] Lacoume, J. L., Amblard, P. O., & Comon, P. (1997). Statistiques d'ordre supérieur pour le traitement du signal.

[17] Lévy, P. (1934). Sur les intégrales dont les éléments sont des variables aléatoires indépendantes. Annali della Scuola Normale Superiore di Pisa-Classe di Scienze, 3(3-4), 337-366.

[18] Pickands III, J. (1975). Statistical inference using extreme order statistics. the Annals of Statistics, 3(1), 119-131.

[19] Reiss, R. D., Thomas, M., & Reiss, R. D. (1997). Statistical analysis of extreme values (Vol. 2). Basel : Birkhäuser.

[20] Rietsch, T. (2013). Théorie des valeurs extrêmes et applications en environnement (Doctoral dissertation, Strasbourg).

[21] Rydman, M. (2018). Application of the Peaks-Over-Threshold Method on Insurance Data.

[22] Saporta, G. (2006). Probabilités, analyse des données et statistique. Editions Technip.

# Annexe B : Abréviations et Notations

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

| | | |
|---|---|---|
| $F$ | : | fonction de répartition. |
| $\bar{F}$ | : | fonction de survie. |
| $F_n$ | : | fonction.de répartition |
| $f$ | : | densité de probabilité d'une variable aléatoire. |
| $f_{X_{i,n}}$ | : | fonction de densité de probabilité de $X_{i,n}$. |
| $f$ | : | fonction. |
| $v.a$ | : | variale aléatoire. |
| $v.a.r$ | : | variale aléatoire réelle. |
| $Q\,(_p)$ | : | quantile d'ordre $p$. |
| $u$ | : | seuil. |
| $N\,(0,1)$ | : | loi nermale standard. |
| $iid$ | : | indépendante et identiquement distribuée. |
| $x_F$ | : | point terminal. |
| $TCL$ | : | théorème centrale limite. |
| $\mu$ | : | espérance, ou moyenne d'une $v.a$ |
| $\sigma^2$ | : | variance d'un $v.a$ |
| $\Lambda_0$ | : | loi de $Gumbel$. |
| $\Phi_\gamma$ | : | loi de $Fréchet$. |
| $:=$ | : | égalité par définition. |
| $X_{1,n}, X_{2,n}, ..., X_{n,n}$ | : | statistique d'ordre associées à $(X_1, ..., X_n)$. |
| $(X_1, ..., X_n)$ | : | échantillons de taille $n$ de $v.a's$. |
| $\Psi_\gamma$ | : | loi de $weibull$. |
| $\mathbb{R}$ | : | ensemble des nombres réelles. |
| $GEV$ | : | distriburtion des valeurs extrêmes généralisée. |
| $F^{\leftarrow}$ | : | l'inverse généralisée de $F$. |
| $GPD$ | : | distribution de $Pareto$ généralisée. |

| | | |
|---|---|---|
| $\Omega$ | : | ensemble fondomentale. |
| $\mathcal{F}$ | : | tribu. |
| $w$ | : | evenement. |
| $P$ | : | probabilité. |
| $B_{\mathbb{R}}$ | : | tribu brolienne. |
| $P_B(A)$ | : | probabilité de $A$ sachant $B$. |
| $E[X]$ | : | espérance de $X$. |
| $V[X]$ | : | variance de $X$. |
| $\sigma$ | : | l'écart type. |
| $X_{i,n}$ | : | la $i^{iéme}$ statistique d'ordre. |
| $\min(X_1, ..., X_n)$ | : | minimum de $X_1, ..., X_n$. |
| $\max(X_1, ..., X_n)$ | : | maximum de $X_1, ..., X_n$. |
| $L$ | : | fonction à variation lente. |
| $\gamma$ | : | l'indice de queue. |
| $E[X^2]$ | : | moyenne d'ordre 2 d'une $v.a$. |
| $N_u$ | : | le nombre de dépassements du seuil $u$. |
| $F_u$ | : | la fonction de répartition des excès au-déla d' un seuil $u$ |
| $\begin{pmatrix} n \\ m \end{pmatrix}$ | : | la combinaison de $m$ objets parmi $n$ objets sans remise. |
| $I_A$ | : | fonction indicatrice de l'ensemble $A$. |
| $EVT$ | : | extrem value theory. |
| $\sup A$ | : | supremum de l'ensemble $A$. |
| $\inf A$ | : | infimum de l'ensemble $A$. |

| | | |
|---|---|---|
| $Q_n$ | : | quantile emprique. |
| $\bar{X}$ | : | moyenne emprique. |
| $\xrightarrow{P}$ | : | convergence en probabilité. |
| $\xrightarrow{P.s}$ | : | convergence en prèsque sûre. |
| $\xrightarrow{D}$ | : | convergence en distribution |
| $G_{\gamma,\beta}$ | : | la distribution des excès et la distribution de Pareto Généralisée. |
| $c-a-d$ | : | c'est -à-dire. |
| $S_n$ | : | somme arithmétique. |
| $i.e$ | : | c'est -à-dire. |
| $P.O.T$ | : | Peaks Over Treçholds. |
| $H\gamma$ | : | domaine d'attraction |