



**UNIVERSITE KASDI MERBAH
OUARGLA**

**Faculté des mathématiques et sciences de la
matière**



DEPARTEMENT DE MATHÉMATIQUES

MASTER

Spécialité : Mathématiques

Option : Statistique et probabilité

Par : ALI BEN ALI

Thème

**Estimation de la prime des dépassements des risques
extrêmes par l'intégral de lyenden-bell sous des données
tronquées à droite**

Soutenu publiquement le : 01/06/2017

Devant le jury composé de :

Mr.AGOUN Rachid	M.A université de KASDI Merbah - Ouargla	Président
Mr.ZIBAR Said	M.A université de KASDI Merbah - Ouargla	Examineur
M. MEDDI Fatima	M.C université de KASDI Merbah - Ouargla	Rapporteur

Dédicaces

Je dédie ce travail à :

Mes parents

A mes frères et

mes sœurs, et toute la famille BEN ALI

A mes chers amis

À tout promotion de probabilité et statistique

Et a tous mes professeurs

Enfin à tous ceux qui m'ont aidé de proche ou de loin

Remerciements

Je remercie Dieu le tout-puissant de m'avoir donné la volonté, la force et le courage pour bien mener et finir mon travail de thèse.

Je voudrais d'abord et avant tout remercier mon encadreur vertueuse MEDDI FATIMA pour tous leurs efforts en vue d'établir ce mémoire, elle a eu le rôle fondamental et essentiel et la grand mérite dans tout ce qui a été réalisé, comme cela avait toujours été, près de moi et me guider et corriger mes erreurs et me donner de précieux conseils et les conseils appropriés et les alertes considérés. Je la répète mes remerciements pour tout ce qu'elle a fait pour moi à travers la mise en plan à toutes les étapes de la préparation de ce mémoire depuis le début jusqu'à la fin, étape par étape elle était que le premier et dernier facteur pour le succès de ce travail, je lui dis encore une fois merci beaucoup pour votre appréciation profonde.

Avec un grand honneur, j'aimerais présenter mes remerciements et ma gratitude aux membres du jury, Monsieur AGOUN RACHID, et Monsieur SAID ZIBAR tout d'abord d'avoir accepté d'examiner mon mémoire, qui sans eux ce mémoire ne pourra jamais voir le jour pour l'intérêt et apport qu'ils ont apporté à mon travail.

J'exprime ma gratitude à ma famille qui m'a toujours soutenue et encouragée dans la voie que je m'étais fixée. Je remercie particulièrement mes parents qui m'ont stimulé et encouragé pendant mes études. qui étaient toujours prêts à fournir tous les moyens physiques et morales pour la réussite de ce projet.

Notation

VR_α	Variation régulière d'indice α
<i>al.</i>	Autres.
(Ω, A, \mathbb{P})	Espace probabilisé
S_n	Ensemble des permutation de $\{1, \dots, n\}$
$D.A(F)$	Doumain d'attraction de Frchet
x_F	Point terminal.
<i>i.e.</i>	En d'autre terme
<i>v.a</i>	Variable aléatoire
<i>v.a.r</i>	Variable aléatoire réelle
<i>i.i.d.</i>	Indépendent identiquement distribué
$F(\cdot)$	Fonction de distribution
$\bar{F}(\cdot)$	Fonction des queues de distribution
$F_n(\cdot)$	Fonction de distribution empirique
$\hat{\gamma}_1^{(JR)}$	L'estimateur de l'indice v.e Julien Worms, Rym Worms(1015)
$\hat{\gamma}_1^{(LB)}$	L'estimateur de l'indice v.e Houas et al. (2016)
$\xrightarrow{\mathcal{D}}$	Converge en distribution
\xrightarrow{P}	Converge en probabilité
$\xrightarrow{\mathcal{L}}$	Converge en loi
$:\stackrel{\mathcal{L}}{=}$	Égalité en loi
Π	La prime
$\hat{\Pi}$	La prime estimer
$M_n = X_{n,n}$	Maxcimum de X_1, \dots, X_n
POT	Pics au-delà d'un seuil
GEV	Distribution des valeurs extrêmes généralisée
GPD	Distribution de pareto généralisée
H_γ	Famille de la loi des valeurs extrêmes généralisée
X_{n-k}	Le seuil aléatoire
<i>v.a</i>	Variable aléatoire
$O_P(\cdot)$	Être borné en probabilité
$\mathbf{W}(s)$	Processus de Wiener standard
$\mathbf{D}_n^{(LB)}$	Processus de Lynden-Bell
$\mathbf{D}_n^{(W)}$	Processus de Woodroofe

Table des matières

Dédicaces	1
Remerciements	2
Notations	4
Introduction	9
1 Préliminaires mathématiques des valeurs extrêmes et tronquées	11
1.1 La théorie des valeurs extrêmes	11
1.1.1 Statistique d'ordre	12
1.1.2 Distributions empiriques	14
1.1.3 Loi des valeurs extrêmes	15
1.1.4 Distribution des valeurs extrêmes généralisée	17
1.1.5 Caractérisation des Domaines d'attraction	18
1.1.6 Estimateur de Hill $\hat{\gamma}_{k_n}^{Hill}$	20
1.2 Quelques généralités sur la troncature	24
1.2.1 Données de survie	24
1.2.2 Données tronquées	25
1.3 Le Modèle POT	26
1.3.1 Loi des excès	27
1.4 Recherche du seuil u	29
2 Estimation de la prime sous des données tronquées	30
2.1 Introduction sur la prime	30

2.1.1	Mesures de risque	30
2.1.2	Principes de calcul de prime	31
2.1.3	Estimateur de la prime de Necir et al. (2015) $\hat{\Pi}_n$	32
2.2	Estimation de la prime avec un seuil fixé	37
2.2.1	Estimateur de l'indice de queue $\hat{\gamma}_1^{(JR)}$	37
2.2.2	Hypothèses et résultats	39
2.2.3	Estimation des quantiles extrêmes x_p	40
2.2.4	Méthode d'estimation de la prime $\hat{\Pi}_1$	42
2.3	Estimation de la prime avec un seuil aléatoire	43
2.3.1	Estimateur de l'indice de queue $\hat{\gamma}_1^{(LB)}$	43
2.3.2	Méthode d'estimation de la prime $\hat{\Pi}_2$	48
3	Étude de simulations	49
3.1	Génération de l'échantillon tronqué	49
3.2	Simulations dans le cas du seuil fixé	51
3.2.1	Simulations de $\hat{\gamma}_1^{(JR)}$	51
3.2.2	Simulations de $\hat{\Pi}_1$	54
3.3	Simulations dans le cas du seuil aléatoire	54
3.3.1	Simulations de $\hat{\gamma}_1^{(LB)}$	54
3.3.2	Simulation de $\hat{\Pi}_2$	56
3.3.3	Représentation graphique :	57
3.4	Résultats des simulations	58
	Conclusion	62
	Bibliographie	64

Table des figures

1.1	Fonction de distribution empirique et théorique	14
1.2	Distributions standard des valeurs extrêmes.	17
1.3	Représentation de la densité de probabilité : Gumbel ($\gamma = 0$), Fréchet($\gamma = 1$), Weibull ($\gamma = -1$)	19
1.4	Représentation de la fonction de répartition Gumbel ($\gamma = 0$), Fréchet ($\gamma =$ -1), Weibull ($\gamma = 1$)	19
1.5	Observations X_1, \dots, X_{11} et excès Y_1, \dots, Y_6 au-delà du seuil u	26
1.6	Graphe de densité, Pareto, Beta et Exponentielle	28
3.1	Distribution de l'échantillon initial X de Pareto ($\gamma_1=0.15$), avant la tron- cature de taille $N=2000$	58
3.2	Distribution de l'échantillon Y de Pareto ($\gamma_2=2.9$), appelée v.a de tronca- ture de taille $N=2000$	58
3.3	Représentation de l'échantillon X après la troncature par l'échantillon Y avec $n < N=2000$	59
3.4	Estimateur $\hat{\Pi}_2$ et la prime pour $\gamma_1 = 0.02, \gamma_2 = 2.9$	59
3.5	Estimateur $\hat{\Pi}_2$ et la prime pour $\gamma_1 = 0.15, \gamma_2 = 2.9$	60
3.6	Estimateur $\hat{\Pi}_2$ et la prime pour $\gamma_1 = 0.5, \gamma_2 = 2.9$	60

Liste des tableaux

- 1.1 Quelques distributions appartenant au domaine d'attraction de Fréchet . . . 20
- 1.2 Quelques distributions appartenant au domaine d'attraction de Gumbel . . . 20
- 1.3 Quelques distributions appartenant au domaine d'attraction de Weibull . . . 20

- 3.1 $X \sim$ Pareto ($\gamma_1 = 0.2$) tronques par $Y \sim$ Pareto ($\gamma_1 = 2.5$), $N=2000$ et après
troncature $n=1713$ 54

Introduction

La théorie des valeurs extrêmes a été développée pour l'estimation de probabilités d'occurrences d'événements 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). Pour plus de détails voir Embrechts et al. (2007).

La troncature s'affiche lorsqu'un temps d'événement est uniquement observé dans une étude, si le temps de la variable de l'événement est plus grand ou plus petit que la variable de troncature. Quelques exemples de données tronquées en astronomie, et en économie se trouvent dans Woodroffe (1985), et pour une application dans l'analyse des données sur le SIDA voir Wang (1989). En matière de fiabilité, un ensemble de données réel, consistant en des temps de vie des plaquettes de frein automobiles est déjà été discuté dans la page 69 par Lawless (2002), et été récemment analysé par Gardes and Stupfier (2015) comme une application dans des modèles à queue lourdes tronqués.

Le premier papier publié sur l'estimation des valeurs extrêmes sous troncature aléatoire a été publié par L.Gardes et G.Stupfier (2015), qui a traité des données tronquées à droites à queue lourde. Worms et Worms (2015) ont proposé un nouvel estimateur de l'indice de valeurs extrêmes qui prend la forme d'une intégrale de Lynden-Bell d'une fonction particulière, et est donc une sorte de version naturelle de l'estimateur de Hill (1975), mais en termes de données tronquées à droite avec un seuil fixé donné. Récemment, Haouas et al. (2016) ont utilisé le même concept pour proposer un estimateur de l'indice de queue sur les k -plus grandes valeurs ensuite ont déterminé ses propriétés asymptotiques, mais avec un seuil aléatoire.

Ce mémoire, est réparti en trois chapitres.

Dans le premier chapitre nous présentons tout d'abord quelques rappels et définitions essentiels sur la théorie des valeurs extrêmes, nous discutons également la notion des domaines d'attraction d'une distribution selon le paramètre de l'indice de queue. Nous présentons aussi quelques rappels sur les données tronquées.

Le deuxième chapitre est réparti en trois sections, dans la première section nous présentons une introduction sur la prime, les mesures de risque et le principe de calcul de prime, en particulier la prime nette pour un échantillon de risque positive suivant une distribution à queue lourde. Sous des données tronquées, Necir et al. (2015) ont présenté un nouvel estimateur de l'indice de queue basé sur une fraction d'échantillon unique des observations supérieures à partir de données tronquées et de troncature. Ils ont aussi établi sa normalité asymptotique. Par l'utilisation de l'estimateur bien connu de Lynden-Bell (1971), ils ont appliqué leur nouvel estimateur de l'indice précité pour développer un estimateur de la prime nette des dépassements de réassurance. Dans la deuxième section nous avons calculé un nouvel estimateur de la prime nette dans le cas d'un seuil fixé, en utilisant l'estimateur de l'indice de valeur extrême proposé récemment par Worms, et Worms (2015). Dans la troisième section, nous avons fait de même que la deuxième pour un seuil aléatoire en se basant sur l'estimateur de l'indice de valeurs extrêmes de Haouas et al. (2016).

Nous avons consacré le troisième chapitre pour des études de simulations sous le logiciel R pour montrer comment se fait la troncature par un échantillon de variables aléatoire, notamment calculer l'estimateur de l'indice de queue dans le cas du seuil fixé et le cas aléatoire pour finir avec des graphiques qui montrent l'estimateur de la prime dans les deux cas contre la prime théorique.

PRÉLIMINAIRES MATHÉMATIQUES DES VALEURS EXTRÊMES ET TRONQUÉES

1.1 La théorie des valeurs extrêmes

LA théorie des valeurs extrêmes est une branche probabiliste fascinante qu'on rencontre dans de nombreux domaines pratiques tels que l'assurance, la réassurance et la finance. Ce domaine de recherche a d'abord attiré aussi bien l'intérêt des probabilistes que celui des ingénieurs de risques financiers et assuranciers. Il est basé sur la connaissance de la loi asymptotique des extrêmes (maximum et minimum) d'un échantillon aléatoire. En 1928, Fisher et Tippett établissent un résultat crucial qui spécifie la forme de la distribution limite du maximum convenablement normalisé. Ils montrent que la distribution limite possible des extrêmes est l'une des trois familles de distributions suivantes : Gumbel, Fréchet ou Weibull, appelées distributions de valeurs extrêmes.

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, \dots, 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 $x \in \mathbb{R}$).

1.1.1 Statistique d'ordre

La statistique d'ordre de l'échantillon (X_1, \dots, X_n) est le réarrangement croissant de (X_1, \dots, X_n) . On le note par $(X_{1;n}, \dots, X_{n;n})$. On a $X_{1;n} \leq \dots \leq X_{n;n}$ et il existe une permutation aléatoire $\sigma \in S_n$ telle que :

$$(X_{1;n}, \dots, X_{n;n}) = (X_{\sigma_n(1)}, \dots, X_{\sigma_n(n)})$$

Le vecteur $(X_{1;n}, \dots, X_{n;n})$ est appelé l'échantillon ordonné associé à l'échantillon (X_1, \dots, X_n) , et $X_{k;n}$ étant la $k^{\text{ième}}$ statistique d'ordre. Dans un échantillon de taille n deux statistiques d'ordre sont particulièrement intéressantes pour l'étude des événements extrêmes ce sont :

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

En utilisant la propriété d'indépendance des variables aléatoires X_1, X_2, \dots, 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, \dots, X_n sont :

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

Alors,

$$F_{X_{n;n}}(x) = [F(X)]^n$$

Puis,

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

$$\begin{aligned}
&= 1 - \prod_{i=1}^n p(X_i > x) \\
&= 1 - \prod_{i=1}^n [1 - p(X_i \leq x)] \\
&= 1 - \prod_{i=1}^n [1 - F_X(x)]
\end{aligned}$$

D'où on déduit :

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

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 :

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

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).

Loi de $X_{i,n}$

$$F_{i,n} = P\{X_{i,n} \leq x\} = \sum_{k=1}^n \binom{n}{k} (F(x))^k (1 - F(x))^{n-k}$$

Nous en déduisons que la fonction de densité est :

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

où $f(x)$ est la densité de probabilité de X_i et F sa fonction de répartition associée. En utilisant la propriété d'indépendance des variables aléatoires X_1, \dots, X_n , on obtient :

Loi de $X_{1,n}$

$$F_{1,n} = P\{X_{1,n} \leq x\} = 1 - (1 - F(x))^n$$

d'où

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

Loi de $X_{n,n}$

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

1.1.2 Distributions empiriques

Définition 1 (La fonction de répartition empirique) notée F_n et donnée par :

$$F_{X_{n,n}}(x) = \begin{cases} 0 & \text{si } x \leq X_{1,n}, \\ \frac{i-1}{n} & \text{si } X_{i-1,n} \leq x \leq X_{i,n}, \quad 2 \leq i \leq n, \\ 1 & \text{si } x > X_{n,n}. \end{cases}$$

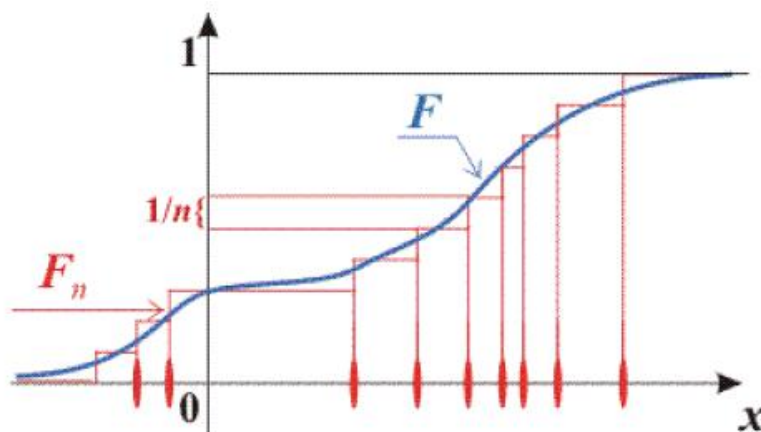


FIGURE 1.1 : Fonction de distribution empirique et théorique

Définition 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 $n \geq 1$

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

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

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

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

Définition 3 (Fonction des quantiles de queue) La fonction des quantiles de queue (tail quantile function), notée U est définie par :

$$\begin{aligned} U(t) &= Q(1 - 1/t) \\ &= (1/\bar{F})^{\leftarrow}(t) \quad , -1 < t < 1. \end{aligned}$$

La fonction des quantiles de queue empirique notée U_n est donnée par :

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

1.1.3 Loi des valeurs extrêmes

Comme la fonction de répartition obtenue précédemment conduit à une loi dégénérée lorsque n tends vers l'infini, on recherche une loi non dégénérée pour le maximum de X . Cette loi limite non dégénérée est fournie par le "théorème des distributions extrêmes" qui donne une condition nécessaire et suffisante pour l'existence d'une loi limite non dégénérée pour le maximum. Ce théorème est proposé par Gnedenko (1943) qui donne la forme des lois limites et Jenkinson (1955) qui en donne l'expression générale.

Théorème 4 (Fisher et Tippett, 1928, Gnedenko, 1943) Soit X_1, \dots, X_n une suite de n variables aléatoires réelles iid de loi continue P et $M_n = \max(X_i)_{1 \leq i \leq n}$. S'il existe deux suites réelles $(a_n)_{n \geq 1}$ et $(b_n)_{n \geq 1}$ avec $b_n > 0$, et une fonction de répartition non-dégénérée H_γ telle que,

$$\frac{M_n - a_n}{b_n} \xrightarrow{d} H_\gamma, \quad \text{lorsque } n \rightarrow \infty \quad (1.1)$$

Alors H_γ est nécessairement l'un des trois types suivant :

Gumbel (type I) :

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

Fréchet (type II) :

$$H_\gamma(x) = H_{1,\gamma}(x) = \begin{cases} 0 & \text{si } x < 0, \\ \exp(-x^{-1/\gamma}) & \text{si } x \geq 0, \text{ Avec } \gamma > 0. \end{cases}$$

Weibull (type III) :

$$H_\gamma(x) = H_{2,\gamma}(x) = \begin{cases} \exp(-(x)^{-\gamma}) & \text{si } x \leq 0, \\ 1 & \text{si } x > 0, \text{ Avec } \gamma < 0 . \end{cases}$$

- 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.

Remarque 5 1. La fonction de répartition H est appelée loi des valeurs extrêmes. (que l'on note (EVD) "Extreme Value Distribution"). Le paramètre est un paramètre γ 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".

Fonctions à variation régulière

Définition 6 Une fonction mesurable $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est à variation régulière à l'infini si et seulement si, il existe un réel tel que, pour tout $x > 0$

$$\lim_{t \rightarrow \infty} \frac{g(tx)}{g(t)} = x^\alpha, \alpha > 0$$

Et on note $g \in VR_\alpha$, α est appelé indice (ou exposant) de la fonction à variation régulière. Dans le cas particulier où $\alpha = 0$, on dit que g est à variation lente à l'infini, c'est à dire,

$$\lim_{t \rightarrow \infty} \frac{g(tx)}{g(t)} = 1.$$

Les fonctions à variation lente sont génériquement notées $\ell(x)$. Pour toute fonction à variation lente ℓ l'infini, on a :

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

1.1.4 Distribution des valeurs extrêmes généralisée

Pour faciliter le travail avec les trois distribution limites, Jenkinson-VonMises (1955) 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 Extrême Value Distribution). Elle est simplement une reparamétrisation des distributions apparaissant dans le théorème 4.

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

Avec $\mu \in \mathbb{R}, \gamma > 0$.

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

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

Où γ est le paramètres de forme. Nous exprimons les trois distributions des valeur extrêmes $H_{0,\gamma}(x), H_{1,\gamma}(x)$ et $H_{2,\gamma}(x)$ en termes de GEVD H_{γ} . La Figure 1.2 ci-dessous illustre le comportement de GEVD standard :

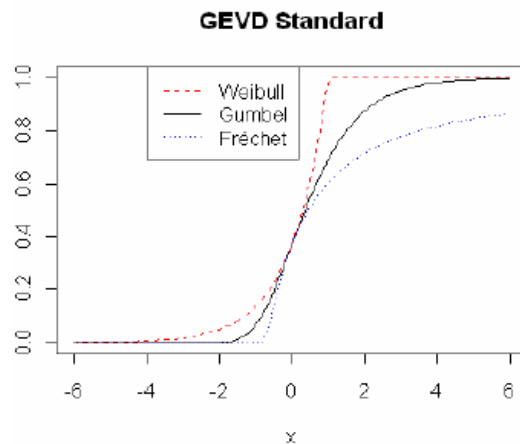


FIGURE 1.2 : Distributions standard des valeurs extrêmes.

1.1.5 Caractérisation des Domaines d'attraction

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

Les Figures 1.3 et 1.4 ci-dessous illustrent le comportement de différentes distributions GEV correspondant à différentes valeurs de γ . Les tableaux 1.3 1.2, 1.1 donnent quelques lois et leur domaine d'attraction.

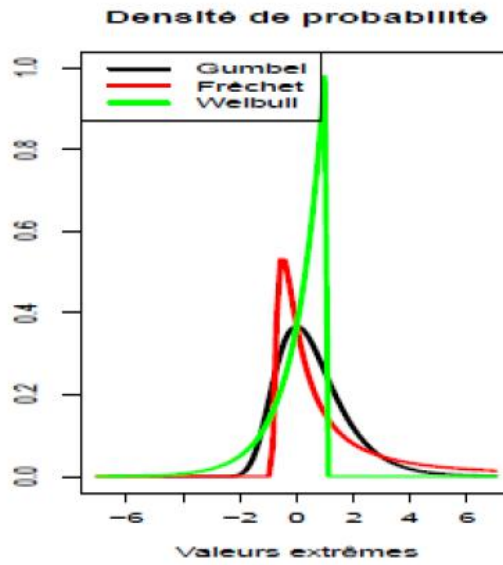


FIGURE 1.3 : Représentation de la densité de probabilité : Gumbel ($\gamma = 0$), Fréchet($\gamma = 1$), Weibull ($\gamma = -1$)

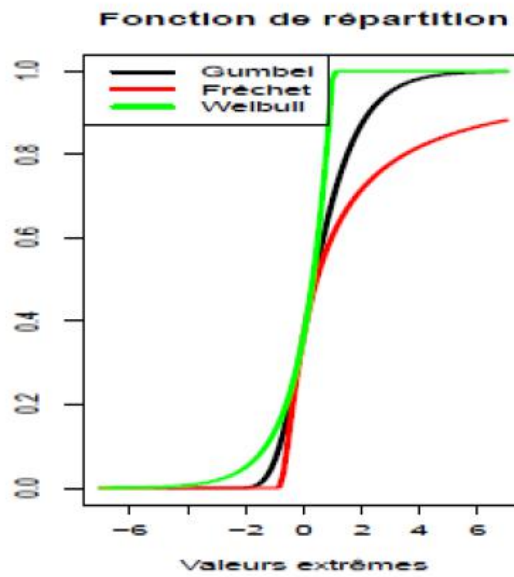


FIGURE 1.4 : Représentation de la fonction de répartition Gumbel ($\gamma = 0$), Fréchet ($\gamma = -1$), Weibull ($\gamma = 1$)

Distribution	La queue $1 - F$ ou la densité f	γ
Burr(β, τ, λ), $\beta > 0, \tau > 0, \lambda > 0$	$1 - F(x) = \left(\frac{\beta}{\beta + x^\tau}\right)^\lambda$	$\frac{1}{\lambda\tau}$
Fréchet($\frac{1}{\alpha}$), $\alpha > 0$	$1 - \exp(-x^{-\alpha})$	$\frac{1}{\alpha}$
Loggamma(m, λ), $m > 0, \lambda > 0$	$f(x) = \frac{\lambda^m}{\Gamma(m)} \int_x^\infty (\log(u))^{m-1} u^{-(\beta+1)} du$	$\frac{1}{\lambda}$
Log-logistic(β, α), $\beta > 0, \alpha > 0$	$1 - F(x) = \frac{1}{1 + \beta x^\alpha}$	$\frac{1}{\alpha}$
Pareto(α), $\alpha > 0$	$1 - F(x) = x^{-\alpha}$	$\frac{1}{\alpha}$

TABLE 1.1 : Quelques distributions appartenant au domaine d'attraction de Fréchet

Distribution	La queue $1 - F$ ou la densité f	γ
Gamma(m, λ), $m \in \mathbb{N}, \lambda > 0$	$f(x) = \frac{\lambda^m}{\Gamma(m)} \int_x^\infty u^{m-1} \exp(-\lambda u) du$	0
Gumble(μ, β), $\mu \in \mathbb{R}, \beta > 0$	$f(x) = \exp(-\exp(-\frac{x-\mu}{\beta}))$	0
Logistic	$1 - F(x) = \frac{2}{1 + \exp(x)}$	0
Log normale(μ, σ), $\mu \in \mathbb{R}, \sigma > 0$	$f(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{1}{u} \exp(-\frac{1}{2\sigma^2}(\log u - \mu)^2) du$	0
Weibull (λ, τ), $\lambda > 0, \tau > 0$	$1 - F(x) = \exp(-\lambda x^\tau)$	0

TABLE 1.2 : Quelques distributions appartenant au domaine d'attraction de Gumbel

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

TABLE 1.3 : Quelques distributions appartenant au domaine d'attraction de Weibull

1.1.6 Estimateur de Hill $\hat{\gamma}_{k_n}^{Hill}$

Soient X_1, \dots, X_n des variables aléatoires iid de fonction de répartition commune F et soit $X_{1,n}, \dots, X_{n,n}$ les statistiques d'ordre correspondantes. On suppose que F satisfait la condition du domaine d'attraction de Fréchet, en terme de variations régulière ceci est équivalent à : pour $\gamma > 0$

$$1 - F(x) = x^{-1/\gamma} \ell(x), x > 0$$

La méthode du maximum de vraisemblance est la plus populaire et qui sous certaines conditions est la plus efficace. On se sert des statistiques d'ordre supérieures à un certain seuil u pour ne garder que les observations les plus grandes de façon à ce qu'elles suivent approximativement une distribution de Pareto. D'autre part on peut écrire :

$$\frac{1 - F(x)}{1 - F(u)} = \left(\frac{x}{u}\right)^{-1/\gamma}, \quad x > u, \quad u \in R.$$

Si $(n - k)$ désigne le nombre de statistiques d'ordre qui dépassent le seuil u , alors on estime u par $X_{n-k,n}$. En utilisant les plus grande statistiques d'ordre $X_{n-k+1,n}, \dots, X_{n,n}$ la fonction de log-vraisemblance sera alors :

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

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

$$\hat{\gamma}_{k_n}^{Hill} = \frac{1}{K} \sum_{i=1}^k \log X_{n-i+1,n} - \log X_{n-k,n} \\ = \frac{1}{K} \sum_{i=1}^k i (\log X_{n-i+1,n} - \log X_{n-i,n}).$$

Plusieurs chercheurs ont essayé de déterminer les propriétés asymptotiques de l'estimateur de Hill. Mason (1982) a prouvé la consistance faible de l'estimateur de *Hill* pour toute suite $k = k(n)$ satisfaisant $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$ appelée suite intermédiaire d'entiers. La condition $k \rightarrow \infty$ assure que la taille de statistiques d'ordre k est assez grande afin d'obtenir des estimateurs stables . Par contre, la condition $k/n \rightarrow 0$ permet de rester dans la queue de distribution. Davis et Resnick (1984) ont proposé sa normalité asymptotique sous les conditions de Von Mises, Csörgö et Mason (1985) ont présenté sa normalité asymptotique en introduisant l'approximation des processus empiriques par les ponts browniens. Dans cette même lancée Resnick et de Haan (1998) ont montré cette propriété asymptotique. Rappelons à présent les propriétés asymptotiques de l'estimateur. Pour cela, nous allons commencer par les conditions du premier et du second ordre avec les fonctions quantiles définies ainsi :

$$Q(s) = F^{\leftarrow}(s) := \inf \{x \in \mathbb{R} : F(x) \geq s\}, 0 < s < 1.$$

et

$$U(t) := Q(1 - 1/t) = (1/\bar{F})^{\leftarrow}(t), 0 < t < 1.$$

Proposition 7 (Conditions du premier ordre, de Haan et Ferreira (2006)) Les assertions suivantes sont équivalentes :

1. F est à queue lourde

$$F \in D.A(\text{Frchet}), \gamma > 0$$

2. $1 - F$ est une fonction à variation régulière à l'infini d'indice $-1/\gamma$

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

3. $Q(1 - s)$ est une fonction à variations régulières à zéro d'indice $-1/\gamma$

$$\lim_{s \rightarrow 0} \frac{Q(1 - sx)}{Q(1 - s)} = x^{-\gamma}, x \gg 0$$

4. U est une fonction à variation régulière à l'infini d'indice γ

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma, \text{ pour } x \gg 0,$$

Proposition 8 (Conditions du second ordre de Haan et Ferreira(2006)) Une fonction de répartition $F(\cdot) \in D.A(\text{Fréchet}), \gamma > 0$, admet une condition du second ordre à l'infini si elle satisfait à l'une des assertions suivantes :

1. S'il existe un paramètre $\rho \leq 0$, et une fonction $A_1(\cdot)$ qui tend vers 0 (ne change pas de signe à l'infini) définie par, $\forall x > 0$

$$\lim_{t \rightarrow \infty} \frac{\frac{(1-F(tx))}{(1-F(t))} - x^{-1/\gamma}}{A_1(t)} = x^{-1/\gamma} \frac{x^\rho - 1}{\rho}, x \gg 0$$

2. S'il existe un paramètre $\rho \leq 0$ et une fonction $A_2(\cdot)$ qui tend vers 0 (ne change pas de signe à zéro) définie par, $\forall x > 0$

$$\lim_{s \rightarrow 0} \frac{\frac{(1-Q(sx))}{(1-Q(s))} - x^{-\gamma}}{A_2(t)} = x^{-\gamma} \frac{x^\rho - 1}{\rho}, x \gg 0$$

3. S'il existe un paramètre $\rho \leq 0$, et une fonction $A(\cdot)$ qui tend vers 0 (ne change pas de signe à l'infini) définie par, $\forall x > 0$

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}, x \geq 0$$

si $\rho = 0$, on remplace $(x^\rho - 1)/\rho$ par $\log(x)$ Les fonctions $A(\cdot), A_1(\cdot), A_2(\cdot)$ sont à variations régulières à l'infini d'indices respectifs $\rho, \rho/\gamma$, et $-\rho$, avec $A_1(t) = A(1/(1 - F(t)))$ et $A_2(s) = A(1/s)$. Ces deux conditions ont permis de déterminer les propriétés asymptotiques de certains estimateurs de l'indice des valeurs extrêmes.

Théorème 9 (Propriétés asymptotiques de l'estimateur de Hill) Soit $k_n, n \geq 1$ une suite d'entiers telle que $1 \leq k_n \leq n, k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$ quand $n \rightarrow \infty$

(a) **Consistance faible** : $\hat{\gamma}_{k_n}^{Hill}$ converge en probabilité vers γ .

(b) **Consistance forte** : Si de plus $k_n/\log n \rightarrow \infty$ quant $n \rightarrow \infty$ alors $\hat{\gamma}_{k_n}^{Hill}$ converge presque sûrement vers γ .

(c) **Normalité asymptotique** : Si la condition

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx)}{U(t)} - x^\gamma}{A(t)} = x^\gamma \frac{x^\rho - 1}{\rho}$$

est satisfaite avec $\sqrt{k_n}A(n/k_n) \rightarrow \lambda \in \mathbb{R}$ quand $n \rightarrow \infty$, alors

$$\sqrt{k_n}(\hat{\gamma}_{k_n}^{Hill} - \gamma) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\frac{\lambda}{1 - \rho}, \gamma^2\right)$$

Programme sous R calculant directement la valeur de l'estimateur $\hat{\gamma}_{k_n}^{hill}$:

```
X<-sort(X) # on ordonne L'échantillon des observations  $X_i$ 
y=sort(abs(X))
z=log10(y)
k<-seq(1,n-1)
hill<-seq(1,n-1)
for(k in 1 :n-1)
{ l<-seq(1,k)
hill[k]<-((1/k)*sum(z[n-l+1]))-z[n-k]} # estimateur de Hill vs k
```



```
plot(k, hill, ylim=c(0.05, 0.2), type="l", col="6",lwd="1",ylab=expression (paste(plain(hat(gamma)^h)))
abline(h=gamma1, col="red", lwd="2")
```

1.2 Quelques généralités sur la troncature

1.2.1 Données de survie

L'analyse de survie, autrement dit la modélisation du temps de survenue d'un événement, apporte un outil principal d'évaluation théorique et pratique. L'analyse de ce type de données possède deux particularités intrinsèques, d'une part, celle-ci ne concerne que des variables aléatoires positives et d'autre part, la présence de données non complètement observées comme nous l'expliquons ci-dessous. Désignons par X une variable d'intérêt, c'est à dire une variable aléatoire positive décrivant le temps qui s'écoule entre deux événements par exemple

1. En médecine : durée de guérison d'un patient, durée de rémission d'un malade.
2. En économie : durée de chômage...
3. En éducation : durée d'apprentissage d'une langue...
4. En assurance : durée de vie d'un contrat qui peut être définie comme la différence entre la date de résiliation et la date de création du contrat.

Définition 10 *La « durée de vie » d'un individu est une variable aléatoire (v.a.) X positive et continue. Sa fonction de répartition*

$$F(t) = \mathbb{P}(X \leq t)$$

est la probabilité que l'événement se produise entre 0 et t .

Définition 11 *La fonction de risque instantané, pour t fixé représente la probabilité de mourir dans un petit intervalle de temps après t , conditionnellement au fait d'avoir survécu jusqu'au temps t (c'est-à-dire le risque de mort instantané pour ceux qui ont survécu) :*

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(t \leq x < t + x | x \geq t)}{h} = \frac{f(t)}{s(t)}$$

où f est la densité de probabilité de X .

Définition 12 La fonction de risque cumulé est l'intégrale du risque instantané définie par :

$$\Lambda = \int_0^t \lambda(u) du.$$

Définition 13 La troncature est une variante de censure ce qui se produit à la conception de l'étude lorsque la nature de l'observation incomplète est due à un inhérent de processus de sélection systématique. On a trois types de troncature, comme suit

1.2.2 Données tronquées

Une autre situation dans laquelle les données incomplètes apparaissent est celle des données tronquées. Les troncatures diffèrent des censures au sens où elles concernent l'échantillonnage lui-même. Une observation est dite tronquée si elle est conditionnelle à un autre événement. On dit que la variable X de durée de vie est tronquée si X n'est observable que sous une certaine condition dépendante de la valeur de X .

Données tronquées à droite

Définition 14 Soit Y une v.a. indépendante de X , on dit qu'il y a troncature à droite lorsque X n'est observable que si $X \leq Y$.

Exemple 1 Si la durée de vie d'une population est étudiée à partir d'une cohorte tirée au sort dans cette population, seule la survie des sujets vivants à l'inclusion pourra être étudiée (il y a troncature à gauche car seuls les sujets ayant survécu jusqu'à la date d'inclusion dans la cohorte sont observables).

Données tronquées à gauche

De même, il y a troncature à gauche lorsque X n'est observable que si $X > Y$.

Données tronquées par intervalle

Définition 15 *Quand une durée est tronquée à droite et à gauche, on dit qu'elle est tronquée par intervalle.*

Exemple 2 *On rencontre ce type de troncature lors de l'étude des patients d'un registre : les patients diagnostiqués avant la mise en place du registre ou répertoriés après la consultation du registre ne seront pas inclus dans l'étude.*

1.3 Le Modèle POT

La méthode des excès encore appelé POT (Peaks Over Threshold) développée dans les années 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. On appelle right-end point ou point terminal de la fonction répartition F , le point x_F tel que :

$$x_F = \{x, F(x) < 1\}$$

La figure ci-dessous représenté la principe de la méthode des excès

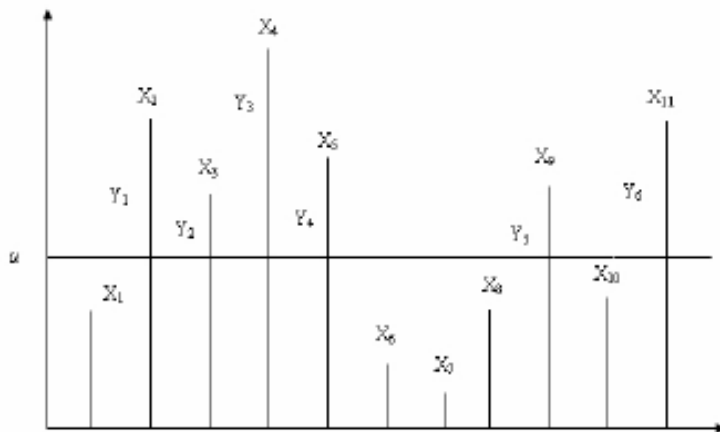


FIGURE 1.5 : Observations X_1, \dots, X_{11} et excès Y_1, \dots, Y_6 au-delà du seuil u .

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 $X > u$:

1.3.1 Loi des excès

Définition 16 *La fonction de répartition des excès au-delà d'un seuil u pour une variable aléatoire X avec fonction de répartition F est définie par*

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(y - u) - F(u)}{1 - F(u)}. \quad (1.2)$$

Cette distribution des excès est un outil largement utilisé en hydrologie, en analyse de survie et en sciences de l'environnement et d'une importance majeure dans la pratique actuarielle reconnue (voir par exemple Hogg et Klugman, 1984; Embrechts et al. 2005). Notamment en matière de réassurance, si on prend par exemple un traité en excédent de perte (excess-of-loss) avec une rétention u sur une demande particulière dans le portefeuille, le réassureur doit payer un montant aléatoire $X - u$, mais seulement si $X > u$. Quand un actuaire veut se prononcer sur un niveau de priorité grâce à la simulation, il a besoin de calculer le montant qui devrait être payé par le client quand un niveau donné u est choisi. Voilà donc une première étape importante pour décider sur la prime. Par exemple, le principe de la prime nette dépend de la taille moyenne des réclamations $E(x)$, où pour le dépassement, l'actuaire a besoin de calculer la fonction moyenne des excès. Cette dernière consistera la base de la construction du graphique mean-excess plot. On donnera plus de détails sur cette fonction dans ce qui suit. Ainsi, on présente tout d'abord, le modèle POT et principalement la fonction GPD qui nous aidera pour examiner la queue de la distribution de l'échantillon réelle en question et notamment la détermination du seuil u .

Le modèle à dépassement de seuil (Peaks Over Threshold)

Au delà d'un seuil u par une classe de distributions de Pareto généralisée GPD dépendant des deux paramètres γ et $\sigma(u)$, étant appropriée pour décrire les propriétés statistiques des excès Une variable aléatoire X a une distribution de Pareto généralisée (GPD) si elle a une fonction de répartition de la forme

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

où $\sigma > 0$, et $x > 0$ quand $\gamma > 0$ et $0 \leq x < -\sigma/\gamma$ si $\gamma < 0$: Les paramètres et

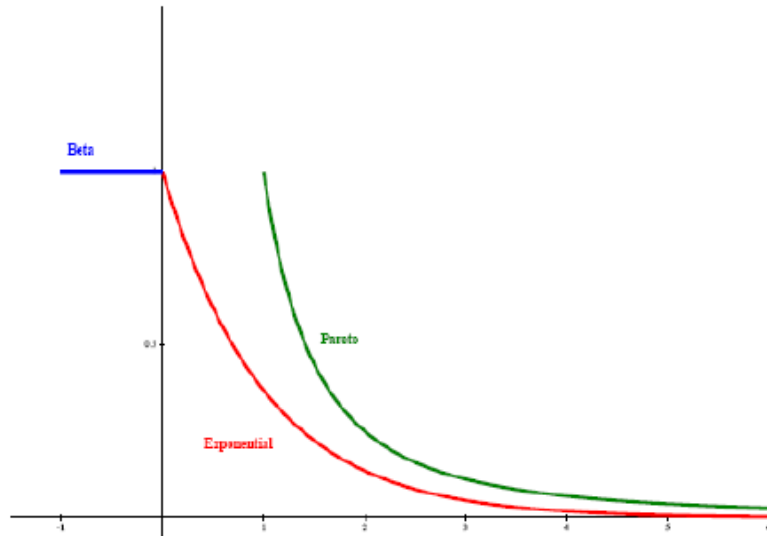


FIGURE 1.6 : Graphe de densité, Pareto, Beta et Exponentielle

σ sont désignés respectivement comme les paramètres de la forme et d'échelle.

On a les correspondances suivantes :

si $\gamma > 0$	$G_{\gamma,\sigma} \longrightarrow$	loi Pareto.
si $\gamma < 0$	$G_{\gamma,\sigma} \longrightarrow$	loi béta.
si $\gamma = 0$	$G_{\gamma,\sigma} \longrightarrow$	loi exponentielle.

Le théorème Pickands-Balkema-de Haan (1975) (Embrechts et al. 2005, Theorem 7.20, page 277) fournit la justification théorique de la centralité de la classe des distributions GPD pour la modélisation POT.

Théorème 17 (*Balkema-de Haan-Pickands*)(1975)

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 :

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\gamma,\sigma(u)}(x)| = 0$$

où σ est une fonction strictement positive, si $F \in D(F)$

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.

1.4 Recherche du seuil u

Le choix du seuil est primordial dans cette approche : afin de pouvoir approcher la queue de distribution par une loi GPD, le seuil doit être suffisamment élevé. Cependant, un seuil trop grand peut réduire le nombre de données disponibles. Nous choisissons d'implémenter plusieurs méthodes afin de sélectionner le seuil adéquat, que nous présentons ici.

Fonction moyenne des excès Afin de déterminer ce seuil, nous pouvons faire appel à la fonction moyenne des excès (ou mean excess function) qui se caractérise par sa linéarité en u pour les lois GPD :

$$e(u) = E[X - u | X > u], \quad 0 < u < \infty$$

Il s'agit donc de chercher le seuil u à partir duquel la fonction est linéaire.

Hill Plot.

Pour cette méthode il faut calculer un estimateur de $\frac{1}{\xi}$ (valable pour $\xi > 0$). Lorsque nous utilisons tout l'échantillon pour cela, sa valeur est biaisée car seuls les dépassements au-dessus d'un seuil élevé suivent une loi GPD. L'idée est de calculer cet estimateur avec un grand échantillon et d'enlever au fur et à mesure les observations les plus petites, jusqu'à ce que l'estimateur se stabilise. Le point où il se stabilise est notre seuil.

Cunnane Test.

Asymptotiquement, le nombre d'excès au-dessus du seuil u suit une loi de Poisson. Nous utilisons ce résultat pour déterminer u , en faisant aussi l'approximation qu'une loi de Poisson est proche d'une loi normale pour un paramètre suffisamment élevé et en testant la proximité entre la loi du nombre d'excès et une loi normale par une statistique du X^2 .

Kolmogorov-Smirnov Test.

Cette méthode est proche de la précédente, à la différence qu'elle ne considère pas le nombre d'excès mais les excès en eux-mêmes, censés suivre une loi de Pareto généralisée. De plus, c'est la statistique de Kolmogorov-Smirnov qui est appliquée).

ESTIMATION DE LA PRIME SOUS DES DONNÉES TRONQUÉES

Des observations extrêmes tronquées sont rencontrées dans l'assurance par exemple, où l'accent est mis sur la modélisation des événements extrêmes, est de protéger une compagnie d'assurance, l'estimation du prix d'un risque est un problème actuariel très important. Une intention particulière est portée sur la prise en compte des valeurs extrêmes sur laquelle repose le calcul des exigences de capitaux propres. Le prix doit refléter la propriété de la distribution de la variable aléatoire décrivant la perte correspondante.

2.1 Introduction sur la prime

La prime de risque des actions, c'est l'écart de rentabilité entre le rendement attendu d'un placement en action (aussi appelé le taux risqué) et le rendement d'un placement sans risque (taux sans risque). L'investisseur n'acceptera de prendre du risque que s'il en espère une meilleure rémunération.

2.1.1 Mesures de risque

Définition 18 (*Mesures de risque*). On appelle mesure de risque toute application ρ associant un risque X à un réel $\rho(x) \in \mathbb{R}_+(+\infty)$.

Définition 19 (*Chargement de sécurité*). Une mesure de risque ρ contient un chargement de sécurité si pour tout risque X on a, $\rho(X) > E[X]$.

Définition 20 (*Constance*). Une mesure de risque ρ contient pas un chargement de risque injustifié si pour toute constante réelle a , $\rho(a) = a$.

Définition 21 (*Non chargement excessif*). Une mesure de risque ρ ne dépasse pas la perte maximale si, $\rho(X) \leq \max[X]$.

Définition 22 (*Objectivité*). Une mesure de risque ρ est dite objective lorsqu'elle dépend que de F et non de X elle-même.

Définition 23 (*L'invariance en loi*). Une mesure de risque ρ est dite invariante en loi si pour tout X et Y telle que $X \stackrel{L}{=} Y$ alors, $\rho(X) = \rho(Y)$.

2.1.2 Principes de calcul de prime

L'activité d'assurance repose sur le concept de transfert de risque : moyennant une prime, l'assuré, se protège d'un aléatoire financier. Mesurer le risque assuré s'avère donc inévitable puisque cette information est nécessaire dans le cadre de la tarification pour déterminer les chargements de sécurité à ajouter à la prime pure et dans une approche de solvabilité pour déterminer le niveau des réserves et des fonds propres dont doit disposer l'assureur pour être solvable. Mesurer le risque pour les compagnies d'assurance c'est la détermination de ses besoin en fonds propre, d'où l'importance d'utiliser des mesures de risques efficaces et efficientes qui peuvent être adaptées dans le secteur d'assurance.

Définition 24 (*Principe de calcul de prime*). Le principe de calcul de prime est une règle qui assigne un nombre non négatif Π (la prime) à n'importe quel risque donné X (une variable aléatoire positive) i.e. déterminer la prime comme un fonctionnel en assignant une valeur $\Pi(F_X) \in \mathbb{R} \cup [\infty]$ la distribution de risque F_X .

Propriétés de meilleures primes Soit X, Y, Z des risques assurables arbitraires pour lesquelles la prime $\Pi(X)$ sera mieux définie.

1. *Aucun chargement de sécurité injustifié* : si, pour tout constant $a \geq 0$, $\Pi(a) = 0$.
2. *Proportionnalité* : si, pour tout constant $a \geq 0$, $\Pi(aX) = a\Pi(X)$.
3. *Sous-additivité* : si, $\Pi(X + Y) = \Pi(X) + \Pi(Y)$.

4. *Consistance* : si, pour tout $a \geq 0$, $\Pi(X + a) = \Pi(X) + a$.
5. *Conservation d'ordre stochastique* : si, $X < Y$ implique que $\Pi(X) \leq \Pi(Y)$.
6. *Compatibilité sous mélange* : pour tout $p \in [0,1]$ et pour tout Z , $\Pi(X) = \Pi(Y)$ implique $\Pi(pF_X + (1 - p)F_Z) = \Pi(pF_Y + (1 - p)F_Z)$.

Quelques principes de base. Parmi les principes de calcul de prime les plus simple on cite :

- principe de la prime nette : pour tout $a > 0$, $\Pi(X) = E(X)$
- principe de la valeur espérée : pour tout $a > 0$, $\Pi(X) = (1 + a)E(X)$
- principe de l'écart-type : pour tout $a > 0$, $\Pi(X) = E(X) + a\sqrt{\text{var}(X)}$
- principe de la variance : pour tout $a > 0$, $\Pi(X) = E(X) + a\text{var}(X)$
- principe de l'utilité nulle : soit $v(x)(x \in \mathbb{R})$ une fonction deux fois différentiable avec $v_0(x) > 0$. Dans ce cas et $w(X)$ est la solution de l'équation $E[v(\Pi(X) - X)] = v(0)$
- principe de la perte maximale : pour $p > 0$, $q = 1 - p$ on a $\Pi(X) = pE(X) + q\max(X)$ où, $\max(X)$ note le point terminal à droite du rang x .
- principe de quantile : $\Pi(X) = F_X^{-1}(1 - \varepsilon)$: i.e., la prime la plus petite tel que la probabilité d'une perte est à la plus part ε .
- principe de risque ajusté : $\Pi(X) = \int_0^\infty (1 - F_X(x))^{1/\rho} dx$, ou $\rho > 0$ appelé coefficient de distorsion.

2.1.3 Estimateur de la prime de Necir et al. (2015) $\hat{\Pi}_n$

Nous nous concentrons dans nos travaux sur l'approche des quantiles extrêmes pour l'estimation de la prime. Nous présenterons celui de Necir et al. (2015), qui est un basé sur l'estimateur de quantiles extrêmes de Weissman (1978) dans le but de calculer estimateur de la prime. Des études empiriques ont montré que les données tronquées présentent des distributions à queues lourdes ou de type pareto comme les distributions de Burr, Student, Lévy-satble et log-gamma. Ces dernières sont connues pour être des modèles appropriés

pour modéliser les grandes réclamations d'assurance, les fluctuations des prix, Reiss et Thomas (2007) et Rolski et al. (1999). Les plus populaires sont la prime nette, la variance et des primes à base d'écart type, la valeur exposée au risque, l'espérance conditionnelle, la prime de risque lorsque les pertes suivent des distributions à queue lourde, sous certaines conditions Necir et al. (2015) ont établi sa normalité asymptotique. Le calcul de cet estimateur dépend du choix du nombre des valeurs extrêmes K utilisés dans l'estimation de l'indice de queue .

Soit X_1, \dots, X_n représentent le montant de sinistre individuel d'un assuré à queue lourde avec moyenne finie de distribution de Pareto, avec indice de queue supérieur ou égal à 1, n'a pas de moyenne finie. Par conséquent, en supposant que $\mathbb{E}[X]$ existe nécessairement implique que $\gamma_1 < 1$. Dans le traité de réassurance excédentaire il cède à la société, pour couvrir les réclamations qui dépassent un nombre (élevé) $u \geq 0$, a rappeler le niveau de rétention, le réassurer paie la partie $(X_i - u)_+ = \max(0, X_i - u)$ de chaque revendication dépassant u . La prime nette défini sur l'intervalle $[u, +\infty]$ est définie par :

$$\Pi = \mathbb{E}[(X_i - u)_+] = \int_u^{+\infty} (1 - \mathbf{F}(x)) dx \quad (2.1)$$

pour calculer l'expression de la prime, on utilise le Théorème de Karamata suivant :

Théorème 25 (Théorème de Karamata :) . nous Supposons $\bar{\mathbf{F}} \in RV_{-1/\gamma_1}$, il existe $t_0 > 0$ tel que $\bar{\mathbf{F}}(t)$ est positive et délimitée localement pour $t \geq t_0$, si $-1/\gamma_1 > -1$ alors

$$\lim_{t \rightarrow \infty} \frac{t\bar{\mathbf{F}}(t)}{\int_{t_0}^t \bar{\mathbf{F}}(s) ds} = \frac{-1}{\gamma_1} + 1. \quad (2.2)$$

si $-1/\gamma_1 < -1$, ou $-1/\gamma_1 = -1$ et $\int_0^{\infty} \bar{\mathbf{F}}(s) ds < \infty$ alors

$$\lim_{t \rightarrow \infty} \frac{t\bar{\mathbf{F}}(t)}{\int_t^{\infty} \bar{\mathbf{F}}(s) ds} = \frac{1}{\gamma_1} - 1. \quad (2.3)$$

inversement si 2.2 est remplis avec $-1 < -1/\gamma_1 < \infty$, alors $\bar{\mathbf{F}} \in RV_{-1/\gamma_1}$, si 2.3 est remplis avec $-\infty < -1/\gamma_1 < -1$.

Preuve. voir de *Haan et Ferreira* (2006) page 364. ■

d'une autre manière, nous pouvons calculer par intégration directe, pour

$$1 - \mathbf{F}(x) = x^{-1/\gamma_1}, \quad \gamma_1 > 0.$$

Nous obtenons

$$\begin{aligned} \Pi &= \int_u^{+\infty} x^{-1/\gamma_1} dx \\ &= \left[\frac{1}{-1/\gamma_1 + 1} x^{-1/\gamma_1 + 1} \right]_u^{+\infty} \\ &= \frac{1}{-1/\gamma_1 + 1} (0 - u^{-1/\gamma_1 + 1}) \end{aligned}$$

$$\Pi = \frac{\gamma_1}{1 - \gamma_1} u \bar{\mathbf{F}}(u) \quad , \quad 0 < \gamma_1 < 1 \quad . \quad (2.4)$$

les distributions marginales des v.a's X et Y respectivement, notés par F et G sont données par :

$$F(x) = p^{-1} \int_0^x \bar{\mathbf{G}}(z) d\mathbf{F}(z)$$

et

$$G(y) = p^{-1} \int_0^y \mathbf{F}(z) d\mathbf{G}(z).$$

Comme on le voit, un estimateur semi-paramétrique pour \mathbf{F} est nécessaire pour estimer la prime Π . Pour ce, nous définissons $C(x)$:

$$C(x) = \mathbf{P}(\mathbf{X} \leq x \leq \mathbf{Y} | \mathbf{X} \leq \mathbf{Y}), \quad (2.5)$$

cette quantité C qui joue un rôle important dans l'analyse des données tronquées. En d'autres termes nous avons,

$$\begin{aligned} C(x) &= P^{-1} \mathbf{F}(x) \bar{\mathbf{G}}(x) \\ &= F(x) - G(x) \\ &= \bar{G}(x) - \bar{F}(x) \end{aligned}$$

nous savons que puisque \mathbf{F} et \mathbf{G} sont lourdes, leurs points d'extrémité droit sont infinies et donc elles sont égales. Par conséquent. D'après de Woodrooffe (1985), les fonctions \mathbf{F} , F et C sont liés par

$$C(x)d\mathbf{F}(x) = \mathbf{F}(x)dF(x)$$

Connue sous le nom d'équation d'auto-cohérence équation (voir, par exemple, *Strzalkowska-Kominiak et Stute, 2009*), dont la solution est

$$\mathbf{F}(x) = \exp - \Lambda(x)$$

où, $\Lambda(x) = \int_x^\infty dF(x)/C(x)$ en remplaçant F et C par leur empirique respectivement :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x)$$

et

$$C_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i \leq x \leq Y_i} \quad (2.6)$$

Nous obtenons l'estimateur connu de limite produit de Lynden-Bell (1971) de \mathbf{F} ,

$$\mathbf{F}_n(x) = \exp - \Lambda_n$$

Où, $\Lambda_n(x) = \int_x^\infty dF_n(x)/C_n(x)$ s'il n'y a pas de lien, \mathbf{F}_n peut être mis sous la forme :

$$\mathbf{F}_n(x) = \prod_{X_{i:n} > x} \left(1 - \frac{1}{nC_n(X_{i:n})} \right) \quad (2.7)$$

Puisque $\bar{\mathbf{F}}$ varie régulièrement avec l'indice à l'infini $-1/\gamma_1$, puis

$$\bar{\mathbf{F}}(x) \sim \bar{\mathbf{F}}(U(n/k))(x/U(n/k))^{-1/\gamma_1}, \text{ quand } x \rightarrow \infty.$$

Cela nous amène à déduire un estimateur de type Weissman (Weissman, 1978)

$$\hat{\bar{\mathbf{F}}}(x) = \left(\frac{x}{X_{n-k:n}} \right)^{-1/\hat{\gamma}_1} \bar{\mathbf{F}}_n(X_{n-k:n}) \quad (2.8)$$

Pour la queue de distribution $\bar{\mathbf{F}}$ avec des données tronquées. Notons que,

$$\mathbf{F}_n(X_{n-k:n}) = \prod_{i=n-k+1}^n \left(1 - \frac{1}{nC_n(X_{i:n})} \right) \quad (2.9)$$

Ainsi, l'estimateur de queue de distribution est sous la forme suivante :

$$\hat{\mathbf{F}}(x) = \left(\frac{x}{X_{n-k:n}} \right)^{-1/\hat{\gamma}_1} \left\{ 1 - \prod_{i=1}^k \left(1 - \frac{1}{nC_n(X_{n-i+1:n})} \right) \right\}.$$

Par conséquent, nous définissons un estimateur de la prime comme suit :

$$\hat{\Pi}_n = \frac{\hat{\gamma}_1}{1 - \hat{\gamma}_1} X_{n-k:n} \left(\frac{u}{X_{n-k:n}} \right)^{-1/\hat{\gamma}_1} \left\{ 1 - \prod_{i=1}^k \left(1 - \frac{1}{nC_n(X_{n-i+1:n})} \right) \right\}$$

le résultat principal de Necir et al. (2015) est donné par le théorème suivant où ils établissent la normalité asymptotique.

Théorème 26 (Normalité asymptotique de $\hat{\Pi}_n(u)$ - Necir et al(2015)) *Supposons que les conditions de second ordre de variation régulière avec $(0 < \gamma_1 < 1$ et $\gamma_1 < \gamma_2$). Soit $k = k_n$ une suite de nombres entiers tels que $k \rightarrow \infty$, $k/n \rightarrow 0$ et $\sqrt{k}\mathbf{A}(U(n/k)) \rightarrow \lambda^* < \infty$. En suite, chaque fois $u/(U(n/k)) \rightarrow a$, nous avons aussi $n \rightarrow \infty$,*

$$\frac{\sqrt{k}(\hat{\Pi}_n - \Pi)}{(u/U(n/k))^{1-\gamma_1}(u/U(n/k))\bar{\mathbf{F}}(u/U(n/k))} \xrightarrow{\mathcal{D}} \mathcal{N} \left(\frac{\lambda^*}{(\gamma_1 - 1 - \tau_1)(\gamma_1 - 1)}, \sigma^{*2} \right)$$

où

$$\sigma^{*2} = \zeta^2 \sigma^2 \frac{\gamma_1^2 \gamma_2^2}{(\gamma_2^2 - \gamma_1^2)(1 - \gamma_1)^2} + 2 \frac{\gamma \gamma_1 \zeta \sigma^*}{1 - \gamma_1}$$

avec σ , ζ et σ^* défini par :

$$\sigma = \int_0^1 \int_0^1 \frac{R(s,t)}{ts} \int_0^1 (R(s,1) - R(1,s)) ds + R(1,1),$$

$$\zeta = ((1 - \gamma_1) \log(a) + \gamma_1) / (\gamma_1(1 - \gamma_1)^2)$$

et

$$\begin{aligned} \sigma^* &= \frac{c}{\gamma \gamma_2} + \frac{c_2}{\gamma_1} \left(R(1,1) - \int_0^1 \frac{R(1,t)}{t} dt \right) \\ &+ \frac{c_2}{\gamma_1 + \gamma_2} \left(\int_0^1 \frac{R(s,1)}{s^{\gamma/\gamma_2+1}} ds - \int_0^1 \int_0^1 \frac{R(s,t)}{ts^{\gamma/\gamma_2+1}} ds dt \right) \end{aligned}$$

où $c := \gamma_1^2/\gamma$, $c_2 := \gamma_1^2/\gamma_2$.

Preuve. Voir Necir et al (2016). ■

Remarque 27 Notez que σ^2 est fini. En effet, le fait que $\ell(x, y)$ une copule fonction de queue, implique que, $\max(x, y) \leq \ell(x, y) \leq x + y$ (voir, par exemple, Gudendorf et Seegers, 2010) et si $R(x, y) = x + y - \ell(x, y)$, alors $0 \leq R(x, y) \leq \min(x, y)$. Il s'ensuit que

$$\int_0^1 \int_0^1 \frac{R(s, t)}{ts} ds dt \leq \int_0^1 \int_0^1 \frac{\min(s, t)}{ts} ds dt = 2$$

Par conséquent $|\sigma| \leq 4$, ce qui donne que $\sigma^2 < \infty$.

2.2 Estimation de la prime avec un seuil fixé

Julien Worms, Rym Worms (2015), ont donnée un estimateur de l'indice des valeurs extrêmes pour les données tronquées à queue lourdes, à droite au hasard tronquée par une autre variable à queue lourde.

2.2.1 Estimateur de l'indice de queue $\hat{\gamma}_1^{(JR)}$

Ce paramètre a été largement étudié et plusieurs bons estimateurs ont été proposés, (voir par exemple Pickands (1975), Hill (1975), Dekkers, Einmahl et de Haan (1989), Csörgö, Deheuvels et Mason (1985), Groeneboom, Lopuhaä, et de Wolf. (2003),...), dont le plus célèbre est celui de Hill (1975). Cet estimateur est basé sur la propriété de la variation régulière dans le cas où la fonction de distribution \mathbf{F} appartient au domaine d'attraction Fréchet (de H_γ avec $\gamma > 0$).

Définition 28 On dit que la fonction de distribution \mathbf{F} est à queue lourde $\bar{\mathbf{F}} = 1 - \mathbf{F}$ si la fonction de queue est à variation régulière à l'infini d'indice $-1/\gamma$, tel que

$$\bar{\mathbf{F}} = x^{-1/\gamma} \ell(x)$$

où γ est l'indice des valeurs extrêmes et ℓ est une fonction à variation lente.

Soit $(\mathbf{X}_i, \mathbf{Y}_i)_{1 \leq i \leq N}$, $N \geq 1$ copies indépendantes d'un couple (\mathbf{X}, \mathbf{Y}) , où \mathbf{X} et \mathbf{Y} sont des variable aléatoire indépendantes positives ayant des fonctions de distribution cumulées respectives \mathbf{F} et \mathbf{G} . Nous supposons que les points d'extrémité de \mathbf{F} et \mathbf{G} sont tous deux égaux à 0 (mais cela n'a aucune influence sur les résultats, puisque seules les données les

plus élevées, des valeurs sont retenues pour l'estimation de la queue). Nous supposons \mathbf{X} et \mathbf{Y} sont à queue lourde, ce qui signifie que $1 - \mathbf{F}$ et $1 - \mathbf{G}$ (également supposés être continues) varient régulièrement avec les indices respectifs $-1/\gamma_1$ et $-1/\gamma_2$ où γ_1 et γ_2 sont > 0 .

Nous n'observons que les couples $(\mathbf{X}_i, \mathbf{Y}_i)$ satisfaisant $\mathbf{X}_i \leq \mathbf{Y}_i$: autrement dit, les données originales \mathbf{X}_i sont aléatoirement tronquée à partir de la droite par \mathbf{Y}_i , et l'échantillon réellement observé est $(X_i, Y_i)_{1 \leq i \leq n}$, où n suit la loi binomial $B(n, p)$ distribution, p dénotant la probabilité (inconnue) de non-troncature $p = \mathbb{P}(X \leq Y)$. Les couples $(X_1, Y_1), \dots, (X_n, Y_n)$ sont indépendants et identiquement distribués.

Remarque 29 pour $\gamma > 0$, on a la limite suivante :

$$\lim_{t \rightarrow \infty} \frac{\int_0^\infty (1 - F(x)) \frac{dx}{x}}{1 - F(t)} = \gamma \quad (2.10)$$

est équivalent à

$$\lim_{t \rightarrow \infty} \mathbb{E}[\log(X) - \log(t) | X > t] = \gamma$$

En fait,

$$\frac{\int_0^\infty (1 - F(x)) \frac{dx}{x}}{1 - F(t)} = \mathbb{E}[\log(X) - \log(t) | X > t]$$

puisque

$$\int_0^\infty (\log(X) - \log(t)) dF(x) = \int_0^\infty (1 - F(x)) \frac{dx}{x}$$

Relation 2.10 Sera la base de la construction de l'estimateur de γ (Section 3.6.2 dans Haan, L. Ferreira, A., 2006., et (1.2.9)) équivalent à

$$\lim_{t \uparrow x^*} \frac{E((X - t)^2 | X > t)}{E^2((X - t) | X > t)} = \lim_{t \uparrow x^*} 2h(t) = 2$$

leur objectif est adapter l'estimateur Hill célèbre dans le contexte de la troncature droite on a

$$\lim_{t \rightarrow \infty} \mathbb{E}[\log(X/t) | X > t] = \frac{1}{\bar{\mathbf{F}}(t)} \int_t^\infty \log(X/t) dF(x) = \gamma_1$$

Si (t_n) est une suite de seuils positifs croissant à l'infini avec n , On peut alors définir une version aléatoire

$$\Phi(x) = (\bar{\mathbf{F}}(t))^{-1} \log(X/t) \mathbb{I}_{x > t}$$

avec un seuil fixée $u = t_n$, l'estimateur empirique de $\Phi(x)$

$$\hat{\Phi}_n(x) = (\bar{\mathbf{F}}_n(t_n))^{-1} \log(X/t_n) \mathbb{I}_{x > t_n}$$

et par conséquent, l'adaptation de l'estimateur *Hill* pour γ_1 est

$$\hat{\gamma}_1^{(JR)} = \int \hat{\Phi}_n(x) d\mathbf{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \hat{\Phi}_n(X_i) \frac{\mathbf{F}_n(X_i)}{C_n(X_i)}$$

qui conduit à l'estimateur de l'indice de valeur extrême

$$\hat{\gamma}_1^{(JR)} = \frac{1}{n\bar{\mathbf{F}}_n(t_n)} \sum_{i=1}^n \log\left(\frac{X_i}{t_n}\right) \frac{\mathbf{F}_n(X_i)}{C_n(X_i)} \mathbb{I}_{X_i > t_n} \quad (2.11)$$

2.2.2 Hypothèses et résultats

La condition de première ordre suppose dans ce travail est la suivante :

$$\bar{\mathbf{F}} \in RV_{-1/\gamma_1} \quad \text{et} \quad \bar{\mathbf{G}} \in RV_{-1/\gamma_2} \quad (2.12)$$

En d'autres termes, nous supposons que la queue de la variable tronconique Y est plus lourde que la queue de la variable X d'intérêt. Cette condition est nécessaire dans les preuves de nos résultats et est due à la présence de l'estimateur de *Lynden Bell*, évalué dans la queue. Notez que cela implique la finitude de l'intégrale $\int_0^\infty dF(x)/\bar{G}(x)$ (Ce qui est une condition suffisante parfois indiquée dans les documents traitant de la normalité asymptotique de \mathbf{F}_n).

De plus, si on note $\ell_{\mathbf{F}}$, la fonction de variation lente associée à \mathbf{F} (i.e $\bar{\mathbf{F}}(x) = x^{-1/\gamma_1} \ell(x)$), la condition de deuxième ordre que nous considérons est la condition SR_2 classique pour $\ell_{\mathbf{F}}$,

$$\forall x > 0, \left(\frac{\ell_{\mathbf{F}}(tx)}{\ell_{\mathbf{F}}(t)} - 1 \right) \stackrel{t \rightarrow 0}{\sim} h_{\rho_1}(x)g(t), \quad (\forall x > 1) \quad (2.13)$$

où g est une fonction mesurable positive, variant lentement avec l'index ρ_1 , et $h_{\rho_1} = \frac{x^{\rho_1-1}}{\rho_1}$ avec $\rho_1 < 0$, ou $h_{\rho_1}(x) = \log(x)$ quand $\rho_1 = 0$.

La première hypothèse sur la séquence de seuil (t_n) , si on note $\bar{H} = \bar{\mathbf{F}}\bar{\mathbf{G}}$ (notez que H est le fonction de distribution de $\min(x, y)$), (t_n) satisfait

$$n\bar{H}(t_n) \xrightarrow{n \rightarrow \infty} +\infty. \quad (2.14)$$

Le résultat de normalité asymptotique exigera alors la condition suivante sur t_n

$$\sqrt{n\bar{H}(t_n)}g(t_n) \xrightarrow{n \rightarrow \infty} \lambda \quad , \quad \text{pour } \lambda > 0 \quad (2.15)$$

Sous ces l'hypothèse, nous avons les résultats asymptotiques des estimateurs.

Théorème 30 *Sous les hypothèses 2.13, 2.14, et 2.15*

$$\sqrt{n\bar{H}(t_n)}(\hat{\gamma}_1^{(JR)} - \gamma_1) \xrightarrow{\mathcal{L}} \mathcal{N}(\lambda m, s^2)$$

La notation $\mathcal{N}(\lambda m, s^2)$ désigne la distribution normale de moyenne λm et de variance s^2 , où

$$m = \begin{cases} \frac{\gamma_1^2}{1-\gamma_1\rho_1} & \text{si } \rho_1 < 0, \\ \gamma_1^2 & \text{si } \rho_1 = 0. \end{cases}$$

et

$$s^2 = p\gamma_1^2 \left(1 + \left(\frac{\gamma_1}{\gamma_2} \right)^2 \right) \left(1 - \frac{\gamma_1}{\gamma_2} \right)^{-3}$$

Preuve. voir Julien Worms et Rym Worms(2015). ■

Passons maintenant aux résultats concernant l'estimateur de quantile extrême. Supposons que la suit (p_n) est l'ordres du quantile, tendant vers 0

$$\bar{\mathbf{F}}(t_n)/p_n \xrightarrow{n \rightarrow \infty} +\infty \quad (2.16)$$

2.2.3 Estimation des quantiles extrêmes x_p

Définition 31 (*Quantile extrême*) soit le X_1, \dots, X_n , n variables aléatoires i.i.d de fonction de répartition caommun \mathbf{F} . On définit quantile extrême par :

$$x_{p_n} = \mathbf{F}^{\leftarrow}(1 - p_n) \quad (2.17)$$

avec $p_n \rightarrow 0$, $np_n \rightarrow c$ pour $n \rightarrow \infty$, $c > 0$.

Méthodes d'estimation d'un quantile extrême

Plusieurs méthodes d'estimation d'un quantile extrême ont été proposées dans la littérature : principalement la méthode des Blocs (Block Maxima notée par BM), qui modélise la distribution des extrêmes par la loi des extrêmes généralisées (GEV) et la méthode POT qui modélise la distribution des excès au-dessus d'un seuil élevé par la distribution de Pareto généralisée (GPD), méthode de régression et méthode du Q.Q-plot ...etc.

Définition 32 (*Estimation naturelle*). Du fait que $X_{(k,n)}$ est un estimateur naturel du quantile d'ordre $1 - (k - 1)/n$, un estimateur naturel de $\mathbf{F}^{-1}(p_n)$, pour $p_n \in]0, 1[$, est donné par

$$\hat{x}_{p_n} = ([p_n n] - p_n n + 1)X_{(n-[p_n n],n)} + ([p_n n] - p_n n)X_{(n-[p_n n],n)}, \quad p_n \in]0, 1[.$$

où $[.]$ désigne la partie entière cette méthode nécessite pour être efficace en pratique un volume de données jamais disponible.

l'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. Nous utiliserons cet estimateur de l'indice de queue γ_1 afin d'estimer un quantile extrême, suite à un classique schéma plus précisément, soit p_n une séquence de quantiles arrêtés tendant à 0, telle que $p_n = o(\bar{\mathbf{F}}(t_n))$, si x_{p_n} désigne le quantile de \mathbf{F} d'ordre $1 - p_n$, i.e résoudre les $\bar{\mathbf{F}}(x_{p_n}) = p_n$, il est facile de voir que nous pouvons estimer x_{p_n} par \hat{x}_{p_n} alors :

$$\hat{x}_{p_n, t_n} = t_n \left(\frac{\bar{\mathbf{F}}_n(t_n)}{p_n} \right)^{\hat{\gamma}_1^{(JR)}}$$

Théorème 33 Sous 2.16 et les hypothèses 2.2.2, ajustement $d_n = \bar{\mathbf{F}}(t_n)/p_n$, si $\rho_1 < 0$ et

$$\sqrt{n\bar{H}(t_n)/\log d_n} \longrightarrow \infty \tag{2.18}$$

quand n tend à ∞

$$\frac{\sqrt{n\bar{H}(t_n)}}{\log d_n} \left(\frac{\hat{x}_{p_n, t_n}}{x_{p_n}} - 1 \right) \xrightarrow{\mathcal{L}} \mathcal{N}(\lambda m, s^2)$$

Preuve. voir Julien Worms, Rym Worms (2015). ■

2.2.4 Méthode d'estimation de la prime $\hat{\Pi}_1$

Nous avons besoin dans ce cas pour déterminer l'estimateur de la prime avec un seuil fixé, à définir la quantile extrême par :

$$x_{p_n} = \mathbf{F}^{\leftarrow}(1 - p_n).$$

Nous pouvons l'écrire d'une autre manière,

$$x_{p_n} = t_n \left(\frac{\bar{\mathbf{F}}(t_n)}{p_n} \right)^{\gamma_1}, \quad \gamma_1 > 0$$

Par simplification, on obtient

$$\begin{aligned} x_{p_n} &= t_n \left(\frac{\bar{\mathbf{F}}(t_n)}{1 - \mathbf{F}(x_{p_n})} \right)^{\gamma_1} \\ x_{p_n} &= t_n \left(\frac{\bar{\mathbf{F}}(t_n)}{\bar{\mathbf{F}}(x_{p_n})} \right)^{\gamma_1} \\ x_{p_n} &= t_n \frac{(\bar{\mathbf{F}}(t_n))^{\gamma_1}}{(\bar{\mathbf{F}}(x_{p_n}))^{\gamma_1}} \\ (\bar{\mathbf{F}}(x_{p_n}))^{\gamma_1} &= \frac{t_n}{x_{p_n}} (\bar{\mathbf{F}}(t_n))^{\gamma_1}, \quad \gamma_1 > 0 \end{aligned}$$

Où,

$$\bar{\mathbf{F}}(x_{p_n}) = \left(\frac{t_n}{x_{p_n}} \right)^{1/\gamma_1} \bar{\mathbf{F}}(t_n)$$

Nous pouvons l'écrire d'une autre manière,

$$\bar{\mathbf{F}}(x_{p_n}) = \left(\frac{x_{p_n}}{t_n} \right)^{-1/\gamma_1} \bar{\mathbf{F}}(t_n)$$

Maintenant, nous obtenons l'estimateur empirique de $\bar{\mathbf{F}}(x_{p_n})$ par :

$$\hat{\bar{\mathbf{F}}}(x_{p_n}) = \left(\frac{x_{p_n}}{t_n} \right)^{-1/\hat{\gamma}_1^{(JR)}} \bar{\mathbf{F}}_n(t_n)$$

si le seuil fixé $u = t_n$, alors,

$$\hat{\bar{\mathbf{F}}}(x_{p_n}) = \left(\frac{t_n}{t_n} \right)^{-1/\hat{\gamma}_1^{(JR)}} \bar{\mathbf{F}}_n(t_n)$$

Nous obtenons

$$\hat{\mathbf{F}}(x_{p_n}) = \bar{\mathbf{F}}_n(t_n)$$

Finalement, nous définissons un estimateur $\hat{\Pi}_1$ de la prime 2.4 en estimer γ_1 par $\hat{\gamma}_1^{(JR)}$ on obtient :

$$\hat{\Pi}_1(t_n) = \frac{\hat{\gamma}_1^{(JR)}}{1 - \hat{\gamma}_1^{(JR)}} t_n \bar{\mathbf{F}}_n(t_n)$$

où

$$\bar{\mathbf{F}}_n(t_n) = \left\{ 1 - \prod_{i, X_i > t_n} \left(1 - \frac{1}{nC_n(X_i)} \right) \right\}$$

Par conséquent, nous déduisent l'estimateur de la prime par relation suivent :

$$\hat{\Pi}_1(t_n) = \frac{\hat{\gamma}_1^{(JR)}}{1 - \hat{\gamma}_1^{(JR)}} t_n \left\{ 1 - \prod_{i, X_i > t_n} \left(1 - \frac{1}{nC_n(X_i)} \right) \right\} \quad (*)$$

2.3 Estimation de la prime avec un seuil aléatoire

Haouas et al. (2016) ont proposé un estimateur de l'indice de valeur extrême adaptée par l'estimateur de Hill sous données tronquées à droite de type Pareto. Nous nous l'appliquons pour calculer un estimateur de la prime avec un seuil aléatoire $X_{n-k:n}$.

2.3.1 Estimateur de l'indice de queue $\hat{\gamma}_1^{(LB)}$

Soit l'échantillon X_1, \dots, X_n , est variables aléatoires non-négatives définie sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$ avec sa fonction de répartition \mathbf{F} , supposons que la queue de distribution $1 - \mathbf{F}$ est à variations régulières à l'infini d'indice $(-1/\gamma_1)$, noté $\bar{\mathbf{F}} = 1 - \mathbf{F} \in RV_{-1/\gamma_1}$, c'est-à-dire, pour tout $x > 0$

$$\lim_{x \rightarrow \infty} \frac{1 - \mathbf{F}(xz)}{1 - \mathbf{F}(z)} = x^{-1/\gamma_1}, \quad \gamma_1 > 0. \quad (2.19)$$

l'échantillon $(X_i)_{1 < i \leq n}$ n'est pas observé mais il est tronqué par un deuxième échantillon Y_1, \dots, Y_n avec fonction de répartition \mathbf{G} , *i.i.d* supposés indépendants de l'échantillon X_i . La queue de la distribution est à variations régulières aussi à l'infini d'indice $(-1/\gamma_2)$, noté $\bar{\mathbf{G}} = 1 - \mathbf{G} \in RV_{-1/\gamma_2}$, c'est-à-dire, pour tout $x > 0$

$$\lim_{x \rightarrow \infty} \frac{1 - \mathbf{G}(xz)}{1 - \mathbf{G}(z)} = x^{-1/\gamma_2} \quad (2.20)$$

Il est bien connu que, dans l'analyse de la valeur extrême, des approximations faibles sont réalisées dans le cadre de second ordre (voir, par exemple, de Haan et Ferreira, 2006, page 48). Il semble tout à fait naturel de supposer que \mathbf{F} et \mathbf{G} satisfont la condition de deuxième ordre de variation régulière, ce que nous exprimons en termes de fonctions de queue appartenant aux deux fd. C'est-à-dire que nous supposons que pour $x > 0$, nous avons

$$\lim_{t \rightarrow \infty} \frac{\mathbb{U}_{\mathbf{F}}(tx)/\mathbb{U}_{\mathbf{F}}(t) - x^{\tau_1}}{\mathbf{A}_{\mathbf{F}}(t)} = x^{\tau_1} \frac{x^{\tau_1-1}}{\tau_1} \quad (2.21)$$

et

$$\lim_{t \rightarrow \infty} \frac{\mathbb{U}_{\mathbf{G}}(tx)/\mathbb{U}_{\mathbf{G}}(t) - x^{\tau_2}}{\mathbf{A}_{\mathbf{G}}(t)} = x^{\tau_2} \frac{x^{\tau_2-1}}{\tau_2} \quad (2.22)$$

où $\tau_1, \tau_2 < 0$, sont les paramètres de second ordre et $\mathbf{A}_{\mathbf{G}}, \mathbf{A}_{\mathbf{F}}$ sont des fonctions tendant à zéro et ne change pas les signes à proximité de l'infini avec des valeurs absolument variables à l'infini avec indices τ_1, τ_2 respectivement.

Prenons maintenant (X_i, Y_i) , $i = 1, \dots, n$ les données observées, comme des copies d'un couple de v.a (X, Y) , correspondant à l'échantillon tronqué $(\mathbf{X}_i, \mathbf{Y}_i)$, $i = 1, \dots, N$, où $n = n_N$ est une suite aléatoire de nombres entiers, en vertu de la loi faible des grands nombres, satisfait

$$n_N/N \xrightarrow{\mathbf{P}} p = \mathbf{P}(\mathbf{X} \leq \mathbf{Y})$$

quand $N \rightarrow \infty$, on désigne la jointe f.d de X et Y par

$$\begin{aligned} H(x, y) &= \mathbf{P}(X \leq x, Y \leq y) \\ &= \mathbf{P}(\mathbf{X} \leq \min(x, \mathbf{Y}), \mathbf{Y} \leq y | \mathbf{X} \leq \mathbf{Y}) \\ &= P^{-1} \int_0^y \mathbf{F}(\min(x, z)) d\mathbf{G}(z) \end{aligned}$$

Puisque \mathbf{F} et \mathbf{G} sont à queue lourde, alors leurs points d'extrémité droit sont infinis et donc ils sont égal. On peut écrire la relation entre \mathbf{F} , F et C :

$$\int_x^\infty d\mathbf{F}(y)/\mathbf{F}(y) = \int_x^\infty dF(y)/C(y).$$

La solution est définie par,

$$\mathbf{F}(x) = \exp \left\{ - \int_x^\infty dF(y)/C(y) \right\}.$$

Ceci conduit à l'estimateur non paramétrique de Woodroofe (1985) de \mathbf{F} , donné par

$$\mathbf{F}_n^{(W)}(x) = \prod_{i: X_i > x} \exp \left\{ \frac{1}{nC_n(X_i)} \right\}$$

Théorème 34 *Supposons que \mathbf{F} et \mathbf{G} satisfaisant les conditions du second ordre, avec $\gamma_1 < \gamma_2$. Soit $k = k_n$ une suite aléatoire d'entiers tels que, $n = m$, $k_m \rightarrow \infty$ et $k_m/m \rightarrow 0$, quand $N \rightarrow \infty$, lors pour tout $x_0 > 0$, on a*

$$\sup_{x \geq x_0} x^{-1/\gamma_1} \frac{|\bar{\mathbf{F}}_n^W(X_{n-k:n}x) - \bar{\mathbf{F}}_n(X_{n-k:n}x)|}{\bar{\mathbf{F}}_n(X_{n-k:n})} = O_P((k/n)^{\gamma_1/\gamma}).$$

Preuve. Voir Houas et al. (2016) ■

Notez que la queue de F , en même temps dépend de $\bar{\mathbf{G}}$ et $\bar{\mathbf{F}}$ alors que celui des \bar{G} ne repose que sur $\bar{\mathbf{G}}$. À l'aide de la proposition B.1.10 de Haan et Ferreira (2006), pour la variation régulière des fonctions des $\bar{\mathbf{G}}$ et $\bar{\mathbf{F}}$, nous montrons que les deux fonctions \bar{F} et \bar{G} sont régulièrement variante à l'infini, avec les indices respectifs $-1/\gamma := (\gamma_1\gamma_2)/(\gamma_1 + \gamma_2)$ et $-1/\gamma_2$.

Gardes et Stupfler (2015) ont obtenu un estimateur cohérent, pour l'indice de valeur extrême γ_1 , dont la normalité asymptotique est établie dans Benchaira et al. (2015), sous les queue lourdes et les conditions du second ordre de variation régulière.

De même, Benchaira et al. (2016a) a considéré une intégration de Woodroofe (avec un seuil aléatoire) pour proposer un nouvel estimateur pour l'indice de queue γ_1 donné par

$$\hat{\gamma}_1^{(W)} = \frac{1}{n\bar{\mathbf{F}}_n^{(W)}(X_{n-k:n})} \sum_{i=1}^n \frac{\mathbf{F}_n^{(W)}(X_{n-i+1:n})}{C_n(X_{n-i+1:n})} \log \frac{X_{n-i+1:n}}{X_{n-k:n}}$$

Où, $n = m = m_N$, $Z_{1:m} \leq \dots \leq Z_{m:m}$, énumérer les statistiques d'ordre relatives à un échantillon $Z_m \leq \dots \leq Z_m$, et $k = k_n$ est une séquence (aléatoire) d'entiers tel que,

$$1 < k_m < m, \quad k_m \rightarrow \infty \quad \text{et} \quad k_m/m \rightarrow 0, \quad \text{quand} \quad N \rightarrow \infty.$$

La consistance et l'asymptotique normalité de $\hat{\gamma}_1^{(W)}$ sont établis à Benchaira et al. (2016a), par une approximation faible au processus de la queue de Woodroofe

$$\mathbf{D}_n^{(W)}(x) = \sqrt{K} \left(\frac{\bar{\mathbf{F}}_n^{(W)}(X_{n-K:n}x)}{\bar{\mathbf{F}}_n^{(W)}(X_{n-K:n})} - x^{-1/\gamma_1} \right), \quad x > 0$$

Plus précisément, les auteurs ont montré que, en dessous de 2.21 et 2.22 avec $\gamma_1 < \gamma_2$, il existe une fonction $\mathbf{A}_0(t) \sim \mathbf{A}_{F^*}(t) = \mathbf{A}_F(1/\bar{F}(\mathbb{U}_F(t)))$, $t \rightarrow 0$, et un processus Wiener standard $\{\mathbf{W}(s), s > 0\}$, défini sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$, de sorte que, pour $0 < \epsilon < 1/2 - \gamma/\gamma_2$ et $x_0 > 0$,

$$\sup_{x \geq x_0} x^{(1/2-\epsilon)/\gamma-1/\gamma_2} \left| \mathbf{D}_n^{(\mathbf{W})}(x) - \Gamma(x, \mathbf{W}) - x^{-1/\gamma_1} \frac{x^{-\tau_1/\gamma_1-1}}{\tau_1 \gamma_1} \sqrt{k} \mathbf{A}_0(n/k) \right| = o_{\mathbf{P}}(1).$$

quand $N \rightarrow \infty$, pourvu que $n = m \sqrt{k_m} \mathbf{A}_0(m/k_m) = o(1)$ où $\{\Gamma(x, \mathbf{W}), x > 0\}$, est un processus gaussien défini par

$$\begin{aligned} \Gamma(x, \mathbf{W}) &= \frac{\gamma}{\gamma_1} x^{-1/\gamma_1} \{x^{-1/\gamma} \mathbf{W}(x^{-1/\gamma}) - \mathbf{W}(1)\} \\ &+ \frac{\gamma}{\gamma_1 + \gamma_2} x^{-1/\gamma_1} \int_0^1 s^{-\gamma/\gamma_2-1} \{x^{1/\gamma} \mathbf{W}(x^{1/\gamma} s) - \mathbf{W}(s)\} ds. \end{aligned}$$

Compte tenu de l'approximation faible antérieure, les auteurs ont également prouvé que si, donné $n = m$, $\sqrt{K_m} \mathbf{A}_{F^*} \rightarrow \lambda$, alors

$$\sqrt{K} (\hat{\gamma}_1^{(JR)} - \gamma_1) \xrightarrow{\mathcal{D}} \mathcal{N} \left(\frac{\lambda}{1 - \tau_1}, \sigma^2 \right)$$

où

$$\sigma^2 := \gamma^2 (1 - \gamma_1/\gamma_2) (1 + (\gamma_1/\gamma_2)^2) / (1 - \gamma_1/\gamma_2)^3$$

Alors *Haouas et al. (2016)* ont introduit un estimateur du noyau à γ_1 qui améliore le biais de $\hat{\gamma}_1^{(W)}$, nous sommes intéressés par l'estimateur de Worm $\hat{\gamma}_1^{(JR)}(t_n)$, qui est supposé pour un seuil aléatoire égale à $X_{n-k:n}$. Cela rend l'estimateur plus commode pour le calcul numérique mise en œuvre que celle avec un seuil déterministe.

$$\hat{\gamma}_1^{(LB)} := \frac{1}{n \bar{\mathbf{F}}_n(X_{n-k:n})} \sum_{i=1}^k \frac{\mathbf{F}_n(X_{n-i+1:n})}{C_n(X_{n-i+1:n})} \log \frac{X_{n-i+1:n}}{X_{n-k:n}} \quad (2.23)$$

Notez que $\mathbf{F}_n(\infty) = 1$ et écrire $\bar{\mathbf{F}}_n(X_{n-k:n}) = \int_{X_{n-k:n}}^{\infty} d\mathbf{F}_n(y)$. D'autre part, nous avons

$$C_n(x) d\mathbf{F}_n(x) = \mathbf{F}_n(x) dF_n(x).$$

(voir, par exemple, *Strzalkowska-Kominiak et Stute, 2009*), alors

$$\bar{\mathbf{F}}_n(X_{n-k:n}) = \int_{X_{n-k:n}}^{\infty} \frac{\mathbf{F}_n(x)}{C_n(x)} dF_n(x) = \frac{1}{n} \sum_{i=1}^k \frac{\mathbf{F}_n(X_{n-i+1:n})}{C_n(X_{n-i+1:n})}.$$

Cela nous permet de réécrire le nouvel estimateur

$$\hat{\gamma}_1^{(LB)} := \sum_{i=1}^k a_n^{(i)} \log \frac{X_{n-i+1:n}}{X_{n-k:n}} \quad (2.24)$$

où

$$a_n^{(i)} = \frac{\mathbf{F}_n(X_{n-i+1:n})}{C_n(X_{n-i+1:n})} / \sum_{i=1}^k \frac{\mathbf{F}_n(X_{n-i+1:n})}{C_n(X_{n-i+1:n})}$$

Remarque 35 *il est à noter que pour des données complètes, nous avons $n \equiv N$ et $\mathbf{F}_n \equiv F_n \equiv C_n$, il s'ensuit que $a_n^{(i)} \equiv k^{-1}$, $i = 1, \dots, k$ et par conséquent les deux $\hat{\gamma}_1^{(LB)}$ et $\hat{\gamma}_1^{(W)}$ réduit à l'estimateur classique de Hill (1975). La cohérence et la normalité asymptotique de $\hat{\gamma}_1^{(LB)}$ seront atteints grâce à une faible approximation de la queue correspondante du processus de Lynden-Bell que nous définissons par :*

$$\mathbf{D}_n^{(LB)}(x) = \sqrt{K} \left(\frac{\bar{\mathbf{F}}_n(X_{n-K:n}x)}{\bar{\mathbf{F}}_n(X_{n-K:n})} - x^{-1/\gamma_1} \right), \quad x > 0$$

Théorème 36 (normalité asymptotique de $\hat{\gamma}_1^{(LB)}$) *Supposons que 2.19 et 2.20 et remplis avec $\gamma_1 < \gamma_2$, $\tau_1 < 0$ et soit $k = k_n$ une suite aléatoire de nombres entiers tels que $n = m$, et $k_m/m \rightarrow 0$, quand $N \rightarrow \infty$, $\hat{\gamma}_1^{(LB)} \xrightarrow{P} \gamma_1$ alors*

$$\begin{aligned} \sqrt{k} \left(\hat{\gamma}_1^{(LB)} - \gamma_1 \right) &= \frac{\sqrt{k} \mathbf{A}_0(n/k)}{1 - \tau_1} - \gamma \mathbf{W}(1) \\ &+ \frac{\gamma}{\gamma_1 + \gamma_2} \int_0^1 (\gamma_2 - \gamma_1 - \gamma \log(s)) s^{-\gamma/\gamma_2 - 1} \mathbf{W}(s) ds + o_{\mathbf{P}}(1) \end{aligned}$$

Si, en outre, on suppose que, étant donné que $n = m$, $\sqrt{k_m} \mathbf{A}_{\mathbf{F}}^(m/K_m) \rightarrow \lambda < \infty$, alors*

$$\sqrt{k} (\hat{\gamma}_1^{(LB)} - \gamma_1) \xrightarrow{\mathcal{D}} \mathcal{N} \left(\frac{\lambda}{1 - \tau_1}, \sigma^2 \right)$$

où,

$$\sigma^2 := \gamma^2 (1 - \gamma_1/\gamma_2) (1 + (\gamma_1/\gamma_2)^2) / (1 - \gamma_1/\gamma_2)^3$$

Preuve. La preuve est semblable, à celle du Corollaire 3.1 dans Benchaira et al. (2016a).

■

2.3.2 Méthode d'estimation de la prime $\hat{\Pi}_2$

Nous voudrions estimer la prime des dépassement de risque extrêmes avec un seuil aléatoire $X_{n-k:n}$. L'estimateur de type Weissman (1978) est défini par :

$$\hat{\mathbf{F}}(x) = \left(\frac{x}{X_{n-k:n}} \right)^{-1/\hat{\gamma}_1^{(LB)}} \bar{\mathbf{F}}_n(X_{n-k:n}) \quad (2.25)$$

La queue de distribution $\bar{\mathbf{F}}_n$ empirique avec des données tronquées, donne la relation entre $\bar{\mathbf{F}}$ et \mathbf{F} par :

$$\bar{\mathbf{F}}(x) = 1 - \mathbf{F}(x)$$

En utilisant la fonction empirique de $\bar{\mathbf{F}}$ et \mathbf{F} avec un seuil aléatoire, nous trouvons :

$$\bar{\mathbf{F}}_n(X_{n-k:n}) = 1 - \mathbf{F}_n(X_{n-k:n})$$

Où \mathbf{F}_n est donnée par la équation 2.9,

$$\mathbf{F}_n(X_{n-k:n}) = \prod_{i=n-k+1}^n \left(1 - \frac{1}{nC_n(X_{i:n})} \right)$$

Nous avons à partir de la relation précédente :

$$\bar{\mathbf{F}}_n(X_{n-k:n}) = 1 - \prod_{i=n-k+1}^n \left(1 - \frac{1}{nC_n(X_{i:n})} \right)$$

Alors en remplaçant $\bar{\mathbf{F}}_n(X_{n-k:n})$ par son estimateur donne par 2.25 on obtient,

$$\begin{aligned} \hat{\mathbf{F}}(X_{n-k:n}) &= \left(\frac{X_{n-k:n}}{X_{n-k:n}} \right)^{-1/\hat{\gamma}_1^{(LB)}} \left\{ 1 - \prod_{i=n-k+1}^n \left(1 - \frac{1}{nC_n(X_{i:n})} \right) \right\} \\ \hat{\mathbf{F}}(X_{n-k:n}) &= \left\{ 1 - \prod_{i=n-k+1}^n \left(1 - \frac{1}{nC_n(X_{i:n})} \right) \right\} \end{aligned} \quad (2.26)$$

Finalement, avec une seuil aléatoire $u = X_{n-k:n}$ en remplaceant 2.26 dans 2.4, et γ_1 par $\hat{\gamma}_1^{(LB)}$

On obtient notre nouvel estimateur,

$$\hat{\Pi}_2(X_{n-k:n}) = \frac{\hat{\gamma}_1^{(LB)}}{1 - \hat{\gamma}_1^{(LB)}} X_{n-k:n} \left\{ 1 - \prod_{i=1}^k \left(1 - \frac{1}{nC_n(X_{n-i+1:n})} \right) \right\} \quad (**)$$

Cet estimateur est en fonction du nombre de valeurs extrêmes k et de l'estimateur de $\hat{\gamma}_1^{(LB)}$

ÉTUDE DE SIMULATIONS

Le logiciel R est un logiciel de statistique créé par Ross Ihaka & Robert Gentleman. Il est à la fois un langage informatique et un environnement de travail : les commandes sont exécutées grâce à des instructions codées dans un langage relativement simple, les résultats sont affichés sous forme de texte et les graphiques sont visualisés directement dans une fenêtre propre. C'est un clone du logiciel S-plus qui est fondée sur le langage de programmation S, développée par les laboratoires Bell en 1988. Ce logiciel sert à manipuler des données, à tracer des graphiques et à faire des analyses statistiques sur ces données. Le logiciel R constitue aujourd'hui un langage de programmation intégré d'analyse statistique. Le site Internet <http://www.r-project.org>, est la meilleure source d'informations sur ce logiciel.

Ce chapitre est réparti en deux sections, on présente dans la première section comment déterminer la valeur de l'estimateur de l'indice des valeurs extrêmes de *J. Worms, R. Worms(2015)* dans le cas du seuil fixé ensuite, déterminer la valeur de la prime par simulation sous le logiciel R. La deuxième section avec un seuil aléatoire, nous présente la valeur de l'estimateur de l'indice des valeurs extrêmes de *Haouas et al. (2016)* et la valeur de l'estimateur de la prime correspondante.

3.1 Génération de l'échantillon tronqué

1. La loi de simulation utilisée dans ce cas est une loi de Pareto de paramètre γ de fonction de répartition,

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

2. On suppose que les données $(X_i)_{1 \leq i \leq n}$ sont tronquée par $(Y_i)_{1 \leq i \leq n}$ et que X et Y suivent une distribution à queue lourde.

3. Nous avons généré un échantillon $(X_i)_{1 \leq i \leq n} \sim \text{Pareto}(\gamma_1)$ de taille $n = 2000$, à partir d'une variable u de $U([0, 1])$, le modèle ajusté sera :

$$F^{-1}(t) = (1 - t)^{-\gamma_1}$$

4. L'échantillon $(X_i)_{1 \leq i \leq n}$ est tronqué par un deuxième échantillon $(Y_i)_{1 \leq i \leq n} \sim \text{Pareto}(\gamma_2)$ à partir d'une variable v de $U([0, 1])$:

$$G^{-1}(v) = (1 - v)^{-\gamma_2}$$

5. Les variables que nous observons sont d'une part les $X_i \sim \text{Pareto}(\gamma)$ définies par :

$$X_i = \{X_i, si \quad X_i \leq Y_i\}_{1 \leq i \leq N}$$

PROGRAMME SOUS R

```
# Extraire les données tronquées
N=1500
gamma1<-?
gamma2<-?
# échantillon initial
t<-runif(N,0,1)
x1<-(1-t)^(-gamma1)# généré un échantillon Xi
v<-runif(N,0,1)
y1<-(1-v)^(-gamma2)# généré un échantillon Yi
A<-seq(1,N)
for(i in 1 :N)
{ if(x1[i]<=y1[i])
A[i]<-x1[i]
else
A[i]<-NA}
```

```

B<-seq(1,N)
for(i in 1 :N)
{if(x1[i]<=y1[i])
B[i]<-y1[i]
else
B[i]<-NA}
A;B
X<-na.omit(A)#supprimer la valeur NA dans la échantillon A .
Y<-na.omit(B) #supprimer la valeur NA dans la échantillon B .

```

3.2 Simulations dans le cas du seuil fixé

Pour simuler l'estimateur de la prime, premièrement on détermine les données tronquées : soit l'échantillon $(\mathbf{X}_i)_{1 \leq i \leq N}$ tronquée par un deuxième échantillon $(\mathbf{Y}_i)_{1 \leq i \leq N}$. Ensuite obtenir notre nouvel échantillon $(X_i)_{1 \leq i \leq n}$ ceci après avoir supprimer les valeurs inconnues (NA) dans l'échantillon. Ensuite simuler l'indice des valeurs extrêmes $\hat{\gamma}_1^{(JR)}$ et par conséquent l'estimateur de la prime.

3.2.1 Simulations de $\hat{\gamma}_1^{(JR)}$

Pour estimer l'indice de queue $\hat{\gamma}_1$ on suit les étapes suivantes :

1. Il est important pour analyser les données tronquées de calculer les valeurs C_n l'estimateur de $C(x)$.

$$C_n(X_j) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i \leq X_j \leq Y_i}$$

PROGRAMME SOUS R

```

X<-sort(X)
Y<-sort(Y)
n<-length(X)# taille de X
kk<-0
Cn<-seq(1,n)

```

```

for(j in 1 :n)
{
for(i in 1 :n)
{if(X[i]<=X[j] && X[j]<=Y[i])
kk[i]<-1
else
kk[i]<-0
}
Cn[j]<-sum(kk)/n
}
Cn

```

2. Calcul de l'estimateur F_n donnée par :

$$F_n(X_j) = \prod_{i, X_i > X_j} \left(1 - \frac{1}{nC_n(X_i)} \right)$$

PROGRAMME SOUS R

```

Fn<-seq(1,n)
for(i in 1 :n)
{pr<-1
for(j in 1 :n)
{
if( X[j]>X[i] && i!=j )
pr<-pr*(1-1/(n*Cn[j]))# prod Fn et appel de la fonction Cn
else
pr<-pr*1
}
Fn[i]<-pr
}
Fn # valeur Fn

```

3. On peut fixer le seuil u à partir de l'échantillon tronqué X .
4. Pour un seuil fixé u on calcul la valeur de $\bar{\mathbf{F}}_n(u)$ qui représente l'estimateur empirique des valeurs dépassants le seuil u de queue :

$$\bar{\mathbf{F}}_n(u) = 1 - \prod_{i, X_i > u}^n \left(1 - \frac{1}{nC_n(X_i)} \right)$$

Le programme suivant sous R calcul directement la valeur de $\bar{\mathbf{F}}_n(u)$:

```
Fbar<-0
p<-0
for(i in 1 :n)
{
  {if(X[i]>u )
pr<-(1-1/(n*Cn[i])) # produit Fn ET appel de la fonction Cn
else
pr<-1}
p[i]<-pr
}
Fbar<-1-prod(p)
Fbar
```

5. Donc, on peut calculer la valeur de l'estimateur la l'indice de γ_1 comme suit(Worms et worms (2015)) :

$$\hat{\gamma}_1^{(JR)} = \frac{1}{n\bar{\mathbf{F}}_n(u)} \sum_{i=1}^n \log \left(\frac{X_i}{u} \right) \frac{\mathbf{F}_n(X_i)}{C_n(X_i)} \mathbb{I}_{X_i > u}$$

Le programme suivant sous R calcul directement la valeur de $\hat{\gamma}_1^{(JR)}$ +

```
kk<-0
for(i in 1 :n)
{ {if(X[i]>u)
kk[i]<-(1/(n*Fbar))*(log(X[i]/u))*(Fn[i]/(Cn[i]))# Équation de gamma
else
```

```

kk[i]<-0 }
GAMMA1<-sum(kk) }
GAMMA1 # la valeur de l'estimateur

```

3.2.2 Simulations de $\hat{\Pi}_1$

Pour estimer la prime dans le cas de seuil fixé on utilise

$$\hat{\Pi}_1(t_n) = \frac{\hat{\gamma}_1^{(JR)}}{1 - \hat{\gamma}_1^{(JR)}} t_n \left\{ 1 - \prod_{i, X_i > t_n}^n \left(1 - \frac{1}{nC_n(X_i)} \right) \right\}$$

Après avoir déterminé la valeur de $\hat{\gamma}_1$ et le seuil u , on utilise l'instruction suivante pour donner la valeur de la prime estimée :

```
ESTIMprime1<-(GAMMA1/(1-GAMMA1))*(u)*(Fbar)
```

```
ESTIMprime1
```

```
# valeur la prime théorique :
```

```
prim1<-(gamma1/(1-gamma1))*(u)^(-1/gamma1+1)
```

```
prim1
```

U_i	$U_1=1.066206$	$U_2=1.142513$	$U_3=1.360111$	$U_4=1.564085$	$U_5=1.813226$	$U_6=2.011409$
$\hat{\gamma}_1$	0.1978515	0.1966073	0.1962428	0.1868816	0.1891323	0.1758272
Π	0.193453	0.1467221	0.07305376	0.04177328	0.0231276	0.0152735
$\hat{\Pi}_1$	0.1861763	0.1401618	0.06949668	0.03805058	0.02042173	0.012755

TABLE 3.1 : $X \sim$ Pareto ($\gamma_1 = 0.2$) tronques par $Y \sim$ Pareto ($\gamma_1 = 2.5$), $N=2000$ et après troncature $n=1713$.

D'après nos simulations notre estimateur est stable par rapport à U et présente une allure très proche de sa valeur théorique (voir les tableau 3.1).

3.3 Simulations dans le cas du seuil aléatoire

3.3.1 Simulations de $\hat{\gamma}_1^{(LB)}$

1. Calculer les valeurs C_n et \mathbf{F}_n de la même façon définies précédemment.

2. Nous allons utilisé dans nos simulations la formule :

$$k = [n^\theta], \quad 0 < \theta < 1.$$

pour déterminer le nombre des valeurs extrêmes.

3. Finalement, on peut calculer Haous et al.(2016)

$$\hat{\gamma}_1^{(LB)} = \sum_{i=1}^k a_n^{(i)} \log \frac{X_{n-i+1:n}}{X_{n-k:n}}.$$

telle que

$$a_n^{(i)} = \frac{\mathbf{F}_n(X_{n-i+1:n})}{C_n(X_{n-i+1:n})} / \sum_{i=1}^k \frac{\mathbf{F}_n(X_{n-i+1:n})}{C_n(X_{n-i+1:n})}$$

PROGRAMME SOUS R

#déterministe la valeur $a_n^{(i)}$

som<-0 #vecteur vide

som1<-0

An<-seq(1,n)

for(i in 1 :k)

{ som[i]<-(Fn[n-i+1]/Cn[n-i+1])} # APPLER LES DEUX FONCTION Fn et Cn

som1<-sum(som)

som1

for(i in 1 :k)

{

An[i]<-(Fn[n-i+1]/Cn[n-i+1])/som1

}

An # valeur de $a_n^{(i)}$

Le programme suivant sous R calcul directement la valeur de $\hat{\gamma}_1^{(LB)}$

GAMMALB<-0

for(i in 1 :k)

{

GAMMALB[i]<-An[i]*log(X[n-i+1]/X[n-k])

}

GAMMALB<-sum(GAMMALB)

GAMMALB # Valeur de l'indice de valeur extrême $\hat{\gamma}_1^{(LB)}$

3.3.2 Simulation de $\hat{\Pi}_2$

Pour estimer la prime dans le cas du seuil aléatoire on procède comme suit :

$$\hat{\Pi}_2(X_{n-k:n}) = \frac{\hat{\gamma}_1^{(LB)}}{1 - \hat{\gamma}_1^{(LB)}} X_{n-k:n} \left\{ 1 - \prod_{i=1}^k \left(1 - \frac{1}{nC_n(X_{n-i+1:n})} \right) \right\}$$

Pour calculer l'estimateur de la prime nous utilisons le programme suivant :

PROGRAMME SOUS R

```
pri<-0
```

```
for(i in 1 :k)
```

```
{ pri[i]<-(1-1/(n*Cn[n-i+1])) }
```

```
lapoduit<-1-prod(pri)
```

```
lapoduit
```

```
LPRIMEstim2<-X[n-k]*((GAMMALB)/(1-GAMMALB))*lapoduit
```

```
LPRIMEstim2
```

```
# valeur la prime théorique :
```

```
vallaprim2<-(gamma1/(1-gamma1))*(X[n-k])^(-1/gamma1+1)
```

3.3.3 Représentation graphique :

```
# graphe l'indice de valeur extrême  $\hat{\gamma}_1^{(LB)}$  vs k.
GAMMALB<-seq(1,n-1)
Z<-log(X)
for(k in 1 :n-1)
{ l<-seq(1,k)
GAMMALB[k]<-((Fn[n-k+1]/Cn[n-k+1])/(sum(Fn[n-l+1]/Cn[n-l+1])))*(sum(Z[n-l+1]-Z[n-
k]))
}
GAMMALB #print valeur de GAMMA2 estimer
k<-seq(1,n-1)
plot(k,GAMMALB,ylim=c(0.1,0.35),col="3", type='l',lwd="1",ylab=expression(GAMMALB))
abline(h=gamma1, col="2", lwd="2")
# graphe la ESTIMprime2 et prime2 en fonction de k
prime2<-seq(1,n-1)
ESTIMprime2<-seq(1,n-1)
for(k in 1 :n-1)
{
prime2[k]<-(gamma1/(1-gamma1))*((X[n-k])^(-1/gamma1+1))
l<-seq(1,k)
ESTIMprime2[k]<-(GAMMALB[k]/(1-GAMMALB[k]))*X[n-k]*(1-prod(1-1/(n*Cn[n-l+1])))
}
ESTIMprime2 ;prime2
k<-seq(1,n-1)
plot(k,ESTIMprime2,type='l',ylab="Estimation la prime",col='3')
lines(k,prime2,ESTIMprime2,type='l',col='2',lwd=2)
text(500, 0.01, expression(La prime), col='2')
text(1500, 0.005, expression(La prime estimer), col='3')
text(100, 0.019, expression(n ==1993), col='6')
text(100, 0.018, expression(paste(plain(gamma^1))==0.02), col='6')
text(100, 0.017, expression(paste(plain(gamma^2))==2.9), col='6')
```

3.4 Résultats des simulations

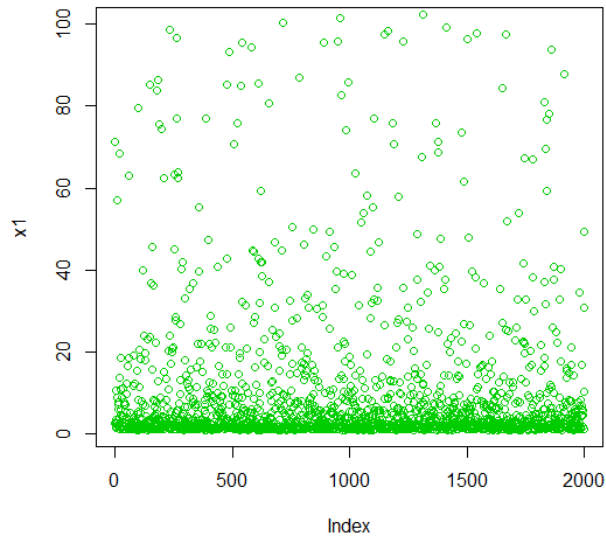


FIGURE 3.1 : Distribution de l'échantillon initial X de Pareto ($\gamma_1=0.15$), avant la troncature de taille $N=2000$.

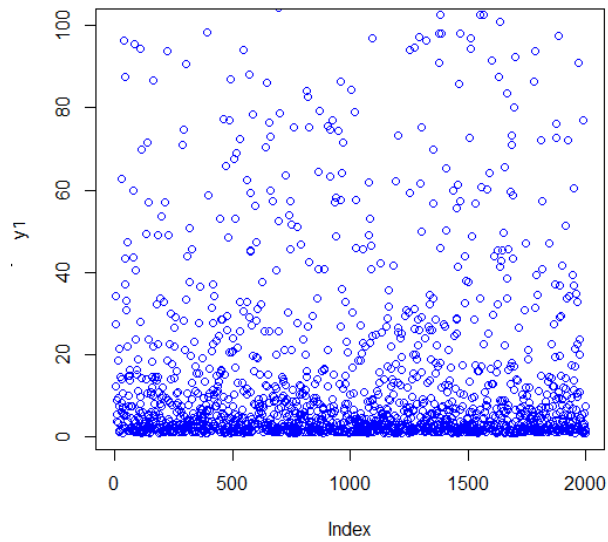


FIGURE 3.2 : Distribution de l'échantillon Y de Pareto ($\gamma_2=2.9$), appelée v.a de troncature de taille $N=2000$

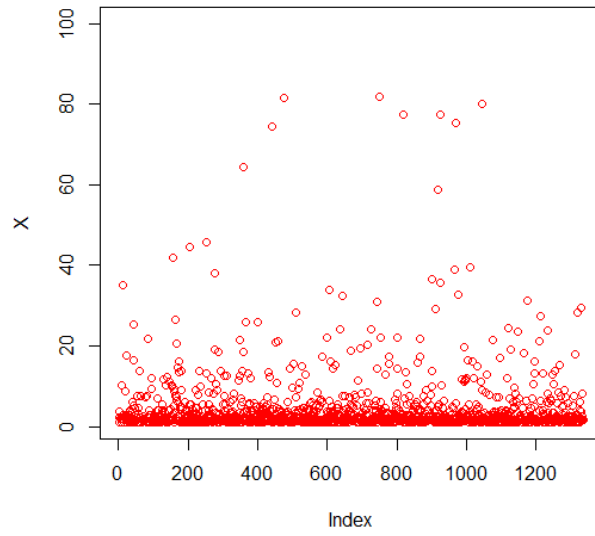


FIGURE 3.3 : Représentation de l'échantillon X après la troncature par l'échantillon Y avec $n < N = 2000$.

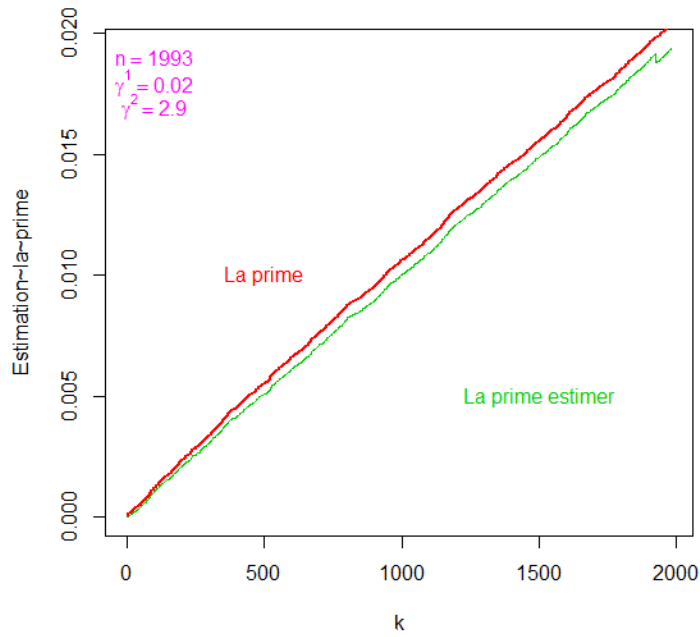


FIGURE 3.4 : Estimateur $\hat{\Pi}_2$ et la prime pour $\gamma_1 = 0.02$, $\gamma_2 = 2.9$

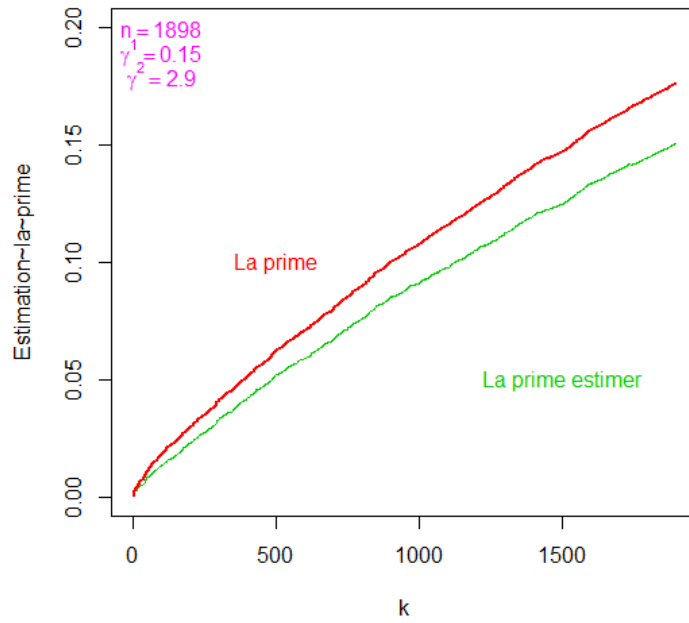


FIGURE 3.5 : Estimateur $\hat{\Pi}_2$ et la prime pour $\gamma_1 = 0.15$, $\gamma_2 = 2.9$

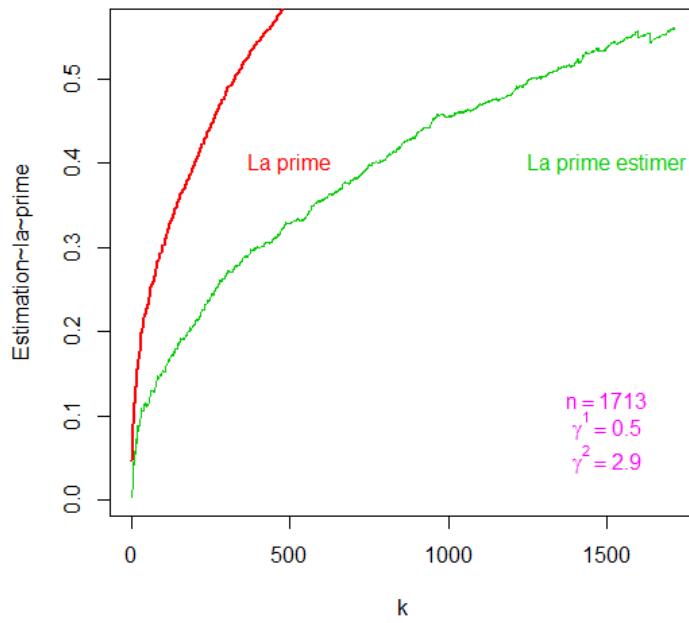


FIGURE 3.6 : Estimateur $\hat{\Pi}_2$ et la prime pour $\gamma_1 = 0.5$, $\gamma_2 = 2.9$

Commentaires.

En générale, l'estimateur de la prime est croissant en fonction du nombre des valeurs extrême k . On observe que, plus γ_1 augmente plus la troncature est remarquable.

Sur la figure 3.4 pour $\gamma_1 = 0.02$, la taille de l'échantillon s'est réduit avec 6 valeurs et l'estimateur est presque collé a la valeur théorique de la prime.

Sur la figure 3.5 pour $\gamma_1 = 0.15$, la taille de l'échantillon s'est réduit avec 102 valeurs et l'estimateur est moins collé que le graphique 3.4 mais il est quand même proche de la valeur théorique.

La figure 3.6 pour $\gamma_1 = 0.5$, la taille de l'échantillon s'est réduit avec 287 valeurs ce qui est relativement grand et qui rends clairement l'estimateur de la prime très éloigné de sa valeur théorique.

Conclusion

La troncature est une sorte de données incomplètes qu'on retrouve dans des phénomènes naturels. Nous disposons d'un échantillon X de Pareto (γ_1) tronqué par un échantillon Y de Pareto (γ_2). Soit

$$\Pi = \mathbb{E} [(X_i - u)_+] = \int_u^{+\infty} (1 - \mathbf{F}(x)) dx$$

appelée prime nette des dépassements. Dans notre travail nous avons présenté deux nouveaux estimateurs de la prime nette pour l'échantillon de risque X à partir d'un seuil appelé U ; le premier $\hat{\Pi}_1$, en utilisant l'estimateur de l'indice de valeur extrême de Worms et Worms (2015) pour un seuil fixé. Le deuxième $\hat{\Pi}_2$ via l'utilisation de l'estimateur de Haouas et al. (2016) pour un seuil aléatoire. D'après nos simulations préliminaires nous avons observé que le premier estimateur pour un seuil fixé représente un bon estimateur quelque soit le seuil pris. Par contre pour celui calculé pour un seuil aléatoire, on a observé clairement que l'estimation est bonne lorsque la troncature est réduite.

Nous avons pu sortir avec les perspectives suivantes :

1. Dans un premier temps utiliser le bootstrap pour calculer la loi limite empirique pour vérifier une éventuelle normalité asymptotique empirique pour une troncature réduite et inversement.
2. Selon les analyses faites et sur des conditions de régularité, une étude des propriétés asymptotiques de notre nouvel estimateur semblent être prometteuse.

Bibliographie

- [1] Benchaira, S., Meraghni, D., Necir, A., 2016a. Tail product-limit process for truncated data with application to extreme value index estimation. *Extremes* 19, 219-251.
- [2] Benchaira, S., Meraghni, D., Necir, A., 2016b. Kernel estimation of the tail index of a right-truncated Pareto-type distribution. *Statist-Probab. Lett.* 119, 186-193.
- [3] Benchaira, S., Meraghni, D., Necir, A., 2015. On the asymptotic normality of the extreme value index for right-truncated data. *Statist. Probab. Lett.* 107, 378-384.
- [4] CASTILLO, E., HADI, A.S. "Fitting the Generalized Pareto Distribution to data". *Journal of the American Statistical Association*, 1997.
- [5] De Haan Laurens and Ferreira Ana., 2006., «Extreme Value Theory. An Introduction», Springer. New York.
- [6] De Haan, L., Stadtmüller, U., 1996. Generalized regular variation of second order. *J. Australian Math. Soc. (Series A)* 61, 381-395.
- [7] Dakkoun, M., El Arrouchi, M. Valeur extrême et l'estimation de la prime pure en traité de réassurance en excédent de sinistre .
- [8] Gardes, L., Stuper, G. Estimating extreme quantiles under random truncation, *TEST* 24 (2015) 207-227.
- [9] Haouas, N., Necir, A., Meraghni, D., Brahimi, B., 2016. A Lynden-Bell integral estimator for the tail index of right-truncated data with a random threshold.

- [10] Hua, L., Harry, J., 2011. Second order regular variation and conditional tail expectation of multiple risks. *Insurance Math. Econom.* 49 : 537-546
- [11] Herbst, T., 1999. An application of randomly truncated data models in reserving IBNR claims. *Insurance Math. Econom.* 25 : 123-131.
- [12] Julien Worms, Rym Worms, 2015. A Lynden-Bell integral estimator for extremes of randomly truncated data.
- [13] Lynden-Bell, D., 1971. A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices Roy. Astronom. Soc.* 155 :95-118.
- [14] Lynden-Bell, D., 1971. A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices Roy. Astronom. Soc.* 155, 95-118.
- [15] Necir, A., Meraghni, D., 2009. Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts. *Insurance Math. Econom.* 45 : 49-58.
- [16] Vandewalle, B., Beirlant, J., 2006. On univariate extreme value statistics and the estimation of reinsurance premiums, statistics and the estimation of reinsurance premiums. *Insurance Math. Econom.* 38 : 441-459.
- [17] Woodroffe, M., 1985. Estimating a distribution function with truncated data. *Ann. Statist.* 13 :163-177.
- [18] Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.* 73 : 812-815.
- [19] Wang, M.C., 1989. A semiparametric model for randomly truncated data. *Statist. Assoc.*, 84 : 742-748.
- [20] Strzalkowska-Kominiak, E., Stute, W., On the probability of holes in truncated samples, *Journal of Statistical Planning and Inference* 140 (2010) 1519-1528.

Résumé.

Récemment, Worms et Worms (2015) ont présenté un estimateur de l'indice des valeurs extrêmes dans le cas des données tronquées à droite. Haouas et al. (2016) ont également présenté un nouvel estimateur pour le même indice mais pour un seuil aléatoire. Nous avons exploité ces deux estimateurs via l'intégrale de Lynden-Bell (1971) pour développer deux nouveaux estimateurs de la prime nette pour des dépassements de réassurance pour un seuil fixé et un autre aléatoire. Nous avons effectué quelques simulations simples pour illustrer nos résultats.

Mots clés : Théorie des valeurs extrêmes, Distribution à queue lourde, Indice des valeurs extrêmes, Données tronquées, Estimateur de Lynden-Bell.

Abstract.

Recently, Worms and Worms (2015) presented an extreme value index estimator for right truncated data. Haouas et al. (2016) also presented a new estimator for the same index but for a random threshold. We exploited these two estimators through the Lynden-Bell integral (1971) to develop two new net premium estimators of excess of loss reinsurance for a fixed and a further random threshold. We performed some simple simulations to illustrate our results.

Key words : Extreme Value Theory, Heavy-tailed Distribution, Extreme Value Index, Truncated Data, Lynden-Bell Estimator.