

Université Abderrahmane Mira de Béjaia

Faculté des Sciences Exactes

Département de Mathématiques



Mémoire de fin de cycle

En vue de l'obtention du diplôme de Master en

Mathématiques

Option :

Statistique et Analyse Décisionnelle

Thème

**Estimation de probabilités d'évènements rares et de
quantiles extrêmes.**

Présenté par

M^{elle} DJOUADI Siham

M^{elle} MERADI Wassila

Devant le jury composé de :

M^{me} TIMERIDJINE.K.	M.C.A.	Présidente	U. A .Mira de Béjaia
M^{elle} SAADI.N.	M.C.B.	Examinatrice	U. A .Mira de Béjaia
M^{me} BARECHE.A.	M.C.A.	Promotrice	U. A .Mira de Béjaia

Année universitaire 2016/2017

Remerciements

Nous tenons tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce modeste travail.

En second lieu, notre encadreur M^{me} BARECHE A. pour ses précieux conseils et son orientation ficelée durant toute la période du travail.

Nos vifs remerciements vont également aux membres du jury, M^{me} Timeridjine K. et M^{elle} Saadi N. pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions.

On n'oublie pas nos familles et nos amis qui par leurs prières et leurs encouragements, on a pu surmonter tous les obstacles.

Enfin, nous tenons également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

DEDICACES

Je dédie ce mémoire à :

Ma mère, qui a œuvré pour ma réussite, par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mon père, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.

Mes frères et sœurs Saadia, Yasmina, Bilal et Ramy qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

Mes professeurs de l'université qui doivent voir dans ce travail la fierté d'un savoir bien acquis.

Wassila

DEDICACES

Je dédie ce mémoire :

A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.

A mes chères sœurs Sonia, Afifa et Ahlam pour leurs encouragements permanents, et leur soutien moral.

A mes chers frères, Mohamed Aziz et Abd El Malek pour leur appui et leur encouragement.

A toute ma famille pour leur soutien tout au long de mon parcours universitaire.

Que ce travail soit l'accomplissement de vos vœux tant allégués, et le fruit de votre soutien infaillible.

Merci d'être toujours là pour moi.

SIHAM

Table des matières

Introduction	2
1 Notions de base sur la théorie des valeurs extrêmes	6
1.1 Introduction	6
1.2 Les évènements rares	6
1.3 Statistiques d'ordres	7
1.4 La théorie des valeurs extrêmes	10
1.4.1 Théorème fondamental des valeurs extrêmes	10
1.5 Caractérisation des domaines d'attraction	14
1.5.1 Fonctions à variations régulières	14
1.5.2 Domaine d'attraction de Fréchet	15
1.5.3 Domaine d'attraction de Weibull	16
1.5.4 Domaine d'attraction de Gumbel	17
1.6 Estimation des paramètres de la loi des valeurs extrêmes	17
1.6.1 Estimation de l'indice des valeurs extrêmes γ	18
1.6.2 Méthode du maximum de vraisemblance	21
1.6.3 Méthode des moments pondérés	22
1.6.4 Méthode des moments pondérés généralisés	23
1.7 Distribution des excès (P.O.T)	24
1.8 Théorème de Pickands	26
1.9 Propriétés de la distribution de pareto généralisée (GPD)	26
1.9.1 La détection du seuil	27
1.10 Estimation des paramètres de la GPD	28
1.10.1 Méthode du maximum de vraisemblance	28
1.10.2 Méthode des moments pondérés	29
1.11 Conclusion	30
2 Estimation de quantile extrême et de niveau de retour	31
2.1 Introduction	31
2.2 Quantile	31
2.2.1 Quantile d'ordre p	32
2.2.2 Fonction quantile de queue	32
2.2.3 Quantile extrême	32

2.2.4	Quantile empirique	32
2.3	Niveau de retour et période de retour	33
2.3.1	Notion de niveaux et périodes de retour	33
2.4	Estimation d'un quantile	33
2.5	Méthodes d'estimation d'un quantile extrême	33
2.5.1	Estimation naturelle	34
2.5.2	Méthode des blocs	34
2.5.3	Méthode des excès ou P.O.T.	36
2.6	Estimation de niveau de retour	37
2.7	Estimation de la période de retour	37
2.8	Conclusion	38
3	Application en pluviométrie	39
3.1	Introduction	39
3.2	Application numérique sur un échantillon des données réelles	39
3.2.1	L'ajustement par la loi normale	42
3.2.2	Détection du seuil	43
3.2.3	L'approximation de la loi des excès par une distribution de Pareto Généralisée (GPD)	45
3.2.4	Estimation des paramètres de la GPD par la méthode de maximum de vraisemblance	46
3.2.5	L'estimation de quantile extrême et de la période de retour	47
3.2.6	Interprétation des résultats	48
3.3	Application numérique sur un échantillon simulé	48
3.3.1	Sélection de seuil	49
3.3.2	Estimation des paramètres de la GPD par la méthode de maximum de vraisemblance	50
3.3.3	L'estimation de quantile extrême	51
3.3.4	L'estimation de la période de retour	51
3.3.5	Interprétation des résultats	51
3.4	Conclusion	52
	Conclusion	52
	Bibliographie	55

Notations et abréviations

Abréviations

- TCL** : Théorème central limite.
- TVE** : Théorie des valeurs extrêmes.
- i.i.d** : Indépendant et identiquement distribué.
- GEV** : Distribution des Valeurs Extrêmes Généralisée.
- EMV** : Méthode du Maximum de Vraisemblance.
- GPD** : Distribution de Pareto Généralisée.
- POT** : L'approche par dépassements de seuil.
- BM** : Block Maxima.

Notations

- F^{\leftarrow} : Inverse généralisée de la fonction de répartition.
- F^{-1} : Inverse de la fonction de répartition.
- x_F : Point terminal.
- F_n : Fonction de répartition empirique.

Liste des figures

1.1	Exemples de densités associées à la loi des valeurs extrêmes	12
1.2	Comparaison de l'erreur d'estimation en fonction de la méthode.....	21
1.3	Dépassement de seuil (POT).....	25
1.4	Distribution GPD.....	27
3.1	L'histogramme de la fréquence des débits annuels maximaux.....	43
3.2	La distribution moyenne des excès.....	44
3.3	Dépassement de seuil	45
3.4	L'approximation des excès par une GPD.....	46
3.4	Mean Excess Plot d'échantillon simulé.....	50

Liste des tableaux

1.1	Quelques lois et leurs domaines d'attraction.....	14
3.1	Débits annuels maximaux de la Garonne à Mas d'Agenais sur la période 1913-1977...	40
3.2	Les paramètres obtenus (cas de l'échantillon réel)	47
3.3	Les paramètres obtenus (cas de l'échantillon simulé)	51

Introduction générale

Au cours des dernières années, nous avons pu observer dans la recherche scientifique, une modélisation des évènements rares. Ces derniers sont des évènements dont la probabilité d'apparition est trop faible, c'est-à-dire ils se trouvent dans les queues des distributions. La modélisation des évènements extrêmes (tels les inondations, les crues, les tempêtes, les pics de pollution etc..) est aujourd'hui un champ de recherche particulièrement actif.

Ces évènements extrêmes peuvent causer des dégâts humains et matériels considérables. De telles catastrophes ne peuvent pas toujours être évitées. Cependant, la société peut prendre des actions préventives pour minimiser leurs effets. Pour cela, les spécialistes ont à leur disposition la théorie statistique des valeurs extrêmes. Elle donne un résultat très intéressant parce qu'il est d'une portée générale.

La théorie des valeurs extrêmes est une branche de la statistique qui essaie d'amener une solution face à des évènements rares. Elle se repose principalement sur des distributions limites des extrêmes et leurs domaines d'attraction. Cependant, on y retrouve deux modèles : loi généralisée des extrêmes (GEV : " Generalized Extreme Value ") et loi de Paréto généralisée (GPD : " Generalized Pareto Distribution ").

Tout a commencé avec les auteurs Fisher et Tippett (1928) [10] quand ils étudiaient la résistance des fils de coton puis plus tard Gnedenko (1943) [15] s'est intéressé à ces distributions. Ils ont énoncé un théorème fondamental avec la création de trois domaines d'attraction : domaine d'attraction de Fréchet, Gumbel et Weibull.

Cette théorie est largement utilisée, elle a trouvé des champs d'application en :

- Climatologie : étude des évènements climatiques extrêmes (précipitations, températures, chutes de neige), modélisation des grands feux de forêt (voir par exemple Alvarado et al. 1998)[1].
- Hydrologie : crues consécutives à des pluies torrentielles : aux Pays-Bas, digues menacées par l'effet conjoint des grandes marées et des conditions climatiques en Mer du Nord (Inondation de 1953).
- Assurance : survenue des sinistres d'intensité exceptionnelle (ouragan Katrina en 2005, importants incendies en risques industriels, sinistres graves en responsabilité civile automobile) qui peuvent avoir des conséquences négatives sur les résultats et la solvabilité des organismes d'assurance.
- Finance : fortes variations du cours d'actifs financiers, gestions du risque opérationnel des banques (la crise fin des années 2000).

Parmi les précurseurs, on peut citer en particulier Gnedenko (1943) [15], Leadbetter (1983) [18], Resnick (1987) [22], Embrechts et al. (1997) [8], de Haan et Ferreira (2006) [9].

La théorie des valeurs extrêmes est fondée sur les lois des probabilités, elle offre un cadre mathématique rigoureux pour l'estimation de quantiles extrêmes et de niveau de retour. Trois différentes techniques d'estimation du quantile extrême sont mentionnées : estimation naturelle , méthode des blocs , méthodes des excès où P.O.T.

Ce travail peut être considéré dans sa globalité comme une contribution à la théorie des valeurs extrêmes et à l'estimation d'un quantile extrême ainsi qu'à l'estimation du niveau et période de retour.

Ce mémoire est composé d'une introduction générale, de trois chapitres, d'une conclusion générale et d'une bibliographie.

Le premier chapitre présente un état de l'art sur la théorie des valeurs extrêmes, nécessaire pour estimer le quantile extrême et la période de retour qui sont la problématique principale de cette thèse. Tout d'abord, on présentera les principaux résultats et définitions de la théorie des valeurs extrêmes utiles dans nos travaux. Après avoir introduit le comportement du maximum d'un échantillon, on présentera les deux principaux outils servant

à modéliser le comportement des valeurs extrêmes : la loi des valeurs extrêmes et la loi des excès. On s'intéressera ensuite à la caractérisation des domaines d'attraction. Enfin on rappellera les différentes méthodes d'estimation des paramètres de la GEV et la loi des excès.

Dans le deuxième chapitre, on présente des méthodes pour estimer le quantile extrême : estimation naturelle basé sur le quantile empirique, méthode des blocs(Bloc Maxima), (méthode P.O.T (Peak Over Threshold)). Nous donnons aussi un aperçu sur les méthodes d'estimation du niveau de retour et période de retour.

Dans le chapitre 3, on présente une application en pluviométrie pour cela on dispose de quelques enregistrements des débits passés de la rivière Garonne au cours de chacune des années de 1913 à 1977. Dans le but d'estimer le quantile extrême et la période de retour du débordement de cette rivière on a mis en pratique la méthode P.O.T en faisant appel à des fonctions et des méthodes prédéfinies sous le logiciel R.

Chapitre 1

Notions de base sur la théorie des valeurs extrêmes

1.1 Introduction

La théorie des valeurs extrêmes (TVE) est une branche des statistiques qui a pour but de modéliser et de décrire la survenue et l'intensité d'évènements dits rares c'est-à-dire qui présentent des variations de très grandes amplitudes.

Ce chapitre introduit la notion de la TVE pour une suite de variables aléatoires et énonce les principaux résultats probabilistes les concernant. Nous y exposerons uniquement la théorie des valeurs extrêmes pour des variables aléatoires indépendantes.

1.2 Les évènements rares

Les évènements rares sont des évènements ayant une faible probabilité d'apparition. Lorsque le comportement de ces évènements est dû au hasard on peut étudier leur loi. Ils sont dits extrêmes quand il s'agit de valeurs beaucoup plus grandes ou plus petites que celles observées habituellement.

Les évènements extrêmes et catastrophiques (tremblements de terre, inondations, accidents nucléaires, crises monétaires ou financières, crachs boursiers, émergence d'un nouveau phénomène endémique, etc.) dominent l'actualité quotidienne par leur caractère imprévisible. Compte tenu de l'importance des enjeux sociaux et scientifiques, aucun débat sérieux sur le hasard ne saurait être mené sans une réflexion sur les évènements rares et extrêmes.

On considère un évènement dont la probabilité d'apparition est p ; dans un modèle à temps discret, la distribution du premier instant de réalisation de l'évènement est une loi géométrique de paramètre p , ie $\mathbb{P}(T = k) = p(1 - p)^{k-1}$. En particulier, le temps moyen d'attente de la réalisation de l'évènement est $\mathbb{E}(T) = \frac{1}{p}$.

On peut donc apprécier le caractère exceptionnel d'un évènement au travers de sa fréquence : un évènement qui se produit une fois par siècle est ainsi associé à une probabilité de survenance dans l'année de 1 % .

1.3 Statistiques d'ordres

Soient n variables aléatoires indépendantes identiquement distribuées (iid) X_1, \dots, X_n de densité commune f , et de fonction de répartition F . Notons $\bar{F}(x) = 1 - F(x)$ avec $x \in \mathbb{R}$ la fonction de survie (ou de queue).

Rangeons ces variables aléatoires par "ordre croissant de grandeur". Pour cela, nous introduisons la notation $X_{i:n}$:

$$X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n-1:n} \leq X_{n:n},$$

$X_{i:n}$ est donc la i -ième statistique d'ordre (ou statistique d'ordre i) dans un échantillon de taille n .

Soient n observations x_1, \dots, x_n . La valeur observée de $X_{i:n}$ est notée $x_{i:n}$ et nous avons

$$x_{i:n} = e_i^T \text{sort}(x),$$

où sort est la fonction de tri ascendant et x le vecteur des observations x_1, \dots, x_n . Deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des évènements extrêmes. Ce sont les statistiques d'ordre extrême qui correspondent à la plus petite statistique d'ordre $X_{1:n}$ (ou statistique du minimum)

$$X_{1:n} = \min(X_1, \dots, X_n),$$

et à la plus grande statistique d'ordre $X_{n:n}$ (ou statistique du maximum)

$$X_{n:n} = \max(X_1, \dots, X_n).$$

Avec : $\max(X_1, \dots, X_n) = -\min(X_1, \dots, X_n)$.

Enfin, nous introduisons l'écart des statistiques d'ordre $W_{i,j:n} = X_{j:n} - X_{i:n}$ ($i < j$). L'écart $W = W_{1,n:n} = X_{n:n} - X_{1:n}$ est appelée la déviation extrême.

L'expression de la distribution de $X_{i:n}$ est :

$$\begin{aligned} F_{i:n}(x) &= \mathbb{P}(X_{i:n} \leq x) \\ &= \sum_{r=i}^n \binom{n}{r} [F(x)]^r [1 - F(x)]^{n-r}. \end{aligned}$$

Nous en déduisons que la fonction de densité est :

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

Pour les statistiques d'ordre extrême, nous obtenons des expressions très simples. En effet, nous avons

pour la statistique du minimum :

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

et :

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

pour la statistique du maximum :

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

et :

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

Remarque 1.3.1. Les expressions de $F_{1:n}$ et $F_{n:n}$ peuvent s'obtenir très facilement en considérant les relations

$$\{X_{1:n} \geq x\} \Leftrightarrow \{\min(X_1, \dots, X_n) \geq x\}$$

$$\Leftrightarrow \bigcap_{i=1}^n \{X_i \geq x\}$$

et

$$\{X_{n:n} \leq x\} \Leftrightarrow \{\max(X_1, \dots, X_n) \leq x\}$$

$$\Leftrightarrow \bigcap_{i=1}^n \{X_i \leq x\}.$$

En utilisant la propriété d'indépendance des variables aléatoires X_1, \dots, X_n , nous en déduisons que

$$\begin{aligned} F_{1:n}(x) &= \mathbb{P}(X_{1:n} \leq x) \\ &= 1 - [1 - F(x)]^n \end{aligned}$$

et

$$\begin{aligned} F_{n:n}(x) &= \mathbb{P}(X_{n:n} \leq x) \\ &= [F(x)]^n. \end{aligned}$$

Définition 1.3.1. Le point terminal d'une distribution F est défini par :

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

Définition 1.3.2. On appelle inverse généralisé de F , l'application notée F^{\leftarrow} définie par :

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

L'inverse généralisé F^{\leftarrow} coïncide avec l'inverse F^{-1} lorsque la fonction F est strictement croissante et continue.

1.4 La théorie des valeurs extrêmes

La loi normale représente l'une des lois de probabilité les plus utilisées pour l'ajustement de très nombreuses variables. Nous pouvons citer à titre d'exemple les rendements agricoles, qui ont constitué la première motivation historique justifiant l'utilisation de la loi normale, les tailles, les poids, les numérations sanguines ou les dosages hormonaux. La loi normale a été introduite par Abraham de Moivre en 1733, comme une loi limite d'une loi binomiale. Elle a été intégrée par la suite, dans les calculs de Gauss en ce qui concerne le mouvement des corps célestes.

L'ample utilisation de la loi normale tient de grande part à sa simplicité et d'autre part au théorème centrale limite (TCL). Ce théorème assure la convergence vers la loi normale, quelle que soit les lois individuelles des échantillons utilisées. Le TCL admet des généralisations qui garantissent la convergence de sommes de variables aléatoires sous les hypothèses faibles.

Malgré sa grande utilité, la loi normale échoue à décrire l'évidence empirique de plusieurs domaines, qui sont caractérisés beaucoup plus par l'existence des extrêmes, telles que les séries financières, hydrologiques ou météorologiques.

Donc si l'on souhaite étudier les valeurs extrêmes d'un t échantillon, le TLC ne présente que peu d'intérêt. On utilise plutôt un résultat établissant la loi asymptotique du maximum de l'échantillon [2].

1.4.1 Théorème fondamental des valeurs extrêmes

Historiquement, l'étude de la loi de probabilité du maximum d'un échantillon de n variables a été la première approche pour décrire les événements extrêmes. Fisher et Tippett [10] en 1928 ont, les premiers, déduit de manière heuristique les lois limites possibles pour le maximum d'une suite de variables aléatoires indépendantes et de même loi, avant que Gnedenko [15] en 1943 n'obtienne rigoureusement la convergence, dont la preuve fut simplifiée par de Haan [12] en 1976. Les travaux de Von Mises [24] en 1936 et Jenkinson [11] en 1955 ont permis de donner une forme unifiée à ce résultat. Les applications ont commencé suite aux travaux de Gumbel [16] en 1954, en particulier en hydrologie.

Le résultat de base de la TVE consiste à décrire la loi asymptotique du maximum de n variables indépendantes et identiquement distribuées (iid). La forme de cette loi (notée

GEV) dépend d'un seul paramètre qui permet de spécifier le comportement de la queue de la loi considérée. Ce paramètre est l'indice des valeurs extrêmes (noté γ). En d'autres termes, le résultat de base de la TVE assure que la loi du maximum de n variables (i.i.d) est toujours dans le domaine d'attraction d'une loi GEV_γ .

Le théorème ci-dessous connu sous le nom de théorème de Fisher-Tippet-Gnedenko est fondamental en théorie des valeurs extrêmes car il établit la loi asymptotique du maximum $X_{n,n}$ convenablement normalisé d'un échantillon.

Théorème 1.4.1. [10] Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F . S'il existe deux suites normalisantes réelles $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1} \in \mathbb{R}$ et une loi non-dégénérée H_γ telle que :

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \mathcal{H}_\gamma(x), \quad (1.1)$$

alors \mathcal{H} est l'une des trois lois limites :

$$\Lambda(x) = \exp\{-\exp(-x)\}, \quad x \in \mathbb{R} \quad (\text{distribution de Gumbel})$$

$$\Phi_\gamma(x) = \begin{cases} \exp\{(-x)^{-\gamma}\} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} \quad (\text{distribution de Fréchet})$$

$$\Psi_\gamma = \begin{cases} \exp\{-(-x)^\gamma\} & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \quad (\text{distribution de Weibull})$$

Le regroupement de ces trois lois limites donne :

$$\mathcal{H}_\gamma(x) = \exp(-(1 + \gamma x)_+^{-1/\gamma}), \quad (1.2)$$

où $\gamma \in \mathbb{R}$ et $z_+ = \max(0, z)$.

Le cas $\gamma = 0$ dans l'équation précédente peut être vu comme le cas limite lorsque $x \rightarrow \infty$. On retrouve alors la loi de Gumbel ayant pour fonction de répartition :

$$\mathcal{H}_0(x) = \exp(-\exp(-x)).$$

Le comportement limite du maximum normalisé est ainsi décrit par la fonction de répartition \mathcal{H}_x pour la plus grande partie des lois usuelles. \mathcal{H}_x est appelée fonction de répartition de la loi des valeurs extrêmes, en anglais “Generalized Extreme Value Distribution” notée GEV. En introduisant les paramètres de localisation μ et de dispersion σ dans la paramétrisation (1.2), on obtient la forme la plus générale de la GEV suivante :

$$\mathcal{H}_{\gamma,\mu,\sigma}(x) = \begin{cases} \exp\left\{-\left(1 + \gamma\frac{x-\mu}{\sigma}\right)^{-1/\gamma}\right\} & \text{si } \gamma \neq 0 \text{ et } 1 + \gamma\frac{x-\mu}{\sigma} > 0, \\ \exp(-\exp(-\frac{x-\mu}{\sigma})) & \text{si } \gamma = 0. \end{cases}$$

La densité GEV de la loi des valeurs extrêmes s'en déduit par simple dérivation. Donc elle est donnée par :

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

La figure 1.1 présente les différentes formes de h_γ possibles en fonction de γ .

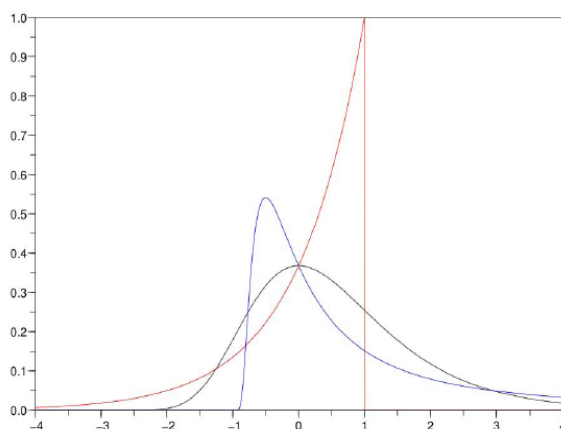


FIG. 1.1 – Exemples de densités associées à la loi des valeurs extrêmes (noir : $\gamma = 0$, bleu : $\gamma = 1$ et rouge : $\gamma = -1$) [7].

L'unification du comportement du maximum en une seule fonction de répartition facilite grandement l'étude du comportement du maximum. Cette loi dépend du seul paramètre

de forme γ appelé indice des valeurs extrêmes ou indice de queue. Comme on le verra ce dernier est le paramètre clé de toute la théorie des valeurs extrêmes. L'estimation de γ nous fournira le comportement de la queue de distribution. En effet, selon son signe, on distingue trois domaines d'attraction dont quelques densités ont été représentées pour γ fixé sur la Figure 1.1.

- si $\gamma > 0$, on dit que F appartient au domaine d'attraction de Fréchet, que l'on notera $D(\text{Fréchet})$. Il contient toutes les lois dont la fonction de survie décroît comme une fonction puissance. Ce sont les lois à « queue lourde ». Les distributions du domaine de Fréchet sont beaucoup utilisées en fiabilité mécanique, dans les phénomènes climatiques tels que la météorologie, l'hydrologie, la vitesse du vent enregistrée en continu dans les aéroports et en finance dans les études de risque.
- si $\gamma < 0$, on dit que F appartient au domaine d'attraction de Weibull, que l'on notera $D(\text{Weibull})$. Toutes les lois de ce domaine d'attraction ont un point terminal x_F fini. Les distributions de type de Weibull sont souvent utilisées pour décrire la résistance mécanique d'un matériau ou encore le temps de fonctionnement d'un appareil électronique ou mécanique.
- si $\gamma = 0$, on dit que F est dans le domaine d'attraction de Gumbel, que l'on notera $D(\text{Gumbel})$. Ce sont les lois dont la fonction de survie décroît vers zéro à une vitesse exponentielle. Ces distributions sont souvent utilisées pour faire des prévisions dans les événements environnementaux tels que le séisme (le tremblement de terre), l'hydrologie (les inondations, la destruction des barrages), etc.

Voici un classement de quelques lois par domaine d'attraction dans le tableau 1.1.

Fréchet ($\gamma > 0$)	Gumbel ($\gamma = 0$)	Weibull ($\gamma < 0$)
Pareto	Normale	Uniforme
Student	Exponentielle	Beta
Burr	Log-normale	
Chi-deux	Gamma	
Fréchet	Weibull	
Log-gamma	Gumbel	
Log-logistique	Logistique	
Cauchy		

TAB. 1.1 – Quelques lois et leurs domaines d’attraction [5].

1.5 Caractérisation des domaines d’attraction

Un problème important revient à définir les conditions (nécessaires et suffisantes) d’appartenance d’une distribution à un domaine d’attraction. La recherche de ce domaine d’attraction peut être considérée comme l’étude réciproque de la recherche de la distribution des valeurs extrêmes associée éventuellement à une distribution. Ceci consiste donc à répondre à la question suivante : étant donnée une loi \mathcal{H} de type extrême (donc appartenant à l’une des trois familles Fréchet, Gumbel et Weibull) quels sont les critères à vérifier pour que la loi du maximum de la suite de variables aléatoires i.i.d. de loi F converge vers \mathcal{H} ?

Différentes caractérisations des trois domaines d’attraction de Fréchet, Gumbel et de Weibull ont été proposées dans de Haan (1985) ; Resnick (1987) ; Embrechts et al. (1997) ; de Haan et Ferreira (2006).

Et comme ces caractérisations sont basées sur la notion de fonctions à variations régulières, on commencera par rappeler la définition de ces fonctions et quelques unes de leurs propriétés.

1.5.1 Fonctions à variations régulières

Définition 1.5.1. [5] Une fonction mesurable positive f est dite à variation régulière d’indice α à l’infini et on note $f \in RV_\alpha$ si pour tout $x > 0$, nous avons

$$\lim_{t \rightarrow +\infty} \frac{f(tx)}{f(t)} = x^\alpha.$$

On dira que f est à variation régulière au voisinage de 0 si $f(x^{-1})$ est à variation régulière à l'infini.

Si $\alpha = 0$, on dit que f est à variation lente, on la note souvent ℓ (ou parfois L).

Si $\alpha = +\infty$, f est dite fonction à variation rapide.

Remarque 1.5.1. Les exemples typiques de fonctions à variation lente sont les constantes positives ou les fonctions convergeant vers une constante positive. Citons, $\ln(1+x)$, $\ln(\ln(e+x))$, etc.

Il est facile de montrer qu'une fonction f à variation régulière d'indice α peut toujours s'écrire sous la forme

$$f(x) = x^\alpha \ell(x),$$

où $\ell \in RV_0$.

1.5.2 Domaine d'attraction de Fréchet

Comme rappel le domaine d'attraction de Fréchet contient les lois dont la fonction de survie est à décroissance polynomiale, i.e. les lois à queues lourdes ou lois de type Pareto. Les lois de ce domaine ont un point terminal x_F infini. En effet, le résultat ci-dessous énoncé par Gnedenko et dont on trouvera une démonstration simple dans le livre de Resnick [22] [Proposition 1.11] assure que toute fonction appartenant au domaine d'attraction de Fréchet est une fonction à variations régulières (voir Définition 1.5) et inversement.

Théorème 1.5.1. [15] *Une fonction de répartition F appartient au domaine d'attraction de Fréchet (avec un indice des valeurs extrêmes $\gamma > 0$) si et seulement si $x_F = \infty$ et si la fonction de survie \bar{F} est à variations régulières d'indice $-1/\gamma$ qui s'écrit $\bar{F} \in RV_{-1/\gamma}$ i.e*

$$\forall x > 0, \lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^{-1/\gamma}.$$

Autrement dit, une fonction de répartition F appartenant au domaine d'attraction de Fréchet s'écrit sous la forme :

$$F(x) = 1 - x^{-1/\gamma} \ell(x) \text{ avec } \ell \in RV_0.$$

Les suites de normalisation (a_n) et (b_n) sont données dans ce cas pour tout $n > 0$ par :

$$a_n = \overline{F}\left(\frac{1}{n}\right) \text{ et } b_n = 0.$$

Le domaine d'attraction de Fréchet regroupe une grande diversité de lois comptant parmi elles des lois usuelles (loi de Student, loi du Chi-deux, loi Log-gamma, loi de Fréchet). Se reporter au tableau 1.1 pour plus d'exemples de lois.

1.5.3 Domaine d'attraction de Weibull

Rappelons que toutes les lois appartenant au domaine d'attraction de Weibull ont un point terminal x_F fini. Le résultat suivant [15],[22], montre que l'on passe du domaine d'attraction de Fréchet à celui de Weibull par un simple changement de variable dans la fonction de répartition.

Théorème 1.5.2. [22]

Une fonction de répartition F appartient au domaine d'attraction de Weibull (avec un indice des valeurs extrêmes $\gamma < 0$) si et seulement si $x_F < 1$ et si la fonction de répartition F^ définie par :*

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

appartient au domaine d'attraction de Fréchet avec un indice des valeurs extrêmes $-\gamma > 0$, c'est-à-dire que $\overline{F^*}$ est une fonction à variation régulière d'indice $1/\gamma$ à l'infini et on note $(\overline{F^*} \in RV_{1/\gamma})$. Dans ce cas un choix possible pour les suites a_n et b_n est :

$$a_n = x_F - F^{\leftarrow}(1 - 1/n) \text{ et } b_n = x_F.$$

La démonstration du théorème (1.5.2) fait appel à la Proposition 1.13 de la référence [22],

Remarque 1.5.2. On déduit du théorème (1.5.2) que F appartient au domaine de Weibull, ($\gamma < 0$) si et seulement si $x_F = \infty$ et $\overline{F}(x) = (x_F - x)^{-1/\gamma} \ell((x_F - x)^{-1})$ où $\ell \in RV_0$.

1.5.4 Domaine d'attraction de Gumbel

On sait que le domaine d'attraction de Gumbel ne contient que les lois à queues légères voir le tableau 1.1. La caractérisation des fonctions de répartition de ce domaine est complexe contrairement aux deux autres domaines.

Le résultat ci-dessous qui est démontré notamment dans Resnick (1987), donne une caractérisation plus précise.

Théorème 1.5.3. [22] *Une fonction de répartition F appartient au domaine d'attraction de Gumbel si et seulement si il existe $t < x_F \leq \infty$ tel que :*

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

$c(x) \rightarrow c > 0$ lorsque $x \rightarrow x_F$ et δ est une fonction positive et dérivable de dérivée δ' telle que $\delta'(x) \rightarrow 0$ lorsque $x \rightarrow x_F$.

Dans ce cas, un choix possible pour les suites (a_n) et (b_n) pour tout $n > 0$ est :

$$a_n = q(1/n) \quad \text{et} \quad b_n = \frac{1}{\bar{F}(a_n)} \int_{a_n}^{x_F} \bar{F}(z) dz.$$

Ce domaine d'attraction regroupe une grande diversité de lois, comptant parmi elles la plupart des lois usuelles (voir Tableau 1.1), ce qui rend la caractérisation de la fonction quantile de lois appartenant au domaine d'attraction de Gumbel difficile.

1.6 Estimation des paramètres de la loi des valeurs extrêmes

Dans cette partie on s'intéresse à l'estimation des paramètres de la GEV, et plus précisément on va s'intéresser à l'estimation de l'indice de queue γ qui joue un rôle essentiel dans le comportement de la loi des extrêmes.

1.6.1 Estimation de l'indice des valeurs extrêmes γ

Les deux estimateurs sans doute les plus populaires dans la littérature sont les estimateurs de Hill (1975) et de Pickands (1975).

On note $X_{1,n}, \dots, X_{n,n}$ les statistiques d'ordre associées à l'échantillon X_1, \dots, X_n . C'est à dire que l'on classe X_1, \dots, X_n par ordre croissant de sorte que :

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

On considère les k valeurs les plus grandes (ou les plus petites). k dépend a priori de n , même si on ne le mentionnera pas dans la notation : l'idée est d'avoir $k \rightarrow \infty$ lorsque $n \rightarrow \infty$, mais sans prendre « trop » de valeurs de l'échantillon, ce qui conduit à imposer $\frac{k}{n} \rightarrow 0$.

Incidentement, cela implique que se posera la question du choix optimal de k . En effet, il est indispensable de calculer ces estimateurs sur les queues de distribution. Choisir un k trop élevé engendre le risque de prendre en compte des valeurs qui ne sont pas extrêmes, à l'inverse, un sous-échantillon trop petit ne permet pas aux estimateurs d'atteindre leur niveau de stabilité.

Enfin, on retiendra que l'approche non paramétrique n'est envisageable que si l'on dispose d'un nombre important d'observations : dans le cas où les échantillons sont de petite taille, on se tournera vers l'approche paramétrique.

L'estimateur de Pickands

Il est défini par la statistique [20] :

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

Il présente l'intérêt d'être valable quelle que soit la distribution des extrêmes (Gumbel, Weibull ou Fréchet). La représentation graphique de cet estimateur en fonction du nombre k d'observations considérées montre un comportement en général très volatil au départ, ce qui nuit à la lisibilité du graphique. De plus, cet estimateur est très sensible à la taille de l'échantillon sélectionné, ce qui le rend peu robuste. Il est donc d'un maniement délicat. On peut noter qu'il est asymptotiquement normal, avec :

$$\sqrt{k} \frac{\hat{\gamma}_{k,n}^P - \gamma}{\sigma(\gamma)} \rightarrow \mathcal{N}(0, 1).$$

Lorsque $k \rightarrow \infty$ l'écart type asymptotique est donnée par :

$$\sigma(\gamma) = \frac{\gamma \sqrt{2^{2\gamma+1} + 1}}{2(2^\gamma - 1) \ln(2)}.$$

Estimateur de Hill

L'estimateur de Hill n'est utilisable que pour les distributions de Fréchet (donc telles que $\gamma > 0$) pour lesquelles il fournit un estimateur de l'indice de queue plus efficace que l'estimateur de Pickands. Il est défini par la statistique suivante [23] :

$$\hat{\gamma}_{k,n}^H = \frac{1}{k-1} \sum_{j=1}^{k-1} \ln\left(\frac{X_{j,n}}{X_{k,n}}\right).$$

Si on choisit $k, n \rightarrow \infty$ de sorte que $\frac{k}{n} \rightarrow 0$ alors on peut montrer que $\lim_{k \rightarrow +\infty} \hat{\gamma}_{k,n}^H = \gamma$ et l'estimateur de Hill est de plus asymptotiquement normal :

$$\sqrt{k} \frac{\hat{\gamma}_{k,n}^H - \gamma}{\hat{\gamma}} \rightarrow \mathcal{N}(0, 1),$$

la convergence étant en loi.

Cet estimateur est l'estimateur du maximum de vraisemblance dans le cas particulier du modèle $S(x) = 1 - F(x) = Cx^{-1/\gamma}$; on reconnaît ici une distribution de Pareto d'indice $\alpha = \frac{1}{\gamma}$.

Dans le cas général du domaine de Fréchet, la fonction de survie est de la forme $S(x) = 1 - F(x) = x^{-1/\gamma} L(x)$ avec L une fonction à variation lente. Cela induit un biais important sur l'estimateur de Hill, qui est donc en pratique d'un maniement délicat. Dans le cas général, la fonction L apparaît comme un paramètre de nuisance de dimension infinie, qui complique l'estimation.

L'estimateur des moments

C'est un estimateur proposé par DEKKERS et al [4]. Comme pour l'estimateur de Hill, il n'est utilisable que pour le domaine d'attraction de Fréchet ($\gamma > 0$).

Cet estimateur est défini par la statistique [23] :

$$\hat{\gamma}_{k,n}^M = 1 + \hat{\gamma}_{k,n}^{(1)} - \frac{1}{2} \left(1 - \frac{(\hat{\gamma}_{k,n}^{(1)})^2}{\hat{\gamma}_{k,n}^{(1)}}\right)^{-1},$$

avec $\hat{\gamma}_{k,n}^{(i)} = \left(\frac{1}{k} \sum_{j=1}^k \ln\left(\frac{X_{j,n}}{X_{k-1,n}}\right)\right)^i$. Cet estimateur est convergent et asymptotiquement gaussien :

$$\sqrt{k} \frac{\hat{\gamma}_{k,n}^H - \gamma}{\sqrt{1 + \gamma^2}} \rightarrow \mathcal{N}(0, 1).$$

Comparaison des différents estimateurs

La figure 1.2 nous montre l'erreur de l'estimation en fonction de la méthode.

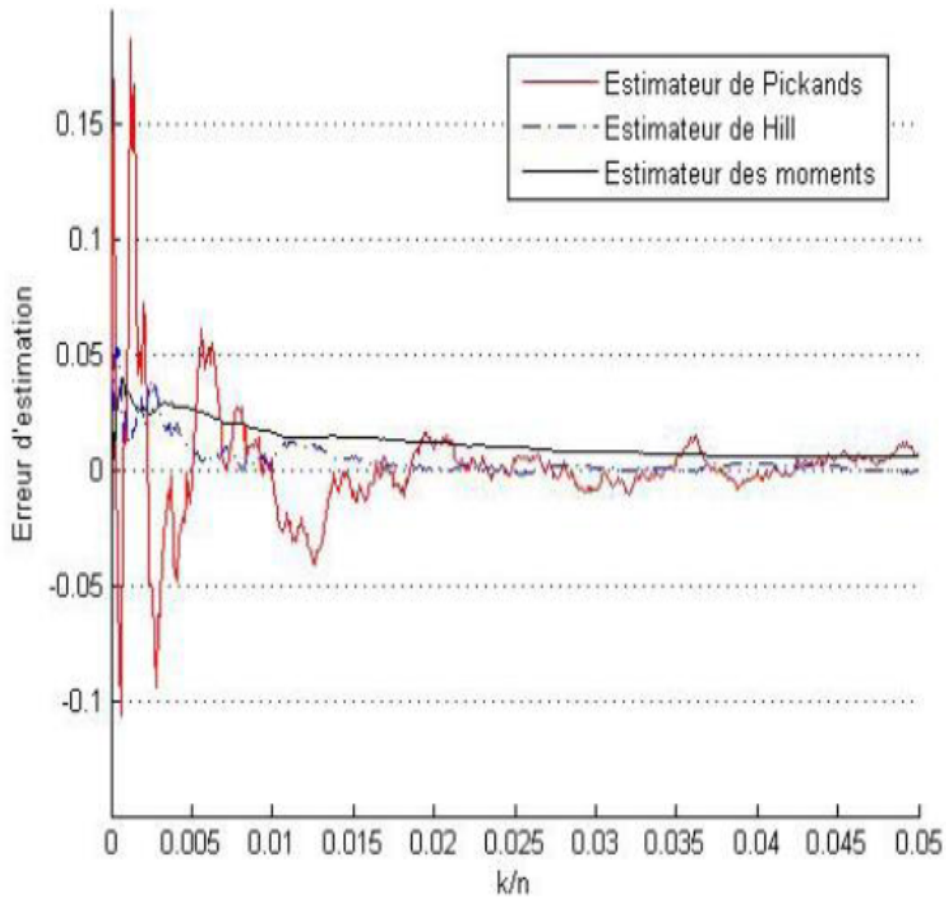


FIG. 1.2 – Comparaison de l’erreur d’estimation en fonction de la méthode [23].

On remarque que l’estimateur de Pickands est moins efficace dans l’estimation de l’indice de queue. Par contre, on observe une efficacité et une suprématie de l’estimateur de Hill sur ceux de Pickands et des moments. Par ailleurs, pour $k < 0.02 \times n$ l’estimateur de Hill est relativement volatile. On serait donc amené à utiliser de l’ordre de 2,5 % des données les plus extrêmes pour estimer l’épaisseur de la queue de distribution.

1.6.2 Méthode du maximum de vraisemblance

La méthode la plus classique est celle du maximum de vraisemblance. L’estimation par le maximum de vraisemblance donne des résultats asymptotiques efficaces, et les estimateurs obtenus convergent sous certaines conditions vers les vraies valeurs des paramètres. Soit X_1, X_2, \dots, X_n un échantillon, de n variables aléatoires supposées indépendantes identiquement distribuées i.i.d, de densité h_θ , où $\theta = (\gamma, \sigma, \mu)$.

L'expression de la fonction de vraisemblance est donnée par :

$$l((X_1; X_2, \dots, X_n); (\gamma, \sigma, \mu)) = \prod_{i=1}^n h_{\theta}(x_i) \text{ avec } \theta = (\gamma, \sigma, \mu).$$

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

$$\begin{cases} \frac{\partial \log l(\theta)}{\partial \theta} = 0, \\ \frac{\partial^2 \log l(\theta)}{\partial^2(\theta)} < 0. \end{cases}$$

Dans le cas où $\gamma = 0$ (loi de Gumbel), la fonction logarithme de vraisemblance est égale à :

$$\log l((\sigma, \mu, 0); (X_1, X_2, \dots, X_n)) = -n \log \sigma - \sum_{i=1}^n \exp\left(-\frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^n \frac{x_i - \mu}{\sigma}.$$

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

$$\frac{\partial \log l}{\partial \sigma} = 0 \Leftrightarrow n + \sum_{i=1}^n \frac{x_i - \mu}{\sigma} [\exp\left(-\frac{x_i - \mu}{\sigma}\right) - 1] = 0.$$

$$\frac{\partial \log l}{\partial \mu} = 0 \Leftrightarrow n - \sum_{i=1}^n \exp\left(-\frac{x_i - \mu}{\sigma}\right) = 0.$$

c'est pas facile de résoudre ce système. En pratique, il est plus aisé d'utiliser des méthodes numériques.

1.6.3 Méthode des moments pondérés

Cette méthode consiste à obtenir les paramètres comme fonctions des moments $\omega_r = \mathbb{E}(M \mathcal{H}_{\gamma, \mu, \sigma}^r(M))$, où M suit une GEV de fonction de répartition $\mathcal{H}_{\gamma, \mu, \sigma}$ pour $r=0,1,2$. En effet, nous pouvons montrer que pour $0 < \gamma < 1$:

$$\omega_r = \frac{1}{r+1} \left(\mu - \frac{\sigma}{\gamma} (1 - (r+1)^\gamma \Gamma(1-\gamma)) \right).$$

Et pour $\gamma = 0$

$$\omega_r = \frac{1}{r+1} (\mu - \sigma(\xi - \ln(r+1))).$$

Γ est la fonction Gamma et ξ est la constante d'Euler. Pour $\gamma \geq 1$, les moments ne sont pas définis (la queue de distribution étant trop épaisse, ils ne sont pas finis).

Hosking, Wallis, et Wood ont montré dans ([13]) que ces quantités pouvaient être approchées par :

$$\hat{\omega}_r = \frac{1}{k} \sum_{j=1}^k \left(\frac{j - 0.35}{k} \right)^r M_{(j)},$$

où $M_{(1)} \leq M_{(j)} \leq M_{(k)}$ sont les maxima ordonnés.

Nous en déduisons alors les valeurs des paramètres :

$$\begin{cases} \frac{3\hat{\omega}_2 - \hat{\omega}_0}{2\hat{\omega}_1 - \hat{\omega}_0} = \frac{3\hat{\gamma} - 1}{2\hat{\gamma} - 1} \\ \hat{\sigma} = \frac{(2\hat{\omega}_1 - \hat{\omega}_0)\hat{\gamma}}{\Gamma(1 - \hat{\gamma})(2\hat{\gamma} - 1)} \\ \hat{\mu} = \hat{\omega}_0 + \frac{\hat{\sigma}}{\hat{\gamma}}(1 - \Gamma(1 - \hat{\gamma})). \end{cases}$$

Ces estimateurs sont définis pour $\gamma < 1$ mais ils ne sont asymptotiquement normaux que pour $\gamma < \frac{1}{2}$.

1.6.4 Méthode des moments pondérés généralisés

Cette méthode vise à généraliser la précédente en étendant son champ d'application. En effet, la méthode des moments pondérés perd de son intérêt lorsque γ dépasse $\frac{1}{2}$ puisqu'il n'y a plus de normalité asymptotique de l'estimateur. Cette généralisation consiste à utiliser les mouvements $v_h = \mathbb{E}(Mh(\mathcal{H}_{\gamma, \mu, \sigma}(M)))$, où M suit une GEV de fonction de répartition $\mathcal{H}_{\gamma, \mu, \sigma}$. h est une fonction définie sur $[0, 1]$ et valant 0 sur les bornes. Ainsi, lorsque M prend une grande valeur, $\mathcal{H}_{\gamma, \mu, \sigma}(M)$ est proche de 1 et la fonction h ramène la valeur à 0. Intuitivement, nous voyons bien que l'impact des queues de distribution va être moins fort et donc que les moments existeront pour des valeurs de γ plus élevées.

Diebolt [6] a montré que si nous choisissons $h(t) = h_{a,b}(t) = t^a(-\log(t))^b$ avec $a > 0$ et $b > 0$ alors v_h vérifie :

$$v_h = \frac{\sigma}{\gamma} \frac{1}{(a+1)^{b-\gamma+1}} \Gamma(b - \gamma + 1) - \left(\frac{\sigma}{\gamma} - \mu \right) \frac{1}{(a+1)^{b+1}} \Gamma(b + 1).$$

En prenant $(a,b)=(1,1),(1,2),(2,1)$, nous pouvons ainsi obtenir des estimateurs des paramètres de la GEV avec les relations suivantes :

$$\frac{\hat{\gamma}}{1 - (\frac{3}{2})^{\hat{\gamma}}} = \frac{2(\hat{h}_{1,1} - \hat{h}_{1,2})}{\hat{h}_{1,1} - 2.25\hat{h}_{2,1}}$$

$$\hat{\sigma} = (\hat{h}_{1,1} - \hat{h}_{1,2}) \frac{2^{3-\hat{\gamma}}}{\Gamma(2 - \hat{\gamma})}$$

$$\hat{\mu} = \frac{\hat{\sigma}}{\hat{\gamma}} - \frac{\hat{\sigma}}{\hat{\gamma}} 2^{\hat{\gamma}} \Gamma(2 - \hat{\gamma}) + 4\hat{h}_{1,1}$$

cette méthode est valable pour $\gamma < \frac{3}{2}$, et la normalité asymptotique est vérifiée pour $\gamma < 1$.

1.7 Distribution des excès (P.O.T)

L'approche par dépassements de seuil, en anglais "Peaks-Over-Threshold approach" notée POT. Cette méthode initialement développée par Pickands (1975) [20] et abondamment étudiée par divers auteurs tels que de Smith (1987), Davison et Smith (1990), ou Reiss et Thomas (2001) pour de plus amples références, repose sur l'utilisation des statistiques d'ordre supérieur de l'échantillon [7]. Elle consiste à ne conserver que les observations dépassant un certain seuil. L'excès au-delà du seuil est défini comme l'écart entre l'observation et le seuil.

Plus précisément, soit un échantillon de variables aléatoires indépendantes, identiquement distribuées X_1, X_2, \dots, X_n de fonction de répartition F et x_F un point terminal. Alors, pour un seuil $u < x_F$ fixé, considérons les n observations x_{i_1}, \dots, x_{i_n} dépassant le seuil u . On définit la variable $Y_j = X_{i_j} - u$, $j=1, \dots, n$, l'excès au-dessus du seuil u .

voir Figure 1.3 ci-dessous :

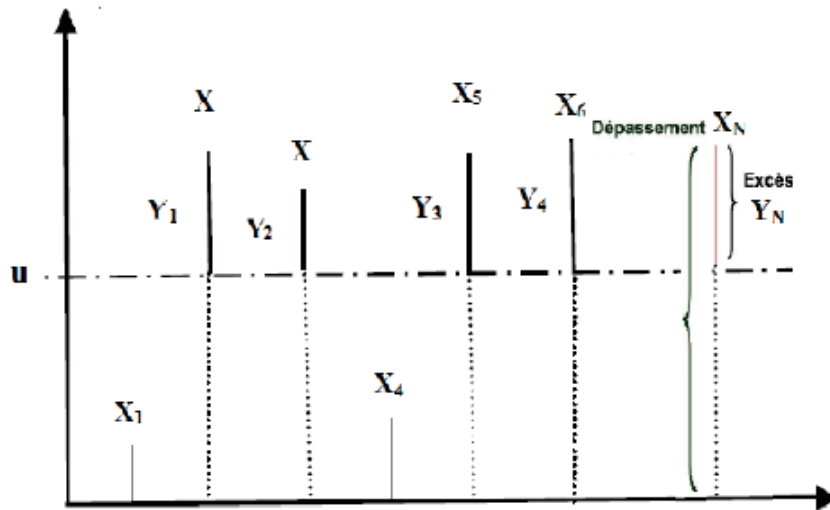


FIG. 1.3 – Dépassement de seuil (POT) [2].

On notera F_u la fonction de répartition de l'excès Y au delà du seuil u . La loi des excès est celle de variables aléatoires i.i.d. admettant pour fonction de répartition :

$$F_u(y) = \mathbb{P}(Y \leq x / X > u),$$

représentant la probabilité que la variable aléatoire X ne dépasse pas le seuil u de au moins une quantité x sachant qu'elle dépasse u . F_u décrit ainsi la loi de Y sachant que $X > u$. On peut la réécrire en fonction de F à l'aide du résultat suivant [7] .

On a pour $x \geq 0$:

$$F_u(x) = \mathbb{P}(X - u \leq x / X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}.$$

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 généralisée que l'on notera GPD (Generalized Pareto Distribution).

1.8 Théorème de Pickands

Théorème 1.8.1. [14] Soit X_1, X_2, \dots, X_n un échantillon de variables aléatoires i.i.d. suivant la loi F . F appartient au domaine d'attraction D si et seulement si il existe une fonction $\sigma(u)$ positive telle que

$$\lim_{x \rightarrow x^*} \sup_{0 < x < x^*} \{ |\mathbb{P}(X - u \leq x / X > u) - H_{\gamma, \sigma(u)}(x)| \} = 0,$$

où $x^* = \sup x$; $F_u < 1$ et $H_{\gamma, \sigma}$ est la distribution de Pareto généralisée (GPD) définie par

$$H_{\gamma, \sigma}(x) = \begin{cases} 1 - (1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0, x \geq 0 \text{ et } x < \frac{-\sigma}{\gamma} \text{ si } \gamma < 0 \\ 1 - \exp(-\frac{x}{\sigma}) & \text{si } \gamma = 0, x \geq 0. \end{cases} \quad (1.3)$$

Ce théorème signifie que si F appartient au domaine d'attraction D alors il existe une fonction $\sigma(u)$ positive et un réel γ tels que la loi des excès F_u peut être uniformément approchée par une distribution de Pareto généralisée (GPD) notée $H_{\gamma, \sigma}$.

1.9 Propriétés de la distribution de pareto généralisée (GPD)

La fonction de répartition ($H_{\gamma, \sigma}$) met en exergue trois paramètres : u, σ, γ . Le paramètre γ , appelé indice des valeurs extrêmes ou indice de queue, renseigne sur le type de la loi asymptotique (Fréchet pour $\gamma > 0$, Gumbel pour $\gamma = 0$ et Weibull pour $\gamma < 0$). En outre, plus γ est élevé en valeur absolue, plus le poids des extrêmes est important. Nous parlons dans ce cas d'une loi à queue épaisse.

La GPD a les propriétés suivantes :

$$\mathbb{E}(X) = \frac{\sigma}{1 - \gamma}, \quad (\gamma < 1) \text{ et } V(X) = \frac{\sigma^2}{(1 - \gamma)^2(1 - 2\gamma)}, \quad \gamma < \frac{1}{2}.$$

L'identification du seuil est la phase la plus déterminante dans l'implémentation de l'approche POT, vu que la qualité du modèle en dépend. C'est à partir de la définition d'un bon seuil u que nous pouvons garantir la convergence des excès vers une GPD, estimer les autres paramètres et évaluer le temps de retour.

La loi des excès au-delà d'un seuil converge vers une loi GPD lorsque la taille de l'échantillon, et donc le seuil, tendent vers l'infini.

L'indice de queue γ de cette distribution est le même que pour la distribution GEV. Le paramètre σ est un indicateur de la taille de la queue à une distance finie. La distribution GPD a l'allure suivante :

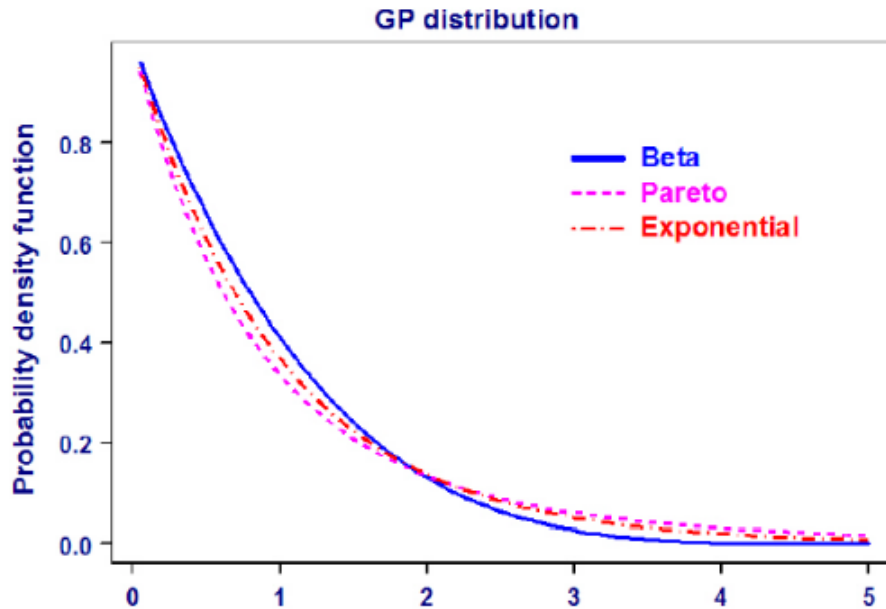


FIG. 1.4 – Distribution GPD [21].

1.9.1 La détection du seuil

Le choix du seuil doit établir un compromis entre biais et variance. Concrètement, le seuil doit être suffisamment grand pour pouvoir utiliser les résultats asymptotiques, mais pas trop élevé pour obtenir des estimations précises. Par contre le choix d'un seuil faible risque de déclarer abusivement des observations extrêmes, introduire un biais dans l'estimation et par conséquent, mal approximer la loi asymptotique. Dans ce sens, plusieurs méthodes de détection du seuil ont été proposées, nous utilisons celle qui est la plus employée à savoir la fonction d'excès en moyenne (mean excess plot ou mean residual life plot (ME-plot)).

Le Mean Excess Plot

Le Mean Excess Plot est le graphe des points $(u, e(u))$ où $e(u)$ est la moyenne des excès au delà du seuil u , définie par :

$$e(u) = \mathbb{E}(Y|X > u).$$

Elle correspond à une fonction de u qui s'exprime à l'aide de 1-F. Nous supposons que pour ce modèle proposé, $\mathbb{E}(X) < \infty$.

En pratique, la fonction des moyennes des excès e est estimée par \hat{e}_n [7] :

$$\hat{e}_n(u) = \frac{\sum_{i=1}^n x_i \mathbf{1}_{(u, \infty)}(x_i)}{\sum_{i=1}^n \mathbf{1}_{(u, \infty)}(x_i)} - u,$$

telle que $\mathbf{1}_{(u, \infty)}$ est égale à 1 si $x_i > u$ et 0 sinon, et nous choisissons la $(k + 1)$ ème grande observation : $u = x_{n-k, n}$ comme seuil, d'où l'estimation

$$\hat{e}(x_{n-k, n}) = \frac{1}{k} \sum_{j=1}^k x_{n-j+1, n} - x_{n-k, n}.$$

Proposition 1.9.1. [7] Si (X_1, \dots, X_{Nu}) suivent une loi $GPD_{\gamma, \sigma}$, alors pour $\gamma < 1$,

$$\mathbb{E}(Y|X > u) = \frac{\gamma}{1-\gamma} u + \frac{\sigma}{1-\gamma}.$$

Dans ce cas, le seuil à retenir u , est celui pour lequel la moyenne des excès est approximativement linéaire.

Plus les queues de la distribution sont épaisses, plus la fonction $e(u)$ tend rapidement vers l'infini.

En effet, la fonction moyenne des excès relative à une distribution de loi Pareto de paramètre $\alpha > 0$, appartenant au domaine d'attraction de Fréchet d'indice des valeurs extrêmes $\gamma = \frac{1}{\alpha}$, s'écrit comme suit : $e(u) = \frac{k+u}{\alpha-1}$.

Alors que pour une distribution de queue moins épaisse, par exemple la distribution de Weibull de paramètres $\lambda > 0$ et $\nu > 0$, appartenant au domaine d'attraction de Gumbel, la fonction moyenne des excès relative à cette loi s'écrit comme suit : $e(u) = \frac{u^{1-\nu}}{\lambda\nu}$.

1.10 Estimation des paramètres de la GPD

1.10.1 Méthode du maximum de vraisemblance

Une fois le seuil optimal choisi, on construit une nouvelle série d'observations au dessus de ce seuil, et la distribution de ces données suit approximativement une distribution généralisée de Pareto. La densité de la distribution GPD s'écrit [21] :

$$H_{\gamma,\sigma}(x) = \begin{cases} \sigma^{\frac{1}{\gamma}}(\sigma + \gamma x)^{-\frac{1}{\gamma}-1} & \text{si } \gamma \neq 0, \\ \sigma^{-1} \exp(-\frac{x}{\sigma}) & \text{si } \gamma = 0. \end{cases} \quad (1.4)$$

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

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

On pose $\tau = \frac{\gamma}{\sigma}$, l'annulation des dérivées partielles des logarithmes de la fonction de vraisemblance conduit au système :

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

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

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

Cette dernière équation se résout numériquement de manière itérative pour autant que l'on dispose d'une valeur initiale τ_0 pas trop éloignée de τ . En pratique cette valeur initiale pourra être obtenue par la méthode des moments ou par la méthode des quantiles .

1.10.2 Méthode des moments pondérés

Définition 1.10.1. [23] On appelle moment pondéré d'ordre r , le moment défini par :

$$\mu_r = \mathbb{E}[Z(\overline{H}_{\gamma,\sigma}(Z))^r], \quad r \in \mathbb{N}.$$

Prenons par exemple Z suit la loi de Paréto généralisée de paramètre (γ, σ) .

Le moment pondéré μ_r est égal à

$$\mu_r = \frac{\sigma}{(r+1)(r+1-\gamma)}, \quad r = 0, 1.$$

En résolvant l'équation précédente pour $r=1$ et $r=2$, on obtient :

$$\hat{\sigma} = \frac{2\mu_0\mu_1}{\mu_0 - 2\mu_1} \quad \text{et} \quad \hat{\gamma} = 2 - \frac{\mu_0}{\mu_0 - 2\mu_1}.$$

On remplace μ_0 et μ_1 par leurs estimateurs des moments empiriques pondérés définis par :

$$M_r = \frac{1}{n} \sum_{j=1}^n \left(\prod_{l=1}^r \frac{n-j-l+1}{n-l} \right) X_{(j,n)} = \frac{1}{n} \sum_{j=1}^n \left(1 - \frac{j}{n+1} \right)^r X_{(j,n)},$$

où $X_{(j,n)}$ est la j ième statistique d'ordre.

On trouve un estimateur de γ par la méthode des moments pondérés

$$\hat{\gamma} = \frac{\frac{1}{n} \sum_{j=1}^n \left(4 \frac{j}{n+1} - 3 \right) X_{(j,n)}}{\frac{1}{n} \sum_{j=1}^n \left(2 \frac{j}{n+1} - 1 \right) X_{(j,n)}}.$$

1.11 Conclusion

Dans ce chapitre, nous avons fait un rappel sur la théorie des valeurs extrêmes, en mentionnant les différentes caractéristiques et les notions de base qui sont très utiles pour l'estimation des quantiles extrêmes et le niveau de retour que nous aborderons dans le chapitre suivant.

Chapitre 2

Estimation de quantile extrême et de niveau de retour

2.1 Introduction

Le choix de la loi pour les valeurs extrêmes et l'estimation de la fonction de répartition en déterminant l'indice de queue de distribution ne sont souvent qu'un objectif intermédiaire, l'objectif réel étant plutôt l'estimation d'un quantile extrême ou d'un niveau de retour ainsi que l'estimation d'une période de retour.

2.2 Quantile

Les quantiles d'une variable aléatoire univariée, discrète (ex : entière) ou continue (réelle), sont les valeurs que prend la variable pour des valeurs de probabilité sous le quantile considéré, valant une valeur remarquable. On les appelle encore fractiles, synonyme complet selon le contexte d'usage, et ce sont les valeurs réciproques de la fonction de répartition de la loi de probabilité considérée.

Les quantiles d'un échantillon statistique de nombres sont des valeurs remarquables permettant de diviser le jeu de ces données ordonnées (i.e. triées) en intervalles consécutifs contenant le même nombre de données [17].

2.2.1 Quantile d'ordre p

Définition 2.2.1. [23] On appelle quantile ou fractile d'ordre p , le nombre x_p défini par :

$$x_p = \inf\{x \in \mathbb{R} : F(x) \geq p\}, \text{ avec } p \in [0, 1]. \quad (2.1)$$

Remarque 2.2.1. Si F est strictement croissante et continue, alors x_p est l'unique nombre réel tel que

$$F(x_p) = p.$$

2.2.2 Fonction quantile de queue

Définition 2.2.2. [23] La fonction de quantile de queue est définie par :

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

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

2.2.3 Quantile extrême

Définition 2.2.3. [23] On appelle quantile extrême le quantile d'ordre $(1 - p)$, défini par

$$\begin{aligned} x_p &= \inf\{x \in \mathbb{R} : F(x) \geq p\} \\ &= F^{-1}(1 - p). \end{aligned}$$

où p est proche de zéro.

2.2.4 Quantile empirique

Définition 2.2.4. [23] La fonction quantile empirique de l'échantillon X_1, \dots, X_n , est donnée par :

$$Q_n(p) = F_n^{-1}(p) = \inf\{x \in \mathbb{R} : F_n(x) \geq p\}, \quad 0 < p < 1. \quad (2.2)$$

La fonction quantile empirique de la queue correspondante est définie par :

$$U_n(t) = F_n^{\leftarrow}\left(1 - \frac{1}{t}\right) \text{ avec } 1 < t < \infty. \quad (2.3)$$

2.3 Niveau de retour et période de retour

Typiquement, quand on considère les valeurs extrêmes d'un échantillon, on est intéressé par leurs niveaux de retour. On peut vouloir déterminer la probabilité d'occurrence d'une probable catastrophe (type tempête, canicule, inondation, etc.), à cette probabilité correspond une période de retour qui est une mesure de récurrence.

2.3.1 Notion de niveaux et périodes de retour

Définition 2.3.1. Pour une variable aléatoire X , le niveau de retour Z , associé à la période de retour t , est la valeur que X dépasse en moyenne une fois au cours de t observations.

Remarque 2.3.1. La période de retour n'a rien à voir avec un phénomène périodique.

Remarque 2.3.2. Pour une série de données finie (de longueur N), le niveau de retour associé à la période $t=N$ est $\max(X)$.

Remarque 2.3.3. D'une manière générale, il est facile de déterminer des niveaux de retour pour des périodes inférieures à la longueur des données (diagrammes de Gumbel) à partir des quantiles.

2.4 Estimation d'un quantile

Différentes méthodes sont utilisées pour estimer un quantile :

- La méthode paramétrique.
- La méthode semi-paramétrique (utilisation du théorème des valeurs extrêmes).
- La méthode non-paramétrique.

Ces méthodes ont été bien présentées dans [17]. L'objectif de notre travail est l'estimation du quantile extrême et du niveau de retour.

2.5 Méthodes d'estimation d'un quantile extrême

Nous estimerons le quantile d'ordre $p_t = 1 - 1/t$ de F , c'est à dire le nombre x_{p_t} défini par $1 - F(x_{p_t}) = 1/t$. Si t est grand, un tel quantile est dit quantile extrême. Plusieurs méthodes d'estimation d'un quantile extrême ont été proposées dans la littérature : principalement la méthode des Blocs (Block Maxima notée par BM), qui modélise la distribution des extrêmes par la loi des extrêmes généralisées (GEV) et la méthode POT qui modélise la

distribution des excès au-dessus d'un seuil élevé par la distribution de Pareto généralisée (GPD), méthode de régression, etc.

2.5.1 Estimation naturelle

Définition 2.5.1. Du fait que $X_{(k,n)}$ est un estimateur naturel du quantile d'ordre $1 - (k - 1)/n$, un estimateur naturel de $F_n^{-1}(p)$, pour $p \in]0, 1[$ est donné par

$$\hat{x}_p = ([pn] - pn + 1)X_{(n-[pn]-1,n)} + ([pn] - pn)X_{(n-[pn],n)} \quad p \in]0, 1[,$$

où $[\cdot]$ désigne la partie entière.

Cette méthode nécessite pour être efficace en pratique une taille de l'échantillon assez grande.

2.5.2 Méthode des blocs

Supposant que l'échantillon de maximums suit exactement une loi GEV, cette méthode est basée sur le théorème de Fisher-Tippett vu précédemment.

Construction des blocs

On divise l'échantillon X_1, X_2, \dots, X_N en k blocs de même taille n (soit $N = n_k$); le j ième bloc est défini par $(X_{(j-1)n+1}, \dots, X_{(j-1)n+n})$, il faut que la taille n et le nombre de bloc k soient suffisamment grands.

On calcule le maximum $M_{n,j}$ du bloc j défini par :

$$M_{n,j} = \text{Max}\{(X_{(j-1)n+1}, \dots, X_{(j-1)n+n})\}, \quad j = \{1, \dots, k\}.$$

Estimation des paramètres de la GEV

Puisque la taille de chaque bloc est supposée grande, alors la condition asymptotique (1.1) du théorème (1.4.1) est vérifiée donc on obtient un échantillon de maximum $(M_{n,1}, \dots, M_{n,k})$ de la GEV.

Il faut aussi que nous ayons un nombre suffisant de maxima pour que l'estimation des paramètres de la GEV soit assez précise.

On propose pour l'estimation des paramètres γ , μ et σ les méthodes vues précédemment : méthode de maximum de vraisemblance, méthode des moments et méthode des moments pondérés.

Estimation du quantile de la GEV

La première méthode utilise un résultat donnant l'expression de la loi asymptotique du maximum d'un échantillon. Cette loi limite est la loi des valeurs extrêmes (GEV).

Nous estimons alors le quantile extrême en inversant la fonction de répartition de la loi des valeurs extrêmes et en estimant les paramètres de cette loi. L'estimateur de quantile extrême obtenu par cette méthode s'écrit sous la forme :

$$\hat{x}_{p_t}^{GEV} = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}} \{1 - [-\log(p_t)]\}^{-\hat{\gamma}} & \text{si } \gamma > 0 \\ \hat{\mu} - \hat{\sigma} \log[-\log(p_t)] & \text{si } \gamma = 0, \end{cases} \quad (2.4)$$

où $\hat{\gamma}$, $\hat{\mu}$, $\hat{\sigma}$ sont les estimateurs des paramètres de la loi des valeurs extrêmes \mathcal{H}_γ :

$$\mathcal{H}_\gamma(x) = \exp\left\{-\left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-1/\gamma}\right\}, \quad \forall \gamma \in \mathbb{R}, \quad \forall \mu \in \mathbb{R}, \quad \text{et } \forall \sigma > 0, \quad (2.5)$$

définie pour tout $x > \mu - \sigma/\gamma$.

– En utilisant l'estimateur de Hill

Considérons les fonctions de répartition appartenant au domaine d'attraction maximal de Fréchet ($\gamma > 0$), alors la fonction de survie peut se mettre sous forme

$$\bar{F} = x^{\frac{-1}{\gamma}} L(x),$$

où L est une fonction à variation lente.

On peut donc écrire que :

$$\frac{\bar{F}(x)}{\bar{F}(X_{n-k})} = \frac{\mathcal{L}(x)}{\mathcal{L}(X_{n-k})} \left(\frac{x}{X_{n-k}}\right)^{\frac{-1}{\gamma}}.$$

Et en considérant que le rapport des fonctions à variation lente et proche de 1, on trouve :

$$\bar{F}(x) \sim \bar{F}(X_{n-k}) \left(\frac{x}{X_{n-k}}\right)^{\frac{-1}{\gamma}}.$$

On déduit de cette expression un estimateur de $F(x)$ pour $x > X_{(n-k)}$:

$$\bar{F}(x) = 1 - \frac{k}{n} \bar{F}(X_{n-k}) \left(\frac{x}{X_{n-k}}\right)^{-1/\gamma_{n,k}^t}.$$

$$\widehat{x}_p^{\mathcal{H}} = X_{(n-k)} \left(\frac{n}{k} (1-p) \right)^{-\widehat{\gamma}_{n,k}^{\mathcal{H}}}. \quad (2.6)$$

– **En utilisant l'estimateur des moments**

Définition 2.5.2. [24] L'estimateur des moments est donné par :

$$\widehat{x}_p^M = \frac{a_n^{\widehat{\gamma}_{(k,n)}^M} - 1}{\widehat{\gamma}_{(k,n)}^M} \cdot \frac{X_{(n-k,n)} \mathcal{H}_{k,n}}{\rho_1(\widehat{\gamma}_{(k,n)}^M)} + X_{(n-k,n)},$$

où $a_n = \frac{k}{n(1-p)}$,

$$\rho_1(\gamma) = \begin{cases} 1 & \text{si } \gamma \geq 0 \\ (1-\gamma)^{-1} & \text{si } \gamma < 0. \end{cases} \quad (2.7)$$

Un estimateur pour le point terminal $\widehat{x}_F = X_{(n-k,n)} \mathcal{H}_{k,n} \left(1 - \frac{1}{\widehat{\gamma}_M} \right) + X_{(k-n,n)}$.

2.5.3 Méthode des excès ou P.O.T.

Cette méthode assure que la loi des observations qui dépassent un seuil u peut être approchée, pour u grand, par une loi de Pareto généralisée GPD. Le quantile extrême est alors estimé en inversant la fonction de répartition de la loi de Pareto généralisée et en estimant les paramètres de cette loi à l'aide des observations supérieures au seuil u . Voir la formule (1.3).

$$F_u(x) = \frac{F(u+x) - F(u)}{1 - F(u)} \quad \text{si } x \geq 0 \quad \Leftrightarrow \quad F_u(x-u) = 1 - \frac{1 - F(x)}{1 - F(u)} \quad \text{si } x \geq u \quad (2.8)$$

$$\Leftrightarrow \quad \overline{F}(x) = \overline{F}(u) \overline{F}_u(x-u) \quad \text{si } x \geq u. \quad (2.9)$$

(avec $\overline{F} = 1 - F$).

Sachant que

$$\overline{F}(u) = 1 - \mathbb{P}[X \leq u] = \mathbb{P}[X > u] = \frac{N_u}{n}, \quad (2.10)$$

et

$$\widehat{\overline{F}}_u \approx 1 - H_{\widehat{\gamma}, \widehat{\sigma}} \quad \text{pour } u \text{ assez grand,} \quad (2.11)$$

où $H_{\widehat{\gamma}, \widehat{\sigma}}$ est la GPD définie en (1.3), N_u est le nombre d'observations au-dessus du seuil u et $\widehat{\gamma}$ et $\widehat{\sigma}$ sont les estimateurs des paramètres de la loi GPD, l'estimateur de \overline{F} peut

alors s'écrire [3] :

$$\widehat{F}(x) = \frac{N_u}{n} \left(1 + \widehat{\gamma} \frac{x - u}{\widehat{\sigma}}\right)^{-\frac{1}{\widehat{\gamma}}}, \quad \forall \widehat{\gamma} \neq 0. \quad (2.12)$$

Par inversion, l'estimateur obtenu par cette méthode s'écrit sous la forme [3] :

$$\widehat{x}_{p_t}^{GPD} = u + \frac{\widehat{\sigma}}{\widehat{\gamma}} \left[\left(\frac{n}{N_u} (1 - p_t) \right)^{-\widehat{\gamma}} - 1 \right], \quad \forall \widehat{\gamma} \neq 0. \quad (2.13)$$

Remarque 2.5.1. Cette méthode présente un avantage par rapport à la méthode des blocs, en ce sens qu'il est plus facile d'avoir un échantillon d'excès que de maxima. Dans la pratique, on remplace u par $X_{(n-k+1,n)}$ qui représente la k plus grande observation de l'échantillon X_1, \dots, X_n .

2.6 Estimation de niveau de retour

Nous définissons le niveau de retour comme la valeur x_t telle que nous espérons détecter en moyenne un seul dépassement de cette quantité au bout de t périodes, c'est à dire :

$$\mathbb{E} \left(\sum_{i=1}^t \mathbf{1}_{(X_i > x_t)} \right) = 1 \Leftrightarrow \mathbb{P}[X_i > x_t] = 1/t \quad \text{avec } i = 1, \dots, t \quad (2.14)$$

$$\Leftrightarrow 1 - F(x_t) = 1/t, \quad (2.15)$$

où $\mathbf{1}$ représente la fonction indicatrice.

Nous remarquons que l'estimation d'un niveau de retour d'ordre t revient à l'estimation d'un quantile extrême d'ordre $p_t = 1 - 1/t$. Comme il est situé au dessus de l'observation maximale avec une probabilité qui tend vers 1 lorsque $t \rightarrow \infty$ on ne peut pas, comme pour l'estimation de quantiles classiques, inverser tout simplement la fonction de répartition empirique. Il faut donc, à l'aide des plus grandes observations, estimer la fonction de répartition au-delà de l'observation maximale.

2.7 Estimation de la période de retour

Supposons maintenant que nous disposions d'un quantile x_{p_T} et que nous désirions estimer la période de retour T , associée à la probabilité d'excès au delà de x_{p_T} . Cette

période s'interprète comme une moyenne d'unités de temps séparant un événement de grandeur donnée x_{p_T} d'un second événement d'une grandeur égale ou supérieure.

Ainsi, estimer une période de retour T d'un quantile x_{p_t} revient à estimer l'ordre de ce quantile $\hat{p}_T = 1 - 1/\hat{T}$.

Selon (2.6), nous estimons l'ordre \hat{p}_T^{GEV} du quantile x_{p_T} pour la loi des valeurs extrêmes GEV par :

$$\hat{p}_T^{GEV} = \begin{cases} \exp\{-[1 + \hat{\gamma}(\frac{x_{p_T} - \hat{\mu}}{\hat{\sigma}})]^{-\frac{1}{\hat{\gamma}}}\} & \text{si } \hat{\gamma} \neq 0 \\ \exp\{-\exp(-\frac{x_{p_T} - \hat{\mu}}{\hat{\sigma}})\} & \text{si } \hat{\gamma} = 0 \end{cases} \quad (2.16)$$

$$\forall \hat{\mu} \in \mathbb{R} \text{ et } \forall \hat{\gamma} > 0$$

$$\begin{aligned} \hat{T}^{GEV} &= \begin{cases} (1 - \exp\{-[1 + \hat{\gamma}(\frac{x_{p_T} - \hat{\mu}}{\hat{\sigma}})]^{-\frac{1}{\hat{\gamma}}}\})^{-1} & \text{si } \hat{\gamma} \neq 0 \\ (1 - \exp\{-\exp(-\frac{x_{p_T} - \hat{\mu}}{\hat{\sigma}})\})^{-1} & \text{si } \hat{\gamma} = 0 \end{cases} \\ &= \frac{1}{1 - \mathcal{H}_{\hat{\gamma}}(x_{p_T})}, \forall \hat{\gamma} \in \mathbb{R}, \forall \hat{\mu} \in \mathbb{R} \text{ et } \forall \hat{\gamma} > 0. \end{aligned}$$

De même, selon (2.13) et (1.3), nous estimons la période de retour pour la loi des valeurs extrêmes GPD par [3] :

$$\hat{T}_r^{GPD} = \frac{n}{N_u} \frac{1}{1 - H_{\hat{\gamma}, \hat{\sigma}}(x_{p_T})}, \quad \forall \hat{\gamma} \in \mathbb{R} \text{ et } \forall \hat{\sigma} > 0. \quad (2.17)$$

2.8 Conclusion

Dans ce chapitre nous avons présenté quelques définitions nécessaires afin d'aborder l'estimation de quantile extrême et de niveau de retour. Puis nous avons cité les différentes méthodes (estimation naturelle, méthode des blocs, méthode des excès ou P.O.T). Enfin, nous avons présenté l'estimation de la période de retour.

Chapitre 3

Application en pluviométrie

3.1 Introduction

Dans cette partie, nous avons appliqué la méthode des excès (POT) dans le domaine de pluviométrie. Nous avons en effet développé une méthodologie à l'aide de cette théorie, permettant le calcul de quantiles extrêmes et la période de retour.

3.2 Application numérique sur un échantillon des données réelles

Comme il est déclaré dans le journal La Petite Gironde du 7 mars 1930, les débits d'une rivière comme la Garonne sont très variables. Le jeudi matin à 10 heures, Cadillac a connu la plus grosse inondation du siècle. Les eaux ont atteint 11 m.77, soit 30 centimètres de plus qu'en 1875, ce qui signifie une vraie catastrophe naturelle (événement rare) et beaucoup de dégâts humains et matériels.

On dispose généralement de quelques enregistrements des débits passés de la rivière. Dans le tableau suivant, on a enregistré le maximum de débit journalier qui s'est écoulé au cours de chacune des années de 1913 à 1977. Les données x sont formées de la collection (jour de la mesure j , $x_j =$ débit enregistré en m^3/s).

année	Max	année	Max	année	Max	année	Max	année	Max
1913	4579	1926	3200	1939	2800	1952	6721	1965	4968
1914	4774	1927	6332	1940	5553	1953	2700	1966	5163
1915	4968	1928	4968	1941	5163	1954	3000	1967	2600
1916	4774	1929	1950	1942	3100	1955	5747	1968	2530
1917	3400	1930	7500	1943	3600	1956	2300	1969	4073
1918	6137	1931	3700	1944	4579	1957	3200	1970	3120
1919	4189	1932	3600	1945	3200	1958	2900	1971	4696
1920	4579	1933	2500	1946	950	1959	4968	1972	5377
1921	2800	1934	3700	1947	1850	1960	3400	1973	3956
1922	4384	1935	6137	1948	2000	1961	4774	1974	4228
1923	5747	1936	4189	1949	1900	1962	2300	1975	3200
1924	3200	1937	5747	1950	2600	1963	2700	1976	4209
1925	3100	1938	3200	1951	2900	1964	3300	1977	4482

TAB. 3.1 – Débits annuels maximaux (en m^3/s) de la Garonne à Mas d’Agenais sur la période 1913-1977 [19].

Dans l’objectif de protéger ce site contre le débordement d’une rivière et de connaître la période de retour de cet évènement rare, on veut construire une digue de protection contre les crues, jusqu’à quelle hauteur h construire cette digue ? et quand est ce que cet évènement se reproduira ?

Plusieurs modélisations des valeurs extrêmes (ici les débits d’une rivière comme la Garonne) sont possibles :

- Par les lois des valeurs extrêmes généralisées (Fréchet, Gumbel et Weibull) dont on peut estimer les paramètres par maximum de vraisemblance ou par moments pondérés. Les hypothèses d’application de ces modèles ne sont pas toujours vérifiées. En dehors de la loi de Weibull, estimer des niveaux de retour à plus de 50 ans donne des valeurs inexploitable.
- Par la méthode des dépassements de seuil (POT) et la loi de Pareto généralisée. Le choix du seuil n’est pas simple et les hypothèses sont contraignantes. On a pu constater des résultats différents selon les méthodes employées pour le même phénomène.

Cela montre que cette modélisation, nécessaire en raison du faible nombre de données extrêmes, implique une mise en oeuvre délicate et une grande prudence dans l'exploitation des résultats, une validation exacte ne semblant pas possible.

Méthode POT (Peaks Over Threshold)

C'est une méthode qui est basée sur l'approximation de la distribution des excès par la loi de Pareto généralisée.

La méthode consiste à suivre les étapes suivantes :

Détermination de seuil avec la méthode Mean Excess Plot.

Estimation des paramètres de la GPD par la méthode de maximum de vraisemblance.

L'estimation de quantile extrême et de la période de retour.

La démarche de l'application est la suivante :

- L'ajustement par la loi normale.
- Détermination de seuil avec la méthode Mean Excess Plot.
- L'approximation de loi des excès par une distribution de pareto généralisée (GPD).
- Estimation des paramètres de la GPD par la méthode de maximum de vraisemblance.
- L'estimation de quantile extrême et de la période de retour.

La réalisation de cette application nécessite de faire appel à des commandes et des fonctions prédéfinies sous le logiciel **R**.

R est un logiciel de traitement statistique des données et un langage de programmation. Il fonctionne sous la forme d'un interpréteur de commandes. Il dispose d'une bibliothèque très large de fonctions statistiques, d'autant plus large qu'il est possible d'en intégrer de nouvelles par le système des "packages", des modules externes compilés (sous forme de DLL sous Windows) que l'on peut télécharger gratuitement sur internet. **R** propose également une palette étendue de fonctionnalités graphiques. Il est possible d'utiliser **R** en mode interactif sans jamais avoir à programmer.

De plus il est libre, donc gratuit. **R** fonctionne sur tout ordinateur (Mac, PC, Windows, Linux), **R** est un logiciel d'excellente qualité, utilisé notamment par de nombreux chercheurs en statistique.

3.2.1 L'ajustement par la loi normale

L'ajustement par la loi normale en utilisant le test de Shapiro Wilk

On teste les hypothèses suivantes :

H_0 : "L'échantillon suit une loi normale" contre H_1 : "L'échantillon ne suit pas une loi normale".

On fixe un seuil de signification $\alpha = 0.05$.

Soit le vecteur :

```
donnée=c(4579 ,4774 ,4968 ,4774 ,3400 ,6137 ,4189 ,4579 ,2800 ,4384 ,5747 ,3200 ,3100
,3200 ,6332 ,4968 ,1950 ,7500 ,3700 ,3600 ,2500 ,3700 ,6137 ,4189 ,5747 ,3200 ,2800 ,5553
,5163 ,3100 ,3600 ,4579 ,3200 ,950 ,1850 ,2000 ,1900 ,2600 ,2900 ,6721 ,2700 ,3000 ,5747
,2300 ,3200 ,2900 ,4968 ,3400 ,4774 ,2300 ,2700 ,3300 ,4968 ,5163 ,2600 ,2530, 4073 ,3120
,4696 ,5377 ,3956 ,4228 ,3200 ,4209 ,4482 )
```

La commande utilisée pour tester la normalité sous **R** est :

```
> shapiro.test(rnorm(65,mean=3910.169,sd=1344.739))
```

le résultat :

W = 0.9856, p-value = 0.6542.

Décision du test :

L'ajustement des données de notre échantillon par la loi normale donne une p-valeur (valeur renvoyée par logiciel **R**) égale à 0.6542. Cette valeur est supérieure au seuil de signification $\alpha=0.05$, donc l'échantillon suit une loi normale.

La représentation graphique de l'échantillon étudié est donnée dans la Figure 3.1.

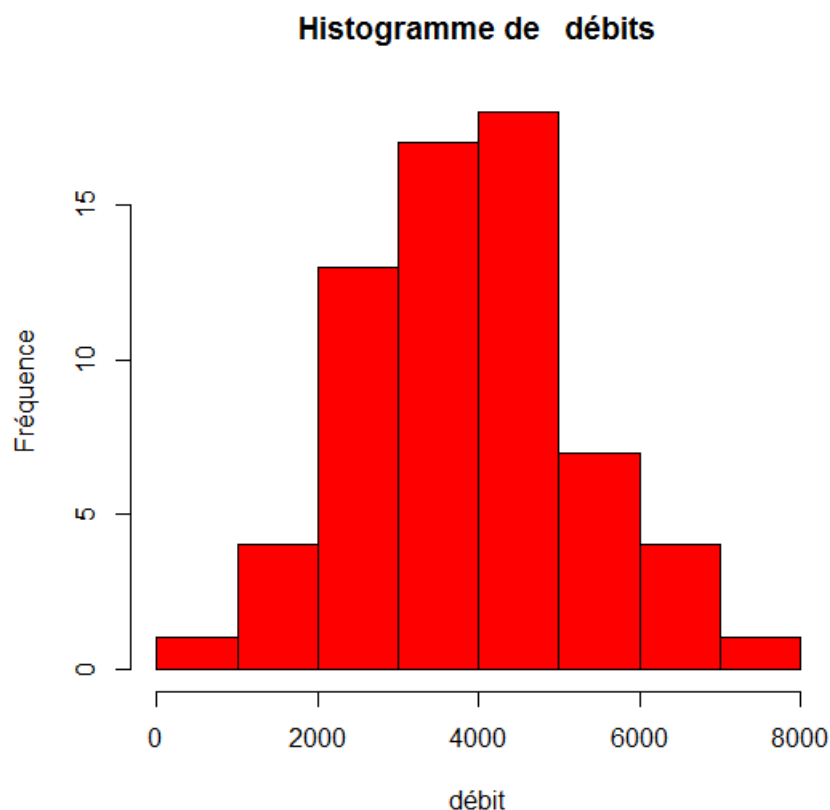


FIG. 3.1 – L’histogramme de la fréquence des débits annuels maximaux.

3.2.2 Détection du seuil

Dans un cadre statistique, le choix du seuil u est très important car il induit une grande variabilité dans l’estimation des quantiles extrêmes et des paramètres de la loi des excès. Il existe différentes approches pour le choix du seuil u de la méthode POT. En effet, le seuil doit être suffisamment grand pour satisfaire la caractéristique asymptotique du modèle, mais pas trop élevé non plus, afin de garder un nombre d’excès suffisant pour estimer convenablement les paramètres du modèle. Généralement, u est déterminé graphiquement par le graphe de la fonction moyenne des excès (Mean Excess Plot), voir la Figure 3.2 :

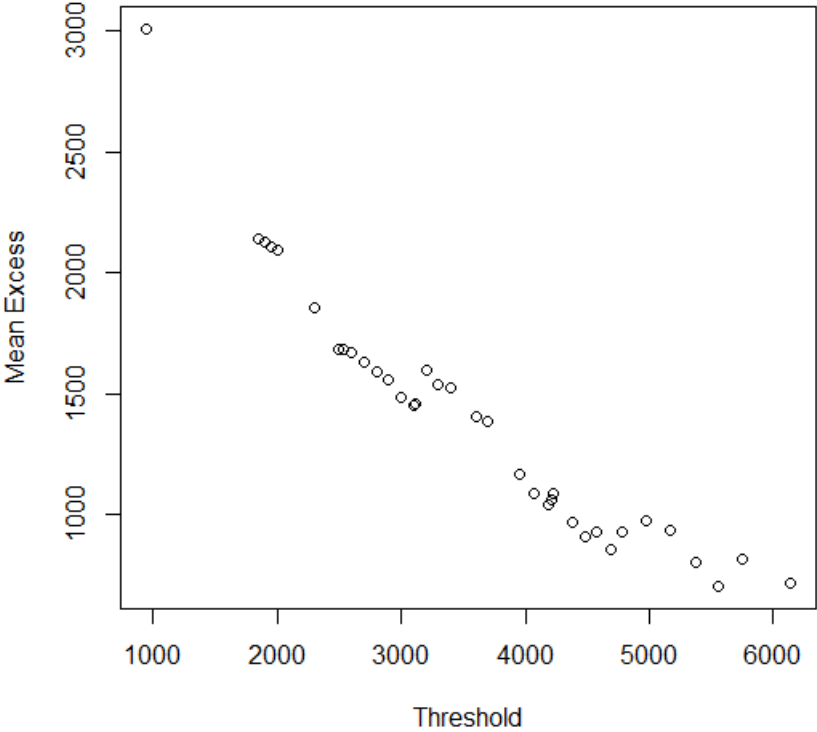


FIG. 3.2 – La distribution moyenne des excès.

D’après le graphe, le seuil a la valeur $u_0 = 4800$ tel que le nuage de points $(u, e(u))$ soit approximativement linéaire pour $u > u_0$.

La Figure 3.3 représente les débits maximaux qui ont dépassé le seuil $u_0 = 4800$.

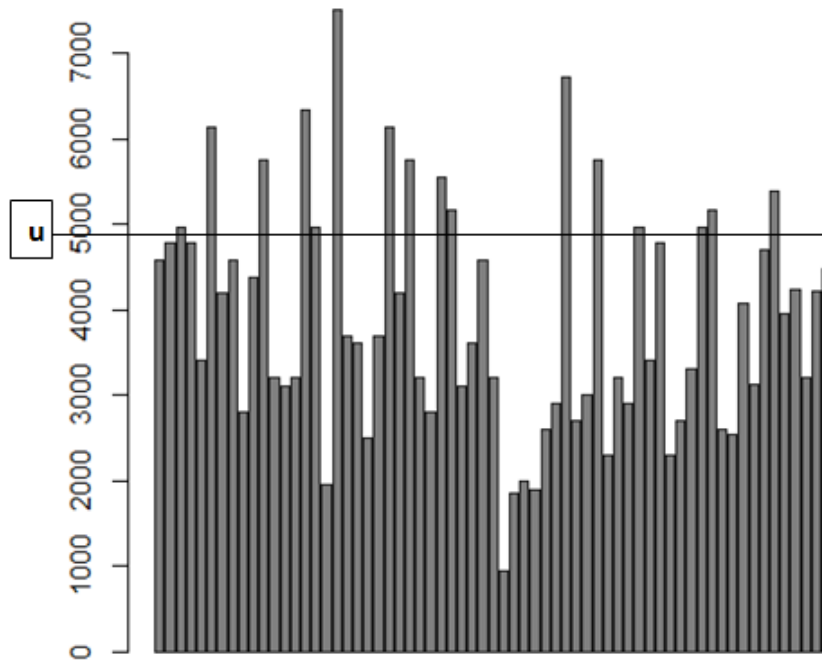


FIG. 3.3 – Dépassement de seuil.

3.2.3 L'approximation de la loi des excès par une distribution de Pareto Généralisée (GPD)

Cette méthode s'appuie sur le fait que la distribution des excès au-delà d'un seuil $u_0 = 4800$ peut être approchée par une loi GPD généralisée (Figure 3.4)

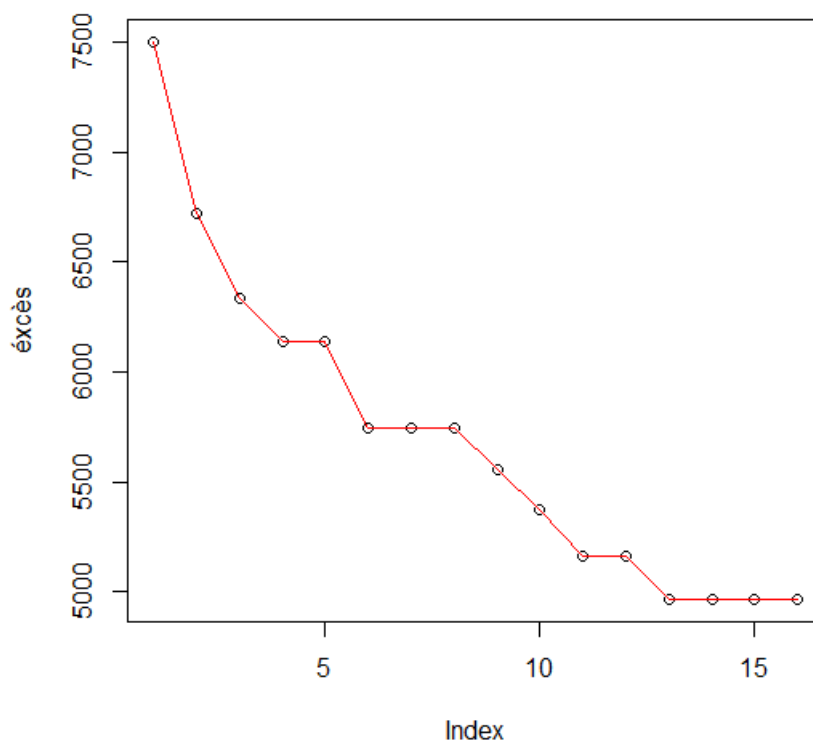


FIG. 3.4 – L'approximation des excès par une GPD.

Interprétation :

D'après le graphe, on remarque que la distribution des excès converge vers une distribution de Pareto Généralisée (l'allure d'une GPD).

3.2.4 Estimation des paramètres de la GPD par la méthode de maximum de vraisemblance

Nous estimons les paramètres γ et σ de la GPD par la méthode de maximum de vraisemblance et nous obtenons le résultat suivant :

La commande utilisée est :

```
> V=gpd(excès,4800,method="ml")
```

Les paramètres estimés sont :

$\hat{\gamma}$	$\hat{\sigma}$
0.2486602	416.9053757

TAB. 3.2 – Les paramètres obtenus (cas de l'échantillon réel).

La loi GPD avec les paramètres estimés

$$H_{\hat{\gamma}, \hat{\sigma}}(x_{p_T}) = 1 - \left(1 + \frac{0.2486602 x_{p_T}}{416.9053757}\right)^{\frac{-1}{0.2486602}} \quad \text{si } \gamma \neq 0.$$

3.2.5 L'estimation de quantile extrême et de la période de retour

Il existe plusieurs approches pour estimer le quantile extrême. Celle basée sur la théorie des valeurs extrêmes est donné par :

$$\hat{x}_{p_t}^{GEV} = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}} \{1 - [-\log(p_t)]\}^{-\hat{\gamma}} & \text{si } \gamma > 0 \\ \hat{\mu} - \hat{\sigma} \log[-\log(p_t)] & \text{si } \gamma = 0, \end{cases}$$

et celle basée sur la méthode des excès et l'approximation de la loi GPD à savoir :

$$\hat{x}_{p_t}^{GPD} = u + \frac{\hat{\sigma}}{\hat{\gamma}} \left[\left(\frac{n}{N_u} (1 - p_t) \right)^{-\hat{\gamma}} - 1 \right], \quad \forall \hat{\gamma} \neq 0.$$

Nous avons utilisé cette dernière pour estimer le quantile extrême :

Avec :

$u=4800$ le seuil.

$\hat{\sigma}=416.9053757$.

$\hat{\gamma}= 0.2486602$.

n : la taille de l'échantillon.

N_u : le nombre d'observation des excès.

p_t : l'ordre de quantile extrême ($0 < p_t < 1$) fixé à 0.90 .

Le résultat obtenu est :

$$\hat{x}_{p_t}^{GPD} = 5202.45.$$

Pour trouver la période de retour on utilise la formule suivante :

$$\widehat{T}^{GPD} = \frac{n}{N_u} \frac{1}{1 - H_{\widehat{\gamma}, \widehat{\sigma}}(x_{p_T})}, \quad \forall \widehat{\gamma} \in \mathbb{R} \text{ et } \forall \widehat{\sigma} > 0.$$

avec $H_{\widehat{\gamma}, \widehat{\sigma}}(x_{p_T})$ est donnée par :

$$H_{\widehat{\gamma}, \widehat{\sigma}}(x_{p_T}) = 1 - \left(1 + \frac{\widehat{\gamma} x_{p_T}}{\widehat{\sigma}}\right)^{\frac{-1}{\widehat{\gamma}}} \text{ si } \gamma \neq 0.$$

Le résultat obtenu est :

$$\widehat{T}_r^{GPD} = 40.625.$$

3.2.6 Interprétation des résultats

Dans cette partie nous avons appliqué la méthode P.O.T sur des données réelles c'est-à-dire, sur le maximum des débits journaliers qui s'est écoulé au cours de chacune des années de 1913 à 1977. Après avoir détecté le seuil et estimé les paramètres de la loi GPD, on passe à l'estimation du quantile extrême et la période de retour. Cette dernière qui vaut approximativement 40.625 ans est l'objectif principal de cette étude vu qu'on cherche quand est ce que le débordement de cette rivière apparaîtra dans le futur. L'estimation de la période de retour nous permettra de conclure que le débordement de cette rivière sera approximativement dans les 40.625 années qui viendront.

3.3 Application numérique sur un échantillon simulé

Simuler un échantillon de 65 valeurs suivant une loi normale (mean=3910.169, sd=1344.739) correspondant à 65 années d'observation d'une hauteur d'eau d'une rivière la Garonne.

La commande utilisée sous **R** est :

```
> rnorm(65, mean=3910.169, sd=1344.739)
```

L'échantillon simulé est :

3665.7568 3598.8677 3232.0131 1293.3543 2638.8463 5865.7270 2404.6775
3119.1860 5073.7616 4095.4834 4755.1030 4100.7288 4251.7834 1929.4935
3479.6197 3806.6131 4966.3864 2726.9952 4003.7482 6141.6024 5470.4770
603.4675 3636.3544 1469.5908 6146.5412 4383.7920 3056.3514 4353.3080
2743.8878 2649.0200 3554.9541 4532.1190 5505.9226 3493.4811 5430.1153
4593.9886 1893.8286 3150.3124 4736.6524 3914.4367 3810.3000 3969.8949
4424.3849 943.9291 4315.6895 4527.2690 3290.9515 4992.1394 2854.8060
4584.5635 5217.1914 4197.0907 4437.8413 5244.3303 4551.4311 3102.2261
2063.9172 5625.9011 2599.8669 4429.5143 2300.5323 3954.5815 4319.8935
4616.4994 3002.6329

3.3.1 Sélection de seuil

L'estimation graphique de u_0 n'est pas aisée dans la pratique, même lorsque nous travaillons avec des données simulées. Nous représentons sur le graphique 3.5 la fonction moyenne des excès (Mean excess plot) :

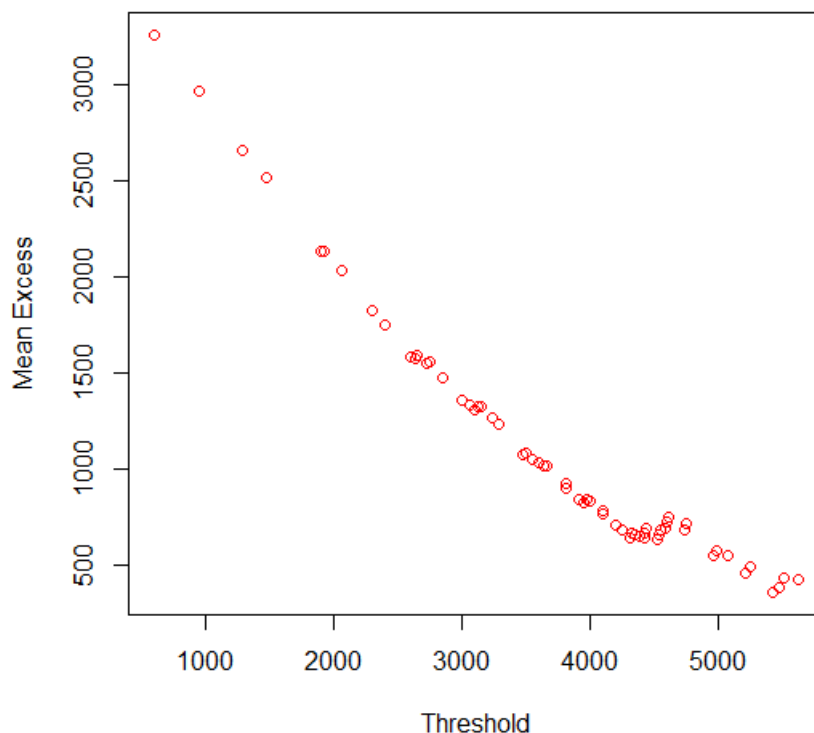


FIG. 3.5 – Mean Excess Plot d'échantillon simulé.

La méthode du maximum de vraisemblance permet d'ajuster la loi GPD. Une difficulté réside cependant dans le choix de la valeur du seuil u_0 . En effet, un biais va apparaître si le seuil est trop petit et on perdra de l'information si le seuil est trop grand. Aussi, est-il préconisé de choisir comme seuil la valeur pour laquelle la fonction moyenne des excès devient approximativement linéaire. Donc d'après le graphique le seuil $u_0 = 4500$.

Remarque 3.3.1. La distribution des excès converge toujours vers une distribution de Pareto Généralisée.

3.3.2 Estimation des paramètres de la GPD par la méthode de maximum de vraisemblance

La loi GPD, a été ajustée par la méthode du maximum de vraisemblance à partir de 20 valeurs dépassant le seuil, parmi les 65 valeurs simulées.

Les paramètres estimés sont :

$\hat{\gamma}$	$\hat{\sigma}$
0.3405395	403.1737622

TAB. 3.3 – Les paramètres obtenus (cas de l'échantillon simulé).

3.3.3 L'estimation de quantile extrême

L'adéquation des données à la loi GPD permet d'estimer un quantile extrême, et d'une période de retour, comme stratégie minimale, sensible à la taille des échantillons, et de prévoir des coûts aléatoires dont la probabilité d'occurrence est très faible. Alors le quantile extrême estimé est :

$$\hat{x}_{pt}^{GPD} = 5051.923532.$$

3.3.4 L'estimation de la période de retour

Pour trouver la période de retour on utilise toujours la formule suivante :

$$\hat{T}^{GPD} = \frac{n}{N_u} \frac{1}{1 - H_{\hat{\gamma}, \hat{\sigma}}(x_{pT})}, \quad \forall \hat{\gamma} \in \mathbb{R} \text{ et } \forall \hat{\sigma} > 0.$$

avec $H_{\hat{\gamma}, \hat{\sigma}}(x_{pT})$ est donnée par :

$$H_{\hat{\gamma}, \hat{\sigma}}(x_{pT}) = 1 - \left(1 + \frac{\hat{\gamma} x_{pT}}{\hat{\sigma}}\right)^{-\frac{1}{\hat{\gamma}}} \text{ si } \gamma \neq 0.$$

D'où la période de retour de l'échantillon simulé est égale à :

$$\hat{T}_s^{GPD} = 32.5.$$

3.3.5 Interprétation des résultats

Cette partie basée sur la simulation d'un échantillon de taille 65 de loi normale (la taille de l'échantillon simulé est équivalente à l'échantillon réel) était faite à cause l'indisponibilité des données réelles et la difficulté qu'on a rencontré dans la récolte de ces dernières. On

veut donc connaître l'utilité de la simulation dans des situations pareilles, pour cela on a réalisé cette étude pour comparer avec l'étude faite sur des données réelles.

La même démarche est suivie avec l'échantillon simulé, on a mis en pratique la méthode P.O.T dans le but d'estimer la période de retour de cet évènement rare (le débordement de la rivière Garonne). Le résultat était proche de l'étude réelle la période de retour de ce débordement est approximativement 32 ans.

3.4 Conclusion

Prévoir certains évènements ou comportements, à partir d'une étude des valeurs extrêmes d'une série, est donc un des principaux objectifs pour ceux qui tentent d'appliquer la théorie des valeurs extrêmes.

Dans cette application nous avons essayé d'estimer la période de retour d'un évènement rare (le débordement d'une rivière), en effectuant deux études comparatives une sur des données réelles et une autre sur des données simulées, on a pu grâce à la première étude de prédire le débordement de cette rivière qui ne sera qu'après 40 ans approximativement. La deuxième étude était faite sur l'échantillon simulé on a estimé aussi la période de retour qui est proche de l'estimation réalisée sur l'échantillon réel (32 ans approximativement).

On peut conclure à travers cette application qu'on peut compter sur la simulation en cas de difficulté dans la récolte des données.

Conclusion générale

Dans ce travail, nous avons introduit la notion des valeurs extrêmes et nous avons énoncé les principaux résultats les concernant, tout en mentionnant les deux principaux outils servant à modéliser le comportement des valeurs extrêmes : la loi des valeurs extrêmes et la loi des excès.

Ensuite, nous avons présenté des méthodes pour estimer le quantile extrême, le niveau de retour et la période de retour.

Enfin, nous avons mis en pratique la méthode P.O.T pour estimer le quantile extrême et la période de retour dans le domaine de pluviométrie.

Nous avons effectué une étude sur le maximum de débit journalier d'une rivière qui s'est écoulé au cours des années de 1913 à 1977 dans l'objectif de prédire la période de retour de cet évènement et de construire une digue de protection, donc éviter des dégâts matériels et humains. Nous avons réalisé cette étude sur un échantillon de données réelles et un échantillon de données simulées de loi normale.

Nous avons tout d'abord appliqué l'approche P.O.T sur le premier échantillon de données réelles. Après avoir détecté le seuil u en utilisant la méthode Mean Excess Plot, nous avons fait une approximation de loi des excès par une distribution de Pareto généralisée (GPD) et nous avons estimé les paramètres de cette loi (par la méthode du maximum de vraisemblance). Dans cette étape, nous avons obtenu les paramètres estimés comme suit : $\hat{\gamma} = 0.2486602$ et $\hat{\sigma} = 416.9053757$. La détermination de ces derniers nous a permis d'estimer le quantile extrême et la période de retour qui vaut 40 ans.

Les mêmes étapes ont été faites sur l'échantillon de données simulées, la période de retour

obtenue était proche de celle de l'échantillon réelle, elle vaut approximativement 32 ans.

À partir de cette application réalisée on peut prédire que cet évènement rare ne reviendra qu' après 40. Ce qui nous permettra de prendre nos précautions et d'éviter des dégâts matériels et humains.

Les points abordés dans ce mémoire nous ouvrent la voie sur plusieurs pistes de recherche intéressantes.

- Comment déterminer précisément les suites (a_n) et (b_n) de normalisation qui apparaissent dans le théorème de Fisher-Tippett pour des processus spécifiques.
- Proposer d'autres techniques et méthodes plus efficaces pour le choix du seuil.
- Investigation dans les techniques d'estimation.

Bibliographie

- [1] E. Alvarado , D. Sand Berg and S. Picford, *Modelling large forest fires as extreme events*. 1998.
- [2] A. Amar, *Apport de la théorie des valeurs extrêmes à la modélisation et la gestion des risques boursiers, financiers et hydro-météorologiques*. Thèse doctorat. Université Mohammed V-Agdal, 2014.
- [3] A. Borchani, *Statistiques des valeurs extrêmes dans le cas de lois discrètes*. Document de recherche ESSEC/Centre de recherche de l'ESSEC ISSN. Ecole Supérieure de la Statistique et de l'Analyse de l'Information, Tunis, December 2010.
- [4] A.L.M. Dekkers, J.H.L.Einmahl, de L.Haan, *A moment estimator for the index of an extrem value distribution*. The Annals of Statistics, 1989.
- [5] E. Deme, *Quelques contributions à la théorie univariée des valeurs extrêmes et estimation des mesures de risque actuariel pour des pertes à queues lourdes : synthèse bibliographique*. Thèse de doctorat. Université Gaston Berger De Saint-Louis, 2013.
- [6] J. Diebolt, A. Guillou, P. Naveau, and P. Ribereau, *Improving Probability-Weighted Moment methods for the Generalized Extreme Value distribution* . Revstat Statistical Journal, 2008.
- [7] J. El Methni, *Contributions à l'estimation de quantiles extrêmes. Applications à des données environnementales*. Thèse de doctorat. Université de Grenoble, 2013.
- [8] P. Embrechts, C. Kluppelberg , and T. Mikosch *Modelling Extremal Events for Insurance and Finance*.Springer Verlag, 1997.
- [9] A. Ferreira , L. Haan , and L. Peng *On optimizing the estimation of high quantiles of a probability distribution*. Statistics 37, 401-434, 2003.

-
- [10] R. Fisher, L. Tippett, *Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proceedings of the Cambridge Philosophical Society.* 1928.
- [11] A. Jenkinson, *The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society.* 1955.
- [12] L. Haan, *Sample extremes : an elementary introduction. Statistica Neerlandica .* 1976.
- [13] J.R.M. Hosking, J.R. Wallis, E.F. Wood, *Estimation of the Generalized Extreme Value distribution by the method of Probability-Weighted Moments . Technometrics,* 1985.
- [14] M. Garrido, *Modélisation des évènements rares et estimation des quantiles extrêmes, Méthodes de sélection de modèles pour les queues de distribution. Modélisation et simulation .* Université Joseph-Fourier - Grenoble, 2002.
- [15] B. Gnedenko, *Sur la distribution limite du terme maximum d'une série aléatoire. The Annals of Mathematics.* 1943.
- [16] E. Gumbel, *Statistical theory of extreme values and some practical applications : a series of lectures. Numéro Applied Mathematics Series, 33.* National Bureau of Standards, Washington, 1954.
- [17] L. Kherraz, N.Samer, *Sur l'estimation de quantiles.* Thèse de master. Université A. Mira, 2016.
- [18] M.R. Leadbetter, *Extremes and local dependence of stationary sequences.* 1983.
- [19] E. Parent, J.Bernier, *Le raisonnement bayésien. Modélisation et inférence.* Springer-Verlag France, Paris, 2007.
- [20] J. Pickands, *Statistical inference using extreme order statistics.* 1975.
- [21] F. Planchet, *Utilisation de la théorie des valeurs extrêmes dans le contexte solvabilité 2.* Support de cours. I.S.F.A, 2015-2016.
- [22] S. Resnick, *Extreme Values, Regular Variation, and Point Processes. Springer Verlag.* 1987.
- [23] N. Temame, *Estimation du quantile extrême et de la VAR.* Thèse de Magistère. Université de Mouloud Mammeri, Tizi-Ouzou, 2011.
- [24] R. Von Mises, *La distribution de la plus grande de n valeurs. American Mathematical Society, Providence.* 1954.

Résumé

Dans ce mémoire, nous donnons dans un premier temps, les principaux résultats et définitions de la théorie des valeurs extrêmes utiles dans nos travaux, en présentant les deux principaux outils servant à modéliser le comportement des valeurs extrêmes : la loi des valeurs extrêmes et la loi des excès. Dans un second temps, nous présentons des méthodes pour estimer le quantile extrême : estimation naturelle basée sur le quantile empirique, méthode des blocs (Bloc Maxima), méthode P.O.T (Peak Over Threshold). Nous donnons aussi un aperçu sur les méthodes d'estimation du niveau de retour et de la période de retour. Une application en pluviométrie est donnée pour estimer le quantile extrême et la période de retour d'un débordement d'une rivière en mettant en pratique la méthode P.O.T.

Mots-clés: Valeurs extrêmes, Quantile extrême, évènement rare, Statistique d'ordre, période de retour.

Abstract

In this paper, we first give the main results and definitions of the extreme value theory useful in our work, presenting the two main tools used to model the behavior of extreme values: the law of extreme values and the law of excesses (peaks). In a second step, methods are proposed to estimate the extreme quantile: natural estimation based on the empirical quantile, block method (Bloc Maxima), P.O.T (Peak Over Threshold) method. We also provide an overview of methods for estimating the return level and the return period. An application in rainfall is given to estimate the extreme quantile and the return period of an overflow of a river by putting into practice the method P.O.T.

Keywords: Extreme values, Extreme quantile, rare event, Order statistics, return period.