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

UNIVERSITE MOHAMED KHIDER BISKRA

Faculté des Sciences Exactes et Sciences de la Nature et de la Vie
Département de Mathématiques

MEMOIRE

Présenté en vue de l'obtention du diplôme de

Magister en Mathématiques

Par

Samah BATEKA

Thème

DETERMINATION DU NOMBRE DE STATISTIQUES D'ORDRE EXTREMES

Option

Probabilités & Statistique

Soutenu publiquement le : 28/04/2010

Devant le jury :

Président: Brahim MEZERDI	**Pr.**	**U.M.K BISKRA**
Rapporteur: Abdelhakim NECIR	**Pr.**	**U.M.K BISKRA**
Examinateur: Seïd BAHLALI	**M.C. (A)**	**U.M.K BISKRA**
Examinateur: Khaled MELKEMI	**M.C. (A)**	**U.M.K BISKRA**
Examinateur: Djamel MERAGHNI	**M.C. (A)**	**U.M.K BISKRA**

بِسْمِ اللهِ الرحْمنِ الرحِيمِ

« رَب أَوْزِعْنِي أَنْ أَشْكُرَ
نِعْمَتَكَ الَتِي أَنْعَمْتَ عَلَي
وَعَلَى وَالِدَي وَأَنْ أَعْمَلَ صَالِحاً
تَرْضَاهُ وَأَدْخِلْنِي بِرَحْمَتِكَ فِي
عِبَادِكَ الصالِحِينَ »

سورة النمل «19»

Dédicace

A mes parents,
A mon cher oncle Chaouch Boukhalfa,
A l'esprit de Monsieur Seid Bahlali,
A mon cher neveu aniss.

Remerciments

Je tiens à remercier tout d'abord DIEU le tout puissant qui m'a donné durant toutes ces années la santé, le courage et la fois en moi même pour arriver à ce jour.

Il m'est particulièrement agréable aujourd'hui de remercier toutes les personnes qui m'ont aidé de près ou de loin à mener à bien ce travail.

La première personne que je tiens à remercier très chaleureusement est mon encadreur Monsieur ***Necir Abdelhakim****, professeur à l'université de Biskra, pour la confiance et les conseils m'ont été si précieux dans ma carrière scientifique. J'espère que ce mémoire sera un remerciement suffisant au soutien et à la confiance sans cesse renouvelée dont il a fait preuve en mon égard. Merci pour votre soutien et votre attention au moral de votre étudiante !*

Je tiens à remercier les membres du jury qui m'ont fait l'honneur de m'accompagner dans la phase finale de ce mémoire...

Je remercie Monsieur ***Mezerdi Brahim****, professeur à l'université de Biskra, pour sa gentillesse et sa simplicité. Je suis très honorée qu'il me fait en acceptant de présider le jury de mon mémoire.*

Je suis très honorée que Monsieur ***Melkemi Khaled****, Maître de Conférence à l'université de Biskra,* ***Bahlali Seïd*** *Maître de Conférence à l'université de Biskra et Monsieur* ***Meraghni Djamel****, Maître de Conférence à l'université de Biskra, aient accepté la si difficile tâche d'Examinateur. Merci sincèrement du temps et de l'énergie que vous avez consacré à la lecture de mon travail.*

J'adresse mes plus vifs remerciements à Monsieur ***Meraghni Djamel*** *pour sa disponibilité sa rigueur scientifique et pour ses conseils utiles et commentaires judicieux qui ont entraîné une importante l'amélioration du mémoire. Je le remercie aussi d'avoir accepter de faire partie du jury.*

Merci à tous les membres du département de Mathématiques.

Enfin, j'en profite aussi pour exprimer ma gratitude à ***mes parents*** *et mon oncle* ***Boukhalfa Chaouch****, qui est devenu au fil des années comme un père et ma soeur* ***Soraya*** *qui m'ont fourni un soutien morale et matériel durant ces années d'études et je n'oublie pas mes Collègues et mes chères amies* ***Dallal Makhloufi****,* ***Abir Achour****.*

Table des matières

Table des figures

Liste des tableaux

Introduction

L'apparition de valeurs extrêmes (aussi bien supérieur qu'inférieur) dans une série d'observations relatives à un certain phénomène témoigne de l'occurrence d'événements rares, qui malgré leur faible probabilité ont des répercussions (souvent négatives) sur les décideurs (individus ou institutions). D'où l'importance de la construction de models statistique décrivant le mieux possible ces observations. A cet effet, la théorie de valeurs extrêmes (TVE) représente un outil approprié permettant d'extrapoler le comportement des queues de distributions à partir des plus grand (ou plus petits) valeurs observées.

Les modèles des valeurs extrêmes sont appliquées à une grande variété de problèmes tels l'environnement (vitesse du vent, extrêmes pluviométriques et de températures,...), la pluviométrie, la finance et l'assurance (Mesure du risque, Valeur à risque VaR, modèle de volatilité stochastique) et en hydrologie pour calculer la probabilité que la hauteur d'eau d'un fleuve dépasse un certain seuil, télécommunications, physique,...

La modélisation des distributions à queues lourdes et fortement liée à un nombre réel appelé indice de queues ou indice des valeurs extrêmes (IVE) et dont la valeur gouverne le degré d'épaisseur des queues. L'estimation de cet indice, primordiale dans le processus de modélisation, dépend largement du nombre de statistiques d'ordre extrêmes observées. Ce nombre détermine les valeurs, qui parmi les données, qui sont réellement extrêmes. En d'autres termes, il permet de définir le seuil où les observations commencent à devenir extrêmes.

La sélection du nombre optimal de statistiques d'ordre extrêmes cruciale pour l'estimation de l'IVE et permet d'améliorer la performance des estimateurs est alors notre but dans ce mémoire. Le présent mémoire est alors une synthèse des travaux de recherches concernant la théorie des valeurs extrêmes, il s'articule autour du plan du travail suivant :

Chapitre1 : Dans ce chapitre, nous rappelons quelques éléments théoriques essentiels de la théorie des valeurs extrêmes (TVE). Il contient des rappels sur la statistique d'ordre, qui est très utile en théorie des valeurs extrêmes et on fait une introduction sur l'étude du comportement asymptotique du maximum d'un échantillon. Cette étude faisant appel à la notion de fonctions à variations régulières, on rappelle préalablement la définition de

telles fonctions et on en donne quelques propriétés. On donne ensuite des résultats décrivant les limites possibles de la loi du maximum d'un échantillon. Deux théorèmes sont essentiels à la compréhension de la Théorie des Valeurs Extrêmes : celui de Fisher-Tippett et celui de Balkema-de Haan-Pickands.

Chapitre 2 : Dans ce chapitre, on passe en revue les différentes méthodes d'estimation de l'indice des valeurs extrêmes et des quantiles extrêmes. On distingue deux Modèles d'estimation : la première, appelée "modèle EVT", est basée sur la distribution des valeurs extrêmes généralisée (GEVD) et la deuxième, appelée "modèle POT", est basée sur la distribution de Pareto généralisée (GPD). La vaste collection des estimateurs de l'indice des valeurs extrêmes γ qui caractérise les queues de distributions, à été la question centrale dans ce chapitre. Nous avons présenté les principales approches pour l'estimation de γ, Les plus utilisés sont les estimateurs du maximum de vraisemblance et des moments pondérés. Avec une attention particulière à l'estimation semi-paramétrique (Hill, Moment, Pickands parmi autres). Le plus connu est l'estimateur de Hill (1975) où nous basons notre travail.

Chapitre 3 : Dans ce chapitre, nous nous intéressons à répondre à la question comment séléctionner le nombre de statistiques d'ordre extrêmes impliqué dans le calcul de l'estimateur ? Puisque l'estimation de l'indice de queue est nécessairement liée du nombre de statistiques d'ordre supérieurs utilisés dans cette estimation alors nous avons proposé dans ce chapitre certaines des méthodes essayez de choisir la valeur de k qui minimise l'erreur moyenne quadratique (MSE) de l'estimateur, puisque cette quantité équilibre deux effets : le biais et la variance. Le problème du choix de la valeur optimale des statistiques d'ordre supérieur a reçu beaucoup d'attention dans la littérature, (voir [8], [12], [17], [19], [28], [40] et [36]).

Une bonne référence pour la théorie et les applications des valeurs extrêmes est le livre de Embrechts, Kluppelberg et Mikosch [20].

Nous ne finirions pas cette introduction sans mentionner que le logiciel statistique R, présenté dans l'annexe A, est utilisé dans le traitement des exemples présentés dans toute ce mémoire. Les illustrations sont basées sur des modèles théoriques et des données réelles en respectivement résumés dans le tableau 1.

Données	Taille	Description	Application/Source
Danish Fire	2167	Large réclamations d'assurance incendie au Danemark (plus de 1 million couronnes danoises) du 01/01/1980 au 31/12/1990	Assurance /R package evir
BMW	6146	Rendements journaliers de BMW cours de l'action du 2/1/1973 au 23/7/1996	Assurance /R package evir

TAB. 1 – Ensembles des données réelles utilisées pour les illustrations.

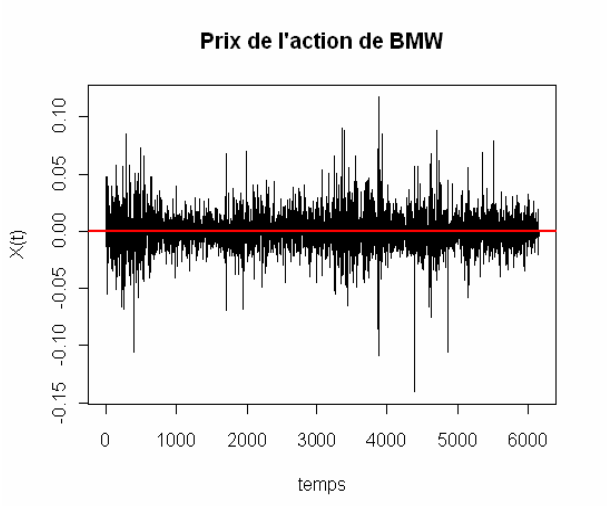


FIG. 1 – Série des rendements quotidiens des parts de BMW de la période allant du 2/1/1973 au 23/7/1996.

Notations et abréviations

$\xrightarrow{d}$	convergence en distribution
$\xrightarrow{P}$	convergence en probabilité
$\xrightarrow{p.s}$	convergence presque sûre
H_γ	distribution des valeurs extrêmes (EVD)
GEVD	distribution des valeurs extrêmes généralisée
GPD	distribution de Paréto généralisée
$D(H_\gamma)$	domaine d'attraction de H_γ
(Ω, F, P)	espace de probabilité
MSE	erreur moyenne quadratique
EMV	estimation par maximum de vraisemblance
EMP	estimation par les moments pondérés
$\stackrel{D}{=}$	égalite en distribution
F	fonction de répartition
$\bar{F}$	fonction de survie
Q	fonction inverse généralisée de F
F_n	fonction de répartition empirique
Q_n	fonction des quantiles empiriques
1_A	fonction indicatrice de l'événement A
l	fonction à variation lente
i.i.d	identiquement et identiquement distribuée
IVE	indice des valeurs extêmes
$\mathcal{N}(\mu, \sigma^2)$	loi normale d'espérence μ et de variance σ^2
k	nombre de statisqtique d'ordre extrêmes
#	nombre de
POT	Peaks-Over-Therhold
x_F	point terminal de F
s.o	statistique d'ordre
TVE	théorie des valeurs extrêmes
GEV	valeurs extrêmes généralisé

Chapitre 1

Théorie des valeurs extrêmes

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

Ce chapitre réservé a une introduction sur les statistiques d'ordre, les valeurs extrêmes et les queues de distributions. Nous énonçons les principaux résultats concernant les distributions limites des plus grandes observations d'un échantillon ainsi que les domaines d'attractions.

En règle générale, les résultats de minima peuvent être déduits des résultats correspondant à maxima par la relation triviale

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

Pour plus de détails sur la théorie des valeurs extrêmes, on peut se référer aux ouvrages de Embrechts, Klüpperg, et Milkosch [20], Reiss et Thomas [36], de Hann et Ferreira [27] et beaucoup de révues.

1.1 Définitions et caractéristiques de bases

Nous présentons tout d'abord quelques rappels et définitions essentiels dans notre étude du comportement asymptotique du maximum d'un échantillon.

Définition 1.1 (Fonction de distribution emprique) *Soit $X_1, ..., X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) définies sur le même espace de probabilité (Ω, F, P)[1] d'une fonction de répartition commune F telle que,*

$$F(x) := P\{w \in \Omega / \ X(w) \leq x\} = P\{X \leq x\}, \qquad x \in \mathbb{R}.$$

La fonction de répartition empirique : notée F_n est définie par :

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}}, \quad x \in \mathbb{R}.$$

En plus notons par $\overline{F}$ la fonction de survie (ou la fonction des queues) :

$$\overline{F}(x) := P\{w \in \Omega / \ X(w) > x\} = 1 - F(x).$$

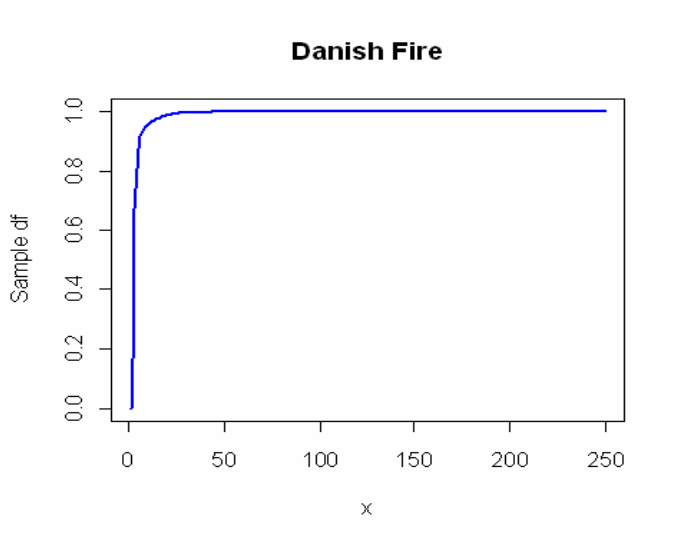


FIG. 1.1 – Fonction de distribution empirique des données de réclamations des pertes dues au feu au Danemark de la période allant du 1/1/1980 au 31/12/1190 (2167 observations).

Théorème 1.1 (Glivenko-Cantelli, 1933)

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{p.s} 0 \qquad quand \qquad n \to \infty.$$

[1] On appelle espace de probabilité un triplet $(\Omega, \mathcal{F}, P)$ où $(\Omega, \mathcal{F})$ est un espace mesurable et P une probabilité sur $\mathcal{F}$.

Définition 1.2 (Fonction des quantiles empiriques) *La fonction des quantiles ou l'inverse généralisé de la fonction de distribution F notée par Q : pour chaque entier $n \geq 1$*

$$Q(t) := F^{\leftarrow}(t) := \inf\{s : \ F(s) \geq t\}, \quad 0 < t < 1.$$

La fonction des quantiles empiriques : notée Q_n est définie par :

$$Q_n(t) := F_n^{\leftarrow}(t) = \inf\{s : \ F_n(s) \geq t\}, \quad 0 < t < 1.$$

Où $F^{\leftarrow}$est l'inverse généralisée de la fonction de distribution F.

Définition 1.3 (Fonction des quantiles de queues) *La fonction des quantiles de queues (tail quantile function), notée U est définie par :*

$$U(t) := Q(1 - 1/t) = (1/\overline{F})^{\leftarrow}(t), \quad 1 < t < \infty.$$

Et la fonction des quantiles de queues empirique notée U_n est donnée par :

$$U_n(t) := Q_n(1 - 1/t), \quad 1 < t < \infty.$$

Définition 1.4 (Somme et moyenne arithmétique) *Soit $X_1, X_2, ...$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) de fonction de répartition commune F. Pour un entier $n \geq 1$, on définit la somme partielle et la moyenne arithmétique correspondante respectivement par*

$$S_n := \sum_{i=1}^{n} X_i \quad et \quad \overline{X}_n := S_n/n.$$

$\overline{X}_n$ est appelé moyenne de l'échantillon ou moyenne empirique.

1.1.1 Lois des grands nombres

Ces lois décrivent le comportement asymptotique de la moyenne de l'échantillon. Elles sont de deux types : loi faible[2] mettant en jeu la convergence en probabilité et loi forte[3] relative à la convergence presque sûre.

[2] *La loi faible des grands nombres est également appelée théorème de Khintchine est sur la convergence en probabilité de $\overline{X}_n$.*

[3] *La loi forte des grands nombres (Kolmogorov, 1929) basée sur la convergence presque sûrement de $\overline{X}_n$.*

Théorème 1.2 (Lois des grands nombres) *Si* $(X_1, ..., X_n)$ *est une suite d'une v.a* X *tel que* $E|X| < \infty$, *Alors*

$$\begin{aligned} &\textit{La loi faible} \quad && \overline{X}_n \xrightarrow{P} \mu \quad && \textit{quand } n \to \infty, \\ &\textit{La loi forte} \quad && \overline{X}_n \xrightarrow{p.s} \mu \quad && \textit{quand } n \to \infty, \end{aligned}$$

où $\mu := E(X)$.

1.1.2 Théorème central limite

L'étude de somme de variables indépendantes et de même loi joue un rôle capitale en statistique. Le théorème suivant connu sous le nom de théorème central limite (TCL) établit la convergence vers la loi de Guass.

Théorème 1.3 (TCL) *Si* $(X_1, ..., X_n)$ *est une suite de variable aléatoire définie sur le même espace de probabilité de variance* σ^2 *finie et de moyenne* μ *alors*

$$\frac{1}{\sqrt{n}}\left(\frac{S_n - n\mu}{\sigma}\right) \xrightarrow{d} \mathcal{N}(0,1) \quad \textit{quand} \quad n \to \infty.$$

Dans les applications pratiques, ce théorème permet en particulier de remplacer une somme de variables aléatoires en nombre assez grand mais fini par une approximation normale.

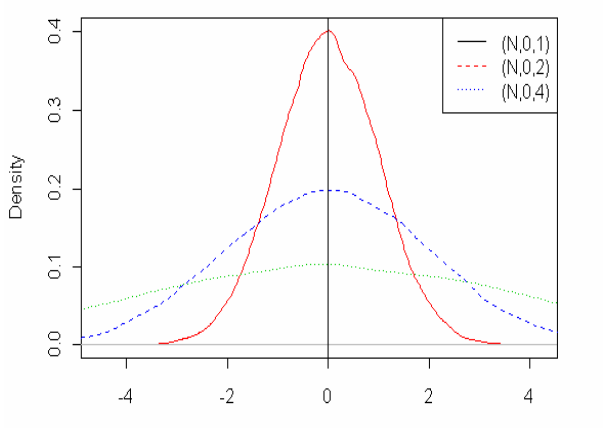


FIG. 1.2 – Densités des distributions normales de moyennes nulles et de varinaces 1, 2 et 4 respectivement.

1.2 Statistique d'ordre

Pour commencer notre étude et les explications de la théorie des valeurs extrêmes, il faut avoir un grand bagage, alors notre point de départ sera les statistiques d'ordre. Celles-ci sont un outil essentiel de modélisation des risques.

Soit $(X_1, ..., X_n)$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) de fonction de répartition F ($F(x) = P(X_n \leq x)$, pour $x \in \mathbb{R}$). Soit S_n l'ensemble des permutation de $\{1, ..., n\}$.

1.2.1 Définition de la statistique d'ordre

La statistique d'ordre de l'échantillon $(X_1, ..., X_n)$ est le réarrangement croissant de $(X_1, ..., X_n)$. On le note par $(X_{1,n}, ..., X_{n,n})$.

On a $X_{1,n} \leq ... \leq X_{n,n}$, et il existe une permutation aléatoire $\sigma_n \in S_n$ telle que

$$(X_{1,n}, ..., X_{n,n}) = \left(X_{\sigma_n(1)}, ..., X_{\sigma_n(n)}\right).$$

Le vecteur $(X_{1,n}, ..., X_{n,n})$ est appelé l'échantillon ordonné associé à l'échantillon $(X_1, ..., X_n)$, et $X_{k,n}$ étant la $k^{ième}$ statistique d'ordre.

Dans un échantillon de taille n deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des événements extrêmes ce sont :

$$X_{1,n} = \min(X_1, ..., X_n) \quad \text{et} \quad X_{n,n} = \max(X_1, ..., X_n).$$

En utilisant la propriété d'indépendance des variables aléatoires $X_1, X_2, ..., X_n$ nous en déduisons que les lois de maximum $X_{n,n}$ et de minimum $X_{1,n}$ de la statistique d'ordre associées à l'échantillon $X_1, X_2, ..., X_n$ sont

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

Alors

$$\boxed{F_{X_{n,n}}(x) = [F_X(x)]^n .}$$

Puis

$$\begin{aligned} F_{X_{1,n}}(x) \quad = P(X_{1,n} \leq x) \quad &= \quad 1 - P(X_{1,n} > x) \\ &= \quad P(\min(X_1, ..., X_n) > x) \\ &= \quad 1 - P\left[\bigcap_{i=1}^{n} (X_i > x)\right] \\ &= \quad 1 - \prod_{i=1}^{n} P(X_i > x) \\ &= \quad 1 - \prod_{i=1}^{n} [1 - P(X_i \leq x)] \\ &= \quad 1 - \prod_{i=1}^{n} [1 - F_X(x)] \end{aligned}$$

D'où on déduit :

$$\boxed{F_{X_{1,n}}(x) = 1 - [1 - F_X(x)]^n .}$$

De ces résultats, nous en tirons la conclusion que le maximum $F_{X_{n,n}}$ est une variable aléatoire dont la fonction de répartition correspond à F^n.

La fonction de répartition de X n'étant pas souvent connue, il n'est généralement pas possible de déterminer la distribution du maximum à partir de ce résultat. On s'intéresse alors à la distribution asymptotique du maximum en faisant tendre n vers l'infini. On a :

$$\lim_{n\to\infty} F_{X_{n,n}} = \lim_{n\to\infty} [F(x)]^n = \begin{cases} 1 & \text{si} \quad F(x) = 1, \\ 0 & \text{si} \quad F(x) < 1. \end{cases}$$

On constate que la distribution asymptotique du maximum, déterminée en faisant n tendre vers l'infini, donne une loi dégénérée (ils prennent des valeurs de 0 et 1 seulement).

Définition 1.5 (Point extrême) *On note par x_F (resp. x_F^*) le point extrême supérieur (resp. inférieur) de la distribution F (i.e. la plus grande valeur possible pour $X_{k,n}$ peut prendre la valeur $+\infty$ (resp. $-\infty$)) au sens où*

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

respectivement

$$x_F^* := \inf\{x : \ F(x) > 0\}.$$

Proposition 1.1 (Limite de $X_{n,n}$) $X_{n,n} \overset{p.s}{\to} x_F \quad$ *quand* $\quad n \to \infty$.

D'autre part, le comportement asymptotique de la loi de $X_{n,n}$ est donné, dans certaines conditions sur la queue de distribution.

Soit $X_1, ..., X_n$ un échantillon de n variables aléatoires indépendantes et de même fonction de répartition F. Dans la suite on note $X_{1,n}, ..., X_{n,n}$ échantillon ordonné associé à cette échantillon.

Définition 1.6 (Distribution empirique et quantiles empirique) *Il existe une autre version de la définition de F_n en utilisant les statistiques d'ordre (s.o) comme suit :*

$$F_n(x) := \begin{cases} 0 & \text{si} \quad x < X_{1,n} \\ \dfrac{i-1}{n} & \text{si} \quad X_{i-1,n} \le x < X_{i,n}, \quad 2 \le i \le n \\ 1 & \text{si} \quad x \ge X_{n,n}. \end{cases}$$

De même on obtient une autre version de Q_n en utilisant les statistiques d'ordre (s.o) comme suit :

$$Q_n(t) := \begin{cases} X_{i,n} & \text{si} \quad \dfrac{(i-1)}{n} < t \le \dfrac{i}{n} \\ X_{n,n} & \text{si} \quad 0 < t \le 1. \end{cases}$$

Proposition 1.2 *Soit $X_1, ..., X_n$ des variables aléatoires indépendantes et de fonction de répartition F. Soit $U_1, ..., U_n$ des variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Alors $(F^{-1}(U_{1,n}), ..., F^{-1}(U_{n,n}))$ a même loi que $(X_{1,n}, ..., X_{n,n})$.*

Lemme 1.1 *Si F est continue, alors presque sûrement on a $X_{1,n} < ... < X_{n,n}$.*

Preuve. Il suffit de vérifier que $P(\exists i \neq j \quad \text{tel que} \quad X_i = X_j) = 0$.
On a

$$\begin{aligned} P(\exists i \neq j \text{ tel que } X_i = X_j) &\le P(\exists i \neq j \quad \text{tel que} \quad F(X_i) = F(X_j)) \\ &\le \sum_{i \neq j} P(F(X_i) = F(X_j)). \end{aligned}$$

Les variables $F(X_i)$ et $F(X_j)$ sont pour $i \neq j$ des variables aléatoires indépendantes de loi uniforme sur $[0,1]$. On déduit que

$$P(F(X_i) = F(X_j)) = \int\limits_{[0,1]^2} 1_{\{u=v\}} dudv = 0.$$

Donc p.s. pour tous $i \neq j$, on a $X_i \neq X_j$. ■

Lemme 1.2 *On suppose F continue. La loi de σ_n est la loi uniforme sur S_n. De plus la permutation σ_n est indépendante de la statistique d'ordre.*

Preuve. Soit $\sigma_n \in S_n$. On a $P(\sigma_n = \sigma) = P(X_{\sigma_n(1)} < ... < X_{\sigma_n(n)})$.
Les variables $(X_{\sigma_n(1)}, ..., X_{\sigma_n(n)})$ sont indépendantes et de même loi.
En particulier le vecteur $(X_{\sigma_n(1)}, ..., X_{\sigma_n(n)})$ a même loi que $(X_1, ..., X_n)$.
Il vient

$$P(\sigma_n = \sigma) = P(X_1 < ... < X_n).$$

Le membre de droite est indépendant de σ. La loi de σ_n est donc la loi uniforme sur S_n, et on a $P(\sigma_n = \sigma) = 1/n!$. Soit g une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$, mesurable bornée.
Comme le vecteur $(X_{\sigma_n(1)}, ..., X_{\sigma_n(n)})$ a même loi que $(X_1, ..., X_n)$, on a

$$\begin{aligned} E\left[1_{\{\sigma_n=\sigma\}} g(X_{1,n}, ..., X_{n,n})\right] &= E\left[1_{\left\{X_{\sigma_n(1)}<...<X_{\sigma_n(n)}\right\}} g(X_{\sigma_n(1)}, ..., X_{\sigma_n(n)})\right] \\ &= E\left[1_{\{X_1<...<X_n\}} g(X_1, ..., X_n)\right]. \end{aligned}$$

On en déduit, en sommant sur $\sigma_n \in S_n$, que

$$E\left[g(X_{1,n}, ..., X_{n,n})\right] = n! E\left[1_{\{X_1<...<X_n\}} g(X_1, ..., X_n)\right].$$

Enfin, on remarque que

$$E\left[1_{\{\sigma_n=\sigma\}} g(X_{1,n}, ..., X_{n,n})\right] = P(\sigma_n = \sigma) E\left[g(X_{1,n}, ..., X_{n,n})\right].$$

Cela implique que la permutation σ_n est indépendante de la statistique d'ordre.
■

Proposition 1.3 *Le $m^{ième}$ $(m = 1, 2, ...)$ moment de la $i^{ième}$ $(i = 1, ..., n)$ statistique d'ordre est*

$$\begin{aligned}\mu_{i,n}^{(m)} &= \frac{n!}{(i-1)!(n-i)!}\int_{-\infty}^{\infty} x^m \left[F(x)\right]^{i-1}\left[1-F(x)\right]^{n-i} f(x)dx \\ &= \frac{n!}{(i-1)!(n-i)!}\int_0^1 \left[F^{\leftarrow}(u)\right]^m u^{i-1}(1-u)^{n-i}du.\end{aligned}$$

Proposition 1.4 (Propriété de Markov) *Quand F est continue, l'échantillon ordonné $(X_{1,n},...,X_{n,n})$ forme une chaîne de Markov*[4]*. En d'autres termes, nous avons pour $i = 2,...,n$*

$$P(X_{i,n} \leq x / X_{1,n} = x_1, ..., X_{i-1,n} = x_{i-1}) = P(X_{i,n} \leq x / X_{i-1,n} = x_{i-1}).$$

Les preuves de ces résultats sont simples et pourraient être trouvées dans [1].

1.2.2 Loi de la statistique d'ordre

Corollaire 1.1 *Si la loi de X_1 possède une densité f, alors la statistique d'ordre $(X_{1,n},...,X_{n,n})$ possède la densité*

$$f_{X_{1,n},...,X_{n,n}}(x_1,x_2,...,x_n) = n!\prod_{i=1}^{n} f(x_i),\ -\infty < x_1 < ... < x_n < \infty.$$

Preuve. On rappelle que, puisque $f(x)$ est la dérivée de $F(x)$

$$f(x) = \lim_{\delta x \to 0} \frac{F(x+\delta x) - F(x)}{\delta x}.$$

Soit $f(x) = \lim_{\delta x \to 0}\left(\dfrac{P(x \leq X < x+\delta x)}{\delta x}\right)$

$P(x_1 \leq X_{1,1} < x_1 + \delta x_1, ..., x_n \leq X_{n,n} < x_n + \delta x_n)$

$= n!P(x_1 \leq X_{1,1} < x_1 + \delta x_1)...P(x_n \leq X_{n,n} < x_n + \delta x_n),$

d'où

$$\frac{P(x_1 \leq X_{1,1} < x_1 + \delta x_1 ... x_n \leq X_{n,n} < x_n + \delta x_n)}{\delta x_1 ... \delta x_n}$$

$$= n!\frac{P(x_1 \leq X_{1,1} < x_1 + \delta x_1)}{\delta x_1} ... \frac{P(x_n \leq X_{n,n} < x_n + \delta x_n)}{\delta x_n}.$$

[4] *On appelle chaine de Markov une suite de variables aléatoires (X_n) telle que, pour chaque n, connaissant la valeur de X_n, X_{n+1} soit indépendante de X_k, pour k inférieur ou égal à $n-1$.*

Alors

$$f_{X_{1,1},...,X_{n,n}}(x_1, x_2, ..., x_n) = n! \prod_{i=1}^{n} f(x_i), \quad -\infty < x_1 < ... < x_n < \infty.$$

■

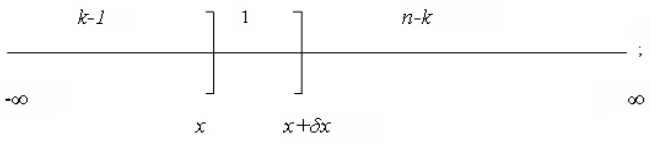
Notation 1.1 *On note* $f_{X_{k,n}}(x)$ *par* $f_k(x)$ *et* $F_{X_{k,n}}(x)$ *par* $F_k(x)$.

1.2.3 Loi de $X_{k,n}$

Soit $X_{k,n}(1 \leq k \leq n)$ la $k^{ième}$ statistique d'ordre, l'événement

$$\{x < X_{k,n} \leq x + \delta x\},$$

peut être représenté comme suit :



$X_i \leq x$ pour $k-1$ des X_i, $x < X_i \leq x + \delta x$ pour exactement l'un des X_i et $X_i > x + \delta x$ pour $n-k$ restes des X_i. Il y a C_n^{k-1} manière de réaliser le premier événement et C_{n-k+1}^1 manière de réaliser le deuxième événement, et il reste une seule manière de réaliser le dernier événement.

Considérons δx assez petit, on peut écrire

$$\begin{aligned} P(x \leq X_{k,n} < x + \delta x) &= \frac{n!}{(k-1)!1!(n-k)!} [P(X_i < x)]^{k-1} \\ &\quad P(x \leq X_{i,n} < x + \delta x) [P(X_i \geq x)]^{n-k}. \end{aligned}$$

Alors, la densité marginale $f_k(x)$ de la coordonnée $X_{k,n}$ $(k = 1, ..., n)$ de l'échantillon ordonné s'obtient comme suit **:**

$$\begin{aligned} f_k(x) &= \lim_{\delta x \to \infty} \left\{ \frac{P(x \leq X_{k,n} < x + \delta x)}{\delta x} \right\} \\ &= \frac{n!}{(k-1)!(n-k)!} \left(F(x)^{k-1}\right) (1 - F(x))^{n-k} f(x),\ x \in \mathbb{R}. \end{aligned}$$

Le calcul de la fonction de répartition $F_k(x)$ de $X_{k,n}$ est immédiat :

$$\begin{aligned} F_k(x) &= P(X_{k,n} \leq x) \\ &= P(\text{au moins } k \text{ des } X_i \text{ sont inférieurs à } x) \\ &= \sum_{i=k}^{n} P(\text{exactement } i \text{ de } X_1, ..., X_n \text{ sont inférieurs à } x) \end{aligned}$$

$$\boxed{F_k(x) = \sum_{i=k}^{n} C_n^i \left[F(x)\right]^i \left[1 - F(x)\right]^{n-i},\ x \in \mathbb{R}.}$$

Cas particulier des valeurs extrêmes $X_{1,n}$ et $X_{n,n}$:

A partir de ce qui précède :

$$\begin{aligned} F_1(x) &= \sum_{i=1}^{n} C_n^i \left[F(x)\right]^i \left[1 - F(x)\right]^{n-i} \\ &= 1 - (1 - F(x))^n. \end{aligned}$$

Et

$$f_1(x) = nf(x)(1 - F(x))^{n-1}.$$

De même pour $X_{n,n}$:

$$\begin{aligned} F_n(x) &= F^n(x) \\ f_n(x) &= nf(x)F^{n-1}(x). \end{aligned}$$

Lemme 1.3 *La variable aléatoire $Y_k = F(X_{k,n})$ suit une loi bêta*[5]

[5] Une variable aléatoire X à valeurs dans [0,1] est dite suivre la loi bêta de paramètres $r > 0,\ s > 0$ (ce que l'on note $B(r,s)$) si elle est absolument continue et admet pour densité :

$$f(x) = \begin{cases} \dfrac{1}{\mathcal{B}(r,s)} x^{r-1}(1-x)^{s-1} & \text{si } x \in [0,1] \\ 0 & \text{sinon.} \end{cases}$$

La fonction bêta pour $r, s > 0$ est

$$\mathcal{B}(r,s) = \int_0^1 t^{r-1}(1-t)^{s-1} dt$$

Preuve. On a

$$\begin{aligned} P(F(X_{k,n}) \leq u) &= P(X_{k,n} \leq F^{-1}(u)) \\ &= \frac{n!}{(k-1)!(n-k)!} \int_0^{F(F^{-1}(u))} t^{k-1}(1-t)^{n-k}dt \\ &= \frac{1}{\mathcal{B}(k, n-k+1)} \int_0^u t^{k-1}(1-t)^{n-k+1-1}dt \end{aligned}$$

Si on pose $k-1=r$ et $n-k+1=s$ on remarque que cette distribution est la loi de bêta de paramètres k et $n-k+1$.Alors

$$F(X_{k,n}) \sim \mathcal{B}(k, n-k+1).$$

■

Remarque 1.1 *Il est facile de démonter que*

$$\frac{n!}{(k-1)!(n-k)!} = \frac{1}{\mathcal{B}(k, n-k+1)}.$$

Par l'utilisation de la relation suivante $\mathcal{B}(r,s) = \dfrac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$, *où* Γ *est la fonction Gamma*[6]

1.2.4 Loi jointe d'une couple $(X_{i,n},\ X_{j,n})$

Soit un couple $(X_{i,n}, X_{j,n})$ avec $1 \leq i < j \leq n$, l'événement

$$\left\{x < X_{i,n} \leq x + \delta x,\ y < X_{j,n} \leq y + \delta y\right\},$$

peut être représenté comme suit :

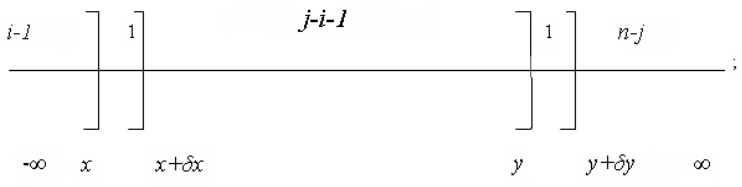
[6]On définit la fonction Gamma comme suit :

$$\Gamma(r) = \int_0^1 t^{r-1}e^{t-1}dt\ , \quad r > 0.$$

On note que pour $n \geq 1$, $\Gamma(n) = n!$.

En générale $\Gamma(r) = (r-1)\Gamma(r-1)$

Pour $r = 1/2$, $\Gamma(1/2) = \sqrt{\pi}$.



$X_r \leq x$ pour $i-1$ des X_r, $x < X_r \leq x+\delta x$ pour l'un des X_r, $x+\delta x < X_r \leq y$ pour $j-i-1$ des X_r. $y < X_r \leq y+\delta y$ pour l'un des X_r et $X_r > y+\delta y$ pour $n-j$ restes des X_r.

Considérons δx et δy assez petits,

$$\begin{aligned} P &= (x < X_{i,n} \leq x+\delta x,\ y < X_{j,n} \leq y+\delta y) \\ &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(x))^{i-1} (F(x+\delta x) - F(x)) \\ &\quad (F(y) - F(x+\delta x))^{j-i-1} (F(y+\delta y) - F(y)) (1 - F(y+\delta y))^{n-j}. \end{aligned}$$

Alors on peut calculer la densité jointe $f_{i,j}(x,y)$ de deux statistiques d'ordre $X_{i,n}$ et $X_{j,n}$ $(1 \leq i < j \leq n)$ comme suit:

$$\begin{aligned} f_{i,j}(x,y) &= \lim_{\substack{\delta x \to 0 \\ \delta y \to 0}} \left\{ \frac{P(x < X_{i,n} \leq x+\delta x,\ x < X_{j,n} \leq x+\delta x)}{\delta x \delta y} \right\} \\ &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(x))^{i-1} (F(y) - F(x))^{j-i-1} \\ &\quad (1 - F(y))^{n-j} f(x) f(y), -\infty < x < y < +\infty. \end{aligned}$$

Pour $i = 1$ et $j = n$, on obtient la densité jointe de $(X_{1,n},\ X_{n,n})$ comme suit:

$$f_{1,n}(x,y) = n(n-1) (F(y) - F(x))^{n-2} f(x) f(y), \quad -\infty < x < y < \infty.$$

Conséquence :

A partir de la densité jointe d'un couple $(X_{i,n},\ X_{j,n})$, on trouve la densité jointe de n statistique d'ordre

$$f_{1,\ldots,n}(x_1, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i), \quad -\infty < x_1 < \ldots < x_n < +\infty.$$

La fonction de distribution jointe de $(X_{i,n}, X_{j,n})$ est

$$F_{i,j}(x,y) = P\left(\{X_{i,n} \leq x\} \cap \{X_{j,n} \leq y\}\right), \qquad x,\ y \in \mathbb{R}.$$

On a deux cas :

1^{ière}cas : $x \geq y$

$$\begin{aligned} F_{i,j}(x,y) &= P(X_{i,n} \leq x,\ X_{j,n} \leq y) \\ &= P(X_{j,n} \leq y) \\ F_{i,j}(x,y) &= F_{i,j}(y) \end{aligned}$$

2^{ième}cas : $x < y$

$$\begin{aligned} F_{i,j}(x,y) &= P(X_{i,n} \leq x,\ X_{j,n} \leq y) \\ &= P(\text{au moins } i \text{ de } X_1, ..., X_n \text{ sont inférieurs à } x \\ &\qquad \text{et au moins } j \text{ de } X_1, ..., X_n \text{ sont inférieurs à } y\ ) \\ &= \sum_{s=j}^{n} \sum_{r=i}^{s} P(\text{exactement } r \text{ de } X_1, ..., X_n \text{ sont inférieurs à } x \\ &\qquad \text{et exactement } s \text{ de } X_1, ..., X_n \text{ sont inférieurs à } y) \\ F_{i,j}(x,y) &= \sum_{s=j}^{n} \sum_{r=i}^{s} \frac{n!}{r!(s-r)!(n-s)!} \left(F(x)\right)^r \left(F(y) - F(x)\right)^{s-r} \left(1 - F(y)\right)^{n-s}. \end{aligned}$$

1.3 Loi des valeurs extrêmes

La distribution de $X_{n,n}$ devrait nous fournir des informations sur des événements extrêmes et comme la limite de cette distribution obtenue précédemment conduit à une loi dégénérée lorsque n tend vers l'infini, on recherche une loi non dégénérée pour le maximum de X. De façon analogue au théorème central limite, la théorie des valeurs extrêmes montre qu'il existe des suites $\{a_n\}$ et $\{b_n\}$ $n \in \mathbb{N}$, avec $a_n > 0$ et $b_n \in \mathbb{R}$, telles que

$$\lim_{n\to\infty} P\left(\frac{X_{n,n} - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = H(x),\ x \in \mathbb{R}, \tag{1.1}$$

avec H est une distribution non dégénérée. Deux questions se posent ici : quelle est la loi H? Et quelles conditions doit vérifier F pour qu'il existe des suites $\{a_n\}$ et $\{b_n\}$ satisfaisant (1.1) ?

Fisher et Tippett (1928) [21], Gendenko (1943) et de Hann (1970) ont montré que les seules distributions limites non dégénérée H possibles sont les distributions de valeurs extrêmes. Le théorème suivant donne une condition nécessaire et suffisante pour l'existence d'une loi limite non dégénérée pour le maximum.

Théorème 1.4 (Fisher et Tippett(1928), Gnedenko (1943)) *Soit* $(X_n)_n$ *une suite de variables aléatoire (i.i.d), s'il existe un réel* γ *et deux suites réelles* (a_n) *et* (b_n) , $n \in \mathbb{N}$, *avec* $a_n > 0$ *et* $b_n \in \mathbb{R}$ *telles que :*

$$\lim_{n\to\infty} P\left[\frac{X_{n,n} - b_n}{a_n} \leq x\right] = H_\gamma(x), \tag{1.2}$$

pour tout x, *où* H *est une fonction de distribution non dégénérée. Alors* H *est du même type que l'une des fonctions suivantes :*

Gumbel (type I) :

$$H_0(x) = \Lambda(x) = \exp[-\exp(-x)], \quad x \in \mathbb{R}.$$

Fréchet (type II) :

$$H_\gamma(x) = \Phi_\gamma(x) = \begin{cases} 0, & x \leq 0, \\ \exp\left(-(x)^{-1/\gamma}\right), & x > 0, \end{cases} \quad \text{avec } \gamma > 0.$$

Weibull (type III) :

$$H_\gamma(x) = \Psi_\gamma(x) = \begin{cases} \exp\left(-(-x)^{-\gamma}\right), & x \leq 0, \\ 1, & x > 0, \end{cases} \quad \text{avec } \gamma < 0.$$

• Ce théorème présente un intérêt important, car si l'ensemble des distributions est grand, l'ensemble des distributions des valeurs extrêmes est très petit. Ce théorème n'est valable que si les suites $\{a_n\}$ et $\{b_n\}$ existent et admettent des limites.

• Ce théorème est un résultat important car il n'est pas nécessaire de faire d'hypothèses paramétriques sur la loi des X_i. La valeur de γ détermine le comportement de la queue de distribution.

Définition 1.7 (Distributions standard des valeurs extrêmes) *Les trois fonctions de distribution du théorème 1.4 s'appellent les distributions standard ou traditionnelle des valeurs extrêmes.*

1.3.1 Remarques

1. La fonction de répartition H_γ est appelée loi des valeurs extrêmes. (que l'on note (EVD) "*Extreme Value Distribution*"). Le paramètre γ est un paramètre de forme (*shape parameter*) encore appelé indice des valeurs extrêmes ou indice de queue, a_n est un paramètre de position et b_n est un paramètre d'échelle.
2. Les suites de normalisation $\{a_n\}$ et $\{b_n\}$ ne sont pas uniques.
3. Plus l'indice γ est élevé en valeur absolue, plus le poids des extrêmes dans la distribution initiale est important. On parle alors de distributions à "queues épaisses".

Proposition 1.5 (Relation entre Λ, Φ_γ et Ψ_γ) *Soit Y une variable aléatoire positive $(Y > 0)$ alors les affirmations suivantes sont équivalentes.*

(a) $Y \sim \Phi_\gamma$.

(b) $\ln Y^\gamma \sim \Lambda$.

(c) $-Y^{-1} \sim \Psi_\gamma$.

Il existe une certaine liberté dans le choix des constantes de normalisation ($\{a_n\}$ et $\{b_n\}$) dans (1.2) parce que l'unicité de la limite de H est seulement jusqu'à affiner les transformations. Réellement, c'est une conséquence du théorème de Khintchine.

Théorème 1.5 (Khintchine) *Soit $\{G_n,\ n \in \mathbb{N}\}$ une suite de fonctions de distributions et G une fonction de distribution non dégénérée. Pour $x \in \mathbb{R}$, soit $\{a_n\}$ et $\{b_n\}$, $n \in \mathbb{N}$, avec $a_n > 0$ et $b_n \in \mathbb{R}$ tels que*

$$\lim_{n\to\infty} G_n(a_n x + b_n) = G(x).$$

Alors, pour certaine fonction de distribution non dégénérée G^* et suites $\{\alpha_n\}$ et $\{\beta_n\}$, $n \in \mathbb{N}$, avec $\alpha_n > 0$ et $\beta_n \in \mathbb{R}$,

$$\lim_{n\to\infty} G_n(\alpha_n x + \beta_n) = G^*(x).$$

ssi

$$\frac{\alpha_n}{a_n} \to A > 0 \qquad et \qquad \frac{\beta_n - b_n}{a_n} \to B \in \mathbb{R}, \quad quand\ n \to \infty.$$

De plus,

$$G^*(x) = G(Ax + B),\ x \in \mathbb{R}.$$

Jenkinson (1955) a généralisé les résultats de *Fréchet* (1927) et *Fisher & Tippett* (1928) en montrant que la loi du maximum de l'échantillon tend vers la loi généralisée des valeurs extrêmes.

1.3.2 Distribution des valeurs extrêmes généralisée

Pour faciliter le travail avec les trois distribution limites, Jenkinson-Von Mises a donné une représentation qui a obtenu en introduisant les paramètres de localisation μ et de dispersion σ dans la paramétrisation des distributions extrêmes, on obtient la forme la plus générale de la distribution des valeurs extrêmes, notée GEVD (*Generalized Extreme Value Distribution*). Elle est simplement une reparamétrisation des distributions apparaissant dans le théorème 1.4.

$$H_{\mu,\sigma,\gamma}(x) = \begin{cases} \exp\left\{-\left[1+\gamma\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\} & \text{pour} \quad 1+\dfrac{\gamma}{\sigma}(x-\mu) > 0, \\ \exp\left(-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right) & \text{pour} \quad \gamma = 0,\ x \in \mathbb{R}. \end{cases}$$

Avec $\mu \in \mathbb{R}$ et $\sigma > 0$.

En remplaçant $\left(\dfrac{x-\mu}{\sigma}\right)$ par x on obtient la forme standard de la GEVD :

$$H_\gamma(x) = \begin{cases} \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\} & \text{pour} \quad \gamma \neq 0,\ 1+\gamma x > 0, \\ \exp\left\{-\exp\left(-x\right)\right\} & \text{pour} \quad \gamma = 0,\ x \in \mathbb{R}. \end{cases}$$

Où γ est le paramètres de forme.

Nous exprimons les trois distributions des valeur extrêmes Λ, Φ_γ et Ψ_γ en termes de GEVD H_γ. La Figure 1.3 ci-dessous illustre le comportement de GEVD standard.

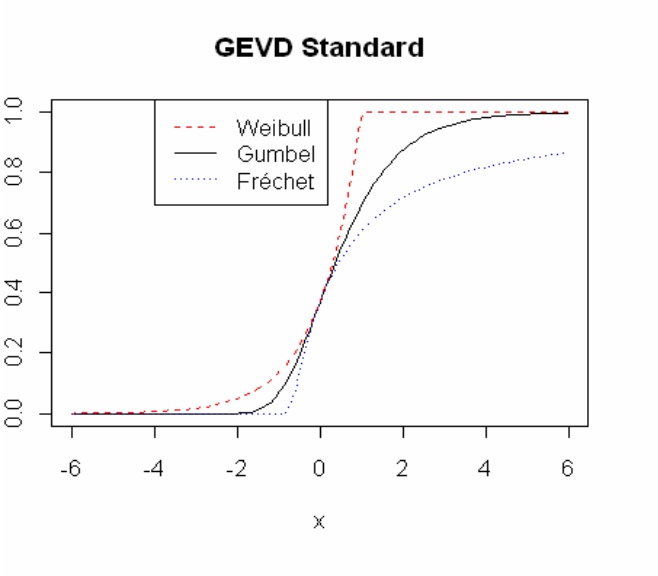


FIG. 1.3 – Distributions standard des valeurs extrêmes.

Proposition 1.6 (Λ, Φ_γ et Ψ_γ en terme de $H_{\mu,\sigma,\gamma}$) *Nous avons, les correspondances suivantes :*

$$H_{1,\frac{1}{\gamma},\frac{1}{\gamma}}(\gamma(x-1)) = \Phi_\gamma(x) \quad \text{si} \quad \gamma > 0.$$

$$H_{0,\frac{1}{\gamma},\frac{-1}{\gamma}}(\gamma(x+1)) = \Psi_\gamma(x) \quad \text{si} \quad \gamma < 0.$$

$$H_{0,1,0}(x) = \Lambda(x) \quad \text{si} \quad \gamma = 0.$$

La densité de la loi GEV s'écrit pour $\gamma \neq 0$ comme suit :

$$h_{\mu,\sigma,\gamma}(x) = \frac{1}{\sigma}\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-(1+\gamma/\gamma)} \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\}.$$

La fonction de densité standard correspondante $h_{\mu,\sigma,\gamma}$ est :

$$h_\gamma(x) = \begin{cases} H_\gamma(x)(1+\gamma x)^{-1/\gamma-1} & \text{si} \quad \gamma \neq 0,\ 1+\gamma x > 0, \\ \exp(-x-e^{-x}) & \text{si} \quad \gamma = 0,\ x \in \mathbb{R}. \end{cases}$$

La Figure 1.4 illustre les densities standard de GEV.

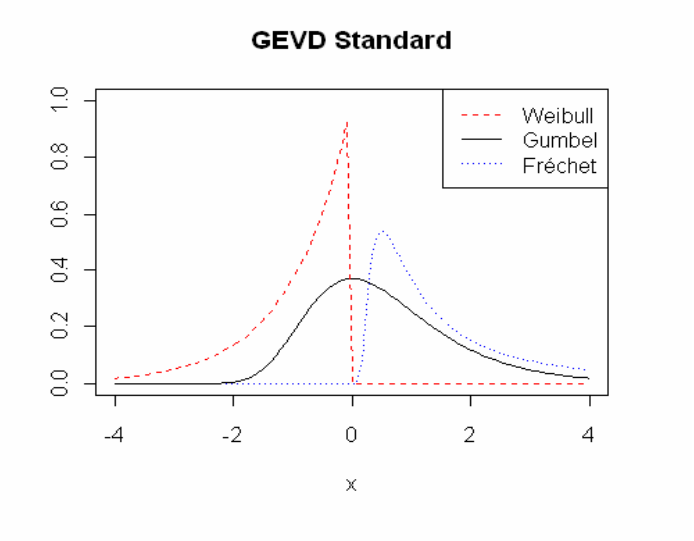


FIG. 1.4 – Densites standard des valeurs extrêmes.

1.3.3 Domaines d'attraction

Définition 1.8 (Domaines d'attraction) *Si F vérifie le Théorème 1.4, on dit alors que F apppartient au domaine d'attraction de H_γ, dénotée par $F \in \mathcal{D}(H_\gamma)$.*

Avant de caractériser les domaines d'attractions, on définit les fonctions à variations.

Définition d'une fonction à variations régulières

Définition 1.9 *On dit qu'une fonction G est à variations régulières d'indice $\rho \in \mathbb{R}$ à l'infini (notation $G \in \mathcal{R}_\rho$) si G est positive à l'infini (i.e. s'il existe A tel que pour tout $x \geq A$, $G(x) > A$) et pour tout $t > 0$*

$$\lim_{x \to +\infty} \frac{G(tx)}{G(x)} = t^\rho. \tag{1.3}$$

Dans le cas particulier où $\rho = 0$, on dit que G est une fonction à variation lentes à l'infini. Dans la suite, les fonctions à variations lentes sont notées l (et parfois L). En remarquant que si G est à variations régulières d'indice ρ alors $G(x)/x^\rho$ est à variations lentes, il est facile de montrer qu'une fonctions régulières d'indice ρ peut toujours s'écrire sous la forme $x^\rho l(x)$. Comme exemple de fonctions à variations lentes, citons la fonctions $\ln(1+x)$, $\ln[1+1/ln(1+x)]$, etc...

Théorème 1.6 (Représentation de Karamata) *l est une fonction à variations lentes si et seulement si pour tout $x > 0$,*

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

où c et ε sont des fonctions positives telles que

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

Remarques

1. Si la fonction c est constante, on dit que l est normalisée.
2. Le théorème 1.6 implique que si l est normalisée alors l est dérivable le dérivé λ avec pour $x > 0$:

$$\lambda(x) = \frac{\varepsilon(x) l(x)}{x}.$$

3. Soit G une fonction à variations régulières d'indice ρ. En utilisant le fait que $G(x) = x^\rho l(x)$, on déduit facilement du théorème 1.6 pour tout $x > 0$,

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

où c et ε sont des fonctions positives telles que

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

Caractérisation des domaines d'attraction

Nous allons donner des conditions sur la fonction de répartition F pour qu'elle appartienne à l'un des trois domaines d'attraction.

• **Domaine d'attraction de Fréchet**

Ce domaine d'attraction regroupe la majorité des distributions à queue lourde comme par exemple la loi de Cauchy, la loi de Pareto, Log-Gamma, et Student, etc...

Théorème 1.7 *$F \in \mathcal{D}(\Phi_\gamma)$ avec $\gamma > 0$ ssi $x_F = +\infty$ et $1 - F$ est une fonction à variations régulière d'indice $-1/\gamma$ (i.e. $1 - F(x) = x^{-1/\gamma} l(x)$, où l est une fonction à variations lentes). Dans ce cas, un choix possible pour les suites a_n et b_n est $a_n = F^{\leftarrow}(1 - \frac{1}{n})$ et $b_n = 0$.*

• **Domaine d'attraction de Weibull**

Ce domaine d'attraction contient la majorité des fonctions de répartition dont le point terminal est fini (loi uniforme, etc...).

Théorème 1.8 *$F \in \mathcal{D}(\Psi_\gamma)$ avec $\gamma < 0$ ssi $x_F < +\infty$ et $1 - F^*$ est une fonction à variations régulière d'indice $1/\gamma$ (i.e. $1 - F(x) = (x_F - x)^{-1/\gamma} [l(x_F - x)^{-1}]$). Avec*

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

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

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

• **Domaine d'attraction de Gumbel**

Ce domaine d'attraction regroupe la majorité des distributions à queue fine. par exemple loi normale, exponentielle, gamma, lognormale, etc...

Rappelons tout d'abord la définition d'une fonction de Von-Mises.

Définition 1.10 (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*

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

où $c > 0$ et a est une fonction positive absolument continue de densité a' vérifiant $\lim_{x \to x_F^+} a'(x) = 0$, alors F est une fonction de Von-Mises et a est sa fonction auxiliaire.

Théorème 1.9 *$F \in D(\Lambda)$ ssi il existe une fonction de Von-Mises F^* telle que pour $z < x < x_F$ on ait :*

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

où g et c sont des fonctions mesurables tels que $g(x) \to 1$ et $c(x) \to c > 0$ quand $x \to x_F$, et a est une fonction positive et absolument continue (par rapport à la mesure de Lebesgue) avec une densité a' satisfaire $\lim_{x \to x_F} a'(x) = 0$. Dans ce cas, on peut choisir $b_n = Q(1 - 1/n)$ et $a_n = a(b_n)$ comme des constantes de normalisation. Un choix possible pour a est

$$a(x) = \int_z^{x_F} \frac{F(t)}{F(x)} dt, \qquad x < x_F.$$

Les Tables 1.1 à 1.3 donnent quant à elles différents exemples de distributions standards dans ces trois domaines d'attraction.

Distribution	$1 - F(x)$	γ
Burr (β, τ, λ), $\beta > 0, \tau >, \lambda > 0$	$\left(\frac{\beta}{\beta + x^\tau}\right)^\lambda$	$1/\lambda\tau$
Fréchet $(1/\alpha)$, $\alpha > 0$	$1 - \exp(-x^{-\alpha})$	$1/\alpha$
Loggamma (m, λ), $m > 0$, $\lambda > 0$	$\frac{\lambda^m}{\Gamma(m)} \int_x^\infty (\log u)^{m-1} u^{-\lambda-1} du$	$1/\lambda$
Loglogistic (β, α), $\beta > 0$, $\alpha > 1$	$\frac{1}{1+\beta x^\alpha}$	$1/\alpha$
Pareto (α), $\alpha > 0$	$x^{-\alpha}$	$1/\alpha$

TAB. 1.1 – Quelques distributions de type Pareto associées à un indice positif.

Distribution	$1-F(x)$	γ
Gamma $(m,\lambda),\ m\in\mathbb{N},\ \lambda>0$	$\frac{\lambda^m}{\Gamma(m)}\int_x^\infty u^{m-1}\exp(-\lambda u)du$	0
Gumbel $(\mu,\beta),\ \mu\in\mathbb{R},\ \beta>0$	$\exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right)$	0
Logistique	$\frac{2}{1+\exp(x)}$	0
Lognormale $(\mu,\sigma),\ \mu\in\mathbb{R},\ \sigma>0$	$\frac{1}{\sqrt{2\pi}}\int_x^\infty \frac{1}{u}\exp\left(-\frac{1}{2\sigma^2}(\log u-\mu)^2\right)du$	0
Weibull $(\lambda,\tau),\ \lambda>0,\tau>0$	$\exp(-\lambda x^\tau)$	0

TAB. 1.2 – Quelques distributions associées à un indice nul.

Distribution	$1-F(x)$	γ
ReverseBurr$(\beta,\tau,\lambda,x_F),\ \beta>0,\tau>0,\lambda>0$	$\left(\frac{\beta}{\beta+(x_F-x)^{-\tau}}\right)^\lambda$	$-1/\lambda\tau$
Uniforme$(0,1)$	$1-x$	-1

TAB. 1.3 – Quelques distributions associées à un indice négatif.

Dans cette partie de la Section, nous allons établir des conditions nécessaires et suffisantes pour une distribution appartenant à la fonction F au domaine d'attraction de H_γ.

Proposition 1.7 *On a $F\in\mathcal{D}(H_\gamma)$ ssi*

$$n\overline{F}(x)(xa_n+b_n)\underset{n\to\infty}{\to}-\log H_\gamma(x),$$

pour une certaine suite $((a_n,b_n),\ n\geq 1)$ où $a_n>0$ et $b_n\in\mathbb{R}$. On a alors la convergence en loi de $(a_n^{-1}(M_n-b_n),n\geq 1)$ vers une variable aléatoire de fonction de répartition H_γ.

Preuve. Voir [14]. ■

Proposition 1.8 (Caractérisations de $\mathcal{D}(H_\gamma)$) *Pour $\gamma\in\mathbb{R}$ les affirmations suivantes sont équivalentes.*

(a) $F\in\mathcal{D}(H_\gamma)$

(b) Pour certaine fonction positive b

$$\lim_{t\to x_F}\frac{\overline{F}(t+xb(t))}{\overline{F}(t)}=\begin{cases}(1+\gamma x)^{-1/\gamma} & \text{si}\quad \gamma\neq 0,\\ e^{-x} & \text{si}\quad \gamma=0,\end{cases}\tag{1.4}$$

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

(c) Pour certaine fonction positive $\tilde{a}$

$$\lim_{s\to 0} \frac{Q(1-sx) - Q(1-s)}{\tilde{a}(s)} = \begin{cases} \dfrac{x^{-\gamma}-1}{\gamma} & \text{si} \quad \gamma \neq 0, \\ \log x & \text{si} \quad \gamma = 0, \end{cases}$$

pour $x > 0$.

(d) Pour certain fonction positive $a(t) = \tilde{a}(1/t)$

$$\lim_{t\to 0} \frac{U(tx) - U(t)}{a(t)} = \begin{cases} \dfrac{x^{\gamma}-1}{\gamma} & \text{si} \quad \gamma \neq 0, \\ \log x & \text{si} \quad \gamma = 0, \end{cases}$$

pour $x > 0$.

Les deux dernières affirmations sont respectivement équivalentes à :

$$\lim_{s\to 0} \frac{Q(1-sx) - Q(1-s)}{Q(1-sy) - Q(1-s)} = \begin{cases} \dfrac{x^{-\gamma}-1}{y^{-\gamma}-1} & \text{si} \quad \gamma \neq 0, \\ \dfrac{\log x}{\log y} & \text{si} \quad \gamma = 0, \end{cases}$$

$$\lim_{t\to 0} \frac{U(tx) - U(t)}{U(ty) - U(t)} = \begin{cases} \dfrac{x^{\gamma}-1}{x^{\gamma}-1} & \text{si} \quad \gamma \neq 0, \\ \dfrac{\log x}{\log y} & \text{si} \quad \gamma = 0, \end{cases} \tag{1.5}$$

pour $x,\ y > 0,\ y \neq 1$.

1.3.4 Conditions de Von Mises

Les conditions nécessaires et suffisantes pour qu'une fonction de répartition F appartienne à un domaine d'attraction d'une loi extrême donnée ci dessus sont difficiles à vérifier. Pour cette raison, pour présenterons ci-après des conditions suffisantes, dues à Von Mises (1936), plus simples à vérifiées et sans grandes restrictions. Néanmoins, elles ne seront applicables que pour les fonctions de répartition ayant une densité.

Théorème 1.10 (*Conditions de Von Mises*) *Soit F une fonction de distribution absolument continue dans l'intervalle $]x_1, x_F[$, de densité f. Alors les*

conditions suffisantes pour que F appartient à l'un des trois domaines d'attraction sont :

1. Si f admet une dérivée f' négative pour tout $x \in]x_1, x_F[$, $f(x) = 0$ pour $x \geq x_F$ et

$$\lim_{t \to x_F} \frac{f'(t)(1 - F(t))}{(f(t))^2} = -1,$$

alors $F \in \mathcal{D}(\Lambda)$.

2. Si $f(x) > 0$ pour tout $x \in]x_1, \infty[$ et pour $\gamma > 0$

$$\lim_{t \to \infty} \frac{tf(t)}{1 - F(t)} = \gamma,$$

alors $F \in \mathcal{D}(\Phi_\gamma)$.

3. Si $f(x) > 0$ pour tout $x \in]x_1, x_F[$, $f(x) = 0$ pour tout $x > x_F$ et pour $\gamma > 0$

$$\lim_{t \to x_F^+} \frac{(x_F - t)f(t)}{1 - F(t)} = \gamma,$$

alors $F \in \mathcal{D}(\Psi_\gamma)$.

1.4 Distribution de Pareto généralisée

La loi de Pareto généralisée notée GPD (*generalized Pareto distribution*), pour $\sigma > 0$, $\gamma \in \mathbb{R}$ est définie par la fonction de répartition $G_{\gamma,\mu,\sigma}$:

$$G_{\gamma,\mu,\sigma}(x) := \begin{cases} 1 - \left(1 + \gamma \left(\dfrac{x - \mu}{\sigma}\right)\right)^{-1/\gamma} & \text{si} \quad \gamma \neq 0, \\ 1 - \exp\left(-\dfrac{x - \mu}{\sigma}\right) & \text{si} \quad \gamma = 0, \end{cases}$$

où

$$\begin{cases} x \geq 0 & \text{si} \quad \gamma \geq 0, \\ x \in [\mu, \mu - \sigma/\gamma] & \text{si} \quad \gamma < 0. \end{cases}$$

Le GPD standard correspond au cas $\mu = 0$ et $\sigma = 1$.
La Figure 1.5 illustre la distibitution et la densité de la loi de GPD standard.
Le GPD avec les paramètres $\mu = 0$ et $\sigma > 0$ joue un rôle important (comme sera vu en chapitres deux), dans l'analyse statistique des événements extrêmes, en fournissant une approximation appropriée pour l'excès d'un grand seuil. Cette famille spéciale sera dénotée par $G_{\gamma,\sigma}$ et définie comme suit :

$$G_{\gamma,\sigma}(x) = \begin{cases} 1 - \left(1 + \gamma\dfrac{x}{\sigma}\right)^{-1/\gamma} & \text{si } \gamma \neq 0 \\ 1 - \exp\left(-\dfrac{x}{\sigma}\right) & \text{si } \gamma = 0 \end{cases}, \quad x \in \mathcal{S}(\gamma, \sigma),$$

où

$$\mathcal{S}(\gamma, \sigma) := \begin{cases} [0, \infty[& \text{si} \quad \gamma \geq 0, \\ x \in [0, -\sigma/\gamma] & \text{si} \quad \gamma < 0, \end{cases}$$

est le support de $G_{\gamma,\sigma}$.

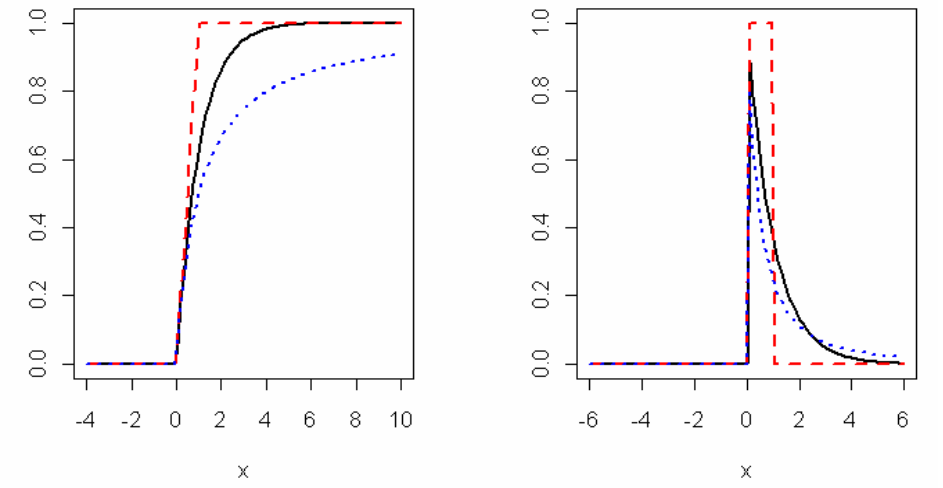


FIG. 1.5– Densité (à gauche) et la distribution (à droite) de GPD standard.

Le GPD avec deux paramètres regroupe trois distributions selon les valeurs du paramètre de forme. Lorsque $\gamma > 0$, c'est la loi Pareto ; lorsque $\gamma < 0$, nous avons la loi Bêta et $\gamma = 0$ donne la loi exponentielle.
Dans la Figure 1.6 ci-dessous, nous illustrons le comportement de différentes GPD pour $\sigma = 1$ et différentes valeurs de γ.

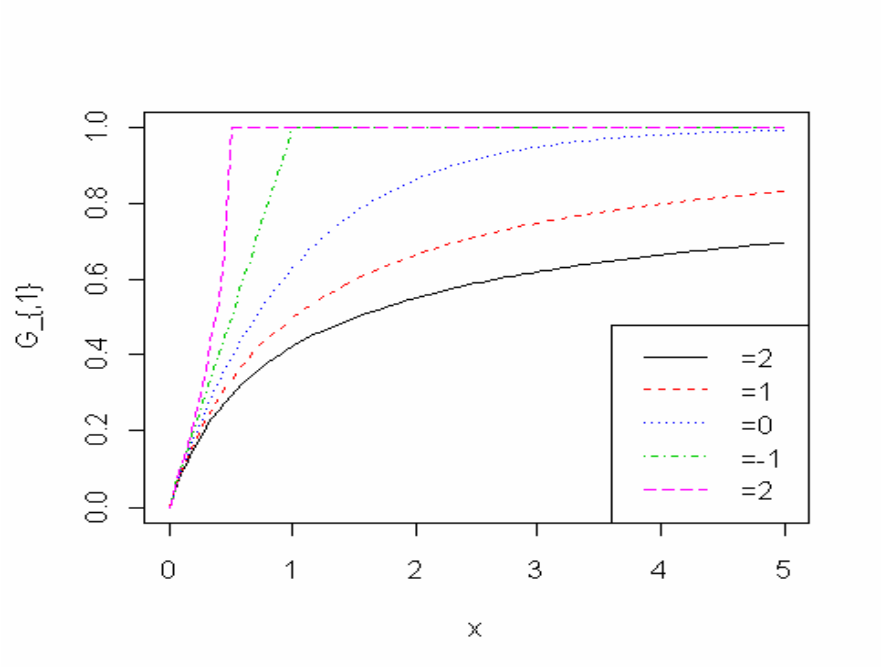


FIG. 1.6– Distributions de Pareto généralisées $G_{\gamma,1}$.

Chapitre 2

Estimation de l'indice des valeurs extrêmes et de quantiles extrêmes

La loi des valeurs extrêmes, lorsqu'elle existe, est indexée par un paramètre appelé indice des valeurs extrêmes, ainsi qu'éventuellement par des paramètres d'échelle et de position. La connaissance de l'indice des valeurs extrêmes est un élément important car il contrôle l'épaisseur (lourdeur) de la queue de distribution.

La question est comment estimer ce paramètre crucial d'un échantillon fini $(X_1, ..., X_n)$: La réponse est que l'estimation peut être effectuée après deux approches diffèrentes. Le premier, celui nous appellerons l'approche d'EVT, est basés sur la GEVD et le second, connu comme approche de POT, utilise la GPD. Dans les deux cas, il existe plusieurs méthodes d'estimation. Pour beaucoup des détails voir [20].

Dans ce chapitre, nous nous concentrons sur certains des estimateurs les plus connus avec leurs propriétés asymptotiques. Aussi, nous passons en revue certaines des méthodes classiques à construire les estimateurs pour des quantiles extrêmes et des queues de distribution. Mais d'abord, nous commençons par quelques techniques exploratoires pour les extrêmes.

2.1 Analyse exploratoire des données

L'information préliminaire utile sur un ensemble proposé de données à analyser, peut être obtenu par plusieurs résultats graphiques et analytiques assez faciles. On commence habituellement l'analyse statistique des données par la détermination des statistiques de base (moyenne, variance ,...) et dessin nuages de points, histogrammes, box-plots, ...en outre, dans l'analyse des extrêmes, la première tâche consiste en étudiant le poids de queue des données.

Dans cette Section, nous présentons quelques outils graphiques qui sont très utiles pour explorer visuellement la qualité de l'ajustement d'un modèle de Pareto à la queue d'une distribution.

2.1.1 Probabilité et Quantile Plots

Supposons que $\hat{F}$ l'estimateur de la fonction de répartition F a été obtenu. La probabilité probabilité plot (PP-plot) et le quantile quantile plot (QQ-plot) peuvent fournir une estimation graphique à la fonction de distribution ajustée $\hat{F}$.

Définition 2.1 (PP-plot) *Le graphique*

$$\left\{ \left(F\left(X_{i,n}\right), \frac{i}{n+1} \right) : \ i = 1, ..., n \right\},$$

est appelé PP-plot.

Définition 2.2 (QQ-plot) *Le graphique*

$$\left\{ \left(X_{i,n}, \hat{F}^{\leftarrow}\left(\frac{i}{n+1} \right) \right) : \ i = 1, ..., n \right\},$$

est appelé QQ-plot.

$X_{i,n}$ est le quantile empirique d'ordre $(i/(n+1))$ de la fonction de distribution F tandis que $\hat{F}^{\leftarrow}(i/(n+1))$ est son estimateur.

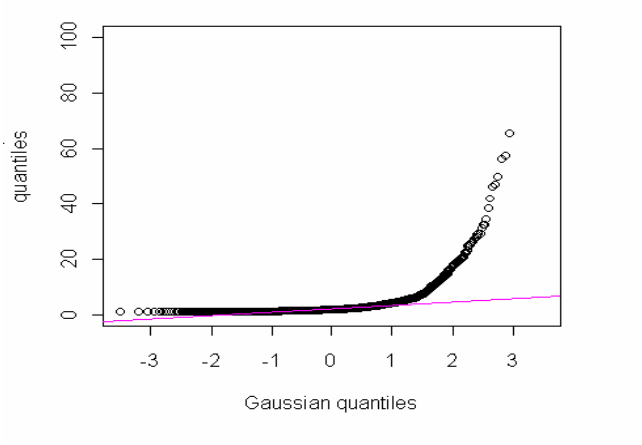


FIG. 2.1 – Quantiles empirique de l'ensemble des données de réclamations danoises contre les quantiles gaussiens.

Le QQ-plot est un graphique qui oppose les quantiles de la distribution empirique aux quantiles de la distribution théorique envisagée. Si l'échantillon provient bien de cette distribution théorique, alors le QQ-plot sera linéaire.

Dans la théorie des valeurs extrêmes, le QQ-plot se base sur la distribution exponentielle. Le QQ-plot sous l'hypothèse d'une distribution exponentielle est la représentation des quantiles de la distribution empirique sur l'axe des X contre les quantiles de la fonction de distribution exponentielle sur l'axe des Y.
Le graphique est l'ensemble des points tel que :

$$\left\{\left(X_{k,n}, G_{0,1}^{\leftarrow}\left(\frac{n-k+1}{n+1}\right)\right),\ k=1,...,n\right\},$$

$X_{k,n}$: Représente le $k^{ième}$ ordre statistique et $G_{0,1}^{\leftarrow}$ est la fonction inverse de la distribution exponentielle.
L'intérêt de ce graphique est de nous permettre d'obtenir la forme de la queue de la distribution. Trois cas de figure sont possibles :

(a) Les données suivent la loi exponentielle : la distribution présente une queue très légère, les points du graphique présentent une forme linéaire.

(b) Les données suivent une distribution à queue épaisse "fat-tailed distribution" : le graphique QQ-plot est concave.

(c) Les données suivent une distribution à queue légère "short-tailed distribution" : le graphique QQ-plot a une forme convexe.

Définition 2.3 (Pareto quantile plot) *Le Pareto quantile plot ou Zipf plot consiste à dessiner les points de corrdonnées*

$$\left\{\left(\log\left(\frac{n+1}{i}\right), \log X_{n-i+1,n}\right),\ i=1,...,n\right\}.$$

Pour avoir une bonne adéquation, le Pareto quantile plot doit être alors approximativement linéaire, et la pente ajustée aux points de Pareto quantile plot coïncide avec l'estimateur de l'index γ.
En d'autres termes, le "Pareto quantile plot" sera approximativement linéaire, avec une pente γ, pour les petites valeurs de i, i.e. les points extrêmes.

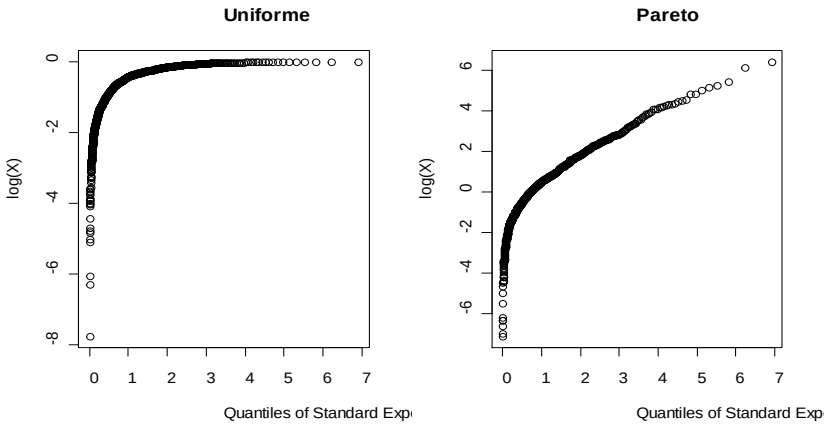


FIG. 2.2 – Pareto-quantile plot de la distribution uniforme standard (à gauche) et de la distribution de Pareto standard (à droite), basé sur 1000 observations.

2.1.2 Quantile plot généralisé

Une approche permettant d'éviter le choix à priori du domaine d'attraction a été proposée par Beirlant et al. (1996). Elle consiste à utiliser un "quantile plot généralisé", défini comme le graphe

$$\left\{\left(\log\frac{n+1}{j}, \log UH_{j,n}\right) : j=1,...,n\right\},$$

avec $UH_{j,n}$ de la forme

$$UH_{j,n} = X_{n-j,n}\left(j^{-1}\sum_{i=1}^{j}\log X_{n-i+1,n} - \log X_{n-j,n}\right).$$

Suivant la courbure de ce graphe, on peut déduire dans quel domaine d'attraction on se situe : si pour les points extrêmes on voit apparaître une droite de pente positive, on est alors dans le domaine de Fréchet, si par contre on est plutôt constant, on est alors dans le domaine de Gumbel, le cas d'une décroissance linéaire signifie que l'on appartient au domaine de Weibull.

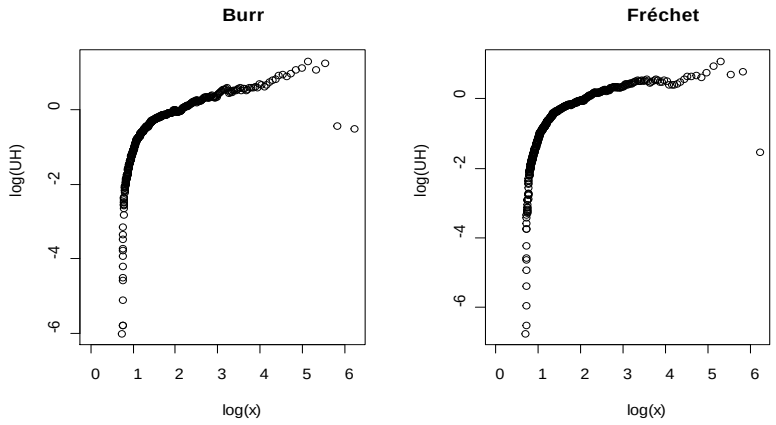


FIG. 2.3 – Pareto-quantile plot de la distribution Burr (à gauche) et de la distribution de Fréchet(1) (à droite), basé sur 1000 observations.

2.1.3 Mean Excess Function Plot

Un test graphique pour le comportement de queue peut être effectué par *mean-excess plot*, ce dernier consiste à représenter $\hat{e}(u)$ en fonction de u avec

$$\hat{e}(u) = e_n(u) := \frac{1}{\bar{F}_n(u)}\int_u^{\infty}\bar{F}_n(x)dx = \frac{1}{N_u}\sum_{i=1}^{n}(X_i - u)\,I_{\{X_i>u\}}, \qquad (2.1)$$

i.e. la somme des excès au-dessus du seuil u divisé par le nombre N_u de données qui excèdent u.

$$N_u = \#\{i \in \{1,...,n\} : X_i > u\}.$$

La fonction moyenne des dépassements empirique (*sample mean excess function*) $e_n(u)$ est l'estimateur empirique de la fonction moyenne des dépassements

$$e(u) = E[X - u/X > u]$$

Définition 2.4 *Le ME-plot est défini de la manière suivante :*

$$\{(X_{k,n}, e_n(X_{k,n})) : k = 1, ..., n\}.$$

Pour nos buts, ME-plot est utilisé seulement comme méthode graphique, principalement pour distinguer les modèles à queue lourde et les modèles à queue légère, voir la Figure 2.4. Rappelons que, pour la GPD la fonction moyenne des dépassements est linéaire.

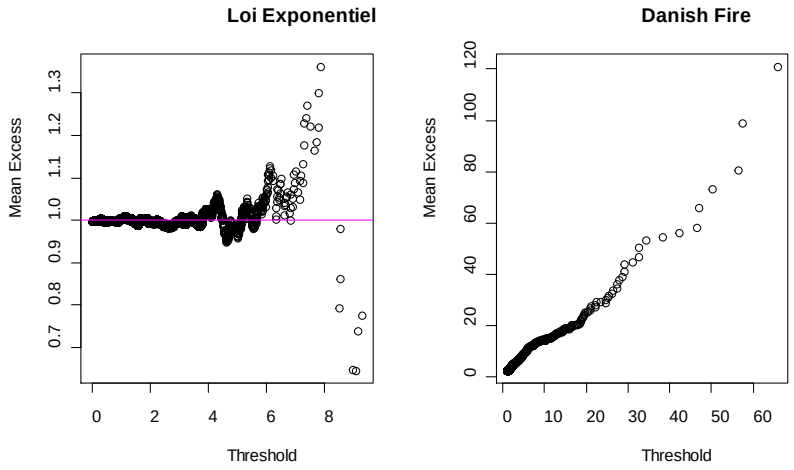


FIG. 2.4 – Mean-plot excès des données simulées à partir d'une distribution exponentielle (à gauche) et de données de Réclamations danoises (à droite).

2.1.4 Exemple Illustratif

On termine cette section avec un exemple explicatif et on utilise les données de Rendements quotidiens des parts de BMW. Les résultats sont résumés dans le tableau 2.1 et illustré par la Figure 2.5.

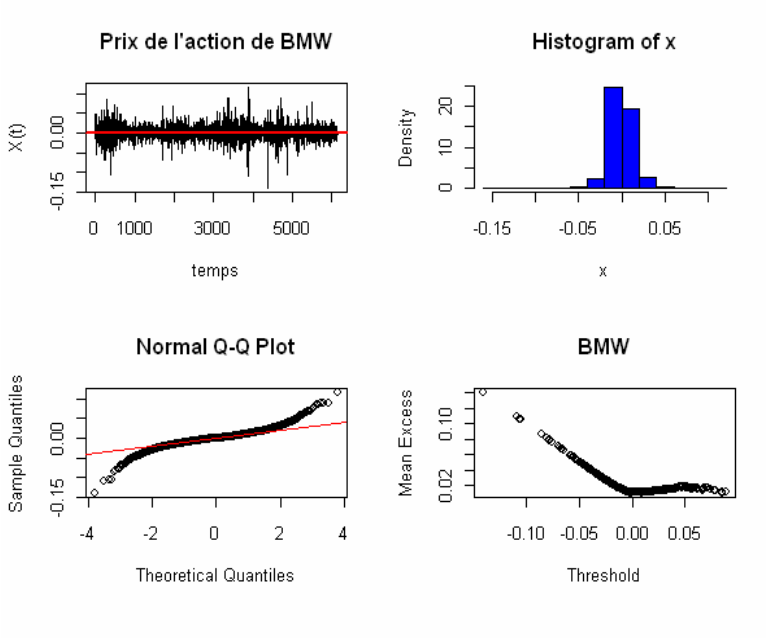


FIG. 2.5 – Analyse exploratoire des données de Rendements quotidiens des parts de BMW.

Min.	$1^{ière}$ Qu.	Médian	Moyenne	$3^{ième}$ Qu.	Max.	$\hat{x}_{0.99}$
-0.1406	-0.0066	0.0000	0.0003	0.0071	0.1172	0.0414

TAB. 2.1 – Statistiques générales sur les données de rendements quotidiens des parts de BMW.

2.2 *Modèle EVT*

On veut estimer les paramètres de la loi des valeurs extrêmes de fonction de répartition de GEV :

$$H_\theta(x) = \begin{cases} \exp\left\{-\left(1+\gamma\left(\dfrac{x-\mu}{\sigma}\right)\right)^{-1/\gamma}\right\}, & \gamma \neq 0,\ 1+\gamma\dfrac{x-\mu}{\sigma} > 0, \\ \exp\left(-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right) & \gamma = 0,\ x \in \mathbb{R}. \end{cases}$$

Où le vecteur $\theta := (\gamma, \mu, \sigma) \in \Theta \subset \mathbb{R}^2 \times \mathbb{R}_+$ se compose d'un paramètre de forme, d'un paramètre de localisation et d'un paramètre d'échelle. Ces para-

mètres doivent être estimés en utilisant un échantillon $X_1, ..., X_n$ de n variables aléatoires indépendantes et de même fonction de répartition F.

Deux situations peuvent être considérées : le premier est quand F est exactement H_θ et le deuxième correspond au cas où F est dans $\mathcal{D}(H_\gamma)$. Dans cette Section, nous traitons la première situation, nous pouvons estimer les paramètres de la GEVD. Deux méthodes d'estimation sont envisageables : l'Estimation par Maximum de Vraisemblance (EMV) et celle par les Moments Pondérés (EMP).

2.2.1 Méthode du Maximum de Vraisemblance (EMV)

Les paramètre d'un modèle statistique peuvent être estimés par la méthode du maximum de vraisemblance. Dans cette méthode, on retient les paramètres qui maximisent la fonction de vraisemblance ou, plus souvent le logarithme de la fonction de vraisemblance en fonction des paramètres de la famille de lois choisie pour l'ajustement.

Dans la théorie l'estimation paramétrique, la méthode de maximum de vraisemblance (EMV) est la technique la plus populaire. Elle rapporte les estimateurs efficaces, consistants et asymptotiquement normaux. Si H_θ a la fonction de densité h_θ alors la fonction de vraisemblance *(likelihood function)* basée sur les données $(X_1, ..., X_n)$ est définie comme suit

$$L(\theta; X_1, ..., X_n) := \prod_{i=1}^{n} h_\theta(X_i) 1_{\{1+\gamma(X_i-\mu)/\sigma>0\}} \qquad \text{(car l'échantillon est i.i.d)},$$

et la fonction log-vraisemblance est

$$l(\theta; X_1, ..., X_n) := \log L(\theta; X_1, ..., X_n).$$

l'estimateurs du maximum de vraisemblance de θ est définie comme suit

$$\hat{\theta}_n = \hat{\theta}_n (X_1, ..., X_n) = \arg\max_{\theta\in\Theta} l(\theta; X_1, ..., X_n),$$

i.e. $\hat{\theta}_n$ maximise $l(\theta; X_1, ..., X_n)$ sur l'espace du paramètre Θ : Une approche à déterminer $\hat{\theta}_n$ est résoudre le système de vraisemblance (*the likelihood system*) obtenu en mettant égal à 0 les dérivés partiels de $l(\theta; X_1, ..., X_n)$ en ce qui concerne γ, μ et σ.

Dans le cas où $\gamma = 0$, la log-vraisemblance est égale à :

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

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

$$\begin{cases} n - \sum_{i=1}^{n}\exp\left\{-\frac{X_i-\mu}{\sigma}\right\} = 0. \\ n + \sum_{i=1}^{n}\frac{X_i-\mu}{\sigma}\left(\exp\left\{-\frac{X_i-\mu}{\sigma}\right\} - 1\right) = 0. \end{cases}$$

C'est un système non linéaire pour lesquels aucune solution explicite existe. Quand $\gamma \neq 0$, la situation est encore plus compliquée.

2.2.2 Estimateurs des Moments Pondérés (EMP)

Cette méthode, qui remonte à *Hosking, Wallis, Wood* (1985) est basée sur la quantité suivante :

$$\omega_r(\theta) := E\left[X\ H_\theta^r(X)\right], \qquad r \in \mathbb{N},$$

et son analogue empirique

$$\hat{\omega}_r(\theta) := \int_{-\infty}^{+\infty} x H_\theta^r(x) dF_n(x), \qquad r \in \mathbb{N},$$

où F_n est la fonction de répartition empirique de l'échantillon $(X_1,...,X_n)$: On note que $\omega_r(\theta)$ est la moyenne de la distribution et $\hat{\omega}_r(\theta)$ est la moyenne de l'échantillon.

Observons que, par un changement de variable, nous avons

$$\hat{\omega}_r(\theta) := \int_0^1 H_\theta^{\leftarrow}(y) y^r dy,$$

où, pour $0 < y < 1$,

$$H_\theta^{\leftarrow}(y) = \begin{cases} \mu - \frac{\sigma}{\gamma}(1-(1-\log y)^{-\gamma}) & \text{si} \quad \gamma \neq 0, \\ \mu - \sigma\log(-\log y) & \text{si} \quad \gamma = 0. \end{cases}$$

D'autre part, nous avons

$$\hat{\omega}_r(\theta) = \frac{1}{n}\sum_{i=1}^{n} X_{i,n} H_\theta^r(X_{i,n}), \tag{2.2}$$

où $X_{1,n}, ..., X_{n,n}$ sont les statistiques d'ordre (s.o) associées à l'échantillon $(X_1, ..., X_n)$.
Rappeler la transformation de quantile du lemme (1.2) :

$$(H_\theta(X_{1,n}), ..., H_\theta(X_{n,n})) \stackrel{d}{=} (U_{1,n}, ..., U_{n,n}),$$

où $U_{1,n}, ..., U_{n,n}$ sont les statistiques d'ordre (s.o) associées à l'échantillon uniforme standard $(U_1, ..., U_n)$. Avec cette interprétation (2.2) peut être écrite comme

$$\hat{\omega}_r(\theta) = \hat{\omega}_r = \frac{1}{n}\sum_{i=1}^{n} X_{i,n} U_{i,n}^r, \ r = 0, 1, 2. \tag{2.3}$$

Dans la pratique $U_{i,n}^r$ est souvent rapproché par son espérance

$$EU_{i,n}^r = \frac{(n-i)(n-i-1)...(n-i-r+1)}{(n-1)(n-2)(n-r)}, \ r = 1, 2.$$

L'estimateur de EMP du θ est obtenu en résolvant le système de trois équations

$$\omega_r(\theta) = \hat{\omega}_r(\theta), \ r = 0, 1, 2.$$

Pour $\gamma < 1$ et $\gamma \neq 0$,

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

Dans ce cas, le système ci-dessus est équivalent à

$$\begin{cases} \hat{\omega}_0(\theta) & = & \mu - \dfrac{\sigma}{\gamma}\left(1 - \Gamma(1-\gamma)\right), \\ 2\hat{\omega}_1(\theta) - \hat{\omega}_0(\theta) & = & \dfrac{\sigma}{\gamma}\Gamma(1-\gamma)(2^\gamma - 1), \\ 3\hat{\omega}_2(\theta) - \hat{\omega}_0(\theta) & = & \dfrac{\sigma}{\gamma}\Gamma(1-\gamma)(3^\gamma - 1), \end{cases}$$

et par conséquent

$$\frac{3\hat{\omega}_2(\theta) - \hat{\omega}_0(\theta)}{2\hat{\omega}_1(\theta) - \hat{\omega}_0(\theta)} = \frac{3^\gamma - 1}{2^\gamma - 1}.$$

La solution de cette équation est l'estimateur de EMP $\hat{\gamma}$ de γ. Les autres paramètres σ et μ sont estimés respectivement par :

$$\hat{\sigma} = \frac{(2\hat{\omega}_1 - \hat{\omega}_0)\,\hat{\gamma}}{\Gamma(1-\hat{\gamma})(2\hat{\gamma}-1)},$$

et

$$\hat{\mu} = \hat{\omega}_0 + \frac{\hat{\sigma}}{\hat{\gamma}}(1 - \Gamma(1-\hat{\gamma})).$$

2.3 Estimation semi-paramétrique

Dans cette Section, nous donnons quelques estimateurs de l'IVE γ construits sous les conditions de domaine d'attraction.

Par opposition aux méthodes paramétriques de la Section précédente, des procédures statistiques semi-paramétrique appropriées à cette situation, n'assument pas la connaissance de la distribution entière mais seulement sur les queues de distribution.

Les estimateurs classiques sont basés sur les plus grandes statistiques d'ordre $X_{n-k,n} \leq ... \leq X_{n,n}$, où k est une suite intermédiaire d'entiers liés à la taille de l'échantillon n de la façon suivante :

$$k = k_n \to \infty \text{ et } k/n \to 0 \qquad \text{quand} \qquad n \to \infty.$$

2.3.1 Estimateur de Pickands

Pickands a proposé l'estimateur suivant pour $\gamma \in \mathbb{R}$ et $1 \leq k \leq [n/4]$

$$\hat{\gamma}_n^{(P)} := (\log 2)^{-1} \log \frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}}.$$

Il a prouvé la consistance faible de l'estimation.

Construction de l'estimateur de Pickands

On déduit de la relation 1.5 que pour $\gamma \in \mathbb{R}$, on a avec le choix $t = 2s,\ x = 2$ et $y = 1/2$,

$$\lim_{t\to\infty}\frac{U(t)-U(t/2)}{U(t/2)-U(t/4)}=2^{\gamma}.$$

En fait, en utilisant la croissance de U qui se déduit de la croissance de F, on obtient

$$\lim_{t\to\infty}\frac{U(t)-U(tc_1(t))}{U(tc_1(t))-U(tc_2(t))}=2^{\gamma}.$$

dès que $\lim_{t\to\infty} c_1(t) = 1/2$ et $\lim_{t\to\infty} c_2(t) = 1/4$. Il reste donc à trouver des estimateurs pour $U(t)$.

Soit $(V_{1,n},...,V_{n,n})$ la statistique d'ordre d'un échantillon de variables aléatoires indépendantes de loi de Pareto. On le note ($F_V(x) = 1 - x^{-1}$, pour $x \geq 1$). On déduit avec certains résultats de base liées à $(V_{1,n},...,V_{n,n})$ que les suites $(\frac{k}{n}V_{n-k+1,n},\ n\geq 1)$, $(\frac{2k}{n}\ V_{n-2k+1,n},\ n\geq 1)$ et $(\frac{4k}{n}V_{n-4k+1,n},\ n\geq 1)$ convergent en probabilité vers 1.

En particulier, on a les convergences en probabilités suivantes :

$$V_{n-k+1,n}\underset{n\to\infty}{\to}\infty,\quad \frac{V_{n-2k+1,n}}{V_{n-k+1,n}}\underset{n\to\infty}{\to}1/2,\quad \text{et}\quad \frac{V_{n-4k+1,n}}{V_{n-k+1,n}}\underset{n\to\infty}{\to}1/4.$$

On en déduit donc que la convergence suivante a lieu en probabilité :

$$\frac{U(V_{n-k+1,n})-U(V_{n-2k+1,n})}{U(V_{n-2k+1,n})-U(V_{n-4k+1,n})}\underset{n\to\infty}{\to}2^{\gamma}.$$

Remarquons que si $x \geq 1$, alors $U(x) = F^{-1}(F_V(x))$.
On déduit de la croissance de F_V que $((F^{-1}(F_V(V_{1,n})),...,F^{-1}(F_V(V_{n,n})))$,a même loi que la s.o de n variables aléatoires uniformes sur $[0,1]$ indépendantes. Alors le vecteur aléatoire $((F^{-1}(F_V(V_{1,n})),...,F^{-1}(F_V(V_{n,n})))$, a même loi que $(X_{1,n},...,X_{n,n})$.

Théorème 2.1 (Propriétés asymptotiques de $\hat{\gamma}_n^{(P)}$) *Supposons que $F \in \mathcal{D}(H_\gamma)$, $\gamma \in \mathbb{R}$, $k \to \infty$ et $k/n \to 0$ quand $n \to \infty$:*

(a) Consistance faible :

$$\hat{\gamma}_n^{(P)}\xrightarrow{p}\gamma\quad \text{quand } n\to\infty.$$

(b) Consistance forte : si $k/\log\log n \to \infty$ quand $n \to \infty$, alors

$$\hat{\gamma}_n^{(P)}\xrightarrow{p.s}\gamma\quad \text{quand } n\to\infty.$$

(c) Normalité asymptotique : sous certaines conditions pour k et F on a:

$$\sqrt{k}\left(\hat{\gamma}_n^{(P)} - \gamma\right) \xrightarrow{d} \mathcal{N}(0, \eta^2), \quad \text{quand } n \to \infty,$$

où

$$\eta^2 := \frac{\gamma^2(2^{2\gamma+1}+1)}{(2(2^\gamma - 1)\log 2)^2}.$$

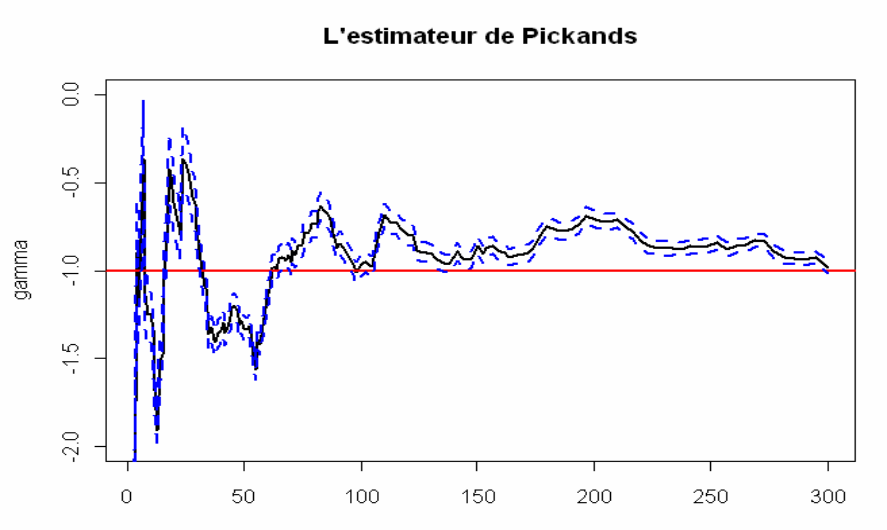


FIG. 2.6 – Estimateur de Pickands, en fonction du nombre des extrêmes (en trait plein) avec l'intervalle de confiance 95% (lignes tirées), pour l'IVE de la distribution uniforme ($\gamma = -1$) basé sur 100 échantillons de 3000 observations.

Une généralisation de l'estimateur de Pickands à été introduite dans [43] comme suit

$$\hat{\gamma}_n^{(Y)} = \hat{\gamma}_n^{(Y)}(k; u, v) := (\log v)^{-1} \log \frac{X_{n-k+1,n} - X_{n-[uk]+1,n}}{X_{n-[vk]+1,n} - X_{n-[uvk]+1,n}},$$

où u, v sont des nombres réels positifs différents de 1 tels que $[uk]$ $[vk]$ et $[uvk]$ ne dépassent pas n. Pour $u = v = 2$, nous avons $\hat{\gamma}_n^{(P)}$.

2.3.2 Estimateur de Hill

Une grande partie de la théorie de l'estimation de l'indice de valeur extrême est développée pour des valeurs extrêmes des indices positifs. Le meilleur estimateur connu d'un indice des valeurs extrêmes positif est l'estimateur de Hill (Hill, 1975) défini de la façon suivante :

$$\hat{\gamma}_n^{(H)} := \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n}, \tag{2.4}$$

avec $X_{1,n}, ..., X_{n,n}$ les statistiques d'ordre associées à l'échantillon $X_1, ..., X_n$.
En premier lieu, si on note $E_{j,u}$ les excès relatifs au-delà de u i.e. $E_{j,u} := X_j/u$ avec $X_j > u$, on peut facilement vérifier que

$$P(E_{j,u} > x/E_{j,u} > 1) \to x^{-1/\gamma} \qquad \text{quand } u \to \infty,\ x > 1.$$

En formant la vraisemblance basée sur cette distribution limite, on vérifie facilement que l'estimateur de Hill n'est rien d'autre que l'estimateur du maximum de vraisemblance dans le cas où le seuil $u = X_{n-k,n}$ et en utilisant les statistiques d'ordre $X_{n-k+1,n}, ..., X_{n,n}$.

la fonction de log-vraisemblance sera alors :

$$L(\gamma; X_{n-k+1,n}, ..., X_{n,n}) =$$
$$-k\log(\gamma u) + \log(1 - F(u)) - \frac{\gamma+1}{\gamma}\sum_{i=1}^{k}\left(\log X_{n-i+1,n} - \log u\right).$$

En maximisant la fonction log-vraisemblance par rapport à γ, on obtient l'estimateur de Hill pour $\gamma > 0$.

En second lieu, un côté très attrayant de l'estimateur de Hill est qu'il est possible de l'interpréter graphiquement. Ceci est particulièrement important pour les praticiens, qui préfèrent souvent des interprétations graphiques à des formules mathématiques. Plus précisément, si on utilise le graphe "Pareto quantile plot" qui définit dans la Section 2.1, dans le cas de distributions de type Pareto, ce graphe sera approximativement linéaire, dans les points extrêmes, avec une pente γ.

Comportement de l'estimateur de Hill

(a) Les propriétés asymptotiques de l'estimateur de Hill ont été établies par Mason (1982) qui a prouvé la consistance faible de l'estimateur de Hill $\hat{\gamma}_n^{(H)}$ pour toute suite vérifiant :

$$k = k_n \to \infty \text{ et } k_n/n \to 0 \text{ quand } n \to \infty.$$

(b) La consistance forte a été prouvée dans Deheuvels, Haeusler et Mason (1985) sous les condition que

$$k/\log\log n \to \infty \text{ et } k_n/n \to 0 \text{ quand } n \to \infty.$$

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

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

On peut associer à l'estimateur de Hill un intervalle de confiance asymptotique $I_N(\alpha)$ de niveau α.

$$I_N(\alpha) = \left[\hat{\gamma} - z_{\alpha/2}\hat{\gamma}\frac{1}{\sqrt{k}}, \hat{\gamma} + z_{\alpha/2}\hat{\gamma}\frac{1}{\sqrt{k}}\right],$$

où $z_{\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite. Pour plus de détails, voir [26]

Classe de Hall de la fonction de distribution

Hall (1982) a considéré les fonctions de distribution F qui satisfont

$$F(x) = 1 - cx^{-1/\gamma}(1 + dx^{-\rho/\gamma} + o(x^{-\rho/\gamma})) \qquad \text{quand } x \to \infty, \tag{2.5}$$

pour $\gamma > 0,\ \rho \leq 0,\ c > 0,\ d \in \mathbb{R}\backslash\{0\}$. Il a prouvé la normalité asymptotique pour l'estimateur de Hill.

Cette sous-classe des distributions à queue lourd contient les distributions Pareto, Burr, Fréchet et t-Student.

La relation (1.5) peut être reformulée en termes de fonctions Q et U comme suit :

$$Q(1-s) = c^{\gamma} s^{-\gamma} \left(1 + \gamma d c^{\rho} s^{-\rho} + o(s^{-\rho})\right) \qquad \text{quand } s \to \infty.$$

et

$$U(t) = c^{\gamma} t^{\gamma} (1 + \gamma d c^{\rho} t^{\rho} + o(t^{\rho})) \qquad \text{quand } t \to \infty.$$

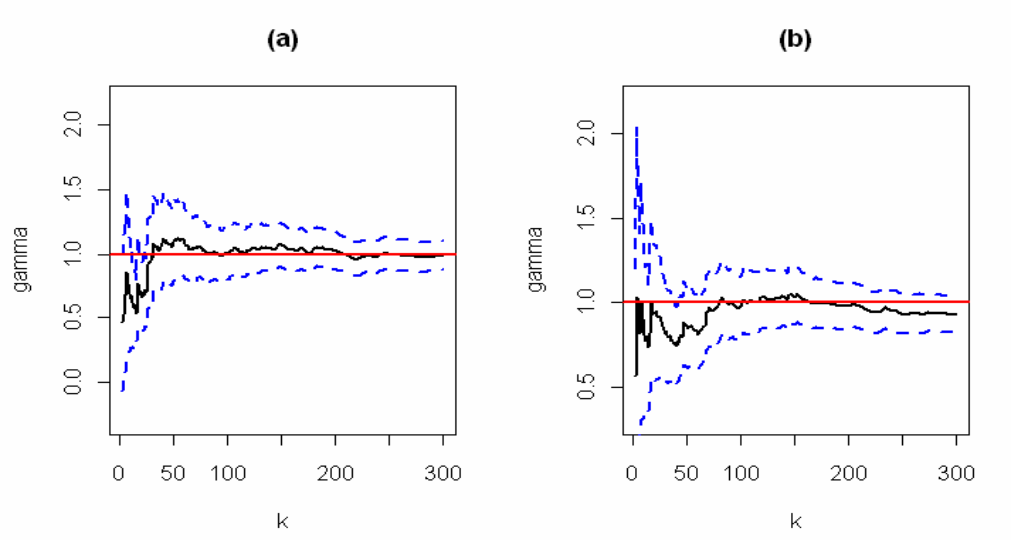


FIG. 2.7 – Estimateur de Hill, en fonction du nombre des extrêmes (en trait plein) avec l'intervalle de confiance 95% (lignes tirées), pour (a) la distribution de Pareto standard et pour (b) la distribution de Fréchet(1) basées sur 100 échantillons de 3000 observations.

Le principal inconvénient de l'estimateur de Hill est qu'il n'est valable que dans le cas d'un indice positif.
Différentes généralisations de l'estimateur de Hill ont été proposées. Parmi elles, on peut citer l'estimateur des moments (voir [16]) ou encore l'estimateur UH (voir [2]) qui vont être brièvement présentés ci-dessous.

2.3.3 Estimateur du moment

Un autre estimateur qui peut être considérée comme une adaptation de l'estimateur de Hill, pour obtenir la consistance pour tout $\gamma \in \mathbb{R}$, a été proposé par Dekkers et al. (1989). C'est l'estimateur de moment, donnée par

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

où

$$M_n^{(r)} := \frac{1}{k}\sum_{i=0}^{k}\left(\log X_{n-i+1,n} - \log X_{n-k,n}\right)^r ;\ r = 1, 2. \tag{2.6}$$

Théorème 2.2 (Propriétés asymptotiques de $\hat{\gamma}_n^{(M)}$) *Supposons que $F \in \mathcal{D}(H_\gamma)$, $\gamma \in \mathbb{R}$, $k \to \infty$ et $k/n \to 0$ quand $n \to \infty$:*

(a) *Consistance faible :*

$$\hat{\gamma}_n^{(M)} \xrightarrow{p} \gamma \qquad \text{quand } n \to \infty.$$

(b) *Consistance forte : Si $k/(logn)^\delta \to \infty$ quand $n \to \infty$ pour certaine $\delta > 0$, alors*

$$\hat{\gamma}_n^{(M)} \xrightarrow{p.s} \gamma \qquad \text{quand } n \to \infty.$$

(c) *Normalité asymptotique :* (voir le théorème 3.1 et le corollaire 3.2 de [16])

$$\sqrt{k}\left(\hat{\gamma}_n^{(M)} - \gamma\right) \xrightarrow{d} \mathcal{N}(0, \eta^2) \qquad \text{quand } n \to \infty,$$

où

$$\eta^2 := \begin{cases} 1 + \gamma^2, & \gamma \geq 0, \\ (1-\gamma^2)(1-2\gamma)\left(4 - 8\dfrac{1-2\gamma}{1-3\gamma} + \dfrac{(5-11\gamma)(1-2\gamma)}{(1-3\gamma)(1-4\gamma)}\right), & \gamma < 0. \end{cases}$$

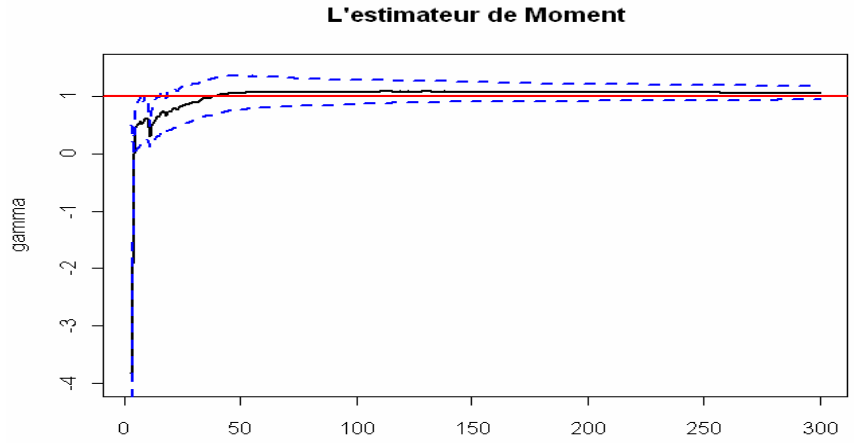


FIG. 2.8– Estimateur de Moment, en fonction du nombre des extrêmes (en trait plein) avec l'intervalle de confiance 95% (lignes tirées), pour la distribution de Burr(1,1,1) basée sur 100 échantillons de 3000 observations.

La normalité de cet estimateur a été établie par Dekkers et al. [16] sous des conditions de régularité convenables. Le problème de cet estimateur est que l'on ne peut pas, contrairement à l'estimateur de Hill, l'interpréter graphiquement. Afin d'apporter une solution à ce problème, une généralisation du "Pareto quantile plot" a été proposée et a donné lieu à l'estimateur UH présenté ci-dessous.

2.3.4 Estimateur basé sur le quantile plot généralisé (UH)

L'estimateur UH proposé par Beirlant et al. [2] est basé sur une extension du "Pareto quantile plot" en un "quantile plot généralisé" qui peut être décrit comme suit.

On considère la fonction UH définie par :

$$UH(x) := U(x)H(x)$$

où U est présentée dans la définition 1.3,
et

$$H(x) = E(\log X - \log U(x)/X > U(x)).$$

L'estimateur empirique de cette fonction évalué en $x = n/k$ est le suivant :

$$UH_{k,n} = X_{n-k,n}\left(k^{-1}\sum_{i=1}^{k}\log X_{n-i+1,n} - \log X_{n-k,n}\right) = X_{n-k,n}\hat{\gamma}_n^{(H)}.$$

En remarquant que la fonction UH est à variation régulière à l'infini, d'indice γ i.e. $UH(x) = x^{\gamma}l(x)$, il est naturel de considérer le "quantile plot généralisé". Beirlant et al. [2] ont alors proposé l'estimateur suivant pour $\gamma \in \mathbb{R}$ basé sur une approximation de la pente de ce "quantile plot généralisé" :

$$\hat{\gamma}_{k,n}^{(UH)} = \frac{1}{k}\sum_{i=1}^{k}\log UH_{i,n} - \log UH_{k+1,n}.$$

Les propriétés de cet estimateur ont été par ailleurs établies dans ce même article.

Tous les estimateurs actuels de l'IVE (par exemple l'estimateur de Hill) reposent sur les plus grandes observations de l'échantillon. Ils supposent en effet que seules ces grandes observations contiennent de l'information sur la queue de

la distribution. L'inconvénient de telles méthodes est une grande sensibilité au nombre d'observations retenues.

En statistique, l'estimation par noyau (ou encore méthode de Parzen-Rozenblatt) est une méthode non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire.

2.3.5 Estimateur de type noyau

En 1985, Csörgö, Deheuvels et Mason [9] ont proposé une nouvelle classe d'estimateurs du noyau de l'indice de queue, dans lequel l'estimateur de Hill était un cas particulier où $K = 1_{[0,1]}$ (le noyau uniforme) Cette classe d'estimateurs (CDM) ne peuvent être utilisés seulement dans le cas où $\gamma > 0$. Sous la condition de von Mises, les estimateurs du noyau (CDM) ont été généralisés par Groeneboon, Lopuhaä et de Wolf (GLW) en (2003) pour tous les indices de queue réels.

On définit d'abord la fonction noyau :

Soit $K(.)$ une fonction noyau vérifiant les conditions suivantes :

(CK1) $K(x) = 0$, si $x \notin [0, 1[$ et $K(x) \geq 0$, si $x \in [0, 1[$.

(CK2) K est deux fois continûment différentiable sur $]0, 1]$.

(CK3) $K(1) = K'(1) = 0$.

(CK4) $\int_0^1 K(x)dx = 1$.

Sélectionner $\alpha > 0$ et $h > 0$, soit

$$\hat{q}_{n,h}^{(1)} := \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^{\alpha} K_h\left(\frac{i}{n}\right) \log(X_{n-i+1,n}/X_{n-i,n}),$$

où $K_h(u) := h^{-1}K(u/h),\ u \in]0, 1]$, et $h > 0$ est appelé paramètre de lissage ou fenêtre (*Bandwidth* en Anglais).

Dans l'estimateur de type noyau le paramètre de lissage h joue un rôle semblable comme nombre de statistiques d'ordre k dans les estimateurs mentionnés dans la Section 2.3.

Pour une telle fonction K, comme (CK2) satisfait, nous avons mis

$$\hat{q}_{n,h}^{(2)} := \sum_{i=1}^{n-1} \frac{d}{du} \left[u^{\alpha+1} K_h(u) \right]_{u=i/n} \left(\log(X_{n-i+1,n}/X_{n-i,n}) \right),$$

L'estimateur $\hat{\gamma}_{n,h}^{(GLW)}$ est défini comme suit pour $\gamma \in \mathbb{R}$:

$$\hat{\gamma}_{n,h}^{(GLW)} := \hat{\gamma}_{n,h}^{(CDM)} - 1 + \frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}},$$

où

$$\hat{\gamma}_{n,h}^{(CDM)} := \sum_{i=1}^{n-1} \frac{i}{n} K_h \left(\frac{i}{n} \right) \log(X_{n-i+1,n}/X_{n-i,n}).$$

est l'estimateur à noyau CDM pour $\gamma > 0$.

Les propriétés (la consistance et la normalité asymptotique) de cet estimateur $\hat{\gamma}_n^{(GLW)}$ont été par ailleurs établies dans [25]. Plus récemment, Necir [34] a proposé une loi fonctionnelle du logarithme itéré pour cet estimateur et a prouvé sa consistance forte.

Il parait naturel, pour estimer l'IVE, d'utiliser uniquement l'information apportée par les "grandes observations". En effet, l'IVE est un paramétre intervenant sur la forme de la queue de distribution. Ainsi, tous les estimateurs cités jusqu'à présent utilisent les k plus grandes observations d'un échantillon. Une autre méthode pour définir les "grandes observations" consiste à prendre les observations qui dépassent un seuil déterministe u. Cette méthode appelée méthode POT pour estimer l'IVE.

2.4 *Modèle POT*

Une deuxième méthode d'estimation la queue de distribution est la méthode des excès encore appelé POT (*Peaks Over Threshold* "pic au dessus d'un seuil") développée dans les année soixante-dix en hydrologie puis abondamment étudiée en statistique. Cette méthode consiste à utiliser les observations qui dépassent un certain seuil déterministe et plus particulièrement les différences entre ces observations et le seuil, appelées excès.

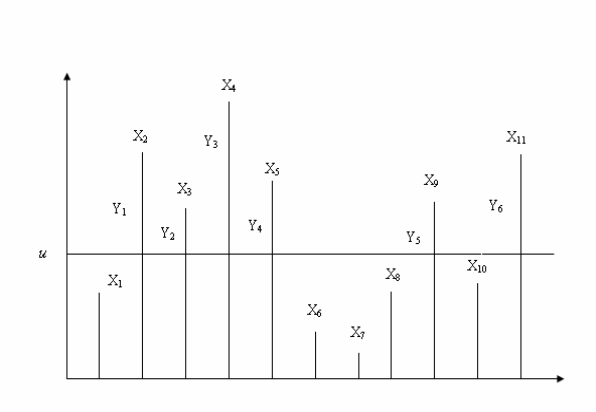


FIG. 2.8 – Observations $X_1, ..., X_{11}$ et excès $Y_1, ..., Y_6$ au-delà du seuil u.

La méthode des excès s'appuie sur une approximation de la loi des excès au dessus du seuil u de la variable aléatoire réelle X, i.e. de la loi conditionnelle de la variable aléatoire réelle positive $X - u$ sachant $X > u$.

2.4.1 Loi des excès

Soit X une variable aléatoire de fonction de répartition F et u un réel suffisamment grand et inférieur au point terminal[1] appelé seuil.

On définit les excès au-delà du seuil u comme l'ensemble des variables aléatoires Y tel que : $y_i = x_i - u, \quad x_i > u$. On cherche à partir de la distribution F de X à définir une distribution conditionnelle F_u par rapport au seuil u pour les variables aléatoires dépassant ce seuil.

Définition 2.5 (Fonction de distribution des excès) *On définit la distribution des excès F_u au dessus du seuil u par :*

$$F_u(y) = P(X - u \leq y / X > u) = \frac{F(y+u) - F(u)}{1 - F(u)} \quad \text{pour } 0 < y < x_F - u.$$

La fonction de distribution des excès représente la probabilité qu'une certaine perte dépasse le seuil u d'au plus une quantité y, sachant qu'elle dépasse u.

[1] *On appelle right-end point ou point terminal de la fonction de répartition F , le point x_F tel que : $x_F = \sup\{x : \ F(x) < 1\}$.*

L'objectif de la méthode POT est de déterminer par quelle loi de probabilité on peut approcher cette distribution conditionnelle. Balkema & de Haan (1974), Pickands (1975), ont proposé le théorème ci-après.
Ce Théorème est très utile lorsque on travaille avec des observations qui dépassent un seuil fixé puisqu'il assure que la loi des excès peut-être approchée par une loi de Pareto généralisée.

2.4.2 Théorème de Balkema-de Haan-Pickands

Le théorème énonce que si F appartient à l'un des trois domaines d'attraction de la loi limite des extrêmes (Fréchet, Gumbel ou Weibull), alors il existe une fonction de répartition des excès au-delà de u, noté F_u qui peut être approchée par une loi de Pareto généralisée (GPD) telle que :

$$\lim_{u \to x_F^+} \sup_{0<y<x_F-u} \left| F_u(y) - G_{\gamma,\sigma(u)}(y) \right| = 0. \tag{2.7}$$

où σ est une fonction strictement positive.

Cette considération théorique suggère que, lorsque nous avons des données issues d'une distribution inconnue, il est possible d'approximer la distribution au-delà d'un certain seuil (assez grand) par une distribution de Pareto généralisée.

Motivation de la Pareto généralisée

La distribution de Pareto généralisée peut être utilisée pour modéliser les observations qui dépassent un seuil u. L'avantage principal de l'utilisation de la Pareto généralisée par rapport aux distributions de valeurs extrêmes est que plusieurs observations qui dépassent le seuil entrent dans la modélisation plutôt qu'une seule observation par période, le maximum ou le minimum. Par contre, il faut maintenant déterminer un seuil adéquat (ce qui parallèle le problème de déterminer k pour les estimateurs de la distribution de valeurs extrêmes généralisée).

2.4.3 Stabilité du seuil

Une variable aléatoire de loi GPD est stable par rapport au seuil i.e., soit $u > 0$, alors pour $x > 0$, on a $P(X - u \leq x/X > u) = G_{\gamma,0,\sigma+\gamma(u-\mu)}(x)$ sera noté $G_{\gamma,\sigma(u)}(x)$.

- Si $\mu = 0$ alors $\sigma(u) = \sigma + \gamma u$.
- Si $\mu \neq 0$ alors $\sigma(u) = \sigma + \gamma(u - \mu)$.

Avant de pouvoir estimer le modèle, il faut construire un échantillon adéquat. Il nous faut trouver un seuil u de sélection des données extrêmes suffisamment élevé pour que l'approximation asymptotique du modèle (2.7), soit applicable.

2.4.4 Détermination du seuil

Le seuil doit être choisi de façon à faire un compromis : plus le seuil est élevé, plus de biais de l'estimateur est réduit ce qui donne un meilleur modèle, par ailleurs, plus le seuil est bas plus la variance de l'estimateur est réduit car plus de données participent à l'estimation. Voici quelques unes des méthodes proposées dans la littérature.

Études diagnostiques

Davison et Smith [13] mentionnent que certains résultats théoriques existent mais ne sont pas applicables en pratique, voir [37]. Ils déterminent le seuil plutôt à l'aide d'études diagnostiques afin de déterminer si le modèle est adéquat pour les données.

Méthode graphique

Davison et Smith [13] proposent également une méthode graphique d'estimation du seuil. Si $\gamma > 1$, $u > 0$ et $\sigma + \gamma u > 0$, alors la moyenne excédentaire est

$$e(u) = E[Y - u | Y > u] = \frac{\sigma + \gamma u}{1 - \gamma}.$$

Si l'hypothèse Pareto généralisée est appropriée, le graphique de la moyenne excédentaire observée en fonction du seuil donne une droite dont l'ordonnée à l'origine est $\sigma/(1-\gamma)$ et la pente est $\gamma/(1-\gamma)$.

Seuil aléatoire

McNeil et Frey choisissent un seuil aléatoire : $N_u = k$ (Il y a k observations excédentaires) où $k \ll N$ et le seuil est donc la $(k+1)^{ième}$ statistique d'ordre.

Soient, $z_{1,n} \leq z_{2,n} \leq ... \leq z_{n,n}$, les excès ordonnés. La loi Pareto généralisée de paramètres γ et σ est ajustée aux données :

$$(z_{1,n} - z_{k+1,n}, ..., z_{k,n} - z_{k+1,n}) .$$

Ces auteurs font une étude simulatoire à partir d'une distribution t de Student pour déterminer une valeur de k appropriée.

Bootstrap

Caers, Beirlant et Maes utilisent une méthode basée sur le bootstrap semi-paramétrique afin de déterminer l'erreur moyenne carrée (MSE). Ceci permet de choisir le seuil qui donne la plus petite MSE. La méthodologie est la suivante :

1. Un seuil u est fixé pour lequel les paramètres de la Pareto généralisée, sont estimés par $\hat{\gamma}_u$, $\hat{\sigma}_u$ avec la méthode de notre choix (par exemple, avec le maximum de vraisemblance). Ceci nous permet d'utiliser la densité semi-paramétrique suivante :

$$\hat{F}_s(x\backslash u) = \begin{cases} (1 - \hat{F}(u)) \left(1 - \left(1 + \dfrac{\hat{\gamma}_u}{\hat{\sigma}_u}(x-u) \right)^{-1/\hat{\gamma}_u} \right) + \hat{F}(u), & x > u, \\ \hat{F}(x), & x \leq u. \end{cases}$$

2. Cette densité estimée permet de faire du bootstrap semi-paramétrique. Donc B échantillons sont tirés à partir de $\hat{F}_s(x\backslash u)$:

$$X^{(b)} = \{X_1^{(b)}, ..., X_n^{(b)}\},$$

où $b = 1, ..., B$.

3. À partir de ces échantillons, un paramètre (par exemple l'indice de queue γ) est estimé. De ces B estimés, le biais et la variance de l'estimateur sont calculés et combinés pour former un estimé de la MSE.
4. Le seuil ayant donné la plus petite MSE pour l'estimation du paramètre qui nous intéresse est retenu.

Il n'est pas clair que cette méthode produise un estimé raisonnable du biais du choix de seuil u, le biais correspondrait plutôt à la méthode d'estimation du paramètre d'intérêt (par exemple γ) pour un seuil fixé.

L'estimation de γ et σ pose le problème de la détermination du seuil u. Il doit être suffisamment grand pour que l'on puisse appliquer le résultat précèdent, mais ne doit pas être trop grand afin d'avoir suffisamment de données pour obtenir des estimateurs de bonne qualité.

2.4.5 Estimation des paramètres de la GPD

Deux méthodes d'estimation sont ici encore réalisables : l'Estimation par Maximum de Vraisemblance (EMV) et celle par les Moments Pondérés (EMP).

Méthode du Maximum de Vraisemblance (EMV)

Supposons que notre échantillon des excès $X = (X_1, ..., X_{N_u})$ est i.i.d avec comme la fonction de distribution la GPD. La fonction de densité $g_{\gamma,\sigma}$ de GPD $G_{\gamma,\sigma}$est

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

La log-vraisemblance est donc égale à

$$l((\gamma,\sigma)\,;X) = -N_u \ln\sigma - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{N_u}\log\left(1+\frac{\gamma}{\sigma}x_i\right)$$

En dérivant cette expression par rapport aux deux paramètres d'intérêt, nous obtenons un système de deux équations à deux inconnues γ et σ. C'est en résolvant ces équations que nous obtenons les estimateurs du maximum de vraisemblance $\left(\hat{\gamma}_{N_u}, \hat{\sigma}_{N_u}\right)$ (à l'aide de méthodes numériques).

Et pour $\gamma = 0$, nous avons

$$\begin{aligned} g_{0,\sigma}(x) &= \frac{1}{\sigma}\exp\left(-\frac{x}{\sigma}\right) \\ l((0,\sigma)\,;X) &= -N_u\ln\sigma - \frac{1}{\sigma}\sum_{i=1}^{N_u}x_i. \end{aligned}$$

Nous obtenons alors $\hat{\sigma}_{N_u} = \sum_{i=1}^{N_u} x_i/N_u$ qui n'est autre que la moyenne empirique des excès (pour $\gamma = 0$, la GPD est la loi exponentielle).

Méthode des Moments Pondérés (EMP)

Il peut arriver que certains moments n'existent pas, ne soient pas finis. Au lieu de la Méthode des Moments, nous utilisons alors la Méthode des Moments Pondérés. Définissons, avec r l'ordre du moment

$$\omega_r(\gamma,\sigma) := E\left[X\bar{G}_{\gamma,\sigma}(X)\right], \quad r \in \mathbb{N},$$

où $\bar{G}_{\gamma,\sigma} = 1 - G_{\gamma,\sigma}$.

Alors

$$\begin{aligned}\omega_r(\gamma,\sigma) &= \int_{-\infty}^{+\infty} x\bar{G}_{\gamma,\sigma}(x)d\bar{G}_{\gamma,\sigma}(x)\\ &= \int_0^1 \bar{G}^{-1}_{\gamma,\sigma}(y)y^r dy\\ &= \int_0^1 \frac{\sigma}{\gamma}(y^{-\gamma}-1)y^r dy.\end{aligned}$$

Nous obtenons grâce à la dernière formulation et après quelques calculs

$$\sigma = \frac{2\omega_0\omega_1}{\omega_0 - 2\omega_1}.$$

$$\gamma = 2 - \frac{\omega_0}{\omega_0 - 2\omega_1}.$$

Nous avons aussi

$$\hat{\omega}_r(\gamma,\sigma) = \frac{1}{n}\sum_{i=1}^{N_u} X_i\hat{F}^r(X_i),\ r = 0,1.$$

où $\hat{F}$ est la fonction de répartition empirique de l'échantillon $X_1,...,X_{N_u}$. Pour estimer γ et σ, nous remplaçons w_r par $\hat{w}_r$ pour $r = 0,1$.

Hosking et Wallis [30] ont montré que lorsque $0 \leq \gamma \leq 0.4$ et pour des échantillons de taille petite, l'EMP obtient des estimateurs plus précis que l'EMV (avec des écarts-types plus faibles). Néanmoins, cette différence s'atténue avec l'augmentation de la taille de l'échantillon. En outre, Rootzén et Tajvidi [37] révèlent que pour $\gamma \geq 1/2$, l'EMP calcule des estimateurs fortement biaisés contrairement aux estimateurs de l'EMV qui sont efficaces. Enfin, pour $\gamma \geq -1/2$, les conditions de régularité de l'EMV sont remplies et les estimateurs du Maximum de Vraisemblance $\left(\hat{\gamma}_{N_u}, \hat{\sigma}_{N_u}\right)$ calculés sur l'échantillon des N_u excès sont asymptotiquement normaux.

Exemple 2.1 *Fixant le seuil u=4.4 nous utilisons le package POT du logicie R pour ajuster le GPD aux données de réclamation danoises. Les résultats sont résumés dans le tableau 2.2. Notez que le paramètre de la forme obtenu estimation (0.69) est proche de l'estimation de Hill de l'IVE (0.7).*

	# Ci-dessus	Proportion Ci-dessus	para. Forme	para. Echelle
EMV	311	14.35%	0.694	3.044
EMP	3.11	14.35%	0.581	3.330
Pickands	311	14.35%	1.047	2.637

TAB. 2.2 – Résultats de l'ajustement de GPD (par l'EMV, EMP et l'estimateur de Pickands) aux excès plus de seuil 4.4, de l'ensemble des donnée du Feu danois.

Estimation de la queue de la distribution

La distribution de Pareto généralisée est utilisée pour modéliser la queue (supérieure ou inférieure) d'une distribution présentant des valeurs extrêmes. Pour pouvoir faire de l'estimation des quantiles, il faut avoir une formulation qui combine la distribution estimée dans la queue et la distribution centrale.
Une telle formulation est donnée par l'égalité suivante, $\forall u < x < x_F$

$$\bar{F}(x) := \bar{F}(u)\bar{F}_u(x-u), \tag{2.8}$$

où $F_u(y) = P(X - u \leq y/X > u)$.
Autrement dit, $\forall x > u$:

$$P(X > x) = P(X > u)P(X > x/X > u).$$

L'estimation possède de la façon suivante :
$\bar{F}(x)$ est estimée avec la probabilité de dépassement empirique N_u/n.

$$\widehat{\bar{F}}(u) = \bar{F}_n(u) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X>u\}} = N_u/n,\ u < x_F. \tag{2.9}$$

La queue conditionnelle $\bar{F}_u$ de F est estimée par

$$\widehat{\bar{F}}_u(x-u) := 1 - G_{\hat{\gamma},\hat{\sigma}}(x-u) = \left(1 + \hat{\gamma}_u \frac{x-u}{\hat{\sigma}_u}\right)^{-1/\hat{\gamma}_u},\ u < x < x_F. \tag{2.10}$$

L'estimateur de queue de la distribution est donc :

$$\widehat{\bar{F}}(x) := \frac{N_u}{n}\left(1+\hat{\gamma}_u\frac{x-u}{\hat{\sigma}_u}\right)^{-1/\hat{\gamma}_u}, \ u<x<x_F. \tag{2.11}$$

2.5 Estimation des quantiles extrêmes

Estimation des quantiles extrêmes joue un rôle important dans le contexte de la gestion des risques où il est crucial d'évaluer de manière adéquate le risque d'une grande perte qui se produit très rarement.

On considère $X_1, X_2, ..., X_n$ la réalisation de n variables aléatoires réelles indépendantes et de fonction de répartition commune F supposée continue.
Pour $0 < p < 1$, le quantile d'ordre $(1-p)$ de la fonction de distribution F noté x_p est défini comme étant la solution de l'équation :

$$1 - F(x) = p.$$

En utilisant les fonctions présentées dans les définitions 1.2 et 1.3, on définit les quantiles extrêmes par :

$$x_p := F^{\leftarrow}(1-p) = Q(1-p) = U(1/p), \qquad \text{quand } p \to 0.$$

On veut estimer le quantile d'ordre p où p est strictement inférieur à $1/n$. La difficulté principale dans l'estimation de quantile extrême est qu'avec une probabilité qui tend vers 1 lorsque n tend vers l'infini, le nombre x_p est strictement supérieur à l'observation maximale de l'échantillon. On ne peut pas, comme pour l'estimation de quantiles "classiques", inverser tout simplement la fonction de répartition empirique.

2.5.1 Approche EVT

La GEVD H_θ définie à la Section 2.2, est utilisée pour dériver les estimateurs pour les quantiles extrêmes.

• **Cas où F est exactement H_θ**

Cette méthode utilise un résultat donnant l'expression de la loi asymptotique du maximum d'un échantillon (voir Théorème 1.4). On estime alors le quantile

extrême en inversant la fonction de répartition de la loi des valeurs extrêmes $x_p = H_\theta^{\leftarrow}(1-p)$ un estimateur naturel de x_p devient $\hat{x}_p = H_{\hat{\theta}}^{\leftarrow}(1-p)$.
L'estimateur de quantile extrême obtenu par cette méthode s'écrit sous la forme

$$\hat{x}_p := \begin{cases} \hat{\mu} - \dfrac{\hat{\sigma}}{\hat{\gamma}}\left(1 - (-\log(1-p))^{-\hat{\gamma}}\right) & \text{si} \quad \gamma \neq 0, \\ \hat{\mu} - \hat{\sigma}\log(-\log(1-p)) & \text{si} \quad \gamma = 0, \end{cases} \tag{2.12}$$

où $\hat{\mu}, \hat{\sigma}$ et $\hat{\gamma}$ sont des estimateurs des paramètres de la loi des valeurs extrêmes.
Quand $\gamma < 0$, le point terminal est fini. Il peut être estimé par

$$\hat{x}_F := \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}.$$

• **Cas où F dans le domaine de l'attraction de H_θ**

Dans le cas où le quantile d'ordre $(1-p)$ est dans la marge des observations (i.e. $p \geq 1/n$), il peut être estimée comme suite :

Utiliser la Proposition (1.5) sous la forme

$$\ln F(a_n x + b_n) = \frac{1}{n}\ln H_\gamma(x).$$

D'aprés la définition de $H_\gamma(x)$ où $\ln F(z) = \bar{F}(z)$ pour z grand on obtient :

$$F(a_n x + b_n) = 1 - \left(\frac{1}{n}(1+\gamma x)^{-1/\gamma}\right),$$

ce qui implique

$$a_n x + b_n = Q\left(1 - \left(\frac{1}{n}(1+\gamma x)^{-1/\gamma}\right)\right).$$

Si on pose $p = \dfrac{1}{n}(1+\gamma x)^{-1/\gamma}$, on obtient :

$$Q(1-p) = b_n + a_n \frac{(np)^{-\gamma} - 1}{\gamma},$$

ce qui permet d'estimer x_p,

$$\hat{x}_p = \hat{b}_n + \hat{a}_n \frac{(np)^{-\hat{\gamma}} - 1}{\hat{\gamma}}.$$

Pour l'estimateur $\hat{\gamma}$ voir la Section (2.3) et pour un choix convenable des suites $\hat{a}_n$ et $\hat{b}_n$, voir les Sections (6.4.1) et (6.4.3) de [20].

Typiquement nous serons intéressés à estimer le quantile d'ordre $(1-p)$ à l'extérieur de l'échantillon (i.e. $p < 1/n$), nous utilisons un subséquence (n/k), où $k = k_n \to \infty$ et $k/n \to 0$ quand $n \to \infty$.

En supposant que n/k est entier, on obtient

$$\hat{x}_p := \hat{b}_{n/k} + \hat{a}_{n/k}\frac{(np/k)^{-\hat{\gamma}} - 1}{\hat{\gamma}}.$$

Quand $\gamma < 0$, le point terminal est fini. Il peut estimer par

$$\hat{x}_F = \hat{b}_{n/k} - \frac{\hat{a}_{n/k}}{\hat{\gamma}}.$$

Nous définissons maintenant les grandes estimateurs de quantile qui sont associées aux estimateurs semi-paramétriques de l'IVE présenté dans la Section 2.3.

Weissman a donné l'estimation des quantiles extrêmes pour chacune des trois distributions standard des valeurs extrêmes de Fisher-Tippett. Pour la classe de Fréchet ($\gamma > 0$), l'estimateur de quantile d'ordre $(1-p)$ de type Weissman prend la forme suivante :

$$\hat{x}_p^{(W)} := X_{n-k,n}\left(\frac{k}{np}\right)^{\hat{\gamma}_n}, \tag{2.13}$$

- L'estimateur classique de quantile de Weissman pour l'estimateur de Hill $\hat{\gamma}_n^{(H)}$ est défini par :

$$\hat{x}_p^{(H)} := X_{n-k,n}\left(\frac{k}{np}\right)^{\hat{\gamma}_n^{(H)}},$$

- L'estimateur $\hat{x}_p^{(P)}$du quantile d'ordre $(1-p)$ lié au l'estimateur de Pickands $\hat{x}_n^{(P)}$ est de la forme suivante :

$$\hat{x}_p^{(P)} := X_{n-k+1,n} + \frac{(np/k)^{-\hat{\gamma}_n^{(P)}} - 1}{1 - 2^{-\hat{\gamma}_n^{(P)}}}(X_{n-k+1,n} - X_{n-2k+1,n}).$$

Quand $\gamma < 0$, le point terminal est fini. Il peut estimer par

$$\hat{x}_F^{(P)} := X_{n-k+1,n} + \frac{(X_{n-k+1,n} - X_{n-2k+1,n})}{1 - 2^{\hat{\gamma}_n^{(P)}}}.$$

- De la même façon, le quantile d'ordre $(1-p)$ est estimé sur la base de l'estimateur du moment $\hat{x}_p^{(M)}$ par

$$\hat{x}_p^{(M)} := X_{n-k,n} + \left(\frac{X_{n-k,n} M_n^{(1)}}{\varphi\left(\hat{\gamma}_n^{(M)}\right)} \right) \frac{(np/k)^{-\hat{\gamma}_n^{(M)}} - 1}{\hat{\gamma}_n^{(M)}},$$

où

$$\varphi(\gamma) := \begin{cases} 1, & \gamma \geq 0, \\ 1/(1-\gamma), & \gamma > 0. \end{cases}$$

Quand $\gamma < 0$, le point terminal est fini. Il peut estimer par

$$\hat{x}_F^{(M)} := X_{n-k,n} + \left(1 - 1/\hat{\gamma}_n^{(M)}\right) X_{n-k,n} M_n^{(1)}.$$

2.5.2 Approche POT

La méthode POT s'appuie sur le théorème de Balkema-de Haan-Pickands pour estimer x_p et cet estimateur est obtenu en inversant $\widehat{\bar{F}}(x)$ dans l'équation (2.11).
L'estimateur obtenu par cette méthode s'écrit sous la forme :

$$\hat{x}_p := u + \frac{\hat{\sigma}_u}{\hat{\gamma}_u} \left[\left(\frac{N_u}{np} \right)^{\hat{\gamma}} - 1 \right].$$

où N_u désigne le nombre d'excès au-delà du seuil u, $\hat{\sigma}_n$ et $\hat{\gamma}_n$ sont des estimateurs des paramètres de la loi GPD.

Quand $\gamma < 0$, le point terminal est fini et il est estimé par :

$$\hat{x}_F := u - \frac{\hat{\sigma}_u}{\hat{\gamma}_u}.$$

En pratique, u est choisi égal à l'une des statistiques d'ordre et en Prenant $u = X_{n-k,n}$ la $(k+1)^{ième}$ observation, donne $N_u = k$.

Les estimateurs résultants des paramètres γ et σ sont respectivement dénotés par $\hat{\gamma}^{(POT)}$ et $\hat{\sigma}^{(POT)}$. Dans ce cas, l'estimateur des quantiles est de la forme suivante :

$$\hat{x}_p^{(POT)} := X_{n-k,n} + \frac{\hat{\sigma}^{(POT)}}{\hat{\gamma}^{(POT)}} \left(\left(\frac{k}{np} \right)^{\hat{\gamma}^{(POT)}} - 1 \right) \qquad \text{pour } p < \frac{k}{n},$$

le point terminal est estimé par :

$$\hat{x}_F := X_{n-k,n} - \frac{\hat{\sigma}^{(POT)}}{\hat{\gamma}^{(POT)}}.$$

En général, si $F \in D(H_{\gamma,\mu,\sigma})$, les paramètres γ, μ et σ peuvent être estimés par maximum de vraisemblance, ou par la méthode des moments pondérés de façon probabiliste. Lorsque μ et σ sont estimés, les observations peuvent être standardisées et le paramètre γ peut être estimé par une des méthodes mentionnées dans la Section 2.3.

Chapitre 3

Choix du nombre optimal de statistiques d'ordre extrêmes

L'estimation de l'indice des valeurs extrêmes d'une distribution à queue lourde dépend du choix du nombre de statistiques d'ordre extrêmes à utiliser dans l'estimation.

Il est bien connu que la façon de choisir la valeur de nombre de statistiques d'ordre extrêmes (on le note k) est toujours un problème difficile même si la forme estimation a été déterminée, ce problème a été longuement abordé dans la littérature (voir [12], [17], [19], [35], ...). La clef de l'estimation de l'indice de queue est le nombre de statistiques d'ordre extrêmes. Si la valeur de k est obtenue, nous pouvons estimer l'indice de queue.

Sélectionner une bonne valeur de k est une tâche sensible. Lorsque k est petit la variance de l'estimateur est grande et l'utilisation de k introduit un grand biais dans l'estimation. L'équilibrage de ces composants (la variance et le biais) est important dans les applications de la théorie des valeurs extrêmes, parce que cela réduit l'erreur moyenne quadratique.

Dans ce chapitre (où nous nous concentrons surtout sur l'estimateur de Hill $\hat{\gamma}_n^{(H)}$), nous avons introduit certaines des méthodes proposées pour équilibrer entre ces deux vices afin d'obtenir un nombre optimal k_{opt} de statistiques d'ordre qui localise où la queue de distribution (vraiment) commence.

3.1 Méthode Graphique

Le problème de l'estimation de l'indice de queue de distributions à queue lourde est très important dans de nombreuses applications. Nous présentons une méthode graphique universelle qui devrait être appliquée avant toutes les recherches numériques. Elle traite ce problème en sélectionnant un certain nombre de statistiques d'ordre extrêmes.

Cette méthode consiste à dessiner les points de coordonnées

$$\{(k,\ \hat{\gamma}_n(k)) : k = 1, \ldots, n\},$$

pour faire un choix optimal de k.
Où $\hat{\gamma}_n(k)$ désigne n'importe quel estimateur introduit dans le chapitre deux.

Sélectionner une bonne valeur pour k est une tâche sensible et cette valeur devrait être prise, où le graphe est stable, i.e. k pas trop petit donc nous avons de grandes fluctuations en raison d'un trop petit nombre de points de données (l'estimateur a une variance forte), mais aussi k pas trop grand si l'estimation est basée sur les points de l'échantillon du centre de la distribution (ce qui introduit un biais). Ceci est illustré graphiquement dans la Figure 3.1.

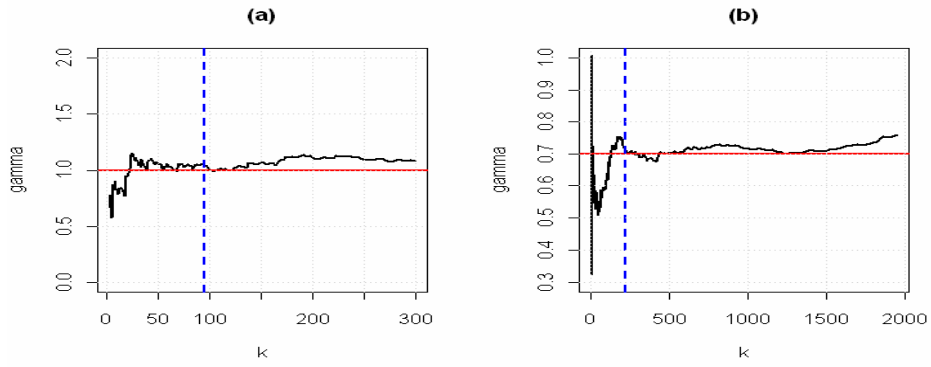


FIG. 3.1 – Estimateur de Hill de l'IVE pour (a) distribution Pareto standard basée sur 300 échantillons de taille 3000 et (b) Danish Fire. La ligne horizontale correspond à la valeur estimée de index de queue et la ligne verticale correspond au nombre optimal.

Dans la Figure 3.1, il semble stable un k autour de 80 de (a) distribution Pareto standard et (b) les données de réclamations danoises on trouve k autour de 220. Nous vous suggérons qu'à partir de l'estimateur de Hill nous obtenons $\hat{\gamma}^{(H)}_{97,3000} = 0.9874$ pour (a) et $\hat{\gamma}^{(H)}_{220,2167} = 0.6899$ pour (b).

Nous présentons une nouvelle procédure graphique, appelée sum-plot, proposée par Sousa [40].

3.1.1 Méthode de sum-plot

La méthode de sum-plot est proposée par Sousa, dans sa thèse de doctorat en 2002. Il détermine la valeur de k par le dessin.
Sa base théorique est que le graphe $\{(k, S_k), 1 \leq k < n\}$ devrait être une ligne. Ensuite, nous avons placé la valeur de k dans l'estimateur de Hill's pour calculer l'indice de queue. Sousa est arrivé à la conclusion que, peu importe que l'indice de queue soit $0 < \gamma < 1/2$ ou $\gamma^{-1} \geq 1/2$, la méthode est supérieure aux autres méthodes par simulation aux différents échantillons et différentes distributions.

Soit la variable aléatoire

$$S_k = \sum_{i=1}^{k} iV_i = \sum_{i=1}^{k} i(\log X_{i,n} - \log X_{k+1,n}). \tag{3.1}$$

Où $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{k+1,n}$ sont les statistiques d'ordre correspondantes.

Si nous choisissons k tel que $X_{k+1,n}$ est assez grand et peut satisfaire la condition suivante

$$\bar{F}(x) = 1 - F(x) = x^{-1/\gamma} l(x), \tag{3.2}$$

où l est une fonction à variations lentes, pour tout $x > X_{k+1,n}$ nous avons $S_k \sim \gamma k$. La formulation approximative a prouvé que la pente de la ligne est γ dans le graphe. Et Sousa a prouvé que la valeur de γ peut être estimée par le modèle de régression linéaire.

$$S_i = \beta_0 + \beta_1 i + \varepsilon_i, \qquad i = 1, ..., k \tag{3.3}$$

Il est facile de voir que la valeur estimant de $\hat{\gamma}$ est la pente $\hat{\beta}_1$ du modèle de régression par l'estimation des moindres carrés généralisés :

$$\hat{\gamma} = \hat{\beta}_1 = \frac{k}{k+1}\hat{\gamma}_n^{(H)} - \frac{k}{k-1}\log X_{1,n}.$$

Si $\beta_0 = 0$, alors $\hat{\gamma} = \hat{\gamma}_n^{(H)}$.

La méthode proposée prouvé être efficace pour les échantillons finis venant d'une distribution de Pareto et un l'inverse de gamma inversé. Pour plus des détails voir [40].

La méthode graphique est un outil adéquat pour choisir k. Toutefois, cette technique sera toujours subjective et il y a des cas où nous avons besoin d'une solution plus objective et d'un choix rapide, automatique, net de k. Ainsi, pour des raisons de perfection, nous présentons quelques méthodes pour le choix de k dans l'estimation de l'indice des valeurs extrêmes.

Nous voulons étudier le problème de savoir comment choisir le nombre k impliqué dans le calcul de l'estimateur. Il est raisonnable que nous nous basons notre choix en minimisant l'erreur moyenne quadratique.

3.2 Erreur moyenne quadratique

L'erreur moyenne quadratique est très utile pour comparer plusieurs estimateurs, notamment lorsque l'un d'eux est biaisé. Si les deux estimateurs à comparer sont sans biais, l'estimateur le plus efficace est simplement celui qui a la variance la plus petite. On peut effectivement exprimer l'erreur moyenne quadratique en fonction du biais de l'estimateur $E(\hat{\gamma}_n - \gamma)$ ainsi que sa variance :

$$MSE := Biais(\hat{\gamma}_n)^2 + var(\hat{\gamma}_n).$$

Le MSE d'un estimateur de l'indice de queue $\hat{\gamma}_n$ est défini par

$$MSE(\hat{\gamma}_n) := E_\infty(\hat{\gamma}_n - \gamma)^2,$$

où E_∞ dénote l'espérance en ce qui concerne la distribution de la limite. On le vérifie facilement que $MSE(\hat{\gamma}_n)$, qui est réellement une fonction de k.

Pour une estimation exacte de l'indice de queue, il est nécessaire pour n'importe quel estimateur classique à faire un compromis entre le biais et la variance. Il

semble raisonnable que le MSE permet de réduire au minimum un compromis entre le biais et la variance rendent l'estimation la plus précise possible. C'est le choix optimal de k, noté k_{opt}, correspond à la plus petite MSE, i.e.

$$k_{opt} := \arg\min_{k} MSE(\hat{\gamma}_n).$$

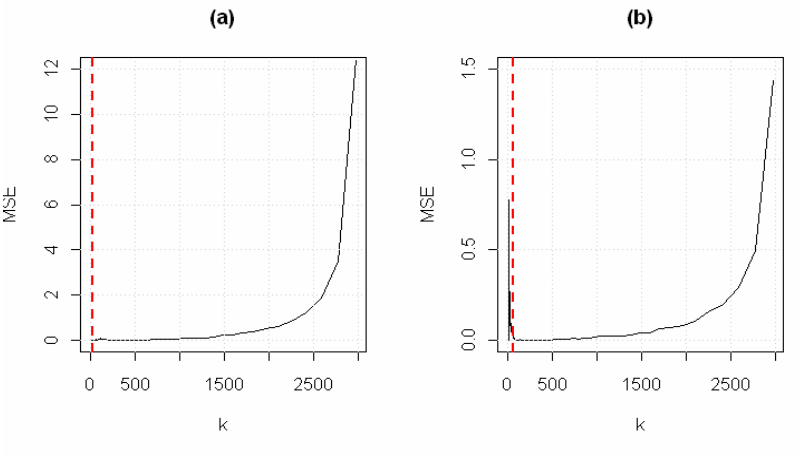


FIG. 3.2 – MSE de l'estimateur de Hill pour l'IVE de (a) distribution de Pareto standard et (b) la distribution de Fréchet(1), basée sur 300 échantillons de 3000 observations. La ligne tirée verticale correspond au minimum du MSE.

Dans la Figure 3.2, Le minimum du MSE est réalisé pour les statistiques d'ordre supérieures k vaut 76, représente presque 2.5% du nombre total d'observations pour (a) et pour (b) k égale 68 représente presque 2.3% du nombre total d'observations. Les résultats sont résumés dans le tableau 3.1.

Algorithme MSE	#des extrêmes	% des extrêmes	Estimateur/Vraie IVE
Fréchet(1)	76	2.5%	1.051/1
Pareto standard	68	2.3%	1.017/1

TAB. 3.1 – Nombres optimaux de statistiques d'ordre supérieurs obtenus par la minimisation de MSE et utilisés dans le calcul de l'estimateur de Hill de l'IVE de la distribution de Fréchet(1) et de la distribution de Pareto standard, basé sur 3000 observations.

Le k *optimal devrait être égal au MSE minimal. Théoriquement, liée à* k. k *croissant, la variance sera diminuée et le biais sera augmenté ;* k *diminue, la variance sera augmentée et le biais sera diminué.*

3.3 Procédures adaptatives

En présence d'un échantillon aléatoire de taille finie, le problème concernant le choix du nombre des extrêmes supérieures n'est pas facile à manipuler.

Le processus de choisir le k optimal est difficile par le fait que le dernier ne dépend pas exclusivement de la taille de l'échantillon et de l'index des valeurs extrêmes mais il dépend également des paramètres inconnus caractérisant F (le paramètre de deuxième ordre ρ parmi autres du fonction de distribution F). Pour surmonter cet obstacle, une grande variété d'algorithmes et de procédures adaptatives ont proposé de calculer $\hat{k}_{opt}$ pour k_{opt} dans le sens

$$\frac{\hat{k}_{opt}}{k_{opt}} \xrightarrow{p} 1 \quad \text{quand} \quad n \to \infty. \tag{3.4}$$

L'estimation correspondante $\hat{\gamma}_n(\hat{k}_{opt})$ est aussi asymptotiquement efficace que $\hat{\gamma}_n(k_{opt})$. Nous décrirons certaines des méthodes les plus connues de choisir le nombre de plus grandes statistiques convenable pour une estimation exacte (précise).

3.3.1 Approche de Hall et Welsh

La performance de l'estimateur de Hill et les autres estimateurs sont fondés essentiellement sur le choix de la fraction de l'échantillon utilisé pour l'estimation. Si k est trop grand, alors les estimateurs ont un grand biais, tandis que d'autre part, leur variance est grande si k est petit. Hall et le Welsh (1985) ont prouvé que l'erreur moyenne quadratique asymptotique de l'estimateur de Hill est minimale pour

$$k_{opt} \sim \left(\frac{c^{2\rho} \left(\rho+1\right)^2}{2d^2\rho^3} \right)^{1/(2\rho+1)} n^{2\rho/(2\rho+1)}. \tag{3.5}$$

Si la fonction de répartition F satisfait la classe de Hall voir la relation (2.5).

Comme les paramètres $\rho > 0,\ c > 0$ et $d \neq 0$ sont inconnus, ce résultat ne peut pas être utilisé directement pour déterminer le nombre optimal de statistiques d'ordre pour un ensemble des donnée.

Hall et Welsh [28] ont construit une estimation cohérente pour choisir k_{opt} avec

$$\hat{\rho} := \left| \log \left| \frac{\left(\hat{\gamma}_n^{(H)}(t_1)\right)^{-1} - \left(\hat{\gamma}_n^{(H)}(s)\right)^{-1}}{\left(\hat{\gamma}_n^{(H)}(t_2)\right)^{-1} - \left(\hat{\gamma}_n^{(H)}(s)\right)^{-1}} \right| / \log \frac{t_1}{t_2} \right|$$

et

$$\hat{\lambda}_0 := \left| (2\hat{\rho})^{-1/2} \left(\frac{n}{t_1}\right)^{\hat{\rho}} \frac{\left(\hat{\gamma}_n^{(H)}(t_1)\right)^{-1} - \left(\hat{\gamma}_n^{(H)}(s)\right)^{-1}}{\left(\hat{\gamma}_n^{(H)}(s)\right)^{-1}} \right|^{2/(2\hat{\rho}+1)}.$$

Alors

$$\hat{k}_{opt} := \left[\hat{\lambda}_0 n^{2\hat{\rho}/(2\hat{\rho}+1)} \right],$$

est un estimateur consistent de k_{opt} dans le sens que $\hat{k}_{opt}/k_{opt} \to 1$ en probabilité si $t_i = [n^{\tau_i}],\ i = 1, 2$ et $s = [n^{\sigma}]$ pour certains $0 < 2\rho(1-\tau_1) < \sigma < 2\rho/(2\rho+1) < \tau_1 < \tau_2 < 1$. Toutefois, noter que, pour un choix de $\sigma < \tau_1 < \tau_2$, on doit limiter le paramètre du deuxième ordre ρ à l'intervalle $[\sigma/(2(1-\sigma)), \sigma/(2(1-\tau_1))]$, i.e. la consistance satisfait simplement dans un sous-modèle de (2.5). Dans ce cas $\hat{\gamma}_{n,\hat{k}_{opt}}$ est aussi asymptotiquement efficace comme $\hat{\gamma}_{n,k_{opt}}$.

L'erreur moyenne quadratique MSE est liée à l'indice de queue de distribution inconnue γ et le paramètre du second ordre ρ. Donc Il ne peut pas être utilisé dans les questions pratiques. Pour cette raison, nous introduisons la méthode bootstrap de Danielsson et al. [12]. Cette méthode ne lie pas à γ et ρ. Il a modifié la méthode bootstrap qui a proposé par Hall et a soulevé la méthode M-bootstrap en même temps.

3.3.2 Approche de Bootstrap

La technique du ré-échantillonnage (*Bootstrap*) à été introduit par B. Efron (1979). C'est la méthode de réplication des échantillons la mieux fondée théoriquement. Elle consiste à créer, à partir d'un échantillon de base, un grand nombre d'échantillon par tirage aléatoire avec remise. Sur chaque échantillon, les statistiques auxquelles on s'intéresse sont calculées, ce qui permet d'approcher leur dispersion. On peut ainsi estimer la variance ou la loi des paramètres caractéristiques de la distribution de l'échantillon et construire des intervalles de confiance.

L'estimation de l'index de queue dépend pour sa précision sur un choix précis de la fraction de l'échantillon , i.e. le nombre de statistiques d'ordre extrêmes sur lesquelles l'estimation est basée. Une solution complète à la sélection de la fraction de l'échantillon est donnée au moyen de deux étapes sous la méthode bootstrap. Cette approche détermine de manière adaptative la fraction de l'échantillon qui minimise le l'erreur moyenne quadratique asymptotique.

Méthode du Bootstrap proposée par Hall

En 1990, Hall a utilisé la méthode du bootstrap dans l'estimation de l'indice de queue pour le nombre de statistiques d'ordre extrêmes. Son fondement théorique est minimiser MSE de l'indice de valeur extrême, voir la Section 3.2.

Nous avons prouvé que quand $k \to \infty,\ k/n \to 0,\ \hat{\gamma}_n^{(H)}$ converge en probabilité vers γ et sa distribution limite est

$$\sqrt{k_{opt}}(\hat{\gamma}_n^{(H)}(k_{opt}) - \gamma) \xrightarrow{d} \mathcal{N}(b, \gamma^2).$$

En fait, la valeur de k_{opt} a équilibré la variance asymptotique et le biais de $E\left(\hat{\gamma}_n^{(H)}(k_{opt}) - \gamma\right)^2$. Quand γ et ρ sont connus, nous pouvons calculer la valeur de k_{opt}.

Notre cadre est une condition de second ordre liée à (3.2). Là existe une fonction A^*satisfait $\lim_{t\to\infty} A^*(t) = 0$ et ne change pas son signe à l'infini ($\alpha = 1/\gamma,\ \rho = \alpha\beta$), tel que

$$\lim_{t\to\infty} \left(\left(\frac{1 - F(x)}{1 - F(t)} - x^{-\alpha} \right) / A^*(t) \right) = x^{-\alpha} \frac{x^{\alpha\beta} - 1}{\alpha\beta}. \tag{3.6}$$

Notre but est de déterminer l'ordre optimal k_{opt} seulement sur la base de l'échantillon i.e. pour déterminer un estimateur $\hat{k}_{opt}$ tel que

$$\sqrt{\hat{k}_{opt}}(\hat{\gamma}_n^{(H)}(\hat{k}_{opt}) - \gamma) \xrightarrow{d} \mathcal{N}(b, \gamma^2).$$

Pour ceci c'est suffisant pour prouver

$$\frac{\hat{k}_{opt}}{k_{opt}} \xrightarrow{p} 1.$$

Hall[1] a précisé la taille n_1 de l'échantillon bootstrap qui doit être inférieure à la taille de l'échantillon total n. Et il avait prouvé que l'estimation n'était pas fiable quand $n_1 = n$.
Le biais est presque zéro sous certaine situation. Par exemple, quand les données présentent linéaire, le biais d'estimation est nulle.

Donc nous avons besoin de retirer un sous-échantillon $\mathcal{X}^*_{n_1} = \{X^*_1, ..., X^*_{n1}\}$ du l'échantillon total $\mathcal{X}_n = \{X_1, ..., X_n\}$ ($n_1 \ll n$) et $\mathcal{X}^*_{n_1}$ est appelé sous-échantillon bootstrap. Nous avons utilisé $X_{1,n_1} \leq X_{2,n_1} \leq \ldots \leq X_{n_1,n_1}$ les statistiques d'ordre de $\mathcal{X}^*_{n_1}$ et nous avons défini :

$$\gamma^*_{n_1}(k_1) := \frac{1}{k_1}\sum_{i=1}^{k_1} \log X^*_{n_1-i+1,n_1} - \log X^*_{n_1-k_1,n_1}.$$

Nous déterminons k et k_1 en minimisant

$$MSE\,(n_1, k_1) = E(\big(\gamma^*_{n_1}(k_1) - \gamma_n(k)\big)^2 \setminus \mathcal{X}_n). \tag{3.7}$$

Et la relation est $k = k_1 \left(\frac{n}{n_1}\right)^{\mu}$. En fait, Hall avait supposé la relation de pouvoir de k et n étais $k = cn^{\mu}$, $(0 < \mu < 1)$ et le rapport de pouvoir n_1 et n étais $n_1 = n^{\beta}$. Il a suggéré que μ et β étais $1/2$.

Cette méthode dépend de γ et ρ et MSE est liée à k.

Méthode du Bootstrap proposée par Danielsson et al.

Danielsson a amélioré la méthode bootstrap qui est proposée par Hall. Danielsson a proposé une nouvelle méthode que le statistique $\hat{\gamma}^{(H)}_n(k)$ a été remplacée par une nouvelle statistique $M_n(k)$, où

$$M_n(k) = \frac{1}{k}\sum_{i=1}^{k}(\log X_{i,n} - \log X_{k+1,n})^2. \tag{3.8}$$

Nous avons prouvé $M_n(k)/2\gamma_n(k)$ converge en probabilité vers γ quand $k \to \infty$. Il a aussi équilibré la variance et le biais.

Les statistiques $M_n(k)/2\gamma_n(k) - \gamma_n(k)$ et $\gamma_n(k) - \gamma$ ont la même propriété asymptotique que leur moyenne asymptotique est 0. Sous quelques situations, il peut recevoir la même valeur de k pour minimiser MSE et $E_\infty(M_n(k) - 2(\gamma_n(k))^2)^2$.

[1] *Hall (1990) utilise également la même idée de choisir le paramètre de lissage dans les procédures d'estimation de type noyau.*

Selon le sous-échantillon bootstrap $\mathcal{X}_{n_1}^*$, Nous sélectionnons les statistiques pour déterminer k_1 en minimisant $Q(n_1, k_1)$.

$$Q(n_1, k_1) = E((M_{n_1}^*(k_1) - 2(\gamma_{n_1}^*(k_1))^2)^2 \backslash \mathcal{X}_n). \tag{3.9}$$

Afin de déterminer k nous avons besoin d'un autre sous-échantillon bootstrap $\mathcal{X}_{n_2}^*$ où $(n_2 = n_1^2/n)$ pour déterminer k_2 par la même méthode de détermination k_1. Ensuite, nous pouvons calculer k avec le rapport de k_1 et k_2.

Nous expliquons, étape par étape, comment mettre en œuvre la procédure :

Etape 1 : Sélectionner un échantillon de taille n à partir d'une série de rapport de retour comme échantillon naissant et arranger par convenable.

Etape 2 : Choisir d'abord n_1 et k_1. Echantillons aléatoires de taille n_1, m fois de la taille naissante d'échantillon n_1 $(n_1 = O(n^{1-\varepsilon}),\ 0 \leq \varepsilon \leq 1)$ et déterminer n_1 et k_1 en réduisant au minimum avec la méthode du bootstrap

$$Q(n_1, k_1) = E((M_{n_1}(k_1) - 2(\gamma_{n_1}(k_1))^2)^2 \backslash \mathcal{X}_n) = \frac{1}{m} \sum_{i=1}^{m} (M_{n_1}^*(k_1) - 2(\gamma_{n_1}^*(k_1))^2)^2.$$

Etape 3 : Sélectionner k_2. Nous supposons $n_2 = n_1^2/n$ et répéter l'étape (2) pour déterminer k_2.

Etape 4 : Déterminer $\hat{k}$. Nous pouvons calculer $\hat{k}$ avec

$$\hat{k} = \frac{k_1^2}{k_2} \left(\frac{(\log k_1)^2}{(2 \log n_1 - \log k_1)^2} \right)^{(\log n_1 - \log k_1(n_1))/\log(n_1)}. \tag{3.10}$$

Donc on peut calculer l'indice de queue avec l'estimateur de Hill.

Méthode M-bootstrap

Nous avons proposé une nouvelle méthode M-bootstrap (éclairant le bootstrap proposé par Danielsson) que γ a été remplacé par $\tilde{\gamma}_n(k)$ dans (3.7) et on obtient :

$$MSE_M\,(n_1, k_1) = E(\left(\gamma_{n_1}^*(k_1) - \tilde{\gamma}_n(k)\right)^2 \backslash \mathcal{X}_n). \tag{3.11}$$

Selon le sous-échantillon bootstrap $\mathcal{X}_{n_1}^*$, nous pouvons déterminer k_1 en minimisant $MSE_M\,(n_1, k_1)$.

Théorème 3.1 *Supposons (3.6) satisfait et $k \to \infty$, $k/n \to 0$. Déterminons $k(n)$ tel que $MSE(n,k)$ est minimal. Alors*

$$k = \frac{n}{s^-\left(\frac{\gamma^2(1-\rho)^2}{n}\right)}(1+o(1)), \qquad \text{quand } n \to \infty,$$

où s^- est la fonction inverse de s, avec s donnée par

$$A^2(t) = \int_t^\infty s(u)du(1+o(1)) \qquad \text{quand } t \to \infty. \tag{3.12}$$

Supposons $A(t) = ct^\rho,\ \rho < 0$ alors

$$k = H(\rho)n^\mu(1+o(1)), \quad \mu = \frac{2\rho}{2\rho-1}. \tag{3.13}$$

Théorème 3.2 *Supposons (3.6) est satisfait et $k \to \infty$, $k/n \to 0$. Supposons $A(t) = ct^\rho,\ c \neq 0,\ \rho < 0,\ n_1 = O(n^{1-\varepsilon}),\ (0 < \varepsilon < 1)$. Nous déterminons k_1 tel que $\widehat{MSE}_n(n_1,k_1)$ est minimal. Alors*

$$k_1 = H(\rho)n_1^\mu(1+o(1)),\ \mu = \frac{2\rho}{2\rho-1}. \tag{3.14}$$

De Théorème 3.1 et Théorème 3.2, il existe des rapports du pouvoir de k et n, k_1 et n_1. Il est consistant avec l'hypothèse proposée par Hall. Nous sélectionnons encore $\mu = 2/3,\ \beta = 1/2$. Nous déterminons k à partir de $k = k_1(n/n_1)^\mu$. $(\gamma_n([\sqrt{n}])$ est l'estimation préliminaire de γ (Voir [19]). $\tilde{\gamma}_n(k) = \gamma_n([2\sqrt{n}])$.

Soit $\mu = 2/3$, cela suppose le paramètre de second ordre $\rho = -1$. Il a montré que la méthode du bootstrap proposée par Hall est liée à $\rho = -1$.

Une double méthode analogue de bootstrap est appliquée à l'estimateur plus général l'estimateur de moment $\hat{\gamma}_n^{(M)}$ aussi bien que l'estimateur de Pickands $\hat{\gamma}_n^{(P)}$ par Draisma, de Haan, Peng et Perreira (1999) [35] où une exécution étape par étape du procédure est donnée.

3.3.3 Approche séquentielle

Drees et Kaufmaan [19] proposent une approche séquentielle pour construire un estimateur consistent de k_{opt} sans aucune connaissance préalable sur la fonction de distribution F.

Le $\hat{k}_{opt}$ proposé est défini sur la base des temps d'arrêt suivant $\tilde{k}(r_n)$ et $\tilde{k}(r_n^{\zeta})$ avec $0 < \zeta < 1$ pour une suite d'estimateur de Hill par :

$$\tilde{k}(r_n) := min\{k \in \{2, ..., n\} : \max_{2 \leq i \leq k} i^{1/2} \left| \hat{\gamma}_{i,n}^{(H)} - \hat{\gamma}_{k,n}^{(H)} \right| > r_n\}, \tag{3.15}$$

où les seuils r_n constituent une suite qui est de plus grand ordre que $(\log \log n)^{1/2}$ mais de plus petit ordre que $n^{1/2}$.

L'estimation donnée dans le théorème 1 de [19] dépend de quelque estimation cohérente de paramètre du deuxième ordre ρ. Mais comme Drees et Kaufmann mentionnent dans leurs résultats de simulation, que pour beaucoup de distributions, les meilleurs résultats sont obtenus si ρ est fixé (généralement on prend ρ_0 égale -1) au lieu d'être estimé.

Pour fixer $\rho_0 = -1$, la méthodologie de Drees et Kaufmann peut être résumée dans les trois étapes suivantes :

Etape 1 : Poser $r_n = 2.5 \times n^{0.25} \times \tilde{\gamma}_n^{(H)}$, avec $\tilde{\gamma}_n^{(H)} := \hat{\gamma}_{2\sqrt{n},n}^{(H)}$.

Etape 2 : Obtenir $\tilde{k}(r_n)$. Si la condition $\max_{i<k} i^{1/2} \left| \hat{\gamma}_{i,n}^{(H)} - \hat{\gamma}_{k,n}^{(H)} \right| > r_n$ est satisfait pour certains k tel que $k_n(r_n)$ bien définis, puis passer à l'étape 3. Sinon, assigner $0.9 \times r_n$ à r_n et répéter l'étape 2.

Etape 3 : Pour $\xi \in (0, 1)$ (en particulier $\xi = 0.7$), déterminer

$$\hat{k}_{opt} := \left[\frac{1}{3} \left(2 \left(\tilde{\gamma}_n^{(H)} \right)^2 \right)^{1/3} \left(\frac{\tilde{k}(r_n^{\xi})}{\left(\tilde{k}(r_n) \right)^{\xi}} \right)^{1/(1-\xi)} \right],$$

où $[x]$ dénote le plus grand entier inférieur ou égal à x.

Dans [19] ils ont mentionné que la procédure fonctionne également pour une classe large des estimateurs de l'indice de queue, y compris l'estimateur de Pickands, l'estimateur de moment et l'estimateur par la méthode de EMP.

L'avantage Découvert de la méthode de Drees et Kaufmann (1998) est que le paramètre du second ordre ρ peut être fixé à une valeur fixe ρ_0 ou peut être estimé. Elle a construit un estimateur convergent de k qui fonctionne asymptotiquement sans aucune connaissance préalable sur la fonction de répartition sous-jacente.

Le théorème 1 de Drees et Kaufmann (1998) assure que $\hat{k}_{opt}$ est un estimateur convergent de k_{opt} si la fonction de distribution sous-jacente satisfait la condition Hall. Malheureusement, $\hat{k}_{opt}$ n'est pas toujours bien définie, car la procédure peut revenir $\hat{k}_{opt} < 2$.

3.3.4 Approche de couverture de précision

Un des estimateurs les plus connus pour l'index de queue d'une distribution à queue lourde est l'estimateur de Hill. Les intervalles de confiance basés sur l'approximation normale asymptotique de l'estimateur de Hill sont étudiés.

La proposition suivante résulte facilement de Peng et Qi (1997) par noter la relation entre la variation régulière du second ordre pour $1 - F(x)$ et l'autre pour l'inverse de $1/(1 - F(x))$.

Proposition 3.1 *Supposer que (2.5) satisfait et* $k \to \infty$, $k/n \to 0$ *puis*

$$\sqrt{k}\left(\hat{\gamma}_n^{(H)} - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\right) \qquad quand \qquad n \to \infty,$$

ssi $k = o(n^{-2\rho/(1-2\rho)})$.

Une constatation surprenante est que l'ordre de précision pour une couverture optimale de la face d'un intervalle de confiance dépend du signe de la variation régulier du deuxième ordre.

Ainsi, pour $0 < \alpha < 1$ les intervalles de confiance unilatéral et bilatéral pour γ de niveau $(1 - \alpha)$ sont respectivement

$$I_1(\alpha) = \left[0, \hat{\gamma}_n^{(H)} + \frac{z_\alpha \hat{\gamma}_n^{(H)}}{\sqrt{k}}\right]$$

et

$$I_2(\alpha) = \left[\hat{\gamma}_n^{(H)} - z_{\alpha/2}\frac{\hat{\gamma}_n^{(H)}}{\sqrt{k}}, \hat{\gamma}_n^{(H)} + z_{\alpha/2}\frac{\hat{\gamma}_n^{(H)}}{\sqrt{k}}\right],$$

où z_w $(0 \leq w \leq 1)$ est défini par $P(\mathcal{N}(0,1) \leq z_w) = 1 - w$. z_w est le quantile d'ordre $(1 - w)$ de la distribution normale standard.

Nous étudions la précision de la couverture des intervalles de confiance $I_1(\alpha)$ et $I_2(\alpha)$ et donnons le choix optimal théorique de fraction de l'échantillon k pour $I_1(\alpha)$ dans le sens de minimiser l'erreur absolue de couverture.

Il est montré dans [8] quand $n \to \infty$ les probabilités de couverture pour $I_1(\alpha)$ et $I_2(\alpha)$.

Théorème 3.3 *Supposons que (2.5) satisfait et $k \to \infty, k/n \to 0$. Alors*

$$P(\gamma \in I_1(\alpha)) = \alpha - \phi(z_\alpha)\left\{\frac{1+2z_\alpha^2}{3\sqrt{k}} - \frac{\rho d c^\rho}{(1-\rho)}\sqrt{k}(\frac{n}{k})^\rho\right\} + o\left(\frac{1}{\sqrt{k}} + \sqrt{k}(\frac{n}{k})^\rho\right) \tag{3.16}$$

et

$$P(\gamma \in I_2(\alpha)) = \alpha + o\left(\frac{1}{\sqrt{k}} + \sqrt{k}(\frac{n}{k})^\rho\right). \tag{3.17}$$

où ϕ désigne la fonction de densité normale standard.
Par conséquent la valeur optimale de k qui minimise l'erreur absolue de la couverture pour $I_1(\alpha)$ est

$$k_{opt} := \begin{cases} \left(\dfrac{(1+2z_\alpha^2)(1-\rho)}{-3dc^\rho\rho(1-2\rho)}\right)^{1/(1-\rho)} n^{-\rho/(1-\rho)} & \text{si} \quad d > 0, \\ \left(\dfrac{(1+2z_\alpha^2)(1-\rho)}{3dc^\rho\rho}\right)^{1/(1-\rho)} n^{-\rho/(1-\rho)} & \text{si} \quad d < 0, \end{cases} \tag{3.18}$$

qui satisfait automatiquement la condition $k = o(n^{-2\rho/(1-2\rho)})$ dans la proposition 3.1. En outre, la précision de la couverture optimale pour $I_1(\alpha)$ est

$$P(\gamma \in I_1(\alpha)) = \begin{cases} \alpha - 2\left\{\frac{(1-\rho)(1+2z_\alpha^2)}{3(1-2\rho)}\right\}^{(1-2\rho)/(2(1-2\rho))}\left\{\frac{-d\rho}{c^{-\rho}}\right\} & \\ \times\phi(z_\alpha)n^{\rho/(2(1-2\rho))}(1+o(1)) & \text{si } d > 0 \\ \alpha + o(n^{\rho/(2(1-2\rho))}) & \text{si } d < 0. \end{cases} \tag{3.19}$$

Remarques

1. On remarque à partir de (3.18) et (3.19) que l'ordre de la précision de couverture optimale pour $I_1(\alpha)$ dépend du signe de la variation régulier du second ordre.

2. Pour obtenir le choix optimal de k pour $I_2(\alpha)$, nous pouvons avoir besoin d'une condition plus stricte que (2.5) qui est, - la variation régulière de troisième ordre- nous conjecturons que le choix optimal de k dépend aussi de paramètre du troisième ordre. Par conséquent le choix de k devient beaucoup plus difficile.

Comme la fraction de l'échantillon optimale en termes de probabilité de couverture dépend de quelques quantités inconnues, Cheng et peng proposent un estimateur plug-in (numéric) pour la fraction de l'échantillon optimale en se concentrant sur un intervalle de confiance unilatéral

$$\hat{k}_{opt} := \begin{cases} \left(\dfrac{(1+2z_\alpha^2)}{3\hat{\delta}(1+2\hat{\rho})}\right)^{1/(1+\hat{\rho})} n^{\hat{\rho}/(1+\hat{\rho})} & \text{si} \quad \hat{\delta} > 0, \\ \left(\dfrac{(1+2z_\alpha^2)}{-3\hat{\delta}}\right)^{1/(1-\rho)} n^{-\hat{\rho}/(1-\hat{\rho})} & \text{si} \quad \hat{\delta} < 0, \end{cases}$$

où

$$\hat{\rho} := -\log\left(\left|\frac{M_n^{(2)}\left(n/\left(2\sqrt{\log n}\right)\right) - 2\left\{M_n^{(1)}\left(n/\left(2\sqrt{\log n}\right)\right)\right\}^2}{M_n^{(2)}\left(n/\sqrt{\log n}\right) - 2\left\{M_n^{(1)}\left(n/\sqrt{\log n}\right)\right\}}\right|\right) / \log 2,$$

et

$$\hat{\delta} := (1+\hat{\rho})\,(\log n)^{\hat{\rho}/2}\,\frac{M_n^{(2)}\left(n/\sqrt{\log n}\right) - 2\left\{M_n^{(1)}\left(n/\sqrt{\log n}\right)\right\}^2}{2\hat{\rho}\left\{M_n^{(1)}\left(n/\sqrt{\log n}\right)\right\}^2},$$

avec, $M_n^{(r)}(k)$ est définie dans (1.12).

Quelques simulations

1. Dans le tableau 3.2 nous résumons les résultats de la méthode Cheng et Peng pour les données de réclamations danoises et quelques distributions basés sur 300 échantillons de taille 3000.

Dans la Figure 3.3, on obtient le k_{opt} vaut 182 par la méthode de Cheng et Peng, de la loi de Pareto standard.

Cheng & Peng	#des extrêmes	% des extrêmes	Estimateur/Vraie IVE
Pareto standard	182	6.33%	1.14/1
Burr(1,1,1)	195	6.52%	1.05/1
Loggamma(1)	177	5.92%	0.93/1
Fréchet(1)	182	6.09%	0.98/1
Danish Fire	96	4.43%	0.64/0.7

TAB. 3.2 – Nombres optimaux de statistiques d'ordre supérieurs obtenus par la méthode Cheng & Peng et utilisés dans le calcul de l'estimateur de Hill de l'IVE.

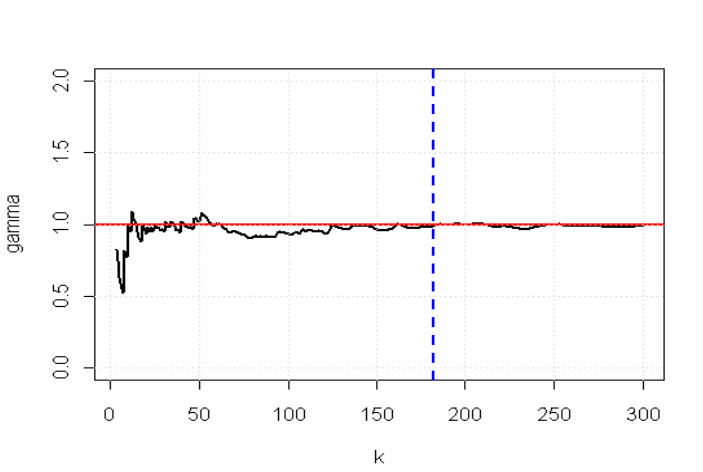


FIG. 3.3 – Estimateur de Hill de l'IVE de la loi Pareto standard. La ligne horizontale représente la vraie valeur de l'index de queue alors que la ligne verticale correspond au nombre optimal des extrêmes par la méthode Cheng et Peng.

2. Nous utilisons 300 échantillons de 3000 observations simulées provenant de plusieurs GEVD avec différents IVE $\{0.3,\ 1,\ 1.5\}$ pour appliquer l'algorithme de Cheng et Peng. Les résultats de cette étude de simulation sont résumés dans le tableau 3.3 et illustré par la Figure 3. 4.

IVE	0.3	1	1.5
Valeur estimée	0.346	1.010	1.500
Erreur absolue	0.046	0.011	0.000
MSE	0.003	0.012	0.015
Erreur relative	0.153	0.011	0.000
Nombre des extrêmes	67	194	264
% des extrêmes	2.23%	6.47%	8.8%

TAB. 3.3 – Résultats de simulation, de l'estimation de l'IVE à l'aide de l'estimateur de Hill et de l'algorithme de Cheng et Peng.

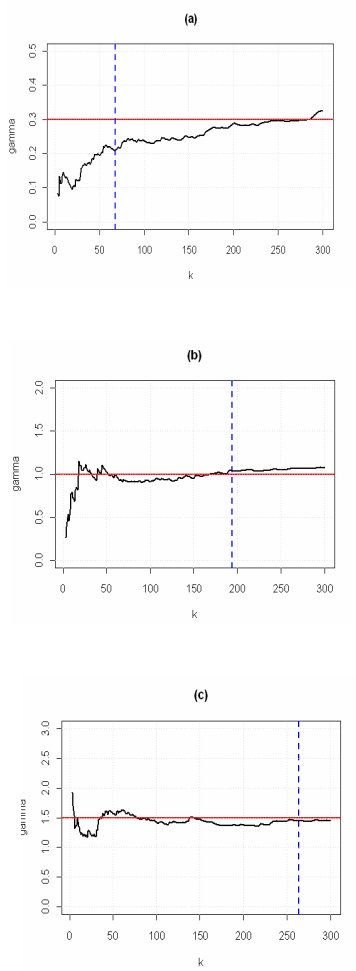


FIG. 3.4–Estimateur de Hill de l'IVE de GEVD, (a) pour $\gamma = 0.3$, (b) pour $\gamma = 1$ et (c) pour $\gamma = 1.5$. La ligne horizontale représente la vraie valeur de l'IVE. La ligne verticale correspond au nombre optimal des extrêmes obtenu par la méthode Cheng & Peng.

3.3.5 Approche de Reiss et Thomas

Reiss et Thomas (1997) [35] ont proposé une méthode heuristique de choisir le nombre des extrêmes pour utiliser dans l'estimation de l'indice de queue.

Reiss et Thomas ont basé leur approche de choisir le nombre adéquat de plus grandes observations sur un moyen de minimiser la distance résumant un terme de pénalité. Dans un certain sens, ce coefficient est prévu pour être plus sévère en ce qui concerne des estimations de γ avec l'origine dans les observations prises plus loin de la queue réelle.
Ils proposent une manière automatique de choisir $\hat{k}_{opt}$ en minimisant

$$\frac{1}{k}\sum_{i\leq k} i^{\beta}\left|\hat{\gamma}_n(i) - med(\hat{\gamma}_{1,n},...,\hat{\gamma}_{k,n})\right|, \qquad 0\leq\beta\leq 1/2. \tag{3.20}$$

Reiss et Thomas ont aussi suggéré de minimiser la modification de (3.20) :

$$\frac{1}{k-1}\sum_{i<k} i^{\beta}\left(\hat{\gamma}_{i,n} - \hat{\gamma}_{k,n}\right)^2, \qquad 0\leq\beta\leq 1/2. \tag{3.21}$$

En ce qui concerne la connaissance des auteurs, pas de méthodologie concernant la spécification de β a été disponible. La motivation principale de ce travail doit alors réduire substantiellement l'importance de variation de β, i.e., trouver bons ensembles de poids dans lesquels la méthode calcule souvent $\hat{k}_{opt} = \hat{k}_{opt}(\beta)$ capable de localiser des estimations raisonnables de γ, quand adopter les estimateurs semi-paramétriques précédemment mentionnés : Hill et Moment.

Sur la base des informations fournie par l'étude de simulation est devenu clairement que le nombre moyen de $\hat{k}_{opt} = \hat{k}_{opt}(\beta)$ a rendu apparaît souvent associée à $\beta = 0$. En outre, il est facile de reconnaître la capacité des valeurs proches de zéro de favoriser des tailles plus élevées de sous-échantillon parce que le terme de pénalité i^{β} devient approximativement 1.

Prenant tout en considération, nous avons trouvé des combinaisons plutôt faisables de l'estimateur semi-paramétrique/la valeur de β/la version (3.20) ou (3.21) de la méthode Reiss et Thomas :

Pour l'estimateur de Hill si les seules informations existantes sur l'indice de queue est que $\gamma > 0$, alors choisi $\beta = 0$ encapsulé dans la version (3.20) de la méthode, alors que pour $0 < \gamma \leq 1$ choisir $\beta = 0.3$ dans (3.20).

La méthode Reiss et Thomas ne tient pas compte des connaissances à priori sur la fonction de distribution sous-jacente mais il est encore capable de produire de meilleurs résultats que la méthode de Drees et Kaufmaan.

Quelques simulations

Nous appliquons l'algorithme de Reiss et Thomas sur 3000 observations simulées de la distribution Pareto standard et quelques distributions et les données de réclamations danoises. Les résultats sont résumés dans le tableau 3.4 et illustrés par les Figures 3.5 et 3.6.

Reiss et Thomas	#des extrêmes	% des extrêmes	Estimateur/Vraie IVE
Pareto standard	202	6.7%	1.14/1
GEVD(1.5)	174	5.8%	1.55/1.5
GEVD(2)	170	5.7%	2.12/2
Burr(1,2,2)	152	5.1%	0.50/0.5

TAB. 3.4 – Nombres optimaux de statistiques d'ordre supérieurs obtenus par la méthode de Reiss et Thomas.et utilisés dans le calcul de l'estimateur de Hill de l'IVE.

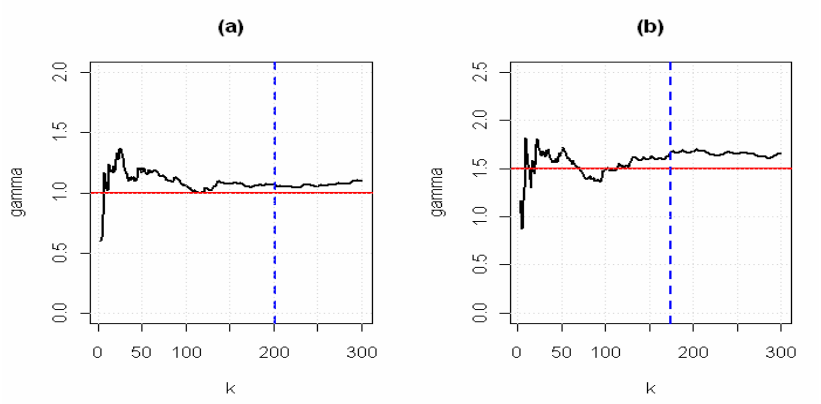


FIG. 3.4 – Estimateur de Hill de l'IVE de distributions (a) Pareto standard et (b) GEVD(1.5). La ligne horizontale représente la vraie valeur de l'index de queue et la ligne verticale correspond au nombre optimal des extrêmes obtenu par méthode Reiss et Thomas.

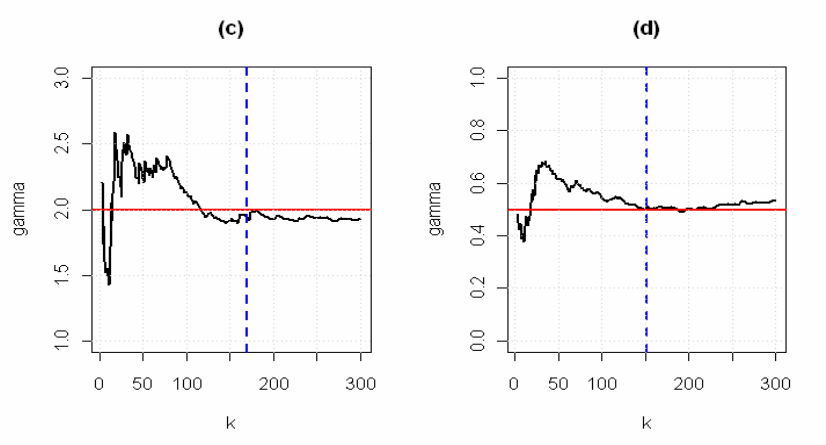


FIG. 3.5 – Estimateur de Hill de l'IVE de (c) GEVD(2) et (d) Burr(1,1,1). La ligne horizontale représente la vraie valeur de l'index de queue alors que la ligne verticale correspond au nombre optimal des extrêmes de Reiss et Thomas.

Comparaison

Nous appliquons les algorithmes de Cheng et Peng [8] et Reiss et Thomas [35] sur 4000 observations simulées de la distribution standard de Pareto. Les résultats sont résumés dans le tableau 3.5 et illustré par la Figure 3.6.

Algorithme	#des extrêmes	% des extrêmes	Estimateur/Vraie EVI
Reiss & Thomas	183	4.57%	1.07/1
Cheng & Peng	236	5.9%	1.00/1

TAB. 3.5 – Nombres optimaux de statistiques d'ordre supérieurs utilisés dans le calcul de l'estimateur de Hill de l'IVE de la distribution de Fréchet(1) et de la distribution de Pareto standard, basé sur 4000 observations.

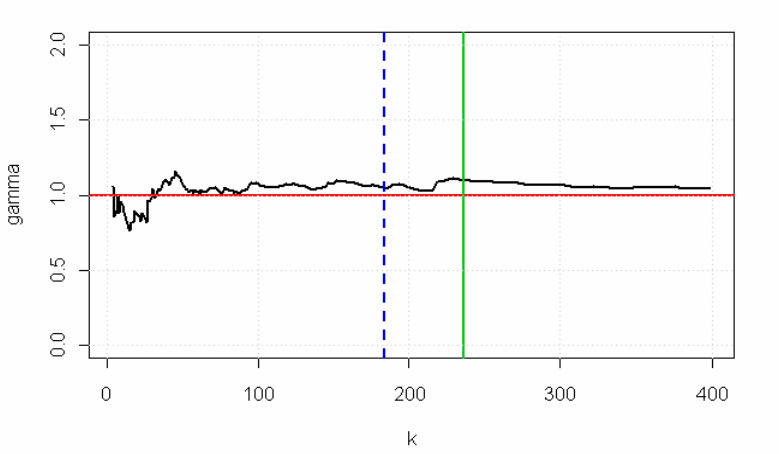


FIG. 3.6 – Estimateur de Hill de l'indice de l'IVE de la distribution de Pareto standard, basé sur 3000 observations. La ligne horizontale représente la la vraie valeur de l'indice de queue alors que les lignes verticales correspondent aux numéros optimale des extrêmes de Cheng et Peng (solide) et Reiss et Thomas (pointillés).

D'après le tableau 3.5 on peut dire que les méthodes adaptives pour sélectinner une bonne valeur de statistiques d'ordre extrêmes (Cheng & Peng et Reiss & Thomas) donnent des résultats très proches de la vraie valeur de l'IVE mais la méthode de Cheng & Peng est plus rapide que l'autre méthode. La méthode de Reiss & Thomas est long puisque elle est basée sur le calcul de la médiane.

3.3.6 Choix automatique de paramètre de lissage

Dans le cas des estimateurs de type noyau, le problème devient une question de choix de paramètre de lissage optimale h_{opt} pour laquelle un estimation consistance $\hat{h}_{opt}$ doit être calculé de manière adaptative. Groeneboom, Lopuhaä et de Wolf [25] suivre une approche similaire à celle dans [18]. Ils fondent leurs estimateurs $\hat{\gamma}_n^{(W1)}$ et $\hat{\gamma}_n^{(W2)}$ sur les noyaux respectifs suivants :

$$K_1(u) = \frac{15}{8}(1-u^2)^2 I_{[0,1]}(u)$$

et

$$K_2(u) = \frac{35}{16}(1-u^2)^3 I_{[0,1]}(u).$$

Nous présentons d'abord la méthode, exposée dans Draisma, de Haan, Peng et Pereira [12], comme elle s'appliquerait à notre estimateur de type noyau.

Soit $X_1, ..., X_n$ un échantillon de une distribution pour laquelle nous voulons estimer l'indice des valeurs extrêmes.

L'algorithme de cinq étapes menant au $\hat{h}_{opt}$ est comme suit :

Etape1 : Sélectionner un échantillon de bootstarp $(X_1^*, \ldots, X_{n1}^*)$ de taille $(n_1 \ll n)$ de l'échantillon initial ($X_1, \ldots, X_n$) et calculer $\hat{\gamma}_{n_1,h}^{(K_1^*)}$ et $\hat{\gamma}_{n_1,h}^{(K_2^*)}$ en termes des statistiques d'ordre $X_{1,n_1}^* \leq \ldots \leq X_{n_1,n1}^*$ concernant l'échantillon bootstrap. Ensuite calculer

$$\delta_{n_1,h}^* := \gamma_{n_1,h}^{(W_1^*)} - \gamma_{n_1,h}^{(W_2^*)}.$$

Etape 2 : Répéter l'étape 1, r fois de façon indépendante. Avec la séquence obtenue $\delta_{n_1,h,1}^*, ..., \delta_{n_1,h,r}^*$ calculer

$$\widehat{MSE}^*(\delta_{n_1,h}^*) := \frac{1}{r}\sum_{i=1}^{r}(\delta_{n_1,h,i}^*)^2.$$

qu'est l'estimation de l'erreur moyenne quadratique du bootstrap de $\delta_{n_1,h}^*$.

Etape 3 : Calculer

$$h^*(n_1) := \arg\min_h \widehat{MSE}^*(\delta_{n_1,h}^*)$$

Dans la pratique, on pourrait calculer $\widehat{MSE}^*(\delta_{n_1,h}^*)$ sur une grille des valeurs du h_i, par exemple avec la distance 0.01 entre les valeurs successives (la distance exacte pourrait être choisie dépend de la taille de l'échantillon n_1), et alors prendre pour $h^*(n_1)$ minimiser les valeurs de $\widehat{MSE}^*(\delta_{n_1,h_i}^*)$.

Etape 4 : Répéter les étapes 1-3 indépendamment avec $n_2 = [n_1^2/n]$ au lieu n_1 : soit

$$h^*(n_2) := \arg\min_h \widehat{MSE}^*(\delta_{n_2,h}^*).$$

Etape 5 : Estimter la fenêtre optimale h_{opt} par

$$\hat{h}_{opt} := c(h^*(n1), h^*(n2))\frac{(h^*(n1))^2}{h^*(n2)},$$

où $c(h_1, h_2)$ est une fonction de h_1 et h_2 selon les noyaux K_1 et K_2 et les tailles de l'échantillon n_1 et n_2 respectivement.

3.4 Discussion

- Le problème du choix de la valeur optimale de statistiques d'ordre extrêmes k a reçu beaucoup d'attention des chercheurs. Ils ont déjà soulevé de nombreuses méthodes.

- Nous pourrions construire quelques règles naïves, tels que :

  1. Pour $\alpha < 1.5$ prendre 5% à 10% des plus grandes observations pour estimer l'indice de queue.

  2. Pour $\alpha \geq 1.5$ prendre entre 2% et 5% des plus grandes observations pour estimer l'indice de queue.

Ces types de règles fixes ont été utilisés dans le passé, mais ils peuvent mener aux résultats très erronés, car il n'y a pas une règle optimale unique pour toutes les valeurs de $\gamma \in [1, 2]$. Cela pourrait suggérer que le choix optimal pour le nombre de statistiques d'ordre supérieur pour l'utilisation dans l'estimation de l'indice de queue devrait se situer entre 2% et 10% de l'ensemble des données.

Conclusion

Pour les estimateurs semi-paramétrique de l'indice des valeurs extrêmes et leurs dépendance et de leurs sensibilité sur le nombre k de statistiques d'ordre supérieurs utilisés dans l'estimation. Aucune règle stricte et rapide n'existe pour affronter ce problème.

Plus généralement, le choix du k meilleur provient de la compétition entre le biais et la variance. D'un côté, la tendance naturelle serait, à n fixé, d'accroître k pour diminuer la variance. Mais d'un autre côté, il faut tenir compte du biais des estimateurs. L'arbitrage entre les deux effets contraires se fait usuellement en calculant l'erreur moyenne quadratique de l'estimateur (dépendant de k) puis en le minimisant en k.

Habituellement, le scientifique décide subjectivement sur le nombre k à utiliser, en regardant les graphiques appropriés. Des manières plus objectives pour faire ceci sont les méthodes adaptives mentionnées dans la Section 3.3.

Enfin, il faut mentionner que une nouvelle branche et prometteuse de l'analyse des valeurs extrêmes est celle de méthodes des valeurs extrêmes multivariées.

Bibliographie

[1] Arnold, B.C., Balakrishnan, N. et Nagaraja, H.N. (1992). A First Course in Order Statistics. Wiley, New York.

[2] Beirlant, J., Vynckier, P. et Teugels, J.L. (1996). Tail index estimation, Pareto quantile plots, and regression diagnostics. J. Amer. Statist. Assoc., **91**, 1659-1667.

[3] Beirlant, J., Vynckier, P. et Teugels, J.L. (1996). Excess function and estimation of the extreme-value index. Bernoulli **2**, 293-318.

[4] Beirlant, J., Goegebeur, Y. et Matthys, G. (1999). Tail Index Estimation and an Exponential Regression Model. Extremes **2**, 177-200.

[5] Beirlant, J., Goegebeur, Y., Segers, J. et Teugels, J. (2004). Statistics of Extremes - Theory and Applications. Wiley.

[6] Beirlant, J., Dierckx, G. et Guillou, A. (2005). Estimation of the extreme-value index and generalized quantile plots. Bernoulli **11**, 949–970

[7] Beirlant, J., Bouquiaux, C. et Werker, B.J.M. (2006). Semiparametric Lower Bounds for Tail Index Estimation. Journal of Statistical Planning and Inference **136**, 705-729

[8] Cheng, S. et Peng, L. (2001). Confidence Intervals for the Tail Index. Bernoulli **7**, 751-760.

[9] Csörgö, S., Deheuvels, P. et Mason, D. (1985). Kernel Estimates of the Tail Index of a Distribution. Annals of Statistics **13**, 1050-1077.

[10] Csörgö, S. et Viharos, L. (1998). Estimating the Tail Index. Asymptotic Methods in Probability and Statistics. B. Szyszkowicz, ed., North Holland, Amsterdam, 833-881.

[11] Danielsson, J. et de Vries, C.G. (1997). Tail Index and Quantile Estimation with Very High Frequency Data. Journal of Empirical Finance **4**, 241-257.

[12] Danielsson, J., de Haan, L., Peng, L. et de Vries, C.G. (2001). Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. Journal of Multivariate Analysis **76**, 226-248.

[13] Davison, A.C. et Smith R.L. (1990). Models for Exceedances over High Thresholds (with discussion). Journal of the Royal Statistical Society, Series B **52**, 393-442.

[14] Delmas, J.F. et Jourdain , B. (2006). Modèles aléatoires. Applications aux sciences de l'ingénieur et du vivant. Springer.

[15] Dekkers, A.L.M. et de Haan, L. (1989). On the Estimation of the Extreme-Value Index and Large Quantile Estimation. Annals of Statistics **17**, 1795-1832.

[16] Dekkers, A.L.M., Einmahl, J.H.J. et de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. Ann. Statist **17**, 1833-1855.

[17] Dekkers, A. L. M. et de Haan, L. (1993). Optimal choice of sample fraction in extreme-value estimation. J. Multivariate Anal. **47**, 173–195.

[18] Draisma, G., de Haan, L., Peng, L. et Perreira, T.T. (1999). A Bootstrap-Based Method to Achieve Optimality in Estimating the Extereme-Value Index. Extremes **2**, 367-404.

[19] Drees, H. et Kaufmann, E. (1998). Selection of the Optimal Sample Fraction in Univariate Extreme Value Estimation. Stochastic Processes and their Applications **75**, 149-195.

[20] Embrechts P., Klüppelberg C., Mikosh T. (1997). Modelling Extremal Events for Insurance and Finance, Springer, Berlin.

[21] Fisher, R.A. et Tippett, L.H.C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. Proceedings of the Cambridge Philosophical Society **24**, 180-190.

[22] Fraga Alves, M.I. (1995). Estimation of the Tail Parameter in the Domain of Attraction of an Extremal Distribution. Journal of Statistical Planning and inference **45**, 143-173.

[23] Gardes ,L. (2003). Estimation d'une fonction quantile extrême. Thèse de doctorat, Université de Montpellier II.

[24] Garrido, M. (2002). Modélisation des évènements rares et estimation des quantiles extrêmes, Méthodes de sélection de modèles pour les queues de distribution, Thèse de doctorat, Université de Grenoble I.

[25] Groeneboom, P., Lopuhaä, H.P.et de Wolf, P.P. (2003). Kernel Estimators for the Extreme Value Index. Annals of Statistics **31**, 1956-1995.

[26] Guillou. A. et Willems. P. (2006). Application de la théorie des valeurs extrêmes en hydrologie. Statistique Appliquée, **LIV** (2), 5-31

[27] de Haan. L et Ferreira. A. (2006). Extreme value theory : an introduction. Springer.

[28] Hall, P. et Welsh, A.H. (1985). Adaptive Estimates of Parameters of Regular Variation. Annals of Statistics **13**, 331-341.

[29] Hill, B. (1975). A Simple Approach to Inference About the Tail of a Distribution. Annals of Statistics **3**, 1163-1174.

[30] Hosking, J. et Wallis, J. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. Technometrics 29, 339-349.

[31] Matthys, G. et Beirlant, J. (2003). Estimating the Extreme Value Index and High Quantiles with Exponential Regression Models. Statistica Sinica **13**, 853-880.

[32] Meraghni, D. (2008). Modelling distribution tails. Thèse de doctorat, Université de Biskra.

[33] Nagaraja, H. N. et David, H. A. (2003). Order Statistics. Third Edition. Wiley.

[34] Necir, A. (2006). A Functional Law of the Iterated Logarithm for Kernel-type Estimators of the Tail Index. Journal of Statistical Planning and Inference **136**, 780-802.

[35] Neves, C. et Fraga Alves, M.I. (2004). Reiss and Thomas' Automatic Selection of the Number of Extremes. Computational Statistics and Data Analysis **47**, 689-704.

[36] Reiss, R.D. et Thomas, M. (1997). Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields. Birkhäuser, Basel.

[37] Rootzen, H. et Tajvidi, N. (1996). Extreme Value Statistics and Wind Storm Losses : a Case Study, Scandinavian Actuarial Journal, 70-94.

[38] Saporta, G. (1990). Probabilités, Analyse des Données et Statistique. Editions Technip, Paris.

[39] Smith, R.L**.** (1987), Estimating tails of probability distributions, Annals of Statistics **15**, No. 3, 1174–1207.

[40] Sousa, B.C. (2002). A Contribution to the Estimation of the Tail Index of Heavy-Tailed Distributions. Thèse de doctorat, Université de Michigan, www.utstat.toronto.edu/~desousa.

[41] Tassi, Ph. et Legait, S. (1990). Théorie des probabilités en vue des applications statistiques. Edition Technip.

[42] de Wolf, P.P. (1999). Estimating the Extreme Value Index. Thèse de doctorat, Delft University of Technology.

[43] Yun, S. (2002). On a Generalized Pickands Estimator of the Extreme Value Index. Journal of Statistical Planning and Inference **102**, 389-409.

Annexe A

Logiciel statistique R

R est un système d'analyse statistique et graphique crée par Ross Ihaka et Robert Gentleman. R est à la fois un logiciel et un langage qualifié de dialecte du langage S crée par AT&T Bell Laboratoires.

R comporte de nombreuses fonctions pour les analyses statistiques et les graphiques ; ceux-ci sont visualisés immédiatement dans une fenêtre propre et peuvent être exportes sous divers formats (jpg, png, bmp, ps, pdf, emf, pictex, xfig ; les formats disponibles peuvent dépendre du système d'exploitation).

Les résultats des analyses statistiques sont affichés à l'écran, certains résultats partiels (valeurs de P, coefficients de régression, résidus, ...) peuvent être sauvés à part, exportés dans un fichier ou utilisés dans des analyses ultérieures.

Le langage R permet, par exemple, de programmer des boucles qui vont analyser successivement différents jeux de données. Il est aussi possible de combiner dans le même programme différentes fonctions statistiques pour réaliser des analyses plus complexes. Les utilisateurs de R peuvent bénéficier des nombreux programmes écrits pour S et disponibles sur internet, la plupart de ces programmes étant directement utilisables avec R.

Aux fins de ce mémoire, nous avons écrit des programmes appropriés que nous avons stocké dans des fichiers texte et exécuté en utilisant les paquets suivants :

- boot : pour le bootstrap dans les exemples basés sur des données simulées.
- actuar : pour simuler les distribution des pertes, des risques.
- evd, evir , fExtremes et ismev : pour la modélisation de GEVD et GPD.
- POT : pour l'approximation GPD.

Le site officiel pour se procurer le logiciel R et les documents d'aide est :

http ://cran.r-project.org/

Résumé

La motivation principale de notre mémoire de Magister est de sélectionner le nombre optimal de statistiques d'ordre extrêmes cruciale pour l'estimation de l'IVE et permet d'améliorer la performance des estimateurs. Dans ce mémoire, nous exposons les différentes méthodes de détermination de ce nombre. Des applications sur des données aussi bien réelles que simulées permettront d'illustrer les résultats obtenus.

Mots-clés: Valeur extrême, Normalité asymptotique, Estimateur Hill, Point terminal, Variation régulière, Domaine d'attraction, Statistique d'ordre, paramètre du second ordre, Bootstrap, Erreur moyenne quadratique, Queues lourdes, Indice de queue, Fraction de l'échantillon.

Abstract

The main motivation of our memory Magister is to select the optimal number of extreme order statistics crucial for the estimation of EVI and improves the performance of estimators. in this paper, we outline the different methods for determining this number. Applications on both real data that will illustrate simulated results.

Key words: Extreme value, Asymptotic normality, Hill Estimator, Upper endpoint, Regular Variation, Domain of attraction, Order statistics, Second order parameter, Bootstrap, Mean Square Error, Heavy tails, Tail index, Fraction sample.

ملخص

الهدف الرئيسي من مذكرة الماجستير هو تحديد العدد الأمثل للإحصاءات القيم القصوى الذي هو عنصر مهم جدا في تقدير قيمة مؤشر القيم القصوى وتحسين أداء المقدرات. في هذا الصدد طرحنا في هذه المذكرة أساليب مختلفة لتحديد هذا العدد. التطبيقات على كل البيانات الحقيقية التي من شأنها أن توضح النتائج التي تم الحصول عليها.

المصطلحات: القيمة القصوى، المقارب الطبيعي، مقدر هيل، نقطة النهاية، الاختلاف المضبوط، منطقة جذب، إحصائيات الترتيب، معامل الدرجة الثانية، *إعادة* المعاينة، الخطأ التربيعي المتوسط، الذيول الثقيلة، مؤشرالذيل، جزء من العينة.