



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et La Recherche
Scientifique



Faculté des Sciences et de Technologie
Département de Mathématiques et Informatique
Laboratoire de Mathématiques et Sciences Appliquées
Université De Ghardaïa

Mémoire de fin d'étude
en vue d'obtention du diplôme de

MASTER

Domaine: Mathématiques et Informatiques
Spécialité: Analyse Fonctionnelle et Applications

THÈME

**Théorie des valeurs extrêmes et son importance
dans la modélisation du comportement extrême**

Présenté par:

Nanna Neffaa

Soutenu publiquement le : 17/ 09/ 2020

Devant le jury:

M.CHIKH SALAH Abdelouhab	MCB(Univ. Ghardaia)	Président
M.CHIKH SALAH Aboukacem	MAA(Univ. Ghardaia)	Encadreur
M.EL HADJ MOUSSA Yacine	MAA(Univ. Ghardaia)	Examineur

Année universitaire 2019/2020

Remerciements

Gloire soit rendue à Allah tout puissant, qui m'a donné force et patience d'accomplir mon mémoire de fin d'étude.

Mes sincères remerciements à mon encadreur **M.Aboukacem CHIKH SALAH** pour son soutien et son attention exceptionnels durant ces derniers mois. Je tiens à lui témoigner ma gratitude pour le choix du thème. Et pour son suivi attentif et pertinent qui a mené à l'acheminement de ce travail.

Je remercie également tous les membres de jury, **M.Abdelouhab CHIKH SALAH**, et **M.Yacine HADJ MOUSSA**, pour avoir accepté d'évaluer mon travail.

Mes vifs remerciements à tout le corps professoral qui a contribué à l'acheminement de cette formation pluridisciplinaire, avec dévouement et grande patience

Enfin, à toute personne ayant contribué, de près ou de loin, à l'élaboration de ce mémoire de fin d'études. Veuillez bien trouver ici l'expression de mes sincères remerciements.

Dédicaces

Mes remerciements les plus sincère vont à ma famille, surtout mes parents qui m'ont donné naissance.

À ma très chère mère Faffa mon amour, ma joie et raison de mon existence.

À mon très cher père Yacine ma source d'inspiration et de patience qui m'a incessamment épaulée tout au long du chemin de savoir et de la science.

Mes très chers aimés, sœur et frères

Imane, Omar, Mohamed, Anas. Ma grande source de confiance qui m'ont aidée à surmonter toute épreuve et circonstance.

Mon cher fiancé Ayman Nedjmeddine et sa très honorable mère Mme Samia, pour m'avoir prodigué de leurs précieux encouragements et assistance.

Mes chers grands-parents ainsi que ma grande famille, pour leurs conseils et bienveillance.

L'âme de mon défunt professeur et ex-encadreur M.Abdennour LANI,
Qu'Allah le bénisse de sa miséricorde et clémence.

A ma sœur Acya pour sa fidélité et assistance.

A ma chère amie Imane, ma main droite.

Enfin, je ne peux passer outre ma reconnaissance envers mes très chères amies pour leurs affection et complaisance et envers tout ceux qui m'ont inculqué les bonnes valeurs et la bonne conscience.

A vous tous, je dédié mon projet de fin d'étude.

Nanna NEFFAA

Résumé

L'objectif de ce mémoire est de donner une synthèse sur les différentes théories et résultats sur les valeurs extrêmes. Le théorème de Fisher-Tippet-Gnedenko qui est rapproché du théorème de la limite centrale. Et aussi le théorème de Pickands qui établit l'équivalence entre la convergence d'une loi du maximum vers une loi des valeurs extrêmes et la convergence d'une loi d'un excès vers une distribution généralisée de Paréto.

Les domaines d'applications de la théorie des valeurs extrême sont très variés. Nous présentons lors de ce mémoire deux échantillons concernant deux domaines différents : un cas d'étude en hydrologie sur des données réelles de débits horaires pour la rivière de Molenbeek en Belgique entre 1986-1996, ainsi qu'une modélisation du taux de change quotidien entre le dollar canadien et le dollar américain dans la période 1995-2000, représentant un cas d'étude effectuée en économie.

Mots clés : Statistique d'ordre ; valeurs extrêmes ; distribution asymptotique ; domaines d'attraction, estimations.

Abstract

The objective of this thesis is to make a synthesis on the different theories and results on extreme values. The theorem of Fisher-Tippet-Gnedenko which is related to the central limit theorem. And also the theorem of Pickands which establishes the equivalence between the convergence in law of the maximum towards a law of extreme values and the convergence in law of an excess towards a generalized Pareto distribution.

The fields of application of extreme values theory are varied. We present during this thesis two samples for two different fields: A case study in hydrology on real data of hourly flows for the Molenbeek river in Belgium between 1986-1996, as well as a modeling of the daily exchange rate between the Canadian dollar and American dollar in the period 1995-2000 representing a case study carried out in economics.

Keywords: Order statistics; extreme values; asymptotic distribution; domains of attraction, estimates.

ملخص

الهدف من هذه الأطروحة هو تجميع النظريات المختلفة والنتائج على القيم المتطرفة. نظرية فيشر- تيببت- غيدنكو المرتبطة بنظرية الحد المركزي ونظرية بيكاندس التي تؤسس التكافؤ بين التقارب في قانون الحد الأقصى تجاه قانون القيم المتطرفة والتقارب في قانون التجاوز نحو توزيع واسع النطاق لباريتو. إن مجالات تطبيق نظرية القيم المتطرفة متنوعة للغاية، نقدم خلال هذه الرسالة عينتين من مجالين مختلفين: دراسة حالة في مجال الري على بيانات حقيقة للتدفقات في الساعة لنهر مولينبيك في بلجيكا بين عامي 1986-1996، بالإضافة إلى نمذجة سعر الصرف اليومي بين الدولار الكندي والدولار الأمريكي في الفترة 1995-2000، تمثل دراسة حالة أجريت في الاقتصاد.

الكلمات المفتاحية: ترتيب الإحصائيات، القيم المتطرفة، توزيع مقارب، مناطق الجذب، التقديرات.

Notations et Abréviations

les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous.

Abréviations

- EVD* : Extrême Value Distribution.
- GEV* : Distribution des Valeurs Extrêmes Généralisée.
- GPD* : Distribution de Paréto Généralisée.
- i.i.d* : independant et identiquement distribué.
- MLE* : L'estimation du maximum de vraisemblance.
- POT* : Peaks-Over-Threshold.
- TLC* : Théorème limite centrale.
- TVE* : Théorie des Valeurs Extrêmes.
- VaR* : Valeur à Risk.

Notations

\xrightarrow{d} :	Convergence en distribution.
$\xrightarrow{p.s.}$:	Convergence presque sûrement.
\xrightarrow{p} :	Convergence en probabilité.
Φ_α :	Distribution de Fréchet.
Λ :	Distribution de Gumbel.
Ψ_α :	Distribution de Weibull.
$\mathcal{D}(\cdot)$:	Domaine d'attraction.
X_1, \dots, X_n :	Échantillon de taille n de X .
\mathbb{N} :	Ensemble des nombres naturel.
\mathbb{R} :	Ensemble des valeurs réelles
Ω :	Espace probabilisé.
μ :	Espérance d'une variable aléatoire.
$\hat{\xi}_{k_n}^H$:	Estimateur de Hill.
$\hat{\xi}_{k_n}^P$:	Estimateur de Pickands.
$\hat{\xi}_{k_n}^Z$:	Estimateur de Zipf.
G :	Fonction à variations régulières.
ℓ :	Fonction à variations lente.
F :	Fonction de répartition
f :	Fonction de probabilité d'une variable aléatoire.
F_n :	Fonction de répartition empirique.
ξ :	Indice des valeurs extrêmes.
$X_{(k,n)}$:	$k^{\text{ème}}$ statistique d'ordre.
$N(\cdot, \cdot)$:	Loi normale
$X_{(1,n)}$:	Minimum de X_1, \dots, X_n .
$X_{(n,n)}$:	Maximum de X_1, \dots, X_n .
$X_{(1,n)}, \dots, X_{(n,n)}$:	n -Statistique d'ordre.
T :	Période de retour.
x_F :	Point terminal.
u :	Seuil.
a_n :	Suite numérique.
$v.a.$:	Variables aléatoires.
σ^2 :	Variance d'une variable aléatoire.

Table des matières

Remerciements

Dédicaces

Notations et Abréviations i

Table des figures v

Liste des tableaux vi

Introduction générale 1

1 Statistique d'ordre et convergence 3

1.1 Variables aléatoires et leurs distributions 4

1.2 Statistique d'ordre 4

1.2.1 Fonction de répartition empirique 5

1.2.2 Densité de probabilité 5

1.2.3 Distribution d'une statistique d'ordre 5

1.2.4 Densité conjointe de n-statistique d'ordre 6

1.2.5 Densité conjointe de la $K^{\text{ème}}$ statistique d'ordre 6

1.3 La convergence 6

1.3.1 Convergence en probabilité 6

1.3.2 Convergence en distribution 7

1.3.3 Convergence de F_n vers F 7

1.3.4 Le point terminal 7

1.3.5 Théorème limite centrale 7

2 Théorie des valeurs extrêmes 9

2.1 Quantile-Quantile plot 10

2.2 Distributions à variations régulières 10

2.2.1 Représentation de Karamata 10

2.2.2 Quelques résultats sur les fonctions à variations régulières 11

2.3 Distribution asymptotique d'un extrême 12

2.3.1 Distribution du maximum d'un échantillon 12

2.3.2 Distributions généralisées des valeurs extrêmes 14

2.4 Domaine d'attraction 15

2.5 Caractérisation des domaines d'attraction 15

2.5.1 Domaine d'attraction de Fréchet 15

2.5.2 Domaine d'attraction de Weibull 16

2.5.3 Domaine d'attraction de Gumbel 17

2.6 Distribution conditionnelle des excès 17

2.6.1	Distribution de Paréto généralisée GPD	18
2.6.2	Théorème de Pickands	19
2.7	Estimation de l'indice des valeurs extrêmes	19
2.7.1	Estimateur de Hill	19
2.7.2	Estimateur de Pickands	21
2.7.3	Estimateur de Zipf	22
2.7.4	Estimation de vraisemblance maximale	22
2.7.5	Estimation de périodes de retour	23
3	Les champs pratiques de la TVE	26
3.1	Application en hydrologie	27
3.1.1	Présentation des données hydrologiques	27
3.1.2	Sélection d'observations indépendantes	28
3.1.3	Simulations	30
3.2	Application en économie	31
3.2.1	Simulation	32
3.2.2	Illustrations graphiques	32
	Conclusion générale	36
	Bibliographie	37

Table des figures

2.1	Représentation graphique de la densité des valeurs extrêmes avec $\alpha = 1$. . .	14
2.2	Méthode des excès : u réel suffisamment élevé appelé seuil, Y :excès de X au-delà de u	18
2.3	Graphique de Hill à $n = 1000$ et la vraie valeur de $\xi = 1$	21
2.4	Graphique de Hill en moyenne sur 100 réalisations à $n = 1000$	21
2.5	La variance asymptotique de l'estimateur $\hat{\xi}_{k_n}^P$ (noir) et de $\hat{\xi}_{k_n}^H$ (rouge) en fonction de ξ	23
3.1	Influence des inondations sur la courbe de tarage pour la rivière Molenbeek à Erpe-Mere avant 1996	27
3.2	Comparaison des débits horaires de la rivière et de l'écoulement pour la période d'inondation de décembre 1993	28
3.3	Explication schématique de l'influence de l'inondation sur les hydrogrammes	28
3.4	Valeurs indépendantes POT sélectionnées à partir de la série de débits horaires	29
3.5	Paramètres utilisés dans le critère de sélection des valeurs POT indépendantes	29
3.6	les différences entre les valeurs POT sélectionnées pour différentes valeurs de f	29
3.7	Comparaison de la distribution empirique pour les débits horaires <i>POT</i> et les débits d'écoulement en amont pour la rivière	30
3.8	Distribution des valeurs extrêmes décomposée en deux sous-populations pour les débits horaires de la rivière	30
3.9	Estimation de type Hill de la pente en fonction du seuil considéré pour les débits horaires pour une valeur de $f = 0.3$	31
3.10	Les valeurs POT translattées dans "l'exponentiel quantile plot	31
3.11	comparaison des extrêmes POT translattés pour différents niveaux d'indé- pendance	32
3.12	Histogramme de la queue de la distribution empirique	33
3.13	estimation de Hill avec indice d'extrême autour de valeurs de 3 à 4	34
3.14	Q-Q pour la distribution EVD, à partir de blocs de 28 observations	34
3.15	Q-Q pour la distribution EVD, à partir de blocs de 90 observations	34
3.16	la distribution excédentaire pour un GPD : 90 dépassements	35
3.17	Le 0.999 quantile de (0,001%) pour différents seuils	35

Liste des tableaux

3.1	les résultats de la VaR pour différentes distributions	32
-----	--	----

Introduction générale

La théorie des valeurs extrêmes TVE est un sujet classique de la théorie des probabilités et des statistiques mathématiques qui s'intéresse aux valeurs extrêmes des distributions de probabilité.

Beaucoup de questions ont été traitées dans la littérature, telles que : Que sont les valeurs extrêmes ? Pourquoi posent-elles problème ? Quel est son contexte mathématique ? Quel est son comportement asymptotique ? Quel est l'exploit de ses thèmes de recherches dans les différents domaines possédant de comportements extrêmes ?

Les extrêmes sont des événements rares qui conduisent à des pertes importantes. Depuis quelques années, la TVE a reçu beaucoup d'attention aussi bien sur le plan théorique que sur le plan pratique. Les domaines d'applications sont en effet très variés : Hydrologie, météorologie, biologie, ingénierie, gestion de l'environnement, finance, assurance, sciences sociales, etc.

Les origines de la théorie des valeurs extrêmes remontent entre 1920 et 1940 grâce aux travaux de Fisher et Tippett en 1928 [14] puis plus tard Gnedenko 1943 [17] dont ils établissent un résultat crucial qui spécifie la forme de la distribution limite du maximum convenablement normalisé. Ils montrent que la distribution limite possible des extrêmes est l'une des trois familles de distributions suivantes : Gumbel, Fréchet ou Weibull, appelées distributions de valeurs extrêmes. Citons en référence pour la théorie Embrechts et al. (1997) [13], S.Coles [9], J.Beirlant et al[3]...

D'ailleurs la loi des valeurs extrêmes, lorsqu'elle existe, est indexée par un paramètre appelé indice de la valeur extrême ou indice de queue, et éventuellement par des paramètres d'échelle et de position. Cette loi limite est notée GEV proposée par Jenkinson (1955) [19].

Par ailleurs, si on considère la valeur extrême d'une variable aléatoire comme un excès au dessus d'un seuil élevé, alors la distribution asymptotique de ce type d'excès, quand le seuil tend vers le point terminal, est donnée par le théorème de Pickands 1975 [25].

L'objectif de ce mémoire est de faire une synthèse sur les différentes théories et résultats sur les valeurs extrêmes, ainsi que leurs champs d'application.

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

Le premier chapitre est consacré à des rappels sur la statistique d'ordre et les différentes initiations de convergence, tels convergence en probabilité et la convergence en distribution, on rappelle préalablement le théorème de la limite centrale [9].

Dans le deuxième chapitre on présente la théorie des valeurs extrêmes. On commence par les différentes distributions utilisées dans la littérature (distribution à variations régulières, distribution asymptotique d'un extrême, distribution conditionnelle des excès). De

plus, on parle sur les trois domaines d'attraction (Fréchet, Weibull, Gumbel). Ensuite, on donne les méthodes d'estimations des extrêmes tels l'estimation de Hill, de Pickands, de Zipf, estimation de vraisemblance maximale, et l'estimation de période de retour).

Par ailleurs, les domaines d'applications sont en effet très variés ce qui fera l'objet du troisième et dernier chapitre de ce mémoire. Dont on a présenté deux exemples; l'un traite un cas d'étude d'hydrologie synthétisé depuis l'article de Guillou et Willems(2006) [18], et l'autre aborde une question de la finance reprise depuis l'étude de Bensalah et al[4].

Statistique d'ordre et convergence

Les références utilisées dans ce chapitre sont : [1],[9],[16],[27],[28]

Introduction

Dans ce chapitre nous présentons quelques notions essentielles de la théorie des valeurs extrêmes, dont elle fournit une base mathématique et probabiliste rigoureuse sur laquelle il est possible de construire des modèles statistiques pour prévoir la fréquence de ces phénomènes rares. En effet, l'idée derrière l'utilisation de la théorie des valeurs extrêmes pour la modélisation des queues de distributions est similaire à celle qui apparaît lorsqu'on approche la moyenne de variables aléatoires indépendantes et identiquement distribuées par une loi normale.

Le théorème de la limite centrale dit que lorsque n tend vers l'infini, la moyenne tend à être distribuée suivant une loi normale si la variance de X est fini. La théorie des valeurs extrêmes se situe dans la même logique, mais étudie la queue de la loi au lieu de sa moyenne.

On représente dans ce chapitre les notions de la statistique d'ordre, en plus les définitions des différents types de convergences nécessaires pour l'étude du comportement asymptotique et le théorème de la limite centrale.

1.1 Variables aléatoires et leurs distributions

Une expérience est dite aléatoire si l'on ne peut pas prédire de manière certaine son résultat, bien qu'elle soit soumise à certaines lois. On considère qu'il est possible d'associer à une expérience aléatoire un ensemble qui contient tous les résultats possibles de l'expérience, cet ensemble est appelé univers et est noté en général Ω .

On distingue deux distributions de probabilités.

Une distribution de probabilité déterminée par la fonction de masse de probabilité qui prend la formule suivante associée à une v.a. discrète X :

$$f(x) = P(X = x)$$

Une distribution de probabilité spécifiée par la fonction de distribution de probabilité associée à une v.a. continue X , définie comme :

$$F(x) = P(X \leq x)$$

Définition 1.1.1

Une variable aléatoire (v.a.) sur Ω est une application X définie sur Ω à valeurs dans un ensemble E , dont elles sont réparties en deux : v.a. discrètes prennent des valeurs entières discontinues sur un intervalle donné et v.a. continues prennent des valeurs continues sur un intervalle donné.

Définition 1.1.2 (loi normale)

Une variable aléatoire absolument continue X suit une loi normale de paramètres (μ, σ) si sa densité de probabilité est donnée par :

$$f : \mathbb{R} \rightarrow \mathbb{R}$$

$$x \mapsto f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \quad \text{avec } \mu \in \mathbb{R} \quad \text{et } \sigma \in \mathbb{R}^+ \quad (1.1)$$

Notons $X \rightarrow N(\mu, \sigma)$

Remarque 1.1

- La loi normale (loi de Gauss-Laplace) prend une place particulière grâce au théorème centrale limite.
- la loi normale permet d'approcher d'autres lois ainsi de modélisation statistique.

1.2 Statistique d'ordre

Les statistiques d'ordre jouent un rôle de plus en plus important dans la théorie des valeurs extrêmes, dont elles fournissent des informations sur la distribution de probabilité.

Définition 1.2.1

On appelle **statistique d'ordre** notées $X_{(1,n)}, \dots, X_{(n,n)}$, les v.a. ordonnées :

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

Définition 1.2.2

Les statistiques d'ordre extrême sont définies comme termes du maximum et du minimum, telle que la plus petite statistique d'ordre $X_{(1,n)}$ (ou statistique du minimum) :

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

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

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

Avec : $\max(X_{(1)}, \dots, X_{(n)}) = -\min(X_{(1)}, \dots, X_{(n)})$

1.2.1 Fonction de répartition empirique

$F_n(x)$ est la proportion des n variables qui sont inférieurs ou égales à x .

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

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}\{X_k \leq x\} = \begin{cases} 0 & \text{si } x < X_{(1,n)} \\ \frac{i}{n} & \text{si } X_{(i,n)} \leq x \leq X_{(i+1,n)} \\ 1 & \text{si } x \geq X_{(n,n)} \end{cases} \quad 1 \leq i \leq n$$

où \mathbb{I} est la fonction indicatrice.

1.2.2 Densité de probabilité

Définition 1.2.3 (fonction de densité)

Si la fonction de distribution (ou répartition) F est différentiable, il est également utile de définir la **fonction de densité** de la probabilité de X comme :

$$f(x) = \frac{dF}{dx}$$

et si la fonction de distribution F admet une dérivée au point x alors :

$$f(x) = \lim_{\delta \rightarrow 0} \frac{P(x \leq X \leq x + \delta)}{\delta}$$

1.2.3 Distribution d'une statistique d'ordre

Soit $X_{(1)}, \dots, X_{(n)}$ n -variables aléatoires $X_{(k,n)}$ i.i.d fonction de répartition F continue. Soit $x < X_{(k,n)} < x + \delta x$, en considérant δ comme étant petit, on a :

$$\begin{aligned} P(x < X_{(k,n)} \leq x + \delta x) &= \frac{n!}{(k-1)!(n-k)!} [F(x)^{k-1}] [1 - F(x + \delta x)]^{n-k} \\ &\quad \times [F(x + \delta x) - F(x)] + O((\delta x)^2) \end{aligned} \quad (1.2)$$

Où $O((\delta x)^2)$ le terme de l'ordre $(\delta x)^2$, est la probabilité correspondant à l'évènement d'avoir plus d'un X_i dans l'intervalle $[x, x + \delta x]$.

De la formule 1.2 on obtient la fonction de densité de $X_{(k,n)}$ ($1 \leq k \leq n$) :

$$\begin{aligned} f_{(k,n)}(x) &= \lim_{\delta x \rightarrow 0} \left\{ \frac{P(x < x_{k,n} \leq x + \delta x)}{\delta x} \right\} \\ &= \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x), \quad -\infty < x < +\infty \end{aligned} \quad (1.3)$$

1.2.4 Densité conjointe de n-statistique d'ordre

Soit $X_{(1)}, \dots, X_{(n)}$ n-variables aléatoires $X_{(k,n)}$ i.i.d fonction de répartition F continue, alors la densité conjointe de statistique d'ordre $(X_{(1,n)}, \dots, X_{(n,n)})$ est donnée par :

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

1.2.5 Densité conjointe de la $K^{\text{ème}}$ statistique d'ordre

La loi de la v.a. $X_{(k,n)}$, pour $1 \leq k \leq n$, est donnée par :

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

Sa densité est donnée par :

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

En effet, nous avons la fonction de distribution de statistique du minimum :

$$F_{X_{(1,n)}}(x) = 1 - [1 - F(x)]^n \quad (1.4)$$

et sa fonction de densité :

$$f_{X_{(1,n)}}(x) = n[1 - F(x)]^{n-1} f(x) \quad (1.5)$$

la fonction de distribution de statistique du maximum :

$$F_{X_{(n,n)}}(x) = P[X_{(n,n)} \leq x] = [F(x)]^n \quad (1.6)$$

et sa fonction de densité :

$$f_{X_{(n,n)}}(x) = n[F(x)]^{n-1} f(x) \quad (1.7)$$

La formule 1.6 présente un intérêt très limité car la loi d'une variable aléatoire parente X est rarement connue avec précision, et même si la loi de cette variable parente X est connue avec exactitude, la loi du maximum n'est pas toujours facilement calculable. Comme les valeurs extrêmes se trouvent à droite et à la fin du support de la distribution, intuitivement le comportement asymptotique de M_n doit permettre de rendre compte de queue de la distribution.

1.3 La convergence

1.3.1 Convergence en probabilité

Définition 1.3.1

On dit que la suite de v.a. $(X_n, n \in \mathbb{N}^*)$ converge en probabilité vers la variable aléatoire X si $\forall \varepsilon > 0$ on a :

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

Ou bien :

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$$

On écrit :

$$X_n \xrightarrow{p} X$$

1.3.2 Convergence en distribution

Définition 1.3.2

On dit que la suite de v.a. $(X_{(n)}, n \in \mathbb{N}^*)$ de fonction de répartition F_n converge en distribution vers la variable aléatoire X de fonction de répartition F si $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ en tout point de continuité de F on note :

$$X_n \xrightarrow{d} X$$

1.3.3 Convergence de F_n vers F

Théorème 1.1 (Glivenko-Cantelli) [28]

La convergence de F_n vers F est presque sûrement uniforme, c'est-à-dire que :

$$\sup_x |F_n(x) - F(x)| \rightarrow 0$$

1.3.4 Le point terminal

Définition 1.3.3 On appelle point terminal (en anglais "upper endpoint") de la fonction F le réel x_F défini par :

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

Le point terminal est la plus grande valeur possible pour la réalisation d'un échantillon issu d'une fonction F . Ce point terminal peut être infini ou fini.

Proposition 1.3.1

La suite des maximums $\{M_n = \max(X_{(1)}, X_{(2)}, \dots, X_{(n)}), n \geq 1\}$ converge presque sûrement vers x_F quand $n \rightarrow \infty$, i.e. :

$$M_n \xrightarrow{p.s.} x_F \quad n \rightarrow \infty \quad (1.8)$$

Le résultat suivant découle automatiquement de la formule (1.4)

Corollaire 1.3.1

La suite des maximums $\{M_n, n \in \mathbb{N}^*\}$ converge en loi vers une v.a. dégénérée concentrée en x_F , car

$$\lim_{n \rightarrow \infty} F_{M_n}(x) = \lim_{n \rightarrow \infty} P(M_n \leq x) = \lim_{n \rightarrow \infty} F^n(x) = \begin{cases} 0 & \text{si } x < x_F \\ 1 & \text{si } x \geq x_F \end{cases} \quad (1.9)$$

On constate que la distribution asymptotique du maximum donne une loi dégénérée, une masse de Dirac en x_F , puisque pour certaines valeurs de x , la probabilité peut être égale à 1 dans le cas où x_F est fini

1.3.5 Théorème limite centrale

Le théorème de la limite centrale établit la convergence en loi de la somme d'une suite de v.a. vers la loi normale.

Théorème 1.2 [9]

Soient $(X_{(n)}, n \in \mathbb{N}^*)$ une suite de variables aléatoires i.i.d avec l'espérance μ et de variance σ^2 . On définit :

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

alors,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} Z$$

En tant que $n \rightarrow \infty$, où $Z \sim N(0, 1)$ (loi de probabilité normale)

Conclusion

Dans ce chapitre nous avons donner une introduction sur les notions préliminaires de la théories des valeurs extrêmes dont on a mentionné les variables aléatoires et leurs distributions, ainsi on a vu les notions de la statistique d'ordre, en plus les définitions des différents types de convergences nécessaires à l'étude du comportement asymptotique et le théorème de la limite centrale. Ces notions de base vont servir à aborder les valeurs extrêmes que nous allons traiter dans le chapitre suivant.

Théorie des valeurs extrêmes

Les références utilisées dans ce chapitre sont : [3],[7],[9],[11],[15],[17],[22],[23],[27]

Introduction

La théorie des valeurs extrêmes communément appelée " Extreme Value Theory" (*EVT*) en anglais, est une vaste théorie dont le but est d'évaluer les évènements rares c'est à dire les 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 des valeurs beaucoup plus grandes ou plus petites que celles observées habituellement.

On compte présenter dans ce chapitre : le théorème de Gnedenko sur la loi des valeurs extrêmes qui est pour le maximum de n observations, un analogue du théorème limite centrale pour la moyenne. Cela il introduit alors des suites numériques $(a_n)_{n \geq 1}$ et $(b_n)_{n \geq 1}$ de normalisation permettant d'obtenir une loi limite pour $a_n(X_{(n,n)} - b_n)$. Ce théorème fait appel aux notions de fonction à variation régulière et normalisée. Dans ce contexte, deux approches sont utilisées pour caractériser l'appartenance à un domaine d'attraction. La première repose sur la loi limite du maximum d'un échantillon de variable aléatoires i.i.d et la deuxième repose sur la loi des excès communément appelée (approche POT, "Peaks-Over-Threshold"). Cette dernière est la plus utilisée dans les applications que l'on peut trouver dans la littérature des valeurs extrêmes, notamment en hydrologie (Davison et Smith 1990, McNeil et Saladin 1997).

2.1 Quantile-Quantile plot

Le Quantile-Quantile plot ou plus succinctement QQ-plot est très utilisés pour tester l'adéquation des modèles des valeurs extrêmes, cet outil graphique permet de mesurer graphiquement la proximité d'une distribution théorique à une distribution empirique. En effet, à partir de l'échantillon observé, on définit alors un certain nombre de quantiles $x_{i,n}$. Si la distribution théorique choisie est pertinente, le graphe de cet ensemble de points sera confondu avec la première diagonale.

2.2 Distributions à variations régulières

La théorie des fonctions à variations régulières apparaît dans un grand nombre d'applications mathématiques, c'est un outil essentiel pour traiter les queues lourdes, la dépendance à longue portée et les domaines d'attraction. Les fonctions à variations régulières ont été lancées par Jovan Karamata (1930) [6], et développées par lui-même et par beaucoup d'autres recherches.

Définition 2.2.1

une fonction G est dite à **variations régulières** d'indice $\rho \in \mathbb{R}$ (à l'infini) si G est positive à l'infini (i.e. s'il existe A tel que pour tout $x \geq A$, $G(x) > 0$) et si pour tout $t > 0$:

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

Dans le cas particulier où $\rho = 0$, on dit que G est une **fonction à variations lentes**. Une fonction à variations régulières d'indice ρ peut toujours s'écrire sous la forme $x^\rho \ell(x)$ où ℓ est une fonction à variations lentes à l'infini.

Par exemple, la fonction constante, les fonctions qui tendent vers une constante et $\log(x)$ sont des fonctions à variations lentes à l'infini. Les fonctions à variations lentes jouent un rôle prépondérant dans l'étude des lois de valeurs extrêmes.

Exemple. $f(x) = \log(x)$, $x > 0$

$$\begin{aligned} \lim_{t \rightarrow +\infty} \frac{\log(tx)}{\log(t)} &= \lim_{t \rightarrow +\infty} \frac{\log(t) + \log(x)}{\log(t)} \\ &= \lim_{t \rightarrow +\infty} \left(1 + \frac{\log x}{\log t}\right) \\ &= 1 \end{aligned}$$

2.2.1 Représentation de Karamata

Théorème 2.1 (Corollaire du théorème 0.6) [27]

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

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

où c et ε sont des fonctions positives telle que :

$$\lim_{x \rightarrow \infty} c(x) = c \in]0, +\infty[\text{ et } \lim_{t \rightarrow \infty} \varepsilon(t) = 0 \quad (2.1)$$

Si la fonction c est constante, on dit que ℓ est normalisée. l'équation 2.1 implique que si ℓ est normalisée alors ℓ est dérivable de dérivé ℓ' pour tout $x > 0$:

$$\ell'(x) = \frac{\epsilon(x)\ell(x)}{x}$$

En particulier, on a :

$$\lim_{x \rightarrow \infty} x \frac{\ell'(x)}{\ell(x)} = 0$$

Remarque 2.1 Soit G une fonction à variations régulières d'indice ρ .

En utilisant le fait que $G(x) = x^\rho \ell(x)$, on déduit facilement du théorème (2.1) que pour tout $x > 0$:

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

où c et ρ sont des fonctions positives telles que :

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

2.2.2 Quelques résultats sur les fonctions à variations régulières

Le résultat du lemme suivant montre que la convergence du rapport de deux fonctions à variations régulières d'indice ρ de la définition 2.2.1 est localement uniforme lorsque x tend vers l'infini.

Lemme 2.1 [15]

Si G est une fonction à variations régulières d'indice ρ alors, pour tout $0 < a < b$:

$$\lim_{x \rightarrow \infty} \sup_{t \in [a, b]} \left| \frac{G(tx)}{G(x)} - t^\rho \right| = 0$$

Lemme 2.2 [15]

Soit G une fonction à variations régulières d'indice $\rho > 0$. Il existe une fonction H à variations régulières d'indice $\frac{1}{\rho}$ telle que :

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

Définition 2.2.2

On dit que a_n et b_n sont deux suites équivalentes (et on note $a_n \sim b_n$) si $\frac{a_n}{b_n} \rightarrow 1$ lorsque $n \rightarrow \infty$

Lemme 2.3 [15]

Soit ℓ une fonction à variations lentes.

1. pour tout $\gamma > 0$:

$$\lim_{n \rightarrow \infty} n^\gamma \ell(n) = +\infty \text{ et } \lim_{n \rightarrow \infty} n^{-\gamma} \ell(n) = 0.$$

2. Si a_n et b_n sont deux suites qui convergent vers $+\infty$ lorsque $n \rightarrow \infty$ et telle que :

$$\liminf_{n \rightarrow \infty} \frac{a_n}{b_n} > 0 \text{ et } \limsup_{n \rightarrow \infty} \frac{a_n}{b_n} < +\infty$$

alors $\ell(a_n) \sim \ell(b_n)$

Le lemme 2.3, qui est une conséquence du lemme 2.2, montre que les fonctions à variations lentes conservent les équivalents. Par conséquent, ce résultat implique que si G est à variations régulières d'indice ρ et si $a_n \sim b_n$, alors $G(a_n) \sim G(b_n)$ et on dit que les fonctions à variations régulières conservent aussi les équivalents.

2.3 Distribution asymptotique d'un extrême

2.3.1 Distribution du maximum d'un échantillon

Nous donnons dans cette section le résultat fondamental de la théorie des valeurs extrêmes. Ce résultat est rapproché du théorème de la limite centrale. Il établit la loi asymptotique du maximum $X_{(n,n)}$ convenablement normalisé d'un échantillon.

Les origines de la caractérisation maximale de l'échantillon asymptotique remontent à Fisher et Tippet (1928)[14]. Leurs arguments ont été complétés par Gnedenko (1943)[17]. Gumbel (1958) a eu une influence dans la promotion de la méthodologie. La paramétrisation de l'extrême valeur généralisé à été proposée de manière indépendante par Von-Mises (1954) et Jenkinson (1955).

Définition 2.3.1

Deux fonctions de distributions H et G sont de même type s'il existe $a > 0$ $b \in \mathbb{R}$ tels que pour tout $x \in \mathbb{R}$:

$$H(x) = G(ax + b)$$

Proposition 2.3.1

Soient H et G deux fonctions de répartitions non-dégénérées. Si $F_n \in \mathcal{D}(H)$ (c'est-à-dire s'il existe deux suites $a_n > 0$ et $b_n \in \mathbb{R}$ telles que $F_n(a_n x + b_n) \rightarrow H(x)$). et si $F_n \in \mathcal{D}(G)$ (c'est-à-dire s'il existe deux suites $\alpha_n > 0$ et $\beta_n \in \mathbb{R}$ telles que $F_n(\alpha_n x + \beta_n) \rightarrow G(x)$). Alors, les fonctions de répartitions H et G sont de même type $G(x) = H(ax + b)$, de plus

$$\frac{\alpha_n}{a_n} \rightarrow a > 0, \quad \frac{(\beta_n - b_n)}{a_n} \rightarrow b \in \mathbb{R}$$

Théorème 2.2 (Théorème de Fisher-Tippet-Gnedenko) [17]

Soit $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ un échantillon aléatoire d'une loi F quelconque et soit $X_{(n,n)} = \max(X_{(1)}, X_{(2)}, \dots, X_{(n)})$. S'il existe deux suites réelles $a_n > 0$ et $b_n \in \mathbb{R}$:

$$\lim_{n \rightarrow \infty} P\left(\frac{X_{(n,n)} - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x) \quad (2.2)$$

où H est non-dégénérée. Alors H est du type de l'une des trois classes suivantes :

$$H(x) = \Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp[-(x)^{-\alpha}] & \text{si } x > 0 \end{cases} \quad (2.3)$$

$$H(x) = \Psi_\alpha(x) = \begin{cases} \exp[-(-x)^\alpha] & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \quad (2.4)$$

où α est une constante positive.

$$H(x) = \Lambda(x) = \exp[-\exp(-x)], \quad x \in \mathbb{R}. \quad (2.5)$$

On se réfère à $\Phi_\alpha, \Psi_\alpha, \Lambda$: EVD (extreme value distributions)

Preuve :

Pour une démonstration de ce théorème. On se réfère à l'ouvrage suivant [17] qui a donner une version complète (Gnedenko 1943). Vous trouvez aussi une autre méthode dans la section 0.3 d'ouvrage suivant [27]. Nous indiquons ici les lignes directrices. Nous avons

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H(x) \quad (2.6)$$

Soit $[t]$ le plus grand entier inférieur ou égal à t . Pour tout $t > 0$ on a

$$\lim_{n \rightarrow \infty} F^{[nt]}(a_{[nt]}x + b_{[nt]}) = H(x) \quad (2.7)$$

De plus

$$\begin{aligned} \lim_{n \rightarrow \infty} F^{[nt]}(a_n x + b_n) &= \lim_{n \rightarrow \infty} [F^n(a_n x + b_n)]^{[nt]/n} \\ &= H^t(x) \end{aligned} \quad (2.8)$$

En utilisons le théorème de convergence des types (Proposition 0.2 dans [27]), nous savons que ces limites $\alpha(t) > 0$, $\beta(t) \in \mathbb{R}$ existent tel que

$$\alpha(t) = \lim_{n \rightarrow \infty} a_n/a_{[nt]} \quad ; \quad \beta(t) = \lim_{n \rightarrow \infty} (b_n - b_{[nt]})/a_{[nt]} \quad (2.9)$$

En déduit que

$$H^t(x) = H(\alpha(t)x + \beta(t)) \quad (2.10)$$

Le théorème de Fisher-Tippett dérive alors de la résolution de cette équation fonctionnelle. Nous avons

$$H^{ts}(x) = H(\alpha(ts)x + \beta(ts)) \quad (2.11)$$

De plus

$$\begin{aligned} H^{ts}(x) &= [H^s(x)]^t = [H(\alpha(s)x + \beta(s))]^t \\ &= H(\alpha(t) \{ \alpha(s)x + \beta(s) \} + \beta(t)) \\ &= H(\alpha(t)\alpha(s)x + \alpha(t)\beta(s)\beta(t)) \end{aligned} \quad (2.12)$$

Nous obtenons un système d'équations fonctionnelles :

$$\begin{cases} \alpha(ts) = \alpha(t)\alpha(s) \\ \beta(ts) = \alpha(t)\beta(s)\beta(t) \end{cases} \quad (2.13)$$

La solution de la première équation fonctionnelle est $\alpha(t) = t^{-\theta}$ avec $\theta \in \mathbb{R}$. Selon les valeurs de θ , nous obtenons les trois types de distribution de valeurs extrêmes :

$$Gumbel \quad \theta = 0$$

$$Frechet \quad \theta > 0$$

$$Weibull \quad \theta < 0$$

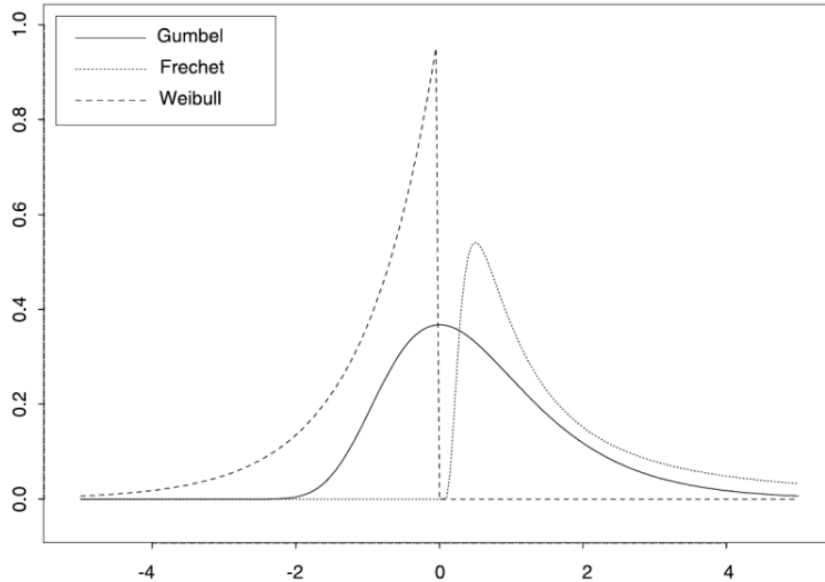
Proposition 2.3.2

Les fonctions de la densité des distributions des valeurs extrêmes, sont les suivantes :

$$\Phi_\alpha(x) = \begin{cases} 0 & \text{sinon} \\ \alpha x^{-(\alpha+1)} \exp[-(x)^{-\alpha}] & \text{si } x > 0, \alpha > 0 \end{cases} \quad (2.14)$$

$$\Psi_\alpha(x) = \begin{cases} \alpha(-x)^{\alpha-1} \exp[-(-x)^\alpha] & \text{si } x \leq 0, \alpha > 0 \\ 0 & \text{sinon} \end{cases} \quad (2.15)$$

$$\Lambda(x) = \exp[-(x + \exp(-x))], \quad x \in \mathbb{R}. \quad (2.16)$$

FIGURE 2.1 – Représentation graphique de la densité des valeurs extrêmes avec $\alpha = 1$

2.3.2 Distributions généralisées des valeurs extrêmes

la distribution **généralisée des valeurs extrêmes** (*GEV*) a l'avantage de combiner les trois types de lois des valeurs extrêmes : distributions de Fréchet, de Weibull et de Gumbel. Cette dernière a été proposée par Jenkinson (1955) [19]. Le théorème précédent montre alors que la loi limite des extrêmes a toujours la même forme.

Définition 2.3.2

Soit $\xi \in \mathbb{R}$, $\mu \in \mathbb{R}$ et $\sigma > 0$.

La forme générale de la fonction de répartition d'une *GEV* est :

$$H(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{\frac{-1}{\xi}} \right\}, \quad -\infty < x < +\infty \quad (2.17)$$

défini sur l'ensemble $x : 1 + \xi \frac{(x-\mu)}{\sigma} > 0$

Les trois paramètres de H , μ , σ , et ξ sont respectivement ses paramètres de position, d'échelle, et de queue. Chacun des trois types de loi des valeurs extrêmes est caractérisé par le signe du paramètre de forme ξ de la loi *GEV* : pour $\xi > 0$ on a une loi de type Fréchet, pour $\xi < 0$ on a une de type Weibull. En effet pour $\xi \rightarrow 0$, $H(x)$ correspond à la loi de Gumbel donnée par l'équation 2.5 qui est en fait la limite, lorsque ξ tend vers zéro de l'équation :

$$H(x) = \exp \left[- \exp \left\{ - \left(\frac{x - \mu}{\sigma} \right) \right\} \right], \quad -\infty < x < +\infty \quad (2.18)$$

Définition 2.3.3 (Distribution max-stable)

On dit que la distribution H est **max-stable** si pour tout $n = 2, 3, \dots$, il existe des constantes $\alpha_n > 0$ et $\beta_n \in \mathbb{R}$ tel que :

$$H^n(\alpha_n x + \beta_n) = H(x)$$

H^n est la fonction de distribution de $M_n = \max X_{(1)}, \dots, X_{(n)}$, où $X_{(i)}$ sont des variables indépendantes, chacune ayant une fonction de distribution H .

Max-stabilité est une propriété satisfaite par les distributions pour lesquelles le fonctionnement de l'échantillon maxima conduit à une distribution identique, à part un changement

d'échelle et d'emplacement.

La relation avec les lois limites de valeur extrême est le résultat suivant.

Théorème 2.3 [9]

Une distribution est à max-stable si et seulement si c'est une distribution de valeur extrême généralisée.

2.4 Domaine d'attraction

Définition 2.4.1

On dit qu'une distribution F appartient au domaine d'attraction du maximum de la distribution H , et on note $F \in \mathcal{D}(H_\xi)$, s'il existe deux suites normalisées $(a_n) \geq 0$ et $(b_n) \in \mathbb{R}$ tels que la condition soit vérifiée :

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H_\xi(x); \forall x \in \mathbb{R} \quad (2.19)$$

Selon le signe de ξ (indice de valeur extrême) on distingue trois domaines d'attraction. On dit que la fonction de répartition F est dans le domaine d'attraction de Fréchet, de Gumbel ou de Weibull selon que $\xi > 0$, $\xi = 0$, $\xi < 0$.

- *si $\xi > 0$, on dit que F appartient au domaine d'attraction de **Fréchet** et on notera $F \in \mathcal{D}(\text{Fréchet})$. Ce domaine d'attraction est celui des distributions à queues lourdes, i.e qui ont une fonction de survie à décroissance polynomiale. Comme exemple de lois appartenant à ce domaine d'attraction on a les lois de cauchy, de Pareto, de Chi-deux, de student, de Burr, de Fréchet, etc.*
- *si $\xi < 0$, on dit que F appartient au domaine d'attraction de **Weibull** et on notera $F \in \mathcal{D}(\text{Weibull})$. Ce domaine d'attraction est celui des fonction de survie dont le point terminal est fini. Pour le domaine d'attraction de Weibull on trouve les lois Uniforme, Beta de Weibull, etc.*
- *si $\xi = 0$, on dit que F appartient au domaine d'attraction de **Gumbel** et on notera $F \in \mathcal{D}(\text{Gumbel})$. Ce domaine d'attraction est celui des distributions à queues légères, i.e qui ont une fonction de survie à décroissance exponentielle. Dans ce domaine d'attraction, on regroupe les lois Normale, Exponentielle, Log-normale, Gamma, etc.*

2.5 Caractérisation des domaines d'attraction

Un problème important est celui de la caractérisation des domaines d'attraction. Un domaine d'attraction est l'ensemble des distributions pour lesquelles la distribution limite du maximum converge vers la même limite *GEV*.

Dans cette partie, on se contentera de donner des conditions sur la fonction de répartition F pour qu'elle appartienne à l'un des domaines d'attraction définis précédemment.

2.5.1 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. Les lois de ce domaines ont un point terminal x_F infini. Le domaine d'attraction de Fréchet est relié avec la variation régulière

Théorème 2.4 [15]

$F \in \mathcal{D}(\text{Fréchet})$ avec un indice de valeur extrême $\xi > 0$ si et seulement si $x_F = +\infty$ et $1 - F$ est une fonction à variations régulières d'indice $-1/\xi$. Dans ce cas, un choix possible pour les suites a_n et b_n est :

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

De ce théorème 2.4, on déduit que $F \in \mathcal{D}(\text{Fréchet})$ avec un indice de valeur extrême $\xi > 0$ si et seulement si le point terminal x_F est infini et $1 - F(x) = x^{-1/\xi} \ell(x)$, où ℓ est une fonction à variations lentes à l'infini et ξ un réel strictement positif.

En effet, une distribution appartient au domaine d'attraction de Fréchet si et seulement si elle est à variations régulières.

Exemple.

Si X_1, X_2, \dots c'est une séquence des variables indépendante de Fréchet. $F(x) = \exp(-\frac{1}{x})$ pour $x > 0$; $a_n = n$ et $b_n = 0$

$$\begin{aligned} P(M_n - b_n)/a_n \leq z &= F^n(nz) \\ &= [\exp -1/(nz)]^n \\ &= \exp(-1/z) \end{aligned}$$

comme $n \rightarrow \infty$, pour chaque z fixe. Par conséquent, la limite dans ce cas.

Ce qui est un résultat exact pour tout n , en raison de la stabilité maximale de F est aussi la distribution Fréchet.

2.5.2 Domaine d'attraction de Weibull

Rappelons que toutes les lois appartenant au domaine d'attraction de Weibull ont un point terminal x_F fini. La caractérisation des fonctions de répartition de ce domaine passe du domaine d'attraction de Fréchet par un simple changement de variable dans la fonction de distribution.

Théorème 2.5 [15]

$F \in \mathcal{D}(\text{Weibull})$ avec un indice de valeur extrême $\xi < 0$ si et seulement si $x_F < +\infty$ et $1 - \tilde{F}$ est une fonction à variations régulières d'indice $1/\xi$ avec :

$$\tilde{F}(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ F(x_F - x^{-1}) & \text{si } x > 0. \end{cases} \quad (2.20)$$

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

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

Du théorème 2.5 on en déduit que $F \in \mathcal{D}(\text{Weibull})$ si et seulement si le point terminal x_F est fini et $1 - F(x) = (x_F - x)^{-1/\xi} \ell[(x_F - x)^{-1}]$ avec ℓ est une fonction à variations lentes à l'infini et ξ un réel strictement négatif.

Exemple.

On suppose que X est une variable aléatoire suivant une loi uniforme sur $[0, 1]$, on a donc

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

avec $x_F = 1 \leq \infty$

D'autre part $1 - \tilde{F} = 1 - F(x_F - x^{-1}) = 1 - (1 - x^{-1}) = x^{-1}\ell(x)$, avec $\ell(x) = 1$

Donc les conditions du théorème 2.5 sont vérifiées alors $F \in \mathcal{D}(\text{Weibull})$

Les constantes de normalisation sont données par

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

2.5.3 Domaine d'attraction de Gumbel

On sait que le domaine d'attraction de Gumbel ne contient que les lois à queues légères. Les résultats concernant de ce domaine sont plus délicats contrairement aux deux autres domaines.

Définition 2.5.1 (*Fonction de Von-Mises*)

Soit F une fonction de répartition de point terminal x_F fini ou infini. S'il existe $z < x_F$ tel que :

$$1 - F(x) = c \exp \left\{ - \int_z^x \frac{1}{a(t)} dt \right\}, \quad z < x < x_F, \quad (2.21)$$

où $c > 0$ et a est une fonction positive absolument continue de densité $a'(x) = 0$, alors F est une **fonction de Von-Mises** et a est sa fonction auxiliaire.

Théorème 2.6 [15]

$F \in \mathcal{D}(\text{Gumbel})$ si et seulement si il existe une fonction de Von-Mises F^* telle que pour $z < x < x_F$ on ait :

$$1 - F(x) = c(x)[1 - F^*] = c(x) \exp \left\{ - \int_z^x \frac{1}{a(t)} dt \right\} \quad (2.22)$$

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

Exemple.

Soit X une variable aléatoire de distribution $F(x) = 1 - \exp(-\lambda x)$ pour $x \geq 0$ et $\lambda > 0$

On a $1 - F(x) = 1 - [1 - \exp(-\lambda x)] = \exp(-\lambda x)$

F est une fonction de Von-Mises de fonction auxiliaire $a(x) = \lambda^{-1}$

Donc d'après le théorème 2.1 on a $F \in \mathcal{D}(\text{Gumbel})$

Proposition 2.5.1

Les variables aléatoires dans les trois domaines d'attraction de Gumbel, de Fréchet et de Weibull sont liées par la relation suivante :

$$X \in \mathcal{D}(\Phi) \Leftrightarrow \log(X^{\frac{1}{\xi}}) \in \mathcal{D}(\Lambda) \Leftrightarrow -X^{-1} \in \mathcal{D}(\Psi)$$

2.6 Distribution conditionnelle des excès

La méthode des excès initialement présentée par Pickands(1975)[25]. Elle préconise de ne retenir que les observations dépassant un seuil fixé u . La loi des n observations ainsi retenues que l'on note par $Y_i, i = 1, \dots, n$ peut-être approchée, si u est assez grand par une loi de Pareto généralisée (*GPD*) (voir Définition 2.6.2). Pour estimer les extrêmes, il suffit d'utiliser le résultat de (Balkema et de Hahn(1974) [2], Pickands[25]) qui établit l'équivalence entre la convergence en loi du maximum vers une loi des valeurs extrêmes et la convergence en loi d'un excès vers une *GPD*.

Définition 2.6.1

On appelle excès de la variable aléatoire X au dessus d'un seuil $u \leq x_F$ la variable aléatoire Y , qui prend ses valeurs dans $]0, x_F - u[$, définie par :

$$Y = \{X - u \mid X > u\}, \quad u \leq x_F$$

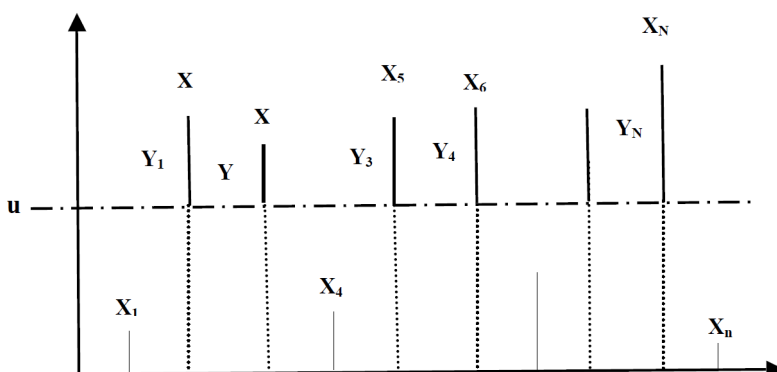


FIGURE 2.2 – Méthode des excès : u réel suffisamment élevé appelé seuil, Y : excès de X au-delà de u

2.6.1 Distribution de Paréto généralisée GPD

La distribution de Paréto généralisée (*GPD*) joue un rôle essentiel dans la modélisation des excès.

Définition 2.6.2

Soit $\sigma > 0$ et $\xi \in \mathbb{R}$. La loi de *Paréto généralisée* est définie par sa fonction de répartition :

$$G_{\xi, \sigma} = \begin{cases} 1 - (1 + \xi x / \sigma)^{-1/\xi} & \text{si } \xi \neq 0, \\ 1 - e^{-x} & \text{si } \xi = 0, \end{cases}$$

où $x \in \mathbb{R}_+$ si $\xi \geq 0$ et $x \in [0, -\sigma/\xi[$ si $\xi < 0$.

Dans l'expression précédente, σ représente le paramètre d'échelle et ξ l'indice des valeurs extrêmes.

La loi *GPD* présente quelques particularités. En voici une liste non exhaustive :

- Si $\sigma = 1$, on parle la loi *GPD* standard.
- Si $\xi = 0$, la *GPD* correspond à une loi exponentielle d'espérance σ .
- Si $\xi = -1$, elle correspond à une loi uniforme sur $[0, \sigma]$.
- Si $\xi > 0$, on retrouve la loi de Paréto décentrée.

Définition 2.6.3 (Distribution des excès)

On appelle *distribution des excès* de la variable aléatoire X par rapport à un seuil $u < x_F$ la loi de probabilité de la variable aléatoire Y excès de X au-delà du seuil $u < x_F$, donnée par sa fonction de répartition F_u , qu'on appelle *fonction de distribution des excès* :

$$\forall y \in \mathbb{R}, \quad F_u(y) = P[X - u \leq y \mid X > u] = \begin{cases} 0 & \text{si } y \leq 0 \\ 1 - \frac{1 - F(u+y)}{1 - F(u)} & \text{si } 0 < y < x_F - u, \\ 1 & \text{si } y \geq x_F - u \end{cases}$$

Lorsque le seuil u est grand, on peut approcher cette quantité par la fonction de survie d'une loi *GPD*.

2.6.2 Théorème de Pickands

Le théorème suivant montre qu'on peut approcher la distribution des excès au-delà d'un seuil u , sous certaines conditions de convergence par une loi de Pareto généralisée GDP

Théorème 2.7 [25]

F appartient au domaine d'attraction de H_ξ si et seulement si :

$$\lim_{u \rightarrow x_F} \sup_{x \in]0, x_F - u[} |F_u(x) - G_{\xi, \sigma(u)}(x)| = 0$$

où σ est une fonction strictement positive.

D'après ce résultat, si pour une fonction de répartition F inconnue, l'échantillon des maxima normalisés converge en loi vers une distribution non dégénérée, alors il s'en déduit que la distribution des excès au-dessus d'un seuil élevé converge vers une GPD lorsque le seuil tend vers la limite supérieure du support de F . Cette caractérisation est à la base des méthodes d'estimation de type Peaks Over Threshold (POT).

2.7 Estimation de l'indice des valeurs extrêmes

Dans cette section on s'intéresse à l'estimation de l'indice de queue ξ qui joue un rôle essentiel dans le comportement de la loi des extrêmes. On utilise les références suivants [3],[9],[11],[22]

Dans la littérature de la théorie des valeurs extrêmes on trouve plusieurs techniques pour l'estimation de l'indice de queue. On peut citer l'estimateur de Hill qu'il est valable pour $\xi > 0$. Il est considéré comme le plus simple des estimateurs de l'indice de queue. Pour pallier les limitations de l'estimateur Hill, mais aussi pour l'étendre aux deux autres domaines d'attractions Dekkers et al. (1989) [11] ont proposé un estimateur des moments valable quel que soit $\xi \in \mathbb{R}$

On présente ici deux estimateurs les plus populaires dans la littérature sont les estimateurs de Hill et de Pickands. Tous basés sur les statistiques d'ordre $X_{(1,n)} \leq \dots \leq X_{(n,n)}$ associées à l'échantillon $X_{(1)}, \dots, X_{(n)}$. 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 $\frac{k}{n} \rightarrow 0$.

2.7.1 Estimateur de Hill

Cet estimateur a été introduit par Hill (1975) pour estimer d'une manière non-paramétrique le paramètre de queue des lois appartenant au $\mathcal{D}(\text{Fréchet})$. Pour construire son estimateur, Hill utilise la méthode du maximum de vraisemblance sur l'ensemble des k_n plus grandes observations d'un échantillon.

Définition 2.7.1

Soit $(k_n)_{n \geq 1}$ une suite d'entiers avec $1 < k_n \leq n$, l'estimateur de Hill est défini par :

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

Théorème 2.8 (Propriétés de l'estimateur de Hill)[22]

Soit $(k_n)_{n \geq 1}$ une suite d'entiers telle que $1 < k_n \leq n$, $k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$ quand $n \rightarrow \infty$

- Alors, $\hat{\xi}_{k_n}^H$ converge en probabilité vers ξ .
- Si de plus $\frac{k_n}{\log \log n} \rightarrow \infty$ quand $n \rightarrow \infty$, alors $\hat{\xi}_{k_n}^H$ converge presque sûrement vers ξ

Preuve : Pour la démonstration de ce théorème. On se réfère au ouvrage suivant [11]

Pour établir la normalité asymptotique de l'estimateur $\hat{\xi}_{k_n}^H$, on a besoin d'une hypothèse sur la fonction à variations lentes ℓ . Il est en effet nécessaire d'imposer une condition qui spécifie la vitesse de convergence du rapport des fonctions à variations lentes vers 1 telle que défini au théorème 2.1

(C.1) : Il existe une constante réelle $\rho < 0$ et une fonction $\varepsilon(x) \rightarrow 0$ quand $x \rightarrow \infty$, telle que pour tout $\lambda > 1$:

$$\log \frac{\ell(\lambda x)}{\ell(x)} \sim \varepsilon(x) \frac{\lambda^\rho - 1}{\rho} \text{ quand } x \rightarrow \infty$$

Cette condition appelée « condition du second ordre » est satisfaite pour la plupart des lois appartenant au \mathcal{D} (Fréchet). Plus la constante $\rho < 0$ de la condition **(C.1)** est proche de zéro, plus difficile est l'estimation de l'indice de queue ξ .

Remarque 2.2

La condition **(C.1)** implique que $\forall \varepsilon > 0, \exists x_0$, tel que $\forall x \geq x_0, \forall \lambda > 1$:

$$\frac{(1 + \varepsilon)\lambda^{\rho+\varepsilon} - 1}{\rho} \leq \frac{1}{\varepsilon(x_n)} \log \left(\frac{\ell(\lambda x)}{\ell(x)} \right) \leq \frac{(1 - \varepsilon)\lambda^{\rho-\varepsilon} - 1}{\rho}$$

Théorème 2.9 (Normalité asymptotique de l'estimateur de Hill)[22]

Soit $(k_n)_{n \geq 1}$ une suite d'entiers telle que $1 < k_n \leq n$, $k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$. Si la condition **(C.1)** est satisfaite avec $\sqrt{k_n} \varepsilon(\frac{n}{k_n}) \rightarrow 0$ quand $n \rightarrow \infty$, alors :

$$\sqrt{k_n} \left(\hat{\xi}_{k_n}^H - \xi \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \xi^2)$$

En pratique, le choix du paramètre k_n pose des problèmes. Si l'on trace le diagramme de Hill (voir Figure 2.3) i.e la fonction $k_n \rightarrow \hat{\xi}_{k_n}^H$, on observe une extrême volatilité qui rend difficile l'utilisation de cet estimateur en pratique si l'on n'a aucune indication sur le choix de k_n . De plus, cet estimateur est biaisé (voir Figure 2.4). Ce biais est de l'ordre de $\varepsilon(\frac{n}{k_n})$. La condition $\sqrt{k_n} \varepsilon(\frac{n}{k_n}) \rightarrow 0$ impose au biais d'être négligeable devant l'écart type de l'estimateur qui est quand à lui égal à $\sqrt{k_n}$.

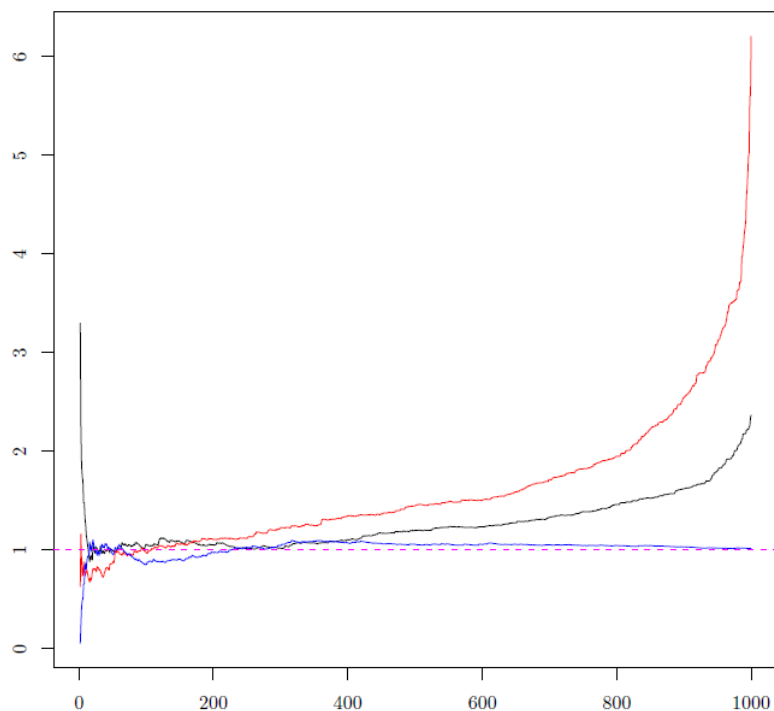
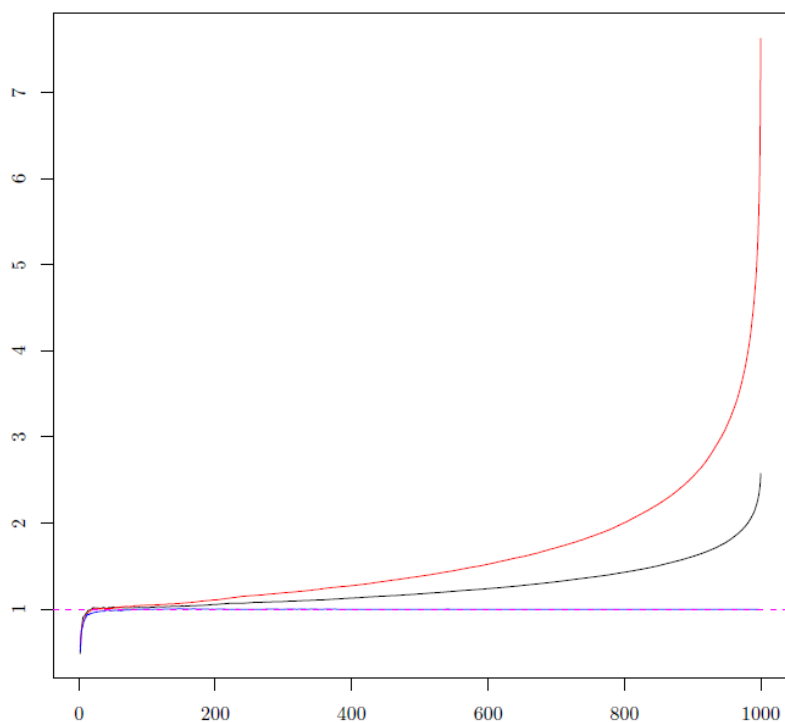
Le résultat sur la normalité asymptotique de l'estimation de Hill permet de donner un intervalle de confiance pour l'estimation. En pratique on se contentera de remplacer ξ par sa valeur estimée. Par conséquent, si k_n est petit, on aura a fortiori, compte tenu des remarques faites précédemment, une estimation avec un intervalle de confiance large et a contrario, si k_n est grand, on aura un intervalle de confiance plus étroit mais pas centrée sur la vraie valeur.

Remarque 2.3

Le graphe de la Figure 2.3 contient trois lois appartenant au domaine d'attraction de Fréchet. La loi de Pareto (bleu), Burr (rouge), Fréchet (noir)

Remarque 2.4

Le graphe de la Figure 2.4 contient trois lois appartenant au domaine d'attraction de Fréchet. La loi de Pareto (bleu), Burr (rouge), Fréchet (noir) et la vraie valeur de $\xi = 1$ en trait interrompu. En ordonnée on a l'indice de queue estimé et en abscisse le seuil k_n . Pour une loi de Pareto comme $\varepsilon = 0$ alors, il n'y a pas de biais asymptotique.

FIGURE 2.3 – Graphique de Hill à $n = 1000$ et la vraie valeur de $\xi = 1$ FIGURE 2.4 – Graphique de Hill en moyenne sur 100 réalisations à $n = 1000$.

2.7.2 Estimateur de Pickands

L'estimateur de Pickands est construit en utilisant trois statistique d'ordres. Cet estimateur a l'avantage d'être valable quel que soit le domaine d'attraction de la distribution et par conséquent, du domaine de définition de l'indice des valeurs extrêmes. Pickands (1975)[25] démontre la consistance faible de son estimateur. La convergence forte ainsi que la normalité asymptotique ont été démontrées par Dekkers et de Haan (1989)[10].

Définition 2.7.2

Soit $X_i, i = 1, \dots, n$ une suite de variables aléatoires indépendantes de loi F 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 :

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

Théorème 2.10 [10]

Soit $(k_n)_{n \geq 1}$ une suite d'entiers telle que $1 \leq k_n < n$, $k_n \rightarrow \infty$ et $\frac{k_n}{n} \rightarrow 0$ quand $n \rightarrow \infty$

- Alors, $\hat{\xi}_{k_n}^P$ converge en probabilité vers ξ .
- Si de plus $\frac{k_n}{\log \log n} \rightarrow \infty$ quand $n \rightarrow \infty$, alors $\hat{\xi}_{k_n}^P$ converge presque sûrement vers ξ .
- Sous des conditions additionnelles sur la suite k_n et la fonction de répartition F que l'on pourra consulter dans Dekkers et Hahn (1989) [10]

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

Comme l'estimateur de Hill, cet estimateur est biaisé et le résultat sur sa normalité asymptotique permet de donner un intervalle de confiance pour l'estimation.

Compte tenu de la variance asymptotique de $\hat{\xi}_{k_n}^P$ qui est assez importante comparativement à $\hat{\xi}_{k_n}^H$ (voir Figure 2.5), certains auteurs ont proposés des estimateurs à « variance minimale » construits à partir de combinaisons linéaires des logarithmes des accroissements des statistiques d'ordres. Par exemple Drees (1995)[12] propose de faire la moyenne de plusieurs estimateurs de Pickands utilisant un nombre de plus grandes observations différents dans le but d'obtenir un estimateur moins sensible au choix de k_n .

2.7.3 Estimateur de Zipf

Dans le but d'améliorer le biais asymptotique des estimateurs précédents, Kratz et Resnick (1996) [20] et Schultze et Steinebach (1996) [29] ont indépendamment proposé d'estimer l'indice de queue par la méthode des moindres carrés classique. Leur estimateur connu sous le nom de Zipf est asymptotiquement gaussien. Il est défini par :

$$\hat{\xi}_{k_n}^Z = \frac{\frac{1}{k_n} \sum_{j=1}^{k_n} \log \frac{k_n+1}{j} \log X_{n-j+1,n} - \frac{1}{k_n} \sum_{j=1}^{k_n} \log \frac{k_n+1}{j} \left(\frac{1}{k_n} \sum_{j=1}^{k_n} \log X_{n-i+1,n} \right)}{\frac{1}{k_n} \sum_{j=1}^{k_n} \left(\log \frac{k_n+1}{j} \right)^2 - \left(\frac{1}{k_n} \sum_{j=1}^{k_n} \log \frac{k_n+1}{j} \right)^2}$$

Toutefois, sa variance asymptotique deux fois supérieur à celle de l'estimateur de Hill.

2.7.4 Estimation de vraisemblance maximale

Le principe de l'estimation du maximum de vraisemblance (*MLE*) indique que la distribution de probabilité souhaitée est celui qui rend les données observées «les plus probables». Supposant que $X_{(1)}, \dots, X_{(n)}$ sont des variables indépendantes ayant la distribution GEV, la log-vraisemblance des paramètres GEV lorsque $\xi \neq 0$ est

$$L(\mu, \sigma, \xi) = -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^m \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^m \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi} \quad (2.23)$$

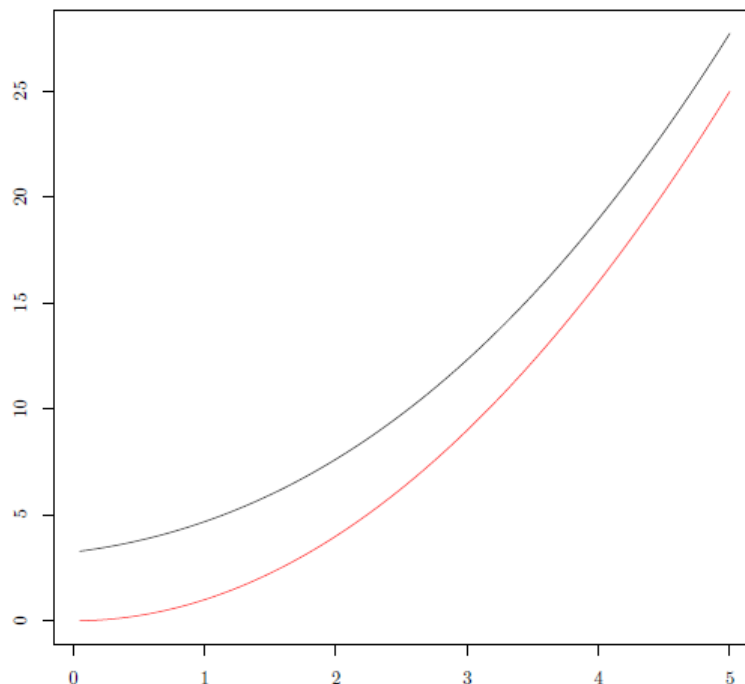


FIGURE 2.5 – La variance asymptotique de l'estimateur $\hat{\xi}_{k_n}^P$ (noir) et de $\hat{\xi}_{k_n}^H$ (rouge) en fonction de ξ

à condition que

$$1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) > 0, \quad i = 1, \dots, m \quad (2.24)$$

Aux combinaisons de paramètres pour lesquelles 2.24 est violé, correspondant à une configuration pour laquelle au moins une des données observées dépasse un point final de la distribution, la probabilité est nulle et la log-vraisemblance est égal à $-\infty$

Le cas $\xi = 0$ nécessite un traitement séparé en utilisant la limite de Gumbel de la distribution GEV. Cela conduit à la log-vraisemblance

$$L(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^m \left(\frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^m \exp \left\{ - \left(\frac{x_i - \mu}{\sigma} \right) \right\} \quad (2.25)$$

2.7.5 Estimation de périodes de retour

En hydrologie par exemple, on s'intéresse particulièrement à l'estimation d'une période de retour T associée à une probabilité $p (= \frac{1}{10}, \frac{1}{100}, \dots)$ représentant la probabilité d'excès au-delà d'un niveau de retour x_F . Cette période s'interprète comme une moyenne du temps ou du nombre d'années séparant un événement de grandeur donnée (x_F) d'un second événement d'une grandeur égale ou supérieure.

D'un point de vue mathématique, suivant qu'on utilise l'approche *GPD* ou *GEV*, deux définitions de la période de retour sont possibles :

- si on utilise l'approche *GPD* : on utilise le fait que la loi des excès au-delà d'un seuil u peut être uniformément approchée par une loi *GPD* (ξ, σ) . En prenant pour seuil $X_{n-k+1,n}$, on a

$$T = \frac{n}{k+1} \frac{1}{1 - G_{\xi, \sigma}(x_F)}$$

- si on utilise l'approche *GEV* : on a directement

$$T = \frac{1}{1 - H(x_F)}$$

Dans le but d'estimer cette période de retour, si on utilise l'approche *GPD*, L'équation de cette droite passant par le point $(\log \frac{n}{k+1}, \log X_{n-k+1,n})$ de pente ξ est donnée par

$$y = \log X_{n-k+1,n} + \xi \left(x - \log \left(\frac{n}{k+1} \right) \right)$$

Nous avons donc la relation suivante entre la période de retour T et le débit X :

- si $\xi > 0$:

$$\log(X) = \log X_{n-k+1,n} + \hat{\xi} \left(\log(T) - \left(\log \frac{n}{k+1} \right) \right) \quad (2.26)$$

- si $\xi = 0$:

$$X = \log X_{n-k+1,n} + \hat{\sigma} \left(\log(T) - \left(\log \frac{n}{k+1} \right) \right) \quad (2.27)$$

Dans le cas de l'approche *GEV*, on peut estimer l'indice des valeurs extrêmes ξ par l'un des estimateurs précédente et obtenir ainsi un estimateur de la période de retour.

Conclusion

Dans ce chapitre, nous avons fait un aperçu général sur la théorie des valeurs extrêmes, en mentionnant la distribution asymptotique d'un extrême ainsi que la distribution conditionnelle des excès et les différentes caractéristiques de domaines d'attraction.

Le chapitre suivant s'intéresse aux champs pratiques de la *TVE*.

Les champs pratiques de la TVE

Les références utilisées dans ce chapitre sont : [4],[18]

Introduction

La modélisation des évènements extrêmes (ouragan, tremblement de terre ou inondation, crues, crises financières, krachs, chocs pétroliers) est aujourd'hui un champ de recherches particulièrement actif, notamment par l'importance de leurs impacts économiques et sociaux. En particulier, depuis quelques années, on note un intérêt croissant pour l'application de la Théorie des Valeurs Extrêmes (*TVE*) pour la modélisation de tels évènements. On présente dans ce chapitre une application en hydrologie, spécifiquement sur des données réelles de débits horaires pour la rivière de Molenbeek à Erpe-Mere (Belgique) de 1986 à 1996. Ensuite, on applique des techniques univariées s'inspirant de la TVE à une série de taux de change quotidiens entre le dollars canadien et le dollars américain sur une période de 5 ans (1995-2000).

3.1 Application en hydrologie

3.1.1 Présentation des données hydrologiques

L'analyse des valeurs extrêmes a beaucoup d'applications en hydrologie. Elle est utilisée pour évaluer les risques d'inondation, pour l'élaboration de cartes d'inondations, ...etc. L'analyse est le plus souvent basée sur les débits de rivière, spécifiquement sur des données réelles de débits horaires pour la rivière de Molenbeek à Erpe-Mere (Belgique) de 1986 à 1996.

Les différentes méthodes d'estimation présentées dans le chapitre 3 seront appliquées à des débits horaires concernant la rivière Molenbeek. Des observations de débits limniographiques sont disponibles dans la station Erpe-Mere qui a une superficie de 47km^2 .

La Figure 3.1 illustre l'influence des inondations sur la courbe de tarage pour la rivière Molenbeek à Erpe-Mere avant 1996. Cette influence a été étudiée dans Willems et al.(2002)[32] au moyen d'un modèle hydraulique pour la rivière et les zones inondables avoisinantes.

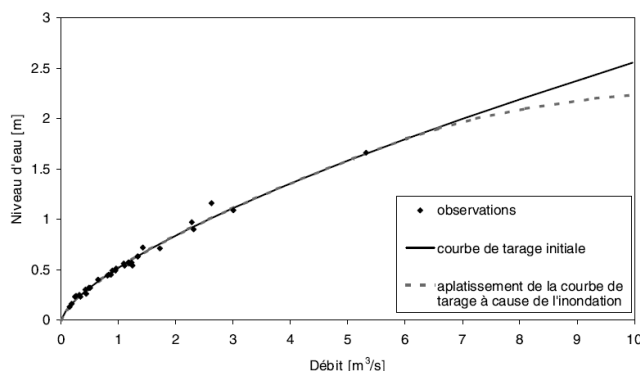


FIGURE 3.1 – Influence des inondations sur la courbe de tarage pour la rivière Molenbeek à Erpe-Mere avant 1996

Dans les études de modélisation des inondations, une description précise des débits de la rivière et des périodes de retour est indispensable. Le calcul d'une période de retour peut être basé sur une série temporelle des débits de rivière ou des mesures du niveau d'eau. Il peut également être basé sur des résultats de simulations découlant d'un modèle hydrologique et/ou hydraulique.

La Figure 3.2 donne un exemple des différences entre le débit des rivières d'une part et l'écoulement d'autre part pour la rivière Molenbeek à Erpe-Mere en Belgique. Le phénomène peut physiquement s'expliquer ici d'une façon conceptuelle par l'accumulation de l'eau dans la zone inondable de la rivière dès que le débit dépasse la capacité du lit de la rivière. Une illustration schématique de ce concept est donnée dans la Figure 3.3

Dans la mesure où les débits disponibles sont horaires, parfois seulement journaliers, il est peu réaliste de considérer la série comme formée de variables aléatoires indépendantes. Or si cette hypothèse, sous-jacente à de nombreuses théories statistiques et en particulier à celle relative à la théorie des valeurs extrêmes, n'est pas satisfaite, des résultats aberrants peuvent en résulter. La plupart des travaux théoriques dans l'approche POT. (Shane et Lynn 1964 [30], Bernier 1967 [5], Todorovic 1970[31]) suppose simplement qu'une série de pic est disponible. Bien que cette hypothèse puisse être appropriée dans les études théoriques, les praticiens ont besoin de lignes de conduite pour l'obtention de cette série.

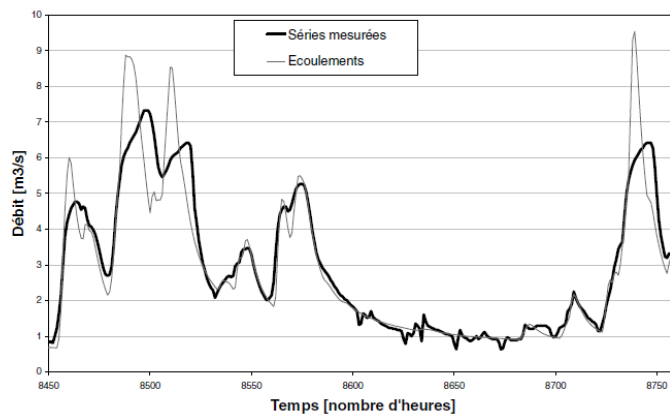


FIGURE 3.2 – Comparaison des débits horaires de la rivière et de l'écoulement pour la période d'inondation de décembre 1993

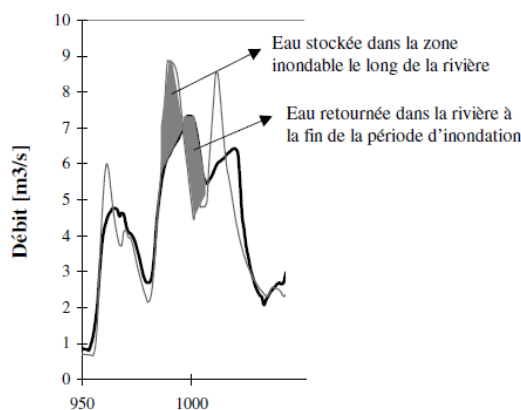


FIGURE 3.3 – Explication schématique de l'influence de l'inondation sur les hydrogrammes

Nous nous proposons donc, dans le suivant, de décrire le critère de sélection que nous utiliserons

3.1.2 Sélection d'observations indépendantes

Des pics d'inondations consécutifs seront considérés comme indépendants si l'intervalle de temps entre ces deux pics dépasse un temps critique et si le débit entre ces deux événements passe en dessous d'un niveau proche du débit de base, qui peut s'interpréter comme un débit «normal» au sens où il s'agit ni d'une période d'inondation, ni d'une période de sécheresse. Différents critères de sélection ont été proposés dans la littérature (Lang et al.1999 [21], Claps et Laio 2003 [8]). Nous décrivons ci-dessous plus en détail celui que nous allons utiliser et nous illustrons dans la Figure 3.4,

Plus spécifiquement, deux pics consécutifs seront considérés comme indépendants quand le temps p entre ces deux pics est plus long que le facteur de récession k , et quand le débit minimum entre ces deux pics est plus petit qu'une fraction f du pic du débit. En utilisant ce critère, pour éviter que des petits pics ne soient sélectionnés : il faut donc supposer que q_{max} soit supérieur à une valeur limite notée q_{lim} . La méthode doit donc vérifier les 3 critères suivants (Figure 3.5 pour la définition des paramètres) :

$$p > k, \quad \frac{q_{min}}{q_{max}} < f, \quad q_{max} > q_{lim}$$

la figure 3.6 représente une comparaison POT empiriques sélectionnés de la série de débits horaires pour différents niveaux d'indépendance. Dans la plupart des cas, il est pré-

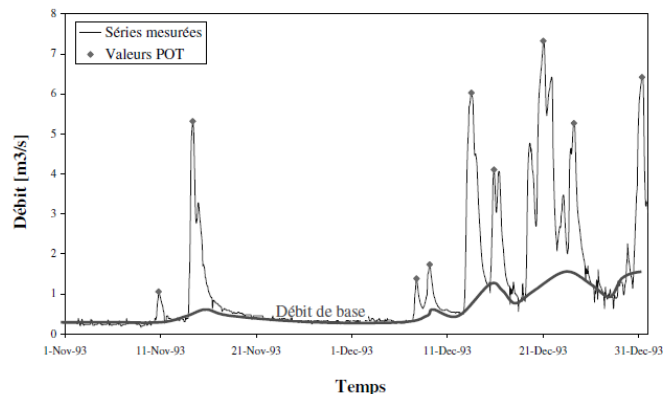


FIGURE 3.4 – Valeurs indépendantes POT sélectionnées à partir de la série de débits horaires

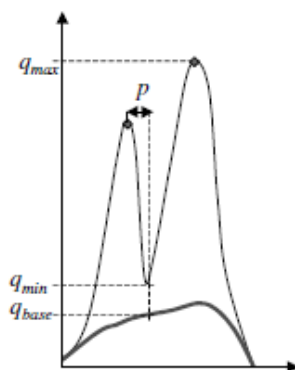


FIGURE 3.5 – Paramètres utilisés dans le critère de sélection des valeurs POT indépendantes

féral de sélectionner un grand nombre d'extrêmes *POT* de la série et donc de disposer de plus d'extrêmes indépendants.

Cette technique d'extraction de variables aléatoires indépendantes a été utilisée, de façon plus ou moins modifiée suivant le type d'applications, un tel critère d'indépendance influence le nombre d'extrêmes et aussi l'interprétation de la période de retour d'un événement extrême. Dans la plupart des cas pratiques cependant le critère subjectif aboutira à une grande indépendance, et ceci garantira asymptotiquement l'existence d'une distribution GPD dans la queue de distribution.

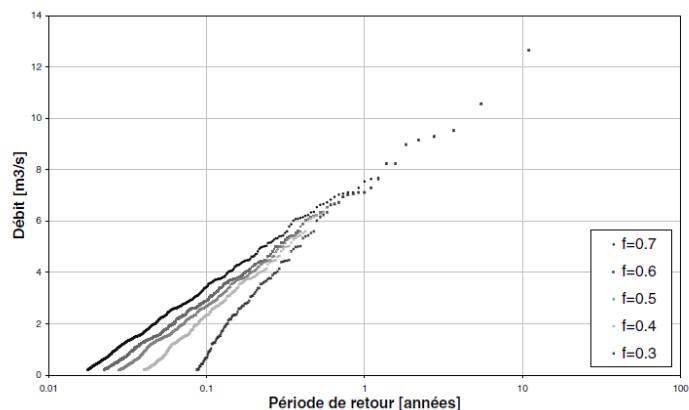


FIGURE 3.6 – les différences entre les valeurs POT sélectionnées pour différentes valeurs de f

3.1.3 Simulations

Dans le cas de Molenbeek, les valeurs POT indépendantes suivent une distribution des valeurs extrêmes exponentielle (loi du Gumbel), i.e indice des valeurs extrêmes $\xi = 0$. Ceci est illustré dans la Figure 3.9.

les inondations commencent à avoir lieu pour un débit supérieur à $5.3m^3/s$ et c'est pour cette raison que pour les débits supérieurs à ce seuil, nous voyons une certaine courbure apparaître dans le graphe. De cette analyse les débits doivent être décomposés en au moins deux sous-populations. Les débits *POT* plus petits et plus grands que $5.3m^3/s$ suivent une loi exponentielle avec une pente différente dans la Figure 3.7 et la Figure 3.8.

Dans la Figure 3.7, la pente (droite) de la loi exponentielle pour les points non inondés est plus ou moins parallèle à celle pour les débits d'écoulement. Le petit décalage est expliqué par la propagation des débits le long de la rivière.

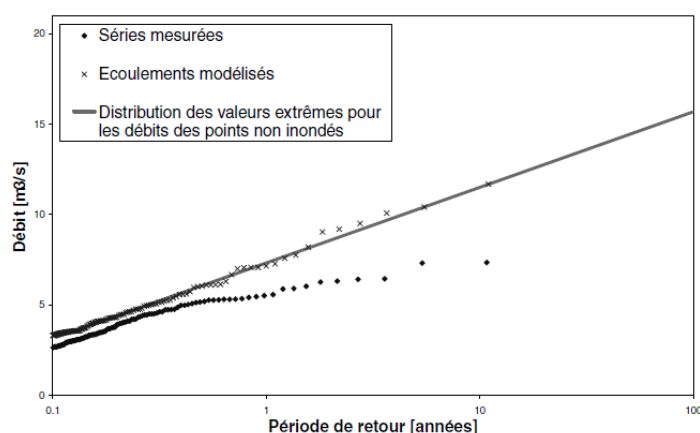


FIGURE 3.7 – Comparaison de la distribution empirique pour les débits horaires *POT* et les débits d'écoulement en amont pour la rivière

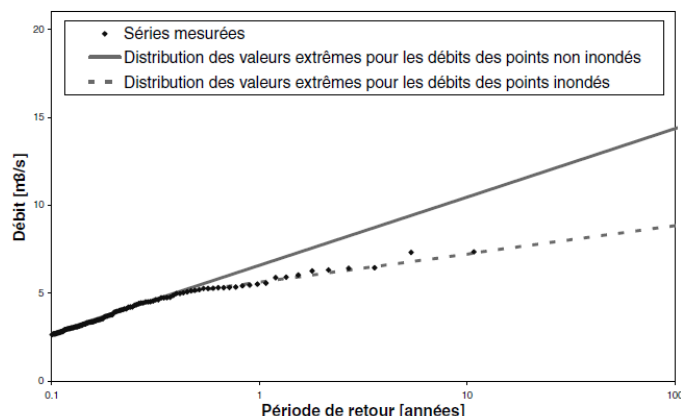


FIGURE 3.8 – Distribution des valeurs extrêmes décomposée en deux sous-populations pour les débits horaires de la rivière

La Figure 3.9 montre les différentes estimations pour des seuils k allant de 1 au nombre maximal de valeurs POT extraites (623). Pour des seuils élevés, le calcul de la pente est seulement basé sur un nombre limité d'observations. Les niveaux les plus bas, la MSE est plus petite et les estimations de la pente plus stables. C'est la gamme de seuils dans laquelle doit être sélectionné le seuil optimal (de préférence au rang ayant la MSE la plus

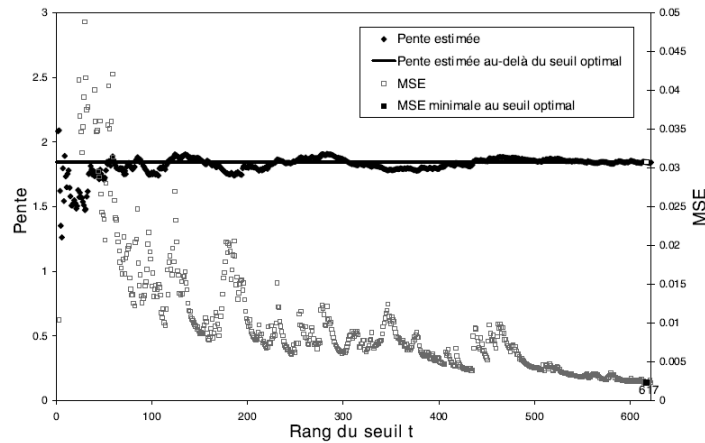


FIGURE 3.9 – Estimation de type Hill de la pente en fonction du seuil considéré pour les débits horaires pour une valeur de $f = 0.3$

faible, ici 617).

On a dans la figure 3.10 une translation des observations par moyenne de la fonction à variations lentes pour $f = 0.3$ pour déterminer la pente asymptotique.

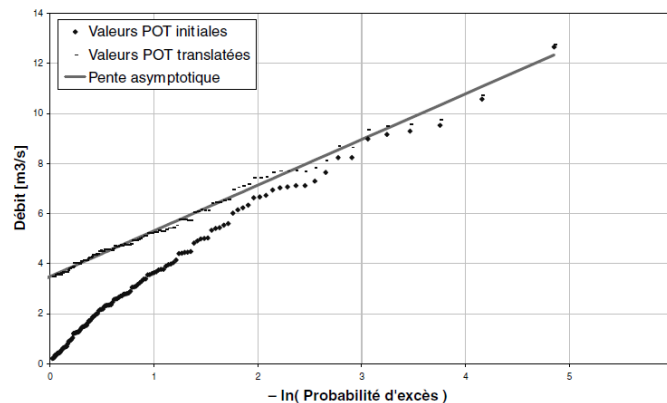


FIGURE 3.10 – Les valeurs POT translattées dans l'exponentiel quantile plot

Les valeurs POT représentées dans la figure 3.11 pour différents niveaux d'indépendance. Les différences sont faibles.

Conclusion

Les inondations peuvent biaiser la courbe de tarage des débits d'une rivière mesurés à partir d'une station limnigraphique, tel de la rivière de Molenbeek. Ce biais doit être éliminé pour éviter qu'une mauvaise distribution des valeurs extrêmes ne soit sélectionnée. Plusieurs estimateurs ont été présentés pour l'indice des valeurs extrêmes et les paramètres de la GPD. Des techniques de réduction de biais ont été appliquées pour tenir compte de la différence entre la pente dans le «quantile plot» basée sur un nombre limité d'observations et la pente asymptotique dans la distribution des valeurs extrêmes.

3.2 Application en économie

la théorie des valeurs extrêmes permet d'estimer la probabilité associée à des événements rares et modéliser l'incidence de krachs boursiers ou de tensions extrêmes sur les porte-

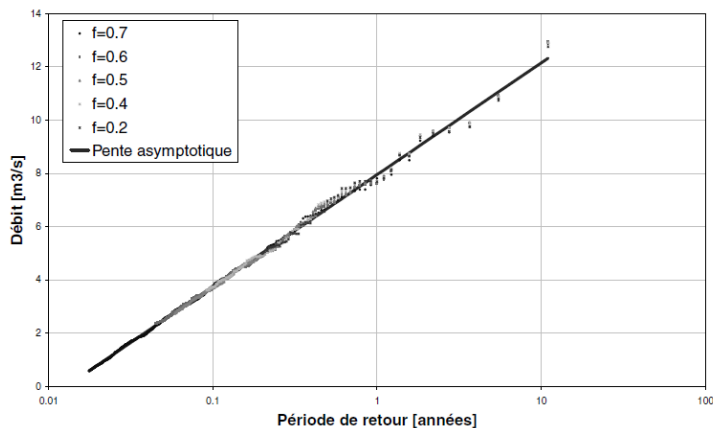


FIGURE 3.11 – comparaison des extrêmes POT translatés pour différents niveaux d'indépendance

feuilles des investisseurs. Dans cette section on applique des techniques univariées s'inspirant de la TVE à une série de taux de change quotidiens entre le dollars Canadien et le dollars américain sur une période de 5 ans (1995-2000).

3.2.1 Simulation

La Valeur à Risk VaR est définie par rapport à un horizon de temps T et le seuil de confiance u .

Le tableau suivant répertorie les résultats de la VaR pour différentes distributions. La VaR indiquée est la perte maximale qui ne peut être dépassé sur un horizon d'une journée avec un niveau de confiance (probabilité) donné. La VaR en dollars est donnée pour 10 milliards de dollars américains et canadiens.

Niveau de confiance	VaR-Normale ^a	VaR-HS ^b	VaR-GEV ^c	VaR-GPD ^d
95% (dans l'échantillon)	22	20	72	NA
97.5%(dans l'échantillon)	26	25	89	NA
99% (hors échantillon)	31	37	116	70
99.9%(hors échantillon)	42	61	NA	136
99.99%(hors échantillon)	50	70	NA	257
99.999%(hors échantillon)	584	71	NA	479

TABLE 3.1 – les résultats de la VaR pour différentes distributions

a. VaR en supposant une distribution normale.

b. VaR en supposant une distribution empirique

c. VaR en supposant une distribution GEV : les paramètres de GEV sont $\mu(0.0028)$, $\sigma(0.0009)$, et $\xi(0.2747)$

d. VaR utilisant la distribution GPD. Le GPD a été estimée en utilisant 90 observations au-dessus du seuil $u(0.0018)$ avec les paramètres $\xi(0.2643)$ et $\beta(0.0006)$.

3.2.2 Illustrations graphiques

On va illustrer dans cette section les techniques graphiques et décrivent la qualité de l'ajustement du différents modèles.

La figure 3.12 représente la queue de la distribution empirique.

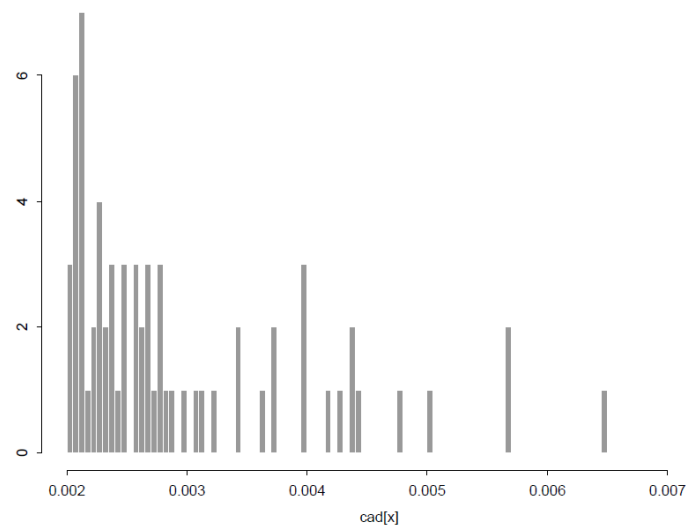


FIGURE 3.12 – Histogramme de la queue de la distribution empirique

La figure 3.13 montre le plot d'estimation de Hill avec indice d'extrême stable autour de valeurs de 3 à 4, ce qui est typique des séries financières.

Les figures 3.14, 3.15 illustre le Graphique Q-Q pour la distribution EVD, maxima à partir de blocs de différents observations.

D'ailleurs, Pour estimer le GPD, on commence par la distribution des excès. La forme du graphique correspond exactement à la forme d'un GPD illustré dans la figure 3.16.

Le graphique suivant 3.17 de la quantile est présenté pour différents seuils. Il n'est généralement pas satisfaisant d'avoir un seul résultat (une VaR extrême) ; on peut choisir le seuil dans une zone stable du graphe des quantiles.

conclusion

Les techniques EVT permettent de se concentrer sur le comportement de ces observations extrêmes. De plus, la perte au-dessus d'un très grand seuil peut être estimée. Bien que ces résultats s'appliquent bien dans le cas univarié.

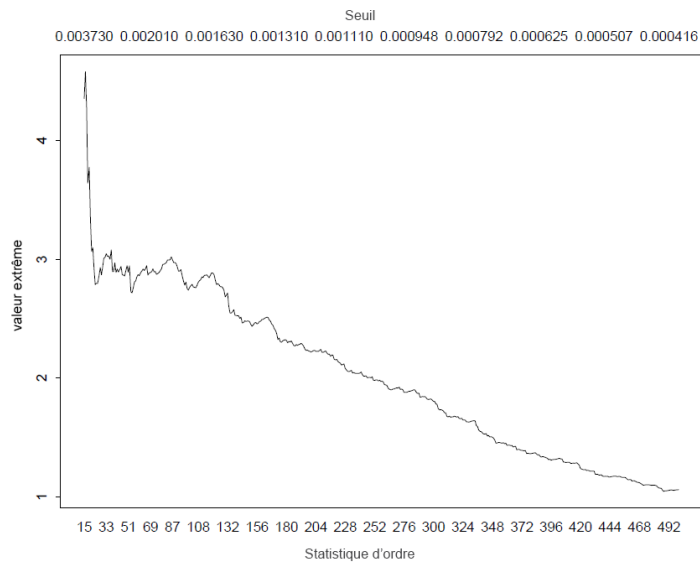


FIGURE 3.13 – estimation de Hill avec indice d’extrême autour de valeurs de 3 à 4

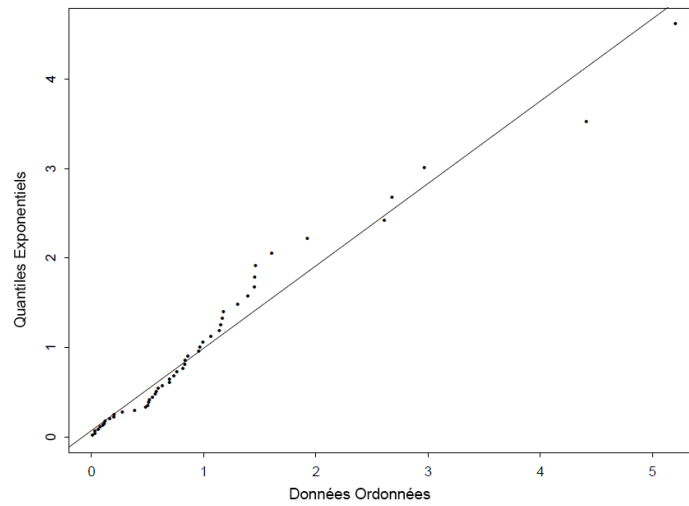


FIGURE 3.14 – Q-Q pour la distribution EVD, à partir de blocs de 28 observations

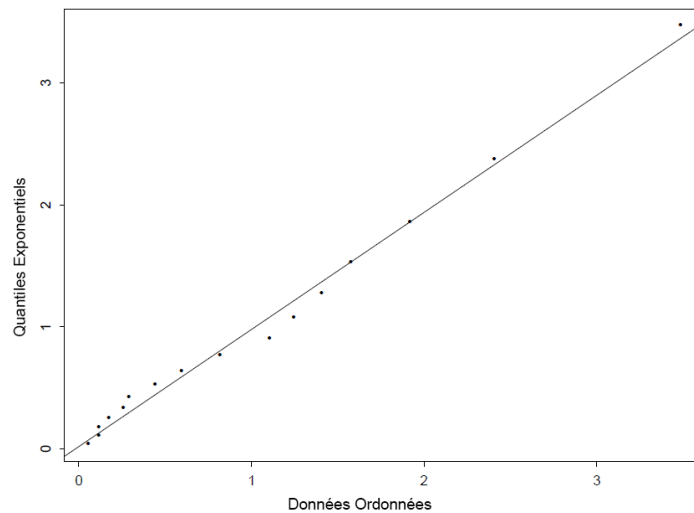


FIGURE 3.15 – Q-Q pour la distribution EVD, à partir de blocs de 90 observations

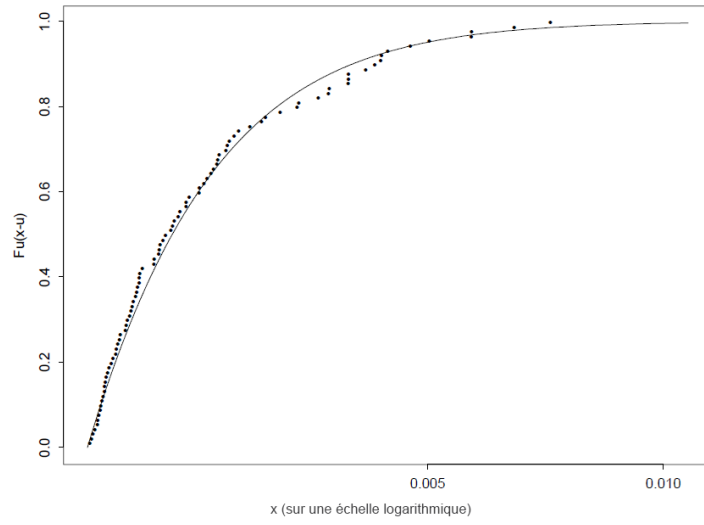


FIGURE 3.16 – la distribution excédentaire pour un GPD : 90 dépassements

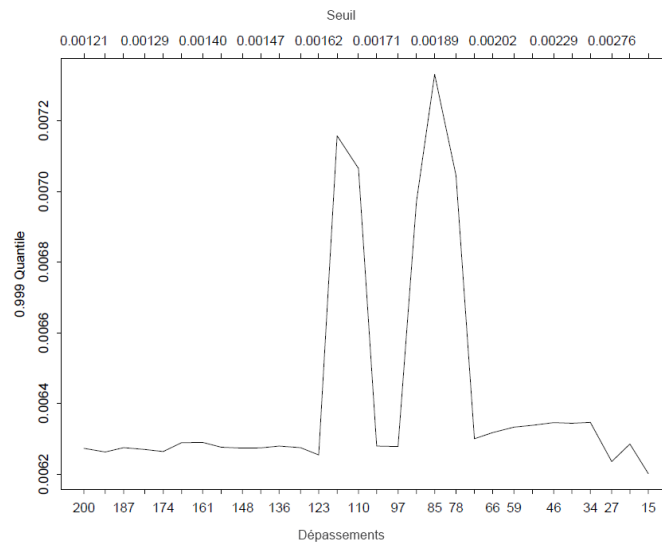


FIGURE 3.17 – Le 0.999 quantile de (0,001%) pour différents seuils

Conclusion générale

Dans ce travail, on est parti dans un premier temps par quelques rappels sur les notions de la statistique d'ordre et les différents types de convergences nécessaires pour l'étude du comportement asymptotique des estimateurs, ainsi que le théorème de la limite centrale.

Ensuite, nous avons effectué une synthèse sur la théorie des valeurs extrême dont on a mentionné les résultats fondamentaux, théorème de Fisher-Tippet-Gnedenko, qui est rapproché du théorème de la limite centrale et le théorème de Pickands qui établit l'équivalence entre la convergence en loi du maximum vers une loi des valeurs extrêmes et la convergence en loi d'un excès vers une GPD.

Par ailleurs, on a présenté les différentes distributions : distribution à variations régulières, distribution asymptotique d'un extrême, distribution conditionnelle des excès. On a traité les différents domaines d'attraction (Fréchet, Weibull, Gumbel). De plus, on a donné les méthodes d'estimations des extrêmes tels l'estimation de Hill, de Pickands, de Zipf, estimation de vraisemblance maximale, et l'estimation de période de retour.

Enfin pour le cas d'étude, on a choisi deux exemples de deux champs d'application différents, dont on a repris des données réelles de débits horaires pour la rivière de Molenbeek en Belgique entre 1986-1996, pour le domaine d'hydrologie, avec une modélisation du taux de change quotidiens entre le dollar canadien et le dollar américain dans la période 1995-2000, qui représente un cas d'étude effectuée en économie.

Bibliographie

- [1] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A first course in order statistics*, volume 54. Siam, 1992.
- [2] A. A. Balkema and L. De Haan. Residual life time at great age. *The Annals of probability*, pages 792–804, 1974.
- [3] J. Beirlant, Y. Goegebeur, J. Segers, and J. L. Teugels. *Statistics of extremes : theory and applications*. John Wiley & Sons, 2006.
- [4] Y. Bensalah et al. *Steps in applying extreme value theory to finance : a review*. Citeseer, 2000.
- [5] J. Bernier. Sur la théorie du renouvellement et son application en hydrologie. *Electricité de France*, (HYD67) :10, 1967.
- [6] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27. Cambridge university press, 1989.
- [7] A. Borchani. Statistiques des valeurs extrêmes dans le cas de lois discrettes. 2010.
- [8] P. Claps and F. Laio. Can continuous streamflow data support flood frequency analysis? an alternative to the partial duration series approach. *Water Resources Research*, 39(8), 2003.
- [9] S. Coles, J. Bawa, L. Trenner, and P. Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.
- [10] A. L. Dekkers, L. De Haan, et al. On the estimation of the extreme-value index and large quantile estimation. *The annals of statistics*, 17(4) :1795–1832, 1989.
- [11] A. L. Dekkers, J. H. Einmahl, and L. De Haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, pages 1833–1855, 1989.
- [12] H. Drees. Refined pickands estimators of the extreme value index. *The Annals of Statistics*, pages 2059–2080, 1995.
- [13] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events : for insurance and finance*, volume 33. Springer Science & Business Media, 2013.
- [14] R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge University Press, 1928.

- [15] L. Gardes. *Estimation d'une fonction quantile extrême*. PhD thesis, Université Montpellier II-Sciences et Techniques du Languedoc, 2003.
- [16] L. Gardes and S. Girard. Estimation de quantiles extrêmes pour les lois à queue de type weibull : une synthèse bibliographique. 2013.
- [17] B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.
- [18] A. Guillou and P. Willems. Application de la théorie des valeurs extrêmes en hydrologie. *Revue de statistique appliquée*, 54(2) :5–31, 2006.
- [19] A. F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348) :158–171, 1955.
- [20] M. Kratz and S. I. Resnick. The qq-estimator and heavy tails. *Stochastic Models*, 12(4) :699–724, 1996.
- [21] M. Lang, T. Ouarda, and B. Bobée. Towards operational guidelines for over-threshold modeling. *Journal of hydrology*, 225(3-4) :103–117, 1999.
- [22] A. Lekina. *Estimation non-paramétrique des quantiles extrêmes conditionnels*. PhD thesis, 2010.
- [23] B. Oussama and D. Meghlaoui. Théorie des valeurs extrêmes : Application au calcul de risques. 2016.
- [24] J. Peng. Value at risk and tail value at risk in uncertain environment. In *Proceedings of the 8th International Conference on Information and Management Sciences*, pages 787–793, 2009.
- [25] J. Pickands III et al. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1) :119–131, 1975.
- [26] S. Resnick. Extreme values, regular variation, and point processes, appl. *Probab. Ser. Appl. Probab. Trust*, 4 :33, 1987.
- [27] S. I. Resnick. *Extreme values, regular variation and point processes*. Springer, 2013.
- [28] G. Saporta. *Probabilités, analyse des données et statistique*. Editions Technip, 2006.
- [29] J. Schultze and J. Steinebach. On least squares estimates of an exponential tail coefficient. *Statistics & Risk Modeling*, 14(4) :353–372, 1996.
- [30] R. M. Shane and W. R. Lynn. Mathematical model for flood risk evaluation. *Journal of the Hydraulics Division*, 90(6) :1–20, 1964.
- [31] P. Todorovic. On some problems involving random number of random variables. *The Annals of Mathematical Statistics*, 41(3) :1059–1063, 1970.
- [32] P. Willems, G. Vaes, D. Popa, L. Timbe, and J. Berlamont. Quasi 2d river flood modelling.