

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire

وزارة التعليم العالي و البحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique  
Centre Universitaire Salhi Ahmed- Naama  
Institut des sciences et technologies  
Département de Mathématiques et Informatique



## *Mémoire de fin d'études*

En vue de l'obtention du diplôme de Master  
En : Mathématiques

Spécialité : Probabilités, Statistique et Application

## Intitulé

---

**Estimation des quantiles extrêmes pour des distributions  
à queues lourdes**

---

Présenté par :  
AIDA TALHA

Soutenu : Juillet 2022

Devant le jury composé de :

M.KHALWATI HAFIDA	MAA	C-Univ Naâma	Président
Dr.kANOUSA DJAMAMEL	MCA	C-Univ Naâma	Examinateur
Dr.BARI AMINA	MCB	C-Univ Tindouf	Encadreur

Année universitaire 2021/2022



## Dédicace

---

*Je dédie ce travail :*

*à l'esprit de l'ancien chef de mon département au Centre Universitaire  
de Tindouf, M. Hawassin, que Dieu tu fasse miséricorde, mon  
honorabile professeur  
à mon cher père Ibrahim talhia et à ma chère mère, qui ont été mon  
premier soutien après Dieu*

*Je dédie ce travail à mes chers soeurs Aisha, Naima et ma soeur Fatima  
Al-Zahra*

*Je dédie ce travail à mon cher frère Ezz El-Din et mon cher frère  
Nasrallah, que Dieu vous bénisse pour le reste de ma vie*

*Je dédie ce travail à toute la famille Talhia et Khachai  
à mes amis proches Nasreen, Zineb, Maryam et à mon cher ami Ismail  
Airou. Je vous souhaite du succès et une remise des diplômes réussie  
l'année prochaine..*

*Je dédie ce travail à tous ceux qui m'aiment et me souhaitent plus de  
succès.*

*Merci!*



## Remerciement



*Grâce à Dieu, qui m'a honoré de la santé et de la patience, j'ai pu terminer ce travail :*

*Il s'agit maintenant à remercier toutes les personnes qui ont contribué à la réalisation de ce travail. Avant tout, je voudrais remercier Dieu tout-puissant, qui m'a aidé et m'a donné la force et la patience tout au long de ces années scolaires et la force de mener à bien ce travail.*

*premièrement je tiens à exprimer mes remerciements et ma gratitude au Dr Bari Amina, pour la confiance qu'il m'a accordée en acceptant d'encadrer ma thèse et de l'avoir si bien fait. Je lui suis particulièrement reconnaissante pour ses conseils, vraiment j'ai pris beaucoup de plaisir à travailler avec elle.*

*Merci infiniment, Madame Bari.*

*Mes sincères remerciements aux enseignants du département de mathématiques et d'informatique qui nous ont encadrés depuis notre première année universitaire aussi es membres du jury qui J'aimerais remercier les membres du jury qui m'a fait l'honneur de m'accompagner jusqu'à la fin*

*j'en profite aussi pour exprimer ma gratitude s mes parents et mon oncle Houssin qui est devenu au fil des années comme mon père, a mes frères, Je tiens également à remercier mon ami Ismail Airou qui m'a aidé et toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.*

# Abréviation et notations

$F(x)$ :	Fonction de répartition de $x$ .
$f(x)$ :	Densité de probabilité de $x$ .
$\mu$ :	L'espérance ou moyenne
$F_n$ :	Fonction de répartition empirique.
$v.a$ :	Variable aléatoire.
$iid$ :	Indépendante et identiquement distribué
$\xrightarrow{\text{p.s.}}$ :	Convergence presque sûre
$\xrightarrow{\mathbb{P}}$ :	Convergence en probabilité
$\xrightarrow{\mathbb{L}}$ :	Convergence en loi
$N(0, 1)$ :	La loi normale standard
$TCL$ :	Théorème centrale limite
$TVE$ :	Théorie des valeurs extrêmes
$x_F$ :	Point terminale
$POT$ :	La méthode de Peak Over Threshold
$BM$ :	La méthode de Bloc Maxima
$IVE$ :	Indice des valeurs extrêmes
$GEV$ :	The generalized extreme value
$GPD$ :	The generalized Pareto distribution
$H_{\gamma, \sigma(u)}$	Distribution de Pareto généralisé
$H_{\mu, \sigma, \gamma}(x)$ :	Distributions généralisées des valeurs extrêmes
$\mathcal{D}(\mathcal{H}_\gamma)$ :	Domaine d'attraction
$l(x)$ :	Fonction à variation lente
$\mathcal{D}(\Phi_\alpha)$ :	Domaine d'attraction de Fréchet
$\mathcal{D}(\psi_\alpha)$ :	Domaine d'attraction de weibul
$\mathcal{D}(\Lambda)$ :	Domaine d'attraction de Gumbell
$EMV$ :	l'estimateur de maximum vraisemblance
$\hat{\gamma}_{k_n}^H$ :	L'estimateur de Hill
$\hat{\gamma}_{k_n}^P$ :	L'estimateur de Pickands
$\hat{\gamma}_n^{(M)}$ :	L'estimateur de moment
$\hat{q}(\alpha_n)$ :	L'estimateur de Weissman

# Résumé

La théorie des valeurs extrêmes est considérée comme un complément aux statistiques classiques, car c'est un moyen approprié d'extrapoler le comportement des queues de distributions en fonction des valeurs observées les plus grandes ou les plus petites, qui sont appelées valeurs extrêmes, de sorte que le comportement de ces valeurs n'est pas similaire à le comportement général d'autres données, ce qui conduit à la survenance d'événements dits extrêmes, sont des catastrophes qui peuvent causer des grands risques dans divers domaines tels que l'hydrologie, la finance...etc Cela reflète l'importance de cette théorie dans le domaine théorique plutôt que dans le domaine pratique, c'est pour ça nous avons présenté dans ce travail la théorie des valeurs extrêmes et sa rôle important dans la modélisation de ces événements extrêmes à travers la distribution généralisée des valeurs extrêmes ou selon la distribution généralisée de Pareto afin que ces lois incluent les paramètres qui peuvent être estimés soit à travers la méthode des valeurs maximales ou bien sur la distribution des données ayant dépassé un certain seuil, nous avons ensuite intéressé à l'estimation d'indice des valeurs extrêmes généralisé et pour l'indice des distributions à queues lourdes comme un cas particulier. à la fin, nous avons choisi deux distributions, nous les simulées puis nous les avons estimées à l'aide du logiciel R.

**Mots Clés :** Statistique d'ordre, Théorie des valeurs extrêmes, Le théorème de Fisher-Tippet, distribution à queues lourdes, estimateur de pickandes, estimateur de Hill.

## Abstract

Extreme value theory is considered a complement to classical statistics, as it is a suitable way to extrapolate the behavior of tails of distributions according to the largest or smallest observed values, which are called extreme values, so that the behavior of these values is not similar to the general behavior of other data, which leads to the occurrence of so-called extreme events, are disasters that can cause great risks in various fields such as hydrology, finance...etc. this reflects the importance of this theory in the theoretical field rather than in the practical field, that is why we have presented in this work the theory of extreme values and their important role in modeling these extreme events through the generalized extreme value distribution or according to the generalized Pareto distribution so that these laws include the parameters which can be estimated either through the method of the maximum values or on the distribution of the data having exceeded a certain threshold we then interesting has the estimate for the generalized extreme value index and for the heavy-tailed distributions index as a special case. At the end, we choose two distributions, we simulate them and then we estimate them using R software.

**Keywords:** Order statistics, extreme values theory, The theorem of Fisher-Tippet, Heavy tails distributions, pickand Estimator, Hill Estimator.

## ملخص

تعتبر نظرية القيمة المتطرفة مكتملة للإحصاءات الكلاسيكية ، لأنها طريقة مناسبة لاستقراء سلوك ذيول التوزيعات وفقاً لأكبر أو أصغر القيم المرصودة والتي تسمى القيم المتطرفة بحيث سلوك هذه القيم الشاذة لا يشبه السلوك العام للبيانات الأخرى مما يؤدي إلي وقوع مايسمي بالأحداث المتطرفة نعني بها الكوارث التي يمكن أن تسبب مخاطر كبيرة في مجالات متعددة كالأرصاد الجوية ، الهيدرولوجيا ، البيولوجيا والتأمين ، والتمويل ، الخ هذا مايعكس أهمية هذه النظرية في المجال النظري أكثر منه في المجال التطبيقي وعليه قدمنا في هذا العمل نظرية القيم المتطرفة ودورها في نمذجة هذه الأحداث المتطرفة من من خلال توزيع القيمة القصوى المعمم أو بموجب قانون باريتو المعمم بحيث تشمل هذه القوانين المعلمات التي يمكن تقديرها إما من خلال طريقة القيم القصوى واما من خلال التركيز علي توزيع البيانات المتجاوزة لعتبة معينة ثم قمنا بالتقدير بالطريقة لابرامترية والنصف برامترية لمؤشر القيم المتطرفة للتوزيع المعمم و للتوزيعات ذات الذيل الثقيلة بصفة خاصة وفي الأخير قمنا باختبار اثنين من هذه التوزيعات وقمنا بمحاكاتها وتقدير قيمها المتطرفة بمقدر هيل باستخدام برمجية R .

**الكلمات المفتاحية:** الاحصاء الكلاسيكية، نظرية القيم المتطرفة، نظرية فيشرتنبات، التوزيعات ثقيلة الذيل، مقدر بيكاند، مقدر هيل.



# Table des matières

<b>Dédicace</b>	<b>i</b>
<b>Remerciement</b>	<b>ii</b>
<b>Abréviations et notations</b>	<b>iii</b>
<b>Résumé</b>	<b>iv</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 La Théorie des valeurs extrêmes</b>	<b>4</b>
1.1 Statistique d'ordre . . . . .	4
1.1.1 Définitions et Caractérisations de bases . . . . .	4
1.1.2 Loi des grandes nombres et théorème central limite . . . . .	7
1.2 Comportement asymptotique des extrêmes . . . . .	8
1.2.1 Concepts des quantiles extrêmes . . . . .	8
1.2.2 Distribution du maximum et théorème de Fisher-Tippet . . . . .	8
1.3 Distribution des valeurs extrêmes généralisées . . . . .	13
1.3.1 La méthode de bloc maxima (BM) . . . . .	15
1.4 Distribution de Pareto généralisée(GPD) . . . . .	16
1.4.1 La méthode de Peak over threshold(POT) . . . . .	17
1.5 Domaines d'attraction . . . . .	19
1.5.1 Fonction à variation régulière . . . . .	19
1.5.2 Représentation de Karamata . . . . .	20
1.5.3 Domaine d'attraction de Fréchet : . . . . .	21
1.5.4 Domaine d'attraction de weibull : . . . . .	22
1.5.5 Domaine d'attraction de Gumbel : . . . . .	22
<b>2 Estimation d'indice des valeurs extrêmes (IVE)</b>	<b>24</b>
2.1 Estimation paramétrique de (IVE) . . . . .	24
2.1.1 Estimateur de maximum vraisemblance (EMV) . . . . .	24
2.2 Estimation semi paramétrique de (IVE) . . . . .	25

2.2.1	Estimateur de Pickands(1975)	25
2.2.2	Estimateur de moments	27
2.3	Estimation d'indice de queue lourde	28
2.3.1	Distribution a queues lourdes	28
2.3.2	Estimateur de Hill(1975)	29
2.3.3	Estimation des quantiles extrêmes	31
2.3.4	Estimateur de weissman(1978)	32
<b>3</b>	<b>Application en logiciel R</b>	<b>33</b>
3.1	Application sur la loi de Pareto	34
3.1.1	Simulation de loi Pareto en logiciel R	34
3.1.2	Estimation de Pareto par Hill	36
3.2	Application sur loi de log-gamma	39
3.2.1	Simulation de loi log-gamma en logiciel R	39
3.2.2	Estimation de loi log-gamma par Hill	41
	<b>Conclusion générale</b>	<b>43</b>
	<b>Bibliographie</b>	<b>44</b>

# Table des figures

1	Niveaux maximums annuels de la mer à Port Pirie, Australie du Sud. . . .	3
1.1	Représentation de la fonction de répartition $F(x)$ et de densité $f(t)$ . . . .	6
1.2	la quantile intermédiaire est en bleu et extrême est en rouge . . . . .	9
1.3	Représentation de la distribution de Fréchet d $\psi_\alpha(x)$ et sa densité . . . . .	12
1.4	Représentation de la distribution de weibull $\psi_\alpha(x)$ et sa densité . . . . .	12
1.5	Représentation de la distribution de Gumbell $\Lambda(x)$ et sa densité . . . . .	13
1.6	Représentation de les trois distributions de GEV . . . . .	14
1.7	Queues des trois lois extrêmes :Weibull,Gumbel,Fréchet . . . . .	15
1.8	Représentation de la méthode de Block Maxima(BM) . . . . .	15
1.9	Densités et distributions de loi Pareto Généralisées avec différentes valeurs du paramètre $\gamma$ . . . . .	17
1.10	Représentation de la méthode de Peak over threshold(POT) . . . . .	17
2.1	Représentation de l'estimateur de Pickands . . . . .	26
2.2	Représentation de l'estimateur de moments . . . . .	28
2.3	Comparaison du Comportement de la Queue . . . . .	29
2.4	Représentation de l'estimateur de Hill . . . . .	31
3.1	Simulation de loi de Pareto $P1(100, 1, 1), P2(100, 1, 2)$ . . . . .	35
3.2	Simulation de loi de Pareto $P3(200, 1, 1), P4(200, 1, 2)$ . . . . .	35
3.3	Simulation de loi de Pareto $P5(500, 1, 1), P6(500, 1, 2)$ . . . . .	36
3.4	Estimation de loi de Pareto(P1,P1) par Hill . . . . .	37
3.5	Estimation de loi Pareto(P3,P4) par Hill . . . . .	37
3.6	Estimation de loi de Pareto (P5,P6) par Hill . . . . .	38
3.7	Simulation de loi de log-gamma $L3(200, 1, 1), L4(200, 1, 2)$ . . . . .	40
3.8	Simulation de loi de log-gamma $L5(500, 1, 1), L6(500, 1, 2)$ . . . . .	40
3.9	Estimation de loi de log-gamma(L1,L2) par Hil . . . . .	41
3.10	Estimation de loi de log-gamma(L3,L4) par Hill . . . . .	42
3.11	Estimation de loi de log-gamma(L5,L6) par Hill . . . . .	42

# Liste des tableaux

1.1	Les trois distributions de GEV selon l'indice $\gamma$ . . . . .	14
1.2	Les trois distribution de GEP selon l'indice $\gamma$ . . . . .	17
1.3	Domaines d'attraction de quelque lois usuelles . . . . .	23

# Introduction générale

La Théorie des valeurs extrêmes s'est imposée comme l'une des disciplines statistiques les plus importantes pour les sciences appliquées au cours des 50 dernières années, malgré est assez récente et ce n'est que dans les années 1980 que les contours de la théorie statistique se sont dessinés et la théorie probabiliste des valeurs extrêmes a été développée par M. Fréchet (1927), R. Fisher et L. Tippett (1928), et R. von Mises (1936), et a culminé dans les travaux de B. Gnedenko (1943); en effet la théorie statistique n'a été initiée que en (1975) par J. Pickands tel elle a reçu beaucoup d'attention aussi bien sur le plan théorique que sur le plan pratique...!

Ce qui nous amène à nous demander pourquoi déjà la théorie des valeurs extrêmes ?? quels sont leur champs d'applications et quelle est leur importance pour éviter de nombreuses menaces naturelles extrêmes ?

Au début, la pertinence de la théorie des valeurs extrêmes est l'analyse de la fréquence des inondations, les sciences de l'environnement, la finance et l'assurance et d'autres domaines très important, nous commençons par le domaine de l'assurance qui est l'une des applications les plus marquantes de la valeur extrême se retrouve dans l'assurance non-vie; en dehors des accidents majeurs tels que les tremblements de terre, les ouragans, les accidents d'avion et ainsi de suite, il existe un grand nombre d'occasions où des sinistres importants surviennent. de temps à autre, l'assurance automobile donne lieu à des sinistres excessifs.. les incendies industriels, en particulier, provoquent de nombreux effets secondaires tels que la perte de biens, le chômage temporaire et la perte de contrats, Une compagnie d'assurance se protégera toujours contre la contamination du portefeuille causée par des sinistres qui doivent être considérés comme extrêmes plutôt que moyens. et grâce à la théorie des valeurs extrêmes l'estimation de la queue supérieure de la distribution de la taille du sinistre est d'un intérêt majeur pour déterminer la prime nette d'un contrat de réassurance, En recherche environnementale et météorologie les données météorologiques ne présentent généralement pas d'aspects alarmants tant qu'elles se situent dans une bande étroite autour de la moyenne, la situation change par exemple lorsque concentrations qui dépassent un seuil écologique spécifique.

Généralement les techniques de valeur extrême sont également de plus en plus utilisées dans de nombreuses autres disciplines. Par exemple, l'évaluation des risques sur les marchés financiers et la prévision du trafic dans les télécommunications tel que il n'y a pas si long temps où des applications de la modélisation des valeurs extrêmes ont été publiées dans les domaines suivants prédiction de la résistance des alliages la modélisation des vagues océaniques (Dawson, 2000), de la défaillance des cellules de mémoire (McNulty et al., 2000), de l'ingénierie éolienne (Harris, 2001), du traitement des données traitement des données biomédicales (Roberts, 2000); l'évaluation des changements météorologiques (Thompson et al., 2001).

Donc nous constatons que le domaine de la statistique des valeurs extrêmes, offre une grande variété de problèmes. Hormis le problème classique de l'analyse de la distribution d'une seule variable aléatoire sur la base d'un échantillon aléatoire nous trouvons des structures de données pour lesquelles les modèles de séries temporelles, la régression et les paramètres multivariés sont appropriés, après les approches paramétriques et non-paramétriques, les techniques bayésiennes d'estimation des paramètres sont aussi désormais utilisées, aussi en hydrologie ou l'intérêt ultime de l'analyse de la fréquence des crues est l'estimation du débit (niveau d'eau) de la crue de l'année  $T$  (niveau d'eau), qui est le niveau dépassé tous les  $T$  ans en moyenne. Habituellement, on prend une période de 100 ans, mais l'estimation est souvent effectuée sur la base des débits de crue pour une période plus courte. des débits de crue pour une période plus courte, les conséquences des inondations dépassant un tel niveau peuvent être désastreuses. Par exemple, les niveaux de crue centennale ont été dépassés par la crue américaine de 1993 et ont causé des dégâts considérables et a provoqué une dévastation généralisée dans les états du le Mid-West aussi les inondations de 1953 aux Pays-Bas ont été vraiment catastrophiques et ont déclenché le plan Delta de digues de construction de digues, Un autre paramètre hydrologique pour lequel la queue de la distribution correspondante présente un intérêt particulier est l'intensité des précipitations. Ce paramètre est important pour modélisation des systèmes de cours d'eau, du drainage urbain et du ruissellement des eaux Grâce au étude d'extrêmes, d'autre exemple la figure (1) montre les niveaux maximums annuels de la mer enregistrés à Port Pirie, un endroit situé juste au nord d'Adélaïde, en Australie-Méridionale, sur la période 1923-1987. A partir de ces données, il peut être nécessaire d'obtenir une estimation du niveau maximum de la mer qui est susceptible de se produire dans la région au cours des 100 ou 1000 prochaines années. Cela soulève un point important - comment pouvons-nous estimer les niveaux qui pourraient se produire dans les 1000 prochaines années sans connaître!!

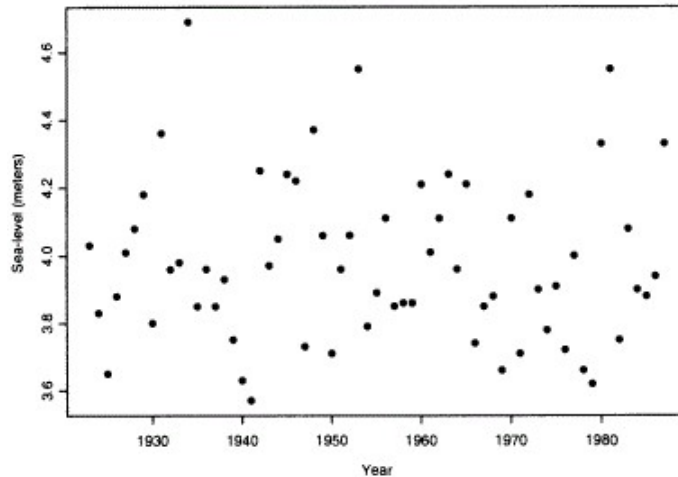


FIGURE 1 – Niveaux maximums annuels de la mer à Port Pirie, Australie du Sud.

Donc l'objectif de notre mémoire est de donner une idée globale de la théorie des valeurs extrêmes, puis une présentation complète des estimateurs des valeurs extrêmes dans le cas général et dans le cas des distributions à queue lourde en particulier, pour le faire de manière compréhensible nous avons divisé le travail en une introduction, trois chapitres et une conclusion générale, plus précisément la thèse que nous l'avons fait suivait le plan de travail Suivant :

**Chapitre1** : Dans le premier chapitre nous concentrerons Sur la théorie des valeurs extrêmes, où nous présenterons une idée complète de cette théorie à partir d'un rappel sur la statistique d'ordre, la théorème centrale limite, les différentes initiations de convergence et tous les concepts statistique essentielle qui nous aideront a construire notre thèse puis on fait une étude sur le comportement asymptotique du maximum d'un échantillon, Puis on décrivant les limites possibles de la loi du maximum d'un échantillon, là deux théorèmes sont essentiels : celui de Fisher-Tippett et celui de Balkema-de Haan Pickands, à la fin de ce chapitre on parle sur les trois domaines d'attraction (Fréchet, Weibull, Gumbel) et les approches les plus utilisées de blocs maxima (BM) et la méthode Peak-over.

**Chapitre2** : Dans le deuxième chapitre nous avons présenté les estimateur d'indice des queues  $\gamma$  les plus utilisés qui sont l'estimateur du maximum de vraisemblance quand on parle sur l'estimation non-paramétrique et les deux estimateurs semi-paramétrique de Hill (1975) et de Pickands (1975).

**Chapitre3** : Concernant le dernier chapitre nous prendrons deux distributions, nous allons les simuler puis les estimer avec l'estimateur de Hill à l'aide du logiciel R

Enfin nous avons présenté une conclusion générale sur cette théorie importante et ce que nous avons conclue à travers nos trois chapitre concernant les valeurs extrêmes et leurs estimations.

# La Théorie des valeurs extrêmes

Ce chapitre est consacré à la théorie des valeurs extrêmes que l'on appelle extrême value théorie (EVT) en anglais, cette théorie est de but d'étudier la loi du maximum d'une suite des variables aléatoires qui n'est pas toujours facilement à calculable, le TVE donne la solution à partir de proposé un analogue au théorème limite centrale (TCL) sous le nom de Fisher-Tippett qui est intéressant à considérer le comportement asymptotique de maximum convenablement normalisé, donc nous présenterons ici la base statistique et tous les principaux résultats classiques sur la théorie des valeurs extrêmes (GEV) et (GPD) et même si les deux approches de modélisation (POT) et (BM).

## 1.1 Statistique d'ordre

La théorie des valeurs extrêmes ne peut pas exister s'il n'y avait pas de statistique d'ordre, c'est parce que cette dernière fournit une base d'informations sur les queues de distribution qui s'intéresse principalement à la théorie des valeurs extrêmes.

Dans cette section, nous fournirons toutes les définitions de base de statistique d'ordre et les deux théorèmes fondamentaux de la statistique asymptotique, connu sous le nom de la loi des grands nombres et le théorème limite centrale.

### 1.1.1 Définitions et Caractérisations de bases

On considère  $(X_1, X_2, X_3, \dots, X_n)$   $n$  variables aléatoires indépendantes et identiquement distribuées de densité  $f$  et de fonction de répartition  $F$ , nous allons maintenant introduire un ensemble de concepts de base.

#### Définition 1.1.1. (Statistique d'ordre)

On appelle statistique d'ordre notées  $(X_{(1,n)}, \dots, X_{(n,n)})$  les variables aléatoires ordonnées comme suite :

$$X_{(1,n)} \leq X_{(2,n)} \leq X_{(3,n)} \leq \dots \leq X_{(n-1,n)} \leq X_{(n,n)}$$

**Remarque 1.1.1.** On a  $X_{(i,n)}$  est la  $i^{eme}$  statistique d'ordre ou statistique d'ordre  $i$  pour tout  $1 \leq i \leq n$

**Définition 1.1.2. (Statistique d'extrême)**

Soit la statistique d'ordre notée  $X_{(1,n)}, \dots, X_{(n,n)}$  les statistiques d'extrêmes sont définies comme suivant :

- la plus petite statistique d'ordre  $X_{(1,n)}$  (ou statistique du minimum) est :

$$X_{(1,n)} = m_n = \min(X_{(1)}, \dots, X_{(n)})$$

- la plus grande statistique d'ordre  $X_{(n,n)}$  (ou statistique du maximum) est :

$$X_{(n,n)} = M_n = \max(X_{(1)}, \dots, X_{(n)})$$

**Remarque 1.1.2.** On a  $\max(X_{(1)}, \dots, X_{(n)}) = -\min(X_{(1)}, \dots, X_{(n)})$

**Exemple : 1.1.** Si  $X$  est le niveau d'eau d'une rivière à un endroit donné  $J$  au jour  $j$  alors  $X_{(n,n)}$  est la seule valeur nécessaire pour l'analyse des inondations, et  $X_{(1,n)}$  pour les sécheresses.

**Définition 1.1.3. (Fonction de répartition)**

On appelle fonction de répartition d'une variable aléatoire  $X : \Omega \rightarrow \mathbb{R}$  la fonction  $F_X$  définie sur  $\mathbb{R}$  par :

$$F_X(x) = \mathbb{P}_X (] - \infty, x])$$

**Proposition 1.1.1. (Propriétés de la fonction de répartition)**

- pour tout  $x \in \mathbb{R}$ ,  $F_X(x) \in [0, 1]$  :

$$0 \leq F_X(x) \leq 1 \text{ avec } \lim_{x \rightarrow -\infty} F_X(x) = 0 \text{ et } \lim_{x \rightarrow +\infty} F_X(x) = 1$$

En effet,  $F(x)$  est une probabilité et quand  $x \rightarrow -\infty$  l'intervalle  $] - \infty, x[$  devient  $\emptyset$  et  $X^{-1}(\emptyset) = \emptyset$ ; quand  $x \rightarrow +\infty$ , l'intervalle  $] - \infty, x[$  devient  $X^{-1}(\mathbb{R}) = \Omega$ .

- Pour tous  $x \in \mathbb{R}$ ,  $F_X(x)$  est continue à gauche et croissante :

En effet, pour  $x < y$  on a  $(X < x) \Rightarrow (X < y)$  et par conséquent

$$F(x) = P_X(X < x) \leq P_X(X < y) = F(y).$$

- Pour tous  $a < b \in \mathbb{R}$ , on calcule la probabilité de l'intervalle  $[a, b[$  par :  $P_X(a \leq X < b) = F(b) - F(a)$  On a en effet :  $] - \infty, b[ = ] - \infty, a[ \cup [a, b[$  donc :  $P_X(X < b) = P_X(X < a) + P_X(a \leq X < b)$ .

**Définition 1.1.4. (Fonction de densité)**

Si la fonction de répartition  $F$  est différentiable, on définit la fonction de densité de probabilité  $X$  par :  $f(x) = \frac{dF}{dx}$

**Proposition 1.1.2. (Propriété de la fonction de densité)**

- Puisque  $f$  est la dérivée d'une fonction croissante  $F$  au sens large donc  $f$  est une densité est positive :  $f(x) \geq 0$

- Une fonction de densité  $f$  est intégrable sur  $\mathbb{R}$ , d'intégrale égale à un :

$$\int_{-\infty}^{+\infty} f(t) dt = 1$$

— En intégrant la densité  $f$  sur un intervalle on obtient la probabilité de ce intervalle :

$$P_X \{X \in [x_1, x_2]\} = \int_{x_1}^{x_2} f(t)dt. \text{ En effet :}$$

$$\begin{aligned} P_X (x_1 \leq X \leq x_2) &= F(x_2) - F(x_1) = \int_{-\infty}^{x_2} f(t)dt - \int_{-\infty}^{x_1} f(t)dt \\ &= \int_{-\infty}^{x_2} f(t)dt + \int_{x_1}^{-\infty} f(t)dt \end{aligned}$$

**Exemple : 1.2.** Supposant  $X$  une variable aléatoire de densité  $f(x) = x$  pour  $x \in [0, 1]$ ,  $f(x) = 2 - x$  pour  $x \in [1, 2]$  et  $f(x) = 0$  en dehors de ces intervalles.

On calcule la fonction de répartition de  $X$  comme suivant :

on a  $F(x) = 0$  pour  $x \leq 0$  et  $F(x) = 1$  pour  $x \geq 2$

$$\text{pour } 0 \leq x \leq 1 : F(x) = \int_0^x t dt = \frac{x^2}{2}$$

$$\text{pour } 1 \leq x \leq 2 : F(x) = F(1) + \int_1^x (2 - t) dt = \frac{-x^2}{2} + 2x - 1$$

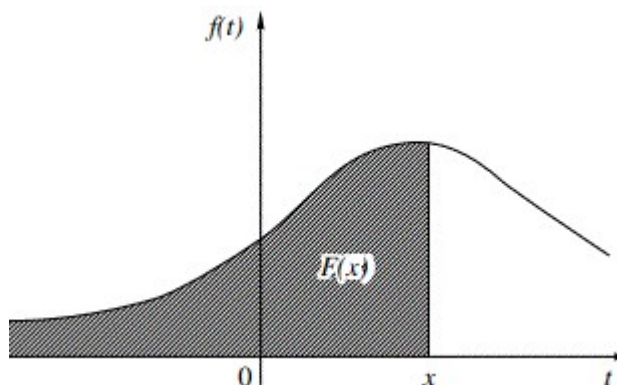


FIGURE 1.1 – Représentation de la fonction de répartition  $F(x)$  et de densité  $f(t)$

**Proposition 1.1.3. (La distribution d'extrêmes)**

— On note  $F_{X_{(1,n)}}(x)$  la distribution de minimum  $m_n$  la plus petite statistique d'ordre et définie par :

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

tel que sa densité est donné par :

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

— On note  $F_{X_{(n,n)}}$  la distribution de maximum  $M_n$  la plus grande statistique d'ordre qui définie par :

$$F_{X_{n,n}}(x) = [F(x)]^n, x \in \mathbb{R} \tag{1.2}$$

tel que sa densité est :

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

**Preuve 1.1.1.** Pour montrer (1.1) et (1.2) en utilisant la propriété iid des  $X_1, X_2, X_3, \dots, X_n$  nous en déduisons :

Démonstration de (1.1) :

$$\begin{aligned} F_{X_{1,n}}(x) &= P(X_{1,n} \leq x) = 1 - P(X_{1,n} \geq x) = 1 - P(X_1 \geq x \dots X_n \geq x) \\ &= 1 - P\left(\bigcap_{i=1}^n \{X_i \geq x\}\right) = 1 - \prod_{i=1}^n (1 - P\{X_i \geq x\}) = 1 - [F(x)]^n \end{aligned}$$

Démonstration de (1.2) :

$$\begin{aligned} F_{X_{n,n}}(x) &= P\{X_{n,n} \leq x\} = P\{X_1 \leq x \dots X_n \leq x\} \\ &= P\left(\bigcap_{i=1}^n \{X_i \leq x\}\right) = \prod_{j=1}^n P(X_j \leq x) = [F(x)]^n \end{aligned}$$

**Lemme 1.1.1.** La loi de variables aléatoire  $X_{k,n}$  pour  $1 < k \leq n$  est donnée par :

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

et sa densité est donnée par :

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

### 1.1.2 Loi des grandes nombres et théorème central limite

Les deux théorèmes fondamentaux de la statistique asymptotique sont premièrement la loi des grands nombres qui comprend la loi faible et loi forte, le deuxième c'est le théorème central limite.

**Théorème 1.1.1. (Loi forte des grande nombre)** Soit  $(X_n)$  une suite des variables aléatoires indépendantes admettent les mêmes moments d'ordres un et deux, c'est-à-dire avec pour tout entier  $n$  tel que  $E(X_n) = \mu$  et  $V(X_n) = \sigma^2$  et quand  $n \rightarrow \infty$ , alors :

$$\bar{X}_n \xrightarrow{\text{p.s.}} \mu$$

**Théorème 1.1.2. (Loi faible des grands nombres)** Soit  $(X_n)$  une suite des variables aléatoires indépendantes admettent les mêmes moments d'ordres un et deux, c'est-à-dire avec pour tout entier  $n$  tel que  $E(X_n) = \mu$  et  $V(X_n) = \sigma^2$  et quand  $n \rightarrow \infty$ , alors :

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mu$$

comme on a vu la loi des grands nombres qui énonce la convergence de la suite  $\overline{X}_n$  des moyennes empiriques vers la moyenne théorique  $\mu$

**Théorème 1.1.3. limite centale** Si  $(X_n)$  une suite des v.a qui sont iid, admettent les mêmes moments d'ordres un et deux, c'est-à-dire avec pour tout entier  $n$  tel que  $E(X_n) = \mu$  et  $V(X_n) = \sigma^2$  en tant que  $n \rightarrow +\infty$  si  $E(X^2) < \infty$  alors :

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{\text{loi}} N(0, 1)$$

## 1.2 Comportement asymptotique des extrêmes

L'objectif de la théorie des valeur extrême (TVE) est l'étude du comportement asymptotique des extrêmes ,s'il faut bien sûr l'étude des valeurs maximales d'un échantillon des variables aléatoire indépendante et identiquement distribuée, dans cette section, nous devons d'abord introduire quelques notions de base des statistiques extrêmes, puis nous présenterons le comportement asymptotique des extrêmes.

### 1.2.1 Concepts des quantiles extrêmes

De même que nous avons présenté toutes définitions statistiques classiques, nous allons maintenant introduire quelques concepts nécessaires aux statique d'extrêmes.

**Définition 1.2.1.** On définit le Point terminal d'un distribution comme suite :

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

**Définition 1.2.2.** On appelle quantile d'ordre  $p$ , avec  $p \in [0, 1]$ , le nombre  $x_p$  défini par :

$$x_p = \inf \{x \in \mathbb{R} : F(x) \geq p\}$$

**Définition 1.2.3.** On note la fonction quantile de la variable aléatoire  $X$   $Q$  pour  $\forall p \in [0, 1]$  qui définie par :

$$Q(p) = F^{\leftarrow}(p) = \inf \{x \in \mathbb{R} : F(x) \geq p\}$$

### 1.2.2 Distribution du maximum et théorème de Fischer-Tippet

L'objectif de la théorie des valeurs extrêmes (TVE) est l'étude du comportement asymptotique des extrêmes qui passe naturellement par l'analyse du maximum d'un échantillon des variables aléatoire iid.

Soit  $(X_1, X_2, X_3, \dots, X_n)$  un échantillon des variables aléatoire indépendante et identiquement distribuées suite un loi  $F$  définie par :  $F(x) = \mathbb{P}(X \leq x)$ .

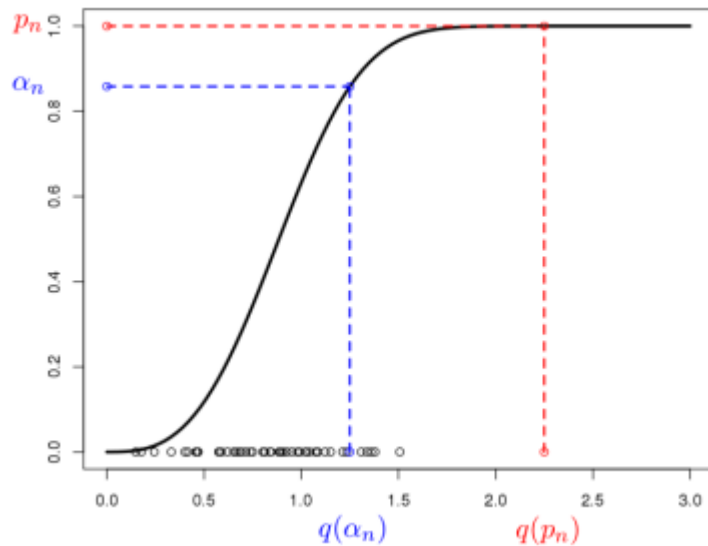


FIGURE 1.2 – la quantile intermédiaire est en bleu et extrême est en rouge

Considérons  $x_1, x_2, X_3..x_n$  un échantillon des variables ,nous les classons par ordre décroissant comme suivante :

$$x_{(1)} \geq x_{(2)} \geq x_{(3)} \geq \dots x_{(n)}$$

Le nième  $x_{(n)}$  peut être considérée un représentant de  $X_{(n)}$  qui une nouvelle variable aléatoire et le même ça marche avec tous les autres observations.

Maintenant,les nouvelles n.variables aléatoires sont s'écrit sous la nouvelle relation :

$$X_{(1)} \geq X_{(2)} \geq X_{(3)} \geq \dots \geq X_{(n)}$$

où la plus grande statistique on a définie par :

$$M_n = \max(X_1, X_2, X_3, \dots X_n)$$

est aussi parfois notée  $X_{(n,n)}$ ,pour étudier la comportement des extrêmes nous intéressons au cette valeur maximale et les même démarches on peut l'utilisée avec le minimum grâce a la premier remarque

Maintenant,puisque ces variables sont indépendante et identiquement distribuées et grâce a la proposition (1,1,3), la fonction de distribution de  $M_n$  est définie par :

$$F_{M_n} = \mathbb{P}(M_n \leq x) = [F(x)]^n \tag{1.3}$$

Donc on remarque que la distribution du terme maximum  $M_n$  n'est pas toujours facile à calculée même si la loi de X parent est connue,pour résoudre cette problématique , le théorème de Fisher-Tippett donne la forme de la possible asymptotique d'un échantillon de taille n,bien sûr nous restons toujours dans le contexte d'un ensemble de n variables

aléatoires  $X_{(1)}, \dots, X_{(n)}$  qui sont iid.

**Théorème 1.2.1. (Théorème de Fisher-Tippett )**

Soit  $(X_i, i \in I)$  n variable aléatoire indépendante et identiquement distribuées (i.i.d) de loi  $F_X(x)$  est dans un domaine d'attraction de maximum ,alors ils existe des suites normalisé sont  $(a_n > 0$  et  $b_n \in \mathbb{R})$  et une loi non-dénégérée  $\mathcal{H}$  te que :

$$\Pr \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} = (F(a_n x + b_n))^n \rightarrow \mathcal{H}(x) \quad (1.4)$$

Alors H appartient à l'un des trois types :

$$\text{Type I (Frchet)} \quad \Phi_\alpha(x) = \begin{cases} 0 & \text{si } x < 0 \\ \exp(-x^{-\alpha}) & \text{si } x \geq 0 (\alpha > 0) \end{cases}$$

$$\text{Type II (Weibull)} \quad \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x < 0 \\ 1 & \text{si } x \geq 0 (\alpha > 0) \end{cases}$$

$$\text{Type III (Gumbel)} \quad \Lambda(x) = \exp(-\exp(-x)), x \in \mathbf{R}$$

On note les distribution par  $\Phi_\alpha, \Psi_\alpha$  et  $\Lambda$  sont appelées les lois des maxima.

La laissons existe entre distribution de Weibul et Frécht est donnees comme suivante :

$$\Phi_\alpha\left(\frac{-1}{z}\right) = \Psi_\alpha(z)$$

avec :  $z = \frac{x - b}{a}, a > 0, b \in \mathbb{R}$  et  $\alpha > 0$

Le théorème [1.2.1] implique que, lorsque  $M_n$  peut être stabilisé avec des suite normelisé  $a_n$  et  $b_n$  , la variable normalisée correspondante  $M_n$  a une distribution limite qui doit être l'un des trois types de distribution de valeurs extrêmes. la caractéristique remarquable de ce résultat est que on peut écrit les trois famille de distributions des valeurs extrêmes en seul forme tel que sont les seules limites possibles pour les distributions de  $M_n$ , quelle que soit la distribution F de la population.en ce sens ce théorème[1.3.1] fournit un du théorème central limite central[1.1.3]et si on fait un parallél avec (TLC) on trouve que :

- $a_n$  est le paramètre d'échelle represente  $\sigma\sqrt{n}$  telque  $\sigma$  est l'écart type de la variable aléatoire X en (TCL).
- $b_n$  est le paramètre de position il represente  $n\mu$  tel que  $\mu$  est l'espérance de la varible aléatoire X en (TCL).

**Proposition 1.2.1.** Les trois distributions des extrêmes Précédent de Fréchet,de weibull,de Gumbel ont les fonctions de densités suivant :

$$(\text{Frchet}) \quad \Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \alpha x^{-(\alpha+1)} \exp(-x^{-\alpha}) & \text{si } x > 0 (\alpha \geq 0) \end{cases}$$

$$\begin{aligned}
 (\textit{Weibull}) \quad \Psi_\alpha(x) &= \begin{cases} \alpha(-x)^{\alpha-1} \exp(-(-x)^\alpha) & \text{si } x \leq 0 \\ 0 & \text{si } x > 0 \quad (\alpha > 0) \end{cases} \\
 (\textit{Gumbel}) \quad \Lambda(x) &= \exp[-(x + \exp(-x))], x \in \mathbf{R}
 \end{aligned}$$

**Proposition 1.2.2.** La relation existe entre les trois distributions des valeurs extrêmes est définies comme suite :

$$X \sim \text{Fréchet} \implies Y = \ln X \sim \text{Gumbel}.$$

$$X \sim \text{Weibull} \implies Y = \frac{1}{X} \sim \text{Fréchet}.$$

$$X \sim \text{Weibull} \implies Y = \ln\left(\frac{1}{X}\right) \sim \text{Gumbel}.$$

**Remarque 1.2.1.** On appelle les trois distributions  $\phi$ ,  $\lambda$  et  $\psi$  les distributions de valeurs extrêmes, et les v.a correspondantes sont appelées les variables aléatoires extrémales.

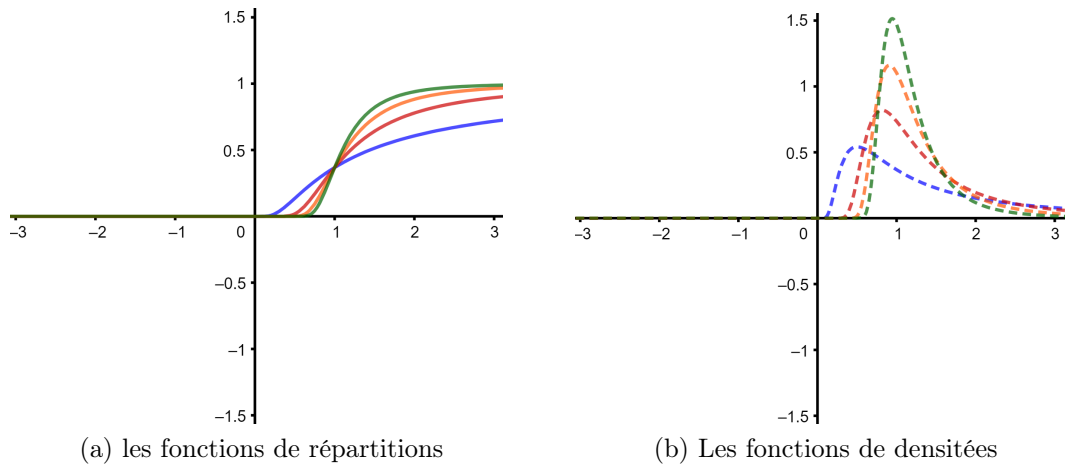


FIGURE 1.3 – Représentation de la distribution de Fréchet  $\psi_\alpha(x)$  et sa densité pour différentes valeurs de  $\alpha$ , le bleu  $\alpha = 1$ , le rouge  $\alpha = 2$ , l'orange  $\alpha = 3$ , le vert  $\alpha = 4$

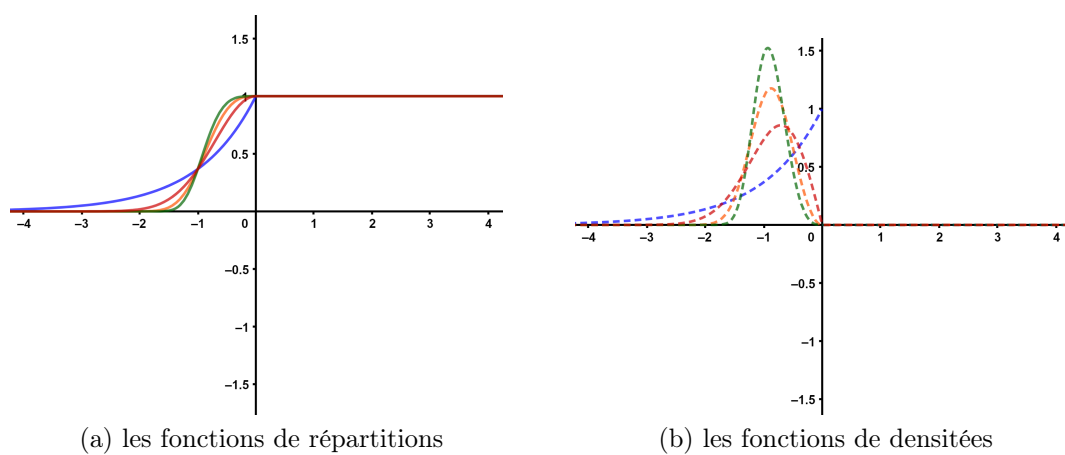


FIGURE 1.4 – Représentation de la distributions de weibull  $\psi_\alpha(x)$  et sa densité pour différentes valeurs de  $\alpha$ , le bleu  $\alpha = 1$ , le rouge  $\alpha = 2$ , l'orange  $\alpha = 3$ , le vert  $\alpha = 4$

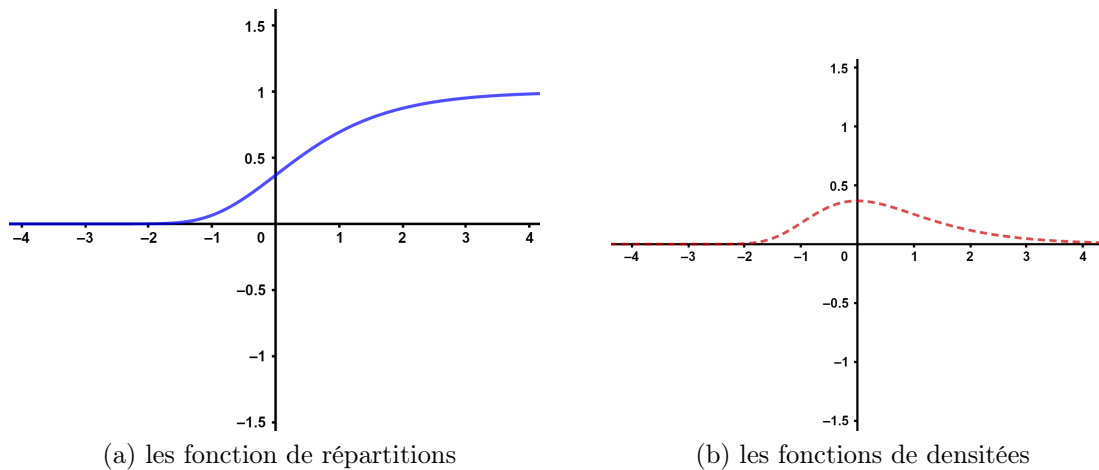


FIGURE 1.5 – Représentation de la distribution de Gumbell  $\Lambda(x)$  et sa densité

### 1.3 Distribution des valeurs extrêmes généralisées

Il était habituel d'adopter l'une des trois types qu'on a vu mais avec le temps cela présente une faiblesse du technique pour choisir laquelle des trois types est la plus appropriée, c'est pour cela Jenkinson-Von Mis en 1955 fournit une caractérisation synthétique des distributions précédentes des extrêmes (Fréchet, Weibull et Gumbel) contenue un unique paramètre  $\gamma$  s'appelle la distribution généralisée des valeurs extrêmes.

**Théorème 1.3.1.** On appelle distribution des valeurs extrêmes généralisées toute fonction de répartition  $\mathcal{H}_\gamma$  non-dégénérée telle que pour tout  $\gamma \in \mathbb{R}$  et  $1 + \gamma x > 0$  On a :

$$\mathcal{H}_\gamma(x) = \begin{cases} \exp \left\{ -(1 + \gamma x)^{-\frac{1}{\gamma}} \right\} & \text{si } \gamma \neq 0 \\ \exp \{ -\exp(-x) \} & \text{si } \gamma = 0 \end{cases} \quad (1.5)$$

On appelle  $\gamma$  l'indice de queue, ou en anglais "extrême values index", on peut écrire cette formule de répartition sous une forme plus générale suivant :

$$H_{\mu, \sigma, \gamma}(x) = \begin{cases} \exp \left[ - \left( 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right)^{-\frac{1}{\gamma}} \right] & \text{si } \gamma \neq 0 \\ \exp \left[ - \exp \left( - \left( \frac{x - \mu}{\sigma} \right) \right) \right] & \text{si } \gamma = 0 \end{cases} \quad (1.6)$$

Où  $\mu$  est un paramètre de localisation,  $\sigma$  est celui de dispersion,  $\gamma$  c'est toujours le paramètre de queue

**Proposition 1.3.1.** On définit la fonction de densité de la fonction de distribution généralisé  $\mathcal{H}_\gamma$  par :

$$h_{\mu,\sigma,\gamma}(x) := \begin{cases} \frac{1}{\sigma} \left[ 1 + \gamma \left( \frac{x - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\gamma}{\gamma}\right)} H_{\mu,\sigma,\gamma}(x) & \text{si } \gamma \neq 0 \\ \frac{1}{\sigma} \exp \left\{ - \left( \frac{x - \mu}{\sigma} \right) - \exp \left[ - \left( \frac{x - \mu}{\sigma} \right) \right] \right\} & \text{si } \gamma = 0 \end{cases} \quad (1.7)$$

Cette loi des valeurs extrêmes généralise implique que le comportement de la queue des trois distributions données par Fisher-Tippett est caractérisée selon le signe de paramètre  $\gamma$  tel que :

valeur d'indice $\gamma$	Domaine d'attraction
$\gamma < 0$	Weibull
$\gamma = 0$	Gumbel
$\gamma > 0$	Fréchet

TABLE 1.1 – Les trois distributions de GEV selon l'indice  $\gamma$

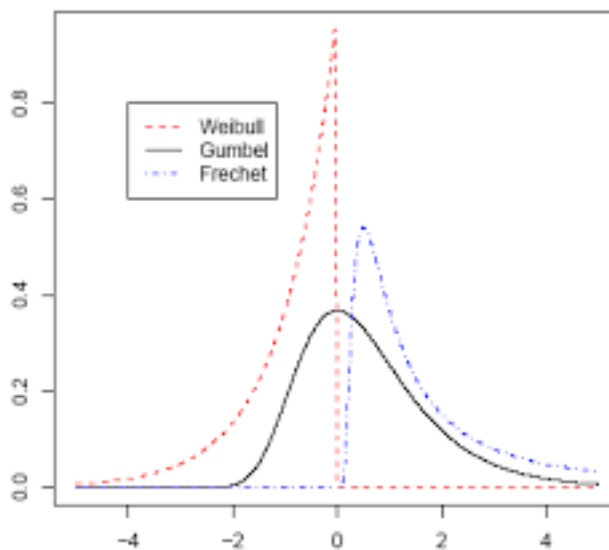


FIGURE 1.6 – Représentation de les trois distributions de GEV

**Remarque 1.3.1.** On obtient ces distributions selon le poids de queues tel que on obtient une loi de Weibull (absence de queue) , queue en loi de Gumbel est finie et queue épaisse en loi de Fréchet ,on représente ces queues de distributions ci-dessous.

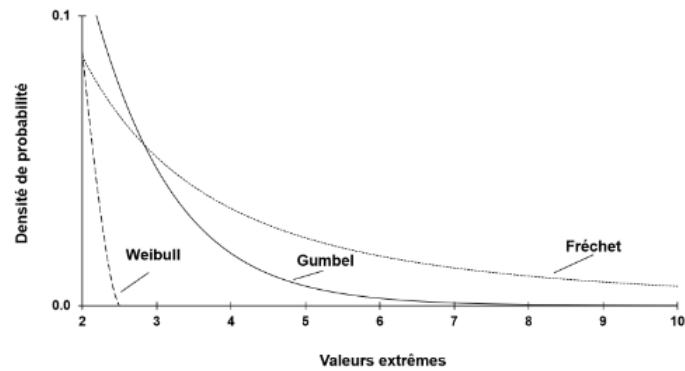


FIGURE 1.7 – Queues des trois lois extrêmes :Weibull,Gumbel,Fréchet

### 1.3.1 La méthode de bloc maxima (BM)

La méthode de Maxima par Bloc c'est une manière traditionnelle de modéliser les évènements extrêmes est utilisée largement dans l'hydrologie et d'autres applications d'ingénierie,le principe de cette méthode c'est divisé l'horizon temporel en blocs ou périodes et considère le maximum de la variable prend en périodes successives, par exemple des mois ou des années,ces observations sélectionnées constituent les évènements extrêmes, galement appelés maxima de bloc (ou par période)

Pratiquement si on supposons  $x_1, x_2, \dots; xn$  n observations, on les regrouper les données en Blocs  $K$  de longueur 1 et on calcule le maximum sur chaque bloc :

$$m_i = \max (x_{(i-1)l+1}, \dots, x_{il}) \text{ pour } i \in 1, \dots, k$$

On approche ensuite la loi de c'est variable aléatoire Mi par une loi GEV qu'on a donnée sa formule au théorème [1.5] ,donc pour que l'approximation par ce loi soit réaliste il faut existe un bon compromise entre la taille des blocs qui doit être assez grande pour bien avoir assez d'informations pour l'estimation des paramètre de (GEV).

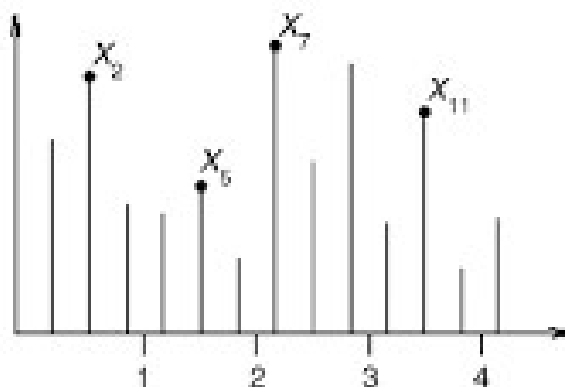


FIGURE 1.8 – Représentation de la méthode de Block Maxima(BM)

**Remarque 1.3.2.** La figure (1.7) illustre l'approche par une division tel que  $(X_2, X_5, X_7; X_{11})$  représentent les maxima de bloc pour quatre périodes.

Cette méthode n'est pas particulièrement adaptée aux séries chronologiques financières en raison de la volatilité regroupement et puisqu'elle ne prend en compte que le rendement maximum de chaque période malgré que on peut observée plusieurs observations extrêmes au sein du même bloc, mais seule la plus grande d'entre elles sera prise en compte donc un grand nombre de les points de données pertinents sont exclus de l'étude.

## 1.4 Distribution de Pareto généralisée(GPD)

La distribution de Pareto généralisée (GPD) est largement utilisée pour modéliser des valeurs extrêmes, par exemple des dépassements de seuil a été appliquée à la classe de distributions de Pickands.

**Définition 1.4.1.** une distribution  $H_{\gamma,\sigma(u)}$  est dite de Pareto généralisés de paramètre  $\gamma \in \mathbb{R}$  et  $\sigma(u) > 0$  si elle s'écrit comme suivant :

$$H_{\gamma,\sigma(u)}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma}{\sigma(u)}x\right)^{-1/\gamma}, & \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma(u)}\right), & \gamma = 0 \end{cases} \quad (1.8)$$

Cette distribution est définie pour :

$$\begin{cases} x > 0 & \text{si } \gamma \geq 0 \\ 0 \leq x \leq \frac{-\sigma(u)}{\xi} & \text{si } \gamma < 0 \end{cases}$$

**Proposition 1.4.1.** La densité de  $H_{\gamma,\sigma(u)}$  Loi de Pareto *GPD* est donné par :

$$\forall x > 0, \quad h(x) = \begin{cases} \frac{1}{\sigma(u)} \left[1 + \gamma \left(\frac{x}{\sigma(u)}\right)\right]^{-\frac{1}{\gamma}-1} & \text{si } \gamma \neq 0 \\ \frac{1}{\sigma(u)} e^{-\frac{x}{\sigma(u)}} & \text{si } \gamma = 0 \end{cases} \quad (1.9)$$

tel que  $\gamma$  est le paramètre de forme de queues et  $\sigma(u)$  c'est le paramètre d'échelle, Selon les valeurs du paramètre de forme  $\gamma$ , la GPD regroupe les trois distributions suivantes :

Le signe de $\gamma$	La distribution
$\gamma < 0$	la loi de Pareto usuelle,
$\gamma = 0$	la loi de Pareto de type II
$\gamma > 0$	la loi exponentielle de paramètre $\sigma$

TABLE 1.2 – Les trois distribution de GEP selon l'indice  $\gamma$

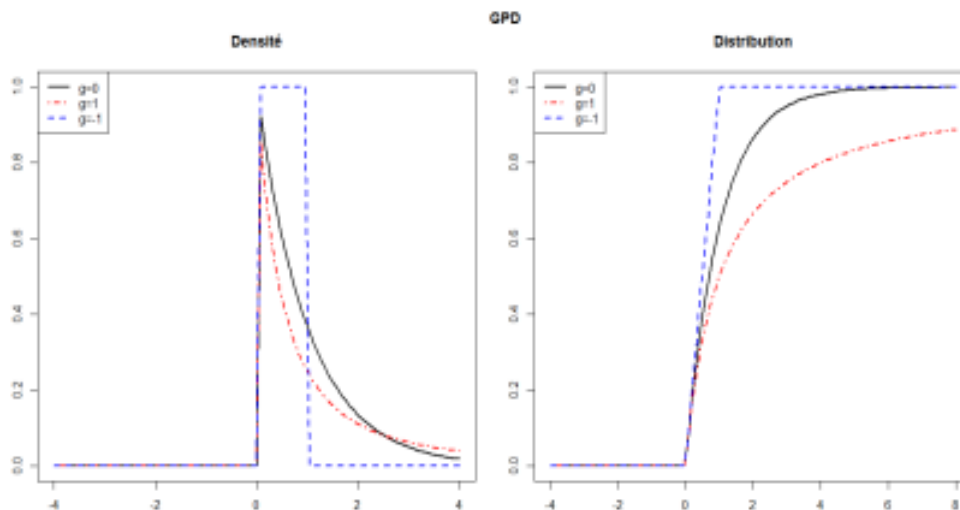


FIGURE 1.9 – Densités et distributions de loi Pareto Généralisées avec différentes valeurs du paramètre  $\gamma$ .

### 1.4.1 La méthode de Peak over threshold(POT)

La méthode Peak over threshold est un moyen qui permet la modélisation de la distribution queues d'un ensemble de données ,a été développée initialement par Pickands (1975),le concept principal de cette méthode de Peak over threshold est sélectionne celles des observations initiales qui dépasser un certain seuil  $u$ .

Pratiquement on concéderont  $X_1, X_2, ..X_n$  une suite des variables aléatoire iid de Loi F la et de point terminal  $x_F$ ,On définie un seuil  $u$  telque  $u < x_F$  ,on définie les excès de la v.a X au-dessus comme suite :  $Y_i = X_i - u > 0$  quand  $x > u$

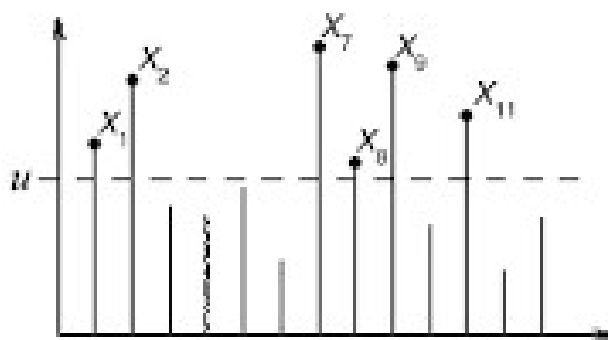


FIGURE 1.10 – Représentation de la méthode de Peak over threshold(POT)

**Définition 1.4.2. (Distribution des excès)** La fonction de répartition des excès de  $X$  au-dessus du seuil  $u$  est définie par :

$$F_u(y) := P(X - u \leq y \mid X > u) \quad (1.10)$$

De plus

$$\begin{aligned} F_u(y) &= \frac{P(X - u \leq y, X > u)}{P(X > u)} = \frac{P(u < X \leq y + u)}{1 - P(X \leq u)} \\ &= \frac{P(X \leq y + u) - P(X \leq u)}{1 - P(X \leq u)} = \frac{F(y + u) - F(u)}{1 - F(u)} \\ &= \frac{F(y + u) - F(u)}{\bar{F}(u)} = 1 - \frac{\bar{F}(u + y)}{\bar{F}(u)}, 0 < y < x_F - u, \end{aligned}$$

Ainsi, pour  $u$  grand, la loi des excès est approchée par une loi Pareto généralisée.

$$F_{ut} \approx H_{\gamma, \sigma(u)}(x) \quad (1.11)$$

Nous intéressons dans cette méthode POT à la fonction de distribution cumulative fonction de répartition  $F_u(x)$  des dépassements de  $X$  sur un seuil haut  $u$ , qui est définie comme suit :

$$F_u(x) \approx H_{\gamma, \sigma(u)}(x), u \rightarrow \infty$$

où la loi de Pareto généralisée est définie par :

$$H_{\gamma, \sigma(u)}(x) = \begin{cases} 1 - \left(1 + \frac{\gamma}{\sigma(u)}x\right)^{-1/\gamma}, & \gamma \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma(u)}\right), & \gamma = 0 \end{cases} \quad (1.12)$$

Le théorème de Pickands-Balkema-de Haan ci-après donne la forme de la loi limite pour les valeurs extrêmes sous certaines conditions de convergence, la loi limite est une loi de Pareto Généralisé on va l'énoncé si-dessus

**Théorème 1.4.1. (Théorème de Pickands-Balkema-de Hann)** Une fonction de répartition  $F$  appartient au domaine d'attraction maximale de  $G_\gamma$  si et seulement si, il existe une fonction positive  $\sigma(u)$  telle que :

$$\lim_{u \uparrow x_F} \sup_{0 \leq x \leq x_F - u} |F_u(x) - H_{\gamma, \sigma(u)}(x)| = 0 \quad (1.13)$$

où  $F_u$  est la fonction de répartition des excès au delà du seuil  $u$ ,  $x_F$  est le point terminal et  $H_{\gamma, \sigma(u)}(x)$  est la fonction de répartition de la loi de Paréto Généralisée tel que Pickands a montré que pour n'importe quelle loi  $F$ , l'approximation GPD définie par (1.8) n'est vérifiée que s'il existe des constantes de normalisation et une loi non dégénérée. Donc il montre l'existence d'une relation étroite entre la GPD et La GEV.

**Exemple : 1.3.** Soit  $f(x) = 1 - cx^\alpha$  on  $c > 0$  et  $\alpha > 0$  est la fonction de répartition de la loi de Pareto on posant :  $a_n = (nc)^{1/\alpha}, b_n = 0$  ; alors on a pour  $x > 0$  :

$$\begin{aligned} F^n(a_n x + b_n) &= (1 - c(a_n x)^{-\alpha})^n = (1 - ca_n^{-\alpha} x^{-\alpha})^n \\ &\rightarrow \exp(-x^{-\alpha}) \\ &\rightarrow \Phi_a(x) \end{aligned}$$

Qui est la loi de Fréchet donc la loi de Pareto appartient au domaine d'attraction maximal de Fréchet, en se basant sur la méthode des excès au-delà d'un seuil  $u$  on considérant  $\sigma(u) = u\sigma$  pour  $\geq 0$ , on a alors :

$$\begin{aligned} F_u(y) &= \frac{F(u + u\sigma y) - F(u)}{1 - F(u)}, \quad y \geq 0 \\ &= \frac{cu^{-\alpha} - c(u + u\sigma y)^{-\alpha}}{cu^{-\alpha}} \\ &= 1 - (1 + \sigma y)^{-\alpha} \end{aligned}$$

Pour  $\gamma = 1/\alpha'$  et  $b\gamma$  la limite est alors la loi GPD de paramètre  $\gamma$

## 1.5 Domaines d'attraction

Dans cette section on va intéresser aux fonctions à variation régulière lente, puis régulière puis on va donner la représentation de Karamata, parce que ce type de fonction permettra de soutenir les diverses caractérisations des domaines d'attraction que nous allons présenter dans cette partie.

### 1.5.1 Fonction à variation régulière

La fonction à variation régulière est le prolongement d'une fonction à variation lente.

**Définition 1.5.1.** soit  $G$  une fonction mesurable, on dit que  $G$  est **a variation régulière** d'indice  $\rho \in \mathbb{R}$  si  $G$  est positive à l'infinie i.e s'il existe  $\kappa$  tel que pour tout  $x \geq \kappa, G(x) > 0$  et si pour tout  $t > 0$  :

$$\lim_{x \rightarrow \infty} \frac{G(tx)}{G(x)} = t^\rho \tag{1.14}$$

**Lemme 1.5.1.** si  $G$  est une fonction a variation réguliere d'indice  $\rho$  alors pour tout  $0 < a < b$  :

$$\sup_{t \in [a, b]} \frac{G(tx)}{G(x)} - t^\rho \rightarrow 0 \text{ lorsque } x \rightarrow \infty$$

**Définition 1.5.2.** Dans le cas particulier de (1.14) quand  $\rho = 0, \lim_{x \rightarrow \infty} \frac{G(tx)}{G(x)} = 1$  on dit que  $G$  est **a variation régulière lente** et on note les fonction a variation régulière lente

par :  $l(x)$

**Proposition 1.5.1.** soit  $\rho \in \mathbb{R}$  et  $g \in \mathbb{R}v_\rho$  alors il existe une fonction  $q$  variation régulier lente a l'infinie telle que :

$$\forall x > 0, g(x) = x^\rho(x)$$

alors l'étude des fonction a variation régulière a l'infinie se ramenée a celle des fonction a variation lente.

**Lemme 1.5.2.** si  $G$  est une fonction a variation régulière d'indice  $\rho > 0$ , donc il existe une fonction  $H$  a variation régulière d'indice  $\frac{1}{\rho}$  telle que :

$$\lim_{x \rightarrow \infty} \frac{H[G(x)]}{x} = 1$$

**Lemme 1.5.3.** si  $G$  est a variation régulière d'indice  $\rho > 0$  a l'infinie , alors  $g \leftarrow (x)$  est a variation régulière d'indice  $\frac{1}{\rho > 0}$  à l'infinie.

si  $G$  est a variation régulière d'indice  $\rho < 0$  a l'infinie , alors  $g \leftarrow (x)$  est a variation régulière d'indice  $\frac{-1}{\rho > 0}$  à l'infinie.

## 1.5.2 Représentation de Karamata

De nombreux résultats concernant les fonctions à variation lente sont dus à Jovan.Karamata tels que l'un des principaux résultats est le théorème de représentation suivant :

**Théorème 1.5.1.**  $l$  est une fonction à variation lente si et seulement si pour tout  $x > 0$  :

$$l(x) = c(x) \exp \left\{ \int_1^x t^{-1} \varepsilon(t) dt \right\}$$

ou  $c$  et  $\varepsilon$  sont des fonction positive telle que :

$$\lim_{x \rightarrow \infty} c(x) = c \in ]0, +\infty[ \text{ et } \lim_{t \rightarrow \infty} \varepsilon(x) = 0$$

Après avoir introduit les fonctions qui varient régulièrement, maintenant présentons les caractérisation du domaines d'attraction de la loi  $F$  et ses constantes de normalisation.

**Définition 1.5.3.** On dit qu'une distribution  $F$  appartient au domaine d'attraction du maximum de la distribution  $\mathcal{H}$ , et on note  $F \in \mathcal{D}(\mathcal{H}_\gamma)$  s'il existe de suites normalisantes  $(a_n) \geq 0$  et  $(b_n) \in \mathbb{R}$  tel que la condition soit vérifiée :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{M_n - b_n}{a_n} \leq x \right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \mathcal{H}_\gamma(x), \forall x \in \mathbb{R} \quad (1.15)$$

On définit trois domaine d'attraction selon le signe de seule paramètre  $\gamma$  qui appelé l'indice de valeurs extrêmes tel que nous avons :

Domaine d'attraction de Fréchet lorsque  $\gamma > 0$ .

Domaine d'attraction de weibull lorsque  $\gamma < 0$

Domaine d'attraction de Gumbel lorsque  $\gamma = 0$ .

Ce paramètre a aussi contrôle la forme de la queue de distribution tel que :

1. Si  $\gamma > 0$ , au domaine d'attraction de Fréchet la fonction de survie  $1 - F(x)$  converge vers 0 lorsque  $x \rightarrow x_F^* = +\infty$  a une vitesse polynomiale (proportionnelle a une puissance de x), Ce domaine regroupe les distributions à queue lourde.

**Exemple : 1.4. Distribution de Pareto :**

$$\bar{F}(x) = x^{-\alpha}, x > 1; \alpha > 0 \text{ tel que } \gamma = \frac{1}{\alpha}$$

**Distribution de Fréchet :**

$$\bar{F}(x) = 1 - \exp(-x^{-\alpha}); \alpha > 0 \text{ tel que } \gamma = \frac{1}{\alpha}$$

**Distribution de Burr :**

$$\bar{F}(x) = \left( \frac{\beta}{\beta + x^\tau} \right)^\lambda, (\beta > 0, \tau > 0, \lambda > 0) \text{ tel que } \gamma = \frac{1}{\lambda\tau}$$

2. Si  $\gamma = 0$ , au domaine d'attraction de Gumbel la fonction de survie  $1 - F(x)$  converge vers 0 lorsque  $x \rightarrow x_F^*$  a une vitesse exponentielle, on parle ici de distributions à queue légère.

**Exemple : 1.5. Distribution de Gamma :**

$$\bar{F}(x) = \frac{\lambda^m}{\Gamma(m)} \int_x^\infty u^{-m-1} \exp(-\lambda u) du, \lambda > 0, m \in \mathbb{R}$$

tel que  $\gamma = 0$

**Distribution de Gumbel :**

$$\bar{F}(x) = \exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right), \beta > 0, \mu \in \mathbb{R} \text{ tel que } \gamma = 0$$

3. Si  $\gamma < 0$ , au domaine d'attraction de weibull la fonction  $1 - F(x)$  converge vers 0 lorsque  $x \rightarrow x_F^* < +\infty$  a une vitesse polynomiale (proportionnelle à une puissance de x), ce domaine regroupe la plupart des distributions ayant un point terminal fini.

**Exemple : 1.6. Distribution Burr inversée :**

$$\bar{F}(x) = \left( \frac{\beta}{\beta + (x_F + x)^{-\lambda}} \right), \beta > 0, \tau > 0, \lambda > 0 \text{ tel que } \gamma = -1/\lambda$$

### 1.5.3 Domaine d'attraction de Fréchet :

Ce domaine d'attraction de Fréchet ( $\gamma > 0$ ) n'admet que les lois aux queues lourdes, ce qui en fait un domaine très homogène. Il est notamment utilisé dans des applications telles que l'hydrologie, la météorologie.. etc.

**Théorème 1.5.2.** La fonction de répartition  $F \in \mathcal{D}(\Phi_\alpha)$  domaine d'attraction de Fréchet si et seulement si  $x_F = +\infty$  et  $1 - F(x)$  est une fonction à variation régulière d'indice  $-1/\gamma$  i.e  $1 - F(x) = x^{-1/\gamma}l(x)$  ou  $l(x)$  est une La fonction à variation lente. dans ce cas le choix possible pour les suite normalisé  $(a_n)$  et  $(b_n)$  est :

$$a_n = F^{\leftarrow} \left( 1 - \frac{1}{n} \right) \text{ et } b_n = 0$$

**Exemple : 1.7.** On sait que la fonction de répartition de loi de Paréto est :

$$F(x) = 1 - cx^{-\alpha} \text{ avec } : \alpha > 0, c > 0$$

on pose :  $a_n = \left( \frac{1}{nc} \right)^{-1/\alpha} b_n = 0$  Donc :

$$\begin{aligned} \lim_{n \rightarrow \infty} P \left( \frac{M_n - b_n}{a_n} \right) &= \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \lim_{n \rightarrow \infty} \left( 1 - c \left( \frac{1}{nc} \right)^{-1/\alpha} \right)^{-\alpha n} \\ &= \lim_{n \rightarrow \infty} \left( 1 - c \left( \frac{x^{-\alpha}}{nc} \right) \right)^n = \lim_{n \rightarrow \infty} \left( 1 - \frac{x^{-\alpha}}{n} \right)^n \rightarrow \exp(-x^{-\alpha}) = \Phi_\alpha(x) \end{aligned}$$

#### 1.5.4 Domaine d'attraction de weibull :

**Théorème 1.5.3.** La fonction de répartition  $F \in \mathcal{D}(\psi_\alpha)$  domaine d'attraction de weibull si et seulement si  $x_F < \infty$  et  $1 - F(x) = (x_F - x)^{-1/\gamma}l[(x_F - x)^{-1}]$  ou  $l$  est une La fonction à variation lente à l'infinie et  $\gamma$  un réel strictement négatif. Dans ce cas le choix possible pour les suite normalisée  $(a_n)$  et  $(b_n)$  est :

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

#### 1.5.5 Domaine d'attraction de Gumbel :

Ce domaine est correspond au cas de  $(\gamma = 0)$ , mais contrairement aux deux domaines précédents, ce domaine n'a pas une représentation simple donc est plus complexes à étudier dans le sens où il comporte une très grande variété de loi.

**Théorème 1.5.4.** La fonction de répartition  $F \in \mathcal{D}(\Lambda)$  domaine d'attraction de Gumbel si et seulement si : il existe une fonction de **Von-Mises**  $F^*$  tel que pour tout  $z < x < x_F$  :

$$1 - F(x) = c(x)[1 - F^*] = c(x) \exp \left\{ - \int_z^x \frac{1}{a(t)} dt \right\} \quad (1.16)$$

ou

$$\lim_{x \rightarrow x_F} c(x) = 0$$

**Exemple : 1.8.** Supposons que  $X$  suite la loi exponentielle de paramètre  $\lambda > 0$  alors  $F(x) = 1 - \exp^{-\lambda x}$  pour  $x \geq 0$ , en posant  $a_n = \lambda^{-1}$  et  $b_n = \lambda^{-1} \log(n)$  on a que :

$$\begin{aligned} \lim_{n \rightarrow \infty} F^n(a_n x + b_n) &= \lim_{n \rightarrow \infty} \left( 1 - \frac{e^{-\lambda a_n x}}{e^{\lambda b_n}} \right)^n \\ &= \lim_{n \rightarrow \infty} \left( 1 - \frac{e^{-x}}{n} \right)^n = \exp(-e^{-x}) \end{aligned}$$

Donc la loi exponentielle appartient au domaine d'attraction de Gumbel.

Le Domaine d'attraction	Gumbel $\gamma = 0$	Fréchet $\gamma > 0$	Weibull $\gamma = 0$
Les loi	Normale ; loi Exponentielle , Lognormale, Gamma, Weibull	Pareto, student, Burr	loi uniforme , Beta

TABLE 1.3 – Domaines d'attraction de quelque lois usuelles

Après avoir répertorié les diverses caractérisations des domaines d'attraction, nous sommes prêts à aborder la notion d'estimation des quantiles extrêmes, voilà c'est l'objet notre chapitre suivante.

# Estimation d'indice des valeurs extrêmes (IVE)

L'un des problèmes les plus importants de la théorie des Valeurs extrêmes est l'estimation d'indice des valeurs extrêmes, en effet il existe plusieurs méthodes pour l'estimation de IVE mais nous intéressons ici à la méthode paramétrique de maximum vraisemblance et semi paramétrique : les deux estimateurs les plus utilisées en pratique qui sont l'estimateur de Pickands et l'estimateur de Hill [1975] puis on va voir l'estimation de quantiles extrêmes par l'estimateur de Weissman dans le cas particulier de lois à queues lourdes exactement nous voulons dire les distributions qui appartenant au domaine d'attraction de Fréchet car ces distributions ont un rôle très important dans la théorie des valeurs extrêmes tel que elles ont été acceptées comme des modèles appropriés de divers phénomènes.

## 2.1 Estimation paramétrique de (IVE)

Dans cette section, nous présenterons la technique paramétrique la plus populaire pour estimer l'indice des valeurs extrêmes ce qui est la méthode paramétrique du Maximum de Vraisemblance.

### 2.1.1 Estimateur de maximum vraisemblance (EMV)

La méthode de maximum vraisemblance donne des résultats efficaces en statistiques, elle est donc considérée comme la méthode paramétrique la plus largement utilisée.

**Définition 2.1.1.** Soit  $(X_1, \dots, X_n)$   $n$ -échantillon, les  $X_i$  sont supposées indépendantes et identiquement distribuées, de densité  $h_\theta$  où  $\theta = (\mu, \sigma, \gamma)$ . L'expression de la fonction de vraisemblance est donnée par :

$$L(\theta = (\mu, \sigma, \gamma); (X_1, \dots, X_n)) = \prod_{i=1}^n h_\theta(X_i)$$

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

$$\begin{cases} \frac{\partial \log L}{\partial \theta} = 0 \\ \frac{\partial^2 \log L}{\partial^2 \theta} < 0 \end{cases}$$

**Exemple : 2.1.** Nous avons vu que la densité de  $H_{\gamma, \sigma(u)}$  Loi de Pareto *GPD* est donné par :

$$\forall x > 0, \quad h(x) = \begin{cases} \frac{1}{\sigma(u)} \left[ 1 + \gamma \left( \frac{x}{\sigma(u)} \right) \right]^{-\frac{1}{\gamma}-1} & \text{si } \gamma \neq 0 \\ \frac{1}{\sigma(u)} e^{-\frac{x}{\sigma(u)}} & \text{si } \gamma = 0 \end{cases} \quad (2.1)$$

Le logarithme de la fonction de vraisemblance que nous maximisons est de la forme :

$$\log l(\gamma, \sigma(u); X_1, \dots, X_n) = -n \log \sigma(u) - \left( 1 + \frac{1}{\gamma} \right) \sum_{i=1}^n \log \left( 1 + \frac{\gamma}{\sigma(u)} X_i \right)$$

Poussons  $\tau = \frac{\gamma}{\sigma(u)}$  l'annulation des drivées partielles des logarithmes de la fonction de vraisemblance conduit au système :

$$\begin{cases} \hat{\gamma} = \frac{1}{n} \sum_{i=1}^n \log (1 + \tau X_i) = \hat{\gamma}(\tau) \\ \frac{1}{\tau} = \frac{1}{n} \left( 1 + \frac{1}{\hat{\gamma}} \right) \sum_{i=1}^n \frac{X_i}{1 + \tau X_i} \end{cases}$$

L'estimateur du maximum de vraisemblance de  $(\gamma, \tau)$  est  $(\hat{\xi} = \hat{\gamma}(\hat{\gamma}), \hat{\tau})$ , où  $\hat{\tau}$  est solution :

$$\frac{1}{\tau} = \frac{1}{n} \left( 1 + \frac{1}{\hat{\gamma}} \right) \sum_{i=1}^n \frac{X_i}{1 + \tau X_i}$$

## 2.2 Estimation semi paramétrique de (IVE)

la théorie des valeurs extrêmes à trouve plusieurs techniques semi paramétriques pour l'estimation de l'indice des valeurs extrêmes  $\gamma$ , nous intéressons dans cette section à exposerons les estimateurs pour  $\gamma \in \mathbb{R}$  qui sont l'estimateur de pickand(1975) et l'estimateurs de moment.

### 2.2.1 Estimateur de Pickands(1975)

L'estimateur de Pickands a été introduit par Pickands (1975),est le premier estimateur suggère pour le paramètre  $\gamma \in \mathbb{R}$  nous présenterons sa formule et ses caractéristiques ci-dessous.

**Définition 2.2.1.** On suppose que  $\{X_i, i = 1, \dots, n\}$  est une suite de variables aléatoires indépendantes de loi  $F$  appartenant à l'un des domaines d'attractions. Soit  $(k_n)_{n \geq 1}$  une

suite d'entiers avec  $1 \leq k_n < n$ , l'estimateur de Pickands est défini par

$$\hat{\gamma}_{k_n}^P = \frac{1}{\log 2} \log \left( \frac{X_{n-k_n+1,n} - X_{n-2k_n+1,n}}{X_{n-2k_n+1,n} - X_{n-4k_n+1,n}} \right).$$

**Proposition 2.2.1. (Propriétés asymptotiques de l'estimateur de Pickands)**

Soit  $X_1, X_2, \dots, X_n$  une suite des variables aléatoires i.i.d de fonction de répartition  $F \in D(H_\gamma)$ , où  $\gamma \in \mathbb{R}$ . si  $k = k(n) \rightarrow \infty$  et  $\frac{k(n)}{n} \rightarrow 0$  quand  $n \rightarrow \infty$ , alors :

(a)**La Consistance Faible** :La cosistance faible a été démontrée par Pikinds en 1975 :

$$\hat{\gamma}_{k_n}^P \text{ converge en probabilité vers } \gamma$$

(b)**La Consistance Forte** : a été démontrées par Dekkers et de Haan en 1986.Si  $k_n / \log \log n \rightarrow \infty$  quand  $n \rightarrow \infty$ , alors :

$$\hat{\gamma}_{k_n}^P \text{ converge presque sûrement vers } \gamma.$$

(c)**Normalité asymptotique** : Sous des conditions additionnelles sur la suite  $k_n$  et la fonction de répartition  $F$  que l'on pourra consulter dans Dekkers et de Haan (1989)

$$\sqrt{k_n} (\hat{\gamma}_{k_n}^P - \gamma) \xrightarrow{\mathcal{H}} \mathcal{N} \left( 0, \frac{\gamma^2 (2^{2\gamma+1} + 1)}{4(\log 2)^2 (2^\gamma - 1)^2} \right)$$

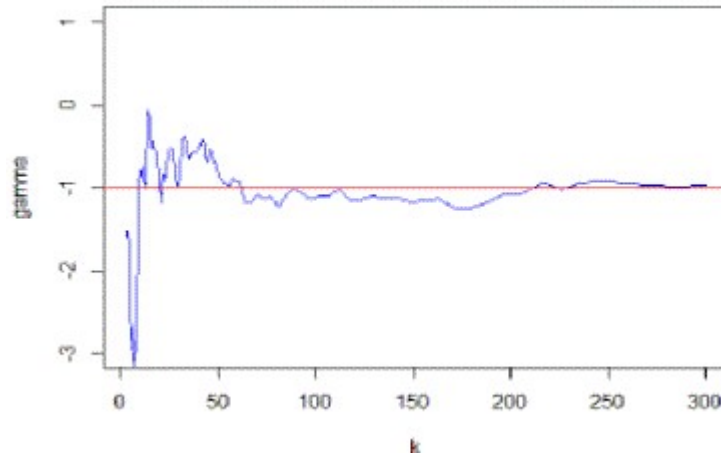


FIGURE 2.1 – Représentation de l'estimateur de Pickands

### 2.2.2 Estimateur de moments

Cette estimateur de moment est considéré comme une adaptation de l'estimateur de Hill, dans cette partie nous allons donner sa définition et présenter ses propriétés asymptotiques.

**Définition 2.2.2. (Estimateur des Moments)** Pour  $\gamma \in \mathbb{R}$ , l'estimateur de moment est défini par :

$$\widehat{\gamma}_n^{(M)} := M_1 + 1 - \frac{1}{2} \left( 1 - \frac{(M_n^{(1)})^2}{M_n^{(2)}} \right)^{-1}$$

Avec :

$$M_n^{(r)} = \frac{1}{k} \sum_{i=0}^k (\log X_{n-i+1:n} - \log X_{n-k=n})^r ; \quad r = 1, 2$$

**Théorème 2.2.1. (Propriétés asymptotiques de  $\widehat{\gamma}_n^{(M)}$ )**

Pour  $\gamma \in \mathbb{R}$ ,  $F \in \mathcal{DA}(\Phi_{1/\gamma})$  et  $k$  que vérifie la condition (2.5) on a :

•(a) Consistance faible :

$$\widehat{\gamma}_n^{(M)} \xrightarrow{P} \gamma \quad \text{quasud } n \rightarrow \infty$$

•(b) Consistance forte : Si  $k/(\log(n))^\delta \rightarrow \infty$  quand  $n \rightarrow \infty$  avec  $\delta > 0$  alors

$$\widehat{\gamma}_n^{(M)} \xrightarrow{p} \gamma \quad \text{quand } n \rightarrow \infty$$

•(c) Normalité asymptotique :

$$\sqrt{k} (\widehat{\gamma}_n^{(M)} - \gamma) \xrightarrow{D} \mathcal{N}(0, \eta^2) \quad \text{quand } n \rightarrow \infty$$

où

$$\eta^2 := \begin{cases} 1 + \gamma^2 & \text{si } \gamma \geq 0 \\ (1 - \gamma)^2(1 - 2\gamma) \left[ 4 - 8 \frac{(1 - 2\gamma)}{(1 - 3\gamma)} + \frac{(5 - 11\gamma)(1 - 2\gamma)}{(1 - 3\gamma)(1 - 4\gamma)} \right] & \text{si } \gamma < 0 \end{cases}$$

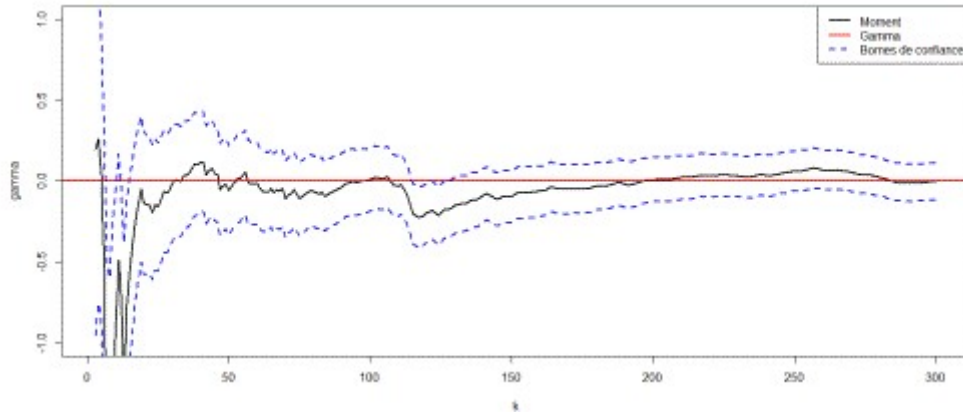


FIGURE 2.2 – Représentation de l'estimateur de moments

## 2.3 Estimation d'indice de queue lourde

Concédons  $(X_1 \dots X_n)$  un échantillon de  $n$  variables aléatoires indépendantes et identiquement distribuées de fonction de répartition  $F$ , appartenant au domaine d'attraction de Fréchet. Pour estimer la quantile extrême qui est définie dans ce cas par  $q(\alpha) \approx q(t\alpha)t^\gamma$ , il faut estimer tout d'abord l'indice de queue lourde, nous le ferons, mais d'abord nous donnerons une idée complète sur les distributions à queue lourde.

### 2.3.1 Distribution à queue lourde

Les distributions à queue lourde sont des distributions qui ont des queues non exponentiellement bornées, i.e. qui ont des queues plus lourdes que celles des distributions.

#### Définition 2.3.1. (Distribution à queue lourde)

Soit une fonction de répartition  $F$  appartenant au domaine d'attraction de Fréchet, On dit que  $F$  est une distribution à queue lourde si :

$$1 - F(x) := x^{-1/\gamma} \ell(x), x \rightarrow \infty, \ell(\cdot) \in RV_0$$

alors  $(1 - F(x))$  est une fonction à variation régulière d'indice  $-\zeta$ .

**Remarque 2.3.1.** La queue converge vers l'axe des  $x_i$  à partir d'un certain seuil à queue lourde c'est pour ça qu'on dit queue lourde.

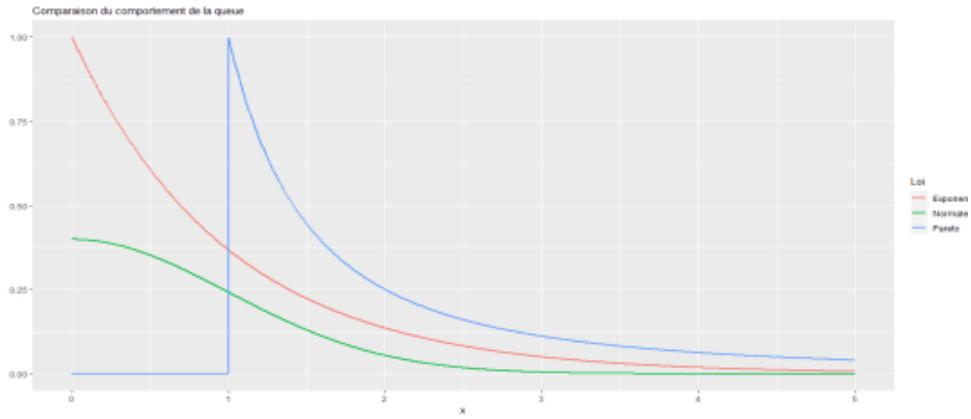


FIGURE 2.3 – Comparaison du Comportement de la Queue

**Remarque 2.3.2.** Au figure(2.3) on remarque que les distributions à queue lourdes, qui sont représentées en verte elle est décroît lentement vers le zéro par rapport la au temps que la distribution exponentielle et la distribution exponentielle que représentée en rouge et la distribution a queue légère représentée en bleu .

### 2.3.2 Estimateur de Hill(1975)

L'estimateur de Hill a été introduit par B.Hill en (1975),cette estimateur n'est utilisable que pour les distributions de Fréchet (donc telles que  $\gamma >$ ) pour lesquelles il fournit un estimateur de l'indice de queue plus efficace que 'estimateur de Pickands.on commence avec la construction de l'estimateur de Hill puis sa définition et ses Propriétés asymptotique.

#### La construction de l'estimateur de Hill(1975)

La construction de l'estimateur de Hill est basée sur la méthode du Maximum de Vraisemblance qu'on a vu, où on se sert des statistiques d'ordre supérieur à un certain seuil  $u$ , pour ne garder que les observations les plus grandes,On comonce a le construire par cette méthode

On considérons les observations  $X_1, \dots, X_n$  suivent une loi de Pareto de paramètres  $c > 0$  et  $\gamma > 0$  ont pour fonction de répartition  $c F(x) = 1 - x^{-1/\gamma}l(x)$  ou  $l$  est une fonction à variations lentes,on remarque que pour  $x$  assez grand,  $\bar{F}(x) \approx (x/c)^{-1/\gamma}$ , la fonction à variations lentes pouvant être considérée comme constante au voisinage de l'infini.

Ainsi on fait comme si la fonction à variations lentes était constante, la vraisemblance de  $(X_1, \dots, X_n)$  est alors

$$\mathcal{L}(\gamma, c) := \prod_{i=1}^n \frac{1}{\gamma c} \left( \frac{X_i}{c} \right)^{-1/\gamma-1} \mathbb{I}\{X_i \geq c\}$$

On maximise dans un premier temps cette vraisemblance par rapport à  $\gamma$ . Il suffit pour ce faire de maximiser la quantité

$$\log \left[ \prod_{i=1}^n \frac{1}{\gamma} \left( \frac{X_i}{c} \right)^{-1/\gamma} \right] = -n \log(\gamma) - \frac{1}{\gamma} \sum_{i=1}^n \log \frac{X_i}{c}$$

En dérivant par rapport à  $\gamma$  et en annulant cette dérivée, on obtient l'équation

$$-\frac{n}{\gamma} + \frac{1}{\gamma^2} \sum_{i=1}^n \log \frac{X_i}{c} = 0$$

On en déduit facilement que la valeur de  $\gamma$  qui maximise la vraisemblance est

$$\gamma = \frac{1}{n} \sum_{i=1}^n \log (X_i) - \log(c)$$

Il reste à présent à maximiser la vraisemblance par rapport à  $c$ . Pour que la vraisemblance  $\mathcal{L}(\gamma, c)$  soit non nulle, il faut que  $\min (X_1, \dots, X_n) \geq c$ . De plus, pour  $c \leq \min (X_1, \dots, X_n)$  la vraisemblance est une fonction croissante de  $c$ . Elle est donc maximale pour  $c = \hat{c}_n := \min (X_1, \dots, X_n) = X_{1,n}$ . En remplaçant dans (4.7), on obtient l'estimateur du maximum de vraisemblance de  $\gamma$  :

$$\tilde{\gamma} := \frac{1}{n} \sum_{i=1}^n \log (X_i) - \log (X_{1,n})$$

les plus grandes observations de l'échantillon  $X_1, \dots, X_n$  suivront approximativement une loi de Pareto de paramètres inconnus  $c$  et  $\gamma$ . En prenant les  $k_n$  plus grandes observations de l'échantillon on obtient  $\tilde{\gamma}$  l'expression de l'estimateur de Hill.

**Théorème 2.3.1. (Propriétés asymptotique de l'estimateur de Hill)**

Un grand nombre de travaux théoriques ont été consacrés à l'étude des propriétés de l'estimateur de Hill, tel que le premier résultat sur les propriétés asymptotiques de cet estimateur est dû à Mason en 1982 qui en a démontré la consistance faible, la consistance forte fut établie en 1988 par Deheuvels et al.

**(a) la consistance faible :**  $k = k_n \rightarrow \infty$  et  $k_n/n \rightarrow 0$  quand  $n \rightarrow \infty$

**(b) La consistance forte :**  $k/\log \log n \rightarrow \infty$  et  $k_n/n \rightarrow 0$  quand  $n \rightarrow \infty$

**(c) Normalité asymptotique :** Sous certaines conditions du second ordre, la normalité asymptotique de l'estimateur de Hill a été démontrée entre autre par Hall (1982), Davis et Resnick (1984), Haeusler et Teugels (1985), Goldie et Smith (1987) et Dekkers (1989)

$$\sqrt{k} (\hat{\gamma}_n^{(H)} - \gamma) \sim \mathcal{N} (0, \gamma^2) .$$

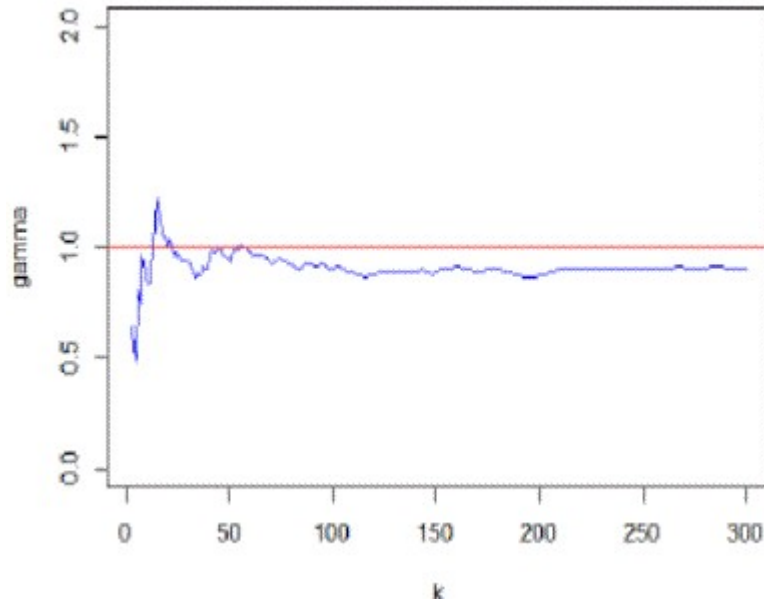


FIGURE 2.4 – Représentation de l'estimateur de Hill

### 2.3.3 Estimation des quantiles extrêmes

Dans le domaine d'attraction de Fréchet on a vu dans le théorème (1.5.2) la fonction de répartition est Donne par :

$$F(x) = x^{-1/\gamma}L(x)$$

tel que  $\gamma > 0$  et  $l(x)$  est la fonction à variation lent c-a-d  $F(x)$  est un fonction a varition régulier d'indice  $-1/\gamma$ .l'approche semi-paramétrique que nous avons entrin de faire est basée surla fonction quantile qu'est donnée par :

$$q(\alpha) = \alpha^{-\gamma}l(\alpha^{-1}) \quad (2.2)$$

où  $l$  est une autre fonction à variations lentes,Pour estimer  $q(\alpha)$  il suffit donc d'estimer l'indice  $\gamma$  et la fonction de variation lente  $l$  mais malheureusement,l'estimation de la fonction à variations lentes n'est pas facile en pratique;donc on utilise l'astuce suivante avec  $t > 0$  et  $l$  toujours une fonction à variations lentes :

$$q(t\alpha) = t^{-\gamma}\alpha^{-\gamma}l(t^{-1}\alpha^{-1}) \quad (2.3)$$

maintenant en divisant (2.2) par (2.3) on obtien :

$$\frac{q(\alpha)}{q(t\alpha)} = t^{\gamma} \frac{l(\alpha^{-1})}{l(t^{-1}\alpha^{-1})}$$

maintenant pour  $\alpha$  proche de 0, en utilisant le fait que  $l$  est une fonction à variations lentes, on obtient l'approximation comme suivante :

$$q(\alpha) \approx q(t\alpha)t^{\gamma} \quad (2.4)$$

Alors, pour estimer la quantile  $q(\alpha)$ , il suffira donc d'estimer l'indice  $\gamma$  et  $q(t\alpha)$  tel que  $t\alpha$  soit un ordre de quantile classique que l'on pourra estimer facilement par la fonction de répartition empirique main il faut le prendre suffisamment grand..

### 2.3.4 Estimateur de weissman(1978)

On rappelle que  $q(\alpha) \approx q(t\alpha)t^\gamma$  donc Pour estimer la quantile  $q(\alpha)$  Pratiquement on prend  $t = t_n = K_n/(n\alpha_n)$  tel que  $k_n$  est une suite intermédiaire.

nous avons aussi :  $q(t\alpha_n) = q(k_n/n)$  que l'on peut estimer en inversant la fonction de répartition empirique par la statistique d'ordre  $X_{n-k_n,n}$ .

l'estimaiton d'indice  $\gamma$  par un estimation quelconque  $\hat{\gamma}$  estimant  $\gamma > 0$  Ce qui nous permet d'obtenir l'estimateur de quantile  $q(\alpha_n)$  qui s'appel de weissman introduit en 1978 Que nous présenterons ci-dessous.

**Définition 2.3.2.** L'estimateur de Weissman est pour  $\gamma > 0$  et défini par :

$$\hat{q}(\alpha_n) = X_{n-k_n,n} \left( \frac{K_n}{n\alpha_n} \right)^{\hat{\gamma}_n}. \quad (2.5)$$

tel que le terme  $\left( \frac{K_n}{n\alpha_n} \right)^{\hat{\gamma}_n}$  est le terme servant.

Pour plus de détails sur les propriétés de cet estimateur de Weissman on les trouve disponibles en Haan et Ferreira [15].

## Application en logiciel R

Dans ce chapitre on prendra deux distributions, on va les simuler et puis on applique l'estimateur de Hill avec les données simulées en logiciel R, les distributions que nous avons choisies sont : la loi de Pareto, la loi de log-gamma. Nous avons fourni les deux packages nécessaires Package 'ReIns' et Package 'extRemes' aussi on a besoin de certaines des bibliothèques que nous allons présenter ci-dessous avec la fonction de chacune comme suite :

```
1 library(evir)
```

Fonctions pour la théorie des valeurs extrêmes, qui peuvent être divisées dans les groupes suivants ; analyse exploratoire des données, maxima de bloc, pics au-dessus des seuils (univariés et bivariés), processus ponctuels, distributions  $gev/gpd$ .

```
1 library(evmix)
```

Fonctions de modélisation de mélange de valeurs extrêmes, d'estimation de seuil et d'estimation de densité de noyau corrigée aux limites

```
1 library(robExtremes)
```

Estimation robuste optimale pour les distributions de valeurs extrêmes à l'aide de classes et de méthodes S4 (basées sur les packages 'disr', 'disrEx', 'disrMod', 'RobAStBase' et 'ROptEst').

```
1 library(VGM)
```

Fonctions pour Modèles linéaires et additifs généralisés vectoriels

```
1 library(evd)
```

Fonctions pour les distributions de valeurs extrêmes et de simulation, de distribution, de quantile et de densité pour des distributions de valeurs extrêmes paramétriques.

```
1 library(actuar)
```

Fonctions et ensembles de données pour la modélisation des distributions de pertes, théorie du risque et théorie de la ruine, simulation de modèles composés et distributions continues à queue lourde

## 3.1 Application sur la loi de Pareto

La distribution de Pareto fait partie d'une distribution de probabilité continue avec le paramètre d'échelle  $\beta$  et paramètre de forme  $\kappa$  où  $\beta > 0$  et  $\kappa > 0$ , ce distribution de Pareto introduite par un Italien économiste Vilfredo Pareto, la fonction de densité de la distribution de Pareto pour  $x \geq \beta, \beta > 0, \kappa > 0$  est la suivante :

$$f(x, \beta, \kappa) = \frac{\kappa \beta^\kappa}{x^{\kappa+1}}$$

Et la distribution cumulative de Pareto pour  $x \geq \beta, \beta > 0, \kappa > 0$  est donnée par :

$$F(x, \beta, \kappa) = 1 - \left(\frac{\beta}{x}\right)^\kappa$$

Certaines méthodes peuvent être utilisées pour estimer ces paramètres : méthodes du moment, méthode du maximum de vraisemblance mais parce que la distribution de Pareto est à queue lourde et appartient au domaine d'attraction de Fréchet nous serons intéressés de l'estimer par l'estimateur de Hill qui est le mieux adapté dans ce cas.

### 3.1.1 Simulation de loi Pareto en logiciel R

Nous avons réalisé une étude de simulation basée sur 500 échantillon d'observations indépendante et identiquement distribuée suite la loi de Pareto donc de densité  $f(x, \beta, \kappa)$  et de Fonction de répartition  $F(x, \beta, \kappa)$ , trois tailles d'échantillons que nous avons choisi ( $n = 50; n = 300, n = 500$ ), nous avons fixé le paramètre d'échelle  $\beta = 1$  et pour le paramètre de forme ( $\kappa = 1, \kappa = 2, \kappa = 3$ )

#### Cod R

```
1 #simulation de Pareto pour n=100
2 par(mfrow=c(1,2))
3 P1=rpareto(100,1,1)
4 plot(P1)
5 P2=rpareto(100,1,2)
6 plot(P2)
```

```
1 #simulation de Pareto pour n=200
2 par(mfrow=c(1,2))
3 P3=rpareto(200,1,1)
4 plot(P3)
5 P4=rpareto(200,1,2)
6 plot(P4)
```

```
1 #simulation de Pareto pour n=500
2 par(mfrow=c(1,2))
3 P5=rpareto(500,1,1)
4 plot(P5)
```

```
5 P6=rpareto(500,1,2)
6 plot(P6)
```

On obtient les résultats suivant :

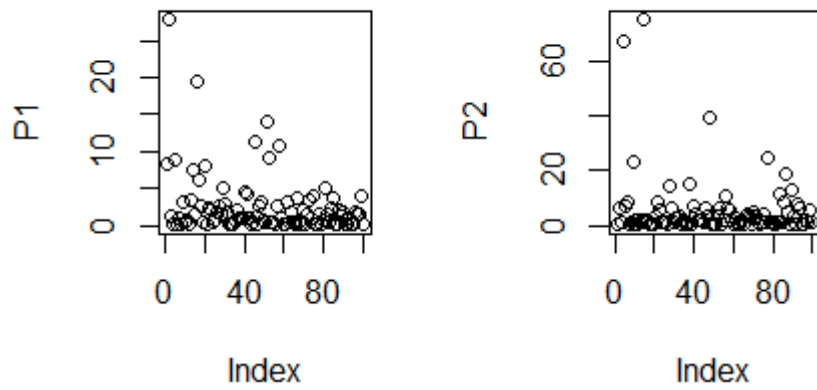


FIGURE 3.1 – Simulation de loi de Pareto  $P1(100, 1, 1)$ ,  $P2(100, 1, 2)$

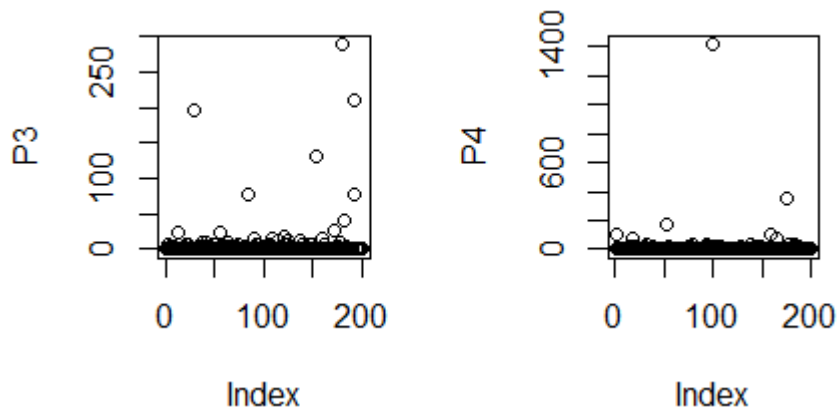
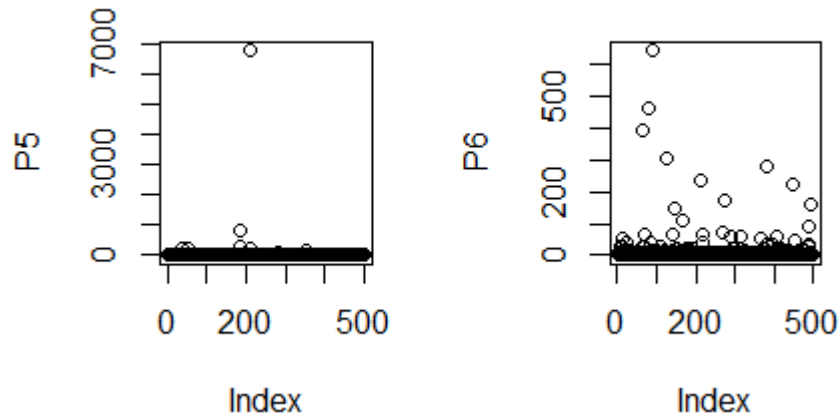


FIGURE 3.2 – Simulation de loi de Pareto  $P3(200, 1, 1)$ ,  $P4(200, 1, 2)$

FIGURE 3.3 – Simulation de loi de Pareto  $P5(500, 1, 1)$ ,  $P6(500, 1, 2)$ 

### 3.1.2 Estimation de Pareto par Hill

Nous ferons un rappel la l'estimateur de Hil, soit  $(k_n)_{n \geq 1}$  une suite d'entiers avec  $1 < k_n \leq n$ , l'estimateur de Hill est défini par

$$\hat{\gamma}_{k_n}^H = \frac{1}{k_n - 1} \sum_{i=1}^{k_n-1} \log X_{n-i+1,n} - \log X_{n-k_n+1,n}.$$

On applique l'estimateur de Hill aux donne simule de Preto

```

1 par(mfrow=c(1,2))
2 hill(P1)
3 hill(P2)
4 hill(P3)
5 hill(P4)
6 hill(P5)
7 hill(P6)

```

On obtient les résultats suivant :



FIGURE 3.4 – Estimation de loi de Pareto(P1,P1) par Hill



FIGURE 3.5 – Estimation de loi Pareto(P3,P4) par Hill

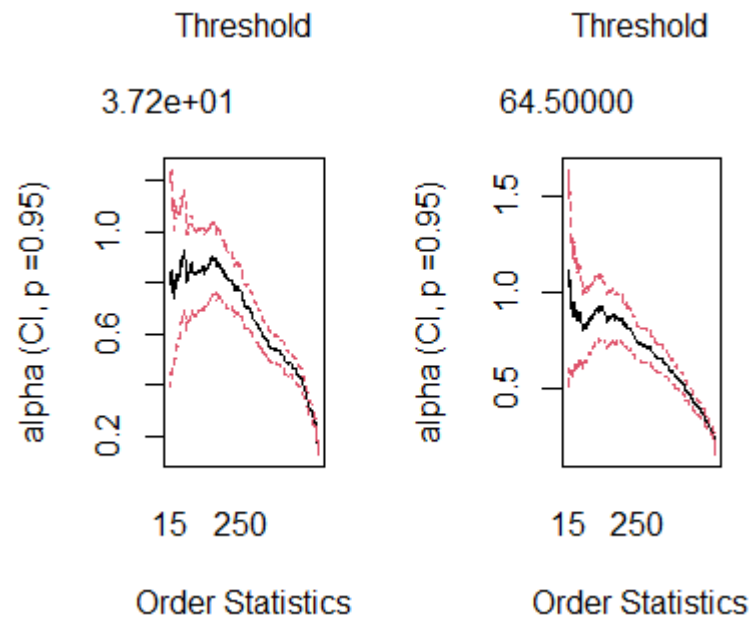


FIGURE 3.6 – Estimation de loi de Pareto (P5,P6) par Hill

«Nous voyons que le graphe se stabilise aux alentours de la valeur 1 ( $1 = 1.5 - 0.5$ ) pour ce qui concerne le paramètre  $\alpha$  et la stabilité démarre plus ou moins pour un seuil qu'on peut calculer à partir de ce graphe, et on peut utiliser ce seuil dans un ajustement généralisé de la distribution de Pareto de la queue supérieur.»

## 3.2 Application sur loi de log-gamma

On rappelle que  $X \sim \text{log-gamma}(\alpha, \beta)$  est utilisé pour indiquer que la variable aléatoire  $X$  a la distribution log-gamma avec un paramètre d'échelle positif  $\alpha$  et un paramètre de forme positif  $\beta$  ou fonction de densité de probabilité est définie par :

$$f(x) = \frac{e^{\beta x} e^{-e^x} / \alpha}{\alpha^{\beta} \tau(\beta)}$$

### 3.2.1 Simulation de loi log-gamma en logiciel R

Maintenant on va réaliser une étude de simulation basée sur 500 échantillon d'observations indépendante et identiquement distribuée suite la loi de Log-gamma donc de densité  $f(x)$ , trois tailles d'échantillons que nous avons choisi  $n = (100, 300, 500)$ , nous avons fixé le paramètre d'échelle  $\alpha = 1$  et pour le paramètre de forme  $\beta = (1, 2, 3)$

#### Cod R

```
1 #simulation de Log-gamma Pour n=100
2 par(mfrow=c(1,2))
3 L1=rlgamma(100,1,1)
4 plot(L1)
5 L2=rlgamma(100,1,2)
6 plot(L2)
```

```
1 #simulation de Log-gamma Pour n=200
2 par(mfrow=c(1,2))
3 L3=rlgamma(200,1,1)
4 plot(L3)
5 L4=rlgamma(200,1,2)
6 plot(L4)
```

```
1 #simulation de Log-gamma Pour n=500
2 par(mfrow=c(1,2))
3 L5=rlgamma(500,1,1)
4 plot(L5)
5 L6=rlgamma(500,1,2)
6 plot(L6)
```

nous avons choisi  $n = (100, 300, 500)$ , nous avons fixé le paramètre d'échelle  $\alpha = 1$  et pour le paramètre de forme  $\beta = (1, 2, 3)$

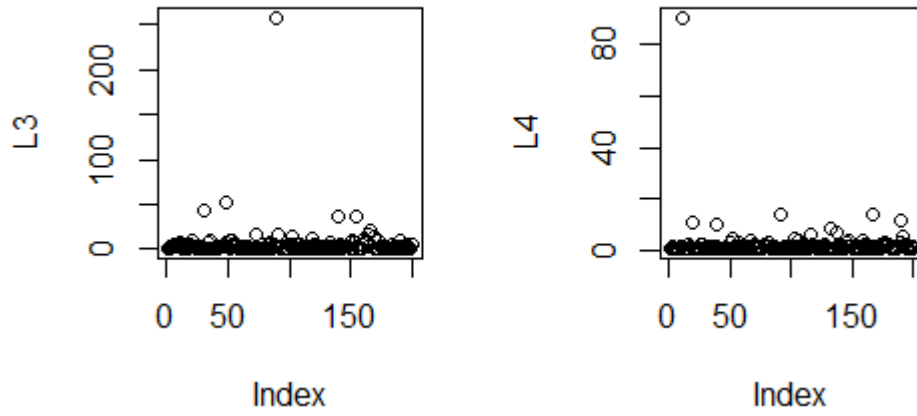


FIGURE 3.7 – Simulation de loi de log-gamma  $L3(200, 1, 1)$ ,  $L4(200, 1, 2)$

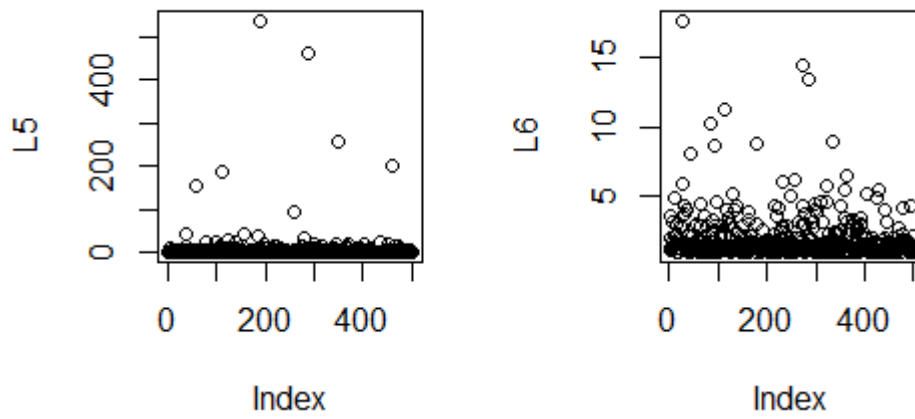


FIGURE 3.8 – Simulation de loi de log-gamma  $L5(500, 1, 1)$ ,  $L6(500, 1, 2)$

### 3.2.2 Estimation de loi log-gamma par Hill

Maintenant On applique l'estimateur de Hill aux données simulées :

```

1 par(mfrow=c(1,2))
2 hill(L1)
3 hill(L2)
4 hill(L3)
5 hill(L4)
6 hill(L5)
7 hill(L6)
8

```

On obtient les résultats suivant :

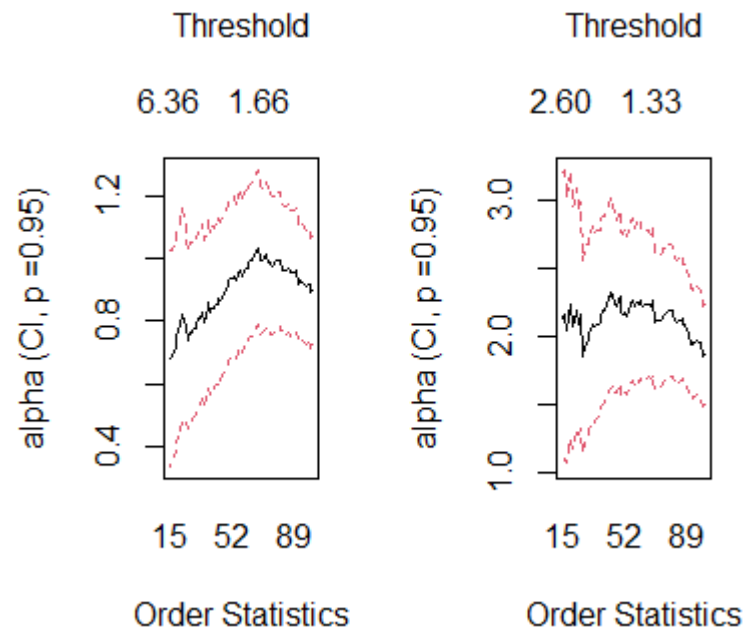


FIGURE 3.9 – Estimation de loi de log-gamma(L1,L2) par Hill

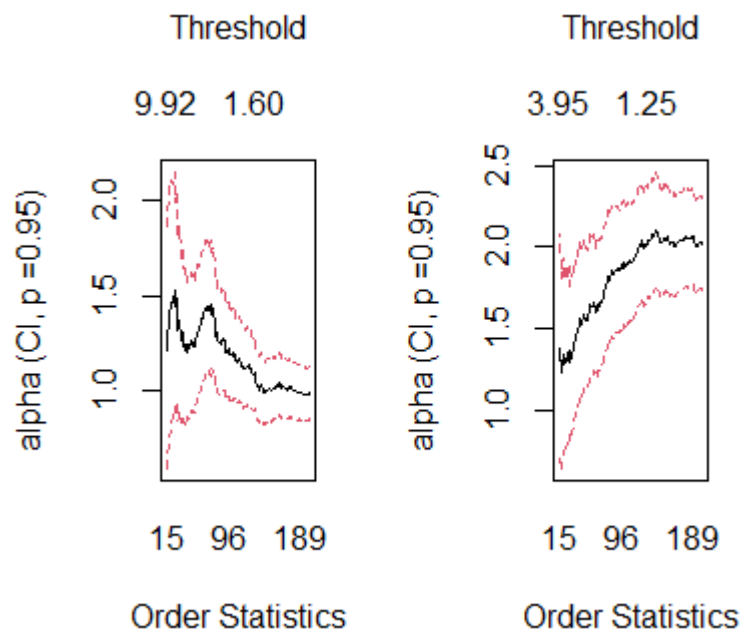


FIGURE 3.10 – Estimation de loi de log-gamma(L3,L4) par Hill

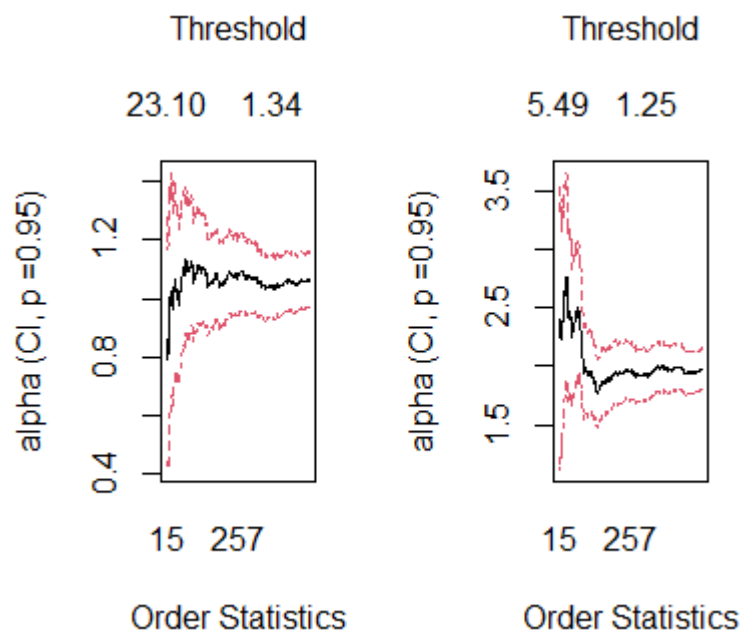


FIGURE 3.11 – Estimation de loi de log-gamma(L5,L6) par Hill

## Conclusion générale

En conclusion, la théorie des valeurs extrêmes nous donne une base large et bien fondée pour extrapoler au-delà de la gamme des données disponibles. Le résultat essentiel obtenu par Fisher-Tippett sur les lois limites possibles du maximum d'un échantillon, a apparemment provoqué l'idée que la théorie des valeurs extrêmes était quelque chose d'assez remarquable et très différent de la théorie classique de la limite centrale. En utilisant soit les maxima de l'échantillon, soit les dépassements d'un seuil, on peut faire des inférences précieuses sur les extrêmes. Celles-ci sont rendues rigoureuses, qui sont étayés par la théorie plus large de la variation régulière. De plus, il y a des techniques permettant de réaliser ces analyses même lorsque les conditions d'indépendance et de stationnarité ne sont pas réunies, ces es résultats ont déjà été adaptés à des domaines tels que la finance, la prévision des inondations, le changement climatique, ce qu'on a précisé dans le premier chapitre.

Dans la littérature sur l'estimation de l'indice des valeurs extrêmes, une variété l'estimateurs est proposée, un bon estimateur doit avoir les propriétés suivante :

- 1) effectuer une plus petite erreur d'estimation ;
- 2) satisfaire des propriétés théoriques telles que le décalage et d'invariance d'échelle
- 3) facile à calculer.

Pour cela on a consacré le chapitre 2 pour présenté quelques estimateurs de l'indice des valeurs extrêmes et de quantiles extrêmes, puis nous avons déterminé leurs propriétés asymptotiques.

Enfin, une simulation a été faite dans le troisième chapitre à l'aide de langage R.

Donc il faut devrai accorder plus d'attention à cette théorie en raison de son évaluation des risques dans un monde soumis à de nombreuses catastrophes soudaines.

# Bibliographie

- [1] Ahsanullah, Mohammad. "Extreme value distributions." *Extreme Value Distributions*. Atlantis Press, Paris, 2016. 73-91
- [2] Beirlant, Jan, et al. *Statistics of extremes : theory and applications*. Vol. 558. John Wiley and Sons, 2004.
- [3] Carrieu, Hervé. *Probabilité (L3M1) : exercices corrigés*. EDP sciences, 2012.
- [4] Borkar, Vivek S. *Probability theory : an advanced course*. Springer Science and Business Media, 1995
- [5] Coles, Stuart, et al. *An introduction to statistical modeling of extreme values*. Vol. 208. London : Springer, 2001.
- [6] D Oliveira, J. Tiago, ed. *Statistical extremes and applications*. Vol. 131. Springer Science and Business Media, 2013.
- [7] El Methni, Jonathan. *Contributions à l'estimation de quantiles extrêmes. Applications à des données environnementales*. Diss. Université de Grenoble, 2013.
- [8] Gardes, Laurent, and Stéphane Girard. "Estimation de quantiles extrêmes pour les lois à queue de type Weibull : une synthèse bibliographique." *Journal de la Société Française de Statistique* 154.2 (2013) : 98-118.
- [9] GHARBI, Djouhaina. "Estimation des quantiles extrêmes;UNIVERSIT... MOHAMMED KHIDER, BISKRA ,2020.
- [10] Hui, Eric Goh Ming. *Learn R for Applied Statistics*. Eric Goh Ming Hui, 2019.
- [11] Jean-Pierre leceteur ,*Statistique et probabilités,7e édition,Dunod,2019*
- [12] Jean-Pierre leceteur ,*Statistique et probabilités,Cours et exercices corrigé, Dunod, 2016*
- [13] Kotz, Samuel, and Saralees Nadarajah. *Extreme value distributions : theory and applications*. world scientific, 2000.
- [14] Lai, Chin-Diew. "Generalized weibull distributions." *Generalized Weibull Distributions*. Springer, Berlin, Heidelberg, 2014. 23-75.
- [15] Laurens de Haan Ana Ferreira,*Extreme Value Theory An Introduction,Springer,2006*.

- 
- [16] Lecoutre, Jean-Pierre. Statistique et probabilités. Dunod, 2019.
- [17] Lecoutre, Jean-Pierre. Statistique et probabilités-6e éd. : Cours et exercices corrigés. Dunod, 2016.
- [18] Mokrani, Tassadit. Distributions à Queues Lourdes et applications. Diss. UMMTO, 2016.
- [19] Munasinghe, Ranjiva, et al. "Fast Tail Index Estimation for Power Law Distributions in R." arXiv preprint arXiv.10308,2020.
- [20] Novak, Serguei Y. "Extreme value methods with applications to finance." Monographs on Statistics and Applied Probability 122,2011.
- [21] Ndao, Pathé. Modélisation de valeurs extrêmes conditionnelles en présence de censure. Diss. PhD thesis, Université Gaston Berger de Saint-Louis, 2015.
- [22] Reiss, Rolf-Dieter, Michael Thomas, and R. D. Reiss. Statistical analysis of extreme values. Vol. 2. Basel : Birkhäuser, 1997.
- [23] Soltane, Louiza. Analyse des valeurs extrêmes en présence de censure. Diss. Université Mohamed Khider-Biskra, 2017.
- [24] Toulemonde, Gwladys. Estimation et tests en théorie des valeurs extrêmes. Diss. Université Pierre et Marie Curie-Paris VI, 2008.