

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Ghardaia

Faculté des Sciences et de la Technologie
Département de Mathématiques et Informatiques



Projet de fin d'étude en vue de l'obtention du diplôme de
MASTER

Domaine : Mathématiques et Informatique
Spécialité : Analyse Fonctionnelle et applications

THÈME

**Théorie des valeurs extrêmes et
Applications**

Présenté par :

Siham LACHEHEB

Soutenu publiquement le : 01/07/2019

Devant le jury :

M.Lahcen Chenini : (Univ. Ghardaia) Président
Mme.Yasmina Khellaf : (Univ. Ghardaia) Examinatrice
M.Abdenour Lani : (Univ. Ghardaia) Rapporteur

Année universitaire : 2018/2019

Dédicaces

Je dédie ce travail à ceux qui m'ont soutenu tout au long de mes études, qui m'ont toujours poussé vers le chemin du savoir, à ma source d'amour et d'affection, les deux êtres les plus chères au monde.

Mes Chèrs parents : Slimane et Zohra.

À mes frères : Senna , Ahmed , Abdelrahmane.

À mes sœurs : Dalal et Aida.

À ma grande famille LACHEHEB.

À mes amis : Hallouma, Saida, Houda, Fatiha, Roumaïssa, Mebarka, Chahra, Bouchra, Mardia.

À mes amis de l'université de GHARDAIA.

À tous les étudiants de math & info .

À tous ceux qui me sont cher(e)s.

*Siham Lacheheb
Ghardaia 2019*

Remerciements

Je tiens à témoigner ma reconnaissance à DIEU tout puissant, de m'avoir donné le courage et la force de mener à terme ce projet. Qui m'a ouvert les portes du savoir.

Je remercie mon encadreur Monsieur *Abdenour Lani* pour le sujet qu'il m'a proposé et pour l'attention et la disponibilité dont il a su faire preuve au long de la préparation de ce mémoire.

Je remercie sincèrement les membres de jury pour avoir accepté de juger ce travail.

Je remercie les enseignants du département des mathématiques et d'informatique qui nous ont encadrés depuis notre première année universitaire.

Merci à tous ceux qui m'ont enseigné au primaire, collège, lycée, université et ailleurs, et tous les personnels de ces établissements, Merci à tous les mathématiciens...

Merci à mes parents, mes amis, ma famille et mes cousines, merci à tous ceux qui ont contribué de près ou de loin dans ce travail.

Résumé, Abstract, ملخص

Résumé

Dans ce mémoire on a travaillé sur la théorie des valeurs extrêmes(EVT) qui est une branche de la statistique.

Cette technique est utilisée pour modéliser les événements rares, elle s'est développée grâce aux travaux de quelques mathématiciens.

La EVT a des applications dans beaucoup de domaines comme la finance où elle permet de calculer la VaR (Value-at-Risk).

La théorie des valeurs extrêmes est utilisée aussi en assurance où on l'utilise pour estimer la prime de réassurance.

Cette théorie a des applications dans d'autres domaines comme l'hydrologie, la géologie, génie civil...etc.

L'utilisation des logiciels comme le logiciel R permettent de procéder à des simulations afin de vérifier les résultats théoriques.

Mots clés : la théorie des valeurs extrême, indice des valeurs extrêmes, la VaR, les quantiles, la distribution.

Abstract

In this memory we worked on the extreme value theory(EVT) which is a branche of statistic.

This technic is used to model rare events, it is developed through the work of a few mathematicians.

The EVT has applied in many areas as finance where it allows as to calculate the VaR (Value-at-Risk).

The theory of extreme values is also used in assurance where it is used to estimate the reinsurance bonus.

This theory has applied in other areas such as hydrology, geology, civil engineeringetc.

The use of software such as R software will make simulations to check the theoretical results.

Keywords : the extreme value theory, the extreme value index, VaR, the quantities, the distribution.

ملخص

في هذه المذكرة ، عملنا على نظرية القيم القصوى التي تعد فرعاً من الإحصائيات. تستخدم هذه التقنية لنمذجة الأحداث النادرة ، وقد تطورت بفضل عمل بعض علماء الرياضيات. نظرية القيم القصوى لها تطبيقات في العديد من المجالات مثل التمويل حيث تسمح بحساب القيمة المعرضة للخطر. تُستخدم نظرية القيم القصوى أيضاً في التأمين حيث يتم استخدامها لتقدير قسط إعادة التأمين. هذه النظرية لها تطبيقات في مجالات أخرى مثل الهيدرولوجيا ، الجيولوجيا ، الهندسة المدنية ... إلخ. يتيح استخدام البرامج مثل برنامج R إجراء عمليات المحاكاة للتحقق من النتائج النظرية.

الكلمات المفتاحية : نظرية القيم القصوى، مؤشر القيم القصوى ،قيمة الخطر، الكميات، التوزيعات .

Table des matières

Introduction	2
1 Préliminaires	3
1.1 Fonction de répartition empirique	3
1.2 Fonction inverse généralisé	3
1.3 Quantiles	4
1.3.1 Fonction quantile de queue	4
1.3.2 Quantile empirique	5
1.4 Distribution de statistique d'ordre	5
1.4.1 Densité conjointe	5
1.4.2 Densité marginale	5
1.4.3 Densité du maximum	6
1.5 Résultats sur la loi exponentielle	6
1.5.1 Représentation de Renyi	6
1.6 Distribution de Paréto Généralisée(GPD)	7
1.7 Distribution généralisée des extrêmes(GEV)	7
1.8 Domaine d'attraction	8
1.9 Caractérisation du domaine d'attraction	9
1.10 Fonction à variation régulière	9
1.10.1 Variation régulière du premier ordre	10
1.10.2 Bornes de Potter	10
1.10.3 Représentation de Karamata	11
1.10.4 Fonction à variation régulière du second ordre	13

2	Estimateurs de l'indice des valeurs extrêmes	16
2.1	Introduction	16
2.2	Estimateur de Hill	16
2.2.1	Construction de l'estimateur de Hill	17
2.2.2	Consistance de l'estimateur de Hill	21
2.3	Estimateur de Pickands	21
2.3.1	construction de l'estimateur de Pickands	21
2.3.2	Consistance	23
2.4	Estimateur des moments	23
2.4.1	Construction de l'estimateur des moments	24
2.4.2	Consistance	26
3	Domaines d'application	29
3.1	Finance	29
3.1.1	La valeur de risque (Var)	29
3.1.2	Niveau de Confiance	30
3.1.3	Horizon de détention	31
3.1.4	Méthodes d'estimation de la VaR	31
3.1.5	Exemple d'application de calcul de la VaR	33
3.2	Assurance	39
3.3	Hydrologie	40
	Conclusion	41
	Bibliographie	42

Table des figures

3.1	Graphique de l'indice S&P500 entre Avril 2014 et Avril 2019 .	33
3.2	Histogramme du S&P500.	34
3.3	Logarithmes des rendements de S&P500	35
3.4	Histogramme des logarithmes des rendements	36
3.5	La fonction de répartition empirique.	39

Notations et abréviations

Notations

F_n : Fonction de répartition empirique.

F^{\leftarrow} : Inverse généralisée de la fonction de répartition.

F^{-1} : Inverse de la fonction de répartition.

x_α : Quantile d'ordre α .

Abréviation

EVT : Théorie des valeur extrême.

i.i.d : indépendant identiquement distribué.

GPD : Distribution de Paréto Généralisée.

GEV : Distribution généralisée des extrêmes.

EMV : Méthode du maximum de vraisemblance.

VaR : Valeur de risque.

P.O.T : L'approche par dépassements de seuil.

Introduction

La théorie des valeurs extrêmes à été développée pour l'estimation de probabilités d'occurrences d'évènements rares. Elle permet d'extrapoler le comportement de la queue de distribution à partir des plus grands données observées.

Fisher et Tippet (1928) ont identifié la loi limite du maximum qui peut être de type Frechet, Gumbel ou weibull.

En pratique, l'évaluation de la queue de la distribution du phénomène étudié est étroitement liée à un paramètre réel appelé l'indice des valeurs extrêmes.

Les premiers estimateur de l'indice des valeurs extrêmes ont été définis par Hill(1975)et par Pickands(1975). Plus tard, Dekkers et Einmahl et De haan(1989) ont défini l'estimateur des moments qui est une généralisation directe de l'estimateur de Hill, d'autre estimateurs sont définis ces dernières années.

Dans le premier chapitre nous présentons tout d'abord quelques résultats de statistique utiles, la distribution généralisée des extrêmes(GEV), la distribution de Paréto généralisée (GPD) et nous discutons également la notion de domaine d'attraction et les fonctions à variation régulière.

Dans le deuxième chapitre on s'intéresse en particulier aux estimateurs les plus utilisés (Hill, Pickands et Dikkers), en on donnera la définition et la consistance de chacun d'eux.

Dans le troisième chapitre on s'intéresse aux domaines d'application de la théorie des valeurs extrême comme par exemple la finance, l'assurance, l'hydrologie etc....

A la fin on a introduit la VaR (Value-at-Risk) et on a procédé à des simulations avec logiciel R.

Chapitre 1

Préliminaires

Dans ce chapitre, on donne quelques résultats de probabilités et statistique utiles pour la suite.

1.1 Fonction de répartition empirique

Soit X_1, X_2, \dots, X_n des variables aléatoires indépendantes identiquement distribuées (i.i.d) et $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ les statistiques d'ordre associées.

Définition 1.1.1. On appelle fonction de répartition empirique associée à un échantillon X_1, X_2, \dots, X_n la fonction F_n définie par :

$$F_n(x) = \begin{cases} 0 & \text{si } x < X_{1,n} \\ \frac{i-1}{n} & \text{si } X_{i-1,n} \leq x < X_{i,n} \quad i = 2, \dots, n \\ 1 & \text{si } x \geq X_{n,n} \end{cases}$$

1.2 Fonction inverse généralisé

Soit F une fonction non décroissante,

Définition 1.2.1. On appelle l'inverse généralisé de F la fonction notée F^{\leftarrow} définie par :

$$F^{\leftarrow}(x) := \sup\{y \in \mathbb{R}; \quad F(y) \leq x\} = \inf\{y \in \mathbb{R}; \quad F(y) \geq x\}$$

Remarque 1.2.1. Si F est une fonction de répartition continue et strictement croissante alors la fonction inverse généralisé F^{\leftarrow} est équivalente à l'application réciproque F^{-1} .

1.3 Quantiles

Soit X une variable aléatoire et F sa fonction de répartition.

Définition 1.3.1. On appelle quantile d'ordre α ($0 < \alpha < 1$) de la loi de X le nombre réel x_α tel que

$$x_\alpha = \sup\{x, \quad F(x) \leq \alpha\}$$

Remarque 1.3.1. Si F est strictement croissante et continue alors, x_α est l'unique nombre réel tel que

$$F(x_\alpha) = \alpha$$

1.3.1 Fonction quantile de queue

Dans la théorie des valeurs extrêmes, on est souvent amené à déterminer la valeur d'un quantile x_α de telle sorte que $P(X > x_\alpha) = \frac{1}{t}$, avec t d'ordre 10^3 , c'est le cas lorsqu'on veut déterminer la hauteur d'une digue de façon qu'un débordement ait lieu soit inférieur à une certaine valeur fixée.

Ceci est équivalent à

$$1 - P(X \leq x_\alpha) = 1 - F(x_\alpha) = \frac{1}{t}$$

D'où

$$x_\alpha = F^{\leftarrow}\left(1 - \frac{1}{t}\right)$$

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

La fonction quantile de queue est la fonction notée U définie par

$$U(t) = F^{\leftarrow}\left(1 - \frac{1}{t}\right) \quad \text{avec } 0 < t < \infty$$

1.3.2 Quantile empirique

La fonction quantile empirique de l'échantillon X_1, X_2, \dots, X_n est la fonction Q_n définie par

$$Q_n(s) = \sup\{x, F_n(x) \leq (s)\} \quad \text{avec } 0 < s < 1$$

La fonction quantile empirique de la queue correspondante est définie par

$$U_n(t) = F_n^{\leftarrow}\left(1 - \frac{1}{t}\right)$$

Il est facile à vérifier que

$$Q_n(s) = X_{n-i+1,n} \quad \text{pour } \frac{n-i}{n} < s \leq \frac{n-i+1}{n}$$

1.4 Distribution de statistique d'ordre

1.4.1 Densité conjointe

La densité f de l'échantillon ordonné $X_{1,n}, X_{2,n}, \dots, X_{n,n}$ est définie par

$$f_{X_{1,n}, X_{2,n}, \dots, X_{n,n}}(x_1, x_2, \dots, x_n) = n! \prod_{i=1}^n f(x_i)$$

Avec $-\infty < x_1 < x_2 < \dots < x_n < +\infty$.

1.4.2 Densité marginale

Soit $X_{i,n}$ une statistique d'ordre, sa fonction densité est définie par

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

Avec x une inconnue réelle.

La fonction de répartition de $X_{i,n}$ est donnée par

$$F_{X_{i,n}} = \sum_{r=i}^n C_r^n (F(x))^r (1 - F(x))^{n-r}$$

1.4.3 Densité du maximum

La densité de la variable aléatoire $X_{n,n} = \max(X_i)$ est définie comme suit

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

Proposition 1.4.1. *Soit un échantillon U_1, U_2, \dots, U_n de la loi uniforme sur $[0, 1]$ et $U_{1,n} \leq U_{2,n} \leq \dots \leq U_{n,n}$ l'échantillon ordonné associé alors,*

$$X_{i,n} \stackrel{d}{=} F^{-1}(U_{i,n})$$

Pour toute fonction de répartition F

Si F est continue alors

$$F(X_{i,n}) \stackrel{d}{=} U_{i,n}$$

Et les variables aléatoires $F(X_1), F(X_2), \dots, F(X_n)$ sont i.i.d et de même loi uniforme sur $[0, 1]$.

1.5 Résultats sur la loi exponentielle

1.5.1 Représentation de Renyi

Proposition 1.5.1. *Soient E_1, E_2, \dots, E_n des variables aléatoires i.i.d de loi exponentielle de paramètre 1 et $E_{1,n} < E_{2,n} < \dots < E_{n,n}$ les statistiques ordonnées associées alors*

$$(E_{1,n}, E_{2,n} - E_{1,n}, \dots, E_{n,n} - E_{n-1,n}) \stackrel{d}{=} \left(\frac{E_n}{n}, \frac{E_{n-1}}{n-1}, \dots, \frac{E_1}{1} \right)$$

Proposition 1.5.2. (*Sukhatme*) Soit $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ des statistiques ordonnées associées à un échantillon de loi exponentielle de paramètre 1, alors les variables aléatoires Z_1, Z_2, \dots, Z_n où

$$Z_i = (n - i + 1)(X_{i,n} - X_{i-1,n}), \quad i = 1, 2, \dots, n$$

Avec $X_{0,n} \equiv 0$, sont i.i.d. de loi exponentielle de paramètre 1, Par conséquent

$$X_{i,n} \stackrel{d}{=} \sum_{r=1}^i \frac{Z_r}{n - r + 1}, \quad i = 1, 2, \dots, n$$

1.6 Distribution de Paréto Généralisée(GPD)

La distribution de la loi de Paréto généralisée est définie par

$$G_{\gamma,\rho}(x) = \begin{cases} 1 - (1 + \frac{x}{\rho}\gamma)^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0 \\ 1 - \exp(-\frac{x}{\rho}) & \text{si } \gamma = 0 \end{cases}$$

Pour $\{x : x > 0 \text{ et } (1 + \gamma\frac{x}{\rho}) > 0\}$ avec $\rho > 0$ et γ des paramètres réels .

Le paramètre ρ est dit d'échelle et le paramètre γ est dit de forme.

1.7 Distribution généralisée des extrêmes(GEV)

Théorème 1.7.1. (*Tippet, Fisher (1928) et Gnedenko (1943)*) Soit X_1, X_2, \dots, X_n des variables aléatoires i.i.d de fonction de répartition commune F et

$$M_n = \max(X_1, X_2, \dots, X_n)$$

si la loi limite de M_n existe alors, on peut trouver deux constantes de normalisation $a_n > 0$ et b_n telles que

$$\lim_{n \rightarrow \infty} F^n(b_n + a_n x) = H_\gamma$$

Avec

$$H_\gamma(x) = \begin{cases} \exp(-x^{\frac{-1}{\gamma}}) & x > 0 & \text{et} & \gamma > 0 & \text{(Fréchet)} \\ \exp(-(-x)^{\frac{-1}{\gamma}}) & x < 0 & \text{et} & \gamma < 0 & \text{(Weibull)} \\ \exp(-\exp(-x)) & x \text{ réel} & \text{et} & \gamma = 0 & \text{(Gumbel)} \end{cases}$$

La loi de Gumbel est dite de type 1, la loi de Fréchet est dite de type 2 et la loi de Weibull est dite de type 3.

La distribution H_γ peut être écrite aussi de la façon suivante

$$H_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{\frac{-1}{\gamma}}) & \text{si} & \gamma \neq 0 \\ \exp(-\exp(-x)) & \text{si} & \gamma = 0 \end{cases}$$

Avec $(1 + \gamma x) > 0$

Cette écriture de la distribution H_γ est due à Von Mises (1936) et Jenkinson (1955). La distribution H_γ est dite distribution généralisée des extrêmes (GEV), c'est une famille de distributions dont les caractéristiques sont assez différentes. Le paramètre réel γ est appelé indice des valeurs extrêmes.

1.8 Domaine d'attraction

Soit X_1, X_2, \dots, X_n des variables aléatoires i.i.d de fonction de répartition commune F .

Définition 1.8.1. On dit que $F \in D(H_\gamma)$ s'il existe deux constantes réelles $a_n > 0$ et b_n telles que :

$$a_n(S_n - b_n) \xrightarrow{d} H_\gamma$$

avec $S_n = \sum_{i=1}^n X_i$

1.9 Caractérisation du domaine d'attraction

Théorème 1.9.1. (*Von Mises*) Soit F une fonction de répartition absolument continue et h une fonction définie par

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

1. Si $h(x) > 0$ pour x assez grand et pour $\alpha > 0$,

$$\lim_{x \rightarrow \infty} xh(x) = \alpha$$

Alors F appartient au domaine de Fréchet.

2. Si $F^{-1}(1) < \infty$, et pour $\alpha > 0$,

$$\lim_{x \rightarrow F^{-1}(1)} (F^{-1}(1) - 1)h(x) = \alpha$$

Alors F appartient au domaine de Weibull.

3. Supposons que $h(x)$ ne s'annule pas et qu'elle est différentiable au voisinage de $F^{-1}(1)$ (ou bien pour x assez grand quand $F^{-1}(1) = \infty$), Alors F appartient au domaine de Gumbel si

$$\lim_{x \rightarrow F^{-1}(1)} \frac{d}{dx} \left(\frac{1}{h(x)} \right) = 0$$

1.10 Fonction à variation régulière

Jovan Karamata (1930), a introduit la notion de fonction à variation régulière qui joue un grand rôle dans la théorie des valeurs extrêmes. Eugène Sénéta a apporté sa contribution dans ce domaine dans sa monographie (1970). La notion de variation régulière du premier ordre est nécessaire pour établir la consistance des estimateurs de l'indice des valeurs extrêmes, leur normalité asymptotique exige la variation régulière du second ordre introduite, pour la première fois, par De Haan et Stadtmuller (1996).

1.10.1 Variation régulière du premier ordre

Définition 1.10.1. *une fonction mesurable L est dite à variation régulière à l'infini d'indice α si et seulement si*

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = t^\alpha \quad \forall t > 0$$

Cette notion est dite aussi variation régulière du premier ordre et on note $L \in \mathfrak{R}_\alpha$ où α est un réel quelconque.

Exemple 1.10.1. *les fonctions $x \mapsto x^\alpha$ et $x \mapsto (\log(1+x))^\alpha$ sont à variation régulière d'indice α . La fonction $x \mapsto \sin x$ n'est pas à variation régulière.*

Remarque 1.10.1. *si $\alpha = 0$ la fonction L est dite à variation lente.*

Exemple 1.10.2. *la fonction $x \mapsto \log x$ est à variation lente. En effet*

$$\lim_{x \rightarrow \infty} \frac{\log(tx)}{\log(x)} = \lim_{x \rightarrow \infty} \frac{\log t + \log x}{\log x} = 1 \quad \text{pour tout } t > 0$$

Proposition 1.10.1. *Soit h une fonction positive mesurable, elle est dite à variation régulière d'indice α si et seulement si*

$$h(x) = x^\alpha L(x) \quad \text{où } L \text{ est à variation lente}$$

1.10.2 Bornes de Potter

Un résultat d'analyse important sur les fonctions à variation régulière sont les bornes de Potter qui est énoncé dans la proposition ci-dessous.

Proposition 1.10.2. *Soit f une fonction à variation régulière d'indice α , $\forall \varepsilon > 0$ et $\forall A > 1 \exists X(A, \alpha)$ tel que*

$$\forall x \geq X \text{ et } \forall y \geq X \quad \frac{f(y)}{f(x)} \leq A \max \left\{ \left(\frac{y}{x} \right)^{\alpha+\varepsilon}, \left(\frac{y}{x} \right)^{\alpha-\varepsilon} \right\}$$

Cette inégalité peut être écrite de la façon suivante

$$\forall \varepsilon > 0 \exists t_0 ; \forall t > t_0 ; (1 - \varepsilon)y^{\alpha-\varepsilon} \leq \frac{f(ty)}{f(t)} \leq (1 + \varepsilon)y^{\alpha+\varepsilon}$$

1.10.3 Représentation de Karamata

Proposition 1.10.3. *Si h est une fonction à variation régulière d'indice α , alors $\forall x \geq A$ tel A un réel strictement positif, on a*

$$h(x) = c(x) \exp \int_A^x \frac{r(t)}{t} dt$$

r et s étant deux fonctions mesurables telles que

$$\lim_{x \rightarrow \infty} c(x) = c_0 \quad c_0 \geq 0 \quad \text{et} \quad \lim_{x \rightarrow \infty} r(x) = \alpha$$

La réciproque de cette proposition est vraie et cette écriture de h est dite de Karamata.

Proposition 1.10.4. (Beirlant) *Soit U la fonction quantile de la queue définie par*

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

F^{\leftarrow} étant l'inverse généralisé de la fonction de répartition F

Si, pour tout $u > 0$ et a une fonction à variation régulière d'indice γ ,

le rapport suivant : $\frac{U(xu) - U(x)}{a(x)}$ est de limite finie quand x tend vers l'infini alors

$$\lim_{x \rightarrow \infty} \frac{U(xu) - U(x)}{a(x)} = h_\gamma(u)$$

avec

$$h_\gamma(u) = \int_1^u v^{\gamma-1} dv = \frac{u^\gamma - 1}{\gamma} \quad \text{si } \gamma \neq 0$$

pour $\gamma = 0$ on a

$$h_0(u) = \log u.$$

La limite

$$\lim_{x \rightarrow \infty} \frac{U(xu) - U(x)}{a(x)} = h_\gamma(u)$$

est notée par $(C_\gamma(a))$, cette notation est relative à la fonction à variation régulière a .

Théorème 1.10.1. (De Haan) Pour $\gamma \in \mathbb{R}$, les résultats suivants sont équivalents

1. Il existe des constantes réelles $a_n > 0$ et b_n telles que

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H_\gamma(x)$$

pour tout réel x avec $(1 + \gamma x) > 0$.

2. Pour tout $x > 0$,

$$\lim_{x \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \quad \text{et } \gamma \neq 0.$$

Pour $\gamma = 0$, le terme de droite de l'égalité ci-dessus vaut $\log x$.

La reformulation de la condition $C_\gamma(a)$ en terme de la fonction de répartition F , supposée continue, est donnée dans la proposition suivante

Proposition 1.10.5. (De Haan (1970)) F appartient au domaine d'attraction de H_γ si et seulement si, pour toute fonction auxiliaire b et $(1 + \gamma v) > 0$ on a

$$\lim_{y \rightarrow x} \frac{1 - F(y + b(y)v)}{1 - F(y)} = (1 + \gamma v)^{\frac{-1}{\gamma}}$$

alors

$$\frac{b(y + vb(y))}{b(y)} \rightarrow u^\gamma = (1 + \gamma v)$$

la fonction auxiliaire b peut être choisie telle que

$$b(y) = a(U^{-1}(y)).$$

Proposition 1.10.6. (Beirlant 2004) la condition $(C_\gamma(a))$ entraîne, quand x tend vers l'infini,

$$\frac{U(x)}{a(x)} (\log U(ux) - \log U(x)) \rightarrow \begin{cases} \log u & \text{si } \gamma \geq 0 \\ \frac{u^\gamma - 1}{\gamma} & \text{si } \gamma < 0 \end{cases} \quad \tilde{C}_\gamma(a)$$

1.10.4 Fonction à variation régulière du second ordre

La variation régulière du second ordre, quoique difficile à vérifier en pratique, est nécessaire pour établir la normalité asymptotique des estimateurs de l'indice des valeurs extrêmes. Elle a été introduite, pour la première fois, par de Haan et Stadtmüller (1996), elle est basée sur la condition $(C_\gamma(a))$, i.e soit une fonction a_2 positive telle que

$$\lim_{x \rightarrow +\infty} a_2(x) = 0$$

et vérifiant

$$\frac{U(xu) - U(x)}{a(x)} - h_\gamma(u) \sim a_2(x)k(u) \quad \text{quand } x \rightarrow \infty, u > 0$$

on a vu précédemment que pour

$$a(x) = x^\gamma L(x) \quad \text{où } L \text{ est à variation lente}$$

$$h_\gamma(u) = \int_1^u v^{\gamma-1} dx$$

on a aussi obtenu pour u et v strictement positifs, les relations suivantes

$$\begin{aligned} h_\gamma(uv) &= u^\gamma h_\gamma(v) + h_\gamma(u) \\ &= v^\gamma h_\gamma(u) + h_\gamma(v) \end{aligned}$$

les deux résultats suivants sont faciles à démontrer à savoir

$$h_{-\gamma}\left(\frac{1}{u}\right) = -h_\gamma(u)$$

Et

$$u^\gamma h_{-\gamma}(u) = h_\gamma(u)$$

Premièrement, déduisons une équation pour la fonction k en remplaçant u par uv , on obtient

$$U(uvx) - U(x) = U(uvx) - U(ux) + U(ux) - U(x)$$

on peut vérifier par un simple calcul que

$$\begin{aligned}
& a_2^{-1}(x) \left\{ \frac{U(uvx) - U(x)}{a(x)} - h_\gamma(uv) \right\} \\
&= a_2^{-1}(ux) \left\{ \frac{U(uvx) - U(ux)}{a(ux)} - h_\gamma(v) \right\} \left\{ \frac{a(ux)a_2(ux)}{a(x)a_2(x)} \right\} + a_2^{-1}(x) \left\{ \frac{U(ux) - U(x)}{a(x)} - h_\gamma(u) \right\} + \\
& a_2^{-1} \left\{ \frac{a(ux)}{a(x)} h_\gamma(u) + h_\gamma(u) - h_\gamma(uv) \right\}
\end{aligned}$$

L'égalité précédente est dite équation fondamentale du second ordre.

Le troisième terme du membre de droite de l'équation fondamentale du second ordre peut être simplifié en utilisant les propriétés de la fonction h_γ

le fait que la fonction a est à variation régulière d'indice γ , nous donne

$$\begin{aligned}
a_2^{-1}(x) h_\gamma(u) \left\{ \frac{a(ux)}{a(x)} - u^\gamma \right\} &= a_2^{-1}(x) h_\gamma(v) \left\{ \frac{(ux)^\gamma l(ux)}{x^\gamma l(x)} - u^\gamma \right\} \\
&= u^\gamma a_2^{-1}(x) h_\gamma(v) \left\{ \frac{l(ux)}{l(x)} - 1 \right\}
\end{aligned}$$

si on suppose que la limite du premier membre de l'équation fondamentale du second ordre existe alors nécessairement, le rapport $\frac{a_2(ux)}{a_2(x)}$ converge ou d'une façon équivalente, l'existence de la limite du rapport

$$a_2^{-1}(x) \left\{ \frac{l(ux)}{l(x)} - 1 \right\}$$

on a

$$a_2^{-1}(x) \left\{ \frac{l(ux)}{l(x)} - 1 \right\} = \frac{l(ux) - l(x)}{a_2(x)l(x)}$$

Donc

$$\lim_{x \rightarrow \infty} \frac{l(ux) - l(x)}{a_2(x)l(x)} = m\rho(u)$$

de la condition $C\rho(\cdot)$ nous déduisons que la fonction $a_2(x)l(x)$ est à variation régulière d'indice ρ , or

$$m\rho(u) = \frac{c(u\rho - 1)}{\rho} \quad \text{avec } \rho \neq 0$$

c étant une constante non nulle.

Ces résultats s'accordent avec ceux donnés par Goldie et Smith (1987) en

supposant que la fonction \log satisfait la condition $C\rho(\cdot)$ on obtient la fonctionnelle suivante

$$k(uv) = u^{\gamma+\rho}k(v) + k(u) + cu^\gamma h_\gamma(v)h_\rho(u) \quad \text{pour } u \text{ et } v > 0$$

l'expression de la fonction k est donnée par De Haan et al.(2006) comme suit

$$k(x) = c_1 \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1} du \quad ds + c_2 \int_1^x s^{\gamma+\rho-1} ds$$

c_1, c_2 sont des constantes réelles.

Définition 1.10.2. *On dit qu'une fonction U est à variation régulière du second ordre si elle vérifie l'approximation suivante*

$$\frac{U(xu) - U(x)}{a(x)} - h_\gamma(u) \sim a_2(x)k(u) \quad \text{quand } x \rightarrow \infty, u > 0$$

ρ est dit paramètre du second ordre.

Chapitre 2

Estimateurs de l'indice des valeurs extrêmes

2.1 Introduction

En pratique l'un des problème de la théorie des valeur extrême (EVT) est d'estimer le paramètre γ (l'indice des valeurs extrême). Les premiers estimateurs de ce paramètre ont été définis par Hill, Pickands et Dekkers.

2.2 Estimateur de Hill

Cet estimateur a été défini par Hill (1975). C'est un estimateur très populaire et beaucoup utilisé mais qui est défini seulement pour les valeurs positives de γ .

Définition 2.2.1. Soit X_1, X_2, \dots, X_n des variables aléatoires indépendantes de même fonction de répartition $F \in D(H_\gamma)$ et $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ les statistiques ordonnées associées, et $\gamma > 0$.

l'estimateur de Hill est définit par :

$$H_{k,n} := \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n}).$$

2.2.1 Construction de l'estimateur de Hill

Il ya plusieurs approches pour construire l'estimateur de Hill.

1.Approche par l'EMV conditionnel

Exemple introductif

On considère une suite de variables aléatoires indépendantes et identiquement distribuées de loi de Paréto de paramètre α ($\alpha > 1$) de fonction de répartition définie par

$$F(x) = 1 - x^{-\alpha} \quad \text{pour tout } x \geq 1$$

sa fonction de densité est définie par

$$f(x) = \alpha x^{-\alpha-1} \quad \text{pour tout } x \geq 1$$

cherchons l'estimateur du maximum de vraisemblance(EMV)du paramètre α la fonction de vraisemblance L et les dérivées de la log-vraisemblance sont

$$L(x_1, x_2, \dots, x_n; \alpha) = \alpha^n \left(\prod_{i=1}^n x_i^{-\alpha-1} \right)$$

$$\frac{d \log L(x_1, x_2, \dots, x_n; \alpha)}{d\alpha} = \frac{n}{\alpha} - \sum_{i=1}^n \log x_i$$

$$\frac{d^2 \log L(x_1, x_2, \dots, x_n; \alpha)}{d^2} = \frac{-n}{\alpha^2} < 0$$

l'estimateur du maximum de vraisemblance de $\frac{1}{\alpha}$ est donc la statistique $H_n = \frac{1}{n} \sum_{i=1}^n \log X_i$
posons $Y_i = \log X_i$, donc $H_n = \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y}$

Soit X une variable aléatoire de fonction de répartition F telle que,
pour $\alpha > 0$

$$P(X > x) = 1 - F(x) = x^{-\alpha}, \quad x \geq 1$$

on déduit que la variable aléatoire $Y = \log X$ est de fonction de répartition

$$P(Y > y) = \exp(-\alpha y), \quad y \geq 0,$$

i.e. Y suit une loi exponentielle de paramètre α et donc l'EMV de α est donné par

$$\widehat{\alpha}_n = (\overline{Y}_n)^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \log X_j \right)^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \log X_{j,n} \right)^{-1}$$

Une généralisation triviale concerne

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

avec u connu. Si on pose $C = u^\alpha$, on obtient comme EMV de α

$$\widehat{\alpha}_n = \left(\frac{1}{n} \sum_{j=1}^n \log \left(\frac{X_{j,n}}{u} \right) \right)^{-1} = \left(\frac{1}{n} \sum_{j=1}^n \log X_{j,n} - \log u \right)^{-1}$$

Souvent, on n'a pas l'expression de la fonction de répartition mais dans le domaine de Fréchet on sait que F suit approximativement une loi de Paréto à partir d'un certain seuil connu u . Soit

$$K = \text{card} \{i : X_{i,n} > u, i = 1, \dots, n\}.$$

conditionnellement, à l'événement $\{K = k\}$, la méthode du maximum de vraisemblance pour estimer α et C consiste à maximiser la densité conjointe de $(X_{k,n}, \dots, X_{1,n})$ définie par

$$f_{X_{k,n}, \dots, X_{1,n}}(x_k, \dots, x_1) = \frac{n!}{(n-k)!} (1 - Cx_k^{-\alpha})^{n-k} C^k \alpha^k \prod_{i=1}^k x_i^{-(\alpha+1)}, \quad u < x_k < \dots < x_1$$

un simple calcul nous donne l'estimateur du maximum de vraisemblance conditionnel du paramètre $\gamma = \frac{1}{\alpha}$

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

2.Approche par les fonctions à variation régulière

Soit X_1, X_2, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées de fonction de répartition commune F et $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ les statistiques ordonnées associées.

pour $\gamma > 0$, $F \in D(H_\gamma)$ si et seulement si

$$\lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\frac{1}{\gamma}} \quad \text{pour tout } x > 0$$

la limite précédente, provient du fait que F est à queue lourde (heavy-tailed), i.e,

$$1 - F(x) = x^{-\frac{1}{\gamma}} L(x) \quad \text{où la fonction } L \text{ est à variation lente}$$

elle est équivalente à

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (1 - F(tx))}{(1 - F(t))} = \gamma$$

en faisant une intégration par partie on obtient

$$\int_t^\infty \frac{(1 - F(x))}{x} dx = \int_t^\infty (\log u - \log t) dF(u)$$

posons

$$u(x) = 1 - F(x) \quad \text{et} \quad dv = \frac{1}{x} dx$$

donc

$$du = -dF(x) \quad \text{et} \quad v(x) = \log x$$

ce qui donne

$$\int_t^\infty \frac{(1 - F(x))}{x} dx = [(1 - F(x)) \log x]_t^\infty + \int_t^\infty (\log x) dF(x)$$

remarquons que lorsque x tend vers l'infini alors

$$[(1 - F(x)) \log x] \rightarrow 0$$

en supposant que $E(X) < \infty$, l'inégalité de Markov nous donne

$$[(1 - F(x)) \log x] = (\log x)[P(X > x)] \leq (\log x) \frac{E(X)}{x}$$

par conséquent,

$$[(1 - F(x)) \log x] \rightarrow 0 \quad x \rightarrow \infty$$

d'où

$$\begin{aligned} \int_t^\infty \frac{(1 - F(x))}{x} dx &= -(1 - F(t)) \log t + \int_t^\infty (\log x) dF(x) \\ &= -(\log t) \int_t^\infty dF(x) + \int_t^\infty (\log x) dF(x) \\ &= \int_t^\infty (\log x - \log t) dF(x) \end{aligned}$$

donc

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty (\log x - \log t) dF(x)}{1 - F(t)} = \gamma$$

en remplaçant dans l'égalité précédente, respectivement, le paramètre t par la statistique d'ordre $X_{n-k,n}$ et F par la fonction de répartition empirique F_n on obtient

$$H_{k,n} := \frac{\int_{X_{n-k,n}}^\infty (\log x - \log X_{n-k,n}) dF_n(x)}{1 - F_n(X_{n-k,n})}$$

nous pouvons approximer l'intégrale du numérateur dans par une somme de Riemann en prenant un pas de $\frac{1}{n}$
remarquons que

$$1 - F_n(X_{n-k,n}) = 1 - \frac{n-k}{n} = \frac{k}{n}$$

nous obtenons finalement

$$H_{k,n} := \frac{1}{k} \sum_{i=0}^{k-1} (\log X_{n-i,n} - \log X_{n-k,n}).$$

La statistique $H_{k,n}$ est dite estimateur de Hill.

2.2.2 Consistance de l'estimateur de Hill

1. Convergence en probabilité

Théorème 2.2.1. (Mason 1982) Soit X_1, X_2, \dots, X_n des variables aléatoires i.i.d de fonction de répartition commune F , avec $F \in D(H_\gamma)$, $\gamma > 0$ et $k(n)$ suit d'entier tq $1 < k(n) \leq n$

$$\text{si } \begin{cases} k(n) \rightarrow \infty \\ \text{et} \\ \frac{k(n)}{n} \rightarrow 0 \end{cases} \quad \text{alors } H_{k,n} \xrightarrow{p} \gamma, \quad \text{lorsque } n \rightarrow \infty$$

2. Convergence forte

Théorème 2.2.2. (Deheuvels, Hausler et Mason 1988) Soit X_1, X_2, \dots, X_n des variables aléatoires i.i.d de fonction de répartition commune F , avec $F \in D(H_\gamma)$.

$$\text{si } \begin{cases} k(n) \rightarrow \infty \\ \text{et} \\ \frac{k(n)}{\log(\log(n))} \rightarrow \infty \end{cases} \quad \text{alors } H_{k,n} \xrightarrow{p.s} \gamma, \quad \text{lorsque } n \rightarrow \infty.$$

2.3 Estimateur de Pickands

Cet estimateur a été défini par James Pickands (1975).

Définition 2.3.1. l'estimateur de Pickands est définie par :

$$\hat{\gamma}_{P,k} = \frac{1}{\log 2} \log \left(\frac{X_{n-\lfloor \frac{k}{4} \rfloor + 1, n} - X_{n-\lfloor \frac{k}{2} \rfloor + 1, n}}{X_{n-\lfloor \frac{k}{2} \rfloor + 1, n} - X_{n-k+1, n}} \right) \quad \text{pour } k = 1, \dots, n$$

2.3.1 construction de l'estimateur de Pickands

Utilisons la condition $(C\gamma(a))$ pour l'expression suivante

$$\left(\frac{1}{\log 2} \right) \log \left[\frac{U(4y) - U(2y)}{U(2y) - U(y)} \right]$$

on peut écrire

$$\left(\frac{1}{\log 2}\right) \log \left[\frac{U(4y) - U(2y)}{U(2y) - U(y)} \right] = \left(\frac{1}{\log 2}\right) \log \left[\frac{U(4y) - U(2y)}{a(2y)} \frac{a(2y)}{a(y)} \frac{a(y)}{U(2y) - U(y)} \right]$$

en prenant a comme fonction à variation régulière d'indice γ , on a

$$\lim_{y \rightarrow \infty} \frac{a(2y)}{a(y)} = 2^\gamma$$

$$\lim_{y \rightarrow \infty} \frac{U(2y) - U(y)}{a(y)} = h_\gamma(2)$$

$$\lim_{y \rightarrow \infty} \frac{U(4y) - U(2y)}{a(2y)} = h_\gamma(2) = \frac{2^\gamma - 1}{\gamma}$$

en utilisant les limites précédentes, quand $y \rightarrow \infty$, on obtient

$$\begin{aligned} \left(\frac{1}{\log 2}\right) \log \left[\frac{U(4y) - U(2y)}{U(2y) - U(y)} \right] &\rightarrow \left(\frac{1}{\log 2}\right) \log \left(h_\gamma(2) 2^\gamma \frac{1}{h_\gamma(2)} \right) \\ &\rightarrow \frac{1}{\log 2} (\log(2^\gamma)) = \gamma \end{aligned}$$

en remplaçant y par $\frac{n}{k}$ et $U(x)$ par sa version empirique, on obtient l'estimateur de Pickands qu'on note $\hat{\gamma}_{P,k}$ défini par

$$\hat{\gamma}_{P,k} = \frac{1}{\log 2} \log \left(\frac{X_{n - [\frac{k}{4}] + 1, n} - X_{n - [\frac{k}{2}] + 1, n}}{X_{n - [\frac{k}{2}] + 1, n} - X_{n - k + 1, n}} \right)$$

$$k = 1, \dots, n \quad \text{et} \quad \hat{U}_n(x) = X_{n - [\frac{n}{x}] + 1, n}$$

$\hat{U}_n(x)$ étant la fonction quantile empirique de queue .

La définition originale de Pickands utilise $4k$ au lieu de k .

2.3.2 Consistance

1. Convergence en probabilité

Théorème 2.3.1. (Dekkers et de Haan 1989) Soit X_1, X_2, \dots, X_n un échantillon de variable aléatoire i.i.d de fonction de répartition F avec $F \in D(H_\gamma)$ et $k(n)$ suit d'entier tq $1 < k(n) \leq n$

$$\text{si } \begin{cases} k(n) \rightarrow \infty \\ \text{et} \\ \frac{k(n)}{n} \rightarrow 0 \end{cases} \quad \text{alors } \hat{\gamma}_{p,k} \xrightarrow{p} \gamma, \quad \text{lorsque } n \rightarrow \infty$$

2. Convergence forte

Théorème 2.3.2. Soit X_1, X_2, \dots, X_n un échantillon de variable aléatoire i.i.d de fonction de répartition F avec $F \in D(H_\gamma)$

$$\text{si } \begin{cases} \frac{k(n)}{\log(\log(n))} \rightarrow \infty \\ \text{et} \\ \frac{k(n)}{n} \rightarrow \infty \end{cases} \quad \text{alors } \hat{\gamma}_{p,k} \xrightarrow{p.s} \gamma, \quad \text{lorsque } n \rightarrow \infty$$

2.4 Estimateur des moments

Cet estimateur a été défini par Dekkers, Einmahl et De Haan (1989). C'est une généralisation directe de l'estimateur de Hill, il est défini pour toutes les valeurs réelles de γ .

Définition 2.4.1. l'estimateur des moments est défini par :

$$M_{k,n} = H_{k,n} + 1 - \frac{1}{2} \left(1 - \frac{H_{k,n}^2}{H_{k,n}^{(2)}} \right)^{-1}$$

avec

$$H_{k,n}^{(2)} = \frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1,n} - \log X_{n-k,n})^2$$

2.4.1 Construction de l'estimateur des moments

. Pour comprendre cet estimateur on peut procéder de la façon suivante pour tout $j \in \{1, \dots, k\}$ on a

$$\log X_{n-j+1,n} - \log X_{n-k,n} = \log \widehat{U}_n \left(\frac{n+1}{j} \right) - \log \widehat{U}_n \left(\frac{n+1}{k+1} \right)$$

et par suite,

$$(\log X_{n-j+1,n} - \log X_{n-k,n})$$

peut être considérée comme une estimation de

$$\log U \left(\frac{n+1}{j} \right) - \log U \left(\frac{n+1}{k+1} \right) = \log U \left(\left(\frac{n+1}{k+1} \right) \left(\frac{k+1}{j} \right) \right) - \log U \left(\frac{n+1}{k+1} \right)$$

en choisissant $x = \frac{n+1}{k+1}$ et $u = \frac{k+1}{j}$ dans $(\widetilde{C}\gamma)$ et quand $\frac{n}{k} \rightarrow \infty$ on a

$$\log X_{n-j+1,n} - \log X_{n-k,n} \sim \begin{cases} \frac{a \left(\frac{n+1}{k+1} \right)}{U \left(\frac{n+1}{k+1} \right)} \log \frac{k+1}{j} & \text{si } \gamma \geq 0 \\ \frac{a \left(\frac{n+1}{k+1} \right)}{U \left(\frac{n+1}{k+1} \right)} \frac{\left(\frac{j}{k+1} \right)^{-\gamma} - 1}{\gamma} & \text{si } \gamma < 0 \end{cases}$$

pour $k \rightarrow \infty$, on a

$$\frac{1}{k} \sum_{j=1}^k \log \left(\frac{k+1}{j} \right) \rightarrow \int_0^1 (\log u) du = 1$$

en effet

$$\frac{1}{k} \sum_{j=1}^k \log \left(\frac{k+1}{j} \right) \sim \int_1^k \frac{1}{k} \log \left(\frac{k+1}{j} \right) dj$$

faisons le changement de variable suivant

$$u = \frac{j}{k+1}$$

si $j = 1$ et $k \rightarrow \infty$ alors $u \rightarrow 0$

et si $j = k$ et $k \rightarrow \infty$ alors $u \rightarrow 1$

de plus, $du = \frac{1}{k+1}dj \sim \frac{1}{k}dj$

donc

$$\frac{1}{k} \sum_{j=1}^k \log \left(\frac{k+1}{j} \right) \rightarrow \int_0^1 \left(\log \frac{1}{u} \right) du = - \int_0^1 (\log u) du = 1$$

on pourra, de la même manière, démontrer le résultat limite suivant

$$\frac{1}{k} \sum_{j=1}^k \left(\log \left(\frac{k+1}{j} \right) \right)^2 \rightarrow \int_0^1 (\log u)^2 du = 2$$

en procédant par changement de variable on montre que

$$\frac{1}{k} \sum_{j=1}^k \left(\left(\frac{j}{k+1} \right)^{-\gamma} - 1 \right) \rightarrow \int_0^1 (u^{-\gamma} - 1) du = \frac{\gamma}{1-\gamma} \quad \text{si } \gamma < 0$$

en effet, posons $u = \frac{j}{k+1}$ donc

$$\begin{aligned} \frac{1}{k} \sum_{j=1}^k \left(\left(\frac{j}{k+1} \right)^{-\gamma} - 1 \right) &= \frac{1}{k} \sum_{j=1}^k (u^{-\gamma} - 1) \\ &\rightarrow \int_0^1 (u^{-\gamma} - 1) du \\ &= \frac{1}{-\gamma + 1} - 1 \\ &= \frac{1 - 1 + \gamma}{1 - \gamma} \\ &= \frac{\gamma}{1 - \gamma} \end{aligned}$$

$$\frac{1}{k} \sum_{j=1}^k \left(\left(\frac{j}{k+1} \right)^{-\gamma} - 1 \right)^2 \rightarrow \int_0^1 (u^{-\gamma} - 1)^2 du = \frac{2\gamma^2}{(1-\gamma)(1-2\gamma)} \quad \text{si } \gamma < 0$$

donc, si n et $k \rightarrow \infty$ et $\frac{k}{n} \rightarrow 0$ alors

$$\frac{H_{k,n}^2}{H_{k,n}^{(2)}} \xrightarrow{p} \begin{cases} \frac{1}{2} & \text{si } \gamma \geq 0 \\ \frac{1-2\gamma}{2(1-\gamma)} & \text{si } \gamma < 0 \end{cases}$$

2.4.2 Consistance

Les résultats limites précédents vont nous servir pour montrer la consistance de l'estimateur des moments, on va étudier deux cas si $\gamma \geq 0$

$$\lim_{x \rightarrow \infty} \frac{a(x)}{U(x)} = \gamma$$

en utilisant les résultats précédents, on peut écrire

$$\frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1,n} - \log X_{n-k,n})^2 \sim \frac{1}{k} \sum_{j=1}^k \left(\frac{a\left(\frac{n+1}{k+1}\right)}{U\left(\frac{n+1}{k+1}\right)} \log \frac{k+1}{j} \right)^2 \sim 2\gamma^2$$

d'où

$$M_{k,n} \xrightarrow{p} \gamma + 1 - \frac{1}{2} \left(1 - \frac{\gamma^2}{2\gamma^2}\right)^{-1} = \gamma$$

donc, pour $\gamma \geq 0$, on a

$$M_{k,n} \xrightarrow{p} \gamma$$

Si $\gamma < 0$, on a l'approximation suivante

$$(\log X_{n-j+1,n} - \log X_{n-k,n}) \sim \frac{a\left(\frac{n+1}{k+1}\right) \left(\frac{j}{k+1}\right)^{-\gamma} - 1}{U\left(\frac{n+1}{k+1}\right) \gamma}$$

donc

$$\begin{aligned} H_{k,n}^2 &= \left(\frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1,n} - \log X_{n-k,n}) \right)^2 \\ &\sim \left(\frac{1}{k} \sum_{j=1}^k \frac{a\left(\frac{n+1}{k+1}\right) \left(\frac{j}{k+1}\right)^{-\gamma} - 1}{U\left(\frac{n+1}{k+1}\right) \gamma} \right)^2 \\ &\sim \frac{a^2\left(\frac{n+1}{k+1}\right)}{U^2\left(\frac{n+1}{k+1}\right)} \left(\frac{1}{k} \sum_{j=1}^k \frac{\left(\frac{j}{k+1}\right)^{-\gamma} - 1}{\gamma} \right)^2 \end{aligned}$$

en utilisant les résultats limites précédents, on obtient

$$H_{k,n}^2 \rightarrow \frac{a^2 \left(\frac{n+1}{k+1}\right)}{U^2 \left(\frac{n+1}{k+1}\right)} \frac{1}{\gamma^2} \left(\frac{\gamma^2}{(1-\gamma)^2}\right)$$

et

$$H_{k,n}^{(2)} \rightarrow \frac{a^2 \left(\frac{n+1}{k+1}\right)}{U^2 \left(\frac{n+1}{k+1}\right)} \frac{2}{(1-\gamma)(1-2\gamma)}$$

d'où

$$\frac{H_{k,n}^2}{H_{k,n}^{(2)}} \xrightarrow{p} \frac{1-2\gamma}{2(1-\gamma)}$$

finalelement,

$$\begin{aligned} M_{k,n} &\xrightarrow{p} 0 + 1 - \frac{1}{2} \left(1 - \frac{1-2\gamma}{2(1-\gamma)}\right)^{-1} = 1 - \frac{1}{2} \left(\frac{1}{2(1-\gamma)}\right)^{-1} \\ &= 1 - \frac{2(1-\gamma)}{2} \\ &= \gamma \end{aligned}$$

donc pour $\gamma < 0$, on a aussi la convergence en probabilité de l'estimateur des moments vers l'indice des valeurs extrêmes.

Remarque 2.4.1. *La difficulté consiste à choisir le nombre k de statistiques d'ordre à utiliser, ce choix est crucial. Si k est trop grand, alors l'approximation par une loi de Pareto sera mauvaise. A l'inverse, si k est trop petit on disposera de très peu d'observations pour le calcul des estimateurs. En théorie, on verra que pour obtenir la consistance des estimateurs il faut prendre $k \rightarrow \infty$ (k assez grand) et $\frac{k}{n} \rightarrow 0$ (k pas trop grand).*

Pour la détermination de k on peut utiliser des méthodes graphiques ou des méthodes analytiques.

l'une des méthodes analytiques les plus utilisées est celle qui consiste à minimiser l'AMSE (asymptotic mean squared error).

Remarque 2.4.2. *Il y'a d'autres estimateurs de l'indice des valeurs extrême comme par exemple : l'estimateur de SMITH (estimateur du maximum de*

vraisemblance), l'estimateur à noyau, estimateur de hill negatif, estimateur rapport des moment, estimateur des records....etc

Chapitre 3

Domaines d'application

3.1 Finance

Les banques pour éviter des faillites (crash) conçoivent des stratégies d'entrée et sortie de l'argent pour cela les financiers fixe un seuil des pertes au dessous du quel il ya un risque pour la banque .

Cette idée a conduit à l'introduction de la VaR (Value-at-Risk).

3.1.1 La valeur de risque (Var)

La notion de value-at-Risk (VaR) est apparue pour la première fois dans le secteur de l'assurance .

A la fin des années 1980 la banque Banker Trust fût l'une des premières institution à utiliser cette notion sur les marchés financiers aux USA, ensuite la banque Jp Morgan a popularisé ce concept durant les années 1990 grâce à son système Risk Matrics.

La VaR est ensuite devenue en quelque années une mesure de référence du risque sur les plans financières, consacrée notamment par la réglementation prudentielle définie dans le cadre des accords de Bâle II.

De façon générale, la Value-at-Risk est définie comme la perte potentielle qui ne devrait être atteinte qu'avec une probabilité donnée sur un horizon temporel donné . La Value at Risk est donc la pire perte attendue sur un

horizon de temps donné pour un niveau de confiance donné. Cette définition très simple constitue l'un des principaux attraits de la Value-at-Risk.

Ainsi, la Value-at-Risk n'est rien d'autre qu'un fractile de la distribution de perte et profit associée à la détention d'un actif ou d'un portefeuille d'actifs sur une période donnée.

la Value-at-Risk correspond tout simplement au fractile de niveau $\alpha\%$ de la distribution de perte et profit pour un horizon donnée :

$$VaR(\alpha) = F^{-1}(\alpha)$$

où F désigne la fonction de répartition associée à la distribution de perte et profit et $(1 - \alpha)$ désigne le niveau de confiance . Généralement α prends les valeurs 5% ou 1%.

De cette définition générale découlent quelque définitions techniques tout aussi simples.

Ainsi, la Value-at-Risk dépend de trois éléments :

1. la distribution des pertes et profits du portefeuille valable pour la période de détention.
2. le niveau de confiance $(1 - \alpha)$ (taux de couverture).
3. la période de détention de l'actif.

3.1.2 Niveau de Confiance

Le niveau de confiance choisi est un paramètre compris entre 0 et 1 (95% ou 99% en général) qui permet de contrôler la probabilité que l'on obtienne un rendement supérieur ou égale à la Value-at-Risk.

Ainsi, la Value-at-Risk correspond généralement à une perte (valeur négative). Toutefois, on trouve souvent une Value-at-Risk définie non pas à partir de la distribution de perte (-) et profit (+), mais à partir au contraire d'une distribution de profit(-) et perte(+). Dit autrement, une telle définition revient à omettre le signe moins devant la perte et donc à afficher une Value-at-Risk positive. Dans ce cas, la définition de la Value-at-Risk correspond à

l'opposé du fractile de la distribution de perte et profit :

$$VaR(\alpha) = -F^{-1}(\alpha)$$

3.1.3 Horizon de détention

Le deuxième élément fondamental dans le calcul de la VaR est la période de détention de l'actif ou du portefeuille d'actifs. La formule de calcul de la Value-at-Risk doit alors être ajustée de façon à tenir compte de la composition des rendements. Il n'existe aucune règle quant au choix de la période de détention dans le calcul de la VaR puisque ce choix dépend fondamentalement de l'horizon de reporting ou d'investissement des opérateurs.

Toutefois, les autorités de régulation peuvent spécifier des horizons de détention spécifiques notamment dans le cadre des procédures de validation de la VaR.

Définition 3.1.1. *Soit $0 < \alpha < 1$, F est la distribution de la perte X d'un investissement dans une période de temps donnée. Les valeurs typiques pour $1 - \alpha$ sont $1 - \alpha = 0.95$ et $1 - \alpha = 0.99$.*

La VaR est le quantile d'ordre $1 - \alpha$ défini par :

$$VaR_\alpha = F^{-1}(1 - \alpha)$$

où F^{-1} est la fonction inverse de F (dans le cas où F n'est pas inversible, on utilise l'inverse généralisé). Notons :

$$\begin{aligned} x_\alpha &= F^{-1}(1 - \alpha) \\ &= \sup \{x \in \mathbb{R} : F(x) \leq 1 - \alpha\} \end{aligned}$$

3.1.4 Méthodes d'estimation de la VaR

Il existe plusieurs méthodes pour estimer la VaR.

1.VaR par la méthode des blocs

C'est une méthode qui est basée sur le théorème de Fisher-Tippet, en supposant que le maximum de l'échantillon suit exactement une loi GEV, et le

quantile extrême x_α est la VaR.

Approche par l'estimateur de Hill

Il suffit d'inverser la fonction H_γ pour obtenir le quantile $x_\alpha = H_\gamma^{-1}(1 - \alpha)$ donc :

$$x_\alpha = \begin{cases} \mu - \frac{\sigma}{\gamma}(1 - (-\log(1 - \alpha))^{-\gamma}) & \text{si } \gamma \neq 0 \\ \mu - \sigma \log(\log(1 - \alpha)) & \text{si } \gamma = 0 \end{cases}$$

donc le quantile extrême est donné par :

$$\begin{aligned} \hat{x}_\alpha^H &= X_{n-k} \left(\frac{n}{k}(1 - \alpha) \right)^{-\gamma_{n,k}^H} \\ &= X_{n-k} + X_{n-k} \left[\left(\frac{n}{k}(1 - \alpha) \right)^{-\gamma_{n,k}^H} - 1 \right] \end{aligned}$$

L'estimation de la Value at Risk est donnée par la formule suivante :

$$\widehat{VaR} = X_{n-k} + X_{n-k} \left[\left(\frac{n}{k}(1 - \alpha) \right)^{-\gamma_{n,k}^H} - 1 \right]$$

où $\gamma_{n,k}^H$ est l'estimateur de Hill.

Approche par l'estimateur de Pickands

$$VaR_\alpha = x_{k,\alpha}^\alpha = \frac{\left(\frac{k}{n(1 - \alpha)} \right)^{\gamma^\alpha} - 1}{1 - 2^{-\gamma^\alpha}} \left(X_{(n-K+1,n)} - X_{(n-2k+1,n)} \right) + X_{(n-k+1,n)}$$

où γ^α est l'estimateur de Pickands.

2.Simulation historique

La simulation historique (Historical Simulation, ou HS) est une méthode très simple d'estimation des mesures de risque fondée sur la distribution empirique des données historiques de rendements. Formellement, la VaR est estimée simplement par lecture directe des fractiles empiriques des rendements passés.

3.1.5 Exemple d'application de calcul de la VaR

Le standard and Poor's 500 (S&P500) est un indice boursier composé de 500 sociétés dominantes cotées sur le marché boursier des États-Unis.

Les sociétés sont sélectionnées par le comité de l'indice S&P500 et cette indice est conçu comme un indicateur important du marché des actions aux États-Unis.

On utilise le logiciel R pour réaliser l'analyse descriptive de l'indice S&P500 et calculer la VaR

3.1.5.1 Analyse descriptive de l'indice financier

On prend les données de l'indice S&P500 dans la période de 01/04/2014 à 30/04/2019.

La figure 3.1 représente la série brute de l'indice S&P500 sur les 5 dernières années.

```
> plot(SP500,xlab="Time",type="l")
```

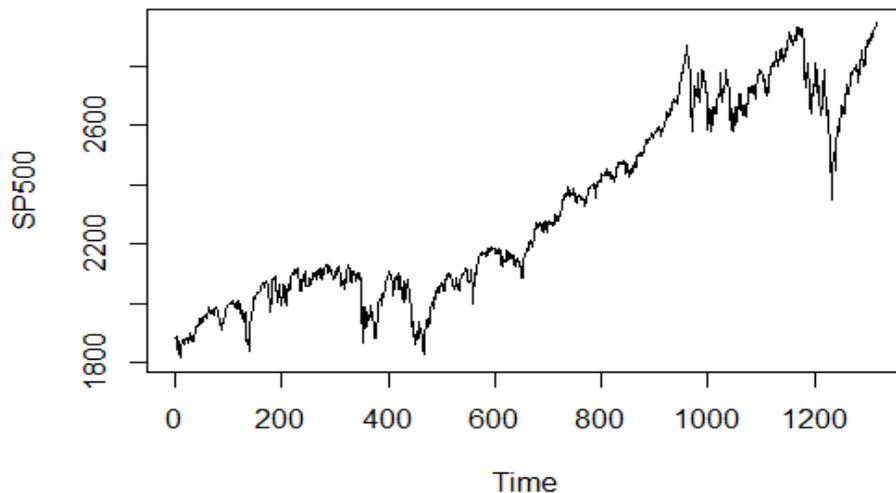


FIGURE 3.1 – Graphique de l'indice S&P500 entre Avril 2014 et Avril 2019

```
> summary(SP500)
```

Min	1st Qu	Median	Mean	3rd Qu	Max
1815	2056	2186	2320	2643	2945

```
> hist(SP500)
```

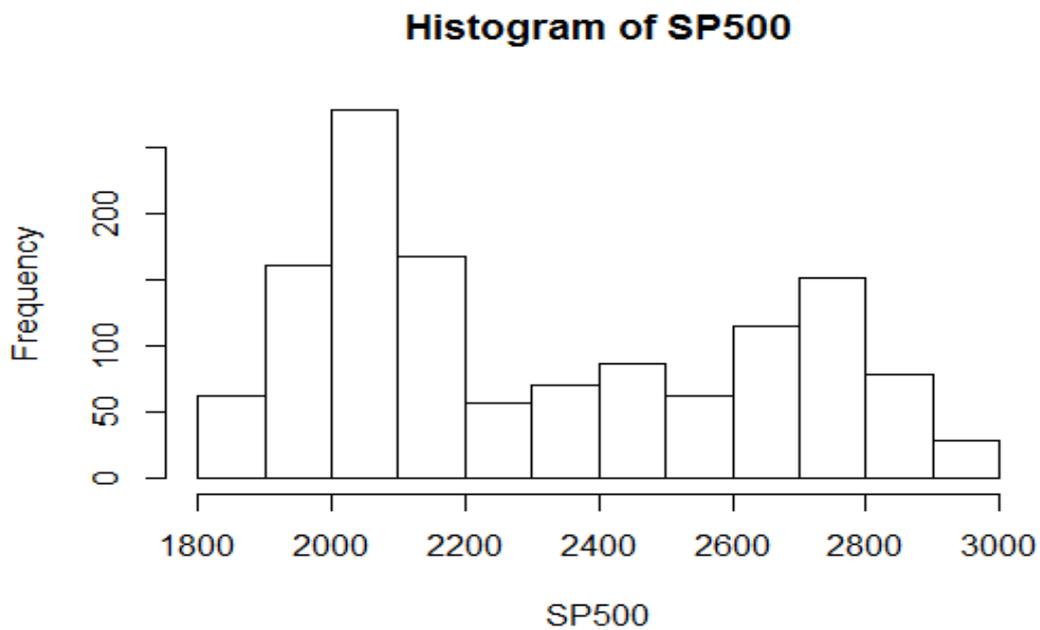


FIGURE 3.2 – Histogramme du S&P500.

3.1.5.2 Analyse descriptive des log des rendements de S&P500

L'objectif de la procédure suivante est d'obtenir une série des logarithmes des rendements puis calculer la VaR. Soit p_t le prix d'un actif à la date t et r_t le logarithme du rendement correspondant :

$$\begin{aligned} r_t &= \log(p_t) - \log(p_{t-1}) \\ &= \log(1 + R_t) \end{aligned}$$

où $R_t = \frac{p_t - p_{t-1}}{p_{t-1}}$ désigne la variation relative des prix.

On considère la série des logarithmes des rendements suivante :

% Insertion des données

```
> log.rend.SP500=diff(log(SP500))
> log.rend.SP500
[1] 0.0026490082 -0.0010587613 -0.0122570146 -0.0107817756 0.0032467561
[6] 0.0112813445 -0.0210534092 -0.0098685011 0.0082304991 0.0065359710
.
.
.
[1306] -0.0006882313 0.0006882313 -0.0024108846 0.0017226533 0.0006882313
[1311] 0.0089041684 -0.0020477823 -0.0003417051 0.0044330848 0.0013600818
[1316] 0.0006793479
```

Le graphe des logarithmes des rendements est donné par :

% Graphe des logs rendements

```
> plot(log.rend.SP500,xlab="Time",type="l")
```

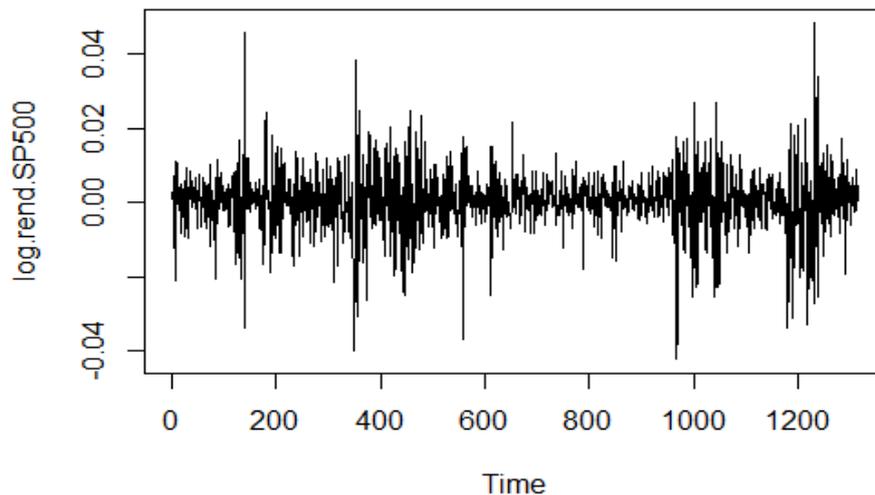


FIGURE 3.3 – Logarithmes des rendements de S&P500

Ainsi que l'histogramme des logarithmes des rendements de l'indice S&P500 et la courbe de loi normale $N(\mu, \sigma^2)$. où $\mu = \text{mean}(\text{log.rend.sp500})$ et $\sigma^2 = \text{sd}(\text{log.rend.sp500})$ sont la moyenne et la variance calculées à partir de la série des logarithmes des rendements.

% Histogramme de logs rendements

```
> hist(log.rend.SP500,prob=T,breaks=500,col=grey(0.9))
> curve(dnorm(x,mean(log.rend.SP500),sd(log.rend.SP500)),add=T,lwd=2)
```

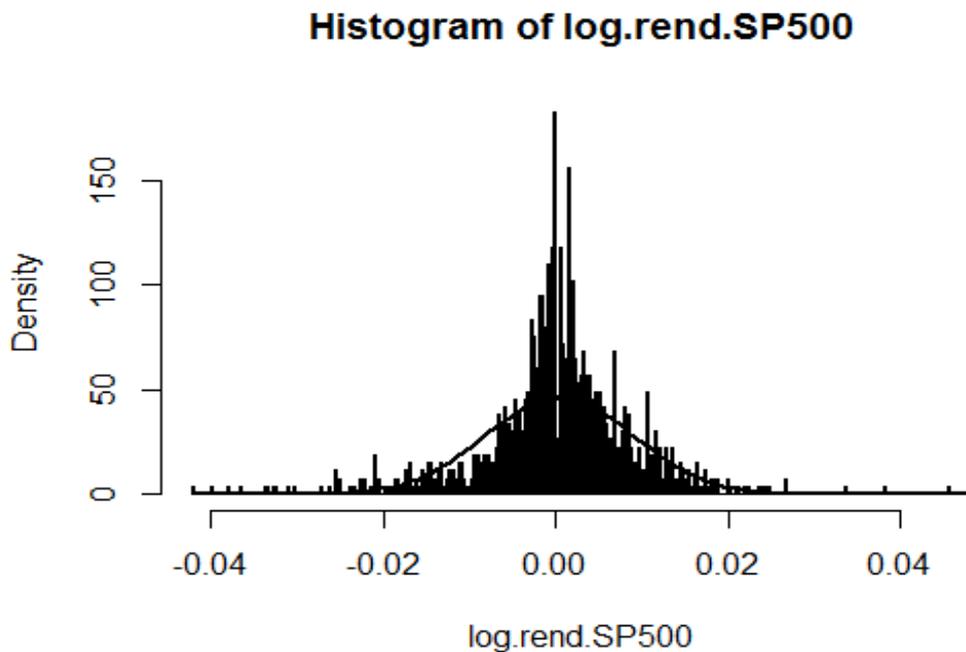


FIGURE 3.4 – Histogramme des logarithmes des rendements

% Analyse descriptive de la série des logs rendements

```
> summary(log.rend.SP500)
```

Min	1st Qu	Median	Mean	3rd Qu	Max
-0.0421504	-0.0028794	0.0004672	0.0003390	0.0046130	0.0481621

L'analyse descriptive des logarithmes des rendements ne permet pas de tirer des conclusions sur la queue de la distribution.

3.1.5.3 La VaR normale

Supposons que les logarithmes des rendements suivent une loi normale $N(\mu, \sigma^2)$. Pour calculer la VaR de ces rendements, il faut calculer le quantile $x_{0.95}$ de $N(0,1)$, et estimer μ et σ^2 par la moyenne et la variance empirique de la série des logarithmes des rendements.

```
> N=length(log.rend.SP500)
> N
[1] 1316

% Simulation d'un échantillon de n variables suivent la loi normale

> x=rnorm(N,mean=0,sd=1)
> x_0.95=quantile(x,0.95)
> x_0.95
95%
1.6424 %Quantile au niveau 95%
> sum=numeric(N);
> sum=0;
> sum2=0;

%Calcul de  $\sigma^2 = sum$  et  $\mu = mean$ 

> for(i in 1:N){
+ sum=sum+x[i]
+ sum2=sum2+x[i]^2}
> mean=sum/N;
> s=sum2/N-mean^2;
> s
[1] 0.9949403
> mean
```

```
[1] -0.01552924
```

On rappelle que la VaR normale est donnée par : $VaR_{normale} = \mu + \sigma \cdot x_{0,95}$

```
> VaR=mean+x_0.95*sqrt(s)
> VaR
95%
1.622711 %var normale au niveau 95%
```

3.1.5.4 La VaR historique

Cette technique utilise la fonction de répartition empirique.

```
> Fn_SP500=ecdf(SP500)
```

Puisque la loi de distribution des rendements est inconnue, on va estimer la VaR des logarithmes des rendements par un quantile empirique, avec un seuil $1 - \alpha = 0,95$.

```
> quantile(log.rend.SP500,0.95)
95%
0.01334147 %Quantile empirique au niveau 95%
```

Pour calculer la VaR sur tout l'historique du S&P500, on utilise :

```
> varh=c()
> for(i in 31:N)
+ {
+ varh[i-30]=quantile(log.rend.SP500[(i-30):i],prob=0.95,na.rm=TRUE)}
> Var_hist=mean(varh)
> Var_hist
[1] 0.01200872 %VaR historique au niveau 95%

> plot(Fn_SP500,main="Fonction de répartition empirique del'indice S&P500")
```

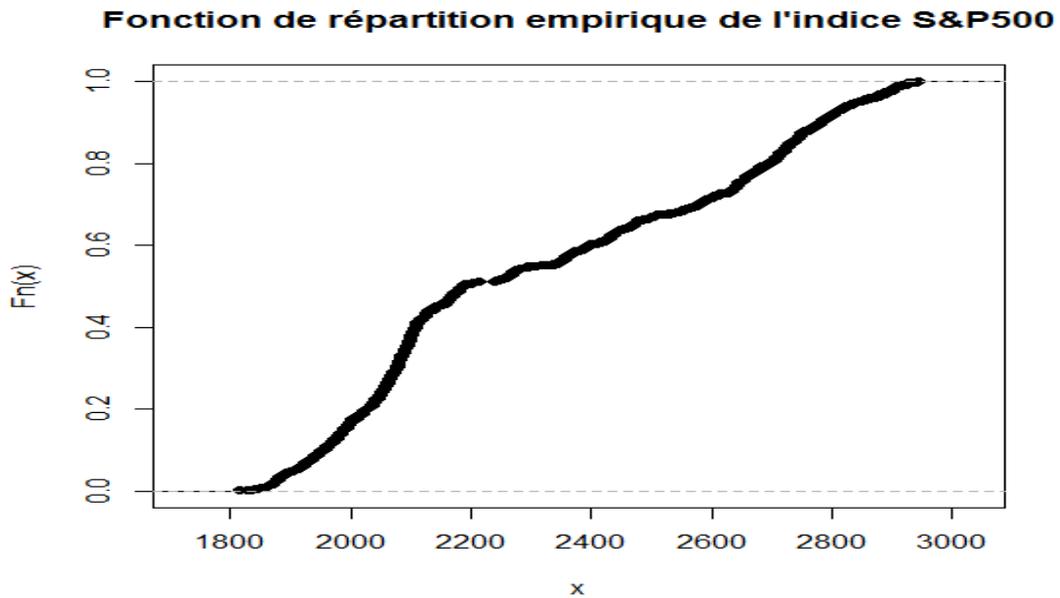


FIGURE 3.5 – La fonction de répartition empirique.

Remarque 3.1.1. *On remarque que les valeurs de la VaR calculées par les deux méthodes sont différentes.*

Ceci est dû au fait que les données ne suivent pas exactement la loi normale. La valeur à garder est la valeur historique.

Pour choisir la bonne valeur de la VaR on peut utiliser plusieurs méthodes comme la méthode P.O.T ou la méthode des blocs.

Remarque 3.1.2. *Les données de l'indice S&P500 sont disponible sur le site suivante : <http://m.investing.com/indices/us-spx-500-historical-data>*

3.2 Assurance

Des compagnies d'assurance ne peuvent rembourser leurs clients quand le montant des dommages dépasse un certain seuil. Pour se protéger, elle font appel à d'autre compagnie qui peuvent payer la différence, cette technique appelée la réassurance et on dit que une assurance au second degré .

En réassurance, l'estimation de la prime pure a pour but d'évaluer, pour chaque assuré ou prospect, le montant attendu des sinistres pour la période d'assurance étudiée. Cette évaluation se fait souvent par des méthodes statistiques. Cependant, les gros coûts de sinistres nécessitent une modélisation d'événements rares, c'est-à-dire les événements avec une faible probabilité d'occurrence, mais avec un montant de sinistre élevé et souvent des effets désastreux.

L'analyse de ces événements extrêmes peut être réalisée à l'aide de la méthodologie des valeurs extrêmes où le comportement de la queue d'une distribution est principalement caractérisé par son indice de la valeur extrême qui permet d'indiquer la taille et la fréquence de certains phénomènes extrêmes dans le cadre d'une distribution de probabilité donnée.

3.3 Hydrologie

Les conséquences des inondations peuvent être désastreuses. Par exemple, en 1953 les inondations furent catastrophiques en Hollande dont 40% du sol est situé au dessous du niveau de mer. Le gouvernement a décidé de construire des digues le long de côtes pour se protéger contre les tempêtes. La hauteur des digues est telle que la probabilité d'occurrence d'une inondation durant une année est de 10^{-4} .

Dans la localité de Delfzijl au nord de la Hollande, 1877 tempêtes furent relevées pendant une période de 111 années. Les observations sont supposées i.i.d.

La hauteur x des digues doit vérifier :

$$P(X > x) = \frac{111}{1877} 10^{-4}$$

En hydrologie, on s'intéresse aussi à l'intensité des précipitations ; ce paramètre est important pour concevoir le système des eaux. Plus précisément, la capacité effective de chaque système est déterminée par des valeurs extrêmes de l'intensité.

Conclusion

Dans ce mémoire, nous avons défini les principaux estimateurs de l'indice des valeurs extrêmes dans le cas de variables aléatoires indépendantes et identiquement distribuées.

De cette étude, il en ressort qu'on ne peut opter pour un des trois estimateurs étudiés. Ces dernières années d'autres estimateurs de l'indice des valeurs extrêmes ont été définis sans parvenir à élire le meilleur estimateur. En pratique, il est conseillé d'utiliser plusieurs estimateurs et non pas se contenter d'un seul.

Ces estimateurs surtout Hill et Pickands sont utilisés pour l'estimation de la VaR.

Plusieurs méthodes sont utilisées pour estimer la VaR.

Il est conseillé d'utiliser beaucoup de méthodes pour choisir la meilleure valeur de la VaR.

Bibliographie

- [1] De Haan Laurent, Ferreira Ana (2006). *Extreme Value Theory an introduction*. Springer-Verlag series in operations research and financial Engineering.
- [2] Delmas Jean François, Benjamin Jourdain(2005). *Modèles aléatoires. Applications aux sciences de l'ingénieur et du vivant*. Champs sur marne.
- [3] Embrechts P, Kluppelberg C, Mikosch T (1997). *Modeling Extremal Events for Insurance and Finance*. Springer, Berlin.
- [4] Hurlin.Ch(2007). *Master Econometrie Et Statistique Appliquée (ESA)*. Université d'Orléans.
- [5] Gardes Laurent(2017). *Théorie des valeurs extrêmes*. Université de Strasbourg.
- [6] Kaidi.O, Belarb.R(2017). *Distribution des valeurs extrêmes et Estimation de la Value at Risk*. Université M'hamed Bougerra,Boumerdès.
- [7] Lani Abdnour(2017). *Etude et comparaison d'estimateurs de l'indice des valeurs extrêmes* . édition universitaires européennes.
- [8] Pickands,J(1975). *Statistical Inference Using Extreme Order Statistics*. The Annals of Statistics,3.
- [9] Reiss.R.D, Thomas.M(2007)*Statistical Analysis of Extreme Value, with applications to insurance, finance, l'hydrologie and other fields*. FB mathematik universitat slegan.walter-flex-str.Slegan Germany.