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

LARDJANI Mohamed

Titre :

Distribution des Statistiques d'Ordre et des Valeurs Extrêmes

Membres du Comité d'Examen :

Dr.	BENAMEUR Sana	UMKB	Encadreur
Dr.	YAHIA Djabrane	UMKB	Co-Encadreur
Dr.	CHERFAOUI Mouloud	UMKB	Président
Dr.	BENBRIKA Ghozlane	UMKB	Examinateur

Juin 2018

Dédicace

Je dedie ce modeste travail

A mes parents,

A mes frères,

A mes soeurs,

A mes amis,

A mes cousins et mes cousines.

A mes collegues de mathématiques 2017-2018.

REMERCIEMENTS

Je glorifie ALLAH le tout puissant de m'avoir donnée la patience, la santé et le courage pour terminer ce modeste travail.

Je remercie chaleureusement mes encadreurs Mme. Benameur Sana et Mr. Yahia Djabrane, pour tous le temps qu'ils m'ont consacrés durant cette année. Et qui grâce à leurs aide, remarques et conseils constructifs j'ai arrivé à atteindre convenablement les objectifs du travail. Mes remerciements vont également aux Membres de Jury : Mr. Cherfaoui Mouloud et Mme. Benbrika Ghozlane, pour l'honneur qu'ils m'ont fait en acceptant de juger mon mémoire.

Je remercie chaleureusement pour tous le temps qu'il nous a consacré durant cette mémoire. Et qui grâce à son aide, ses remarques et ses conseils. Je tiens également à remercie mes enseignants : Mr. Meraghni Mr. Necir, Mr. Benatia, et tous les enseignants du département de mathématiques à l'université de Biskra. Mes vifs remerciements s'adressent également à mes amis de l'université Mohamed Khider Biskra.

Enfin, je souhaite remercier mes parents, mes frères et sœurs.

Merci à Toutes et à tous.

Table des matières

Table des figures

Introduction

Depuis quelques années, on voit se développer des méthodes statistiques, utilisant des observations ordonnées par ordre croissant et appelés statistiques d'ordre. Ces derniers jouent un rôle très important dans la pratique comme l'astronomie l'économétrie et dans les sciences techniques,...

Les statistiques d'ordre sont très utiles en théorie des valeurs extrêmes, ils ont été développés pour l'estimation de probabilités d'occurrences d'évènement rares ou événement ayant une faible probabilité d'apparition. Ils sont dits extrêmes quand il s'agit de valeurs beaucoup plus grandes ou plus petites que celles observées habituellement.

Le mémoire et organisé comme suit :

Le premier chapitre est consacré à l'étude des distributions d'une statistique d'ordre dont je commence par présenter les définitions et les propriétés générales des statistiques d'ordre associées à un échantillons, ainsi que les lois d'une statistique d'ordre et la distribution jointes des statistiques d'ordre. Puis j'étudié quelques propriétés des espacements uniformes et exponentielles. Ensuite, je regroupe les différentes techniques de simulation permettent de générer des échantillons ordonnés.

Dans le deuxième chapitre, nous énonçons des caractérisations générales en présentan le résultat essentiel de la théorie des valeurs extrêmes, décrit les limites possible de la loi du maximum d'un échantillon. Ces lois sont indexées par un paramètre ζ appelé indice de valeurs extrêmes et selon son signe, nous distinguons trois domaines d'attraction :

Weibull $(\zeta < 0)$*, Fréchet* $(\zeta > 0)$ *et Gumbel* $(\zeta = 0)$*. Nous énonçons ensuite les conditions d'appartenance d'une fonction de répartition à l'un de ces trois domaines d'attraction.*

Je présente aussi dans ce dernier chapitre, la distribution des valeurs extrêmes généralisées (GEV)*, la distribution de Pareto généralisée* (GPD) *et le théorème de Balkema-de Haan et Pickands.*

Chapitre 1

Statistiques d'ordre

La statistique d'ordre fait partie des outils fondamentaux de la statistique non paramétrique et de l'inférence statistique. Dans ce chapitre, nous présentons les définitions et les propriétés générales des statistiques d'ordre, et nous réservons une partie sur la simulation des statistiques d'ordre.

De plus amples détails pour les propriétés générales des statistiques d'ordre peuvent être trouvés dans les références [1], [3].

1.1 Définitions des statistiques d'ordre

Soit $X_1, ..., X_n$ n variables aléatoires (va's) indépendantes et identiquement distribuées (iid) de densité commune f et de fonction de répartition $F(x) = P(X \leq x)$, pour $x \in \mathbb{R}$.

Définition 1.1.1 (Statistique d'ordre) *On appelle statistiques d'ordre notées $X_{1,n}, ..., X_{n,n}$ les variables aléatoires ordonnées en sens croissant :*

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

Deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des événements extrêmes, sont : la plus petite $X_{1,n}$ (ou statistique du minimum notée m_n) et la plus grande *statistique d'ordre $X_{n,n}$ (ou statistique du maximum notée M_n), telles que :*

$$m_n := X_{1,n} = \min(X_1, ..., X_n) \quad \textit{et} \quad M_n := X_{n,n} = \max(X_1, ..., X_n).$$

où pour $k = 1, ..., n$, la va $X_{k,n}$ appelée la $k^{ième}$ statistique d'ordre (ou statistique d'ordre du rang k).

1.2 Distribution d'une statistique d'ordre

Nous présentons dans cette partie la fonction de répartition et la densité de l'une des statistiques d'ordre suivantes : $X_{1,n}$, $X_{n,n}$ et $X_{k,n}$.

1.2.1 Distribution du minimum

Proposition 1.2.1 *Soit $X_1, ..., X_n$ n va's iid de densité commune f et de fonction de répartition F, alors la loi de la va $X_{1,n}$ est donnée par :*

$$F_{X_{1,n}}(x) = 1 - [1 - F(x)]^n, \quad \forall x \in \mathbb{R} \tag{1.1}$$

et sa densité est :

$$f_{X_{1,n}}(x) = nf(x)[1 - F(x)]^{n-1}. \tag{1.2}$$

Preuve. On a

$$\begin{aligned} F_{X_{1,n}}(x) &= P(X_{1,n} \leq x) = P(\min(X_1, ..., X_n) \leq x) \\ &= 1 - P(\min(X_1, ..., X_n) > x) \\ &= 1 - P\left[(X_1 > x) \cap ... \cap (X_n > x)\right] \end{aligned}$$

En utilisant le fait que les X_j sont *iid* :

$$F_{X_{1,n}}(x) = 1 - P\left[\bigcap_{j=1}^{n}(X_j > x)\right] = 1 - \prod_{j=1}^{n} P\left[(X_j > x)\right]$$
$$= 1 - \prod_{j=1}^{n}\left[1 - P\left(X_j \leq x\right)\right] = 1 - \prod_{j=1}^{n}[1 - F(x)]$$
$$= 1 - [1 - F(x)]^n, \quad \forall x \in \mathbb{R}.$$

En dérivant $F_{X_{1,n}}(x)$, en déduit la densité :

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

Nous obtenons le résultat. ■

1.2.2 Distribution du maximum

Proposition 1.2.2 *Soit $X_1, ..., X_n$ n va's iid de densité commune f et de fonction de répartition F, alors la loi de la va $X_{n,n}$ est donnée par :*

$$F_{X_{n,n}}(x) = [F(x)]^n, \quad \forall x \in \mathbb{R} \tag{1.3}$$

et sa densité est :

$$f_{X_{n,n}}(x) = nf(x)[F(x)]^{n-1}. \tag{1.4}$$

Preuve. Il est clair que,

$$\begin{aligned}F_{X_{n,n}}(x) &= P(X_{n,n} \leq x) = P(\max(X_1, ..., X_n) \leq x) \\ &= P\left[(X_1 \leq x) \cap ... \cap (X_n \leq x)\right] \\ &= P\left[\bigcap_{j=1}^{n} (X_j \leq x)\right] = \prod_{j=1}^{n} P[(X_j \leq x)] \\ &= \prod_{j=1}^{n} F(x) = [F(x)]^n, \quad \forall x \in \mathbb{R}.\end{aligned}$$

En dérivant $F_{X_{n,n}}(x)$, nous obtenons la forme de la densité comme suit ;

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

■

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

Proposition 1.2.3 *Soit $X_1, ..., X_n$ n va's iid de densité commune f et de fonction de répartition F, alors la loi de la va $X_{k,n}$ pour $k = 1, ..., n$, est donnée par :*

$$F_{X_{k,n}}(x) = \sum_{j=k}^{n} C_n^j [F(x)]^j [1 - F(x)]^{n-j}, \quad \forall x \in \mathbb{R} \tag{1.5}$$

où

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

et sa densité est :

$$f_{X_{k,n}}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x). \tag{1.6}$$

Preuve. On a,

$$\begin{aligned}
F_{X_{k,n}}(x) &= P\left(X_{k,n} \leq x\right) \\
&= P\left\{\text{au moins } k \text{ des } X_j \text{ sont inférieur à } x\right\} \\
&= \sum_{j=k}^{n} P\left\{\text{exactement } j \text{ de } X_1, ..., X_n \text{ sont inférieur à } x\right\},
\end{aligned}$$

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

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

Soit $X_{k,n}$, avec $1 \leq k \leq n$ la $k^{ième}$ statistique d'ordre, on peut représenté l'événement $\{x < X_{k,n} \leq x + \delta x\}$ comme suit :

$$\begin{array}{cccc}
\overset{k-1}{\overline{\qquad\qquad\qquad}} & \overset{1}{]\!\!-\!\!-} & \overset{n-k}{]\,\overline{\qquad\qquad\qquad}} & ; \\
-\infty & x & x+\delta x \qquad\qquad +\infty &
\end{array}$$

$X_s \leq x$ pour $k-1$ des X_s, $x < X_s \leq x + \delta x$ pour l'un des X_s, est $X_s > x + \delta x$ pour $n-k$ restes des X_s.

Considérons δx assez petit, on peut écrire

$$\begin{aligned}
P\left(x < X_{k,n} \leq x + \delta x\right) = \frac{n!}{(k-1)!1!(n-k)!} & [F(x)]^{k-1} \left[F(x+\delta x) - F(x)\right] \\
& [1 - F(x+\delta x)]^{n-k} + \bigcirc\left((\delta x)^2\right),
\end{aligned} \tag{1.7}$$

où $\bigcirc\left((\delta x)^2\right)$ est une limite d'ordre $(\delta x)^2$ qui correspondant à la probabilité de l'événement d'avoir plus d'un X_s dans l'intervalle $(x, x + \delta x]$, alors de l'équation (1.7) on peut calculer la densité de $X_{k,n}$ pour $k = 1, ..., n$, comme suit :

$$\begin{aligned} f_{X_{k,n}}(x) &= \lim_{\delta x \to 0} \frac{1}{\delta x} \left[P\left(x < X_{k,n} \leq x + \delta x \right) \right] \\ &= \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} f(x) [1 - F(x)]^{n-k}, 1 \leq k \leq n. \end{aligned}$$

(voir Arnold et al.,1992 [1]). En particulier pour $k = 1$ et $k = n$, on obtient la densité du minimum et maximum comme suit :

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

Ce qui donne la preuve. ■

Exemple 1.2.1 *Soit $(U_1, ..., U_n)$ un échantillon suit la loi uniforme sur $[0, 1]$, de fonction de répartition $F(t) = t$, $t \in [0, 1]$, alors la densité de la va $U_{i,n}$ pour $i = 1, ..., n$, est donnée par :*

$$f_{U_{i,n}}(t) = \frac{n!}{(i-1)!(n-i)!} t^{i-1} \left(1 - t\right)^{n-i} 1\mathrm{I}_{\{0 \leq t \leq 1\}}.$$

1.3 Distribution jointe des statistiques d'ordre

1.3.1 Distribution jointe d'un couple de statistique d'ordre

Proposition 1.3.1 *Soit $X_1, ..., X_n$ n va's iid de fonction de répartition F et densité f, alors la fonction de répartition d'un couple de statistique d'ordre $(X_{r,n}, X_{s,n})$ pour $1 \leq r < s \leq n$ est définit par : $\forall x, y \in \mathbb{R}$*

$$F_{(X_{rn}, X_{s,n})}(x, y) = P(X_{r,n} \leq x, X_{s,n} \leq y)$$

On a deux cas :

1**[er] **cas : $x < y$

$$F_{(X_{r,n},X_{s,n})}(x,y) = \sum_{v=s}^{n}\sum_{u=r}^{v}\frac{n!}{u!(v-u)!(n-v)!}[F(x)]^{u}$$
$$[F(y)-F(x)]^{v-u}[1-F(y)]^{n-v}, \tag{1.8}$$

2**[ième] **cas : $x \geq y$

$$F_{(X_{r,n},X_{s,n})}(x,y) = F_{X_{s,n}}(y), \tag{1.9}$$

et la densité jointe d'un couple de statistique d'ordre $(X_{r,n}, X_{s,n})$ *pour* $1 \leq r < s \leq n$ *est donnée par :*

$$f_{(X_{r,n},X_{s,n})}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}F(x)]^{r-1}f(x)f(y)$$
$$[F(y)-F(x)]^{s-r-1}[1-F(y)]^{n-s}, \tag{1.10}$$

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

Preuve. Considérons le **cas ou** $x < y$:

$$\begin{aligned}
F_{(X_{r,n},X_{s,n})}(x,y) &= P(X_{r,n} \leq x, X_{s,n} \leq y) \\
&= P\{\text{au moins } r \text{ de } X_1,...,X_n \text{ sont inférieur à } x \text{ et au moins } s \text{ de } X_1,...,X_n \\
&\qquad \text{sont inférieur à } y\} \\
&= \sum_{v=s}^{n}\sum_{u=r}^{v} P\{\text{exactement } u \text{ de } X_1,...,X_n \text{ sont inférieur à } x \text{ et exactement } v \text{ de} \\
&\qquad\qquad X_1,...,X_n \text{ sont inférieur à } y\} \\
&= \sum_{v=s}^{n}\sum_{u=r}^{v}\frac{n!}{u!(v-u)!(n-v)!}[F(x)]^{u}[F(y)-F(x)]^{v-u}[1-F(y)]^{n-v}.
\end{aligned}$$

De même pour le **cas ou** $x \geq y$:

$$\begin{aligned} F_{(X_{rn},X_{s,n})}(x,y) &= P(X_{r,n} \leq x, X_{s,n} \leq y) \\ &= P\left(X_{s,n} \leq y\right) = F_{X_{s,n}}(y). \end{aligned}$$

Concernant la densité jointe du couple $(X_{r,n}, X_{s,n})$ pour $1 \leq r < s \leq n$, on peut représenté l'événement $\{x < X_{r,n} \leq x + \delta x, y < X_{s,n} \leq y + \delta y\}$ comme suit :

$$\underset{-\infty}{\overset{r-1}{\rule{6em}{0.4pt}}} \quad \underset{x}{]}\overset{1}{\rule{1.5em}{0.4pt}} \quad \underset{x+\delta x}{]}\overset{s-r-1}{\rule{6em}{0.4pt}} \quad \underset{y}{]}\overset{1}{\rule{1.5em}{0.4pt}} \quad \underset{y+\delta y}{]}\overset{n-s}{\underset{+\infty}{\rule{6em}{0.4pt}}} \quad ;$$

$X_j \leq x$ pour $r-1$ des X_j, $x < X_j \leq x + \delta x$ pour l'un des X_j, $x + \delta x < X_j \leq y$ pour $s - r - 1$ des X_j, $y < X_j \leq y + \delta y$ pour l'un des X_j et $X_j > y + \delta y$ pour $n - s$ restes des X_j.

Considérons δx et δy assez petits, on peut écrire

$$\begin{aligned} &P(x < X_{r,n} \leq x + \delta x, y < X_{s,n} \leq y + \delta y) \\ &\quad = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \left[F(x)\right]^{r-1} \left[F(x+\delta x) - F(x)\right] \left[F(y) - F(x+\delta x)\right]^{s-r-1} \\ &\quad \times \left\{\left[F(y+\delta y) - F(y)\right]\left[1 - F(y+\delta y)\right]^{n-s}\right\} + \bigcirc\left((\delta x)^2 \delta y\right) + \bigcirc\left(\delta x \left(\delta y\right)^2\right), \qquad (1.11) \end{aligned}$$

où $\bigcirc\left((\delta x)^2 \delta y\right)$ et $\bigcirc\left(\delta x \left(\delta y\right)^2\right)$ sont des limites d'ordre plus supérieur, qui correspondant aux probabilités de l'événement d'avoir plus d'un X_j dans l'intervalle $(x, x + \delta x]$ et un X_j dans l'intervalle $(y, y + \delta y]$, et de l'événement d'avoir un X_j dans l'intervalle $(x, x + \delta x]$ et plus d'un X_j dans l'intervalle $(y, y + \delta y]$ respectivement. Alors de l'équation (1.11) on peut calculer la densité jointe d'un couple de statistique d'ordre $(X_{r,n}, X_{s,n})$ pour $1 \leq r < s \leq n$, comme suit :

$$f_{(X_{r,n},X_{s,n})}(x,y) = \lim_{\delta x,\delta y \longrightarrow 0} \frac{1}{\delta x \delta y} \{P(x < X_{r,n} \leq x + \delta x, y < X_{s,n} \leq y + \delta y)\}$$
$$= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F(x)]^{r-1} f(x)[F(y) - F(x)]^{s-r-1} f(y)[1 - F(y)]^{n-s},$$

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

Pour plus de détaille, voir Arnold et al., (1992) [1]. ■

Remarque 1.3.1 (Cas particulier) *Pour $r = 1$ et $s = n$, on obtient la densité jointe du maximum et minimum, comme suit :*

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

Remarque 1.3.2 *Les variables aléatoires $X_{1,n}, ..., X_{n,n}$ ne sont pas indépendantes et identiquement distribuées.*

1.3.2 Densité de n statistiques d'ordre

Proposition 1.3.2 *Soit $X_1, ..., X_n$ n va's iid de densité commune f. Alors la densité conjoint de $(X_{1,n}, ..., X_{n,n})$ est donnée par :*

$$f_{(X_{1,n},...,X_{n,n})}(x_{1,...,}x_n) = n! \prod_{j=1}^{n} f_{(x_j)}, \quad -\infty < x_1 < ... < x_n < \infty. \tag{1.12}$$

Preuve. On a,

$$P(x_1 \leq X_{1,n} < x_1 + \delta x_1, ..., x_n \leq X_{n,n} < x_n + \delta x_n)$$
$$= n! P(x_1 \leq X_{1,n} < x_1 + \delta x_1) ... P(x_n \leq X_{n,n} < x_n + \delta x_n)$$

on divisant par $\delta x_1 ... \delta x_n$ on obtient :

$$\begin{aligned}
&\frac{P(x_1 \leq X_{1,n} < x_1 + \delta x_1, ..., x_n \leq X_{n,n} < x_n + \delta x_n)}{\delta x_1 ... \delta x_n} \\
&= n! \frac{P(x_1 \leq X_{1,n} < x_1 + \delta x_1)}{\delta x_1} ... \frac{P(x_n \leq X_{n,n} < x_n + \delta x_n)}{\delta x_n}
\end{aligned}$$

alors,

$$f_{(X_{1,n},...,X_{n,n})}(x_{1,...,}x_n) = n! \prod_{j=1}^{n} f_{(x_j)}, \quad -\infty < x_1 < ... < x_n < \infty.$$

Cette preuve est en détaille dans David et Nagaraja (2003) [3]. ■

1.3.3 Densité conditionnelle de statistique d'ordre

Proposition 1.3.3 *Soit $X_1, ..., X_n$ n va's iid de densité commune f. Alors pour $1 \leq r < s \leq n$, la densité conditionnelle $f_{(X_{r,n}/X_{s,n})}$ de $X_{r,n}$ sachant l'évènement $\{X_{s,n} = y\}$ est donnée par :*

$$f_{(X_{r,n}/X_{s,n})}(x/y) = \frac{(s-1)!}{(r-1)!(s-r-1)!} [F(x)]^{r-1} f(x) [F(y) - F(x)]^{s-r-1} [F(y)]^{1-s}, \quad (1.13)$$

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

Preuve. On rappelle que

$$f_{(X_{r,n}/X_{s,n})}(x/y) = \frac{\left(f_{(X_{r,n},X_{s,n})}(x,y)\right)}{f_{X_{s,n}}(y)},$$

alors pour $1 \leq r < s \leq n$, on a :

$$\begin{aligned}
f_{(X_{r,n}/X_{s,n})}(x/y) &= \frac{n!}{(r-1)!(s-r-1)!(n-s)!} F(x)]^{r-1} f(x) [F(y) - F(x)]^{s-r-1} f(y) [1 - F(y)]^{n-s} \\
&\times \frac{(s-1)!(n-s)!}{n!} [F(y)]^{1-s} [f(y)]^{-1} [1 - F(y)]^{s-n} \\
&= \frac{(s-1)!}{(r-1)!(s-r-1)!} [F(x)]^{r-1} f(x) [F(y) - F(x)]^{s-r-1} [F(y)]^{1-s},
\end{aligned}$$

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

Pour plus de détails sur les statistiques d'ordre voir [1] et [3]. ■

1.4 Estimation des quantiles

Définition 1.4.1 (Fonction de répartition empirique) *Soit $X_1, ..., X_n$ échantillon de n va's iid de fonction de répartition F, alors la fonction de répartition empirique est donnée par :*

$$F_n(x) := \frac{1}{n} \sum_{j=1}^{n} 1\mathrm{I}_{\{X_j \leq x\}}, \quad -\infty < x < \infty. \tag{1.14}$$

$1\mathrm{I}_{\{X_j \leq x\}}$ *: c'est la fonction indicatrice, tell que*

$$1\mathrm{I}_{\{X_j \leq x\}} := \begin{cases} 1 \quad si \qquad X_j \leq x, \\ 0 \quad sinon \quad . \end{cases}$$

Définition 1.4.2 (Fonction des quantiles) *L'inverse généralisé de la fonction de répartition appelée la fonction des quantiles (notée Q) telle que :*

$$Q(t) = F^{-1}(t) := \inf\{s : F(s) \geq t\},\ 0 < t < 1. \tag{1.15}$$

Définition 1.4.3 (Fonction des quantiles empirique) *La fonction des quantiles empirique : c'est l'inverse généralisé de la fonction de répartition empirique F_n (notée Q_n), pour $n \geq 1$ ona :*

$$Q_n(t) = F_n^{-1}(t) := \inf\{s : F_n(s) \geq t\},\ 0 < t < 1. \tag{1.16}$$

On peut représenté la fonction des quantiles empirique en terme des statistiques d'ordre *comme suit :*

$$Q_n(t) := \begin{cases} X_{j,n} & si \quad \frac{j-1}{n} < t \leq \frac{j}{n}, \\ X_{n,n} & si \quad 0 < t < 1. \end{cases} \tag{1.17}$$

Définition 1.4.4 (Quantile d'ordre p) *Soit X une variable aléatoire de loi de probabilité F. On définit le quantile d'ordre p ($p \in]0,1[$) par :*

$$x_p := F^{-1}(p),\ 0 < p < 1. \tag{1.18}$$

Proposition 1.4.1 *Soit $p \in]0,1[$. Supposons que F est continue et qu'il existe une seule solution x_p à l'équation $F(x) = p$. Soit $(k(n), n \geq 1)$ une suite d'entiers telle que*

$$1 \leq k(n) \leq n \quad et \quad k(n)/n \to p \qquad quand \quad n \to \infty.$$

Alors la suite des quantiles empiriques $\left(X_{k(n),n}\right)_{n \geq 1}$ converge presque sûrement vers x_p.

Proposition 1.4.2 *Soit $p \in]0,1[$ supposons que la loi de X_1 possède une densité f continue en x_p et telle que $f(x_p) > 0$. On suppose de plus que $k(n) = np + \circ(\sqrt{n})$. On a la convergence en loi suivante*

$$\sqrt{n}\left(X_{k(n),n} - x_p\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f^2(x_p)}\right) \ quand\ n \to \infty. \tag{1.19}$$

où $\mathcal{N}(0,a)$ designe la loi normale centré de variance a.

1.5 Espacements

1.5.1 Propriétés des espacements uniformes

Soit $U_1, ..., U_n$ une suite de va's *iid* uniformes sur $[0, 1]$, avec les statistiques d'ordre associées $U_{1,n} \leq U_{2,n} \leq ..., U_{n,n}$. Les statistiques S_i définies par

$$S_i = U_{i,n} - U_{i-1,n}, \ 1 \leq i \leq n+1, \tag{1.20}$$

où par convention $U_{0,n} = 0$, $U_{n+1,n} = 1$, s'appellent les espacements uniformes pour cet échantillon.

Théorème 1.5.1 *Les espacements $S_1, ..., S_n$ sont uniformément distribuées sur le simplex*

$$A_n := \left\{ (x_1, ..., x_n) : x_i \geq 0, \sum_{i=1}^{n} x_i \leq 1 \right\}. \tag{1.21}$$

Théorème 1.5.2 (Pyke, 1965, 1972) *Soit $E_1, ..., E_{n+1}$ une suite des va's iid exponentielle de paramètre 1.*

$$\{S_1, ..., S_{n+1}\} \stackrel{d}{=} \left\{ \frac{E_1}{\sum_{i=1}^{n+1} E_i}, ..., \frac{E_{n+1}}{\sum_{i=1}^{n+1} E_i} \right\}. \tag{1.22}$$

Preuve. Voir Devroye (1986) [4, Chapitre 5]. ∎

1.5.2 Propriétés des espacements exponentiels

Théorème 1.5.3 (Sukhatme,1937) *Soient $E_{1,n} \leq ... \leq E_{n,n}$ les statistiques d'ordre correspondantes à une suite des va's exponentielles iid $E_1, ..., E_n$. Si $E_{0,n} = 0$, alors les espacements exponentiels*

$$(n - i + 1) (E_{i,n} - E_{i-1,n}), 1 \leq i \leq n, \tag{1.23}$$

sont des va's iid de distribution exponentielle standard.

Corollaire 1.5.1 (Malmquist, 1950) *Soient* $0 \leq U_{1,n} \leq ... \leq U_{n,n} \leq 1$ *les statistiques d'ordre associées à* $U_1, ..., U_n$ *; une suite des va's iid uniformément distribuées sur* $[0,1]$*, si* $U_{n+1,n} = 1$*, alors*

$$\left\{ \left(\frac{U_{i,n}}{U_{i+1,n}} \right)^i, 1 \leq i \leq n \right\} \stackrel{d}{=} \{U_1, U_2, ..., U_n\}. \tag{1.24}$$

Preuve. Voir Devroye (1986) [4, Chapitre 5]. ■

Proposition 1.5.1 (Transformation des quantiles) *Soit* $U_1, ..., U_n$ *une suite de va's uniformes sur* $[0,1]$*, avec les statistiques d'ordre associées* $U_{1,n} \leq U_{2,n} \leq ..., U_{n,n}$.

(a) *Pour toute fonction de répartition* F*, nous avons*

$$X_{k,n} \stackrel{d}{=} F^{-1}(U_{k,n}), \; k = 1, ..., n. \tag{1.25}$$

(b) *Si* F *est continue, nous avons*

$$F\left(X_{k,n}\right) \stackrel{d}{=} U_{k,n}, \; k = 1, ..., n. \tag{1.26}$$

1.6 Simulation des statistiques d'ordre

Dans cette partie, nous allons utiliser trois méthodes à l'aide du logiciel **R**, pour générer un échantillon ordonné. Soit $(U_1, ..., U_n)$ un échantillon suit la loi uniforme sur $[0,1]$, de fonction de répartition $F(x) = x$, $x \in [0,1]$.

1.6.1 Méthode de la fonction de répartition empirique

On peut représenté la fonction de répartition empirique en terme des statistiques d'ordre comme suit :

$$F_n(x) := \frac{1}{n}\sum_{j=1}^{n} 1\mathrm{I}_{\{X_{j,n} \leq x\}} = \begin{cases} 0 & si \quad x < X_{1,n}, \\ \frac{j-1}{n} & si \quad X_{j-1,n} \leq x < X_{j,n},\ 2 \leq j \leq n, \\ 1 & si \quad x > X_{n,n}. \end{cases} \tag{1.27}$$

Par conséquent les statistiques d'ordre sont la solution de cette équation

$$F_n(x) - \frac{j}{n} = 0.$$

Le code suivant permet de résoudre cette équation, afin de simuler un échantillon ordonné issue d'un échantillon uniforme standard.

$N = 8$

$XO = numeric\,(N)$

$U = runif\,(N)$

$FN = function\,(x, U)\,\{$

$Y = numeric\,(N)$

$for\,(j\ in\ 1 : N)\,\{$

$if\,(U\,[j] <= x)\,Y\,[j] = 1$

$else$

$Y\,[j] = 0\}$

$mean\,(Y)\}$

$for\,(j\ in\ 1 : N)\,\{$

$H = function\,(x)\,\{FN\}\,(x, U) - (j/N)\}$

$X = uniroot\,(H,\ lower = 0,\ upper = 1)$

$XO\,[j] = X\$root\}.$

1.6.2 Méthode directe

Soit $(U_1, ..., U_n)$ un échantillon suit la loi uniforme sur $[0, 1]$ et $(U_{1,n}, ..., U_{n,n})$ la statistique d'ordre associée.

Le code suivant nous donne une suite ordonnée de taille $N = 100$.

$N = 100$

$U = runif(N)$

$UO = sort(U)$

1.6.3 Représentation de Reny

Soit $(U_1, ..., U_n)$ un échantillon suit la loi uniforme sur $[0, 1]$ et $(U_{1,n}, ..., U_{n,n})$ la statistique d'ordre associée, et soit $E_1, ..., E_{n+1}$ un échantillon de taille $n + 1$ suit la loi exponentielle de paramètre 1, alors :

$$U_{i,n} \stackrel{d}{=} \frac{E_1 + ... + E_i}{E_1 + ... + E_{n+1}}, \; i = 1, ..., n. \tag{1.28}$$

$N = 100$

$U = runif(N)$

$UO = sort(U)$

$US = numeric(N)$

$E = r\exp(N + 1)$

$S = 0$

$for\,(i\ in\ 1 : N)\,\{$

$S = S + E\,[i]$

$US\,[i] = S/sum\,(E)\}$

$qqplot\,(UO, US)$

$ks.test\,(UO, US)\,.$

L'exécution de ces commandes nous donne FiG 1.1, cette figure montre que le quantile associée à l'échantillon ordonné par la représentation de Reny et le quantile relative à l'échantillon ordonné par la méthode directe sont alignés, se qui preuve l'égalité (1.28).

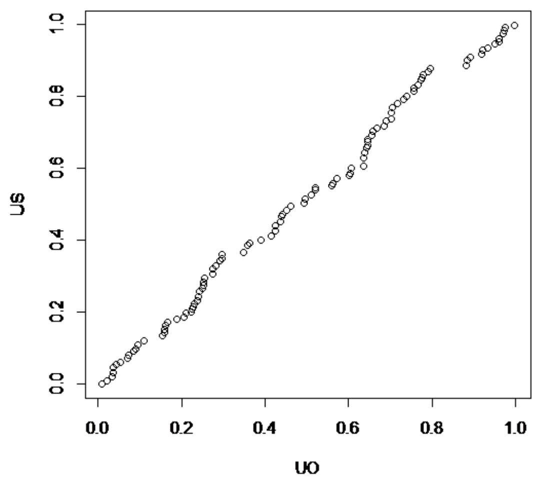

FIG. 1.1 – Comparaison des statistiques d'ordre entre la méthode directe et représentation de Reny

Nous avons appliqués le test de Kolmogorov-Smirnov sur les deux échantillons, en raison de la plus grande valeur de "$P - value = 0.69$", les deux distributions sont identiques à n'importe quel niveau de signification standard.

Chapitre 2

Théorie des valeurs extrêmes

La théorie des valeurs extrêmes (EVT, en anglais Extreme Value Theory) a été développée pour l'estimation des probabilités d'occurrences d'événement rares. Par définition, les événements rares sont des événements ayant une faible probabilité d'apparition.

Dans ce chapitre nous énonçons théorèmes limites et nous parlons sur des distributions limites des extrêmes et leurs domaines d'attraction. Cependant, on y retrouve deux modèles : loi des valeurs extrêmes généralisée et loi de Pareto généralisée. De plus amples détails pour le théorie des valeurs extrêmes peuvent être trouvés dans les références : de Haan, Ferreria (2006) [5], Embrechts et al. [6] et Reiss, Thomas (1997) [14].

On a la relation entre la statistique du minimum (min) et celle du maximum (max) comme suit :

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

Les distributions asymptotiques du minimum et du maximum quand n tendre vers l'infini sont :

$$\lim_{n\to\infty} F_{X_{1,n}}(x) = \lim_{n\to\infty} 1 - [1 - F(x)]^n = \begin{cases} 0 & si \quad F(x) = 0 \\ 1 & si \quad F(x) > 0 \end{cases}$$

et

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

2.1 Caractérisations générales

Dans cette partie, nous présentons quelques définitions et rappels.

Définition 2.1.1 (Fonction de survie) *Soit X est une variable aléatoire définie sur un espace de probabilité $(\Omega, \mathcal{F}, \mathcal{P})$, avec de fonction de répartition F $\{F(x) = P(X \leq x), \forall x \in \mathbb{R}\}$ alors la fonction de survie est donnée comme suit :*

$$\bar{F}(x) = 1 - F(x) = P(X > x), \forall x \in \mathbb{R}. \tag{2.1}$$

Définition 2.1.2 (Fonction quantile de queue) *On définit la fonction quantile de queue (notée U) par :*

$$U(t) = F^{-1}\left(1 - \frac{1}{t}\right),\ 0 < t < \infty. \tag{2.2}$$

Définition 2.1.3 (Fonction quantile de queue empirique) *On définit la fonction quantile de queue empirique (notée U_n), pour $n \geq 1$ par :*

$$U_n(t) = F_n^{-1}\left(1 - \frac{1}{t}\right),\ 0 < t < \infty. \tag{2.3}$$

Théorème 2.1.1 (Glifenko-Contelli, 1933) *Soit $(X_n, n \in \mathbb{N})$ une suite des va's réelles indépendantes et de même loi F, avec la fonction de répartition empirique F_n, alors :*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0 \quad p.s. \tag{2.4}$$

2.2 Théorèmes limites

Dans cette partie, nous énonçons deux théorèmes très importants en statistique : la lois des grandes nombres (L.G.N) et le théorème central limite (T.C.L).

2.2.1 Lois des grandes nombres

Les lois des grands nombres (L.G.N) indiquent que l'on fait un tirage aléatoire dans une série de grandes tailles. Elles sont de deux types : lois faibles mettant en jeu la convergence en probabilité P et lois fortes relatives à la convergence presque sûre $p.s$.

Théorème 2.2.1 (L.G.N) *Si $(X_1, ..., X_n)$ est une suite d'une va X tel que $\mu = E(X) < \infty$, alors :*

$$\begin{aligned} \textit{La loi faible} \quad &: \; \bar{X}_n \xrightarrow{P} \mu \quad \textit{quand } n \to \infty, \\ \textit{La loi forte} \quad &: \; \bar{X}_n \xrightarrow{Ps} \mu \quad \textit{quand } n \to \infty. \end{aligned}$$

2.2.2 Théorème central limite

Le théorème central limite (T.C.L) établit, sous des hypothèses très peu contraignantes, la convergence en loi d'une somme de va's *iid* vers la loi normale.

Théorème 2.2.2 (T.C.L) *Soit $X_1, ..., X_n$ est une suite des va's iid de moyenne μ $(E(X) = \mu)$ et de variance σ^2 finie. Considérons la somme $S_n := X_1 + X_2 + ... + X_n$, alors l'espérence de S_n est $n\mu$ et son écart-type est $\sigma / \sqrt{n}$, donc*

$$\frac{\sqrt{n}\,(S_n - n\mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1) \qquad \textit{quand } n \to \infty.$$

Remarque 2.2.1 *Pour plus de détails sur les théorèmes L.G.N et T.C.L en peut cité le livre Saporta [15].*

2.3 Loi des valeurs extrêmes

Le théorème ci-dessous est fondamental en théorie des valeurs extrêmes.

Théorème 2.3.1 (Fisher-Tippet 1928**, Gnedenko** 1943**)** *Soit* $(X_i, i \geq 1)$ *une suite des va's iid et* $M_n = \max(X_1, ..., X_n)$*, s'il existe deux suites normalisantes réelles* $\{c_n > 0, n \geq 1\}$ *et* $\{d_n \in \mathbb{R}, n \geq 1\}$ *et une distribution limite non dégénérée* H *telles que :*

$$\lim_{n\to\infty} P\left(c_n^{-1}(M_n - d_n) \leq x\right) = H(x),\ \ x \in \mathbb{R}, \tag{2.5}$$

alors H *appartient à l'un des trois types des distributions suivants :*

1) Distribution de Weibull :

$$\Psi_\alpha(x) = \begin{cases} 1 & si \quad x > 0 \\ \exp(-(-x)^\alpha) & si \quad x \leq 0 \end{cases},\ \ \alpha < 0 \tag{2.6}$$

2) Distribution de Fréchet :

$$\Phi_\alpha(x) = \begin{cases} 0 & si \quad x \leq 0 \\ \exp(-x^{-\alpha}) & si \quad x > 0 \end{cases},\ \ \alpha > 0 \tag{2.7}$$

3) Distribution de Gumbel :

$$\Lambda(x) = \exp(-\exp(-x)),\ \ x \in \mathbb{R}. \tag{2.8}$$

Preuve. On peut trouver la démonstration de ce théorème dans le livre Embrechts et al. [6]. ■

Remarque 2.3.1 *1) Les trois distributions ci-dessus sont appellant les distributions de valeurs extrêmes.*

2) Les suites (c_n) et (d_n) ne sont pas unique.

Proposition 2.3.1 *Soit X une va positive $(X > 0)$. Alors la relation entre les trois distributions Fréchet, Weibull et Gumbel est :*

$$X \sim \Phi_\alpha \iff -X^{-1} \sim \Psi_\alpha \iff \ln X^\alpha \sim \Lambda.$$

2.4 Distribution des valeurs extrêmes généralisées

Définition 2.4.1 (Von-Mises 1936, Jenkinson 1955) *Jenkinson [9] en 1955 montre que ces trois distributions peuvent être regroupées dans un unique distribution dite distribution des valeurs extrêmes généralisées (GEV, Generalized Extreme Value), notée H_ζ et donnée par $\forall x \in \mathbb{R}$:*

$$H_\zeta(x) := \begin{cases} \exp\{-\exp(-x)\} & si \quad \zeta = 0 \\ \exp\{-(1+\zeta x)^{-\frac{1}{\zeta}}\} & si \quad \zeta \neq 0, 1+\zeta x > 0 \end{cases} \tag{2.9}$$

- *La fonction de distribution H_ζ est appelée loi des valeurs extrêmes.*
- *Le paramètre ζ est appelée indice des valeurs extrêmes.*
- *Si $\zeta = \alpha^{-1}$ on a :*

$$\Psi_{\frac{1}{\zeta}}(x) = H_{-\zeta}\left(\frac{1}{\zeta}(x+1)\right), \ x < 0$$
$$\Phi_{\frac{1}{\zeta}}(x) = H_{\zeta}\left(\frac{1}{\zeta}(x-1)\right), \ x > 0$$
$$\Lambda(x) = H_0(x), \ x \in \mathbb{R}$$

On peut écrire $H_\zeta(x)$ sous une forme générale :

$$H_{\zeta,\mu,\delta}(x) := \begin{cases} \exp\{-\exp(-(\frac{x-\mu}{\delta}))\} & si \quad \zeta = 0, \\ \exp\{-(1+\zeta(\frac{x-\mu}{\delta}))^{-\frac{1}{\zeta}}\} & si \quad \zeta \neq 0, 1+\zeta(\frac{x-\mu}{\delta}) > 0. \end{cases} \tag{2.10}$$

– *Le paramètre μ, $\mu \in \mathbb{R}$ est appelé le paramètre de localisation.*

– *Le paramètre δ, $\delta > 0$ est appelé le paramètre d'échelle.*

– *Si nous dérivons $H_{\zeta,\mu,\delta}$ on obtient la densité de la loi des valeurs extrêmes pour $\zeta \in \mathbb{R}$, et pour tout x tel que $1 + \zeta(\frac{x-\mu}{\delta}) > 0$,*

$$h_{\zeta,\mu,\delta}(x) := \begin{cases} \frac{1}{\delta}\exp\{-(\frac{x-\mu}{\delta}) - \exp(-(\frac{x-\mu}{\delta}))\} & si \quad \zeta = 0, \\ \frac{1}{\delta}(1+\zeta(\frac{x-\mu}{\delta}))^{-\frac{1+\zeta}{\zeta}}\exp\{-(1+\zeta(\frac{x-\mu}{\delta}))^{-\frac{1}{\zeta}}\} & si \quad \zeta \neq 0, \end{cases} \tag{2.11}$$

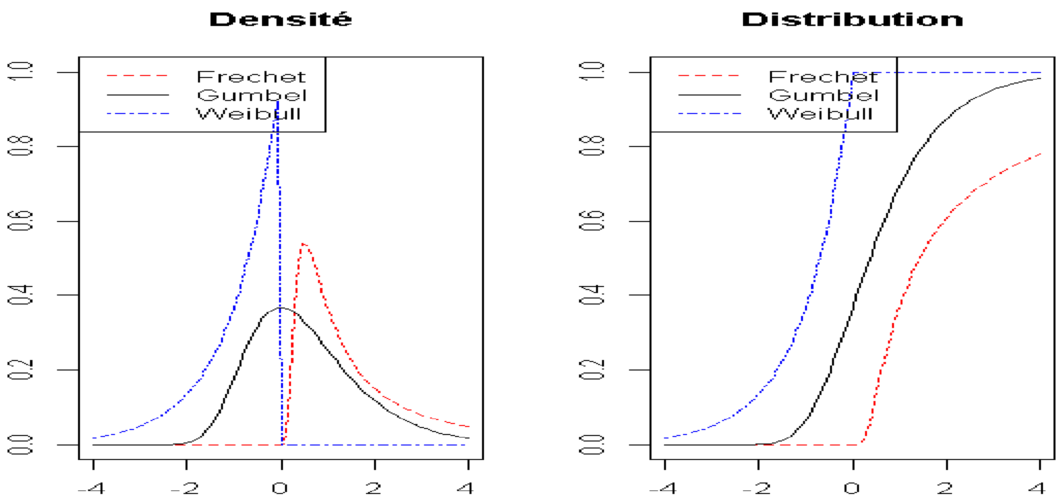

FIG. 2.1 – Distributions et densités des valeurs extrêmes généralisées Weibull ($\zeta = 1$), Fréchet ($\zeta = 1$) et Gumbull ($\zeta = 0$)

2.4.1 Loi max-stable

Définition 2.4.2 (Loi max-stable) *Une loi H non dégénérée est dite max-stable s'il existe des constantes $c_n \in \mathbb{R}_+^*$ et $d_n \in \mathbb{R}$ telles que*

$$H^n\left(c_n x + d_n\right) = H\left(x\right), \forall x \in \mathbb{R}.$$

Exemple 2.4.1 *La loi de Weibull (ψ_α) est une loi max stable, en effet,*

$$\begin{aligned}
\psi_\alpha^n\left(n^{\frac{-1}{\alpha}} x\right) &= \exp(-(-n^{\frac{-1}{\alpha}} x)^\alpha)^n \\
&= \exp(-(-n^{-1} x^\alpha) n) \\
&= \exp(-(-x^\alpha)) \\
&= \psi_\alpha\left(x\right).
\end{aligned}$$

2.5 Domaines d'attraction

Définition 2.5.1 (Domaines d'attraction) *Soit F la fonction de répartition, on dit que F appartient au domaine d'attraction du maximum de la distribution H_ζ $\left(\text{notée } F \in DA(H_\zeta)\right)$, si F vérifie le théorème de Fisher-Tippet.*

Nous présentons tout d'abord la défénition de fonction à variation lente et fonction à variation régulière.

2.5.1 Fonction à variation régulière

Définition 2.5.2 (Fonction à variation lente) *Une fonction positive L est dite à variation lente au voisinage de l'infini pour tout $x > 0$, si et seulement si*

$$\lim_{t \to \infty} \frac{L\left(tx\right)}{L\left(t\right)} = 1. \tag{2.12}$$

Définition 2.5.3 (Fonction à variation régulière) *Une fonction positive G est dite à variation régulière d'indice $\alpha \in \mathbb{R}$ au voisinage de l'infini (notée $G \in RV(\alpha)$), pour tout $x > 0$ si et seulement si :*

$$\lim_{t\to\infty} \frac{G(tx)}{G(t)} = x^{\alpha}. \tag{2.13}$$

On a aussi $G \in RV(\alpha)$ si et seulement si :

$$G(x) = x^{\alpha} L(x), \tag{2.14}$$

avec L est une fonction à variation lente (ou $L \in RV(0)$).

Exemple 2.5.1 *i) La fonction $\log(x)$ est une fonction à variation lente.*
ii) La fonction $x^{\alpha}\log(x)$ est une fonction à variation régulière.

Définition 2.5.4 (Représentation de Karamata) *Si L est une fonction à variation lente si et seulement si pour tout $t > 0$:*

$$L(t) = c(t) \exp\left[\int_a^t \frac{b(z)}{z} dz\right], \tag{2.15}$$

avec b et c sont des fonctions positives telles que :

$$\lim_{t\to+\infty} c(t) = c,\ c > 0 \ \mathit{et} \ \lim_{z\to+\infty} b(z) = 0.$$

Définition 2.5.5 (Le point terminal de F) *On définit le point terminal de la fonction de répartition F (noté x_F) par :*

$$x_F := \sup\{x \in \mathbb{R} : F(x) < 1\}. \tag{2.16}$$

2.5.2 Domaines d'attraction de Weibull

Ce domaine d'attraction regroupe la majorité des distributions dont le point terminal est fini, par exemple la loi Uniforme et la loi Beta.

Théorème 2.5.1 *F est appartient au domaine d'attraction de Weibull noté $F \in DA(\psi_{\zeta})$ avec $\zeta < 0$ si et seulement si $x_F < +\infty$ et $1 - F^*$ est une fonction à variation regulière d'indice $\frac{1}{\zeta}$ avec*

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

et le choix possible pour les constantes de normalisation c_n et d_n est :

$$c_n = x_F - F^{-1}(1 - \frac{1}{n}) \text{ et } d_n = x_F.$$

Exemple 2.5.2 *Soit $U = (U_1, ..., U_n)$ un échantillon suit la loi uniforme sur $[0, 1]$, de fonction de répartition $F(x) = x$, $x \in [0, 1]$.*
Puisque $F \in DA(\psi_1)$ alors les constantes de normalisation sont :

$$c_n = \frac{1}{n} \text{ et } d_n = 1,$$

d'où

$$\begin{aligned} \lim_{n\to\infty} P\left(c_n^{-1}\left(X_{n,n} - d_n\right) \leq x\right) &= \lim_{n\to\infty} P\left(X_{n,n} \leq \frac{x}{n} + 1\right) \\ &= \lim_{n\to\infty} F^n\left(\frac{x}{n} + 1\right) \\ &= \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n \\ &= \exp(x). \end{aligned}$$

2.5.3 Domaines d'attraction de Fréchet

Ce domaine d'attraction contient la majorité des fonctions de répartition à queue lourde, par exemple : Pareto, Cauchy et Student.

Théorème 2.5.2 *F appartient au domaine d'attraction de Fréchet avec $\zeta > 0$ noté $F \in DA(\Phi_\zeta)$ si et seulement si $x_F = +\infty$ et $\bar{F}$ est une fonction à variation regulier d'indice $-\frac{1}{\zeta}$ $\left(i.e : \bar{F}(x) = x^{\frac{-1}{\zeta}} L(x)\right)$, avec L est une fonction a variation lente. Le choix possible pour les constantes de normalisation c_n et d_n est :*

$$c_n = F^{-1}(1 - \frac{1}{n}) \text{ et } d_n = 0.$$

Exemple 2.5.3 *Soit X une va de loi de Pareto, de paramètre 1 et α et F sa fonction de répartition telle que :*

$$F(x) = 1 - x^{-\alpha},\ x,\ \alpha > 0,$$

puisque $F \in DA(\Phi_\alpha)$ alors les constantes de normalisation sont :

$$c_n = F^{-1}(1 - \frac{1}{n}) = n^{\frac{1}{\alpha}} \text{ et } d_n = 0,$$

alors,

$$\begin{aligned}
\lim_{n\to\infty} P\left(c_n^{-1}\left(X_{n,n} - d_n\right) \le x\right) &= \lim_{n\to\infty} P\left(X_{n,n} \le n^{\frac{1}{\alpha}} x\right) \\
&= \lim_{n\to\infty} F^n\left(n^{\frac{1}{\alpha}} x\right) \\
&= \lim_{n\to\infty} \left(1 - \frac{x^{-\alpha}}{n}\right)^n \\
&= \exp(-x^{-\alpha}).
\end{aligned}$$

2.5.4 Domaines d'attraction de Gumbel

Ce domaine d'attraction contient la majorité des fonctions de répartition à queue finie comme la loi Normale, Exponentielle et Gamma.

Théorème 2.5.3 *F appartient au domaine d'attraction de Gumbel noté $F \in DA(\Lambda)$ si et seulement s'il existe une fonction de Von-Mises*[1] *F^* telle que*

$$\bar{F}(x) = c(x)[1 - F^*(x)] = c(x)\exp\{-\int_t^x \frac{g(s)}{\alpha(s)} ds\}, t < x < x_F,$$

où c et g sont deux fonctions mesurables satisfaisantes $\lim\limits_{x \longrightarrow x_F} c(x) = c > 0$ et $\lim\limits_{x \longrightarrow x_F} g(x) = 1$, et α est une fonction positive absolument continue de densité α' vérifiant $\lim\limits_{x \longrightarrow x_F} \alpha'(x) = 0$.

Dans ce cas, le choix possible pour les constantes de normalisation c_n et d_n est :

$$c_n = c(d_n) \text{ et } d_n = F^{-1}(1 - \frac{1}{n}).$$

Exemple 2.5.4 *Soit $E = (E_1, ..., E_n)$ un échantillon suit la loi exponentielle de paramètre 1, avec la fonction de répartition*

$$F(x) = 1 - e^{-x}, x > 0.$$

puisque $F \in DA(\Lambda)$, alors les constantes de normalisation sont :

$$c_n = 1 \text{ et } d_n = \log n.$$

---

[1]

Définition 2.5.6 (Fonction de Von-Mises) *F est une fonction de Von-Mises s'il existe $t < x_F$ telle que*

$$\bar{F}(x) = c\exp\{-\int_t^x \frac{ds}{\alpha(s)}\}, t < x < x_F, \tag{2.17}$$

avec $c > 0$ et α est une fonction positive absolument continue de densité α' satisfaire $\lim\limits_{x \longrightarrow x_F} \alpha'(x) = 0$.

Donc,

$$\begin{aligned}
\lim_{n\to\infty} P\left(c_n^{-1}\left(X_{n,n}-d_n\right)\leq x\right) &= \lim_{n\to\infty} P\left(X_{n,n}\leq x+\log n\right)\\
&= \lim_{n\to\infty} F^n\left(x+\log n\right)\\
&= \lim_{n\to\infty}\left(1-\frac{e^{-x}}{n}\right)^n\\
&= \exp(-\exp\left(-x\right)).
\end{aligned}$$

Proposition 2.5.1 (Conditions de Von-Mises) *Soit F la fonction de répartition continue avec de densité f alors :*

i) Si $\alpha>0$ et

$$\lim_{x\to\infty}\frac{\left(x_{F}-x\right)f(x)}{\bar{F}(x)}=\alpha,$$

alors, $F\in DA(\Psi_\alpha)$ (ie : F appartient au domaine d'attraction de la loi de Weibull de paramètre α).

ii) Si $\alpha>0$ et

$$\lim_{x\to\infty}\frac{xf(x)}{\bar{F}(x)}=\alpha,$$

alors $F\in DA(\Phi_\alpha)$ (ie : F appartient au domaine d'attraction de la loi de Fréchet de paramètre α).

iii) Si f' la dérivée négative de f pour tout $x\in\left]t,x_{F}\right[$ et $f(x)=0$ pour $x\geq x_{F}$ et

$$\lim_{x\to x_{F}}\frac{f'(x)\bar{F}(x)}{\left(f(x)\right)^2}=-1,$$

alors $F\in DA(\Lambda)$ (ie : F appartient au domaine d'attraction de la loi de Gumbel).

Distribution des excès

Définition 2.5.7 (Fonction de répartition des excès) *Soient $X_1,...,X_n$ une suite d'observations indépendantes et de même loi de probabilité F et x_F un point terminal, alors* *un seuil $u < x_F$ fixé, on définit la variable $X_i - u$ pour $i = 1, ..., n$, l'excès au dessus du seuil u. La fonction de répartition des excès est :*

$$F_u(y) = P(X - u \leq y / X > u) = \frac{F(u+y) - F(u)}{\bar{F}(u)}, \, 0 \leq y \leq x_F - u. \tag{2.18}$$

Définition 2.5.8 (Fonction de moyenne des axcès) *On définit la fonction de moyenne des excès, (notée $e(u)$) comme suit :*

$$e(u) = E[X - u / X > u], \; u < x_F. \tag{2.19}$$

2.6 Distribution de Pareto généralisée

La fonction de distribution de Pareto généralisée standard (GPD, en anglais generalized Pareto distribution), pour $\zeta \in \mathbb{R}$ est définie par :

$$G_\zeta(x) := \begin{cases} 1 - (1 + \zeta x)^{\frac{-1}{\zeta}} & si \quad \zeta \neq 0, \\ 1 - \exp(-x) & si \quad \zeta = 0, \end{cases} \tag{2.20}$$

cette distribution est définie pour :

$$\begin{array}{ll} 0 \leq x \leq -\frac{1}{\zeta} & si \quad \zeta < 0, \\ x \geq 0 & si \quad \zeta \geq 0. \end{array}$$

Une forme générale de GPD, notée par $G_{\zeta,\nu,\beta}(x)$, est obtenue en remplaçant l'argument x par $\frac{x-\nu}{\beta}$ dans (2.20) , où $\nu \in \mathbb{R}$ et $\beta > 0$ sont respectivement les paramètres de localisation et d'échelle.

Le GPD standard correspond au cas $\nu = 0$ et $\beta = 1$.

Le GPD avec les paramètres $\nu = 0$ et $\beta > 0$ joue un rôle important dans l'analyse statistique des événements extrêmes, cette famille notée par $G_{\zeta,\beta}$ et définie comme suit :

$$G_{\zeta,\beta}(x) := \begin{cases} 1-(1+\zeta\frac{x}{\beta})^{\frac{-1}{\zeta}} & si \quad \zeta \neq 0, \\ 1-\exp\left(-\frac{x}{\beta}\right) & si \quad \zeta = 0, \end{cases} \tag{2.21}$$

avec,

$$\begin{aligned} 0 \leq x \leq -\tfrac{\beta}{\zeta} \quad & si \quad \zeta < 0, \\ x \geq 0 \quad & si \quad \zeta \geq 0. \end{aligned}$$

La densité de $G_{\zeta,\beta}(x)$ sécrit comme suit :

$$g_{\zeta,\beta}(x) := \begin{cases} \frac{1}{\beta}(1+\zeta\frac{x}{\beta})^{-\frac{1}{\zeta}-1} & si \quad \zeta \neq 0, \\ \frac{1}{\beta}\exp\left(-\frac{x}{\beta}\right) & si \quad \zeta = 0. \end{cases} \tag{2.22}$$

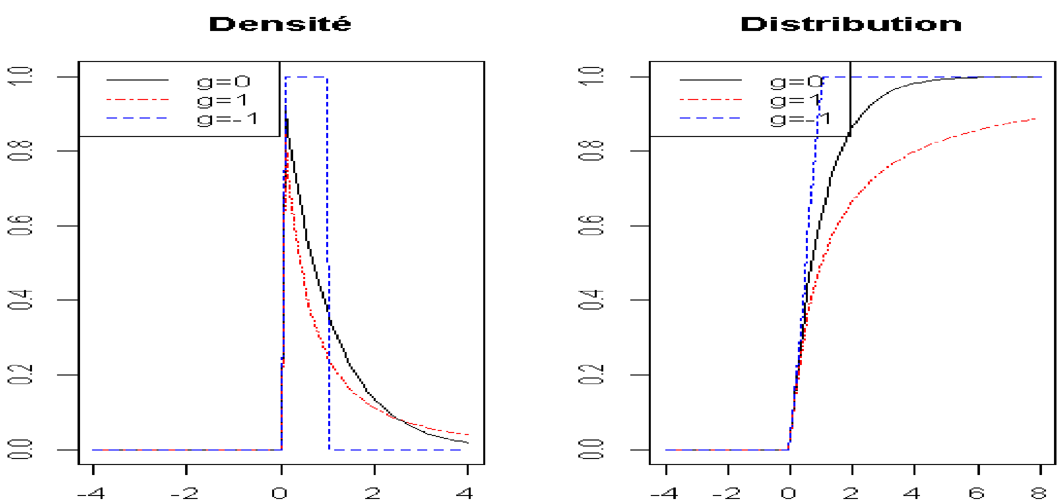

FIG. 2.2 – Distributions et densité de GPD standard avec différentes valeurs de ζ

Remarque 2.6.1 *La relation entre la loi de Pareto généralisée standard $G_\zeta(x)$ et la loi des valeurs extrêmes généralisées standard $H_\zeta(x)$ est :*

$$G_\zeta(x) = 1 + \log H_\zeta(x) \quad \text{si} \quad \log H_\zeta(x) > -1. \tag{2.23}$$

(Pour plus détail sur GPD voir Embrechts et al. [6]).

Théorème 2.6.1 (Balkema-de Haan et Pickands) *Si F appartient à l'un des trois domaines d'attraction de la loi des valeurs extrêmes, alors il existe une fonction strictement positive $\beta(u)$ telle que :*

$$\lim_{u \longrightarrow x_F} \sup_{0<x<x_F-u} \left| F_u(x) - G_{\zeta,\beta(u)}(x) \right| = 0, \tag{2.24}$$

où F_u est la fonction de répartition des excès au dessus du seuil u et $G_{\zeta,\beta(u)}$ est la fonction de répartition de la loi de Pareto généralisée.

Conclusion

Les statistiques d'ordre constituent le point de départ dans la théorie des valeurs extrêmes. Cette théorie est basée sur la connaissance de la loi asymptotique du maximum d'un échantillon.

Dans mémoire, nous avons donnés un aperçu général sur les statistiques d'ordre associées à un échantillon. Ensuite, nous avons présentés la loi des valeurs extrêmes ainsi que les domaines d'attraction et la distribution des valeurs extrêmes généralisées et l a distribution de Pareto généralisée.

Le but de ce mémoire été l'étude des distributions des statistiques d'ordre ainsi que les expressions des lois asymptotiques possibles des valeurs extrêmes.

Bibliographie

[1] Arnold, B.C, Balakrishan, N., Nagaraja, H.N. (1992). A First Course in Order Statistics. Wiely, New York.

[2] Benameur, S.(2010). Sur L'estimation de L'indice des valeurs extrêmes. Mémoire des magistère. Université de Biskra.

[3] David, H.A, Nagaraja, H.N (2003). Order Statistics Third Edition. Wiely

[4] Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag, New York, Berlin, Heidelberg Tokyo.

[5] De Haan, L. Ferreria, A. (2006). Extreme Values Theory : An introduction. Springer.

[6] Embrechts, P. Klüppelberg, C. Mikosch, T. (1997). Modelling Extremal Events, in : Applications in Mathematics. Springer-Verlag, New York. vol **33**.

[7] Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. Annales de Mathématiques **44**, $423 - 453$.

[8] Ihaka, R., Gentleman, R. (1996) *R : A Language for Data Analysis and Graphics.* Journal of Computational and Graphical Statistics **5** : $299 - 314$.

[9] Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quartely J. R. Methodol. Soc, **81**, $158 - 171$.

[10] Malmquist, S. (1950). On a proprety of order statistics from a rectangular distribution. Skandinavish Aktuarietidkrift **33**, $214 - 222$.

[11] Meraghni, D. (2008). Modelling Distribution Tails. Thèse de Doctorat, Université de Biskra, Algérie.

[12] Pyke, R. (1965). Spacings. J. Roy. Statis. Soc. **B 27**, 395 – 449.

[13] Pyke, R. (1972). Spacings Revisted Proceedings of the Sixth Berkeley Symposium on mathematical statistics and probability, vol **1**, 414 – 427. University of California Press, Berkeley.

[14] Reiss, R.D., Thomas, M. (1997). Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields Birkhäuser, Basel.

[15] Saporta, G. (1990). Probabilité, analyse des données et statistique. Editions Technip, Paris.

[16] Sukatme, P.V. (1937), Tests of significance for sample of the chi square population with two degrees of freedom. Ann, Eugen **8,** 52−56.

Annexe : Notations et abréviations

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

$\stackrel{d}{=}$	égalité en distribution.
$\stackrel{d}{\rightarrow}$	convergence en distribution.
C_n^j	combinaison.
$E_1, ..., E_{n+1}$	suite de va's suit la loi exponentielle standard.
$E_{1,n}, ..., E_{n,n}$	statistiques d'ordre associées à $E_1, ..., E_n$.
EVT	extrem value theory.
$D(H_\zeta)$	domaine d'attraction.
$E[X]$	l'espérene de X.
F	fonction de répartition.
$\bar{F}$	fonction de survie.
F^{-1}	l'inverse généralisée de F.
F_n	fonction de répartition empirique.
f	densité de probabilité d'une variable aléatoire.
$f_{X_{k,n}}$	fonction de densité de probabilité de $X_{k,n}$.
$f_{(X_{r,n},X_{s,n})}$	fonction de densité jointe de $X_{r,n}$ et $X_{s,n}$.

G_n	distribution empirique uniforme.
GEV	distribution des valeurs extrêmes généralisée.
GPD	distribution de Pareto généralisée.
$1\mathrm{I}_A$	fonction indicatrice de l'ensemble A.
ie	en d'autre terme.
iid	indépendantes et identiquement distibuées.
$\inf A$	infimum de l'ensemble A.
L	fonction à variation lente.
LGN	loi des grands nombres.
$M_n = X_{n,n}$	maximum de $X_1, ..., X_n$.
$m_n = X_{1,n}$	minimum de $X_1, ..., X_n$.
$\mathbb{N}$	nombre naturelle.
$\mathcal{N}\,(0.1)$	loi normale standard.
$\overset{p}{\rightarrow}$	convergence en probabilité.
$\overset{ps}{\rightarrow}$	convergence en prèsque sûre.
Q	fonction des quantiles.
Q_n	fonction des quantiles empirique.
$\mathbb{R}$	ensemble des nombres réelles.
$\sup A$	supremum de l'ensemble A.
S_n	somme arithmétique.
$S_{i,n}$	espacements uniformes.
TCL	théorème centrale limite.
μ	espérance, ou moyenne d'une va.
σ^2	variance d'un va.
Φ_α	loi de Fréchet.
ψ_α	loi de Weibull.
Λ	loi de Gumbel.

u	seuil.
$U_1, ..., U_n$	suite de va's uniformément distribées sur $[0, 1]$.
$U_{1,n}, ..., U_{n,n}$	statistiques d'ordre associées à $U_1, ..., U_n$.
va	variable aléatoire.
V_n	fonction des quantiles uniforme.
x_F	point terminale.
x_p	quantile d'ordre p.
$(X_1, ..., X_n)$	échantillons de taille n de va's.
$X_{1,n}, ..., X_{n,n}$	statistiques d'ordre associées à $(X_1, ..., X_n)$.
$:=$	égalité par définition.