ROYAUME DU MAROC
Université Mohammed V - RABAT -
جامعة محمد الخامس – الرباط –
Faculté des sciences
كلية العلوم
CENTRE D'ETUDES DOCTORALES - SCIENCES ET TECHNOLOGIES

N° d'ordre 3314

THESE

En vue de l'obtention du : ***DOCTORAT***

Centre de Recherche : Centre de Recherches Mathématiques et Applications de Rabat
Structure de Recherche : Laboratoire Mathématiques, Statistique et Applications
Discipline : Mathématiques
Spécialité : Statistiques.

Présentée et soutenue le : 17/07/2020 par :

Bouchra LABLOUL

Valeurs Extrêmes sous Normalisation non Linéaire avec Application en Finance

JURY

Khalid EL HIMDI	PES Faculté des Sciences, Université Mohammed V de Rabat	Président
Mohammadine BELBACHIR	PES Faculté des Sciences, Université Mohammed V de Rabat	Directeur de thèse
Saad EL OUARDIRHI	PES Faculté des SJESS, Université Mohammed V de Rabat	Co-Directeur de thèse
Abdelouahid IMLAHI	PES Faculté des Sciences, Université Abdelmalek Essaâdi de Tanger	Rapporteur/Examinateur
Abdelhak ZOGLAT	PES Faculté des Sciences, Université Mohammed V de Rabat	Rapporteur/Examinateur
Abdellah OULDKHAL	PES Faculté des Sciences, Université Mohammed V de Rabat	Rapporteur/Examinateur

.

Année Universitaire : 2019-2020

Résumé

La théorie des valeurs extrêmes (TVE) est une discipline statistique qui développe des techniques et des modèles pour décrire ce qui est inhabituel. Sa naissance remonte au début du 20-ème siècle, elle confirme que sous certaines condition la loi asymptotique des maximums appartient nécessairement à une famille de lois extrêmes (GEV : Généralisée des Valeurs Extrêmes), cette famille contient la loi de Weibull, Gumbel et Fréchet. La TVE pour la normalisation non-linéaire qui est le sujet de thèse est initiée par (Mohan et Ravi, 1993) et (Pancheva, 1984), (Ravi et Mavitha, 2016) ont défini les lois limites pour les extrêmes sous une nouvelle normalisation non linéaire appelée la normalisation exponentielle. On a énoncé des nouveaux résultats principaux en utilisant la distance en variation totale et l'approximation de la loi binomiale par la loi de Poisson, la vitesse de convergence est rapide, une application sous forme d'un exemple est présenté pour comparer notre résultat à celle de (Barakat et al., 2010). Une comparaison de la performance du modèle GEV (Generalized Extrême Value) pour le cas linéaire et non-linéaire (GEVL, GEVP) avec application en finance est présentée dans ce travail, cette étude nous a permis de dire que le modèle GEVP est plus performant pour l'indice du marché CAC 40 et plus réaliste que le modèle GEVL.

Abstract

Extreme value theory (TVE) is a statistical discipline who develops techniques and models to describe what is unusual. Its birth goes back to the 20-th century, confirms that under certain conditions the asymptotic law maximums necessarily belongs to a family of extreme laws (GEV : Generalized Extreme Values), this family contains the law of Weibull , Gumbel and Fréchet. The TVE for non-linear normalization which is the subject of the thesis is initiated by (Mohan and Ravi, 1993) and (Pancheva, 1984), (Ravi and Mavitha , 2016) defined the limit laws for the extremes under a new nonlinear normalization called exponential normalization. We have stated new main results using the total variation distance and the binomial approximation by Poisson's law, the convergence speed is fast, an application in the form of an example is presented to compare our result to that of (Barakat et al., 2010). A comparison of the performance of the GEV (Generalized Extreme Value) for the linear and non-linear case (GEVL, GEVP) with Application in finance is presented in this work, that you study we allowed to say that GEVP model is more efficient for the index CAC 40 market and more realistic than the GEVL model.

Avant propos

Ce travail a été réalisé au Laboratoire de Mathématiques, Statistique et Applications (LMSA) à la Faculté des Sciences de Rabat sous la direction de Monsieur Mohammadine Belbachir, professeur d'enseignement supérieur.

Je tiens d'abord à exprimer mon respect et mes sincères remerciements à mon directeur de thèse Monsieur Belbachir Mohammadine, pour avoir accepté d'encadrer cette thèse. Un grand merci pour son extrême patience au cours de ces cinq dernières années. Je lui suis profondément reconnaissante pour ses nombreux conseils et son soutien indéfectible tout au long de ce travail. Il m'a donné l'occasion de travailler sur un thème original et intéressant, me permettant de mieux comprendre les concepts théoriques de la statistique par leur application à des cas concrets.

Je voudrais également exprimer ma profonde gratitude à mon co-encadrant Monsieur Saad El Ouardirhi, professeur à la Faculté des Sciences Juridiques, Economiques et Sociales pour avoir suivi mon travail et m'avoir fait bénéficier de ses précieuses remarques et conseils.

Je tiens également à remercier Monsieur El Himdi Khalid, professeur à la Faculté des Sciences de Rabat, de m'avoir fait l'honneur de présider le

jury de cette thèse.

Je suis très honorée que Monsieur Abdelhak Zoglat a accepté la si difficile tâche de rapporteur. Merci sincèrement du temps et de l'énergie qu'il a consacré â la lecture de mon travail. Je le remercie d'avoir accepté d'examiner ma thèse

J'exprime mes profonds remerciements à Monsieur Imlahi Abdelouahid, professeur à la Faculté des Sciences de Tanger qui a accepté de juger cette thèse en qualité de rapporteur et examinateur. Je le remercie pour l'intérêt et la considération qu'il a portés à ce travail.

Je souhaite également remercier profondément Monsieur Abdellah Ouldkhal, professeur à la Faculté des Sciences de Rabat d'avoir accepté d'examiner cette thèse. Je le remercie d'avoir accepté de juger cette thèse en qualité de rapporteur.

Je remercie également Monsieur Guei Cyrille Okou, professeur à l'Université Jean Lorougnon Guédé de Daloa Côte d'ivoire pour ses précieux conseils.

Je souhaiterais exprimer toute ma reconnaissance envers les professeurs de Mathématiques de la Faculté des Sciences de Rabat

Enfin, je ne voudrais pas terminer sans exprimer ma gratitude à mes soeurs et à mes parents, pour m'avoir soutenue moralement et matériellement tout au long de mes études. Je leur dédie ce travail.

Bouchra Labloul

Table des matières

Table des matières **5**

Introduction **7**

1 La théorie des valeurs extrêmes (T.V.E) **11**

1.1 Les distributions limites pour les maxima 11

1.2 Domaines d'attraction 14

1.3 Distribution de Pareto Généralisée GPD 19

1.4 Estimation des paramètres 21

2 Théorie des valeurs extrêmes pour une normalisation non linéaire **26**

2.1 Les lois limites possibles du maximum 27

2.2 Domaine d'attraction dans le cas de la normalisation puissance 29

2.3 Un estimateur pour les paramètres des lois extrêmes 33

3 Notions de convergence des sommes de variables aléatoires i.i.d **36**

3.1 Propriétés de la distance de variation totale 37

3.2 Convergence de sommes de variables aléatoires i.i.d de Bernouilli vers la loi de Poisson 39

3.3 Distance en variation totale pour le cas multidimensionnel. . 40

4 Sur la vitesse de convergence sous la normalisation exponentielle 43

4.1 Introduction 43

4.2 Résultat principaux 49

4.3 Remarques 52

4.4 Exemple 53

5 Fonctions de normalisation et fiabilité de l'estimation du niveau de retour GEV : Une étude comparative basée sur l'indice CAC 40 55

5.1 Introduction 55

5.2 Méthodologie 59

5.3 Estimation du niveau de retour pour une distribution extrême dans les deux cas de normalisation 64

5.4 Résultats et discussion 68

6 Conclusion Générale 78

Bibliographie 80

Introduction Générale

La théorie des valeurs extrêmes est une discipline statistique qui développe des techniques et des modèles pour décrire ce qui est inhabituel. Sa naissance remonte au début du 20-ème siècle. Les premiers développements de cette théorie sont attribués à Nicolas Bernoulli en 1709. Ce n'est pas un hasard si les premières applications des modèles de valeurs extrêmes s'appliquaient principalement dans le domaine du génie civil.
Les ingénieurs avaient toujours été tenus de concevoir leurs structures de manière à résister aux forces susceptibles de les affecter. La théorie des valeurs extrêmes a fourni un cadre dans lequel une estimation des forces anticipées pouvait être réalisée à l'aide de données historiques.

Les domaines d'application sont en effet très variés : hydrologie, météorologie, biologie, ingénierie, gestion de l'environnement, finance, assurance, sciences sociales, etc..
Citons quelques domaines d'application des valeurs extrêmes :

Catastrophes naturelles : Des inondations, des pluies fortes, des températures extrêmes, des pressions atmosphériques extrêmes, des vents forts et d'autres phénomènes pouvant causer des grosses pertes humaines et matérielles.

La pollution de l'air : La concentration de polluants atmosphériques est exprimée en termes de proportion d'un polluant spécifique dans l'air. Les concentrations sont enregistrées à intervalles de temps réguliers, le devoir des autorités est de prendre les mesures nécessaires pour que ces taux de pollution restent en dessous d'un certain seuil raisonnable.

FIGURE 0.1 – Vues aériennes d'Oude Tonge, au nord de la Zélande, le 3 février 1953. Source : Wikipédia.

Le 4 octobre 1986 marqua l'achèvement du plan Delta aux Pays-Bas. Il s'agit du plus grand chantier de génie civil de tous les temps. Il permit de relier toutes les îles côtières de la province de Zélande par des digues.
Le but serait d'être en mesure de prédire l'apparition de tels phénomènes, leurs impacts et le retour de ces phénomènes. Pour cela, il convient de définir précisément ce qu'est un événement extrême. Il s'agit d'événement qui a une faible probabilité de se produire mais qui, lorsqu'il se produit, prend de très petites ou de très grandes valeurs et a un grand impact. La différence avec un événement rare est que ce dernier a une probabilité d'occurrence faible. Le fait qu'un événement soit rare n'implique pas qu'il soit extrême. Tout événement extrême est rare au sens où il a une faible probabilité de se produire.

La théorie des valeurs extrêmes fournit une base mathématique probabiliste, qui nous offre des outils très puissants permettant de construire des modèles statistiques aidant à prévoir l'intensité et la fréquence des événements extrêmes.

Dans **le premier chapitre** nous présentons les principaux résultats et définitions de la théorie des valeurs extrêmes pour la normalisation linéaire en se basant sur les travaux de Fisher-tippett [22], Kotz et Nadarajah [31] et Reiss [47]. Après avoir rappelé les lois limites extrêmes du maximum dûment normalisé, nous étudions les domaines d'attraction de ces lois limites. Nous donnons ensuite différents estimateurs des paramètres de ces lois. A la fin du chapitre, nous évoquons la méthode POT (Peak Over Threshold) consistant à considérer les dépassements d'un seuil élevé.

Le second chapitre s'intéresse aux lois limites asymptotiques du maximum d'un échantillon de variables aléatoires i.i.d, dans le cas d'une normalisation non linéaire, en l'occurrence la normalisation exponentielle. Cette dernière a été initiée par Pancheva [38] et largement étudiée par Barakat [2]. L'étude des domaines d'attraction pour ces différentes lois limites est menée dans ce chapitre.

Le **troisième chapitre** traite les notions de base de la distance à variation totale pour calculer la vitesse de convergence.

Dans le **quatrième chapitre**, nous évaluons la vitesse de convergence du maximum vers les différentes lois limites, dans le cas de la normalisation puissance, l'outil utilisé à cette fin est la distance à variation totale. Nous donnons des exemples illustrant cette vitesse.

Le cinquième et dernier chapitre est consacré à l'estimation du niveau de retour dans le cas d'une normalisation linéaire et celui d'une normalisation non linéaire. Une application de ces résultats au CAC 40 est

présentée. Une comparaison significative fournit les avantages et les inconvénients de chacune de ses deux normalisations.
La conclusion de ce mémoire exposera les perspectives à donner à nos travaux.

Chapitre 1

La théorie des valeurs extrêmes (T.V.E)

Dans ce chapitre, nous exposons des théorèmes de base de la T.V.E. Nous donnons les différentes lois limites du maximum dûment normalisé, pour une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.). Cette loi, notée GEV (Generalized Extreme Value) dépend principalement d'un paramètre, appelé indice des valeurs extrêmes et noté ξ, qui décrit le comportement de la queue de la loi considérée. Les lois limites pour le minimum peuvent être déduites de celles du maximum en utilisant le fait que
$min(X_1, ..., X_n) = -max(-X_1, ..., -X_n)$.

1.1 Les distributions limites pour les maxima

Nous commençons par définir la notion de la loi max-stable.

Définition 1.1.1. *Une variable aléatoire X est appelée max-stable si pour*

tout $n \geq 2$, il existe deux constantes appropriées $a_n > 0$, $b_n \in \mathbb{R}$ telles que

$$M_n =^d a_n X + b_n \tag{1.1}$$

où $M_n = max(X_1, X_2...., X_n)$ et $X =^d Y$ signifie que X et Y ont même loi de probabilité.

Théoréme 1.1.1. *La classe des distributions max-stables coïncide avec la classe de toutes les lois limites possibles pour les maxima des variables aléatoires i.i.d.*

Pour une démonstration de ce théorème voir (Embrechts et al. [20]).
Le résultat suivant est la base de la théorie classique des valeurs extrêmes.

Théoréme 1.1.2. *(**Fisher-Tippett theorem**) Soit (X_n) une suite de variables aléatoires i.i.d. S'il existe des constantes de normalisation $a_n > 0$, $b_n \in \mathbb{R}$ et H une fonction de répartition non-dégénérée telles que :*

$$a_n^{-1}(M_n - b_n) \to^d H, \tag{1.2}$$

avec $M_1 = X_1$, $M_n = max(X_1, ..., X_n)$, $n \geq 2$ et $\to^d$ désigne la convergence en loi, alors H peut prendre l'un des trois types des distributions suivantes :

-Loi de Fréchet :

$$\phi_\alpha(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\alpha}\}, & x > 0 \end{cases} \tag{1.3}$$

-Loi de Weibull :

$$\psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}, & x \leq 0 \\ 1, & x > 0 \end{cases} \tag{1.4}$$

-Loi de Gumbel :

$$\Lambda(x) = \exp\{-e^{-x}\}, x \in \mathbb{R} \tag{1.5}$$

Ces trois distributions $\Lambda(x)$, $\psi_\alpha(x)$, $\phi_\alpha(x)$ sont appelées les distributions des valeurs extrêmes standard. Une démonstration détaillée peut être trouvée dans Resnick [48] et Embrechts et al. [19].

La Figure 1.1 ci-dessous fournit une idée sur la représentation graphique de ces lois.

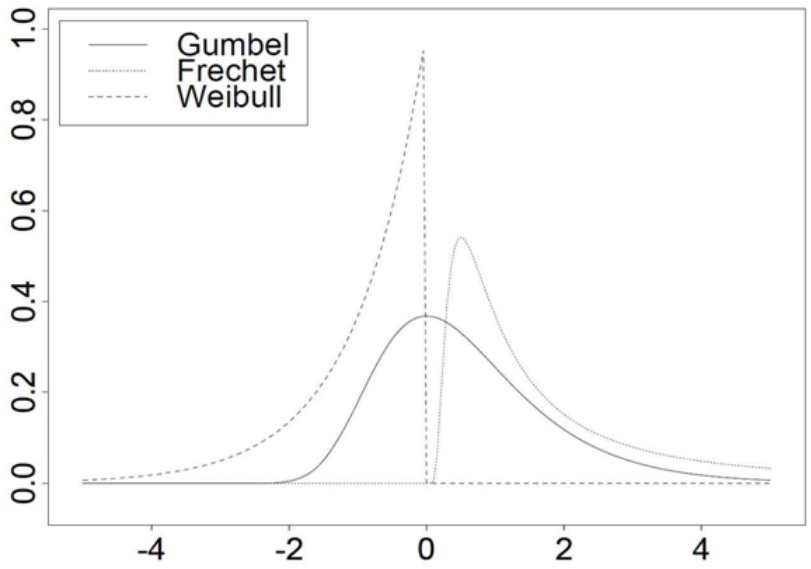

FIGURE 1.1 – Densités de Lois des Valeurs Extrêmes

Von Mises en 1936 et Jenkison en 1955 ont fourni une caractérisation synthétique des lois extrêmes : la distribution généralisée des valeurs extrêmes ou en anglais "Generalized Extreme Value, GEV" est définie par :

$$H_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-\frac{1}{\xi}}\} & \text{si} \quad \xi \neq 0. \\ \exp\{-\exp(-x)\} & si \quad \xi = 0. \end{cases} \tag{1.6}$$

avec $1 + \xi x > 0$, le paramètre ξ est appelé "indice de queue".

1.2 Domaines d'attraction

Définition 1.2.1. *(Domaine d'attraction). On dit qu'une distribution F appartient au domaine d'attraction du maximum de la distribution H_ξ, et on note $F \in D(H_\xi)$, s'il existe deux suites de constantes $a_n > 0$ et $b_n \in \mathbb{R}$ telles que $a_n^{-1}(M_n - b_n) \to^d H_\xi$.*

Proposition 1.2.1. *La fonction de répartition F appartient au max-domaine d'attraction de la distribution des valeurs extrêmes H, si et seulement si*

$$\lim_{x\to\infty} n\bar{F}(a_n x + b_n) = -\ln H(x), x \in \mathbb{R},$$

avec $0 < H(x) < 1$. $a_n > 0$ et b_n étant les constantes de normalisation. Où $\bar{F} = 1 - F$ est la fonction de survie.

Le théorème suivant donne des conditions nécessaires et suffisantes pour l'appartenance de F à un domaine d'attraction des lois extrêmes.
Pour plus de détail voir (Leadbetter et al. [33]).

Théoréme 1.2.1. *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d de fonction de répartition F.*
Les conditions nécessaires et suffisantes pour que $F \in D_l\{\phi_\alpha, \psi_\alpha, \Lambda\}$ sont :

1. *Loi de Fréchet :*
   $x_F = \infty$ *et* $\lim_{t\to\infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-\alpha}, \alpha > 0$, *pour tout* $x > 0$;
2. *Loi de Weibull :*
   $x_F < \infty$ *et* $\lim_{h\to 0} \frac{\bar{F}(x_F - xh)}{\bar{F}(x_F - h)} = x^{\alpha}, \alpha > 0$, *pour tout* $x > 0$;
3. *Loi de Gumbel :*
   il existe une fonction $g(t)$ strictement positive telle que

$$\lim_{t\to x_F} \frac{\bar{F}(t + xg(t))}{\bar{F}(t)} = \exp(-x)$$

où $x_F = sup\{x \in \mathbb{R} : F(x) < 1\}$ *est le point terminal de* F.

Pour caractériser le domaine d'attraction de chacune des lois limites, nous donnons tout d'abord la définition d'une fonction à variation régulière :

Définition 1.2.2. *Une fonction* $f : \mathbb{R}^+ \to \mathbb{R}$, *qui peut être éventuellement positive, est à variation régulière (à l'infini) si pour un réel* α, $\lim_{t\to+\infty} \frac{f(tx)}{f(t)} = x^\alpha$, $x > 0$. *On note* $f \in RV_\alpha$.
Lorsque $\alpha = 0$, f *est dite à variation lente à l'infini.*

Nous fournissons des conditions sur F pour qu'elle appartienne à l'un des trois domaines d'attraction définis précédemment.

Théoréme 1.2.2. *(Domaine d'attraction de la loi de Fréchet* ϕ_α*). La fonction de répartition* F *appartient au max-domaine d'attraction de la loi de Fréchet* ϕ_α, $\alpha > 0$, *si et seulement si* $\bar{F}(x) = x^{-\alpha}L(x)$ *avec* L *fonction à variation lente. On note* $F \in MDA(\phi_\alpha)$ *et alors*

$$a_n^{-1} M_n \to^d \phi_\alpha \tag{1.7}$$

avec $\alpha > 0$, $a_n = F^{\leftarrow}(1 - n^{-1})$. $F^{\leftarrow}$, *la fonction inverse généralisée de* F, *est définie par* $F^{\leftarrow}(y) = inf\{x : F(x) \geq y\}$

Théoréme 1.2.3. *(Domaine d'attraction de la loi de Weibull). La fonction de distribution* F *appartient au max-domaine d'attraction de la loi Weibull* ψ_α, $\alpha > 0$, *si et seulement si* $x_F < \infty$ *et* $\bar{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$ *pour une fonction à variation lente* L. *On note* $F \in MDA(\psi_\alpha)$ *et alors*

$$a_n^{-1}(M_n - b_n) \to^d \psi_\alpha, \tag{1.8}$$

avec $a_n = x_F - F^{\leftarrow}(1 - n^{-1})$ *et* $b_n = x_F$ *des constantes de normalisation.*

Les résultats concernant le domaine d'attraction de la loi de Gumbel sont plus délicats, puisqu'il n'y a pas de représentation simple pour les lois appartenant au domaine d'attraction de Gumbel. Nous rappelons tout d'abord la définition d'une fonction de Von Mises.

Définition 1.2.3. *(**Fonction de Von Mises**) La fonction de répartition F est dite fonction de Von Mises de fonction auxiliaire a, s'il existe un réel $z < x_F$ tel que*

$$\bar{F}(x) = c \exp\left\{ - \int_z^x \frac{1}{a(t)} \, \mathrm{d}t \right\}, z < x < x_F$$

où c est une constante positive et a une fonction positive absolument continue de dérivée a' vérifiant, $\lim_{x \to x_F} a'(x) = 0$.

Théoréme 1.2.4. *(Domaine d'attraction de la loi de Gumbel). La fonction de répartition F appartient au max-domaine d'attraction de la loi Gumbel Λ, si et seulement si*

$$\bar{F}(x) = c(x) \exp\left\{ - \int_z^x \frac{g(t)}{a(t)} \, \mathrm{d}t \right\}, z < x < x_F$$

où c et g sont deux fonctions mesurables satisfaisant $c(x) \to c > 0$ et $g(x) \to 1$, quand $x \to x_F$ et a est une fonction positive, absolument continue (par rapport à la mesure de Lebesgue) de dérivée a' ayant $\lim_{x \to x_F} a'(x) = 0$. Dans ce cas, un choix possible pour les constantes de normalisation est $a_n = x_F - F^{-1}(1 - \frac{1}{n})$ et $b_n = \frac{1}{\bar{F}(a_n)} \int_{a_n}^{x_F} \bar{F}(y) \mathrm{d}y$.

Les tableaux ci-dessous présentent quelques lois qui appartiennent au max-domaine d'attraction de Fréchet, Weibull et Gumbel.

Maximum domain of attraction of the Fréchet distribution.

Fréchet	$\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}, \quad x > 0, \alpha > 0.$
MDA (Φ_α)	$x_F = \infty, \qquad \overline{F}(x) = x^{-\alpha} L(x), \quad L \in \mathcal{R}_0.$
Norming constants	$c_n = F^{\leftarrow}(1 - n^{-1}) = n^{1/\alpha} L_1(n), \quad L_1 \in \mathcal{R}_0, \quad d_n = 0.$
Limit result	$c_n^{-1} M_n \xrightarrow{d} \Phi_\alpha$
Examples	
Cauchy	$f(x) = (\pi(1 + x^2))^{-1}, \quad x \in \mathbb{R}.$ $c_n = n/\pi$
Pareto Burr stable with index $\alpha < 2$	$\overline{F}(x) \sim K x^{-\alpha}, \quad K, \alpha > 0.$ $c_n = (Kn)^{1/\alpha}$
Loggamma	$f(x) = \frac{\alpha^\beta}{\Gamma(\beta)} (\ln x)^{\beta-1} x^{-\alpha-1}, \quad x > 1, \quad \alpha, \beta > 0.$ $c_n = \left((\Gamma(\beta))^{-1} (\ln n)^{\beta-1} n\right)^{1/\alpha}$

Maximum domain of attraction of the Weibull distribution.

Weibull	$\Psi_\alpha(x) = \exp\{-(-x)^\alpha\}, \quad x < 0, \quad \alpha > 0.$
MDA (Ψ_α)	$x_F < \infty, \quad \overline{F}\left(x_F - x^{-1}\right) = x^{-\alpha} L(x), \quad L \in \mathcal{R}_0.$
Norming constants	$c_n = x_F - F^{\leftarrow}(1 - n^{-1}) = n^{-1/\alpha} L_1(n), \quad L_1 \in \mathcal{R}_0, \quad d_n = x_F.$
Limit result	$c_n^{-1}(M_n - x_F) \xrightarrow{d} \Psi_\alpha$
Examples	
Uniform	$f(x) = 1, \quad x \in (0,1).$ $c_n = n^{-1}, \quad d_n = 1$
Power law behaviour at x_F	$\overline{F}(x) = K(x_F - x)^\alpha, \quad x_F - K^{-1/\alpha} \le x \le x_F, \quad K, \alpha > 0.$ $c_n = (K\,n)^{-1/\alpha}, \quad d_n = x_F$
Beta	$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}, \quad 0 < x < 1, \quad a, b > 0.$ $c_n = \left(n \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b+1)}\right)^{-1/b}, \quad d_n = 1$

Maximum domain of attraction of the Gumbel distribution.

Gumbel	$\Lambda(x) = \exp\{-e^{-x}\}\,, \quad x \in \mathbb{R}\,.$
MDA(Λ)	$x_F \le \infty\,, \quad \overline{F}(x) = c(x)\exp\left\{-\int_{x_0}^{x} \frac{g(t)}{a(t)}\,dt\right\}, \quad x_0 < x < x_F\,,$ where $c(x) \to c > 0\,, \quad g(x) \to 1\,, \quad a'(x) \to 0 \quad$ as $x \uparrow x_F\,.$
Norming constants	$d_n = F^{\leftarrow}(1-n^{-1})\,, \quad c_n = a\,(d_n)\,.$
Limit result	$c_n^{-1}(M_n - d_n) \xrightarrow{\mathrm{d}} \Lambda$
Examples	
Exponential–like	$\overline{F}(x) \sim Ke^{-\lambda x}\,, \quad K\,, \lambda > 0\,.$ $c_n = \lambda^{-1}$ $d_n = \lambda^{-1}\ln(Kn)$
Weibull–like	$\overline{F}(x) \sim Kx^{\alpha}\exp\{-cx^{\tau}\}\,, \quad K\,, c\,, \tau > 0\,, \alpha \in \mathbb{R}.$ $c_n = (c\tau)^{-1}\ \left(c^{-1}\ln n\right)^{1/\tau - 1}$ $d_n = \left(c^{-1}\ln n\right)^{1/\tau} + \frac{1}{\tau}\left(c^{-1}\ln n\right)^{1/\tau-1}\left\{\frac{\alpha}{c\tau}\ln(c^{-1}\ln n) + \frac{\ln K}{c}\right\}$
Gamma	$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\,x^{\alpha-1}\,e^{-\beta x}\,, \quad x > 0\,, \quad \alpha\,, \beta > 0\,.$ $c_n = \beta^{-1}$ $d_n = \beta^{-1}(\ln n + (\alpha-1)\ln\ln n - \ln\Gamma(\alpha))$
Normal	$\varphi(x) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,, \quad x \in \mathbb{R}.$ $c_n = (2\ln n)^{-1/2}$ $d_n = \sqrt{2\ln n} - \frac{\ln(4\pi) + \ln\ln n}{2(2\ln n)^{1/2}}$
Lognormal	$f(x) = \frac{1}{\sqrt{2\pi}\sigma x}\,e^{-(\ln x - \mu)^2/2\sigma^2}\,, \quad x > 0\,, \quad \mu \in \mathbb{R}\,, \sigma > 0\,.$ $c_n = \sigma(2\ln n)^{-1/2}d_n$ $d_n = \exp\left\{\mu + \sigma\left(\sqrt{2\ln n} - \frac{\ln(4\pi) + \ln\ln n}{2(2\ln n)^{1/2}}\right)\right\}$

1.3 Distribution de Pareto Généralisée GPD

L'approche basée sur la distribution GEV peut être réductrice du fait que l'utilisation d'un seul maximum conduit à une perte d'information dans les autres grandes valeurs de l'échantillon.

Pickands [43] a introduit la méthode des dépassements de seuil appelée méthode POT (Peak Over Threshold). Elle est aussi appelée méthode des excès au-delà d'un certain seuil réel u si $u < x_F$, x_F étant le point terminal.

Cette approche POT permet de conserver plus d'information que dans la méthode des maxima par blocs. Tout d'abord il faut choisir un seuil $u \in \mathbb{R}$ et considérer les variables Y_j comme les excès correspondants définis, à partir des v.a i.i.d $X_1, ..., X_n$, par $Y_j = X - u > 0$ pour $1 \leq j \leq N_u$, où N_u est le nombre de dépassements du seuil u, c'est-à-dire $N_u = Card\{X_i > u : i = 1, ..., n\}$.

La figure ci-dessous représente les excès Y issus des dépassements X au-delà d'un seuil u.

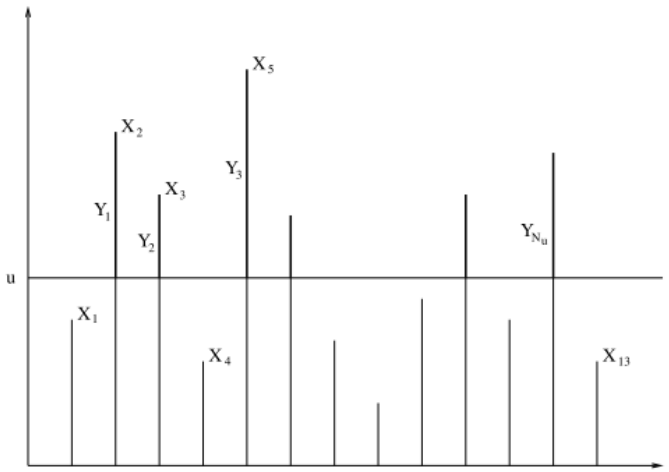

FIGURE 1.2 – Les excès Y issus des dépassements X au-delà d'un seuil u.

Nous utilisons les propriétés de l'approximation de la loi binomiale par la loi de Poisson pour déterminer la loi de probabilité des excès. Soit $\{X_i > u_n\}$ l'événement "Dépassement du seuil u_n par la variable aléatoire X_i". Posons :

$$\mathbb{I}_{\{X_i>u_n\}} = \begin{cases} 1 \quad si \quad X_i > u_n \\ 0 \quad sinon \end{cases} \tag{1.9}$$

Alors le nombre de dépassements de u_n par les v.a X_i, $S_{u_n} := \sum_{i=1}^{n} \mathbb{I}_{\{X_i>u_n\}}$ suit la loi binomiale $B(n, \bar{F}(u_n))$ qui peut être approchée par la loi de Poisson $P(\tau)$. Dans ce cas $S_n \to^d Y_\tau$ où $Y_\tau \sim \mathbf{P}(\tau)$ si et seulement si $n\bar{F}(u_n) \to \tau$, lorsque $n \to +\infty$, $\mathbf{P}(\tau)$ étant la loi de Poisson de paramètre $\tau > 0$

Définition 1.3.1. *Soit X une variable aléatoire de fonction de répartition F. Pour $u < x_F$ fixé, on définit la fonction de répartition conditionnelle des excès au-delà du seuil u par*

$$F_u(x) = P(X - u \leq x | X > u), x \geq 0, \tag{1.10}$$

avec $x_F = sup\{x : F(x) < 1\}$ le point terminal.

Cette approche a été introduite par de Haan et Rootzén [14]. Cette loi peut être approchée par une distribution de Pareto (GPD) notée $F_{\xi,\mu,\sigma}$ définie comme suit

$$F_{\xi,\mu,\sigma}(x) = \begin{cases} 1 - \left(1 + \xi\frac{x-\mu}{\sigma}\right)^{\frac{-1}{\xi}} & \text{pour} \quad \xi \neq 0. \\ 1 - \exp\left\{-\frac{x-\mu}{\sigma}\right\} & pour \quad \xi = 0. \end{cases} \tag{1.11}$$

avec $1 + \xi\frac{x-\mu}{\sigma} > 0$. Les paramètres μ, σ et ξ sont respectivement, les paramètres de position, d'échelle et de forme.

On peut se référer par exemple à l'ouvrage de Embrechts et al. [20] pour plus de détails.

Cette distribution joue un rôle important, dans l'analyse statistique des événements extrêmes, en fournissant une approximation appropriée pour l'excès au-delà d'un grand seuil.
Balkema et de Haan [1] et Pickands [43] ont proposé le théorème ci-après, qui précise la distribution conditionnelle des excès lorsque le seuil déterministe tend vers le point terminal x_F.

Théoréme 1.3.1. *F appartient à l'un des trois domaines d'attraction de la loi des valeurs extrêmes (Fréchet, Gumbel ou Weibull), si et seulement si il existe une fonction σ_u strictement positive et un réel ξ tels que*

$$\lim_{u \to x_F} sup_{0<y<x_F-u} |F_u(y) - G_{\xi,0,\sigma_u}(y)| = 0, \tag{1.12}$$

où $G_{\xi,0,\sigma_u}$ est la fonction de répartition de la loi de Pareto généralisée et F_u est la fonction de répartition conditionnelle des excès au-delà du seuil u.

1.4 Estimation des paramètres

Dans la littérature, différentes méthodes ont été proposées pour estimer les paramètres ξ et σ. Nous allons présenter dans ce qui suit quelques méthodes d'estimation. On notera désormais $(X_{1,n}, ..., X_{n,n})$ les statistiques d'ordre correspondant à l'échantillon $(X_1, ..., X_n)$ où $X_{1,n} \leq .. \leq X_{k,n} \leq ... \leq X_{n,n} = M_n$

Estimateur de Pickands(1975)

L'estimateur de Pickands a été introduit par Pickands [43], pour estimer tout $\xi \in \mathbb{R}$. Il est défini par :

$$\hat{\xi}^P = \hat{\xi}^P(k) := \frac{1}{\ln 2} \ln \frac{X_{n-k+1,n} - X_{n-2k+1:n}}{X_{n-2k+1,n} - X_{n-4k+1:n}}$$

où $X_1, ..., X_n$, sont n v.a. i.i.d de fonction de répartition F et $k = k_n$ une suite d'entiers avec $1 < k < n$.

Sous certaines conditions sur la suite entière (k_n) et la fonction de répartition F, l'estimateur de Pickands possède de bonnes propriétés asymptotiques :

(i) Convergence faible : si $k \to \infty$ et $\frac{k}{n} \to 0$, alors

$$\hat{\xi}^P \to^{\mathbb{P}} \xi, \qquad quand \qquad n \to \infty$$

$\to^{\mathbb{P}}$ désigne la convergence en probabilité.

(ii) Convergence forte : si en plus $k/\log\log n \to \infty$ quand $n \to \infty$, alors

$$\hat{\xi}^P \to^{p.s.} \xi, \qquad quand \qquad n \to \infty$$

$\to^{p.s.}$ désigne la convergence presque sûre.

(iii) Normalité asymptotique : sous d'autres conditions

$$\sqrt{k}(\hat{\xi}^P - \xi) \to^d \mathbf{N}(0, \sigma^2(\xi)) \qquad quand \qquad n \to \infty.$$

où :

$$\sigma^2(\xi) = \frac{\xi^2(2^{2\xi+1}+1)}{(2(2^\xi - 1)ln2)^2}.$$

Pour plus de détail voir Dekkers et de Haan [18].

Estimateur de Hill(1975)

Plusieurs estimateurs sont proposés pour ξ via l'approche des valeurs extrêmes. Le plus célèbre, mais non nécessairement le meilleur, de ces estimateurs est celui défini par Hill [29], comme suit

$$\hat{\xi}^H := \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n}$$

avec $X_{1,n}, ..., X_{n,n}$ les statistiques d'ordre associées à l'échantillon $(X_1, ..., X_n)$. Les propriétés asymptotiques de l'estimateur de Hill ont été établies par Deheuvels et al. [15] pour la consistance et Beirlant et Teugels [7], entre autres, pour la normalité asymptotique. Mais le principal inconvénient de cet estimateur est qu'il n'est valable que pour les distributions appartenant au domaine d'attraction de Fréchet (où $\xi > 0$). Ses propriétés de convergence sont les suivantes :

(i) Convergence faible : si $k \to \infty$ et $\frac{k}{n} \to 0$, alors

$$\hat{\xi}^H \to^{\mathbb{P}} \xi, \qquad quand \qquad n \to \infty.$$

(ii) Convergence forte : si en plus $k/\log\log n \to \infty$ quand $n \to \infty$, alors

$$\hat{\xi}^H \to^{p.s.} \xi, \qquad quand \qquad n \to \infty.$$

(iii) Normalité asymptotique : sous d'autres conditions sur k, Häusler et Teugels [28]

$$\sqrt{k}(\hat{\xi}^H - \xi) \to^d \mathbf{N}(0, \xi^2) \qquad quand \qquad n \to \infty.$$

Estimateur des moments (Estimateur de Dekkers, Einmahl et de Haan(1989))

La généralisation de l'estimateur de Hill pour estimer un indice de queue ξ quelconque a été introduite par Dekkers et al. [18]. Leur estimateur, appelé Estimateur des Moments est défini comme suit :

$$\hat{\xi}^M := \hat{\xi}^H + 1 - \frac{1}{2}(1 - \frac{(\hat{\xi}^H)^2}{S})^{-1}$$

où

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

où $\hat{\xi}^H$ est l'estimateur de Hill de ξ, qui n'est convergent que si $\xi > 0$ et le terme S est la correction à apporter pour que l'estimateur soit convergent quel que soit ξ.
Ses propriétés de convergence sont :

(i) Convergence faible : si $x_F > 0$, $k \to \infty$ et $\frac{k}{n} \to 0$, alors

$$\hat{\xi}^M \to^{\mathbb{P}} \xi, \qquad quand \qquad n \to \infty.$$

(ii) Convergence forte : si $x_F > 0$, $\frac{k}{n} \to 0$ et en plus $k/(\log(n))^\eta \to \infty$ pour un certain $\eta > 0$ quand $n \to \infty$, alors

$$\hat{\xi}^M \to^{p.s.} \xi, \qquad quand \qquad n \to \infty.$$

(iii) Normalité asymptotique : sous d'autres conditions (voir Dekkers, Einmahl et de Haan [18])

$$\sqrt{k}(\hat{\xi}^M - \xi) \to^d \mathbf{N}(0, \sigma^2(\xi)) \qquad quand \qquad n \to \infty$$

avec

$$\sigma^2(\xi) = \begin{cases} \xi^2 + 1 & si \quad \xi \geq 0 \\ \frac{(1-\xi)^2(1-2\xi)(1-\xi+6\xi^2)}{(1-3\xi)(1-4\xi)} & si \quad \xi < 0 \end{cases}$$

Estimateur du maximum de vraisemblance

Si on utilise l'approche POT, les estimateurs des paramètres de la distribution de Pareto sont obtenus par la maximisation de la fonction de log-vraisemblance.
On suppose que les valeurs $X_1 ....., X_k$ sont issues d'une distribution de Pareto généralisée. La fonction log-vraisemblance est

$$L(\sigma, \xi) = \begin{cases} -k \log \sigma - (1 + \frac{1}{\xi}) \sum_{i=1}^k \log(1 + \frac{\xi X_i}{\sigma}), pour \quad \xi \neq 0 \\ -k \log \sigma - \sigma^{-1} \sum_{i=1}^k X_i, pour \qquad \xi = 0 \end{cases} \tag{1.13}$$

à condition que $(1 + \sigma^{-1}\xi X_i) > 0$ pour $i = 1, ..., k$. C'est en dérivant cette expression par rapport aux deux paramètres d'intérêt ξ et σ que nous obtenons un système de deux équations à deux inconnues ξ et σ (pour le cas où $\xi \neq 0$ et $\xi = 0$). En résolvant ces équations nous obtenons les estimateurs du maximum de vraisemblance pour lesquels Smith [52] a établi la normalité asymptotique sous la condition $\xi > -1/2$.

Chapitre 2

Théorie des valeurs extrêmes pour une normalisation non linéaire

La théorie de valeurs extrêmes sous la normalisation non linéaire est un domaine récent initié par Pancheva [40], au début des années quatre vingt du 20^{eme} siècle. Barakat et al. [3] et Barakat [2] ont discuté la propriété de stabilité des extrêmes sous la normalisation non linéaire. Barakat et Omar [5] ont obtenu des propriétés pour les statistiques d'ordre intermédiaires et centrales dans le cas non linéaire. Ravi [44] et Ravi et Praveena [46] ont étudié le comportement de queue des lois max-stables. Plusieurs études sur la normalisation non linéaire ont été menées par Nasri-Roudsari [36] et Peng et al. [41]. Ils ont déterminé les distributions limites de la s-ème statistique d'ordre $M^s_{N_n}$ où N_n est un indice aléatoire à valeur entière qui suit une loi binomiale négative. Grigelionis [26] a décrit les lois limites des valeurs de records sous la normalisation non linéaire et donné les critères d'appartenance de F au domaine d'attraction correspondant. Ravi et Mavitha [45]

ont défini les lois limites pour les extrêmes sous une nouvelle normalisation non linéaire appelée normalisation exponentielle. Le but de ce chapitre est de caractériser les lois limites pour la normalisation non linéaire et de déterminer les différentes propriétés de base. Nous présentons certaines méthodes pour déterminer les constantes de normalisation dans la section 2.2. Nous énonçons les classes de distributions extrêmes, appelées distributions des valeurs extrêmes généralisées et notées GEVP.

2.1 Les lois limites possibles du maximum

Pancheva [39] a montré qu'on peut déterminer la convergence de type Khinchin, pour la normalisation puissance.

Théoréme 2.1.1. *(Théorème de Khinchin). Soient (F_n) une suite de fonctions de répartition, (a_n) et (b_n), deux suites de constantes strictement positives et F une fonction de répartition non dégénérée telles que*

$$F_n(b_n |x|^{a_n} Sign(x)) \to F(x), \quad n \to +\infty.$$

Alors, pour deux suites de constantes strictement positives (α_n) et (β_n) et une fonction de répartition F^* non dégénérée,
$F_n(\beta_n |x|^{\alpha_n} Sign(x)) \to F^*(x)$, si et seulement si $\frac{\alpha_n}{a_n} \to A > 0$ et
$(\frac{\beta_n}{b_n})^{\frac{1}{\alpha_n}} \to B > 0$, lorsque $n \to \infty$ et on a $F^*(x) = F(B |x|^A S(x))$.

Définition 2.1.1. *Une variable aléatoire X (ou sa fonction de distribution) est dite max-stable s'il existe des constantes $a_n, b_n > 0$ telles que*

$$max(X_1, ..., X_n) =^d a_n |X|^{b_n} Sign(X)$$

où $X_1, X_2, ..., X_n$ sont des variables aléatoires i.i.d.

Proposition 2.1.1. *(Propriété limite pour les lois max-stables) La classe des lois max-stables coïncide avec la classe des lois limites pour les maxima (adéquatement normalisés) de variables aléatoires i.i.d.*

Mohan et Ravi [35] et Pancheva [40] ont caractérisé la distribution du maximum $M_n := X_{n,n}$ sous la normalisation non linéaire qui donne six types de lois limites.

Théoréme 2.1.2. *Soient $X_1, ..., X_n$ des variables aléatoires (i.i.d) de fonction de répartition commune F et $X_{1:n} \leq ..... \leq X_{n:n}$ les statistiques d'ordre correspondantes. S'il existe des constantes $a_n, b_n > 0$ et une fonction de répartition non dégénérée H telles que :*

$$\lim_{n\to\infty} P\left[\left|\frac{M_n}{a_n}\right|^{\frac{1}{b_n}} S\left(M_n\right) \leq x\right] = \lim_{n\to\infty} F^n\left(a_n |x|^{b_n} S\left(x\right)\right) = H(x), \quad (2.1)$$

pour tout x, point de continuité de H, alors les lois limites sont données par $H_{i,\beta}(x) = exp[-\mu_{i,\beta}(x)], i = 1, ..., 6, \beta > 0$, où

$$\mu_{1:\beta}(x) = \begin{cases} +\infty, & x \leq 1; \\ (\log x)^{-\beta}, & x > 1; \end{cases} \qquad \mu_{2:\beta}(x) = \begin{cases} +\infty, & x \leq 0. \\ (-\log x)^{\beta}, & 0 < x < 1. \\ 0, & x > 1. \end{cases}$$

$$\mu_{3:\beta}(x) = \begin{cases} +\infty, & x \leq -1. \\ (-\log(-x))^{-\beta}, & -1 \leq x < 0. \\ 0, & x > 0. \end{cases} \quad \mu_{4:\beta}(x) = \begin{cases} (\log(-x))^{\beta}, & x \leq -1. \\ 0, & x > -1. \end{cases}$$

$$\mu_5(x) = \begin{cases} +\infty, & x \leq 0. \\ \frac{1}{x}, & x > 0. \end{cases} \qquad \mu_6(x) = \begin{cases} |x|, & x \leq 0. \\ 0, & x > 0. \end{cases} \quad (2.2)$$

les notations $\mu_{i:\beta}(x) = \mu_i(x), i = 5, 6.$

Nasri-Roudsari et Cramer [37] ont résumé ces lois limites définies en (2.2) par la représentation unifiée, dite représentation de Von Mises et formulée ainsi :

$$H_{1;\gamma}(x,a,b)=\exp\left[-\left(1+\gamma\log\left(ax^{b}\right)\right)^{\frac{-1}{\gamma}}\right],x>0,1+\gamma\log\left(ax^{b}\right)>0. \tag{2.3}$$

et

$$H_{2;\gamma}(x,a,b)=\exp\left[-\left(1-\gamma\log\left(a\left(-x\right)^{b}\right)\right)^{\frac{-1}{\gamma}}\right],x<0,1-\gamma\log\left(a\left(-x\right)^{b}\right)>0. \tag{2.4}$$

Maintenant, nous allons caractériser les domaines d'attraction des lois limites $H_{i,\beta}$ précisées dans (2.2).

2.2 Domaine d'attraction dans le cas de la normalisation puissance

Définition 2.2.1. *Soit F une fonction de répartition non dégénérée, nous dirons que F appartient au domaine d'attraction de H et nous notons $F\in D_p(H)$, s'il existe deux suites de constantes $a_n>0$ et $b_n>0$ telles que*

$$F^{n}\left(a_n\left|x\right|^{b_n}S\left(x\right)\right)=P\left[\left|\frac{M_n}{a_n}\right|^{\frac{1}{b_n}}S\left(M_n\right)\leq x\right]\rightarrow H(x),n\rightarrow+\infty \tag{2.5}$$

où $H\in\{\phi_1,\psi_1,H_{i,\beta},i=1,2,3,4\}$.

Pancheva [[38], [40]] et Mohan et Ravi [35] ont caractérisé les domaines d'attraction de ces lois limites dans le cas de la normalisation puissance.
Les conditions d'appartenance aux domaines d'attraction des lois limites ϕ_1, $H_{1,\beta}$ et $H_{2,\beta}$ tiennent en compte le comportement de queue de la fonction de répartition F.
En notant $x_F=sup\left\{x\in\mathbb{R},F(x)<1\right\}$, nous avons :

Théoréme 2.2.1.

i) $F \in D_p(\phi_1)$ *ssi* $x_F = +\infty$ *et* $1 - F(x) = x^{-1}L_1(x)$.

ii) $F \in D_p(H_{1,\beta})$ *ssi* $x_F = +\infty$ *et* $1 - F(x) = (\log x)^{-\beta}L_2(x)$.

iii) $F \in D_p(H_{2,\beta})$ *ssi* $x_F < 1$ *et* $1 - F(x_F \exp(-\frac{1}{x})) = (\log x_F - \frac{1}{x})^{-\beta}L_3(x)$.

où L_1, L_2 *et* L_3 *sont des fonctions à variation lente au voisinage de* $+\infty$.

Des conditions nécessaires et suffisantes d'appartenance de F aux domaines d'attraction des lois limites de Fréchet et Weibull, (notées respectivement $D_p(\phi_1)$, $D_p(\psi_1)$) sont données par Mohan et Ravi [35].

Théoréme 2.2.2.

i) $F \in D_p(\phi_1)$ *si et seulement si* $x_F > 0$ *et il existe une fonction positive* g *telle que*

$$\lim_{t \to x_F} \frac{1 - F(t \exp(xg(t)))}{1 - F(t)} = \exp(-x), x > 0.$$

Si cette condition est vérifiée, alors

$$\int_a^{x_F} \frac{\bar{F}(s)}{s} ds < \infty, 0 < a < x_F,$$

où $\bar{F} = 1 - F$ *et nous pouvons choisir*

$$g(t) = \frac{1}{\bar{F}(t)} \int_t^{x_F} \frac{\bar{F}(s)}{s} ds.$$

Dans ce cas on prend : $a_n = F^{\leftarrow}(1 - \frac{1}{n}) = inf\{x : F(x) \geq 1 - \frac{1}{n}\}$ *et* $b_n = g(a_n)$.

ii) $F \in D_p(\psi_1)$ *si et seulement si* $x_F \leq 0$ *et il existe une fonction positive* g *telle que* $\lim_{t \to x_F} \frac{1 - F(t \exp(xg(t)))}{1 - F(t)} = \exp(x), x < 0$.

Si cette condition est vérifiée, alors

$$-\int_a^{x_F} \frac{\bar{F}(s)}{s} ds < \infty, a < x_F,$$

et nous pouvons choisir

$$g(t) = -(\frac{1}{\bar{F}(t)}) \int_t^{x_F} \frac{\bar{F}(s)}{s} ds.$$

Dans ce cas, on peut prendre

$a_n = -F^{\leftarrow}(1 - \frac{1}{n}) = -inf\{x : F(x) \geq 1 - \frac{1}{n}\}$ *et* $b_n = g(-a_n)$.

Ci-dessous une comparaison est établie par Mohan et Ravi [35], entre des domaines d'attraction sous les normalisations linéaire et non linéaire. D_l est le domaine d'attraction sous normalisation linéaire.

Théoréme 2.2.3.

$$1) \left.\begin{array}{ll} i) & Si \quad F \in D_l(\phi_\beta) \\ ii) & Si \quad F \in D_l(\Lambda), x_F = \infty \end{array}\right\} \Rightarrow F \in D_p(\phi_1)$$

$$2) \quad F \in D_l(\Lambda), 0 < x_F < \infty \iff F \in D_p(\phi_1), 0 < x_F < \infty$$

$$3) \quad F \in D_l(\Lambda), x_F < 0 \iff F \in D_p(\psi_1), x_F < 0$$

$$4) \left.\begin{array}{ll} i) & Si \quad F \in D_l(\Lambda), x_F = 0 \\ ii) & Si \quad F \in D_l(\psi_\beta), x_F = 0 \end{array}\right\} \Rightarrow F \in D_p(\psi_1) \tag{2.6}$$

$$5) \quad F \in D_l(\psi_\beta), 0 < x_F \iff F \in D_p(H_{2,\beta})$$

$$6) \quad F \in D_l(\psi_\beta), x_F < 0 \iff F \in D_p(H_{4,\beta}).$$

Il nous reste à déterminer d'autres choix possibles pour les constantes de normalisation $a_n > 0$ et $b_n > 0$. En définissant la fonction inverse généralisée de F par

$$F^{\leftarrow}(y) = inf\left\{x \in \mathbb{R}, F(x) > y\right\}, y \in]0, 1[,$$

nous obtenons le résultat suivant :

Théoréme 2.2.4.

i) Lorsque $F \in D_p(\phi_1)$, *un choix possible des constantes de normalisation est* $b_n = 1$ *et* $a_n = F^{\leftarrow}(1 - n^{-1})$.

ii) Lorsque $F \in D_p(H_{1,\beta})$, *un choix possible des constantes de normalisation est* $a_n = \log F^{\leftarrow}(1 - n^{-1})$ *et* $b_n = 1$.

iii) Lorsque $F \in D_p(H_{2,\beta})$, *un choix possible des constantes de normalisation est* $a_n = \log x_F - \log F^{-1}(1 - n^{-1})$ *et* $b_n = x_F$.

Ce théorème est basé sur le résultat suivant

Lemme 2.2.1. *Soient* $X_1, ..., X_n$ n *variables aléatoires i.i.d. positives de fonction de répartition commune* F. *Notons* $Y_i = \log X_i$, $1 \leq i \leq n$ *de fonction de répartition commune* F_1. *Nous avons*

$$F \in D_p \{\phi_1, H_{1,\beta}, H_{2,\beta}\} \iff F_1 \in D_l \{\Lambda, \phi_\beta, \psi_\beta\}.$$

Exemples

Exemple 2.2.1. *Soit* X *une variable aléatoire qui suit la loi de Pareto de paramètre 1 telle que*

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

en prenant $b_n = 1$ *et* $a_n = F^{\leftarrow}(1 - n^{-1}) = n$, *nous obtenons*
$F^n(a_n x^{b_n}) = (1 - (a_n x^{b_n})^{-1})^n = (1 - (nx)^{-1})^n \to \phi_1(x)$, *quand* $n \to \infty$.
Ainsi $F \in D_p(\phi_1)$.

Exemple 2.2.2. *Soit F la fonction de répartition définie par*

$$1 - F(x) = (\log x)^{-\beta} \mathbb{I}_{[e,\infty[}(x).$$

En prenant $a_n = \log F^{\leftarrow}(1 - n^{-1}) = n^{\frac{1}{\beta}}$ *et* $b_n = 1$, *nous obtenons pour* $n > e$,

$\lim_{n\to\infty} F^n(a_n x^{b_n}) = \lim_{n\to\infty}(1 - (\log a_n + b_n \log x)^{-\beta})^n$

$$= \exp(-(\log x)^{-\beta}) = H_{1,\beta}(x).$$

Ainsi $F \in D_p(H_{1,\beta})$

Nous finissons cette section en évoquant la vitesse de convergence pour la normalisation non linéaire. A partir de quelles valeurs de n pouvons nous "approcher" $F^n(a_n |x|^{b_n} S(x))$ par sa limite $H(x)$? Il n'y a pas de réponse générale. Tout va dépendre de la loi F. Une étude bien détaillée sur la vitesse de convergence sous la normalisation non linéaire est présentée par Belbachir et al. [32] où ils utilisent la distance de variation totale ; ce travail fera l'objet du Chapitre 4 de cette thèse. Barakat et al. [6] ont comparé les vitesses de convergence en utilisant la métrique uniforme et la métrique de variation totale, des exemples illustratifs sont donnés pour comparer les vitesses de convergence.

2.3 Un estimateur pour les paramètres des lois extrêmes

Dans la littérature, deux classes de distributions des valeurs extrêmes sont utilisées pour la modélisation des événements extrêmes. La première classe est la distribution des valeurs extrêmes généralisée (GEVL), la seconde est la distribution de Pareto généralisée (GPDL) sous la normalisation linéaire

introduite par Pickands [43].

Pour la normalisation puissance les deux familles (2.3) et (2.4) sont appelées distributions des valeurs extrêmes généralisées (GEVP). Elles satisfont la propriété de max-stabilité définie dans (2.1.1). La distribution des valeurs extrêmes généralisée (GPDP) a été développée par Barakat et al. [4].

Dans cette section nous rappelons des résultats qui résument de nouvelles méthodes de calcul de la (GPDP) dans le cas de la normalisation puissance.

Théoréme 2.3.1. *(GPDP). Soit* $F^{[u]}(x) := P\left(X \leq x | x > u\right)$.
Si l'hypothèse (2.1) est satisfaite avec $H\left(x\right) = H_{1;\gamma}\left(x, a, b\right)$. *Alors il existe* $\alpha(u) > 0$ *tel que*

$$F^{[u]}\left(ux^{\alpha(u)}\right) \to Q_{1,\gamma}\left(x, b\right) \tag{2.7}$$

lorsque $u \uparrow x_F > 0$, *où* $Q_{1,\gamma}\left(x, b\right) = 1 + \log H_{1:\gamma}\left(x, 1, b\right)$.
D'autre part, si (2.1) est satisfaite avec $H\left(x\right) = H_{2;\gamma}\left(x, a, b\right)$ *alors il existe* $\alpha(u) > 0$ *tel que*

$$F^{[u]}\left(ux^{\alpha(u)}\right) \to Q_{2,\gamma}\left(x, b\right) \tag{2.8}$$

lorsque $u \uparrow x_F \leq 0$, *où* $Q_{2,\gamma}\left(x, b\right) = 1 + \log H_{2:\gamma}\left(x, 1, b\right)$.

Le théorème suivant montre que la distribution de Pareto généralisée (2.7) et (2.8) sous la normalisation non linéaire, satisfait la propriété de stabilité pour la méthode de dépassement d'un seuil, appelée méthode POT ; propriété dont nous avons besoin pour estimer les paramètres des lois limites.

Théoréme 2.3.2 (Propriété de stabilité (POT)). .
Les données tronquées à gauche pour la GPDP donnent à nouveau une GPDP. Cela signifie que pour tous $0 < k < x$, *on a*

$$Q_{1,\gamma}^{[k]}(x; \sigma) = Q_{1,\gamma}\left(\frac{x}{k}, \bar{\sigma}\right),$$

où $\bar{\sigma} = \frac{\sigma}{c}$ *avec* $c = 1 + \gamma \log k$.

Dans le cas où $-1 < k < x < 0$, *on a*

$$Q_{2,\gamma}^{[k]}(x;\sigma) = Q_{2,\gamma}\left(\frac{x}{k}, \bar{\sigma}\right).$$

où $\bar{\sigma} = \frac{\sigma}{\acute{c}}$ *avec* $\acute{c} = 1 - \gamma \log(-k)$.

Pour plus de détails voir Barakat et al. [4]. Dans le théorème suivant, nous utilisons la méthode de Pickands [43] pour estimer les paramètres d'échelle et de forme.

Théoréme 2.3.3. *Soient* n *la taille de l'échantillon* $(X_1, ..., X_n)$ *et* $m = m(n)$, *un entier tel que* $1 \leq m(n) < n$. *Soient* $\{Y_i, i = 1, 2, ..., n\}$ *les statistiques d'ordre décroissantes,* $Y_i = X_{n-i+1:n}$ *étant la* i*ème plus grande observation. Nous considérons les variables aléatoires* $\frac{Y_i}{Y_{4m}}, i = 1, 2, 3, ..., 4m-1$, *les statistiques d'ordre décroissantes d'un échantillon de taille* $4m-1$ *issu d'une population de fonction de répartition* $Q_{1,\gamma}(x, \bar{\sigma})$ *où* $0 < \bar{\sigma} < \infty$ *et* $-\infty < \gamma < +\infty$.

Alors les estimateurs de γ *et* $\bar{\sigma}$ *sont donnés par*

$$\hat{\gamma} = (\log 2)^{-1} \log \frac{\log|Y_m| - \log|Y_{2m}|}{\log|Y_{2m}| - \log|Y_{4m}|}$$

et

$$\hat{\bar{\sigma}} = \frac{1 - 2^{\hat{\gamma}}}{\hat{\gamma}(\log|Y_{2m}| - \log|Y_{4m}|)},$$

l'entier $m = m(n)$ *doit satisfaire les deux conditions* $\lim_{n\to\infty} m = \infty$ *et* $\lim_{n\to\infty} \frac{m}{n} = 0$

Pour la démonstration voir Barakat et al. [4].

Chapitre 3

Notions de convergence des sommes de variables aléatoires i.i.d

Dans ce chapitre, nous examinerons des questions de convergence des variables aléatoires i.i.d à valeurs dans l'ensemble des entiers $\mathbb{N}$. Pour comparer leurs lois de probabilités, nous allons utiliser la distance à variation totale qui mesure l'écart entre deux lois ou mesures de probabilité, définie comme suit.

Définition 3.0.1. *Soient μ et ν deux mesures de probabilité sur $\mathbb{R}$. On appelle distance à variation totale entre μ et ν la quantité*

$$d(\mu,\nu) = Sup\left\{\left|\mu(A)-\nu(A)\right|, A \subset \mathbb{R}\right\}. \tag{3.1}$$

Théoréme 3.0.1. *Soient μ et ν deux mesures de probabilité sur $\mathbb{N}$.*
Alors il y'a équivalence entre
(ii) $d(\mu,\nu) = max\left\{\left|\mu(A)-\nu(A)\right|, A \subset \mathbb{N}\right\} = \frac{1}{2}\sum_{x\in\mathbb{N}}\left|\mu(x)-\nu(x)\right|,$

et

(ii) $d(\mu,\nu) = Sup_{|h\leq 1|}\left|E(h(\mu)) - E(h(\nu))\right|,$
où h *est une fonction mesurable telle que* $h \leq 1$.

Pour donner un sens précis aux approximations de lois de probabilité, une approche est d'introduire des distances entre ces lois. Dans le cas discret, on utilise souvent la distance en variation totale, notée :

$$d_v(X,X') := \frac{1}{2}\sum_k \left|P\left[X=k\right] - P\left[X'=k\right]\right|$$

3.1 Propriétés de la distance de variation totale

Serfling [51] a établi les propriétés suivantes de la distance à variation totale.

Propriété 3.1.1. *Soit* (X,Y) *un couple de variables aléatoires. Alors on a,* $d(X,Y) \leq P(X \neq Y)$.
En effet, pour tout $\forall A \subset \mathbb{N}$ *:*
il vient que $P(X \in A) - P(Y \in A)$

$$\begin{aligned} &\leq P(X \in A) - P(Y \in A \quad et \quad X = Y) \\ &= P(\{X \in A\} \setminus \{Y \in A \quad et \quad X = Y\}). \\ &\leq P(X \neq Y). \end{aligned}$$

d'où $d(X,Y) \leq P(X \neq Y)$.

Un exemple, qui illustre cette propriété et qui est dû à Vervaat [49], est le suivant

Lemme 3.1.1. *Soit $p \in [0,1]$. Il existe alors un couple de v.a. (X,Y) tel que X suit une loi de Bernoulli de paramètre p et Y suit une loi de Poisson de paramètre p et $P(X \neq Y) \leq p^2$.*

Propriété 3.1.2. *Soient $X_1, X_2, ..., X_n$ et $Y_1, Y_2, ..., Y_n$ des variables aléatoires à valeurs dans $\mathbb{N}$, mutuellement indépendantes.*
Alors

$$d(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} Y_i) \leq \sum_{i=1}^{n} d(X_i, Y_i).$$

Un résultat important de Freedman [23], qui nous servira par la suite, est le suivant

Lemme 3.1.2. *Soient Y et Y' deux variables aléatoires de Poisson de moyennes λ_1 et λ_2 respectivement, $\lambda_1 \neq \lambda_2$. Alors*

$$d(Y, Y') \leq |\lambda_2 - \lambda_1|.$$

Démonstration 3.1.1. *Supposons que $\lambda_1 < \lambda_2$*
Posons qu'il existe une variable aléatoire Z indépendante de Y et de loi de Poisson de paramètre $\lambda_2 - \lambda_1$. La variable aléatoire Y suit la même loi que Y'. Nous obtenons
$d(Y,Y') = d(Y, Y+Z) : d(Y, Y+Z) \leq P(Y \neq Y+Z) = P(Z \neq 0) = 1 - e^{-(\lambda_2 - \lambda_1)}$ *par utilisation de la propriété 3.1.1.*
Maintenant $1 - e^{-(\lambda_2 - \lambda_1)} \leq \lambda_2 - \lambda_1$ puisque pour tout $x \in \mathbb{R}$, $1 - e^{-x} \leq x$. D'où le résultat.

3.2 Convergence de sommes de variables aléatoires i.i.d de Bernouilli vers la loi de Poisson

Dans cette section, nous nous intéressons à l'approximation de la loi de la somme de n v.a. indépendantes de Bernoulli par la loi de Poisson.
Soit $S_n = \sum_{i=1}^{n} X_i$. Nous noterons désormais $Y(\lambda)$ la variable aléatoire de Poisson de moyenne $\lambda > 0$.
Vervaat [49] a étudié cette approximation lorsque les v.a. $X_1, ..., X_n$ de Bernoulli sont i.i.d.

Théoréme 3.2.1. *Soient $X_1, ..., X_n$ n variables aléatoires de Bernoulli i.i.d. où $p = P(X_i = 1)$ pour tout $1 \leq i \leq n$.*
Alors

$$d(S_n, Y(np)) \leq \frac{1}{2}\frac{p}{\sqrt{1-p}}.$$

Dans le cas où les v.a. de Bernoulli sont indépendantes de paramètre $p_i = P(X_i = 1) = 1 - P(X_i = 0)$, $0 < p_i < 1, i = 1, ..., n$, Serfling [51] a établi ce résultat.

Théoréme 3.2.2. *Pour $X_1, ..., X_n$ n v.a indépendantes où $X_i \sim Ber(p_i), 1 \leq i \leq n$,*

$$d(S_n, Y(\sum_{i=1}^{n} p_i)) \leq \sum_{i=1}^{n} p_i^2$$

et

$$d(S_n, Y(\sum_{i=1}^{n} \lambda_i)) \leq \frac{1}{2}\sum_{i=1}^{n} \lambda_i^2$$

où $\lambda_i = -\log(1 - p_i), 1 \leq i \leq n$.

Pour plus de détail voir aussi Freedman [23].

3.3 Distance en variation totale pour le cas multidimensionnel.

Deheuvels et Pfeifer [17] ont obtenu des convergences poissonniennes dans le cas multidimensionnel. Pour $k \geq 1$ un entier fixé et $n = 1, 2, ...$, considérons $U_n = (U_n(1), ..., U_n(k))$ une suite de vecteurs aléatoires indépendants de $\mathbb{R}^k$ telle que, $j = 1, ..., k$

$$P(U_i(1) = 0, ..., U_i(j-1) = 0, U_i(j) = 1, U_i(j+1) = 0, ..., U_i(k) = 0) = p_{ij} \geq 0.$$

On considère $S_n = (S_n(1), ..., S_n(k))$ où $S_n(j) = \sum_{i=1}^{n} U_i(j)$.
Soit pour $j = 1, ..., k$, $\lambda_j = \sum_{i=1}^{n} p_{ij}$ et $T_n = (T_n(1), ..., T_n(k))$ où pour $1 \leq j \leq k$, les $T_n(j)$ sont des variables aléatoires indépendantes de loi de Poisson de moyenne $\lambda_j = \sum_{i=1}^{n} p_{ij}$. La distance en variation totale entre les lois de S_n et T_n est définie comme suit :

$$\begin{aligned} \Delta_n = d(S_n, T_n) &= Sup_{E \subset \mathbb{N}} |P(S_n \in E) - P(T_n \in E)| \\ &= \frac{1}{2} \sum_{k=0}^{+\infty} |P(S_n = k) - P(T_n = k)| . \end{aligned}$$

Moyennant ces notations, on obtient le résultat suivant dans le cas multidimensionnel.

Théoréme 3.3.1. *Supposons que $p_1 = p_{i1}....., p_k = p_{ik}$ soient indépendants de $i = 1, 2, ..., n$. Soit $P = \sum_{j=1}^{k} p_j$ et $\theta = nP$. Alors*

$$\Delta_n = \frac{1}{2} P\theta \left\{ \frac{\theta^{\alpha-1}(\alpha - \theta)}{\alpha!} - \frac{\theta^{\beta-1}(\beta - \theta)}{\beta!} \right\} e^{-\theta} + R_n$$

$$= \frac{1}{4} d_n + R_n = \frac{P}{4} D(\theta) + R_n, \tag{3.2}$$

avec $\alpha = \left[\theta + \frac{1}{2} + (\theta + \frac{1}{4})^{\frac{1}{2}}\right]$ *et* $\beta = \left[\theta + \frac{1}{2} - (\theta + \frac{1}{4})^{\frac{1}{2}}\right]$

où $[u]$ *est la partie entière de* u *et*

$$|R_n| \leq \min\left\{5nP^3, 16P^2\right\}.$$

En plus, si $n \geq 1$ *et* $P \to 0$, *lorsque* $n \to \infty$ *nous avons toujours*

$$\Delta_n = \frac{1}{4}d_n(1 + O(P)) \sim \frac{1}{4}d_n. \tag{3.3}$$

-Si $\theta = nP \to \infty$, $P \to 0$, *alors*

$$\Delta_n = \frac{P}{\sqrt{2\pi e}}\left\{1 + O(\max\{P, \frac{1}{\sqrt{nP}}\})\right\}. \tag{3.4}$$

En d'autre termes $\theta = nP \to 0$, $n \to +\infty$, *alors*

$$\Delta_n = nP^2\left(1 + O(P)\right). \tag{3.5}$$

Démonstration voir Deheuvels et Pfeifer [16].

Le théorème suivant traite le cas général.

Théoréme 3.3.2. *Supposons que* p_{ij} *où* $j = 1, ..., k$ *et* $i = 1, ..., n$ *soient arbitraires, nous avons alors*

$$\Delta_n = \frac{1}{4}d_n + R_n, \tag{3.6}$$

où

$$\begin{aligned} d_n &= E\left(\left|\sum_{i=1}^{n}\left\{\sum_{j=1}^{k}p_{ij}(1 - \frac{\tau_j}{\lambda_j})\right\}^2 - \sum_{i=1}^{n}\sum_{j=1}^{k}\frac{p_{ij}^2}{\lambda_j^2}\tau_j\right|\right). \\ &\leq \min\left(4\sum_{i=1}^{n}\left\{\sum_{j=1}^{k}p_{ij}\right\}^2, 2\sum_{j=1}^{k}\left\{\sum_{i=1}^{n}p_{ij}^2\right\}\left\{\sum_{i=1}^{n}p_{ij}\right\}^{-1}\right), \end{aligned}$$

et

$$
\begin{aligned}
|R_n| &\leq \exp\left(4K\sum_{i=1}^{n}\left\{\sum_{j=1}^{k}p_{ij}\right\}^2\right) \\
&+ \left(\frac{8}{3}(K-1)^2\sum_{i=1}^{n}\left\{\sum_{j=1}^{k}p_{ij}\right\}^3 + 8K(K-1)\left\{\sum_{i=1}^{n}\left\{\sum_{j=1}^{k}p_{ij}\right\}^2\right\}^2\right) \\
&+ 8(K-1)\sum_{i=1}^{n}\left\{\sum_{j=1}^{k}p_{ij}\right\}^3,
\end{aligned}
$$

$R_n = o(d_n)$ *si* $\sum_{i=1}^{n}\left\{\sum_{j=1}^{k}p_{ij}\right\}^2 = O(1)$.

$$K = 1 + \frac{1}{2}\exp\left\{2\max_{1\leq i\leq n}\sum_{j=1}^{k}p_{ij}\right\} \leq 1 + \frac{e^2}{2} < 4,70.$$

$\tau_1, ..., \tau_k$ *sont des variables aléatoires indépendantes de Poisson de moyenne* $E(\tau_j) = \sum_{i=1}^{n} p_{ij}, 1 \leq j \leq k$.

Démonstration voir Deheuvels et Pfeifer [16].

Nous appliquerons les résultats de ce chapitre, pour le calcul de la vitesse de convergence en théorie des valeurs extrêmes sous normalisation non linéaire.

Chapitre 4

Sur la vitesse de convergence sous la normalisation exponentielle

4.1 Introduction

Soient $X_1, X_2, ..., X_n$ n variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de fonction de répartition commune F. Soient $X_{1:n} \leq X_{2:n} \leq ... \leq X_{n:n}$ les statistiques d'ordre correspondantes.

La théorie des valeurs extrêmes sous une normalisation linéaire est basée sur le résultat de Fischer et Tippett [22] et Gnedenko [25]. S'il existe des constantes de normalisation $(a_n > 0)$ et $(b_n \in \mathbb{R})$ pour $n = 1, 2, ...$ et une fonction de répartition non dégénérée G telles que :

$$\begin{aligned} \lim_{n\to\infty} P(X_{n,n} \leq a_n x + b_n) &= \lim_{n\to\infty} F^n(a_n x + b_n) \\ &= G(x), \end{aligned}$$

pour tout x point de continuité de G, alors G est l'un des trois types de lois suivantes :

- Loi de Fréchet :

$$\phi_\alpha(x) = \exp(-(x)^{-\alpha})\mathbb{I}_{(x\geq 0)}, \alpha > 0$$

- Loi de Gumbel :

$$\Lambda(x) = \exp(-e^{-x}), \forall x \in \mathbb{R}$$

- Loi de Weibull :

$$\psi_\alpha(x) = \exp(-(-x)^{\alpha})\mathbb{I}_{(x<0)} + \mathbb{I}_{(x\geq 0)}, \alpha > 0$$

où $\mathbb{I}_A$ est la fonction indicatrice de l'événement A.

Une autre approche, utilisant une normalisation non linéaire, est initiée par Pancheva [40] et Weinstein [54], permet d'obtenir d'autres lois limites asymptotiques pour le maximum. Mohan, Ravi [35] et Pancheva [40] ont caractérisé la distribution du maximum $M_n := X_{n,n}$ sous la normalisation non linéaire qui donne six types de lois limites possibles. S'il existe des constantes $a_n, b_n > 0$ et une fonction de répartition non dégénérée H telles que :

$$\lim_{n\to\infty} P\left[\left|\frac{M_n}{a_n}\right|^{\frac{1}{b_n}} S(M_n) \leq x\right] = \lim_{n\to\infty} F^n\left(a_n |x|^{b_n} S(x)\right) = H(x), \quad (4.1)$$

pour tout x point de continuité de H et où $S(x) = -1, 0, 1$, quand respectivement $x < 0$, $x = 0$ ou $x > 0$, alors $H \in \{\phi_1, \psi_1, H_{i,\alpha}, i = 1, 2, 3, 4\}$ où

$$H_{1:\alpha}(x) = \begin{cases} 0, & x \leq 1; \\ \exp\left\{-\left(\log x\right)^{-\alpha}\right\}, & x > 1, \end{cases}$$

$$H_{2:\alpha}(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left\{-\left(-\log x\right)^{\alpha}\right\}, & 0 < x < 1. \\ 1, & x \geq 1. \end{cases}$$

$$H_{3:\alpha}(x) = \begin{cases} 0, & x \leq -1. \\ \exp\left\{-\left(-\log(-x)\right)^{-\alpha}\right\}, & -1 < x < 0. \\ 0, & x \geq 0. \end{cases}$$

$$H_{4:\alpha}(x) = \begin{cases} \exp\left\{-\left(-\log(-x)\right)^{\alpha}\right\}, & x \leq -1. \\ 1, & x \geq -1. \end{cases}$$

$$\phi(x) = \phi_1(x) = \begin{cases} 0, & x \leq 0. \\ \exp\left\{-x^{-1}\right\}, & x > 0. \end{cases}$$

$$\psi(x) = \psi_1(x) = \begin{cases} \exp\left\{x\right\}, & x \leq 0. \\ 1, & x \geq 0. \end{cases}$$

Ravi et Mavitha [45] ont étudié une nouvelle forme de la normalisation non linéaire

$$\exp\left\{u_n(|\log|x||)^{v_n} sign(\log|x|)\right\} sign(x),$$

qu'ils ont appelée "la normalisation exponentielle".

Théoréme 4.1.1. *Soient $X_1, ..., X_n$ des variables aléatoires (i.i.d) de fonction de répartition commune F et $X_{1:n} \leq ..... \leq X_{n:n}$ les statistiques d'ordre correspondantes. S'il existe des constantes $u_n, v_n > 0$ et une fonction de répartition non dégénérée U telles que :*

$$\lim_{n\to\infty} P\left(\exp\left\{\left(\frac{|\log|M_n||}{u_n}\right)^{\frac{1}{v_n}} S\left(\log|M_n|\right)\right\} S(M_n) \leq x\right)$$
$$= \lim_{n\to\infty} F^n\left(\exp\left\{u_n(|\log|x||)^{v_n} S(\log|x|)\right\} S(x)\right)$$
$$= U(x) \tag{4.2}$$

où $S(x) = -1, 0, 1$, quand respectivement $x < 0$, $x = 0$ et $x > 0$,
alors U est l'un douze types de lois suivantes :

1. *Loglog - Fréchet :*

$$U_{1,\alpha}(x) = \exp(-(\log\log x)^{-\alpha}), x \geq e.$$

2. *Loglog - Weibull :*

$$U_{2,\alpha}(x) = \exp(-(-\log\log x)^{\alpha}), 1 \leq x < e.$$

3. *Standard log - Fréchet :*

$$H_{1,1}(x) = \exp(-(\log x)^{-1}), x \geq 1.$$

4. *Inverse loglog - Fréchet :*

$$U_{3,\alpha}(x) = \exp(-(-\log(-\log x))^{-\alpha}), \frac{1}{e} \leq x < 1.$$

5. *Inverse loglog - Weibull :*

$$U_{4,\alpha}(x) = \exp(-(\log(-\log x))^{-\alpha}), 0 \leq x < \frac{1}{e}.$$

6. *Standard Uniform :*

$$H_{2,1}(x) = x, 0 \geq x < 1.$$

7. *Negative loglog-Fréchet :*

$$U_{5,\alpha}(x) = \exp(-(\log(-\log(-x)))^{-\alpha}), \frac{-1}{e} \leq x < 0.$$

8. *Negative loglog-Weibull :*

$$U_{6,\alpha}(x) = \exp(-(-\log(-\log(-x)))^{\alpha}), -1 \leq x < -\frac{1}{e}.$$

9. *Standard inverse log-Fréchet :*

$$H_{3,1}(x) = exp(-(-\log(-x))^{-1}), -1 \leq x < 0.$$

10. *Negative inverse loglog-Fréchet :*

$$U_{7,\alpha}(x) = \exp(-(-\log\log(-x))^{-\alpha}), -e \leq x < -1.$$

11. *Negative inverse loglog-Weibull :*

$$U_{8,\alpha}(x) = \exp(-(\log\log(-x))^{\alpha}), x \leq -e.$$

12. *Standard inverse log - Weibull :*

$$H_{4,1}(x) = \frac{-1}{x}, x \leq -1.$$

Définition 4.1.1. *Nous dirons que F appartient au domaine d'attraction de U et nous notons $f \in D_e(U)$ si et seulement si (4.2) est satisfaite.*

Maintenant, si nous posons
$N_n(x) = \sum_{s=1}^{n} \mathbb{I}_{\{X_s \geq \exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)\}}$, $N_n(x)$ suit une loi binomiale $B(n, 1 - F(\exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)))$. Les événements $\{X_{n-k+1,n} \leq \exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)\}$ et $\{N_n(x) \leq k-1\}$ sont identiques puisque si le premier est réalisé la $(n-k+1)$-ème statistique d'ordre $X_{n-k+1,n}$ ne doit pas dépasser $\exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)$. Donc au plus $k-1$ variables parmi $X_1, X_2, ..., X_n$ peuvent dépasser $\exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)$ et inversement.
On en déduit

$$P(X_{n-k+1,n} \leq \exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)) = P(N_n(x) \leq k-1). \tag{4.3}$$

Nous avons (4.2) si seulement si,
$n \ln(1 - (1 - F(\exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)))) \to \ln U(x)$ qui équivaut à

$$n[(1 - F(\exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)))] \to -\ln U(x), \tag{4.4}$$

puisque $\lim_{n\to\infty} F(\exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)) = 1.$

Donc

$$\lim_{n\to\infty} P(N_n(x) \leq k-1) = U(x) \sum_{j=1}^{k-1} \frac{(-\ln U(x))^j}{j!}$$

par l'approximation de loi binomiale par la loi de Poisson.
D'où, en utilisant (4.3),

$$\lim_{n\to\infty} P(X_{n-k+1,n} \leq \exp\{u_n(|\log|x||)^{v_n} S(\log|x|)\} S(x)) \tag{4.5}$$

$$= U(x) \sum_{j=1}^{k-1} \frac{(-\ln U(x))^j}{j!}.$$

Barakat et al. [6] ont comparé la vitesse de convergence entre les normalisations linéaire et non linéaire. Peng et al. [42] ont étudié la vitesse de convergence dans le cas non linéaire en utilisant différentes méthodes.
Dans ce travail nous allons évaluer la vitesse de convergence définie dans (4.5) en utilisant la distance de variation totale définie par $d(X,Y) = \frac{1}{2}\sum_{k=0}^{\infty}\left[P\left\{X=k\right\} - P\left\{Y=k\right\}\right]$, X et Y étant deux variables aléatoires discrètes à valeurs dans $\mathbb{N}$.

4.2 Résultat principaux

Dans ce qui suit nous donnons une série de résultats sur la vitesse de convergence de la loi de la statistique d'ordre $X_{n-k+1,n}$-ème statistique d'ordre vers sa loi limite. $Y(\lambda)$ désignera désormais une loi de Poisson de moyenne λ.

Théoréme 4.2.1. *Soient $X_1, X_2, ..., X_n$ n variables aléatoires (i.i.d.) de fonction de répartition commune F telles que*
$F^n\left(\exp\left\{u_n(|\log|x||)^{v_n}S(\log|x|)\right\}S(x)\right) \to U(x)$, *$U$ étant l'un des douze types de lois définies dans (4.2). Alors*

(i) si $0 \leq -\ln U(x) \leq 1$, *on pose*
$\exp\left\{u_n(|\log|x||)^{v_n}S(\log|x|)\right\}S(x) =: Z_n(x)$, *nous avons*

$$\begin{aligned}&\left|P\left(X_{n-k+1,n} \leq Z_n(x)\right) - U(x)\sum_{j=1}^{k-1}\frac{(-\ln U(x))^j}{j!}\right| \\ &\leq \frac{(-\ln U(x))^2}{n} + \left|n\left(1 - F\left(Z_n(x)\right)\right) + \ln U(x)\right|\end{aligned} \tag{4.6}$$

(ii) si $-\ln U(x) \geq 1$, *nous avons :*

$$\begin{aligned}&\left|P\left(X_{n-k+1,n} \leq Z_n(x)\right) - U(x)\sum_{j=1}^{k-1}\frac{(-\ln U(x))^j}{j!}\right| \\ &\leq \frac{(-\ln U(x))}{n} + \left|n\left(1 - F\left(Z_n(x)\right)\right) + \ln U(x)\right|.\end{aligned} \tag{4.7}$$

Démonstration : L'application du résultat du Théorème 3.3.2, en prenant $p_i = 1 - F\left(Z_n(x)\right)$ pou tout $i = 1, ..., n$, nous donne

$$d\left(N_n(x), Y\left(n\left(1 - F\left(Z_n(x)\right)\right)\right)\right) \leq n\left(1 - F\left(Z_n(x)\right)\right)^2.$$

Pour n assez grand, nous avons selon (4.4)

$$n\left(1 - F(Z_n(x))\right)^2 = \frac{1}{n}\left(n(1 - F\left(Z_n(x)\right))\right)^2 \simeq \frac{\left(-\ln U\left(x\right)\right)^2}{n}. \tag{4.8}$$

En utilisant l'inégalité triangulaire, nous avons

$$\begin{aligned} &d\left(N_n\left(x\right), Y\left(-\ln U\left(x\right)\right)\right) \leq d\left(N_n(x), Y\left(n\left(1 - F\left(Z_n(x)\right)\right)\right)\right) + \\ &d\left(Y\left(n\left(1 - F\left(Z\right)\right), Y\left(-\ln U\left(x\right)\right)\right)\right). \end{aligned} \tag{4.9}$$

D'après le lemme 3.1.2

$$d\left(Y\left(n\left(1 - F\left(Z_n(x)\right)\right)\right), Y\left(-\ln U\left(x\right)\right)\right) \leq \left|n\left(1 - F\left(Z_n(x)\right) + \ln U\left(x\right)\right)\right|. \tag{4.10}$$

En combinant (4.8), (4.9) et (4.10), nous obtenons

$$d\left(N_n\left(x\right), Y\left(-\ln U\left(x\right)\right)\right) \leq \frac{\left(-\ln U\left(x\right)\right)^2}{n} + \left|n\left(1 - F\left(Z_n(x)\right) + \ln U\left(x\right)\right)\right|. \tag{4.11}$$

(4.6) est une conséquence de (4.11) puisque

$$\begin{aligned} &\left|P\left(X_{n-k+1,n} \leq Z_n(x)\right) - U\left(x\right)\sum_{j=1}^{k-1}\frac{\left(-\ln U(x)\right)^j}{j!}\right| \\ &= \left|P\left(N_n\left(x\right) \leq k-1\right) - U\left(x\right)\sum_{j=1}^{k-1}\frac{\left(-\ln U(x)\right)^j}{j!}\right| \\ &\leq d\left(N_n\left(x\right), Y\left(-\ln U\left(x\right)\right)\right). \end{aligned} \tag{4.12}$$

De la même manière, le principal résultat de Vervaat [49] permet d'obtenir

$$d\left(N_n\left(x\right), Y\left(n\left(1 - F\left(Z_n(x)\right)\right)\right)\right) \leq \frac{1 - F\left(Z_n(x)\right)}{\sqrt{F\left(Z_n(x)\right)}} = \frac{1}{\sqrt{n}}\frac{n\left(1 - F\left(Z_n(x)\right)\right)}{\sqrt{n\{1 - [1 - F\left(Z_n(x)\right)]\}}}.$$

Pour n assez grand, ce dernier terme s'écrit

$$\frac{1}{\sqrt{n}}\frac{n\left(1-F\left(Z_n(x)\right)\right)}{\sqrt{n\{1-[1-F\left(Z_n(x)\right)]\}}}\simeq\frac{1}{n}\frac{-\ln U\left(x\right)}{\sqrt{1+\frac{\ln U(x)}{n}}}\simeq\frac{-\ln U\left(x\right)}{n}.$$

Nous avons donc

$$d\left(N_n\left(x\right),Y\left(n\left(1-F\left(Z_n(x)\right)\right)\right)\right)\leq\frac{-\ln U\left(x\right)}{n}. \tag{4.13}$$

En utilisant l'inégalité triangulaire et le lemme 3.1.2, nous obtenons

$$d\left(N_n\left(x\right),Y\left(-\ln U\left(x\right)\right)\right)\leq\frac{-\ln U\left(x\right)}{n}+\left|n\left(1-F\left(Z_n(x)\right)\right)+\ln U\left(x\right)\right|. \tag{4.14}$$

La partie (4.7) est déduite de (4.14) en utilisant (4.12).

Dans le théorème suivant, nous désignons par $[x]$ la partie entière de x.

Théoréme 4.2.2. *Sous les hypothèses du Théorème 4.2.1, nous avons*

$$\begin{aligned}&\left|P\left(X_{n-k+1,n}\leq Z_n(x)\right)-U\left(x\right)\sum_{j=1}^{k-1}\frac{\left(-\ln U(x)\right)^j}{j!}\right|\\&\leq\frac{1}{2n}U\left(x\right)\frac{\left(-\ln U(x)\right)^{[-\ln U(x)]+2}}{[-\ln U(x)]!}+\left|n\left(1-F\left(Z_n(x)\right)\right)+\ln U\left(x\right)\right|+O\left(\frac{1}{n}\right)\end{aligned} \tag{4.15}$$

Démonstration : En prenant $p_i=p=1-F\left(Z_n(x)\right)$

pour tout $1\leq i\leq n$ dans ([17], Theorem 1.3]) , et en tenant compte de (4.4) et du fait que $1-F\left(Z_n(x)\right)\to 0$, lorsque $n\to\infty$, nous avons

$$d\left(N_n\left(x\right),Y\left(-n\ lnF\left(Z_n(x)\right)\right)\right)\simeq\frac{1}{2n}U\left(x\right)\frac{\left(-\ln U\left(x\right)\right)^{[-\ln U(x)]+2}}{[-\ln U\left(x\right)]!}. \tag{4.16}$$

Puisque

$$\frac{n}{2}(1-F\left(Z_n(x)\right))^2=\frac{(n(1-F\left(Z_n(x)\right)))^2}{2n}\simeq\frac{(-\ln U(x))^2}{2n}.$$

Nous avons aussi

$$d\left(N_n\left(x\right), Y\left(-\ln U\left(x\right)\right)\right) \leq d(N_n\left(x\right), Y\left(-n\ln F\left(Z_n(x)\right)\right)) \tag{4.17}$$
$$+d\left(Y\left(-n\ln F\left(Z_n(x)\right)\right), Y\left(-\ln U\left(x\right)\right)\right).$$

En appliquant le Lemme 3.1.2 de (4.16) et (4.17), on a

$$d\left(N_n\left(x\right), Y\left(-\ln U\left(x\right)\right)\right) \leq \tfrac{1}{2n}U\left(x\right)\tfrac{(-\ln U(x))^{[-\ln U(x)]+2}}{[-\ln U(x)]!}$$
$$+\left|-n\ln F\left(Z_n(x)\right)+\ln U\left(x\right)\right|.$$

Mais

$-n\ln F\left(Z_n(x)\right)+\ln U\left(x\right) = n\left(1-F\left(Z_n(x)\right)+\ln U\left(x\right)\right)+O\left(\frac{1}{n}\right)$

pour n assez grand. Donc :

$$d\left(N_n\left(x\right), Y\left(-\ln U\left(x\right)\right)\right) \leq \tfrac{1}{2n}U\left(x\right)\tfrac{(-\ln U(x))^{[-\ln U(x)]+2}}{[-\ln U(x)]!}+$$
$$\left|n\left(1-F\left(Z_n(x)\right)\right)+\ln U\left(x\right)\right|+O\left(\tfrac{1}{n}\right). \tag{4.18}$$

La partie (4.15) du Théorème (4.2.2) se déduit de (4.18) en utilisant (4.12).

4.3 Remarques

Remarque 4.3.1. *Les résultats (4.6) et (4.7) sont valables pour tout x fixé, de telle sorte que $0 < U(x) < 1$. Le choix de x vérifiant $0 < -\ln U(x) < 1$ pour (4.6) et $-\ln U(x) > 1$ pour (4.7) est fait pour améliorer les liens entre $-\ln U(x)$ et $(-\ln U(x))^2$: $(-\ln U(x))^2 \leq -\ln U(x)$ pour le premier cas et $-\ln U(x) \leq (-\ln U(x))^2$ pour le second cas.*

Remarque 4.3.2. *Dans les Théorèmes 4.2.1 et 4.2.2, nous avons obtenu l'approximation suivante :*

$$\left|P\left(X_{n-k+1,n} \leq Z_n(x)\right) - U\left(x\right)\sum_{j=1}^{k-1}\frac{\left(-\ln U\left(x\right)\right)^j}{j!}\right|$$

$$= |n\left(1 - F\left(Z_n(x)\right)\right) + \ln U\left(x\right)| + O\left(\frac{1}{n}\right).$$

La vitesse de convergence dans (4.5) dépend ainsi de la vitesse de convergence de $n\left(1 - F\left(Z_n(x)\right)\right) + \ln U\left(x\right)$ *vers* 0, *donc du choix adéquat des constantes de normalisation* u_n *et* v_n. *En outre, même lorsque le "bon choix" de ces constantes est fait, les vitesses de convergence peuvent être complètement différentes pour des lois différentes, la vitesse de convergence* $O\left(\frac{1}{n}\right)$ *trouvée est rapide.*

4.4 Exemple

Exemple 4.4.1. *Considérons* $F(x) = 1 - (\log x)^{-1}\ \mathbb{I}_{[e,\infty[}(x)$ *où* $F \in D_e(H_{1,1})$. *D'après Ravi et Mavitha [45] [Théorème (4.1)], il existe deux constantes* $u_n = n$ *et* $v_n = 1$ *telles que pour* $x > e$

$$\left|P\left(X_{n-k+1,n} \leq \exp(n\log x)\right) - \exp(-(\log x)^{-1}) \sum_{j=1}^{k-1} \frac{(-\ln(x))^j}{j!}\left(x\right)\right|$$

$$= O\left(\tfrac{1}{n}\right) + |n\left(1 - F\left(\exp(n\log x)\right)\right) - (\log x)^{-1})|$$

$$= O\left(\tfrac{1}{n}\right) + \left|n\left(\tfrac{1}{n}\left(\log x\right)^{-1}\right) - (\log x)^{-1}\right|.$$

Donc

$$\left|P\left(X_{n,n} \leq \exp(n\log x)\right) - \exp(-(\log x)^{-1}) \sum_{j=1}^{k-1} \frac{(-\ln(x))^j}{j!}\left(x\right)\right| = O\left(\tfrac{1}{n}\right).$$

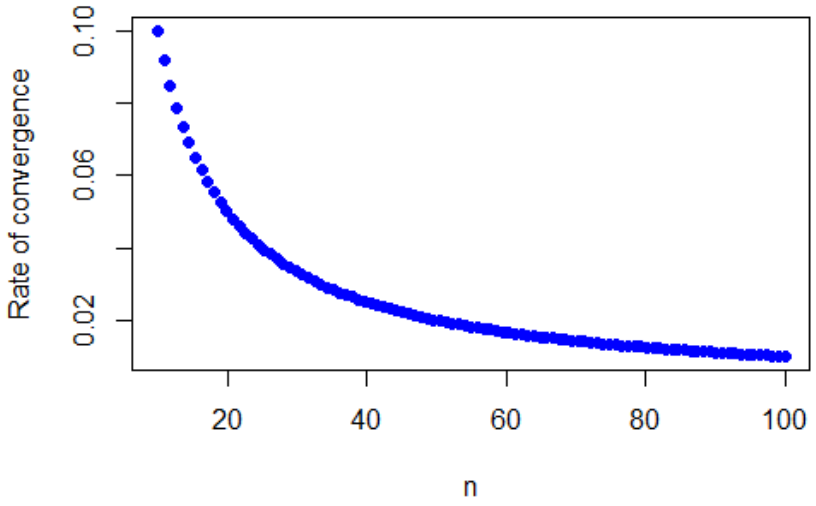

FIGURE 4.1 – La vitesse de convergence sous la normalisation exponentielle.

la vitesse de convergence sous la normalisation exponentielle est rapide sous certaine condition par rapport à celle sous la normalisation non linéaire (voir Barakat et al. [6]).

Chapitre 5

Fonctions de normalisation et fiabilité de l'estimation du niveau de retour GEV : Une étude comparative basée sur l'indice CAC 40

5.1 Introduction

La crise financière et économique qui a durement frappé le monde, a conduit au développement de nouveaux outils de protection contre les risques. Le théorème central limite (TCL), qui décrit le comportement asymptotique de la moyenne d'un grand nombre de variables indépendantes et qui est largement utilisé pour établir les résultats probabilistes requis, ne semble plus adéquat lorsque les événements extrêmes surviennent, comme lors des krachs boursiers et des crises financières. Ainsi, une théorie est apparue

entre 1920 et 1940, appelée théorie des valeurs extrêmes (TVE), pour la modélisation de tels événements. Pour une présentation complète du sujet, nous renvoyons au livre d' Embrechts et al. [20] qui expose les principaux résultats de la théorie des valeurs extrêmes. Reiss [47] pour sa part a proposé un certain nombre d'exemples pratiques. Les domaines d'application de la TVE n'a cessé de croître lors de ces dernières années dans différents domaines : en hydrologie pour la prévision des inondations voir (Davison et Smith [13] et Katz [30]), dans l'assurance où l'une des préoccupations est la prise en compte de grandes créances [voir McNeil et al. [34], Rootzén et Tajvidi [50]]. Leur introduction en finance (voir Embrechts et al. [20], Danielsson et de Vries [12], McNeil [34], Embrechts [19]) est une réponse particulière à la remise en cause de l'hypothèse de normalité avec les observations à hautes fréquences. En météorologie, de nombreux auteurs dont Stuart et Walshaw [11] ont étudié la vitesse du vent, et par exemple, évalué le degré de résistance à la pression exercée par le vent (lors d'une tempête) sur des bâtiments ou des ouvrages de génie civil. La théorie des valeurs extrêmes développée à ce jour suggère deux approches différentes. La première, appelée approche des maxima par blocs (MB), est basée sur le théorème de Fisher-Tippet [22], qui assure la convergence de la fonction de répartition des maxima par blocs vers une fonction de répartition cumulative de valeurs extrêmes généralisées (GEV). La deuxième approche est appelée approche des dépassements au-delà d'un seuil u (POT). La méthode POT est basée sur le théorème de Balkema et de Haan [1], Pickands [43] et Embrechts et al. [19]. Ce théorème montre que la fonction de survie d'une variable aléatoire au delà d'un seuil u assez grand, s'identifie asymptotiquement à une distribution de Pareto généralisée (GPD : Generalized Pareto Distribution). La théorie des valeurs extrêmes (TVE) concerne, en partie, la

détermination de la fonction de répartition du maximum pour n grand. Elle suggère deux types de normalisation pour la convergence asymptotique, à savoir la normalisation linéaire et la normalisation non linéaire. La normalisation linéaire est une méthode traditionnelle utilisée depuis longtemps pour caractériser les distributions marginales asymptotiques. Elle fournit une approximation suffisamment simple pour les fonctions de répartition des statistiques d'ordre. La normalisation linéaire a fait l'objet de nombreuses études initiées par Fisher et Tipett [22], Gnedenko [25], Balkema et de Haan [1], Pickands [43]. En utilisant la normalisation linéaire, ces auteurs ont montré qu'il existe exactement trois types de lois limites de valeurs extrêmes : Weibull, Gumbel et Fréchet. Par contre, la normalisation puissance peut construire des approximations simplifiées de fonction de répartition pour les extrêmes, en utilisant une classe plus large de lois limites que celle trouvée pour la normalisation linéaire. La normalisation non linéaire(puissance) a été rarement étudiée. Une brève étude de ce cas a été réalisée par Pancheva [40]. Récemment Barakat [2] a développé la distribution de Pareto généralisée DPGP basée sur le théorème de Balkema et de Haan [1], Pickands [43] et Embrechts et al. [20].

L'intérêt de la TVE est d'estimer la probabilité d'occurrence des événements extrêmes par la modélisation statistique d'événements rares. C'est un outil approprié pour la gestion des risques car il permet de mieux caractériser les niveaux de retour et les quantiles extrêmes. L'estimation des quantiles extrêmes a fait l'objet de plusieurs recherches dont celle de Weismann [55]. Les domaines d'application des quantiles extrêmes sont nombreux parmi lesquels on peut citer : La finance par Embrechts et al. [20], l'assurrance par Beirlant et Teugels [9] et la climatologie par Rootzén et Tajvidi [50].

Le niveau de retour est défini comme une valeur qui devrait être dépassée

en moyenne une fois pendant un intervalle de temps T avec une probabilité de $\frac{1}{T}$. Plusieurs auteurs ont abordé l'estimation du niveau de retour pour la normalisation linéaire. On peut citer le travail de Gardes et Girard [24] qui ont utilisé la distribution de Weibull pour estimer la période de retour dans le cas de la normalisation linéaire.

Dans ce chapitre, nous proposons une approche théorique pour estimer le niveau de retour ou le quantile extrême d'une fonction de répartition d'une loi limite dans le cas de la normalisation puissance. Ensuite, nous comparons les périodes de retour et le quantile extrême pour les normalisations puissance et linéaire et nous testons la performance de ces modèles. La question que nous nous posons est la suivante : des deux modèles de normalisation (linéaire ou non linéaire) lequel fournit-il plus de précision sur le niveau de retour ?

La seconde section de ce chapitre donne un aperçu de la théorie des valeurs extrêmes pour la normalisation linéaire et non linéaire. Dans la troisième section nous rappelons l'estimation du niveau de retour dans le cas de la normalisation linéaire et donnons une nouvelle expression de celle du niveau de retour dans le cas de la normalisation puissance. Les résultats théoriques de cette dernière sont basés sur les travaux de Barakat et al. [4]. La dernière section est une application de ces méthodes à l'indice du CAC 40 et vérifie la pertinence de la distribution des valeurs extrêmes généralisée (GEV) pour les normalisations puissance et linéaire notées respectivement GEVP et GEVL. Elle met en évidence les résultats des différentes périodes de retour et d'estimations des quantiles obtenus par GEVP et GEVL.

5.2 Méthodologie

Soit $(X_1, \cdots, X_n)$ un échantillon de variables aléatoires indépendantes et identiquement distribuées (i.i.d) de fonction de répartition commune F. Pour étudier le comportement des événements extrêmes, on considère le maximum d'un échantillon de taille n dans chaque bloc (un bloc peut être une année, un trimestre, un mois, etc.), noté :

$$M_n = \sup(X_1, \cdots, X_n)$$

La fonction de répartition de M_n est donnée par :

$$\begin{aligned} P(M_n \leq x) &= P(X_1 \leq x, \cdots, X_n \leq x) \\ &= [F(x)]^n. \end{aligned}$$

le fait que $lim_{n\to\infty} F^n(x) = 0$ pour $0 \leq F(x) < 1$, ne présente aucun intérêt, il est intéressant de tenir compte du comportement asymptotique du maximum normalisé, de façon analogue à celle du théorème central limite qui décrit le comportement d'une distribution limite d'une somme de variables aléatoires i.i.d. L'objectif de la théorie des valeurs extrêmes est de trouver un résultat similaire pour M_n. Plusieurs auteurs ont étudié le comportement asymptotique du maximum pour $n \to +\infty$ en utilisant la normalisation linéaire et non linéaire(puissance).

Premièrement, nous donnons un aperçu des résultats fondamentaux de la théorie des valeurs extrêmes pour la normalisation non linéaire. Ensuite, nous développons une approche théorique du calcul de la période de retour ou quantiles extrêmes pour la normalisation non linéaire.

Théorie des valeurs extrêmes pour la normalisation linéaire

La théorie des valeurs extrêmes est basée sur le travail de Fisher et Tippett [22] et Gnedenko [25] où la loi du maximum est déterminée pour un échantillon indépendant et identiquement distribué, de fonction de répartition commune F. Les trois lois limites trouvées sont de types : $\Lambda(x) = \exp(-\exp(-x))$, $\phi(x) = \exp(-x^{-\alpha})$, pour $x > 0$ et $\psi(x) = \exp(-(-x)^{\alpha})$, pour $x < 0$. Ce sont respectivement les lois de Gumbel, Fréchet et Weibull. Celles-ci peuvent être combinées en une seule famille de distribution limite (dite représentation de Jenkinson-Von Mises) définie par la fonction de distribution H_ξ qui est donnée par :

$$H_\xi(x) = \begin{cases} \exp\left\{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{\frac{-1}{\xi}}\right\}, & \text{si} \quad \xi \neq 0 \\ \exp\left\{-\exp\left\{-\left(\frac{x-\mu}{\sigma}\right)\right\}\right\}, & \text{si} \quad \xi = 0 \end{cases} \tag{5.1}$$

où $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$ et μ, σ et ξ sont les paramètres de position, d'échelle et de forme respectivement.

La loi des excès pour la normalisation linéaire (GPDL)

La méthode de dépassement d'un seuil, appelée Peaks Over Threshold (POT), consiste à observer les valeurs qui dépassent un seuil u et dont la loi de probabilité peut être approchée par une loi de Pareto Généralisée lorsque u est assez grand. L'approche POT permet d'améliorer l'analyse des événements extrêmes, car elle utilise plus d'observations que la méthode des blocs de maxima. Cette approche garantit que la fonction de distribution des observations dépassant un seuil u assez grand, peut être approximée par la distribution de Pareto généralisée (GPD). Ce résultat est initié par Balkema et de Haan [1] et Pickands [43]. La distribution de Pareto généralisée,

notée G_ξ, est donnée par :

$$G_\xi(x) = \begin{cases} 1 - \left(1 + \xi \frac{x-\mu}{\sigma}\right)^{\frac{-1}{\xi}}, & \text{si} \quad \xi \neq 0 \\ 1 - \exp\left\{-\frac{x-\mu}{\sigma}\right\}, & si \quad \xi = 0 \end{cases} \tag{5.2}$$

avec $1 + \xi \frac{x-\mu}{\sigma} > 0$ où μ, σ, ξ sont respectivement des paramétres de localisation, d'échelle et de forme.

Théorie des valeurs extrêmes pour la normalisation non linéaire (GEVP)

Pancheva [41] a caractérisé les limites de la fonction de répartition dans la théorie des valeurs extrêmes pour la normalisation non linéaire, elle obtient une large classe de lois limites. Soient $X_1, ..., X_n$ des variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F continue. En utilisant le résultat de Ferignac [21] sur la limite de la loi de probabilité de la statistique d'ordre centrale sous normalisation linéaire, Pancheva [40] définit la distribution limite pour la normalisation non linéaire par :

$$H_{i:\beta} = \exp\left[-\mu_{i:\beta}(x)\right], \quad i = 1, 2...., 6, and\ \beta > 0.$$

où

$$\mu_{1:\beta}(x) = \begin{cases} \infty, & x \leq 1 \\ (\log x)^{-\beta}, & x > 1, \end{cases} \qquad \mu_{2:\beta}(x) = \begin{cases} \infty, & x \leq 1 \\ (-\log x)^{\beta}, & 0 < x < 1 \\ 0, & x > 1, \end{cases}$$

$$\mu_{3:\beta}(x) = \begin{cases} \infty, & x \leq -1 \\ (-\log(-x))^{-\beta}, & -1 < x \leq 0 \\ 0, & x > 0, \end{cases} \qquad \mu_{4:\beta}(x) = \begin{cases} (\log(-x))^{\beta}, & x \leq -1 \\ 0, & x > -1, \end{cases}$$

$$\mu_5(x) = \begin{cases} \infty, & x \leq 0 \\ \frac{1}{x}, & x > 0, \end{cases} \qquad \mu_6(x) = \begin{cases} |x|, & x \leq 0 \\ 0, & x > 0. \end{cases} \tag{5.3}$$

les notations $\mu_{i:\beta}(x) = \mu_i(x), i = 5, 6.$

Nasri et Roudsari [37] ont synthétisé les distributions limites par :

$$H_{1:\xi}(x, a, b) = \exp\left[-\left(1 + \xi \log\left(ax^b\right)\right)^{\frac{-1}{\xi}}\right], x > 0, 1 + \xi \log\left(ax^b\right) > 0, \tag{5.4}$$

et

$$H_{2:\xi}(x, a, b) = \exp\left[-\left(1 - \xi \log\left(a\left(-x\right)^b\right)\right)^{\frac{-1}{\xi}}\right] \; x < 0, 1 - \xi \log\left(a\left(-x\right)^b\right) > 0. \tag{5.5}$$

La distribution des excès pour la normalisation non linéaire (GPDP)

Pour caractériser la distribution limite de la normalisation puissance, nous utilisons la suite définie par : $T_n(x) = a_n |x|^{b_n} S(x)$, avec $S(x) = -1, 0, 1$ si respectivement $x < 0$, $x = 0, x > 0$, et $a_n, b_n > 0$ sont des constantes. La fonction de répartition du maximum M_n est donnée par :

$$P\left(T_n^{-1}\left(M_n\right) \leq x\right) = P\left[\left|\frac{M_n}{a_n}\right|^{\frac{1}{b_n}} S\left(M_n\right) \leq x\right] \to H(x). \tag{5.6}$$

lorsque $n \to \infty$. H est une fonction de distribution non dégénérée qui appartient à la famille de distributions définie dans (5.3).

Barakat et al. [4] ont développé de nouvelles méthodes de calcul de la GPDP résumées dans les deux résultats suivants :

Théoréme 5.2.1. *Soit* $F^{[u]} = P\left(X \leq x | X > u\right)$

— *Si l'hypothèse (5.6) est satisfaite avec* $H\left(x\right) = H_{1;\xi}\left(x, a, b\right)$*, alors il existe* $\alpha(u) > 0$ *tel que lorsque* $u \uparrow x_F > 0$*,*

$$F^{[u]}\left(ux^{\alpha(u)}\right) \to Q_{1,\xi}\left(x, b\right)$$

où $Q_{1,\xi}\left(x, b\right) = 1 + \log H_{1;\xi}\left(x, 1, b\right)$*.*

— *Si (5.6) est satisfaite avec* $H\left(x\right) = H_{2;\xi}\left(x, a, b\right)$*, alors il existe* $\alpha(u) > 0$ *tel que lorsque* $u \uparrow x_F \leq 0$*,*

$$F^{[u]}\left(ux^{\alpha(u)}\right) \to Q_{2,\xi}\left(x, b\right)$$

où $Q_{2,\xi}\left(x, b\right) = 1 + \log H_{2;\xi}\left(x, 1, b\right)$*.*

Théoréme 5.2.2 (Propriété de stabilité POT). .

Les données tronquées à gauche pour la GPDP donnent à nouveau un GPDP. Cela signifie que pour tout $0 < k < x$*, on a*

$$Q_{1,\xi}^{[k]}(x; \sigma) = Q_{1,\xi}\left(\frac{x}{k}, \bar{\sigma}\right),$$

avec $\bar{\sigma} = \frac{\sigma}{c}$ *avec* $\sigma > 0$ *et* $c = 1 + \xi \log k$*.*
Si $-1 < k < x < 0$*, on a*

$$Q_{2,\xi}^{[k]}(x; \sigma) = Q_{2,\xi}\left(\frac{x}{k}, \bar{\sigma}\right),$$

avec $\bar{\xi} = \frac{\sigma}{\acute{c}}$ *et* $\acute{c} = 1 - \xi \log\left(-k\right)$*.*

5.3 Estimation du niveau de retour pour une distribution extrême dans les deux cas de normalisation

Nous nous intéressons au problème de l'estimation du niveau de retour ou quantile pour la normalisation linéaire et non linéaire sur la GEV.

Estimation du niveau de retour de la GEVL

Le niveau de retour noté $\mathbf{z_p}$ est défini par
$H_\xi(z_p) = 1-p \iff \mathbf{z_p} = H_\xi^{-1}(1-p)$ où la fonction de répartition H_ξ est définie dans (5.1).
Le niveau de retour est alors :

$$\mathbf{z_p} = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{si} \quad \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\}, & \text{si} \quad \xi = 0. \end{cases} \tag{5.7}$$

En notant $y_p = -\log(1-p)$, $\mathbf{z}_p$ s'écrit ainsi :

$$\mathbf{z_p} = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1 - y_p^{-\xi}\right], & \text{si} \quad \xi \neq 0 \\ \mu - \sigma \log y_p, & \text{si} \quad \xi = 0. \end{cases} \tag{5.8}$$

Une estimation du niveau de retour de la GEVL est obtenue comme suit :

$$\hat{\mathbf{z}}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - y_p^{-\hat{\xi}}\right], & \text{si} \quad \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log y_p, & \text{si} \quad \hat{\xi} = 0, \end{cases} \tag{5.9}$$

où $\hat{\mu}$, $\hat{\sigma}$ et $\hat{\xi}$ sont respectivement les estimateurs des paramètres de localisation μ, d'échelle σ et de forme ξ.

Estimation du niveau de retour de la GEVP

Nous nous concentrons maintenant sur l'expression du niveau de retour pour la normalisation non linéaire. L'estimation du niveau de retour en cas de normalisation puissance nécessite l'utilisation de la représentation de Von Mises des lois limites définies dans (5.4) et (5.5). Le niveau de retour $\mathbf{z}_p$ est déterminé en inversant la fonction de la représentation de Von Mises.
A partir de (5.4), on calcule l'expression analytique du niveau de retour z_p qui s'énonce comme suit :

$$\mathbf{z_p} = \left(\frac{1}{a} \exp \left\{ \frac{1}{\xi} \left(-\log\left(1-p\right) \right)^{-\xi} - 1 \right\} \right)^{\frac{1}{b}}, \quad \text{si} \quad \xi \neq 0. \tag{5.10}$$

Par utilisation de l'équation (5.5), on aboutit à l'écriture du niveau de retour z_p ainsi :

$$\mathbf{z_p} = \left(-\frac{1}{a} \exp \left\{ \frac{-\left(-\log\left(1-p\right)^{-\xi} + 1 \right)}{\xi} \right\} \right)^{\frac{1}{b}}, \quad \text{si} \quad \xi \neq 0. \tag{5.11}$$

L'estimation du niveau de retour $\mathbf{z_p}$, est :
- Dans le cas de (5.4) :

$$\hat{\mathbf{z}}_{\mathbf{p}} = \left(\frac{1}{\hat{a}} \exp \left\{ \frac{1}{\hat{\xi}} \left(-\log\left(1-p\right) \right)^{-\hat{\xi}} - 1 \right\} \right)^{\frac{1}{\hat{b}}}, \quad \text{pour} \quad \hat{\xi} \neq 0. \tag{5.12}$$

- Dans le cas de (5.5) :

$$\hat{\mathbf{z}}_{\mathbf{p}} = \left(-\frac{1}{\hat{a}} \exp \left\{ \frac{\left(1 - \left(-\log\left(1-p\right) \right)^{-\hat{\xi}} \right)}{\hat{\xi}} \right\} \right)^{\frac{1}{\hat{b}}}, \quad \text{pour} \quad \hat{\xi} \neq 0, \tag{5.13}$$

où $\hat{\xi}$, $\hat{a}$ et $\hat{b}$ sont les estimateurs des paramètres ξ, a et b respectivement.

Estimation du niveau de retour du GPDP

Pour la seconde approche, celle des excès, en utilisant le résultat suivant du Théorème 5.2.1 : lorsque $u \uparrow x_F > 0$

$$F^{[u]}\left(ux^{\alpha(u)}\right) \to Q_{1,\xi}\left(x, \bar{b}\right),$$

avec $F^{[u]}(x) = P(X \leq x/X > u)$. Pour u assez grand et $x > 0$,

$$P(X > x/X > u) = -\log H_{1;\xi}\left(x, a, \bar{b}\right) = \left(1 + \xi \log a x^{\bar{b}}\right)^{\frac{-1}{\xi}}.$$

En notant $P(X > u) =: \xi_u$, on trouve alors $P(X > \mathbf{z_p}) = \xi_u \left(1 + \xi \log a z_p^{\bar{b}}\right)^{\frac{-1}{\xi}}$, $\mathbf{z_p}$ étant le niveau de retour, à partir duquel $\xi_u \left(1 + \xi \log a z_p^{\bar{b}}\right)^{\frac{-1}{\xi}} = p$.

Pour $\xi \neq 0$, si nous fixons $a = 1$, lorsque $u \uparrow x_F > 0$, on obtient

$$\mathbf{z_p} = \exp\left\{\frac{1}{\xi \bar{b}}\left[\left(\frac{\xi_u}{p}\right)^{\xi} - 1\right]\right\}. \tag{5.14}$$

Pour $u \uparrow x_F \leq 0$, on obtient l'expression du niveau de retour :

$$\mathbf{z_p} = -\exp\left\{\frac{1}{\xi \underline{b}}\left(1 - \left(\frac{\xi_u}{p}\right)^{\xi}\right)\right\}.$$

L'estimateur du niveau de retour est

$$\hat{\mathbf{z}}_{\mathbf{p}} = \begin{cases} \exp\left\{\frac{1}{\hat{\xi}\bar{b}}\left[\left(\frac{\xi_u}{p}\right)^{\hat{\xi}} - 1\right]\right\}, & \text{si} \quad u \uparrow x_F > 0 \\ -\exp\left\{\frac{1}{\hat{\xi}\underline{b}}\left(1 - \left(\frac{\xi_u}{p}\right)^{\hat{\xi}}\right)\right\}, & \text{si} \quad u \uparrow x_F \leq 0. \end{cases} \tag{5.15}$$

Selon Barakat et al. [4], $\hat{\xi} = (\log 2)^{-1} \log \frac{\log Y_m - \log Y_{2m}}{\log Y_{2m} - \log Y_{4m}}$ pour la fonction de distribution cumulative $Q_{1,\xi}\left(x, \bar{b}\right)$ et pour $Q_{2,\xi}(x, \underline{b})$,

$\hat{\xi} = (\log 2)^{-1} \log \frac{\log|Y_m| - \log|Y_{2m}|}{\log|Y_{2m}| - \log|Y_{4m}|}$, où $Y_i = X_{n-i+1:n}$ est la i^{me} plus grande observation d'un échantillon.

Dans ce qui précède, nous avons développé différentes approches théoriques pour estimer les quantiles extrêmes ou les distributions de niveau de retour des valeurs extrêmes, pour la normalisation linéaire et non linéaire. Dans la suite de ce chapitre, nous énonçons le résultat caractérisant les lois extrêmes présenté par Belbachir et Azzat [27]. Cette caractérisation permet d'établir un lien entre les distributions extrêmes dans le cas de la normalisation linéaire et celui de la normalisation non linéaire. Nous décrivons la caractérisation des résultats du lemme développés dans Belbachir et Azzat [27] comme suit :

Si on note $Y_i = \log X_i$, $1 \leq i \leq n$, de fonction de répartition commune F_1 et où $(X_1 ...., X_n)$ est un échantillon de variables aléatoires i.i.d de fonction de répartition commune F, alors

$$F \in D_p \{\phi_1, H_{1,\xi}, H_{2,\xi}\} \Leftrightarrow F_1 \in D_l \{\wedge, \phi_\xi, \psi_\xi\}. \tag{5.16}$$

Ce résultat montre que, sous certaines conditions, les lois limites extrêmes dans le cas de la normalisation non linéaire appartiennent au domaine d'attraction des distributions limites pour la normalisation linéaire, développée dans le travail de Gnedenko [25]. Nous utiliserons, cette caractérisation pour estimer le niveau de retour pour les deux normalisations. Nous illustrerons notre approche en utilisant des données réelles de l'indice CAC 40.

5.4 Résultats et discussion

Étude de cas

Dans cette section, la théorie des valeurs extrêmes est appliquée au modèle des événements extrêmes de l'indice CAC 40 à l'aide des modèles GEVL et GEVP. Le CAC40 est le principal indice boursier de la Bourse de Paris, officiellement créé le 15 juin 1988. Sa création fait suite à la catastrophe de 1987 qui a modifié le monopole de la négociation d'actions. Il est déterminé à partir du cours de quarante actions continuellement cotées sur le premier marché de cent entreprises. Les échanges CAC40 sont les plus abondants sur Euronext Paris, qui fait partie d'Euronext. La valeur du CAC40 a subi des changements radicaux depuis sa création. Créé avec 1000 points de base le 31 décembre 1987, il a atteint son plus haut niveau le 4 septembre 2000 à 6 944,77 points et la forte baisse du krach boursier de 2008, lorsque le CAC40 a perdu plus de 43,5. La figure 5.1 souligne la nature

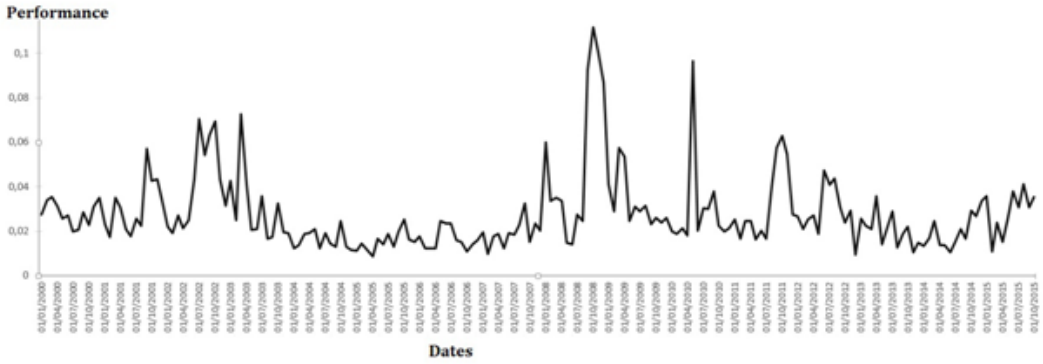

FIGURE 5.1 – Graphique d'évolution de l'indice CAC 40 entre le mois 01/2000 et le mois 10/2015

volatile de l'indice CAC40, qui justifie une étude des données aberrantes. Pour appliquer les travaux de Fisher et Tippett [22], de Gnedenko [25] et de

Pancheva [40] pour modéliser la queue de l'indice CAC40 en normalisation linéaire et non linéaire, il convient de tester l'hypothèse d'indépendance et d'homogénéité pour la valeur maximale mensuelle de l'indice CAC40. Nous utilisons les tests d'indépendance et d'homogénéité non paramétriques car la distribution des données est inconnue. En appliquant le test du point de retournement et de Mann-Whitney aux données brutes, on constate que les maximums mensuels CAC40 ne répondent pas aux exigences d'indépendance et d'homogénéité. Afin de répondre aux exigences de conception, nous avons examiné les rendements mensuels maximaux du CAC40. Selon les mêmes tests, les rendements mensuels maximaux sont indépendants et homogènes.

Pour modéliser les queues de distribution pour la GEVL et la GEVP, nous utilisons le graphique MRL plot (fonction de l" evd " package qui donne le graphique de la durée de vie résiduelle moyenne) pour détecter le type de distribution limite extrême qui peut être ajustée à nos observations. Le MRL plot, connu sous le nom de "parcelle en excès moyen", est couramment utilisé pour analyser les données de précipitation (Coles [10]) et les grandes réclamations en assurance non-vie (Beirlant et al. [8]). Zoglat et al. [57] ont appliqué l'approche POT à l'indice CAC40 qui vise à améliorer l'efficacité des estimateurs des quantiles extrêmes. Le graphique "MRL plot" fournit des informations sur différentes distributions limites extrêmes. Quand $\xi < 0$, la distribution extrême est bornée et on dit que l'on est dans le domaine d'attraction de Weibull. Quand $\xi = 0$, la distribution extrême présente une décroissance exponentielle dans la queue, on dit que l'on se trouve dans le domaine d'attration de Gumbel. Enfin, le domaine d'attration de Fréchet correspond à une distribution extrême non bornée et à une décroissance polynomiale lorsque $\xi > 0$.

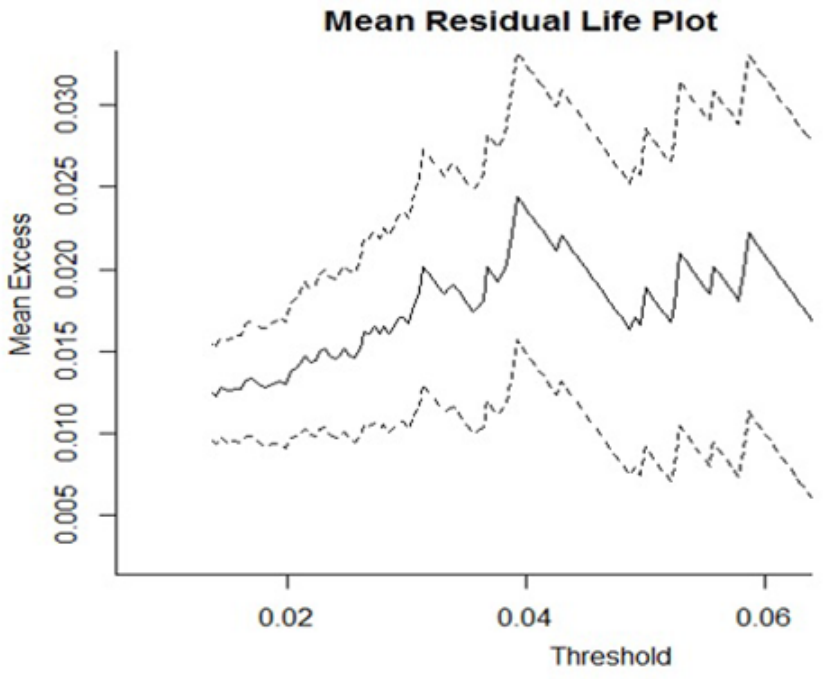

FIGURE 5.2 – MRL plot des données qui ont une performance mensuelle maximale

La Figure 5.2 montre une tendance croissante. Le graphique montre que les données relatives aux rendements maximaux mensuels peuvent être modélisées par la distribution de Fréchet. Pour l'emplacement du seuil 0.04 il y'a une instabilité de la vie résiduelle moyenne qui est due à la petite taille de l'échantillon. Nous avons 29 observations.

Il y a clairement une tendance décroissante dans la figure 5.3. Ce graphique indique que les observations logarithmiques pour la performance mensuelle maximale par rapport à l'indice CAC40, peuvent être ajustées en fonction de la distribution de Weibull à queue épaisse. L'instabilité du graphique à partir de l'emplacement du paramètre -3,5 est due à la petite taille de l'échantillon.

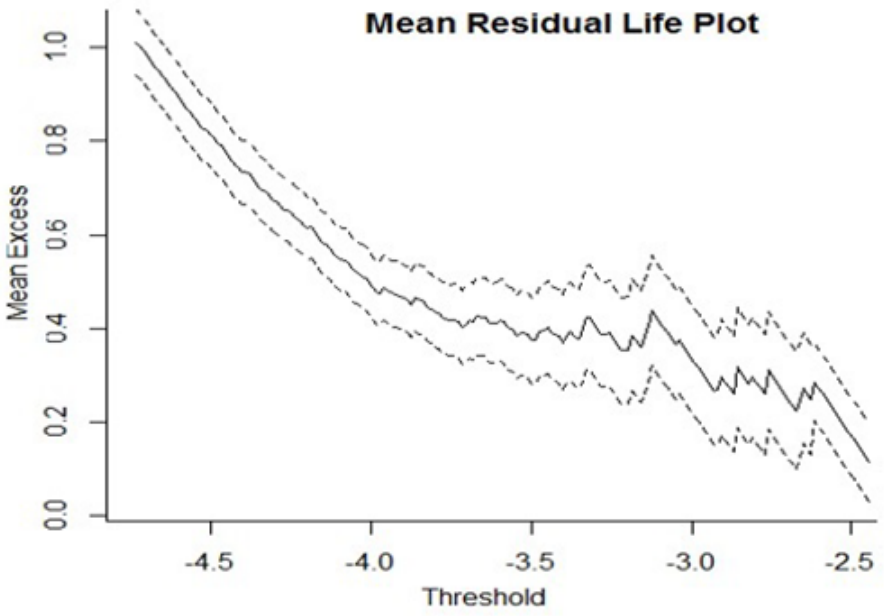

FIGURE 5.3 – MRL plot des données logarithmiques pour une performance mensuelle maximale

Estimation par la méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance a été développée par Fisher en 1922. Couramment utilisée pour établir l'estimation des paramètres de lois extrêmes (voir Smith [53] et Zoglat et al. [56]), elle reste populaire pour estimer les paramètres pour les modèles linéaires et non linéaires.

Dans ce chapitre, nous avons utilisé "evd Package" du logiciel R 3.3.0 pour estimer les paramètres de la distribution asymptotique extrême. Les rendements mensuels du CAC40 sur la période allant de 01/2000 à 01/2015 ont été utilisés pour estimer les paramètres de localisation μ, d'échelle σ et d'indice de queue ξ.

-Cas de la normalisation linéaire

En appliquant cette méthode d'estimation, on construit la distribution de probabilité des rendements :

$$H_{\hat{\xi}}(x) = \exp\left(-\left(1+\hat{\xi}\times\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)\right)^{-\frac{1}{\hat{\xi}}}\right). \tag{5.17}$$

où $\hat{\mu} = 0.019579$, $\hat{\sigma} = 0.008315$ et $\hat{\xi} = 0.301953$ sont les estimations des paramètres de position, d'échelle et de forme.

-Cas de la normalisation non linéaire

De façon similaire, on construit la distribution de probabilité des rendements :

$$H_{2,\hat{\xi}}(x) = \exp\left(-\left(1+\hat{\xi}\times\left(\frac{\log x-\hat{\mu}}{\hat{\sigma}}\right)\right)^{\frac{-1}{\hat{\xi}}}\right). \tag{5.18}$$

où $\hat{\mu} = -3.9333$, $\hat{\sigma} = 0.4341$ et $\hat{\xi} = -0.1051$ sont les estimations des paramètres de position, d'échelle et de forme.

Quantiles extrêmes et période de retour

Dans la section précédente, nous avons donné une estimation de la fonction de distribution extrême en utilisant la méthode du maximum de vraisemblance. En particulier, nous avons estimé ξ, μ et σ, respectivement, les paramètres de forme, de localisation et d'échelle pour la distribution des valeurs extrêmes. Cette étape nous a permis d'ajuster les rendements mensuels maximaux et de construire la distribution de probabilité.

L'étude des rendements mensuels maximaux est généralement motivée par l'estimation du quantile extrême ou la queue de probabilité en fonction du type d'application considéré. C'est l'un des principaux objectifs de la finance car il permet d'estimer le quantile extrême de l'indice boursier et la période de retour.

Dans ce chapitre, nous étudions la performance des modèles GEVL et GEVP. A cette fin, nous comparons les quantiles extrêmes prédits et les périodes de retour des modèles GEVL et GEVP. Nous résumons notre démarche comme suit :

Premièrement, nous estimons le quantile extrême en utilisant les approches GEVL et GEVP définies dans la Section 3.
- Pour la GEVL, le quantile extrême $\mathbf{z_p}$ est estimé au niveau de confiance $1-p$ par la relation suivante :

$$\hat{\mathbf{z}}_{\mathbf{gevl}} = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1 - \{-\log(1-p)\}^{-\hat{\xi}}\right], \tag{5.19}$$

où $\hat{\mu} = 0.019579$, $\hat{\sigma} = 0.008315$, $\hat{\xi} = 0.301953$ sont les estimations des paramètres de localisation, d'échelle et de forme avec $p \in (0,1)$.

- Pour la GEVP, le quantile extrême $\mathbf{z_p}$ est estimé au niveau de confiance $1-p$ par la relation suivante :

$$\hat{\mathbf{z}}_{\mathbf{gevp}} = \exp\left(\hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[-\{-\log(1-p)\}^{-\hat{\xi}} - 1\right]\right). \tag{5.20}$$

où $\hat{\mu} = -3.9333$, $\hat{\sigma} = 0.4341$ et $\hat{\xi} = -0.1051$ sont les estimations des paramètres de localisation, d'échelle et de forme avec $p \in (0,1)$.

Nous avons développé l'algorithme suivant pour estimer les quantiles extrêmes $\hat{\mathbf{z}}_{\mathbf{gevl}}$ et $\hat{\mathbf{z}}_{\mathbf{gevp}}$:

étape 1 : Simuler ν échantillons, de taille n selon la loi uniforme sur (0,1)

étape 2 : Calculer $\hat{\mathbf{z}}^{\mathbf{L}}_{\mathbf{ij}}$ et $\hat{\mathbf{z}}^{\mathbf{P}}_{\mathbf{ij}}$ en remplaçant p, le niveau de probabilité par p_{ij} où $1 \leq i \leq n, 1 \leq j \leq \nu$, dans les relations $\hat{\mathbf{z}}_{\mathbf{gevl}}$ et $\hat{\mathbf{z}}_{\mathbf{gevp}}$.

étape 3 : Calculer $\hat{\mathbf{z}}_{\mathbf{gevl}}$ et $\hat{\mathbf{z}}_{\mathbf{gevp}}$ par :

$$\hat{\mathbf{z}}_{\mathbf{gevp}}(i) = \frac{1}{\nu}\sum_{j=1}^{\nu} \hat{\mathbf{z}}_{\mathbf{ij}}^{\mathbf{P}}, \tag{5.21}$$

et

$$\hat{\mathbf{z}}_{\mathbf{gevl}}(i) = \frac{1}{\nu}\sum_{j=1}^{\nu} \hat{\mathbf{z}}_{\mathbf{ij}}^{\mathbf{L}}. \tag{5.22}$$

Les quantiles extrêmes estimés $\hat{\mathbf{z}}_{\mathbf{gevl}}$ et $\hat{\mathbf{z}}_{\mathbf{gevp}}$ sont comparés aux quantiles empiriques calculés à partir de l'inverse de la fréquence empirique. (voir Tableau 5.1).

Rendements mensuels	Quantile observé	Quantile prédit par	
		GEVL	GEVP
0.008609	0.02865904	0.029096881	0.02499326
0.024605	0.02777911	0.02819042	0.02418369
0.027683	0.02761795	0.02807072	0.02407434
0.030085	0.02844296	0.02913417	0.0246655
0.030974	0.02701921	0.02725079	0.02389681
0.033227	0.02766104	0.02869106	0.02416615

TABLE 5.1 – Quantile observé et prédit (mensuel)

Les résultats du tableau 5.1 et la figure 5.4 montrent que le modèle GEVL surestime les quantiles observés. Contrairement au GEVL, le modèle GEVP sous-estime les quantiles observés.

Deuxièmement, en finance, nous sommes particulièrement intéressés par l'estimation de la période de retour T associée à la probabilité $p \in \left\{\frac{1}{10}, \frac{1}{100}, ...\right\}$, pour les modèles GEVL et GEVP. Dans cette partie, nous estimons les périodes de retour des modèles GEVL et GEVP et nous les

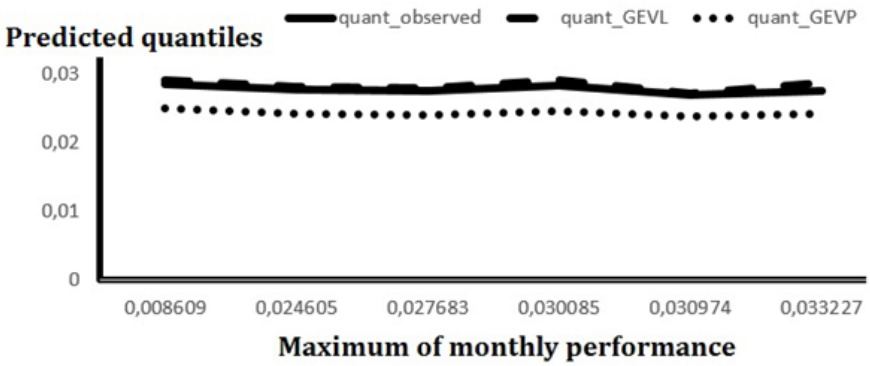

FIGURE 5.4 – Graphique du quantile observé et prédit

comparons aux périodes de retour calculées à partir de la fréquence empirique. La période de retour T_{x_p} d'une valeur critique x_p peut être calculée comme suit :

$$T_{x_p} = \frac{1}{P\left(X \geq x_p\right)} = \frac{1}{1 - P\left(X < x_p\right)}. \tag{5.23}$$

Pour notre étude de cas, nous calculons la période de retour en utilisant deux approches.

— La première utilise la distribution des valeurs extrêmes généralisée dans le cas de la normalisation linéaire (GEVL), la période de retour est alors déterminée par :

$$T_{x_p} = \frac{1}{1 - H_\xi\left(x_p\right)}, \tag{5.24}$$

où H_ξ la distribution de la GEV, est définie par le théorème de Fisher-Tippett [22] et celui de Gnedenko [25], ξ étant le paramètre d'indice de queue.

— La seconde approche utilise le modèle de la GEV pour la normalisation

non linéaire GEVP. La période de retour est calculée par :

$$T_{x_p} = \frac{1}{1 - H_{i,\xi}(x_p)}, \tag{5.25}$$

où $H_{i,\xi}, i = 1, 2$, la GEV, est définie par une distribution non linéaire Nasry-Roudsari [37] et ξ est le paramètre d'indice de queue.

Les périodes de retour calculées à partir des modèles GEVL et GEVP, sont présentées dans le tableau 5.2. Nous utilisons des distributions $H_{\hat{\xi}}(x)$, $H_{2,\hat{\xi}}(x)$ respectivement la GEVL et la GEVP. Les périodes de retour observées sont calculées à partir des fréquences empiriques (voir Tableau 5.2).

Rendement mensuels	Périodes de retour observées	période de retour prédites	
		GEVL	GEVP
0.008609	1.005025	1.004628	1.003634
0.024605	2.150538	2.289856	2.266809
0.027683	2.857143	2.885121	2.83724
0.030085	3.225806	3.444795	3.372636
0.030974	3.571429	3.674551	3.59253
0.033227	4.347826	4.314794	4.206308

TABLE 5.2 – Périodes de retour observées et prédites (mensuellement)

Les résultats du tableau 5.2 et la figure 5.5 montrent que le modèle GEVP donne des périodes de retour très proches de celles calculées à partir de la fréquence empirique. Les estimations obtenues par la méthode GEVP présentent un biais moins important. Nous pouvons en conclure que le modèle GEVP est le plus performant pour l'indice de marché CAC 40, plus réaliste que le modèle GEVL.

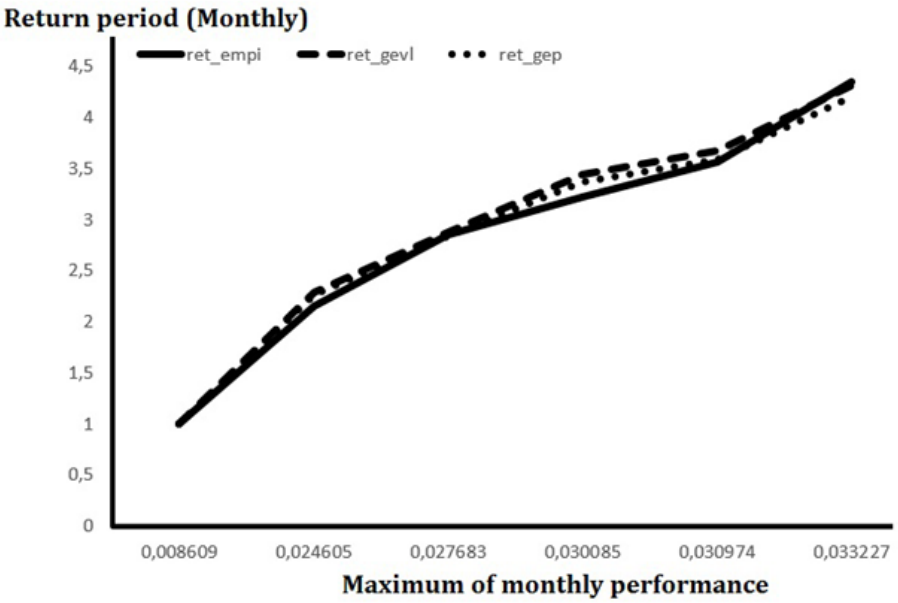

FIGURE 5.5 – Graphique des périodes de retour observées et prédites.

Chapitre 6

Conclusion Générale

La théorie des valeurs extrêmes (TVE) constitue l'une des approches statistiques appliquées en finance. Telle que présentée dans le présent mémoire, elle continue de jouer un rôle important dans la pratique de l'ingénierie pour la gestion financière. Cette théorie peut être appliquée pour améliorer la rigueur des applications financières et pour effectuer des analyses plus significatives. La TVE est fortement recommandée en finance pour l'évaluation des risques de volatilité du marché. Les résultats théoriques de la TVE ne sont valables que si les conditions asymptotiques sont satisfaites. La normalisation des variables est une étape importante pour étudier la convergence asymptotique de distribution de probabilité des valeurs extrêmes. Différentes méthodes ont été proposées pour la normalisation. Il s'agit de la normalisation linéaire et de la normalisation non linéaire (puissance).

Ce travail propose une approche théorique basée sur la fonction de Von Mises. Il consiste à estimer le niveau de retour de la TVE dans le cas de la normalisation non linéaire. Un algorithme basé sur les méthodes

analytiques a été proposé pour estimer les quantiles extrêmes et les périodes de retour de la GEV dans le cas de la normalisation linéaire et non linéaire. Les résultats présentés dans ce travail peuvent être utilisés pour prédire les performances extrêmes et le niveau de rendement. L'intérêt de cette démarche réside dans le fait que les paramètres du GEVP peuvent être influencés par la présence d'une ou de plusieurs variables sur le marché financier.

Ces mesures de risque sont un nouvel outil pour la théorie des valeurs extrêmes sous la normalisation non linéaire. Il serait aussi intéressant d'appliquer notre méthode à un nouveau type de données et de pouvoir estimer le niveau de retour dans d'autres domaines d'application, par exemple en actuariat ou en hydrologie.
Nous souhaitons encore appliquer la vitesse de convergence pour la normalisation non linéaire dans la pratique, pour voir son importance.
Enfin, il est primordial de garder à l'esprit que les résultats obtenus sont basés sur des modèles et qu'ils sont une simplification de la réalité ; il convient en conséquence d'adopter une certaine prudence dans l'interprétation des résultats obtenus.

Bibliographie

[1] Balkema, A. A. et De Haan, L., Residual life time at great age. Residual life time at great age. *Annals of Probability*, 2(5), 792-804, 1974.

[2] Barakat, H. M., The use of power normalization as a new trend in the order statistics limit theory. *Journal of Statistics Applications & Probability*, 2(3), 251, 2013.

[3] Barakat, H. M., Nigm, E. M. et El-Adll, M. E., Continuation theorems of the extremes under power normalization. *Journal of Applied Mathematics and Computing*, 10(1-2), 1, 2002.

[4] Barakat, H. M., Nigm, E. M. et Khaled, O. M., Extreme value modeling under power normalization. *Applied Mathematical Modelling*, 37(24), 10162-10169, 2013.

[5] Barakat, H. M. et Omar, A. R., Limit theorems for intermediate and central order statistics under nonlinear normalization. *Journal of Statistical Planning and Inference*, 141(1), 524-535, 2011.

[6] Barakat, H. M., Nigm, E. M. et El-Adll, M. E., Comparison between the rates of convergence of extremes under linear and under power normalization. *Statistical Papers*, 51(1), 149, 2010.

[7] Beirlant, J. et Teugels, J. L., Asymptotic normality of Hill's estimator. Extreme value theory, 51, pp. 148-155, 1987.

[8] Beirlant, J. , Goegebeur, Y. , Segers, J. et Teugels, J. L., *Statistics of extremes : theory and applications*. John Wiley & Sons, 2006.

[9] Beirlant, J. et Teugels, J. L., Modeling large claims in non-life insurance. *Insurance : Mathematics and Economics*, 11(1), 17-29, 1992.

[10] Coles, S., Bawa, J., Trenner, L. et Pat, D., *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

[11] Stuart, G. C. et Walshaw, D., Directional modelling of extreme wind speeds. Journal of the Royal Society Series C *Applied Statistics*, 43(1), 139-157, 1994.

[12] Danielsson, J. et de Vries, C., Tail index and quantile estimation with very high frequency data. *Journal of empirical Finance*, 4(2-3), 241-257, 1997.

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

[14] de Haan, L. et Rootzén, H., On the estimation of high quantiles. *Journal of Statistical Planning and Inference*, 35(1), 1-13, 1993.

[15] Resnick, S. et de Haan, L., Second-order regular variation and rates of converges in extreme value theory. *Annals of Probability*, 24(1), 97-124, 1996.

[16] Deheuvels, P., Haeusler, E. et Mason, D. M., Almost sure convergence of the Hill estimator. In *Mathematical Proceedings of the Cambridge*

Philosophical Society, (Vol. 104, No 2, pp. 371-381), Cambridge University Press, 1988.

[17] Deheuvels, P. et Pfeifer, D., A semigroup approach to poisson approximation. *Annals of Probability*, 4(2), 663-676, 1986.

[18] Deheuvels, P. et Pfeifer, D., Poisson approximations of multinomial distributions and point processes. *Journal of multivariate analysis*, 25(1), 65-89, 1988.

[19] Dekkers, L. M., Einmahl, H. E. et De Haan, L., A moment estimator for the index of an extreme-value distribution. *Annals of Statistics*, 17(4), 1833-1855, 1989.

[20] Embrechts, P., Kluppelberg, C. et Mikosch, T., Modelling extremal events. *British Actuarial Journal*, 5(2), 465-465, 1999.

[21] Embrechts, P., Kluppelberg, C. et Mikosch, T., *Modelling extremal events : for insurance and finance*, Springer, 1997.

[22] Ferignac, P., Test de Kolmogorov- Smirnov sur la validité d'une fonction de distribution. *Rev. Stat. Appl*, 10, 13-32, 1962

[23] Fisher, R. A. et Tippett, H. C., Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pp. 180-190. Cambridge Univ Press, 1928.

[24] Freedman, D., The poisson approximation for dependent events. *Annals of Probability*, 2(2), 256-269, 1974.

[25] Gardes, L. et Girard, S., Estimation de quantiles extrêmes pour les lois à queue de type weibull : une synthèse bibliographique. *Journal de la Société Française de Statistique*, 154(2) :98-118, 2013.

[26] Gnedenko, B., Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44 , 423-453, 1943.

[27] Grigelionis, B., Limit theorems for record values using power normalization. *Lithuanian Mathematical Journal*, 46(4), 398-405, 2006.

[28] Belbachir, M., Azzat, H., Limit law for maxima of random variables in the domain of attraction of the gumbel law. *International Journal of Tomography*, 17(S11), 39-45, 2011.

[29] Teugels, J. L. et Hausler, E., On asymptotic normaliy of Hill's estimator for exponent of regular variaion. *Annals of Statistics*, 13(2), 743-756, 1985.

[30] Hill, B. M., A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3, 1163-1174, 1975.

[31] Katz, R. W., Parlange, M. B. et Naveau, P., Statistics of extremes in hydrology. *Advances in water resources*, 25(8-12), 1287-1304, 2002.

[32] Kotz, S. et Nadarajah, S., *Extreme value distributions : theory and applications*. World Scientific, 2000.

[33] Le Cam L., An approximation theorem for the poisson binomial distribution. *Pacific J. Math*, 10(4), 1181-1197, 1960.

[34] Belbachir, M., El ouardirhi, S. et Labloul, B., On the rate of convergence in extreme value theory for non linear normalization. *Math-Recherche & Applications*, 16(1), 2018.

[35] Leadbetter, M. R., Lindgren, G. et Rootzen, H., *Extremes and related properties of random sequences and processes*. Springer Science & Business Media, 2012.

[36] McNeil, A. J., Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin : The Journal of the IAA*, 27(1), 117-137, 1997.

[37] Nasri-Roudsari, D., Limit distributions of generalized order statistics under power normalization. *Communications in Statistics-Theory and Methods*, 28(6), 1379-1389, 1999.

[38] Nasri-Roudsari, D. et Cramer, E., On the convergence rates of extreme generalized order statistics. *Extremes*, 2(4),421-447, 1999.

[39] Ravi N. R. et Mohan, S., Max domains of attraction of univariate and multivariate p-max stable laws. *Theory of Probability & Its Applications*, 37(4), 632-643, 1993.

[40] Pancheva, E., Extreme value limit theory with nonlinear normalization. *Extreme Value Theory and Applications*, pp. 305-318, Springer, Boston, MA, 1994.

[41] Pancheva, E., On a problem of a Khinchin type decomposition theorem for extreme values. *Theory of Probability et Its Applications*, 39(3), 329-336, 1995.

[42] Pancheva, E., Limit theorems for extreme order statistics under nonlinear normalization. In *Stability problems for stochastic models*, pp. 284-309, Springer, 1985.

[43] Peng, Z., Jiang, Q. et Nadarajah, S., Limiting distributions of extreme order statistics under power normalization and random index. *Stochastics An International Journal of Probability and Stochastic Processes*, 84(4), 553-560, 2012.

[44] Peng, Z., Shuai, Y. et Nadarajah, S., On convergence of extremes under power normalization. *Extremes*, 16(3), 285-301, 2013.

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

[46] Ravi, S., Tail equivalence of distribution functions belonging to the max domains of attraction of univariate p-max stable laws. *J. Indian Statist. Assoc.*, 29(2), 69-77, 1991.

[47] Ravi, S. et Mavitha, T. S., New limit distributions for extremes under a nonlinear normalization. 09, 1-20, 2016.

[48] Ravi, S. et Praveena, A. S., A note on tail behaviour of distributions in the max domain of attraction of the Frechét/Weibull law under power normalization. In *ProbStat Forum*, volume 3, pp. 1–10, 2010.

[49] Reiss, R. D., Approximate distributions of order statistics : with applications to nonparametric statistics. Springer, New York, 1989.

[50] Resnick, S. I., *Extreme values, regular variation and point processes.* Springer, New York, 1987

[51] Vervaat, W., upper bound for the distance in total variation between the binomial or negative binomial and the poisson distribution. *Statistica Neerlandica*, 23(1) : 79-86, 1969.

[52] Rootzén, H. et Tajvidi, N., Extreme value statistics and wind storm losses : a case study. *Scandinavian Actuarial Journal*, 1997(1), 70-94, 1997.

[53] Weinstein, S., Theory and application of some classical and generalized asymptotic distributions of extreme values. *IEEE Transactions on Information Theory*, 19(2), 148-154, 1973.

[54] Serfling, R. J., Some elementary results on poisson approximation in a sequence of bernoulli trials. *Siam Review*, 20(3), 567-579, 1978.

[55] Smith, R. L., Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1), 67-90, 1985.

[56] Smith, R. L., Estimating tails of probability distributions. *Annals of Statistics*, 15(3), 1174-1207, 1987.

[57] Weissman, I., Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73(364), 812-815, 1978.

[58] Zoglat, A., El Adlouni, S., Badaoui, F, Amar, A. et Okou, C. G., Managing hydrological risks with extreme modeling : application of peaks over threshold model to the Loukkos watershed, Morocco. *Journal of Hydrologic Engineering*, 19(9) :05014010, 2014.

[59] Zoglat, A., El Adlouni, S., Badaoui, F, Amar, A. et Okou, C. G.et Ezzahid, E., Statistical methods to expect extreme values : Application of POT approach to CAC40 return index. 2013.