DÉPARTEMENT DE MATHÉMATIQUES

Master : Mathématiques Appliquées

Processus Empiriques et Statistique D'ordre

BERKANE. H

2021-2022

Chapitre 1

Théorie de base des statistiques d'ordre

Les statistiques d'ordre sont très utiles dans la théorie des valeurs extrêmes, ils fournissent des informations sur la distribution de queue. Le but de ce chapitre est de décrire et de présenter les principaux résultats classiques sur la théorie des valeurs extrêmes dans le cas univarié réel qui permettent de auxquels on fera appel dans les autres chapitres.

1.1 Définitions

Soit $(X_i)_{i\in\mathbb{N}}$ une suite de variables aléatoires réelles indépendantes identiquement distribuées définie sur l'espace $(\Omega, \mathfrak{B})$, d'une densité commune f et d'une fonction de répartition F $(F(x) = P(X_n \leq x)$, pour $x \in \mathbb{R})$.
Soit S_n l'ensemble des permutation de $\{1, ..., n\}$.

Définition 1.1.1 La statistique d'ordre *de l'´echantillon* $(X_1, X_2, ..., X_n)$ *est le réarrangement croissant de* $(X_1, X_2, ..., X_n)$*. On la note* $(X_{1,n}, ..., X_{n,n})$*. On a* $X_{1,n} \leq ... \leq X_{n,n}$*, et il existe une permutation aléatoire* $\sigma_n \in S_n$ *telle que*

$$(X_{1,n}, ..., X_{n,n}) = (X_{\sigma_n(1)}, ..., X_{\sigma_n(n)}).$$

Pour $1 \leq i \leq n$, la v.a $X_{i,n}$ est appelée la $i-^{iéme}$ statistique d'ordre (ou statistique d'ordre i). En effet, les v. a. r. minimum et maximum du n-échantillon iid correspondent mieux à l'idée que l'on se fait d'une valeur extrême :

$$X_{1,n} = m_n = \min(X_1, X_2, ..., X_n) = \min_{1 \leq i \leq n} X_i$$

$$X_{n,n} = M_n = \max(X_1, X_2, ..., X_n) = \max_{1 \leq i \leq n} X_i$$

En fait, on peut se limiter a l'étude du maximum car

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

Il existe différents types de statistiques de rangs, qui nous aident à étudier des variables aléatoires ordonnées.

Définition 1.1.2 *Les variables aléatoires $R(1), R(2), ..., R(n)$ donné par les égalités :*

$$R(m) = \sum_{k=1}^{n} 1_{\{X_m \geq X_k\}} = 1 + \sum_{k=1}^{n} 1_{\{X_m > X_k\}} \qquad m = 1, 2, ..., n$$

sont dits les rangs *correspondant à l'échantillon $X_1, X_2, ..., X_n$*

Les rangs fournissent les égalités suivantes pour les événements :

$$\{R(m) = k\} = \{X_m = X_{k,n}\} \qquad m = 1, 2, ..., n \qquad k = 1, 2, ..., n$$

les réalisations $(r(1), ..., r(n))$ du vecteur de rangs correspondant $(R(1), R(2), ..., R(n))$ représentent toutes les permutations des valeurs $1, 2, ..., n$. Toute réalisation $(r(1), ..., r(n))$ correspond à l'événement $X_{\delta(1)} < X_{\delta(2)} < ... < X_{\delta(n)}$; telleque $\delta(r(k)) = k$

Ici $\delta(k)$ désigne l'indice de X, dont le rang pour cette réalisation prend la valeur k. Pour différentes réalisations du vecteur $(R(1), R(2), ..., R(n))$, $\delta(k)$ peut prendre des valeurs différentes de l'ensemble $\{1, 2, ..., n\}$ et nous avons vraiment affaire à de nouvelles variables aléatoires, dont les réalisations sont $\delta(1), \delta(2), ..., \delta(n)$.

Prise en compte de la symétrie de l'échantillon $X_1, X_2, ..., X_n$, nous obtenons les événements $X_{\delta(1)} < X_{\delta(2)} < ... < X_{\delta(n)}$ ont les mêmes probabilités pour toutes les permutations $(\delta(1), ..., \delta(n))$ de nombres $1, 2, ..., n$

$$\mathbb{P} = \{R(1) = r(1), R(2) = r(2), ..., R(n) = r(n)\} = \mathbb{P}\left\{X_{\delta(1)} < X_{\delta(2)} < ... < X_{\delta(n)}\right\} = \frac{1}{n!}$$

où $r(1), r(2), ..., r(k)$ sont différents numéros tirés de l'ensemble $\{1, 2, ..., n\}$.

Définition 1.1.3 *Soient $X_1, X_2, ..., X_n$ un échantillon aléatoire de taille n issu d'une distribution continue et soient $X_{1,n}, ..., X_{n,n}$, les n statistiques d'ordre correspondant. La variables aléatoires $\Delta(1), \Delta(2), ..., \Delta(n)$, qui satisfont les égalités suivantes :*

$$\{\Delta(k) = m\} = \{X_{k,n} = X_m\} \qquad m = 1, 2, ..., n \qquad k = 1, 2, ..., n$$

sont dits anti-rangs.

Les mêmes arguments, que nous avons utilisés pour les rangs, montrent que toute réalisation $(\delta(1), \delta(2), ..., \delta(n))$ du vecteur $(\Delta(1), \Delta(2), ..., \Delta(n))$ est une permutation de nombres $(1, 2, ..., n)$ et tout les $n!$ réalisations ont des probabilités égales à $1/n!$. En effet, les vecteurs d'anti-rangs sont étroitement liés aux statistiques d'ordre et aux vecteurs de rangs correspondants. En fait, pour toutes les égalités k et m

$$\{\Delta(k) = m\} = \{X_{k,n} = X_m\} = \{R(m) = k\}, \qquad \Delta(R(m)) = m, \qquad R(\Delta(m)) = m$$

Définition 1.1.4 *Soient $X_1, X_2, ..., X_n$ des variables aléatoires indépendantes, ayant des distributions continues (pas nécessairement identiques). Les variables aléatoires $\rho(1), \rho(2), ...,$*

donné par les égalités :

$$\rho(m) = \sum\nolimits_{k=1}^{m} 1_{\{X_m \geq X_k\}} \qquad m = 1, 2, ..., n$$

sont dits rangs séquentiels.

Les rangs séquentiels ne sont pas associés à un échantillon de taille n fixe. En fait, $\rho(m)$ montre la position d'une nouvelle observation X_m à venir parmi ses prédécesseurs $X_1, X_2, ..., X_{m-1}$. En générale, $\rho(m) = k$ impliques que $X_m = X_{k,m}$

$$\mathbb{P}\{\rho(m) = k\} = P\{X_m = X_{k,m}\} = \frac{1}{m}; \qquad k = 1, 2, ..., m$$

Exemple 1.1.1 *Soit les données suivantes représentent les durées de vie (heures) de* 15 *batteries (réalisation d'un échantillon de taille* 15*) :*

$$\begin{array}{lllll} x_1 = 20.3 & x_2 = 17.2 & x_3 = 15.4 & x_4 = 16.8 & x_5 = 24.1 \\ x_6 = 12.6 & x_7 = 15.0 & x_8 = 18.1 & x_9 = 19.1 & x_{10} = 21.3 \\ x_{11} = 22.3 & x_{12} = 16.4 & x_{13} = 13.5 & x_{14} = 25.8 & x_{15} = 16.9 \end{array}$$

Étant ordonnées ces observations nous donnent des réalisations de statistiques d'ordre :

$$\begin{array}{lllll} x_{1,15} = 12.6 & x_{2,15} = 13.5 & x_{3,15} = 15.0 & x_{4,15} = 15.4 & x_{5,15} = 16.4 \\ x_{6,15} = 16.8 & x_{7,15} = 16.9 & x_{8,15} = 17.2 & x_{9,15} = 18.1 & x_{10,15} = 19.1 \\ x_{11,15} = 20.3 & x_{12,15} = 21.3 & x_{13,15} = 22.3 & x_{14,15} = 24.1 & x_{15,15} = 25.8 \end{array}$$

Les réalisations des rangs sont données comme suit :

$$\begin{array}{lllll} r(1) = 11 & r(2) = 8 & r(3) = 4 & r(4) = 6 & r(5) = 14 \\ r(6) = 1 & r(7) = 3 & r(8) = 9 & r(9) = 10 & r(10) = 12 \\ r(11) = 13 & r(12) = 5 & r(13) = 2 & r(14) = 15 & r(15) = 7 \end{array}$$

Les anti-rangs sont présentés par la suite :

$$\begin{array}{lllll} \delta(1)=6 & \delta(2)=13 & \delta(3)=7 & \delta(4)=3 & \delta(5)=12 \\ \delta(6)=4 & \delta(7)=15 & \delta(8)=2 & \delta(9)=8 & \delta(10)=9 \\ \delta(11)=1 & \delta(12)=10 & \delta(13)=11 & \delta(14)=5 & \delta(15)=14 \end{array}$$

Statistiques d'ordres

1.2 Distributions des statistiques d'ordre

Notons par $F_{i,n}(x)$ la distribution de $X_{i,n}$. *David* [*1970*] et *Balakrishnan et Clifford Cohen* [*1991*] montrent que l'expression de la distribution de $X_{i,n}$ est :

$$F_{X_{i,n}}(x) = F_{i,n}(x) = \mathbb{P}(X_{i,n} \leq x) = \sum_{r=i}^{n} C_n^r [F(x)]^r [1 - F(x)]^{n-r} \qquad x \in \mathbb{R}$$

En utilisant la relation :

$$\sum_{r=i}^{n} C_n^r p^r \left(1-p\right)^{n-r} = \int_0^p \frac{n!}{(i-1)!(n-i)!} t^{i-1} [1-t)]^{n-i} dt;$$

La variable aléatoire $Y_i = F(X_{i,n})$ suit une loi *bêta* de paramètre $(i, n-i+1)$.

$$F_{i,n}(x) = P\{X_{i,n} \leq x\} = I_{F(x)}(i, n-i+1),$$

Où

$$I_x(a,b) = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1} dt, \qquad B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \qquad \Gamma(k) = (k-1)!$$

désigne la fonction bêta incomplète.

$$F_{i,n}(x) = \sum_{r=i}^{n} C_n^r [F(x)]^r [1 - F(x)]^{n-r} = \int_0^{F(x)} \frac{n!}{(i-1)!(n-i)!} t^{i-1} [1-t)]^{n-i} dt$$

$$= \int_0^{F(x)} \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} t^{i-1} [1-t)]^{n-i} dt = \frac{1}{B(i, n-i+1)} \int_0^{F(x)} t^{i-1} [1-t)]^{n-i} dt$$

$$= I_{F(x)}(i, n-i+1).$$

Nous en déduisons que la *fonction de densité* est :

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

Preuve. Voir *Reiss* (*1989*), Lemme 1.3.1, page 20. ■

Lemme 1.2.1 (*Barry et al.(1992)*) *La densité jointe de statistique d'ordre* $(X_{1,n}, ..., X_{n,n})$ *est donné par :*

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

On peut donc conclure pour la *statistique du* minimum *que la distribution et la densité* sont respectivement :

$$F_{1,n}(x) = F_{X_{1,n}}(x) = 1 - [1 - F(x)]^n$$

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

De même pour l*a statistique du maximum*, on a :

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

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

Preuve. En utilisant la propriété d'indépendance des variables aléatoires $X_1, X_2, ..., X_n$, nous en déduisons que

$$F_{1,n}(x) = \mathbb{P}\{X_{1,n} \leq x\} = 1 - \mathbb{P}\{X_{1,n} > x\} = 1 - \mathbb{P}\left\{\bigcap_{i=1}^{n} X_i > x\right\}$$

$$= 1 - \prod_{i=1}^{n} \mathbb{P}\{X_i > x\} = 1 - \prod_{i=1}^{n} [1 - \mathbb{P}\{X_i \leq x\}] = 1 - [1 - F(x)]^n$$

$$F_{n,n}(x) = \mathbb{P}\{X_{n,n} \leq x\} = \mathbb{P}\left\{\bigcap_{i=1}^{n} X_i \leq x\right\} = \prod_{i=1}^{n} \mathbb{P}\{X_i \leq x\} = [F(x)]^n$$

■

Soit $X_1, X_2, ..., X_5$ est un échantillon aléatoire simple et $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ sont les statistiques d'ordre et de taille $n = 5$ avec densité

$$f(x) = 2x \qquad 0 < x < 1$$

$$F(x) = \int_0^x f(t)\, dt = x^2 \qquad 0 < x < 1;$$

$$F(y) = P(X \leq y) = y^2 \qquad 0 < y_1 < 1$$

1.

$$f_{1,5}(y_1) = n[1-F(y_1)]^{n-1}f(y_1) = 5\left[1-y_1^2\right]^4 2y_1 = 10y_1\left[1-y_1^2\right]^4$$

$$F_{1,5}(y_1) = 1-[1-F(y_1)]^n = 1-[1-y_1^2]^5 \qquad 0<y_1<1;$$

2.

$$f_{5,5}(y_5) = n[F(y_5)]^{n-1}f(y_5) = 5\left[y_5^2\right]^4 2y_5 = 10y_5^9 \qquad 0<y_5<1;$$

$$F_{5,5}(y_5) = [F(y_5)]^n = \left[y_5^2\right]^5 = y_5^{10} \qquad 0<y_5<1;$$

3.

$$f_{4,5}(y_4) = \frac{5!}{3!1!}[y_4^2]^3[1-y_4^2]2y_4 = 40y_4^7[1-y_4^2] \qquad 0<y_4<1;$$

$$F_{4,5}(y_4) = \sum_{r=4}^{5} C_5^r[F(y_4)]^r[1-F(y_4)]^{5-r} = 8y_4^8[1-y_4^2]+y_4^{10} \qquad 0<y_4<1;$$

4. $\mathbb{P}\left(y_4 \le 1/2\right) = 8\left(1/2\right)^8\left[1-\left(1/2\right)^2\right]+\left(1/2\right)^{10} = 1/64$

Lemme 1.2.2 *(Densité conjointe de deux statistiques d'ordre) La fonction de distribution conjointe de* $(X_{i,n} \le X_{j,n})$ *avec* $1 \le i < j \le n$ *et* $-\infty < x < y < +\infty$ *est donnée par* Arnold et al. (2008) *comme :*

$$F_{i,j:n}(x,y) = \sum_{s=j}^{n}\sum_{r=i}^{s}\frac{n!}{(r-1)!(s-r-1)!(n-s)!}[F(x)]^{r-1}f(x)[F(y)-F(x)]^{s-r-1}f(y)[1-F(y)]^{n-s}$$

Pour $x \geq y$ la fonction de répartition est :

$$F_{i,j:n}(x,y) = F_{(X_{i,n} \leq X_{j,n})}(x,y) = F_{X_{j,n}}(y),$$

La densité conjointe de $(X_{i,n} \leq X_{j,n})$

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

La fonction de densité conjointe de $X_{1,n}$ et $X_{n,n}$

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

Exemple 1.2.1 *Supposons que $X_1, X_2, \ldots, X_n$ sont n i.i.d. variables aléatoires et $Y_1 < Y_2 < Y_3$ sont les statistiques d'ordre d'une densité*

$$f(x) = e^{-x} \qquad x \geq 0.$$

1. $f(y_1, y_2, y_3) = n!\prod_{i=1}^{n} f(y_i) = 3!f(y_1)f(y_2)f(y_3) = 6e^{-y_1}e^{-y_2}e^{-y_3} = 6e^{-(y_1+y_2+y_3)}$.
2. $F(x) = 1 - e^{-x} \qquad x \geq 0.$

$$f_{1,3}(y_1) = n[1-F(y_1)]^{n-1}f(y_1) = 3\left[1-\left(1-e^{-y_1}\right)\right]^2 e^{-y_1} = 3e^{-3y_1} \qquad y_1 \geq 0.$$

3.

$$f_{3,3}(y_3) = 3[1-e^{-y_3}]^{3-1}e^{-y_3} \qquad y_3 \geq 0$$

4. *La fonction de densité conjointe de Y_1 et Y_3*

$$f(y_1, y_3) = \frac{3!}{1!}[F(y_3)-F(y_1)]f(y_1)f(y_3) = 6e^{-(y_1+y_3)}\left[e^{-y_1} - e^{-y_3}\right] \qquad 0 < y_1 < y_3 < \infty$$

5. La densité et la valeur de la médiane $y_{\frac{n+1}{2}} = y_2 = \tilde{x}$

$$f(y_2) = \frac{3!}{1!1!} F(y_2)[1 - F(y_2)] f(y_2) = 6 \left[1 - e^{-y_2}\right] \left[1 - 1 + e^{-y_2}\right] e^{-y_2}$$

$$= 6e^{-2y_2} \left[1 - e^{-y_2}\right] \qquad 0 < y_2 < \infty$$

$$F(y_2) = F(\tilde{x}) = 1/2$$

$$1 - e^{-y_2} = 1/2 \Rightarrow e^{-y_2} = 1/2 \Rightarrow -y_2 = \ln 2$$

$$y_{\frac{n+1}{2}} = y_2 = \tilde{x} = \ln 2$$

$$\mathbb{P}(y_1 > \tilde{x}) = \int_{\tilde{x}}^{\infty} 3e^{-3y_1} dy_1 = \int_{\ln 2}^{\infty} 3e^{-3y_1} dy_1 = \frac{1}{8}.$$

Exemple 1.2.2 *Un échantillon aléatoire est tiré d'une distribution uniforme sur l'intervalle* $[0, 1]$. *Obtenir la distribution des statistiques d'ordre* $i-^{iéme}$ *et la distribution conjointe de deux statistiques d'ordre et*

Solution 1.2.1 *La densité et la fonction de répartition de* $U_{(0,1)}$ *sont :*

$$f(x) = 1, \quad F(x) = x.$$

La distribution des statistiques du $i^{\text{ème}}$ *ordre est*

$$f_{i,n}(x) = \frac{1}{B(i, n-i+1)} [F(x)]^{i-1} [1 - F(x)]^{n-i} f(x) = \frac{1}{B(i, n-i+1)} x^{i-1} [1 - x]^{n-i};$$

la distribution conjointe de deux statistiques d'ordre est :

$$f_{i,j:n}(x, y) = \frac{1}{B(i, j-i, n-j+1)} x^{i-1} [y - x]^{j-i-1} [1 - y]^{n-j}$$

La fonction de densité conjointe de $X_{1,n}$ *et* $X_{n,n}$

$$f_{1,n:n}(x,y) = n\,(n-1)\,[y-x]^{n-2}.$$

Exercise 1.2.1 *Soit* $Y_1 < Y_2 < Y_3$ *est les statistiques d'ordre d'une population uniforme sur* $(0,1)$. *Déterminer le densité de* $Z_1 = Y_3 - Y_1$ *et la médiane.*

Exemple 1.2.3 *Soit* $(X_i)_{1:n}$ *un échantillon aléatoire simple d'une population uniforme sur* $(0,1)$. *Déterminer* $\mathbb{C}ov(X_{(i)}, X_{(j)})$ *des statistiques* $i^{\text{ème}}$ *et* $j^{\text{ème}}$ *tel que* $i < j$

$$f(x_i) = 1, \qquad F(x_i) = x_i, \qquad 0 < x < 1$$

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

$$= cX^{i-1}[Y-X]^{j-i-1}[1-Y]^{n-j} \qquad 0 \leq X < Y \leq 1$$

$$\mathbb{E}\,(XY) = c\int_0^1 \int_0^y xyx^{i-1}[y-x]^{j-i-1}[1-y]^{n-j}dxdy$$

$$= c\int_0^1 y^j\,(1-y)^{n-j}\left[\int_0^y \left(\frac{x}{y}\right)^i [1-\frac{x}{y}]^{j-i-1}dx\right]dy$$

On $u = \dfrac{x}{y} \Rightarrow du = \dfrac{dx}{y}$

$$= c\int_0^1 y^{j+1}\,(1-y)^{n-j}\left[\int_0^1 u^i[1-u]^{j-i-1}du\right]dy = cBeta\,(i+1, j-i)\int_0^1 y^{j+1}\,(1-y)^{n-j}\,dy$$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!}Beta\,(i+1, j-i)\,Beta\,(j+2, n-j+1) = \frac{i\,(j+1)}{(n+1)\,(n+2)}$$

La distribution des statistiques du $i^{\text{ème}}$ ordre est

$$f_{X_{(i)}}(x) = f_{i,n}(x) = \frac{n!}{(i-1)!(n-i)!} X^{i-1}[1-X]^{n-i} \qquad 0 \leq x \leq 1.$$

$$\mathbb{E}\left(X\right) = \frac{n!}{(i-1)!(n-i)!} \int_0^1 x \left\{x^{i-1}[1-x]^{n-i}\right\} dx = \frac{n!}{(i-1)!(n-i)!} Beta\left(i+1, n-i+1\right)$$

$$= \frac{n!}{(i-1)!(n-i)!} \frac{i!(n-i)!}{(n+1)!} = \frac{i}{n+1}$$

$$\mathbb{E}\left(X^2\right) = \frac{n!}{(i-1)!(n-i)!} \int_0^1 x^{i+1}[1-x]^{n-i} dx = \frac{n!}{(i-1)!(n-i)!} Beta\left(i+2, n-i+1\right)$$

$$= \frac{n!}{(i-1)!(n-i)!} \frac{(i+1)!(n-i)!}{(n+2)!} = \frac{i\left(i+1\right)}{\left(n+1\right)\left(n+2\right)}$$

$$\mathbb{V}ar\left(X\right) = \mathbb{E}\left(X^2\right) - \mathbb{E}^2\left(X\right) = \frac{i\left(n+1-i\right)}{\left(n+1\right)^2\left(n+2\right)}$$

$$\mathbb{C}ov(X,Y) = \mathbb{E}\left(XY\right) - \mathbb{E}\left(X\right)\mathbb{E}\left(Y\right)$$

$$= \frac{i\left(j+1\right)}{\left(n+1\right)\left(n+2\right)} - \frac{i}{n+1}\frac{j}{n+1} = \frac{i\left(n+1-j\right)}{\left(n+1\right)^2\left(n+2\right)}$$

Lemme 1.2.3 *Si F est continue, alors p.s. on a $X_{(1,n)} < ... < X_{(n,n)}$.*

Preuve. Il suffit de vérifier que $\mathbb{P}(\exists i \neq j \text{ tel que } X_i = X_j) = 0$. On a :

$$\mathbb{P}(\exists i \neq j \text{ tel que } X_i = X_j) \leq \mathbb{P}(\exists i \neq j \text{ tel que } F(X_i) = F(X_j))$$

$$\leq \sum\nolimits_{i \neq j} \mathbb{P}(F(X_i) = F(X_j))$$

$F(X_i)$ et $F(X_j)$ sont pour $i \neq j$ des variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. On en déduit que

$$\mathbb{P}(F(X_i) = F(X_j)) = \int_{[0,1]^2} 1_{\{u=v\}} du dv = 0.$$

Donc $p.s.$ pour tous $i \neq j$, on a $X_i \neq X_j$. ■

Définition 1.2.1 *(Distribution empirique) La fonction de répartition empirique de l'échantillon ($X_1, X_2, ..., X_n$) est évaluée à l'aide des statistiques d'ordre comme suit :*

$$F_n(x) = \frac{1}{n} \sum\nolimits_{i=1}^{n} 1_{\{X_i \leq x\}} := \begin{cases} 0 & \text{Si} \quad x < X_{1,n}, \\ \dfrac{i-1}{n} & \text{Si} \quad X_{i-1,n} \leq x < X_{i,n}\,, \qquad 2 \leq i \leq n \\ 1 & \text{Si} \quad x \geq X_{n,n}. \end{cases}$$

$$\mathbb{E}\left[F_n(x)\right] = F(x), \qquad \mathbb{V}ar\left[F_n(x)\right] = \frac{F(x)(1 - F(x))}{n}$$

Définition 1.2.2 *(Fonctions quantile et quantile de queue) La fonction quantile de la fonction de distribution F est la fonction inverse généralisée de F définie par :*

$$Q(s) := F^{\leftarrow}(s) = \inf\left\{x \in \mathbb{R} : F(x) \geq s\right\}; \qquad 0 < s < 1.$$

Dans la théorie des extrêmes, une fonction notée U et appelée fonction quantile de queue, est définie par :

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

Définition 1.2.3 *(Fonctions empiriques de quantile et de quantile de queue) La fonction quantile empirique de l'échantillon* $(X_1, X_2, ..., X_n)$ *est définie par :*

$$Q_n(s) := F_n^{\leftarrow}(s) = \inf\{x \in \mathbb{R} : F_n(x) \geq s\}; \qquad 0 < s < 1.$$

La fonction empirique de quantile de queue correspondante est :

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

On Peut être exprimée comme une fonction simple des statistiques d'ordre relatives à l'échantillon $(X_1, X_2, ..., X_n)$ et nous avons :

$$Q_n(s) := \begin{cases} X_{i,n} & \text{si} \quad \frac{(i-1)}{n} < s \leq \frac{i}{n} \\ \\ X_{[np]+1,n} & \text{si} \quad 0 < s \leq 1 \end{cases}$$

Notons que pour $0 < p < 1$; $X_{[np]+1,n}$ est le quantile d'échantillon de l'ordre p.

Théoème 1.2.1 *(Glivenko-Cantelli, 1933)*

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \overset{p.s}{\rightarrow} 0 \qquad \textit{quand} \qquad n \rightarrow \infty.$$

Définition 1.2.4 *(Point extrême) On note par* x_F *(resp.* x_F^**) le point extrême supérieur (resp. inférieur) de la distribution* F *(i.e. la plus grande valeur possible pour* $X_{k,n}$ *peut* *prendre la valeur $+\infty$ (resp. $-\infty$)) au sens où*

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

$$x_F^* := \inf\{x : F(x) > 0\}$$

Proposition 1.2.1 *(Limite de $X_{n,n}$)*

$$X_{n,n} \overset{p.s}{\to} x_F \qquad \text{quand} \qquad n \to \infty.$$

D'autre part, le comportement asymptotique de la loi de $X_{n,n}$ est donné, dans certaines conditions sur la queue de distribution.

Lemme 1.2.4 *(Transformation quantile) Soit $X_1, X_2, ..., X_n$ des variables aléatoires indépendantes et de fonction de répartition F. Soit $U_1, ..., U_n$ des variables aléatoires indépendantes de loi uniforme standard : Alors :*

(a) Pour toute fonction de distribution F, on a

$$X_{i,n} \overset{d}{=} F^{\leftarrow}(U_{i,n}), \qquad i = 1, ..., n$$

(b) Lorsque F est continue, on a

$$F(X_{i,n}) \overset{d}{=} U_{i,n}, \qquad i = 1, ..., n$$

Preuve. Voir *Reiss (1989)*, Théorème 1.2.5, page 17. ■

T.D. N° – 1

Exercice-1

(a) Soit $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ est la série variationnelle basée sur des v.a indépendantes $X_1, X_2, ..., X_n$. Montrez que : $\mathbb{P}\{X_{1,n} < X_{2,n} < ... < X_{n,n}\} = 1$.

(b) Soit $X_1, X_2, ..., X_n$ les n observations indépendantes de la variable aléatoire X, ayant la distribution géométrique, Trouver : $p_n = \mathbb{P}\{X_{1,n} < X_{2,n} < ... < X_{n,n}\}$, $n = 1, 2, ...$

(c) Montrez que $R(1), R(2), ..., R(n)$ sont des variables aléatoires dépendantes pour tout $n = 2, 3, ...$ et trouvez les espérances et les variances de $R(k)$, $1 \leq k \leq n$.

Exercice-2

Soient $X_1, X_2, ..., X_n$ des variables aléatoires i.i.d. de fonction de répartition F. On suppose que F admet une densité f par rapport à la mesure de Lebesgue. On note $X_{(1)} \leq X_{(2)} \leq , ..., \leq X_{(n)}$ les variables aléatoires $X_1, X_2, ..., X_n$ réordonnées par ordre croissant.

1. Donner l'expression de la loi de la statistique d'ordre $X_{(1)}, ..., X_{(n)}$ en fonction de f.
2. Déterminer la fonction de répartition $F_k(x)$ puis la densité $f_k(x)$ de $X_{(k)}$.
3. Sans utiliser les résultats des questions précédentes, calculer les fonctions de répartition de $X_{(1)}, X_{(n)}$, du couple $(X_{(1)}, X_{(n)})$ et la loi de la statistique $W = X_{(n)} - X_{(1)}$. Les variables $X_{(1)}$ et $X_{(n)}$ sont–elles indépendantes ?

Exercice-3

Soit $X_1, X_2, ..., X_n$ des v.a.r. définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, indépendantes et de même loi absolument continue par rapport à la mesure de Lebesgue de densité f. Pour tout ω dans Ω, on peut ordonner les réels $X_1(\omega), ..., X_i(\omega), ..., X_n(\omega)$ sous la forme :

$$X_{(1)}(\omega) \leq X_{(2)}(\omega) \leq ... \leq X_{(i)}(\omega) \leq ... \leq X_{(n)}(\omega).$$

L'application $X_{(i)} : \omega \in \Omega \rightarrow X_{(i)}(\omega)$, ainsi définie pour chaque i est une v.a.r. dite i^{eme} statistique d'ordre.

1. Calculer cette de $X_{(n)} = \sup\{X_1, X_2, ..., X_n\}$ (f.d.r. et densité).
2. Calculer la loi de $X_{(1)} = \inf\{X_1, X_2, ..., X_n\}$ (f.d.r. et densité).
3. Calculer la loi du couple $(X_{(1)}, X_{(n)})$. En déduire celle de l'étendue $R = X_{(n)} - X_{(1)}$.
4. Soit N_y le nombre de X_i inférieurs à y. Quelle est la loi de N_y ? Que dire des événements $\{Ny \geq k\}$ et $\left\{X_{(k)} \leq y\right\}$? En déduire la f.d.r. de $X_{(k)}$.
5. On pourrait du résultat précédent tirer la densité de la v.a. $X_{(k)}$. Mais c'est fastidieux. Il y a bien plus simple en attaquant le problème directement, ce que l'on propose de faire maintenant. On pourra utiliser le résultat suivant : Si f est continue sur un intervalle $[a, b]$, alors, pour tout x dans cet intervalle, on a :

$$f(x) = \lim_{h \to 0^+} \frac{P(X \in]x, x+h])}{h}$$

6. Montrer que si $E(X)$ existe alors $E(X_{(k)})$ aussi.
7. Calculer la densité du vecteur $(X_{(1)}, ..., X_{(n)})$.

Exercice-4

Dans ce qui suit, $X_1, X_2, ..., X_n$ est un échantillon aléatoire simple et $Y_1 \leq ... \leq Y_n$ sont les statistiques d'ordre.

1. Soit $X_1, X_2, ..., X_n$ un échantillon aléatoire simple d'une population exponentielle de paramètre β. Montrer que : Y_1 est de loi exponentielle de paramètre β/n et la densité de Y_n est $g(y_n) = \frac{n}{\beta} e^{-y_n/\beta} \left[1 - e^{-y_n/\beta}\right]^{n-1}$
2. Soit $X_1, X_2, ..., X_n$ un échantillon aléatoire simple d'une population uniforme sur $(0, 1)$. Déterminer les distributions de Y_1 et de Y_n puis la distribution de la médiane $\tilde{x}$, l'espérance et la variance de Y_1.

Chapitre 2

Théorie des Valeurs Extrêmes Univariées

La théorie des valeurs extrêmes (TVE) est basée sur l'approximation asymptotique des lois des maxima convenablement normalisés de vecteurs aléatoires dont les composantes sont des variables supposées $i.i.d.$ Les modèles des valeurs extrêmes sont appliquées à une grande variété de problèmes tels l'environnement (vitesse du vent, extrêmes pluviométriques et de températures,...), la finance et l'assurance (Mesure du risque, Valeur à risque VaR...

Dans ce chapitre, Nous proposons un aperçu des éléments théoriques essentiels de la théorie des valeurs extrêmes. nous étudions le comportement des valeurs extrêmes d'un échantillon de variables aléatoires univariées, nous donnons des critères pour que la limite en loi du maximum suive telle loi des valeurs extrêmes.

2.1 Les théorèmes limites

L'objectif principal de la théorie des valeurs extrêmes (TVE) est de trouver les lois limites possibles pour suite des maximums $\{\max(X_1, X_2, ..., X_n), n \in \mathbb{N}^*\}$ si on connait la loi exacte de v.a X.

2.1.1 Théorème des grands nombres

La loi des grands nombres indique que lorsque l'on fait un tirage aléatoire dans une série de grande taille, plus on augmente la taille de l'échantillon, plus les caractéristiques statistiques du tirage (l'échantillon) se rapprochent des caractéristiques statistiques de la population.

a) Loi faible des grands nombres

Théoème 2.1.1 *Soit $X_1, X_2, ..., X_n$ un ensemble de variables aléatoires i.i.d définies sur le même espace de probabilité de variance finie σ^2 et d'espérance finie μ. On pose :*

$$\bar{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \; \textit{Alors } \bar{X}_n \xrightarrow{P} \mu \quad i.e \forall \varepsilon > 0 \quad \lim_{n\to\infty} \mathbb{P}\left(\left|\bar{X}_n - \mu\right| > \varepsilon\right) = 0$$

b) Loi forte des grands nombres

Considérons n variables aléatoires $i.i.d.$ $X_1, X_2, ..., X_n$ d'espérance finie μ. On pose

$$\bar{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \; \text{Alors } \bar{X}_n \xrightarrow{P.s} \mu \qquad i.e \forall \varepsilon > 0 \qquad \mathbb{P}\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

2.1.2 Théorème de la limite centrale

Théoème 2.1.2 *(TCL) Si $X_1, X_2, ..., X_n$ est une suite de variable aléatoire définie sur le même espace de probabilité de variance σ^2 finie et de moyenne μ On pose $S_n = X_1 + X_2 + ... + X_n$ Alors*

$$S_n \overset{loi}{\to} \mathcal{N}\left(n\mu, \sigma\sqrt{n}\right) \qquad i.e \; \frac{S_n - n\mu}{\sigma\sqrt{n}} \overset{loi}{\to} \mathcal{N}(0,1) \;\; \textit{quand } n \to \infty$$

Définition 2.1.1 Statistique d'ordre extrême *Soient $X_1, X_2, ..., X_n$ des variables aléatoires i.i.d de loi F et de densité f. On a $X_{1,n} \leq ... \leq X_{n,n}$ la statistique d'ordre $i-^{iéme}$.*

En posant

$$W_n = X_{1,n} = \min(X_1, X_2, ..., X_n), \qquad M_n = X_{n,n} = \max(X_1, X_2, ..., X_n)$$

Les variables W_n *et* M_n *définissent les statistique d'ordre extrême et leur écart* $D_n = M_n - W_n$ *est dite déviation extrême. On vérifie la relation :*

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

Proposition 2.1.1 *Les distribution* $F_{_M}$ *et* $F_{_W}$ *des statistiques d'ordres extrêmes* W_n *et* M_n *sont données respectivement par :* $\forall x \in \mathbb{R}$

* $F_{_{M_n}}(x) = [F(x)]^n$ et $f_{M_n}(x) = n[F(x)]^{n-1} f(x)$ Distribution du *maximum.*
* $F_{_{W_n}}(x) = 1 - [1 - F(x)]^n$ et $f_{W_n}(x) = n[1 - F(x)]^{n-1} f(x)$ Pour *minimum*

Définition 2.1.2 *(Le point terminal) On appelle le point terminal ou le point le plus à droite de la fonction de distribution* F*, noté* x_F *la borne supérieure du support de* F *définit par :*

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

Définition 2.1.3 *L'inverse généralisé* $F(y)$ *d'une fonction* F *est définie par :*

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

Proposition 2.1.2 *La suite des maximums* $\{M_n = X_{n,n} = \max(X_1, X_2, ..., X_n), n \geq 1\}$ *converge presque surement vers* x_F *quand* $n \to \infty$ *,i.e*

$$M_n \overset{p.s}{\to} x_F \qquad quand\ n \to \infty$$

Corollaire 2.1.1 *La suite des maximums* $\{M_n, n \in \mathbb{N}^*\}$ *converge en loi vers une variable* *aléatoire dégénérée concentrée en x_F, car :*

$$\lim_{n\to\infty} F_{M_n}(x) = \lim_{n\to\infty} \mathbb{P}\{M_n \leq x\} = \lim_{n\to\infty} F^n(x) = \begin{cases} 0 & si \quad x < x_F \\ \\ 1 & si \quad x \geq x_F \end{cases}$$

2.2 Distributions des valeurs extrêmes

2.2.1 Limites des maximums renormalisés

Définition 2.2.1 *les suites $\{a_n > 0, n \geq 1\}$ et $\{b_n, n \geq 1\}$ sont appelées suites de normalisation, les constantes $a_n \in \mathbb{R}_+^*$ et $b_n \in \mathbb{R}$ sont appelées suites de constantes de normalisation et la variable aléatoire $[a_n^{-1}(M_n - b_n)]$ est appelée maximum renormalisé.*

Loi max-stable

Comme dans la théorie de la limite centrale, nous avons défini les lois stables qui sont les seules lois limites possibles pour la suite des somme normalisées de n variables aléatoires *i.i.d. quand $n \to \infty$.*

Définition 2.2.2 *La variables aléatoires non-dégénérée X ou la loi de probabilité de X ou la fonction de distribution F de X est dite* max-stable*, s'il existe des constantes $a_n \in \mathbb{R}_+^*$ et $b_n \in \mathbb{R}$ telle que :*

$$M_n \xrightarrow{d} a_n X + b_n \qquad n \in \mathbb{N}^* \Leftrightarrow F^n(a_n x + b_n) = F(x) \qquad n \in \mathbb{N}^* \text{ et } x \in \mathbb{R}$$

Exemple 2.2.1 *(Coles 2001) Supposons que X suit une loi de probabilité de Fréchet i.e :*

$$\forall x \in \mathbb{R}, \qquad F(x) = \mathbb{P}(X \leq x) = \begin{cases} 0, & x \leq 0 \\ \\ \exp(-x^{-1}) & x > 0 \end{cases}$$

Posons $a_n = n$ *et* $b_n = 0$*, on a alors :* $\forall x \in \mathbb{R}$, $\forall n \in \mathbb{N}^*$

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

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

$$= F(x)$$

Donc

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

Fisher et Tippett (1928) ont démontré l'existence des suites de normalisation $\{a_n\}$, $\{b_n\}$ $n \in \mathbb{N}$, avec $a_n > 0$ et $b_n \in \mathbb{R}$ et une loi non-dégénérée H telle que :

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{M_n - b_n}{a_n}\right\} = \lim_{n\to\infty} F^n(a_n x + b_n) \to H(x), \qquad x \in \mathbb{R}$$

convergence faible

$$\frac{M_n - b_n}{a_n} \xrightarrow{\mathcal{L}} H(x) \qquad quand\ n \to \infty$$

Le théorème suivant donne une caractérisation de la distribution limite du maximum M_n.

Théoème 2.2.1 (Fisher et Tippett (1928), Gnedenko (1943)*) S'il existe deux suites de constantes de normalisation* $a_n > 0$ *et* $b_n \in \mathbb{R}$ *,* $n \in \mathbb{N}$ *et une loi non-dégénérée de loi* H *telle que* $\dfrac{M_n - b_n}{a_n} \xrightarrow{\mathcal{L}} H$ *et s'il existe un réel* α*, alors* H *appartient à une des trois distributions standard des valeurs extrêmes (à un paramètre d'échelle et de location près) suivants :*

1. Type I : *Gumbel* : $H_0(x) = \Lambda(x) = \exp[-\exp(-x)], \quad x \in \mathbb{R}.$

2. Type II : *Fréchet* : $H_\alpha(x) = \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \\ \exp(-x^{-\alpha}) & x > 0 \end{cases}, \quad \alpha > 0$

3. Type III : *Weibull* : $H_\alpha(x) = \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^{\alpha}) & x \leq 0 \\ \\ 1 & x > 0 \end{cases} \quad \alpha < 0$

Avec $\Lambda(x)$, $\Phi_\alpha(x)$ et $\Psi_\alpha(x)$ sont les lois limites possibles pour le maximum s'appellent les distributions standard ou traditionnelle des valeurs extrêmes.

Preuve. Une preuve détaillée de ce théorème peut être trouvée dans *Embrechts* et *al (1997)*. ■

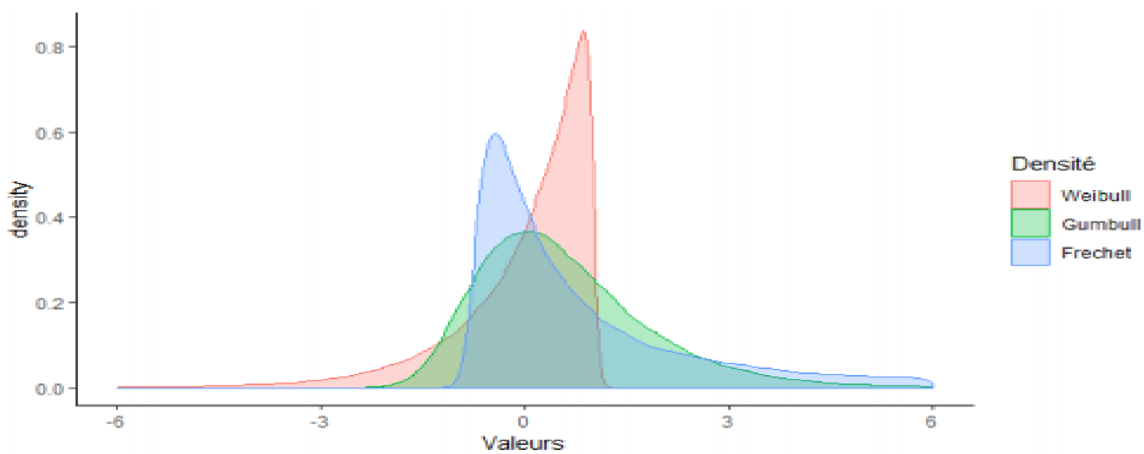

FIG. 2.1 – Densité de probabilité des Valeurs Extrêmes avec différentes valeurs de ζ.

2.2.2 Exemples de convergence du maximum renormalisé

Dans ce paragraphe, on considère des variables aléatoires $(X_n,\, n \geq 1)$ indépendantes de même loi, ainsi que leur maximum $M_n = \max(X_1, X_2, ..., X_n)$. On recherche des suites $(a_n,\, n \geq 0)$ et $(b_n,\, n \geq 1)$, avec $a_n > 0$, telles que la suite $(a_n^{-1}(M_n - b_n),\, n \geq 1)$ converge en loi vers une limite non dégénérée.

Nous considérons des variables de loi uniforme, exponentielle et de Cauchy.

Loi uniforme

a) On suppose que la loi de X_1 est la loi uniforme sur $[0, \theta]$, $\theta > 0$. La fonction de répartition de la loi est $F(x) = x/\theta$ pour $x \in [0, \theta]$, la suite $(M_n,\, n \geq 1)$ converge $p.s.$ vers θ .

Lemme 2.2.1 *La suite* $\left(n\left(\frac{M_n}{\theta} - 1\right); n \geq 1\right)$ *converge en loi vers* W *de fonction de répartition définie par*

$$\mathbb{P}(W \leq x) = e^x \qquad , x \leq 0 :$$

La loi de W est une loi de *Weibull*. Dans ce cas particulier, la loi de $-W$ est la loi exponentielle de paramètre 1.

Preuve. On note F_n la fonction de répartition de $n\left(\frac{M_n}{\theta} - 1\right)$. Comme $M_n < \theta$, on a $F_n(x) = 1$ si $x \geq 0$. Considérons le cas $x < 0$:

$$F_n(x) = \mathbb{P}\left(M_n \leq \theta + \theta\frac{x}{n}\right) = \mathbb{P}\left(X_1 \leq \theta + \theta\frac{x}{n}\right) = \left(1 + \frac{x}{n}\right)^n.$$

Il vient $\lim\limits_{n\to\infty} F_n(x) = e^x$ pour $x < 0$. On en déduit que $n\left(\frac{M_n}{\theta} - 1\right)$ converge en loi vers W de fonction de répartition $x \to min(e^x, 1)$. ■

Exemple 2.2.2 *(Coles 2001) Supposons que* X *suit une loi de probabilité uniforme* $\mathcal{U}[0,1]$ *i.e :*

$$\forall x \in \mathbb{R}, \qquad F(x) = \mathbb{P}(X \leq x) = \begin{cases} 0 & si \quad x < 0 \\ x & si \quad 0 \leq x < 1 \\ 1 & si \quad x \geq 1 \end{cases}$$

Si nous posons $a_n = n^{-1}$ *et* $b_n = 1$*, nous aurons :*

$$\forall x \in \mathbb{R}, \forall n \in \mathbb{N}^* \qquad \mathbb{P}\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = \mathbb{P}\left\{M_n \leq a_n x + b_n\right\} = F^n(1 + \frac{x}{n})$$

$$\mathbb{P}\left\{\frac{M_n-b_n}{a_n}\leq x\right\}=\begin{cases}0 & si \quad 1+\dfrac{x}{n}<0\\ \\ \left(1+\dfrac{x}{n}\right)^n & si \quad 0\leq 1+\dfrac{x}{n}<1\\ \\ 1 & si \quad 1+\dfrac{x}{n}\geq 1\end{cases}=\begin{cases}0 & si \quad x<-n\\ \\ \left(1+\dfrac{x}{n}\right)^n & si \quad -n\leq x<0\\ \\ 1 & si \quad x\geq 0\end{cases}$$

Car $\forall x\in\mathbb{R}\ \underset{n\to\infty}{lim}\left(1+\dfrac{x}{n}\right)^n=e^x$ *. Donc :*

$$\forall x\in\mathbb{R}\qquad \underset{n\to\infty}{lim}\,\mathbb{P}\left\{\frac{M_n-b_n}{a_n}\leq x\right\}=\begin{cases}e^x & si \quad x\leq 0\\ \\ 1 & si \quad x>1\end{cases}=\Psi_1(x)$$

On déduit que la loi uniforme $\mathcal{U}\,[0,1]$ *appartient au max domaine d'attraction de* $\Psi_1(x)$.

Loi exponentielle

On suppose X_1 de loi exponentielle de paramètre $\lambda>0$. La fonction de répartition de cette loi est $F(x)=1-e^{-\lambda x}$ pour $x\geq 0$. Comme $x_F=+\infty$, la suite $(M_n,n\geq 1)$ diverge vers l'infini.

Lemme 2.2.2 *La suite* $(\lambda M_n-\log(n),n\geq 1)$ *converge en loi vers* G *de fonction de répartition définie par*

$$\mathbb{P}(G\leq x)=e^{-e^{-x}}\qquad \forall x\in\mathbb{R}$$

La loi de G *est la loi de* Gumbel.

Preuve. On note F_n la fonction de répartition de $\lambda M_n-\log(n)$. On a

$$F_n(x)=\mathbb{P}\left(\lambda M_n-\log(n)\leq x\right)=\mathbb{P}\left(M_n\leq\left(x+\log(n)\right)/\lambda\right)$$

$$=\mathbb{P}\left(X_1\leq\left(x+\log(n)\right)/\lambda\right)^n=\left(1+\frac{e^{-x}}{n}\right)^n.$$

Alors

$$\underset{n\to\infty}{lim} F_n(x) = e^{-e^{-x}} x \in \mathbb{R}$$

On en déduit que la suite $(\lambda M_n - \log(n), n \geq 1)$ converge en loi vers G, de fonction de répartition $x \to e^{-e^{-x}}$. ■

Exemple 2.2.3 *(Coles 2001) Supposons que X suit une loi de probabilité loi exponentielle standard $\mathcal{E}\,(1)$ i.e :*

$$\forall x \in \mathbb{R}, \qquad F(x) = \mathbb{P}\,(X \leq x) = \begin{cases} 1 - e^{-x} & si \quad x > 0 \\ 0 & si \quad x \leq 0 \end{cases}$$

Si nous posons $a_n = 1$ et $b_n = \ln n$, nous aurons :

$$\forall x \in \mathbb{R}, \forall n \in \mathbb{N}^* \qquad \mathbb{P}\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = \mathbb{P}\left\{M_n \leq a_n x + b_n\right\} = F^n(x + \ln n)$$

Alors :

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

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

Donc :

$$\forall x \in \mathbb{R} \qquad \underset{n\to\infty}{lim} \mathbb{P}\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = \exp\left\{-e^{-x}\right\} = \Lambda(x)$$

On déduit que la loi exponentielle standard $\mathcal{E}\,(1)$ appartient au max domaine d'attraction de $\Lambda(x)$.

Loi de Cauchy

On suppose que X_1 suit la loi de Cauchy (de paramètre $a = 1$). La densité de la loi est $f(x) = \frac{1}{\pi(1+x^2)}$. Comme le support de la densité est non borné, il est clair que la suite $(M_n, n \geq 1)$ diverge.

Lemme 2.2.3 *La suite $(\frac{\pi M_n}{n}, n \geq 1)$ converge en loi vers W de fonction de répartition définie par :*

$$\mathbb{P}(W \leq x) = e^{-1/x}, \qquad x > 0$$

La loi de W appartient à la famille des lois de Fréchet*.*

Preuve. On note F_n la fonction de répartition de $\frac{\pi M_n}{n}$. On a

$$F_n(x) = \mathbb{P}\left(M_n \leq \frac{nx}{\pi}\right) = \mathbb{P}\left(X_1 \leq \frac{nx}{\pi}\right)^n = \left(1 - \int_{\frac{nx}{\pi}}^{\infty} \frac{1}{\pi\left(1+y^2\right)} dy\right)^n$$

Pour $x > 0$, on a :

$$\int_{\frac{nx}{\pi}}^{\infty} \frac{1}{\pi\left(1+y^2\right)} dy = \int_{\frac{nx}{\pi}}^{\infty} \frac{1}{\pi y^2} dy + \int_{\frac{nx}{\pi}}^{\infty} \left[\frac{1}{\pi\left(1+y^2\right)} - \frac{1}{\pi y^2}\right] dy = \frac{1}{nx} + O((nx)^{-3}).$$

On a alors pour $x > 0$, $F_n(x) = (1 - \frac{1}{nx} + O((nx)^{-3})^n$. On en déduit que $\underset{n\to\infty}{lim} F_n(x) = e^{-1/x}$ pour $x > 0$. Ainsi la suite $(\pi M_n = n, n \geq 1)$ converge en loi vers W de fonction de répartition définie par

$$\mathbb{P}(W \leq x) = e^{-1/x}, \qquad x > 0$$

■

Proposition 2.2.1 *(Relation entre Λ, Φ_α et Ψ_α) Soit Y une variable aléatoire positive ($Y > 0$) alors :*

$$(Y \sim \Phi_\alpha) \Leftrightarrow (\ln Y^\alpha \sim \Lambda) \Leftrightarrow \left(-Y^{-1} \sim \Psi_\alpha\right)$$

2.2.3 Distribution des valeurs extrêmes généralisée (GEV)

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

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

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

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

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

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

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

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

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

3. $\xi < 0$, distribution *bornée* (distribution de *Weibull*).

Proposition 2.2.2 *On a, les relations suivantes entre Λ, Φ_ξ et Ψ_ξ en terme de $H_{\xi,\mu,\sigma}$*

$$H_{1,\frac{1}{\xi},\frac{1}{\xi}}\left(\xi\left(x-1\right)\right)=\Phi_\xi\left(x\right)\qquad si\quad \xi>0$$

$$H_{0,\frac{1}{\xi},-\frac{1}{\xi}}\left(\xi\left(x+1\right)\right)=\Psi_\xi\left(x\right)\quad si\quad \xi<0$$

$$H_{0,1,0}\left(x\right)=\Lambda\left(x\right)\qquad\qquad si\quad \xi=0$$

2.3 Domaines d'attraction

Définition 2.3.1 *On dit qu'une distribution F appartient au domaine d'attraction de H, et on note $F\in\mathcal{DA}(H)$ si la distribution du maximum renormalisée converge vers H. Autrement dit, s'il existe des constants réelles $a_n>0$ et b_n tels que :*

$$\forall x\in\mathbb{R}\qquad \lim_{n\to\infty}\mathbb{P}\left\{\frac{M_n-b_n}{a_n}\leq x\right\}=\lim_{n\to\infty}F^n(a_nx+b_n)=H_\zeta(x)=\exp(-(1+\zeta x)^{-1/\zeta})$$

Pour tout x avec $1+\zeta x>0$.

Proposition 2.3.1 *(*Caractérisation de $\mathcal{DA}(H)$*) La fonction de distribution F appartient au domaine d'attraction de la distribution de valeur extrême H avec des constantes de normalisation a_n et $b_n\in\mathbb{R}$ si et seulement si :*

$$\forall x\in\mathbb{R}\qquad \lim_{n\to\infty}n\left[1-F(a_nx+b_n)\right]=\lim_{n\to\infty}n\bar{F}(a_nx+b_n)=-\ln H(x)$$

Lorsque $H(x)=0$ la limite est interprétée comme ∞ ($x\to-\ln x$ quand $x\to 0$ qui égale à ∞).

Voici quelques lois et leurs domaines d'attraction :

Domaine d'attraction	Fréchet $\xi > 0$	Gumbel $\xi = 0$	Weibull $\xi < 0$
Loi	Burr	Gamma	Uniforme
	Student	Normale	Reverse Burr
	Log gamma	Exponentielle	Beta
	Chi-deux	Lognormale	
	Pareto	Weibull	
	Cauchy	Logistique	

TAB. 2.1 – Domaines d'attraction des lois usuelles.

Exemple 2.3.1 *(Resnick* 1987*) Supposons que X suit une loi de probabilité de* Paréto, *i.e :*

$$\forall x \in \mathbb{R}, \qquad 1 - F(x) \;=\; \begin{cases} x^{-\alpha} & si \quad x > 1 \\ & \\ 1 & si \quad x \leq 1 \end{cases} \qquad avec \quad \alpha > 0$$

Si nous posons $a_n = n^{1/\alpha}$ *et* $b_n = 0$*, nous aurons :*

$$\forall x \in \mathbb{R}, \forall n \in \mathbb{N}^*, \qquad n\left[1 - F(a_n x + b_n)\right] = n\left[1 - F(n^{1/\alpha} x)\right] = \begin{cases} x^{-\alpha} & si \quad x > n^{-1/\alpha} \\ n & si \quad x \leq n^{-1/\alpha} \end{cases}$$

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

$$\lim_{n\to\infty} n\left[1 - F(a_n x + b_n)\right] \;=\; \begin{cases} x^{-\alpha} & si \quad x > 0 \\ & \\ \infty & si \quad x \leq 0 \end{cases}$$

$$= \begin{cases} -\ln\left[\exp\left(-x^{-\alpha}\right)\right] & si \quad x > 0 \\ & \\ -\ln 0 & si \quad x \leq 0 \end{cases}$$

$$= -\ln \Phi_\alpha(x)$$

Donc : $F \in \mathcal{DA}(\Phi_\alpha)$

Fonction à variations réguliéres Une fonction U positive est à variations régulières d'indice $\delta \in \mathbb{R}$ à l'infini, et on note : $U(.) \in RV_\delta$, si :

$$\lim_{x\to\infty} \frac{U(\lambda x)}{U(x)} = \lambda^\delta, \qquad \forall \lambda > 0$$

Dans le cas $\delta = 0$ on dit que la fonction $U(.)$ est à variations lentes, et on note : $U \in RV_0$. Toute fonction à variations régulières d'indice $\delta \in \mathbb{R}$, peut s'écrit :

$$U(x) = x^\delta L(x), \qquad L(x) \in RV_0$$

Lemme 2.3.1 Resnick (1987) *Soit U est une fonction à variations régulières d'indice, alors :*

$$\lim_{x\to\infty} \sup_{\lambda\in[a,b]} \left| \frac{U(\lambda x)}{U(x)} - \lambda^\delta \right| = 0, \quad \textit{pour tout } 0 < a < b$$

Lemme 2.3.2 Resnick (1987) *Si U est à variations régulières d'indice $\delta > 0$, alors $U^{\leftarrow}(x)$ est à variations régulières d'indice $1/\delta$. Si U est à variations régulières d'indice $\delta < 0$, alors $U^{\leftarrow}(x)$ est à variations régulières d'indice $-1/\delta$.*

Théoème 2.3.1 (Représentation de Karamata) *Toutes fonction à variations lentes $L(.)$ s'écrive sous la forme :*

$$L(x) = c(x) \exp\left\{ \int_1^x \frac{\Delta(u)}{u} du \right\} \quad c(x) \longrightarrow c > 0 \quad \textit{et } \Delta(x) \longrightarrow 0 \quad \textit{lorsque } x \longrightarrow \infty.$$

Cette représentation des fonctions à variations lentes est connue sous le nom de représentation de Karamata Bingham et al. (1987). *Si la fonction $c(.)$ est constante, la fonction $L(.)$ est dite normalisée.*

2.3.1 Caractérisation des domaines d'attraction

Domaine d'attraction de Fréchet

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

Théoème 2.3.2 *Soit $F \in \mathcal{DA}(\phi_\zeta)$ avec $\zeta > 0$ si et seulement si $x_F = +\infty$ et $1 - F$ est une fonction à variations régulière d'indice $-1/\zeta$ (i.e. $1 - F(x) = x^{-1/\zeta} l(x)$, où l est une fonction à variations lentes). Dans ce cas, un choix possible pour les suites a_n et b_n est $a_n = F^{\leftarrow}(1 - \frac{1}{n})$ et $b_n = 0$.*

Preuve. Voir *Embrechts et al (1997)*, Théorème 3.3.7, page 131. ■

Domaine d'attraction de Weibull

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

Théoème 2.3.3 *Soit $F \in \mathcal{DA}(\Psi_\zeta)$ avec $\zeta < 0$ si et seulement si $x_F < +\infty$ et $1 - F^*$ est une fonction à variations régulière d'indice $1/\zeta$ (i.e. $1 - F(x) = (x_F - x)^{1/\zeta}[l(x_F - x)^{-1}]$) avec :*

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

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

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

Domaine d'attraction de Gumbel

Ce domaine d'attraction regroupe la majorité des distributions à queue fine, par exemple loi normale, exponentielle, gamma, lognormale, etc...

Théoème 2.3.4 *Soit $F \in \mathcal{DA}(\Lambda_\varsigma)$ si et seulement si il existe $t < x_F \leq +\infty$ tel que :*

$$\bar{F}(x) = c(x) \exp\left\{-\int_t^x \frac{1}{a(u)} du\right\}, \quad t < x < x_F$$

où $\lim_{x \to x_F} c(x) = c > 0$ et $a(.)$ est une fonction positive et dérivable de dérivée $a\prime(.)$ telle que : $\lim_{x \to x_F} a'(x) = 0$, dans ce cas, un choix possible pour les suites (a_n) et (b_n) $\forall n > 0$ est :

$$a_n = q(1/n) \qquad b_n = \tfrac{1}{\bar{F}(a_n)} \int_{a_n}^{x_F} \bar{F}(s) ds.$$

Exemple 2.3.2 *La fonction de répartition de la loi exponentielle de paramètre $\lambda > 0$ est :*

$$F(x) = 1 - exp(-\lambda x), \quad 0 \leq x \leq \infty.$$

On pose $a_n = \frac{1}{\lambda}$ et $b_n = \frac{1}{\lambda} \ln(n)$

$$F^n(a_n x + b_n) = (1 - \exp(-\frac{1}{\lambda}\lambda x - \lambda \frac{1}{\lambda} \ln(n))^n = (1 - \frac{\exp(-x)}{n})^n \to \exp(-\exp(-x)) = \Lambda(x).$$

D'où le maximum normalisé de la loi exponentielle converge vers la loi de *Gumbel*, et $F(x)$ appartient au domaine d'attraction de *Gumbel*.

2.3.2 Loi des excès, approche POT

Soit u un réel suffisamment élevé appelé seuil. On définit les excès au-delà du seuil u comme l'ensemble des variables aléatoires conditionnelles $\{Y_j\} = \{X_j - u \mid X_j > u\}$. La loi conditionnelle des excès F_u par rapport au seuil u est :

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

Ou de manière équivalente :

$$\bar{\bar{F}}_u(y) = \mathbb{P}(X - u > y \mid X > u) = 1 - F_u(y) = \frac{\bar{\bar{F}}(u+y)}{\bar{\bar{F}}(u)}, \quad y \geq 0$$

Théoème 2.3.5 *Une fonction de répartition F appartient au domaine d'attraction maximale de H_ζ, si et seulement si, il existe une fonction positive $\beta(u)$ telle que :*

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

où $F_u(y)$ est la fonction de répartition conditionnelle des excès *au delà d'un seuil u, x_F est le point terminal de F et $G_{\zeta,\beta(u)}(y)$ est la* GPD *donnée par :*

$$G_{\zeta,\beta(u)}(y) = \begin{cases} 1 - (1 + \zeta \frac{y}{\beta(u)})^{-1/\zeta} & \zeta \neq 0 \\ \\ 1 - \exp(-\frac{y}{\beta(u)}) & \zeta = 0 \end{cases}$$

Où $y \geq 0$ pour $\zeta \geq 0$ et $y \leq 0 \leq -\frac{\beta(u)}{\zeta}$ pour $\zeta < 0$. On conclut de ce théorème, que si F vérifie le théorème de *Fisher* et *Tippet*, alors il existe une fonction $\beta(.)$ positive et un réel tel que la loi des excès F_u peut être uniformément approchée par une *distribution de Pareto généralisée* (GPD) notée $G_{\zeta,\beta(u)}$. Ainsi, l'indice des valeurs extrêmes donnée par le théorème de *Fisher* et *Tippet* est le même que celui de la *loi des excès.*

Exemple 2.3.3 *Pour le cas de la loi de* Pareto*, la fonction de répartition est :*

$$F(x) = 1 - cx^{-\alpha}, \quad \textit{avec} \quad c > 0 \textit{ et } \alpha > 0,$$

On prend $\beta_u = ub$, avec $b > 0$ et $y \geq 0$, on obtient :

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

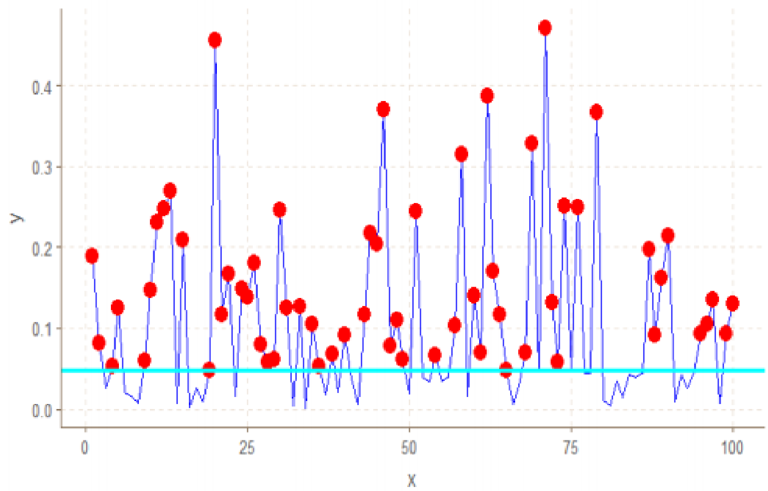

FIG. 2.2 – Méthode des excès au dessus d'un seuil.

D'où, pour $\zeta = 1/\alpha$ et $b = \zeta$ la loi limites est la loi *GPD* de paramètre ζ.

Exemple 2.3.4 *Pour la loi exponentielle de paramètre $\lambda = 1$, la fonction de répartition est $F(x) = 1 - \exp(-x)$, $x \geq 0$: On prend $\beta_u = 1$; alors :*

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{\exp(-u) - \exp(-u-y)}{\exp(-u)} = 1 - \exp(-y).$$

Donc la loi limite est la loi GPD *de paramètre $\zeta = 0$.*

T.D. N° – 2

Exercice-1

Soit (U_n) une suite de variables aléatoires indépendantes suivant toutes la loi *uniforme* sur $[0, 1]$. On note $M_n = \max(U_1, \ldots, U_n)$ et $X_n = n(1 - M_n)$.

1. Quelle est la fonction de répartition de X_n ?
2. Étudier la convergence en loi de la suite (X_n).

Exercice-2 (Domaine d'attraction de Fréchet)

On considère la fonction $F : \mathbb{R} \to \mathbb{R}$ définie pour $\theta > 0$ et $\lambda > 0$ par :

$$F(x) = \left(1 - (1 + x^\theta)^{-\lambda}\right) \mathbb{I}_{\{x>0\}}.$$

1. Montrer que $F(.)$ est une fonction de répartition et donner son point terminal.
2. Montrer que $F(.)$ appartient au domaine d'attraction de *Fréchet* et donner, en fonction de θ et λ , l'indice des valeurs extrêmes γ associé.
3. On pose $L : [0, \infty[\to [0, \infty[$ la fonction définie par $L(x) = x^{-1/\gamma}(1 - F(x))$ pour tout $x > 0$. Quelle est la principale caractéristique de la fonction $L(.)$?
4. On pose $\Delta : [0, \infty[\to [0, \infty[$ la fonction définie par $\Delta(x) = xL^{'}(x)/L(x)$ pour tout $x > 0$ où $L^{'}(.)$ désigne la dérivée de la fonction $L(.)$. Montrer que $\Delta(.)$ est une fonction à variations régulières dont vous préciserez l'indice.
5. On considère à présent n variables aléatoires indépendantes $X_1, ..., X_n$ de même fonction de répartition $F(.)$ et on pose $a_n = F^{\leftarrow}(1 - 1/n)$. Déterminer la loi limite

de la variable aléatoire

$$Y_n := \max\left\{\frac{X_1}{a_n}, ..., \frac{X_n}{a_n}\right\}$$

Expliquer pourquoi on retrouve ainsi le résultat de la question 2 ?

Exercice-3 (Loi de Gumbel)

On considère la fonction g définie pour tout réel x par :

$$g(x) = e^{-e^{-x}}$$

1. Calculer ses limites en $-\infty$ et $+\infty$, sa dérivée.
2. Vérifier que la fonction f définie pour tout réel x par $g(x) = e^{-x-e^{-x}}$ est une densité.
3. Soit X une variable aléatoire de loi exponentielle de paramètre 1. Rappeler ce que vaut la fonction de répartition F de X.
4. Soit X_1 et X_2 des variables aléatoires $i.i.d$ de loi exponentielle de paramètre 1, et soit $M = \max(X_1, X_2)$ la variable aléatoire correspondant au maximum de ces deux variables. Pour tout réel x, calculer $P(M \leq x)$. En déduire la densité de M.
5. On note maintenant $M_n = \max(X_1, ..., X_n)$, où $X_1, ..., X_n$ sont variables aléatoires $i.i.d$ de loi exponentielle de paramètre 1. Pour tout réel x, calculer $F_n(x) = P(M_n \leq x)$.
6. Soit u un réel fixé, que vaut $\underset{n\to+\infty}{lim}(1 - \frac{u}{n})^n$? En déduire que pour tout réel x

$$\underset{n\to+\infty}{lim} F_n(x + \ln n) = g(x).$$

Exercice-4 (Loi de Weibull)

On considère une variable aléatoire X de densité

$$f(x) = \begin{cases} 3x^2e^{-x^3} & si \quad x \geq 0 \\ \\ 0 & si \quad x < 0 \end{cases}$$

1. Vérifier que f est bien une densité de probabilité. On dit que X suit une loi de *Weibull*.
2. Calculer la dérivée de f. En déduire le mode de X.
3. Déterminer la fonction de répartition F de X.

Exercice-5 (Loi de *Pareto*)

Montrer que si X suit une loi de *Pareto* dont le paramètre de seuil a est égal à 1, alors $\ln X$ suit une loi exponentielle.

Chapitre 3

Estimation de l'indice des valeurs extrêmes et de quantiles extrêmes

Dans cette section nous nous intéressons aux différentes méthodes d'estimation du paramètre ζ ou σ qui intervienent dans la distribution asymptotique des valeurs extrêmes. On trouve dans la littérature de la théorie des valeurs extrêmes plusieurs méthodes d'estimation.

3.1 Estimation de l'indice des queues

L'estimation de l'indice des queues, joue un rôle important pour déterminer une loi extrême. Pour cela on a deux approches : paramétrique et semi-paramétrique. Dans le cas paramétrique, les méthodes d'estimations les plus utilisées sont la méthode du maximum de vraisemblance (MV) et la méthode des moments (MM). Pour l'approche semi-paramétrique on trouve plusieurs techniques, on cite les plus célébre : l'estimateur de [*Hill(1975)*] et l'estimateur de [*Pickands(1975)*].

3.1.1 Estimateur du maximum de vraisemblance E.M.V

L'estimateur du maximum de vraisemblance est construit à partir des observations des maxima, il s'agit d'estimer l'indice des valeurs extrêmes ainsi que les deux suites a_n et b_n.

Définition 3.1.1 *Soit $(Y_1, ..., Y_n)$ un n-échantillon, les Y_i sont supposées i.i.d, de denstié h_θ où $\theta = (\mu, \sigma, \zeta)$. L'expression de la fonction de vraisemblance est donnée par :*

$$L(\theta = (\mu, \sigma, \zeta), (Y_1, ..., Y_n)) = \prod_{i=1}^{n} h_\theta(y_i)$$

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

$$\frac{\partial \log L}{\partial \theta} = 0, \qquad \frac{\partial^2 \log L}{\partial^2 \theta} < 0$$

Exemple 3.1.1 *Dans le cas où $\zeta = 0$ (Loi de Gumbel), la fonction de* log-vraisemblance *égale à :*

$$\log L(\theta = (\mu, \sigma, \zeta); (Y_1, ..., Y_n)) = \prod_{i=1}^{n} \exp(-\exp(-\frac{y_i - \mu}{\sigma}))$$

$$= -n \log \sigma - \sum_{i=1}^{n} \exp(-\frac{y_i - \mu}{\sigma}) - \sum_{i=1}^{n} \frac{y_i - \mu}{\sigma}$$

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

$$\begin{cases} \dfrac{\partial \log L}{\partial \sigma} = 0 \Leftrightarrow \quad n + \displaystyle\sum_{i=1}^{n} (\frac{y_i - \mu}{\sigma}) \left[\exp(-\frac{y_i - \mu}{\sigma} - 1) \right] = 0 \\ \\ \dfrac{\partial \log L}{\partial \mu} = 0 \Leftrightarrow \quad n - \displaystyle\sum_{i=1}^{n} \exp(-\frac{y_i - \mu}{\sigma}) = 0 \end{cases}$$

La résolution de ce système est relativement difficile et n'admet pas en général de solutions *explicites. Dans ce cas, on fait appel à des méthodes d'optimisation numériques.*

Soit un échantillon de k maxima $Y_1, ..., Y_k$ *i.i.d.* suivant une loi *GEV*. Pour $\zeta \neq 0$, la fonction de *log-vraisemblance* est donnée par :

$$L(Y, \zeta, a_n, b_n) = -k \log a_n - \left(\frac{1}{\zeta} + 1\right) \sum_{i=1}^{k} \log\left(1 + \zeta\left(\frac{y_i - b_n}{a_n}\right)\right) - \sum_{i=1}^{k} \left(1 + \zeta\left(\frac{y_i - b_n}{a_n}\right)\right)^{-1/\zeta}$$

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

$$L(Y, 0, a_n, b_n) = -k \log a_n - \sum_{i=1}^{k} \exp\left(\frac{y_i - b_n}{a_n}\right) - \sum_{i=1}^{k} \left(\frac{y_i - b_n}{a_n}\right).$$

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

$$\sqrt{m}\left(\left(\hat{a}_n, \hat{\zeta}, \hat{b}_n\right) - (a_n, \zeta, b_n)\right) \to \mathcal{N}\left(0, I^{-1}\right),$$

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

$$I(\Theta) = -\mathbb{E}\left(\frac{\partial^2 L(Y;\Theta)}{\partial^2 \Theta}\right).$$

$L(Y;\Theta)$ est la fonction de log-vraisemblance associée à la loi de la variable aléatoire Y, paramétrée par un ensemble de paramètres Θ.

3.1.2 Estimateur de Hill

L'estimateur le plus connu pour l'*IVE* positif est l'estimateur proposé par *Hill (1975).*

Définition 3.1.2 *Soient $X_{1,n}, ..., X_{n,n}$ les statistiques d'ordre associées à l'échantillon $X_1, ..., X_n$. On note par $X_{n-k,n}$ avec $1 \leq k < n$, la $k-$ième plus grande valeur de l'échantillon. On* *introduit la notation :*

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

L'estimateur de Hill *est défini pour $\zeta > 0$ par :*

$$\hat{\zeta}_{X,k,n}^{(H)} = \log X_{n-i+1,n} - \log X_{n-k,n} = M_n^{(1)}$$

Théoème 3.1.1 *(Propriétés asymptotiques de $\hat{\zeta}_n^{(H)}$) Supposons que $F \in \mathcal{DA}(\Phi_\zeta)$, $\zeta > 0$, $1 \le k < n$, $k \to \infty$ et $k/n \to 0$ quand $n \to \infty$:*

1. Convergence faible :

$$\hat{\zeta}_n^{(H)} \xrightarrow{P} \zeta \qquad \textit{quand} \qquad n \to \infty.$$

2. Convergence forte : Si $k/\log\log n \to \infty$ et $k_n/n \to 0$

$$\hat{\zeta}_n^{(H)} \xrightarrow{P.s} \zeta \qquad \textit{quand} \qquad n \to \infty.$$

3. Normalité asymptotique :*Hill(1975)* a démontrer la normalité asymptotique suivante :

$$\sqrt{k}\left(\hat{\zeta}_{X,k,n}^{(H)} - \zeta\right) \to N\left(0, \zeta^2\right)$$

La propriété de consistance forte a été démontrée par *P.Deheuvels(1988)*.

3.1.3 Estimateur de Pickands

L'estimateur introduit par *Pickands(1975)* est défini pour tout $\zeta \in \mathbb{R}$ par :

$$\hat{\zeta}_{X,k,n}^{(P)} = \frac{1}{\log 2} \log \frac{X_{n-k+1,n} - X_{n-2k+1,n}}{X_{n-2k+1,n} - X_{n-4k+1,n}}$$

Théoème 3.1.2 *(Propriétés asymptotiques de $\hat{\zeta}_n^{(P)}$) Supposons que $F \in \mathcal{DA}(H_\zeta)$, $\zeta \in \mathbb{R}$,*

$k \to \infty$ *et* $k/n \to 0$ *quand* $n \to \infty$ *:*

1. Convergence faible :

$$\hat{\zeta}_n^{(P)} \xrightarrow{P} \zeta \qquad quand \qquad n \to \infty.$$

2. Convergence forte : Si $k/\log\log n \to \infty$ et $k_n/n \to 0$

$$\hat{\zeta}_n^{(P)} \xrightarrow{P.s} \zeta \qquad quand \qquad n \to \infty.$$

3. Normalité asymptotique a été établie par [*Dekkers et De Haan(1989)*] comme suit :

$$\sqrt{k}\left(\hat{\zeta}_{X,k,n}^{(P)} - \zeta\right) \to \mathcal{N}\left(0, \frac{\zeta^2\left(2^{\zeta+1}+1\right)}{\left(2\left(2^{\zeta}-1\right)\log 2\right)^2}\right),$$

lorsque n et k tendent vers l'infini.

3.1.4 Estimateur des moments

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

$$\hat{\zeta}_{X,k,n}^{(M)} = \hat{\zeta}_{X,k,n}^{(H)} + 1 - \frac{1}{2}\left(1 - \frac{\hat{\zeta}_{X,k,n}^{(H)}}{M_n^{(2)}}\right)^{-1}.$$

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

$$\sqrt{k}\left(\hat{\zeta}_{X,k,n}^{(M)} - \zeta\right) \to \mathcal{N}\left(0, 1+\zeta^2\right).$$

3.1.5 Estimateur de type noyau

Csörgö, Deheuvels et *Mason* en *1985* ont proposé une version plus lisse pour l'estimateur de Hill (dénotée par $\hat{\zeta}_{n,h}^{(K)}$) et prouvaient sa consistance et normalité asymptotique. Cet estimateur utilisé seulement dans le cas où $\zeta > 0$.

$$\hat{\zeta}_{n,h}^{(K)} = \sum_{i=1}^{n-1} \left(\frac{i}{nh}\right) K\left(\frac{i}{nh}\right) \log\left(X_{n-i+1,n}/X_{n-i,n}\right) / \int_0^{1/h} K(u)\, du.$$

tel que $\alpha > 0$, $h > 0$ et $K(u) := h^{-1}K(u/h)$, avec $u \in]0,1]$ et $h > 0$ est appelé paramètre de lissage ou fenêtre.

Où $K(.)$ une fonction noyau vérifiant les conditions suivantes :

a $K(s) \geq 0$ pour $0 < s < \infty$.

b est une fonction non-négative et continue à droit dans $]0, \infty[$.

c $\int_0^1 K(u)du = 1$

d $\int_0^1 u^{-1/2}K(u)du < \infty$

Dans l'estimateur de type noyau le paramètre de lissage h joue un rôle semblable comme nombre de statistiques d'ordre k dans les estimateurs mentionnés auparavant (h représente la proportion de statistiques de l'ordre supérieur utilisée).

3.2 Estimation de quantiles extrêmes

Dans ce qui suit, on va supposer que F appartient à l'un des domaines d'attractions définis précédemment, afin d'étudier le problème d'estimation de quantile extrême. Pour cela, on a besoin de faire un rappel de quelques définitions.

Définition 3.2.1 *On appelle quantile d'ordre p, le nombre x_p définie par :*

$$x_p = \inf\{x, F(x) \geq p\}, \quad p \in [0,1].$$

Définition 3.2.2 *L'inverse généralisé $F(y)$ d'une fonction F est définie par :*

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

Définition 3.2.3 *La fonction quantile de queue est définie par :*

$$U(t) = F^{\leftarrow}(1 - \frac{1}{t}), \qquad t \in]1, \infty[.$$

$F^{\leftarrow}$ est l'inverse généralisé de F.

Définition 3.2.4 *Le quantile extrême d'ordre $\alpha = \alpha_n \in (0,1)$ avec $\alpha_n \to 0$ quand $n \to \infty$. est défini par :*

$$q_{\alpha_n} = \bar{F}^{\leftarrow}(\alpha_n) = \inf\left\{x : \bar{F}(x) \leq \alpha_n\right\} = F^{\leftarrow}(1-\alpha_n) = Q(1-\alpha_n) = U(1/\alpha_n).$$

Définition 3.2.5 *(*Quantile empirique*) Soit la statistique d'ordre $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$. On appelle quantile empirique d'ordre p, la variable aléatoire notée $\hat{X}_{(np,n)}$ définie par :*

$$\hat{X}_{(np,n)} = \begin{cases} \dfrac{X_{(np,n)} + X_{(np+1,n)}}{2}, & si \quad np \in \mathbb{N} \\ \\ X_{([np]+1,n)}, & si\ non. \end{cases}$$

$[np]$ est la partie entière de np.

Lemme 3.2.1 Embrechts et all(1997) *Si $\alpha_n \to 0$ et $n\alpha_n \to c$ (non nécessairement fini) quand $n \to \infty$, alors,*

$$\mathbb{P}(X_{n,n} \leq q_{\alpha_n}) \to e^{-c}.$$

Plus généralement si la taille de l'échantillon tend vers l'infini, nous avons :

$$\begin{aligned} \mathbb{P}(X_{n,n} \leq q_{\alpha_n}) &= F^n(q_{\alpha_n}) = (1-\alpha_n)^n \\ &= \exp(n \log(1-\alpha_n)) \\ &= \exp(-n\alpha_n(1+o(1))) \qquad \text{quand } \alpha_n \to 0. \end{aligned}$$

Ainsi, au niveau du dernier terme pour estimer les quantiles extrêmes, deux cas sont envisagés pour q_{α_n} (au sein et dehors de l'échantillon) en fonction de la vitesse de convergence de α_n vers zéro :

D'après le *Lemme 3.2.1* , la probabilité que le quantile extrême soit plus grand que le maximum de l'échantillon dépend du comportement asymptotique de c. Ainsi, lorsque l'on souhaite estimer des quantiles extrêmes d'ordre $\alpha_n \to 0$ quand $n \to \infty$, deux cas doivent être distinguées en fonction de c :

1. Dans la première situation, cas, si $c = \infty$ donc,

$$P(X_{n,n} \leq q_{\alpha_n}) = 0.$$

   Un estimateur naturel de q_{α_n} est le quantile empirique $[n\alpha_n]$ *ième* plus grande observation de l'échantillon $\{X_i, i = 1, ..., n\}$.

2. Dans la deuxième situation, si $c = 0$ alors,

$$\mathbb{P}(X_{n,n} \leq q_{\alpha_n}) = 1,$$

   donc on ne peut pas estimer le quantile q_{α_n} de manière empirique.

Des différentes méthodes d'estimation du quantile extrême ont été proposées dans la littérature. On distingue trois catégories principales de méthodes d'estimation.

3.2.1 L'approche par la loi des valeurs extrêmes

Cette approche a été établit par *Guida et Longo (1988)* dans le but d'estimer les quantiles extrêmes.

D'après le théorème de *Fisher-Tippett*, on a :

$$\lim_{n\to\infty} \mathbb{P}\left\{\frac{X_{n,n} - b_n}{a_n} \leq x\right\} = \lim_{n\to\infty} F^n(a_n x + b_n) \approx H_\zeta(x) = \exp(-(1 + \zeta x)^{-1/\zeta})$$

Cette équation est équivalent à

$$\lim_{n\to\infty} n \log F(a_n x + b_n) = \lim_{n\to\infty} n \log \left(1 - \bar{F}(a_n x + b_n)\right) = \log H_\zeta(x).$$

Cela nous donne

$$\mathbb{P}\left\{X_{n,n} \leq z\right\} \approx H_{_\zeta}(\frac{z - b_n}{a_n}) = H_{\zeta, a_n, b_n}.$$

Donc la loi de $X_{n,n}$ peut être approchée par H_{ζ,a_n,b_n}.

Lorsque $n \to \infty$ on a $\bar{F}(a_n x + b_n) \to 0$, donc le développement limité de $\log\left(1 - \bar{F}(a_n x + b_n)\right)$ au premier ordre est :

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

$$\bar{F}(x) \approx -\frac{1}{n} \log H_{\zeta,a_n,b_n}(x).$$

$$\bar{F}(x) \approx \frac{1}{n}\left(1 + \zeta\left(\frac{x - b_n}{a_n}\right)^{-1/\zeta}\right),$$

d'où on peut approcher le quantile $q(\alpha)$ par :

$$q(\alpha) \simeq b_n + \frac{a_n}{\zeta}\left(\left(\frac{1}{n\alpha}\right)^{\zeta} - 1\right).$$

Définition 3.2.6 *L'estimateur du quantile extrême de la loi* GEV *est défini par :*

$$\hat{q}_n(\alpha_n) \simeq \hat{b}_n + \frac{\hat{a}_n}{\hat{\zeta}_n}\left(\left(\frac{1}{n\alpha_n}\right)^{\hat{\zeta}_n} - 1\right),$$

où $(\hat{a}_n, \hat{b}_n, \hat{\zeta}_n)$ sont respectivement des estimateurs des paramètres (a_n, b_n, ζ_n).

3.2.2 L'approche par la loi des excès

D'après l'équation :

$$\bar{\bar{F}}_u(y) = \Pr(X - u > y \mid X > u) = 1 - F_u(y) = \frac{\bar{\bar{F}}(u+y)}{\bar{\bar{F}}(u)}, \quad y \geq 0$$

on a :

$$\bar{\bar{F}}(u+y) = \bar{\bar{F}}(u)\bar{\bar{F}}_u(y),$$

avec un changement de variable $z = u + y$, on obtient :

$$\bar{\bar{F}}(z) = \bar{\bar{F}}(u)\bar{\bar{F}}_u(z-u) \approx \bar{\bar{F}}(u)\bar{G}_{\zeta,\beta(u)}(z-u),$$

où $\bar{G}_{\zeta,\beta(u)}$ est la *fonction de survie* de la loi de Pareto généralisée $G_{\zeta,\beta(u)}$. la probabilité p pour que X dépasse le seuil u est :

$$p = \mathbb{P}(X > u) = \bar{F}(u),$$

d'où :

$$\bar{\bar{F}}(z) \simeq p\bar{G}_{\zeta,\beta(u)}\left(z - \bar{F}^{\leftarrow}(p)\right),$$

Une approximation de la fonction de survie en queue est :

$$\bar{\bar{F}}(z) \simeq p\left(1 + \zeta\left(\frac{z - \bar{F}^{\leftarrow}(p)}{\beta(u)}\right)\right)^{-1/\zeta},$$

Pour estimer les quantiles extrêmes, on a besoin d'inverser la fonction de survie de l'équation, on obtient :

$$q(\alpha) = \bar{F}^{\leftarrow}(p) + \frac{\beta(u)}{\zeta}\left(\left(\frac{\alpha}{p}\right)^{-\zeta} - 1\right).$$

Donc il faut estimer les paramètres ζ, u et $\beta(u)$ pour pouvoir estimer les quantiles extrêmes.

3.2.3 L'approche semi-paramétrique

On s'intéresse au cas de domaine d'attraction de Fréchet. Rappelons que pour F appartient au *domaine d'attraction de Fréchet,* on a :

$$\bar{\bar{F}}(x) = x^{-1/\zeta} l\left(x\right),$$

où l est une fonction à variations lentes à l'infini et $\zeta > 0$.

$$q_{\alpha_n} = \bar{F}^{\leftarrow}\left(\alpha_n\right) = \alpha_n^{-\zeta} L\left(1/\alpha_n\right) \qquad \alpha_n \leq 1/n,$$

$$q_{\beta_n} = \bar{F}^{\leftarrow}\left(\beta_n\right) = \beta_n^{-\zeta} L\left(1/\beta_n\right) \qquad \beta_n \geq 1/n,$$

telle que L est une fonction à variations lentes à l'infini. Par division et pour β_n suffisamment petit, on obtient

$$\bar{F}^{\leftarrow}\left(\alpha_n\right) \simeq \bar{F}^{\leftarrow}\left(\beta_n\right)\left(\frac{\beta_n}{\alpha_n}\right)^{\zeta}.$$

Après avoir estimer $\bar{F}^{\leftarrow}\left(\beta_n\right)$ et ζ , on obtient l'estimateur de *Weissman (1978)* et *Embrechts et all(1997).*

$$\hat{q}_{\alpha_n}^{W} = X_{n-[n\beta_n]+1,n}\left(\frac{\beta_n}{\alpha_n}\right)^{\hat{\zeta}}$$

où $X_{n-[n\beta_n]+1,n}$ est un estimateur de $\bar{F}^{\leftarrow}\left(\beta_n\right).$

Bibliographie

[1] **Ahsanullah, Mohammad ; Nevzorov, Valery B. ; Shakil, Mohammad**. An introduction to order statistics. Atlantis Studies in Probability and Statistics, 3. Atlantis Press, Paris, 2013.

[2] **Arnold, Barry C. ; Balakrishnan, N. ; Nagaraja, H. N. A** .A first course in order statistics. Wiley Series in Probability and Mathematical Statistics : Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1992

[3] **A.L.M. Dekkers et L. de Haan**. On the estimation of the extreme-value index and large quantile estimation. Ann. Statist., 17(4) : 1795–1832, 1989.

[4] **Balakrishnan, N. et Cohen, A. C**. Order statistics and inference. Estimation methods. Statistical Modeling and Decision Science. Academic Press, Inc., Boston, MA,.1991

[5] **Beirlant. J , Y. Goegebeur, J. Teugels et J. Segers**. Statistics of extremes. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2004.

[6] **Bingham. N , C. Goldie et J. Teugels**. Regular variation. Cambridge University Press, Cambridge, 1987.

[7] **Coles**. **S.** An introduction to statistical modeling of extreme values. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2001.

[8] **David, H. A**.Order statistics. John Wiley & Sons, Inc., New York-LondonSydney. 1970.

[9] **De Haan. L.** Fighting the arch-enemy with mathematics. Stat. Neerl., 44(No.2) : 45–68, 1990.

[10] **De Haan, L.**, **Ferreira, A**. Extreme Values Theory : An introduction. Springer. 2006.

[11] **Embrechts, P., Klüppelberg, C., and Mikosch, T**. Modelling extremal events. Springer. 1997.

[12] **Galambos, J.** The asymptotic theory of extreme order statistics. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-ChichesterBrisbane., 1978.

[13] **P. Embrechts, C. Klueppelberg et T. Mikosch.** Modelling extremal events for insurance and finance, volume 33 d'Applications of Mathematics. Springer, Berlin, 1997.

[14] **M. Falk, J. H¨ usler et R.-D. Reiss**. Laws of small numbers : extremes and rare events, volume 23 de DMV Seminar. Birkh¨auser Verlag, Basel, 1994.

[15] **M.E.J. Newman**. Power laws, Pareto distributions and Zipf's law. Contemporary Physics, 46 : 323–351, 2005.

[16] **PICKANDS J.**, Statistical inference using extreme order statistics, Annals of Statistics 3, p. 119-131, 1975.

[17] **REISS R., THOMAS M.**, Statistical Analysis of Extreme Value with Applications to Assurance, Finance, Hydrology and Other Fields, Basel, Birkhauser, Verlag, 2001.

[18] **S. Resnick**. Extreme values, regular variation, and point processes. Applied Probability. Springer-Verlag, New York, 1987.

[19] **R.L. Smith.** Maximum likelihood estimation in a class of nonregular cases. Biometrika,72(1) :67,1985.