

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE.
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
SCIENTIFIQUE.

Université Mouloud Mammeri, Tizi-Ouzou
Faculté des Sciences
Département de Mathématiques



Mémoire de Master en Mathématiques Appliquées
Option : *Processus Aléatoires et Statistique de la Décision*

Thème :

Distributions à Queues Lourdes et Applications

Présenté par :

MOKRANI Tassadit

Devant le jury d'examen composé de :

Mamou Mohamed	Maître Assistant A	Président
Berkoun Youcef	Professeur	Rapporteur
Boualam Karima	Maître Assistante A	Examinatrice

Soutenu le 23/10/2016

Remerciements

Je tiens à exprimer toute ma reconnaissance à Mr Berkoun Youcef qui a proposé et accepté de diriger ce travail. Je le remercie aussi pour sa grande contribution à l'aboutissement de ce travail et pour sa grande disponibilité.

Je remercie vivement l'ensemble des membres du jury :Mr Mamou Mohamed, Mme Boualam Karima pour l'honneur qu'ils me font en acceptant de juger ce travail.

Mes remerciements chaleureux s'adressent également aux camarades et amis qui m'ont beaucoup aidé dans ce travail.

Je remercie tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste mémoire.

Je suis également reconnaissante à tous les enseignants qui m'ont formé en général et particulièrement à ceux de l'Université Mouloud Mammeri de Tizi- ouzou.

Dédicaces

Je dédie ce mémoire de fin d'études

À

Mon très cher père et ma très chère mère

en témoignage de ma reconnaissance envers le soutien, les sacrifices et tous les efforts qu'ils ont fait pour mon éducation ainsi que ma formation.

À

Mes chers frères, et mes chères sœurs

Djamila, Farida, Mohamed, Rabah, Amar, Idir, Nawel et son mari
pour leur affection, compréhension et patience.

À

Mon neveu et ma petite nièce

Adel et Lidia .

À

La mémoire de ma grand mère

Djouher .

À

toute la famille Mokrani et mes cousines

Cherifa et Fatiha

À

tous ceux qui ont une relation de près ou de loin
avec la réalisation de ce travail.

Table des matières

Remerciements	1
Dédicaces	1
Introduction générale	6
1 Distributions à queues lourdes	8
1.1 Définitions	8
1.2 Exemples de distributions à queues lourdes	9
1.3 Propriétés des distributions à queues lourdes	10
1.4 Caractérisations des distributions à queues lourdes par les moyens de taux de hasard	11
1.4.1 Notation de la fonction $r_e(x)$	12
1.5 Kurtosis	13
1.6 Caractérisation des distributions à queues lourdes	14
1.7 Limite inf des queues de convolution	18
1.8 Les sous-classes de distributions à queues lourdes	22
1.8.1 Les distributions sous-exponentielles	22
1.8.2 Fonctions à variation régulière	32
1.8.3 Les propriétés de base	34
1.8.4 Les distributions à variation dominée	36
2 Applications	38
2.1 Théorie de la ruine	38
2.1.1 Définitions et hypothèses simplificatrices	38
2.1.2 Modèle de Sparre-Andersen	38
2.1.3 Ruines avec des grands sinistres	39
2.2 Application au calcul de la VaR	40
2.2.1 La VaR	40
2.2.2 Distribution de Pareto généralisée (GPD)	41
2.2.3 Estimation des paramètres de la GPD	41
2.2.4 Distribution de l'excès	42
2.2.5 Théorème de Belkema-De Haan-Pickands	43

2.2.6	Quantile extrême	43
2.2.7	Méthode POT (Peaks Over Threshold)	43
2.2.8	Les données	45
2.2.9	Analyse descriptive des données	45
2.2.10	Analyse descriptive des logarithmes des rendements de AirPassengers	46
2.2.11	Estimation de la VaR	49
2.2.12	Modélisation par la méthode POT	49
2.2.13	Sélection de seuil	50
	Conclusion et perspectives	54
	Bibliographie	54

Table des figures

1.1	Illustration de kurtosis	14
1.2	La densité de Weibull	23
1.3	La densité log-normale	24
1.4	la densité de Pareto	33
1.5	la densité de Burr	34
2.1	Graphique de AirPassengers entre 1949 et 1960.	45
2.2	Histogramme de AirPassengers.	46
2.3	Logarithmes des rendements de AirPassengers.	47
2.4	Histogramme des log-rendements.	48
2.5	La distribution de Pareto généralisée $G_{\varepsilon,1}(x)$, pour Pareto ($\varepsilon = 0.5$), exponentielle ($\varepsilon = 0$) et Pareto de typeII ($\varepsilon = -0.5$).	49
2.6	Densité de Pareto généralisée $g_{\varepsilon,1}(x)$, pour Pareto ($\varepsilon = 0.5$), exponentielle ($\varepsilon = 0$) et Pareto de typeII ($\varepsilon = -0.5$).	50
2.7	La distribution moyenne des exès.	51

Liste des tableaux

1.1	Exemples de distributions sous-exponentielles	23
1.2	Exemples des fonctions à variation régulière	33

Introduction générale

De nombreuses distributions utilisées dans la pratique sont des distributions à queues fines, la longueur et le poids des animaux par exemple sont généralement de distributions à queues fines. Certains phénomènes aléatoires ne sont pas modélisés par de telles distributions mais par des distributions à queues lourdes.

Le premier exemple de ce type de distribution a été trouvé dans Mandelbort [15] où il a montré que le changement des prix de coton est une distribution à queue lourde. Depuis lors, de nombreux autres exemples se trouvent, parmi ceux-ci : le rendement des marchés financiers (Rachev [18], Embrechts et al [7]) et l'ampleur des tremblements de terre et les inondations (Latchman et al [13], Malamud et Turcotte [14]). Problème : comment définir les distributions à queues lourdes ? par quel moyen peut on les détecter et dans quel domaine peut on les appliquer ?

Dans ce mémoire nous intéressons à l'étude des distributions à queues lourdes.

Ce travail s'articule autour de deux chapitres. Le premier chapitre donne une définition et des exemples de distributions à queues lourdes ainsi qu'une caractérisation de ces distributions par les moyens de taux de hasard où nous introduisons un critère qui identifie si une distribution est à queue lourde ou à queue légère. Ensuite, on s'intéresse aux propriétés caractéristiques des distributions à queues lourdes, i.e. des distributions dont tous les moments exponentiels positifs sont infinis ; le principal résultat concerne les limites inférieures des queues de convolution.

Ensuite, on présentera les différentes classes des distributions à queues lourdes : on parlera plus profondément des distributions sous-exponentielles et des distributions à variation régulière qui sont les classes les plus importantes de distributions à queues lourdes et les distributions à variation dominée.

Dans le deuxième chapitre, on donnera quelques applications des distributions à queues lourdes : on commencera tout d'abord par la probabilité de ruine, ensuite, on introduit une application sur le calcul de la VaR.

Chapitre 1

Distributions à queues lourdes

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

Soit X une variable aléatoire de fonction de répartition $F(x) = \mathbb{P}[X \leq x]$. Notons $\bar{F}(x) = 1 - F(x)$, $\bar{F}(x)$ est dite fonction de queue.

1.1 Définitions

Définition 1.1.1. Une distribution F sur \mathbb{R} est dite à support non borné si $\bar{F}(x) > 0, \forall x \in \mathbb{R}$.

Définition 1.1.2. Une distribution F (ou X) est dite à queue lourde si et seulement si :

$$\int_{\mathbb{R}} e^{\epsilon x} F(dx) = \infty, \forall \epsilon > 0$$

(Ce qui signifie que la fonction génératrice n'existe pas)
D'une manière équivalente

$$\lim_{x \rightarrow \infty} \mathbb{P}[X > x + y | X > x] = \lim_{x \rightarrow \infty} \frac{\bar{F}(x + y)}{\bar{F}(x)} = 1, \forall x, y \geq 0, \bar{F}(x) > 0.$$

Cette dernière peut s'écrire comme suit :

$$\bar{F}(x + y) \sim \bar{F}(x), \forall y \geq 0 \text{ pour } x \text{ relativement grand}$$

On note L l'ensemble des distributions à queues lourdes.

Proposition 1.1.1. Une fonction f , ($f > 0$) est dite à queue lourde si et seulement si :

$$\limsup_{x \rightarrow \infty} f(x)e^{\epsilon x} = \infty, \forall \epsilon > 0$$

Remarque 1.1.3. On dit qu'une fonction non négative est à queue lourde si elle est bornée par une fonction exponentielle décroissante.

Définition 1.1.4. Une distribution F est dite à queue légère si et seulement si :

$$\int_{\mathbb{R}} e^{\epsilon x} F(dx) < \infty, \epsilon > 0$$

Ou d'une manière équivalente

$$\mathbb{E}[e^{\epsilon x}] < \infty, \epsilon > 0$$

Remarque 1.1.5. Si F est une distribution à queue légère, alors tous ses moments sont finis, i.e., $\int_0^{\infty} x^k F(dx) < \infty, k > 0$.

1.2 Exemples de distributions à queues lourdes

1. Distribution de Pareto :

Sa fonction de queue est donnée par :

$$\bar{F}(x) = \left(\frac{k}{x+k}\right)^a; k, a > 0, x \in \mathbb{R}^+$$

On a :

$$E(X^\alpha) < \infty, \forall \alpha < a$$

$$E(X^\alpha) = +\infty, \forall \alpha \geq a$$

On note la distribution de Pareto par $Pa(a, k)$

2. Distribution de Burr :

Sa fonction de queue est donnée par :

$$\bar{F}(x) = \left(\frac{k}{x^\tau + k}\right)^\alpha; k, \alpha, \tau > 0, x \in \mathbb{R}^+$$

On a :

$$E(X^\alpha) < \infty, \forall \alpha < \alpha\tau$$

$$E(X^\alpha) = \infty, \forall \alpha \geq \alpha\tau$$

Remarque 1.2.1. Pour $\tau = 1$, la distribution de Burr est équivalente dans sa queue à celle de Pareto.

3. **Distribution de Cauchy (sur \mathbb{R}) :**

Sa fonction de densité f est donnée par :

$$f(x) = \frac{k}{\pi((x - \alpha)^2 + k^2)}; k, x, \alpha \in \mathbb{R}$$
$$E(X^\alpha) < \infty, \forall \alpha < 1$$
$$E(X^\alpha) = \infty, \forall \alpha \geq 1$$

4. **Distribution lognormale (sur \mathbb{R}^+) :**

Sa fonction de densité f est donnée par :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\log(x) - \mu)^2}{2\sigma^2}}; \mu, \sigma > 0, x > 0$$
$$E(X^\alpha) < \infty, \forall \alpha.$$

5. **Distribution de Weibull (sur \mathbb{R}) :**

Sa fonction de queue \bar{F} est donnée par :

$$\bar{F}(x) = e^{-\left(\frac{x}{k}\right)^\alpha}, x \geq 0, 0 < \alpha < 1, \forall k > 0$$
$$E(X^\alpha) < \infty, \forall \alpha.$$

Remarque 1.2.2. Si $\alpha = 1$, on retrouve la distribution exponentielle.

Un exemple de distributions à queues non lourdes est la distribution exponentielle :

$$\frac{\bar{F}(x+y)}{\bar{F}(x)} = \frac{e^{-\theta(x+y)}}{e^{-\theta x}} = e^{-\theta y}.$$

1.3 Propriétés des distributions à queues lourdes

Proposition 1.3.1. Soient $X \in L$ et $Y \geq 0$ une variable aléatoire indépendante de X , alors

$$\mathbb{P}[X - Y > x] \sim \mathbb{P}[X > x]$$

preuve :

$$\begin{aligned}
\mathbb{P}[X - Y > x] &= \mathbb{P}[X - Y > x | X > x] \bar{F}(x) \\
&= \bar{F}(x) \int_0^\infty \mathbb{P}[X > x + y | X > x] f_Y(y) dy \\
&\rightarrow \bar{F}(x) \quad \text{puisque } \mathbb{P}[X > x + y | X > x] \rightarrow 1, x \rightarrow \infty
\end{aligned}$$

Lemme 1.3.1. [19]

Soit F la distribution sur $\{a + hn, n \in \mathbb{Z}\}$, $a \in \mathbb{R}$, $h > 0$ telle que $F\{a + hn\} = p_n$, alors F est à queue lourde si et seulement si la suite (p_n) est à queue lourde, i.e.,

$$\limsup_{n \rightarrow \infty} p_n e^{\epsilon n} = \infty, \forall \epsilon > 0.$$

Proposition 1.3.2.

$$X \in L \Leftrightarrow \lim_{x \rightarrow \infty} \frac{\bar{F}(x - y)}{\bar{F}(x)} = 1, \forall y \geq 0. \quad (1.1)$$

Notons que pour toute variable aléatoire X , et $\forall y \geq 0$, on a

$$\frac{\bar{F}(x - y)}{\bar{F}(x)} \geq 1$$

Par conséquent

$$\liminf_{x \rightarrow \infty} \frac{\bar{F}(x - y)}{\bar{F}(x)} \geq 1$$

Alors, (1.1) peut se réécrire comme suit :

$$X \in L \Leftrightarrow \limsup_{x \rightarrow \infty} \frac{\bar{F}(x - y)}{\bar{F}(x)} \leq 1, \forall y \geq 0$$

1.4 Caractérisations des distributions à queues lourdes par les moyens de taux de hasard

Soit $F(x)$ une fonction de distribution sur $[0, \infty)$, avec une fonction de distribution équilibrée $F_e(x)$. Nous étudions dans cette partie la fonction $r_e(x)$, qui est appelée le taux de hasard équilibré de F .

Par le comportement de la limite de $r_e(x)$, nous donnons un critère qui identifie F si une distribution à queue lourde ou bien à queue légère et nous introduisons deux grandes classes de distributions à queues lourdes.

Définition 1.4.1. Soit F une fonction de distribution et f sa densité.

La fonction $R(x) = -\ln \bar{F}(x)$ est appelée fonction de hasard de F , si $R(x)$ est différentiable

alors $r(x) = \frac{f(x)}{1 - F(x)}$ est appelée taux de hasard.

1.4.1 Notation de la fonction $r_e(x)$

Soit X une variable aléatoire de fonction de distribution F et de fonction de queue \bar{F} . Supposons que

$$\begin{cases} \mathbb{P}(X \geq 0) = 1, \bar{F}(x) > 0, \forall x \geq 0; \\ 0 < \mathbb{E}(X) = \int_0^\infty x dF(x) = \int_0^\infty \bar{F}(x) dx = \mu < \infty. \end{cases} \quad (1.2)$$

On note

$$V(x) = \int_0^x \bar{F}(u) du, \bar{V}(x) = \int_x^\infty \bar{F}(u) du \text{ et } F_e(x) = \frac{1}{u} V(x)$$

$F_e(x)$ est appelée distribution équilibrée de F .

La fonction hasard de la distribution F_e est donnée par :

$$R_e(x) = -\ln \bar{F}_e(x)$$

Et puisque $F'_e(x) = \bar{F}(x)/u$ pour tout point continu de F , on a la fonction taux de hasard de $F_e(x)$ qui suit

$$r_e(x) = \frac{\bar{F}(x)}{\int_x^\infty \bar{F}(u) du} = \frac{\bar{F}(x)}{\bar{V}(x)} \quad (1.3)$$

Dans ce qui suit, on note

$$t_F^+ = \sup\{t \geq 0 : \int_0^\infty e^{tx} dF(x) < \infty\}, a^+ = \liminf_{x \rightarrow \infty} \frac{R(x)}{x}, \alpha_F = \limsup_{x \rightarrow \infty} \frac{R(x)}{x}$$

Lemme 1.4.2. [6]

Si $a^+ > 0$, alors F est à queue légère et $t_F^+ = a^+$; si $\alpha_F = 0$, alors F est à queue lourde

Théorème 1.4.3. [6] Pour une distribution F qui satisfait (1.2), on a :

1. $\lim_{x \rightarrow \infty} r_e(x) = 0 \Rightarrow F$ est à queue lourde.
2. $\liminf_{x \rightarrow \infty} r_e(x) > 0 \Rightarrow F$ est à queue légère

preuve :

1. Pour $\epsilon > 0$ arbitraire, il existe $x_0 > 0$ tel que

$$r_e(x) = \frac{\bar{F}(x)}{\int_x^\infty \bar{F}(u) du} < \epsilon, \text{ pour } x \geq x_0$$

et

$$R_e(x) - R_e(x_0) = \log \bar{F}_e(x_0) - \log \bar{F}_e(x) = \int_{x_0}^x r_e(u) du \leq \epsilon(x - x_0),$$

ce qui signifie que

$$\frac{R_e(x) - R_e(x_0)}{x} \leq \epsilon, \quad x \geq x_0,$$

par conséquent

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

De ce fait et par le lemme 1.4.2, on sait que F_e est à queue lourde. Puisque " F est à queue lourde si et seulement si F_e est à queue lourde," on conclut la preuve de (1).

2. Dans ce cas, on a

$$t_0 = t_F^+ = a^+ = \liminf_{x \rightarrow \infty} \frac{R(x)}{x} = \liminf_{x \rightarrow \infty} \frac{\int_0^x r_e(u) du}{x} > 0,$$

et la conclusion est immédiate par le lemme 1.4.2.

Maintenant, on introduit deux classes de distributions à queues lourdes.

1. La classe : $M = \{F / \lim_{x \rightarrow \infty} r_e(x) = 0.\}$
2. La classe : $M^* = \{F / \limsup_{x \rightarrow \infty} x r_e(x) < \infty.\}$

Le kurtosis est une autre manière de déterminer si une distribution est à queue lourde.

1.5 Kurtosis

On rappelle que pour tout entier naturel n , le moment centré d'ordre n d'une variable aléatoire X , s'il existe, est défini par

$$\mu_n(X) = \mathbb{E}([X - \mathbb{E}(X)]^n).$$

Définition 1.5.1. *Le kurtosis (ou le coefficient d'aplatissement) d'une variable aléatoire est défini par :*

$$k = \frac{\mu_4}{\mu_2^2}.$$

k vaut 3 pour une distribution normale.

Un coefficient d'aplatissement élevé indique que la distribution est plutôt pointue à sa moyenne avec nécessairement des queues de distribution (épaisses).

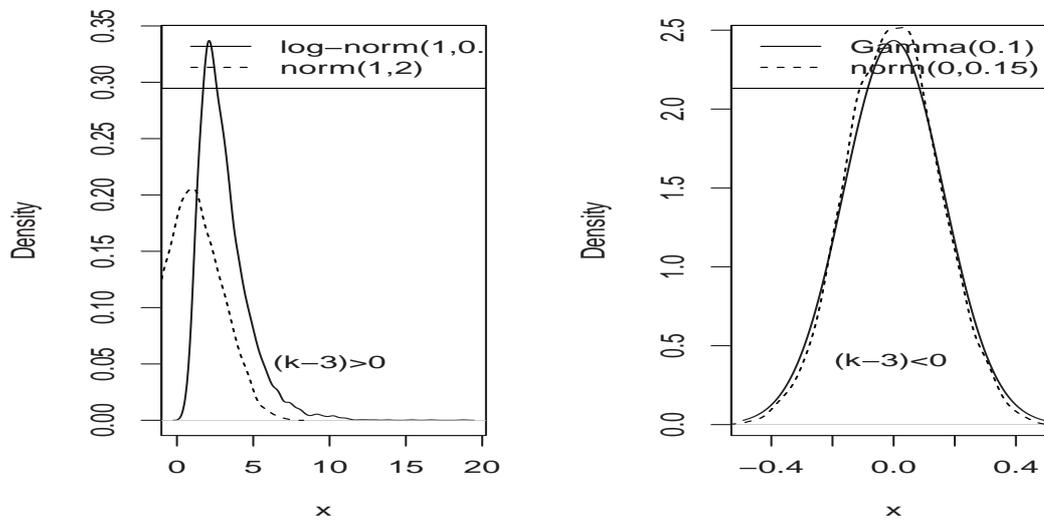


FIG. 1.1 – Illustration de kurtosis

En effet comme l'intégrale sous la densité vaut toujours 1, plus la distribution est pointue près de la moyenne plus les queues de la distribution sont chargées et donc plus le moment d'ordre 4 est important par rapport au carré du moment d'ordre 2.

On dit qu'une distribution est platykurtique si le kurtosis est négatif ($k - 3 < 0$), et elle est leptokurtique si ($k - 3 > 0$).

La figure 1.1 montre que la distribution leptokurtique a une queue plus lourde et un pic plus élevé que la distribution normale (la partie gauche de la figure), tandis que la distribution platykurtique a une queue plus légère et un pic moins élevé que la distribution normale. (La partie droite de la figure)

1.6 Caractérisation des distributions à queues lourdes

Le théorème suivant montre qu'une distribution est à queue lourde si et seulement si sa fonction de queue l'est aussi.

Théorème 1.6.1. [19]

Les assertions suivantes sont équivalentes.

1. F est une distribution à queue lourde.

2. La fonction \overline{F} est à queue lourde.

3. La fonction de hasard correspondante satisfait $\liminf_{x \rightarrow \infty} R(x)/x = 0$

4. La fonction $F[x, x+T]$ est à queue lourde, $\forall T > 0$ avec $F[x, x+T] = F[x+T] - F[x]$.

preuve :

On suppose que $F[x, x+T]$ n'est pas à queue lourde, alors :

$$c = \sup_{\epsilon' > 0} F[x, x+T] e^{\epsilon' x} < \infty.$$

Pour $\epsilon < \epsilon'$

$$\begin{aligned} \int_0^\infty e^{\epsilon x} F(dx) &\leq \sum_{n=0}^\infty e^{\epsilon(n+1)T} F(nT, nT+T) \\ &\leq c \sum_{n=0}^\infty e^{\epsilon(n+1)T} e^{-\epsilon' nT} \\ &= c e^{\epsilon T} \sum_{n=0}^\infty e^{(\epsilon - \epsilon') nT} < \infty. \end{aligned}$$

Cette intégrale est finie $\forall \epsilon \in (0, \epsilon')$, alors F n'est pas à queue lourde.

Les deux théorèmes qui suivent sont essentiels à l'étude de la convolution.

Théorème 1.6.2. Soit $\epsilon \geq 0$ une variable aléatoire de distribution à queue lourde et g une fonction vérifiant $g(x) \rightarrow \infty$ quand $x \rightarrow \infty$. Alors, il existe une fonction concave monotone $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ telle que

$$h(x) = o(x), \quad x \rightarrow \infty, \quad \mathbb{E}(e^{h(\epsilon)}) < \infty \quad \text{et} \quad \mathbb{E}(e^{h(\epsilon)+g(\epsilon)}) = \infty.$$

preuve :

On considère une fonction $h(x)$ linéaire par morceaux, et on construit deux suites (x_n) croissante et (ϵ_n) décroissante telles que

$$h(x) = h(x_{n-1}) + \epsilon_n(x - x_{n-1}), \quad x \in]x_{n-1}, x_n], \quad n \geq 1.$$

Cette fonction est monotone si $\epsilon_n > 0$. De plus elle est concave puisque ϵ_n est monotone.

Soit $x_0 = 0$ et $h(0) = 0$. Puisque ϵ est à queue lourde et $g(x) \rightarrow \infty$, on peut choisir x_1 tel que $e^{g(x)} \geq 2^{x_1}$, $\forall x > x_1$ et

$$\mathbb{E}\{e^\epsilon; \epsilon \in [x_0, x_1]\} + e^{x_1} \bar{F}(x_1) > \bar{F}(x_0) + 1$$

On choisit $\epsilon_1 > 0$ tel que

$$\mathbb{E}\{e^{\epsilon_1 \epsilon}; \epsilon \in [x_0, x_1]\} + e^{\epsilon_1 x_1} \bar{F}(x_1) = \bar{F}(0) + 1/2,$$

qui est équivalent à

$$\mathbb{E}\{e^{h(\epsilon)}; \epsilon \in [x_0, x_1]\} + e^{h(x_1)} \bar{F}(x_1) = e^{h(x_0)} \bar{F}(0) + 1/2$$

Par récurrence, on construit une suite croissante (x_n) et une autre décroissante $(\epsilon_n) > 0$ telle que $e^{g(x)} \geq 2^n$, $\forall x > x_n$ et

$$\mathbb{E}\{e^{h(\epsilon)}; \epsilon \in [x_{n-1}, x_n]\} + e^{h(x_n)} \bar{F}(x_n) = e^{h(x_{n-1})} \bar{F}(x_{n-1}) + 1/2^n, \forall n \geq 2$$

Puisque ϵ est à queue lourde et $g \rightarrow \infty$, il existe x_{n+1} tel que

$$e^{g(x)} \geq 2^{n+1}, \forall x > x_{n+1}$$

et

$$\mathbb{E}\{e^{\epsilon_n(\epsilon - x_n)}; \epsilon \in [x_n, x_{n+1}]\} + e^{\epsilon_n(x_{n+1} - x_n)} \bar{F}(x_{n+1}) > 2$$

Pour une fonction ϵ_{n+1} , la somme

$$\mathbb{E}\{e^{\epsilon_n(\epsilon - x_n)}; \epsilon \in [x_n, x_{n+1}]\} + e^{\epsilon_n(x_{n+1} - x_n)} \bar{F}(x_{n+1})$$

est continument décroissante vers $\bar{F}(x_n)$ quand $\epsilon_{n+1} \rightarrow 0$.

Par conséquent, on peut choisir $\epsilon_{n+1} \in (0, \epsilon_n)$ de façon que

$$\mathbb{E}\{e^{\epsilon_n(\epsilon - x_n)}; \epsilon \in [x_n, x_{n+1}]\} + e^{\epsilon_n(x_{n+1} - x_n)} \bar{F}(x_{n+1}) = \bar{F}(x_n) + 1/(2^{n+1} e^{h(x_n)})$$

qui est équivalent par la définition de $h(x)$ à

$$\mathbb{E}\{e^{h(\epsilon)}; \epsilon \in [x_n, x_{n+1}]\} + e^{h(x_{n+1})} \bar{F}(x_{n+1}) = e^{h(x_n)} \bar{F}(x_n) + \frac{1}{2^{n+1}}.$$

Notre hypothèse de récurrence est vérifiée avec $n + 1$ à la place de n .
Maintenant $\forall N$,

$$\begin{aligned} \mathbb{E}\{e^{h(\epsilon)}; \epsilon \leq x_{N+1}\} &= \sum_{n=0}^N \mathbb{E}\{e^{h(\epsilon)}; \epsilon \in [x_n, x_{n+1}]\} \\ &= \sum_{n=0}^N (e^{h(x_n)} \bar{F}(x_n) - e^{h(x_{n+1})} \bar{F}(x_{n+1}) + 1/2^{n+1}) \\ &\leq e^{h(x_0)} \bar{F}(x_0) + 1 \end{aligned}$$

Ce qui montre que $\mathbb{E}(e^{h(\epsilon)})$ est fini. De l'autre coté, puisque $e^{g(x)} \geq 2^n, \forall x > x_n$,

$$\begin{aligned} \mathbb{E}\{e^{h(\epsilon)+g(\epsilon)}; \epsilon > x_n\} &\geq 2^n \mathbb{E}\{e^{h(\epsilon)}; \epsilon > x_n\} \\ &\geq 2^n (\mathbb{E}\{e^{h(\epsilon)}; \epsilon \in [x_n, x_{n+1}]\} + e^{h(x_{n+1})} \overline{F}(x_{n+1})) \\ &= 2^n (e^{h(x_n)} \overline{F}(x_n) + 1/2^{n+1}). \end{aligned}$$

Alors $\mathbb{E}\{e^{h(\epsilon)+g(\epsilon)}; \epsilon > x_n\} \geq \frac{1}{2}, \forall n$, ce qui implique que $\mathbb{E}(e^{h(\epsilon)+g(\epsilon)}) = \infty$.

Notons aussi que $\lim_{n \rightarrow \infty} \epsilon_n = 0$ est nécessaire, autrement dit $\liminf_{x \rightarrow \infty} h(x)/x > 0$ et ϵ sera à queue légère.

Remarque 1.6.3.

$$\int_0^\infty x e^{(k-1) \ln x} dx = \infty \text{ et } \int_0^\infty e^{(k-1) \ln x} dx < \infty \quad (1.4)$$

(1.4) est un exemple particulier de théorème ci-dessus avec $g(x) = \ln x$. Si les moments de ϵ sont infinis, alors on peut prendre la fonction concave $h(x) = (k-1) \ln x$. Ce théorème garantit l'existence d'une fonction concave $h, \forall g$.

Remarque 1.6.4. Si ϵ a une distribution de Weibull de fonction de queue $\overline{F}(x) = e^{-x^a}, a \in (0, 1)$, et si $g(x) = \ln x$, alors on peut choisir $h(x) = (x+c)^a - \ln(x+c)$, avec $c > 0$ suffisamment grand.

En effet, on considère chaque variable ϵ de distribution à queue légère, alors $\mathbb{E}(e^{\gamma \epsilon}) < \infty, \forall \gamma > 0$.

Si $g(x) = \ln x$ et $h(x) = o(x), x \rightarrow \infty$, alors $h(x) \leq c + \frac{\gamma x}{2}, c < \infty$ d'où $\mathbb{E}(e^{h(\epsilon)+g(\epsilon)}) \leq \mathbb{E}(e^{c+\gamma \epsilon/2}) < \infty$.

Théorème 1.6.5. [19]

Soit $\epsilon \geq 0$ une variable aléatoire de distribution à queue lourde et $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ une fonction concave telle que $\mathbb{E}(e^{f(\epsilon)}) = \infty$.

Soit la fonction $g : \mathbb{R}^+ \rightarrow \mathbb{R}$ telle que $g(x) \rightarrow \infty, x \rightarrow \infty$, alors il existe une fonction concave $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ telle que $h \leq f, \mathbb{E}(e^{h(\epsilon)}) < \infty, \mathbb{E}(e^{h(\epsilon)+g(\epsilon)}) = \infty$.

1.7 Limite inf des queues de convolution

Rappelons que la convolution $F * G$ de deux mesures de probabilité F et G est donnée par :

$$(F * G)(B) = \int_{-\infty}^{\infty} F(B - y)G(dy) = \int_{-\infty}^{\infty} G(B - y)F(dy)$$

pour chaque borélien dans B où $B - y = \{x - y; x \in B\}$.

Si ϵ et η sont des variables aléatoires indépendantes de distributions F et G respectivement, alors

$$(F * G)(B) = \mathbb{P}\{\epsilon + \eta \in B\}.$$

La fonction de queue de la convolution de F et G est donnée par :

$$\begin{aligned} \overline{F * G}(x) &= \mathbb{P}\{\epsilon + \eta > x\} \\ &= \int_{-\infty}^{\infty} \overline{F}(x - y)G(dy) \\ &= \int_{-\infty}^{\infty} \overline{G}(x - y)F(dy), \forall x \in \mathbb{R} \end{aligned}$$

Soit F une distribution sur \mathbb{R}^+ , on s'intéresse à :

$$\liminf_{x \rightarrow \infty} \frac{\overline{F * F}(x)}{\overline{F}(x)},$$

dans le cas où F est à queue lourde. On commence par le résultat suivant.

Théorème 1.7.1. *Soient F_1, \dots, F_n des distributions à supports non bornés, alors*

$$\liminf_{x \rightarrow \infty} \frac{\overline{F_1 * \dots * F_n}(x)}{\overline{F_1}(x) + \dots + \overline{F_n}(x)} \geq 1.$$

preuve :

Soit $\epsilon_1, \dots, \epsilon_n$ des variables aléatoires indépendantes de distributions respectives F_1, \dots, F_n .

Du fait que les évènements $\{\epsilon_k > x, \epsilon_j \in [0, x], \forall j \neq k\}$ sont disjoints $\forall k$, alors

$$\begin{aligned}
\overline{F_1 * \dots * F_n}(x) &\geq \sum_{k=1}^n \mathbb{P}\{\epsilon_k > x, \epsilon_j \in [0, x], \forall j \neq k\} \\
&= \sum_{k=1}^n \overline{F_k}(x) \prod_{j \neq k} F_j(x) \\
&\sim \sum_{k=1}^n \overline{F_k}(x), \quad x \rightarrow \infty
\end{aligned}$$

Pour des distributions sur la droite réelle, la conclusion est généralement fausse. Il s'ensuit que pour chaque distribution F sur \mathbb{R}^+ de support non borné,

$$\liminf_{x \rightarrow \infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \geq n, \forall n \geq 2. \quad (1.5)$$

En particulier

$$\liminf_{x \rightarrow \infty} \frac{\overline{F * F}(x)}{\overline{F}(x)} \geq 2. \quad (1.6)$$

Le théorème suivant montre qu'il est suffisant que F soit à queue lourde pour avoir une égalité dans (1.6).

Théorème 1.7.2. [19]

Pour chaque distribution à queue lourde F sur \mathbb{R}^+

$$\liminf_{x \rightarrow \infty} \frac{\overline{F * F}(x)}{\overline{F}(x)} = 2. \quad (1.7)$$

preuve :

Par (1.6), il suffit de montrer que :

$$\liminf_{x \rightarrow \infty} \frac{\overline{F * F}(x)}{\overline{F}(x)} \leq 2$$

Supposons le contraire, i.e., $\exists \sigma > 0$ et x_0 tel que,

$$\overline{F * F}(x) \geq (2 + \sigma)\overline{F}(x), \quad \forall x > x_0. \quad (1.8)$$

Appliquons le théorème 1.6.2 avec $g(x) = \ln x$, on peut choisir une fonction concave croissante $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ telle que $\mathbb{E}(e^{h(\epsilon)}) < \infty$ et $\mathbb{E}(\epsilon e^{h(\epsilon)}) = \infty$.

On considère la fonction concave suivante

$$h_b(x) = \min\{h(x), bx\}, \forall b > 0$$

Puisque F est à queue lourde, $h(x) = o(x)$, $x \rightarrow \infty$, alors $\forall b$ fixé, $\forall x_1$ tel que $h_b(x) = h(x)$, $\forall x > x_1$, il existe b tel que

$$\mathbb{E}(e^{h_b(\epsilon_1)}) \leq 1 + \sigma/4. \quad (1.9)$$

$\forall \alpha, t$, soit $\alpha^{[t]} = \min\{\alpha, t\}$, alors

$$\begin{aligned} \mathbb{E}(\epsilon_1^{[t]} + \epsilon_2^{[t]})e^{h_b(\epsilon_1 + \epsilon_2)} &= 2\mathbb{E}(\epsilon_1^{[t]}e^{h_b(\epsilon_1 + \epsilon_2)}) \\ &\leq 2\mathbb{E}(\epsilon_1^{[t]}e^{h_b(\epsilon_1) + h_b(\epsilon_2)}) \end{aligned}$$

Et ce grâce à la concavité de la fonction h_b . Donc

$$\begin{aligned} \frac{\mathbb{E}((\epsilon_1^{[t]} + \epsilon_2^{[t]})e^{h_b(\epsilon_1 + \epsilon_2)})}{\mathbb{E}(\epsilon_1^{[t]}e^{h_b(\epsilon_1)})} &\leq 2 \frac{\mathbb{E}(\epsilon_1^{[t]}e^{h_b(\epsilon_1)})\mathbb{E}(e^{h_b(\epsilon_2)})}{\mathbb{E}(\epsilon_1^{[t]}e^{h_b(\epsilon_1)})} \\ &= 2\mathbb{E}(e^{h_b(\epsilon_2)}) \leq 2 + \sigma/2 \end{aligned} \quad (1.10)$$

par (1.9).

De l'autre coté, puisque $(\epsilon_1 + \epsilon_2)^{[t]} \leq \epsilon_1^{[t]} + \epsilon_2^{[t]}$

$$\begin{aligned} \frac{\mathbb{E}((\epsilon_1^{[t]} + \epsilon_2^{[t]})e^{h_b(\epsilon_1 + \epsilon_2)})}{\mathbb{E}(\epsilon_1^{[t]}e^{h_b(\epsilon_1)})} &\geq \frac{\mathbb{E}((\epsilon_1 + \epsilon_2)^{[t]}e^{h_b(\epsilon_1 + \epsilon_2)})}{\mathbb{E}(\epsilon_1^{[t]}e^{h_b(\epsilon_1)})} \\ &= \frac{\int_0^\infty x^{[t]}e^{h_b(x)}(F * F)(dx)}{\int_0^\infty x^{[t]}e^{h_b(x)}F(dx)} \end{aligned} \quad (1.11)$$

Après intégration par parties, on trouve

$$\frac{\int_0^\infty x^{[t]}e^{h_b(x)}(F * F)(dx)}{\int_0^\infty x^{[t]}e^{h_b(x)}F(dx)} = \frac{\int_0^\infty \overline{F * F}(x)d(x^{[t]}e^{h_b(x)})}{\int_0^\infty \overline{F}(x)d(x^{[t]}e^{h_b(x)})}.$$

Puisque $\mathbb{E}((\epsilon_1)e^{h_b(\epsilon_1)}) = \infty$, les deux dernières intégrales tendent vers l'infini, quand $t \rightarrow \infty$. La fonction $h_b(x)$ étant croissante, et avec l'hypothèse (1.8), on obtient :

$$\liminf_{t \rightarrow \infty} \frac{\int_0^\infty \overline{F * F}(x)d(x^{[t]}e^{h_b(x)})}{\int_0^\infty \overline{F}(x)d(x^{[t]}e^{h_b(x)})} \geq 2 + \sigma \quad (1.12)$$

En remplaçant (1.11) dans (1.12), on obtient une contradiction avec (1.10) pour t suffisamment grand.

Ce qui donne $\liminf_{x \rightarrow \infty} \frac{\overline{F * F}(x)}{\overline{F}(x)} = 2$, pas uniquement pour des distributions à queues lourdes mais aussi pour des distributions à queues légères.
On donne l'exemple suivant :

Soit F une distribution atomique aux points $x_n, n = 0, 1, \dots$; de masse p_n , i.e. $F(x_n) = p_n$. On suppose que $x_0 = 1, x_{n+1} > 2x_n, \forall n$. Alors la queue de convolution $F * F$ au point $x_n - 1$ est donnée par

$$\begin{aligned} \overline{F * F}(x_n - 1) &= (F * F)([x_n, \infty[*\mathbb{R}^+) + (F * F)([0, x_{n-1}] * [x_n, \infty[) \\ &\sim 2\overline{F}(x_n - 1), n \rightarrow \infty \end{aligned}$$

Par conséquent

$$\lim_{x \rightarrow \infty} \frac{\overline{F * F}(x_n - 1)}{\overline{F}(x_n - 1)} = 2$$

De cette équivalence et en vertu de (1.6), on a :

$$\liminf_{x \rightarrow \infty} \frac{\overline{F * F}(x_n - 1)}{\overline{F}(x_n - 1)} = 2 \quad (1.13)$$

Prenons maintenant $x_n = 3^n, n = 0, 1, \dots$ et $p_n = ce^{-3^n}$, où c est la constante normalisée, alors F est une distribution à queue légère satisfaisant la relation (1.13).

On conclut cette partie par le résultat suivant sur la convolution de distributions non identiquement distribuées.

Théorème 1.7.3. [19]

Soit F_1 et F_2 deux distributions sur \mathbb{R}^+ . Si la distribution F_1 est à queue lourde, alors

$$\liminf_{x \rightarrow \infty} \frac{\overline{F_1 * F_2}(x)}{\overline{F_1}(x)\overline{F_2}(x)} = 1. \quad (1.14)$$

preuve :

Supposons qu'il $\exists \epsilon > 0$ tel que pour x suffisamment grand

$$\frac{\overline{F_1 * F_2}(x)}{\overline{F_1}(x)\overline{F_2}(x)} \geq 1 + 2\epsilon. \quad (1.15)$$

On considère la distribution $G = (F_1 + F_2)/2$, cette distribution est à queue lourde, par le théorème 1.7.2 on obtient

$$\liminf_{x \rightarrow \infty} \frac{\overline{G * G}(x)}{\overline{G}(x)} = 2 \quad (1.16)$$

Dans l'autre coté, (1.15) et le théorème 1.7.1 impliquent que pour x suffisamment grand

$$\begin{aligned} \overline{G * G}(x) &= \frac{\overline{F_1 * F_1}(x) + \overline{F_2 * F_2}(x) + 2\overline{F_1 * F_2}(x)}{4} \\ &\geq \frac{2(1 - \epsilon)\overline{F_1}(x) + 2(1 - \epsilon)\overline{F_2}(x) + 2(1 + 2\epsilon)(\overline{F_1}(x) + \overline{F_2}(x))}{4} \\ &= 2(1 + \epsilon/2)\overline{G}(x) \end{aligned}$$

Ce qui contredit (1.16)

1.8 Les sous-classes de distributions à queues lourdes

1.8.1 Les distributions sous-exponentielles

Définition 1.8.1. Une fonction de distribution F avec $F(x) > 0, \forall x < 0$ est appelée sous-exponentielle si

$$\lim_{t \rightarrow \infty} \frac{1 - F^{*2}(t)}{1 - F(t)} = 2.$$

Ou d'une manière équivalente

$$\lim_{x \rightarrow \infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} = n, \overline{F}(x) > 0, x \geq 0, n \geq 0 \quad (1.17)$$

On note S la classe de distributions sous-exponentielles.

Exemples de distributions sous-exponentielles

Distribution	La densité f ou la queue \bar{F}	paramètres
Lognormale	$f(x) = \frac{1}{\sigma x \sqrt{2\pi} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}}$	$\mu \in \mathbb{R}, \sigma > 0$
Weibull	$\bar{F}(x) = e^{-cx^\tau}$	$c > 0, 0 < -\tau < 1$

TAB. 1.1 – Exemples de distributions sous-exponentielles

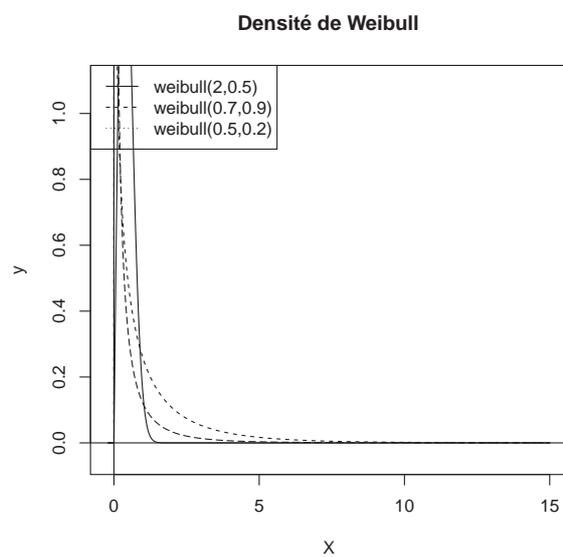


FIG. 1.2 – La densité de Weibull

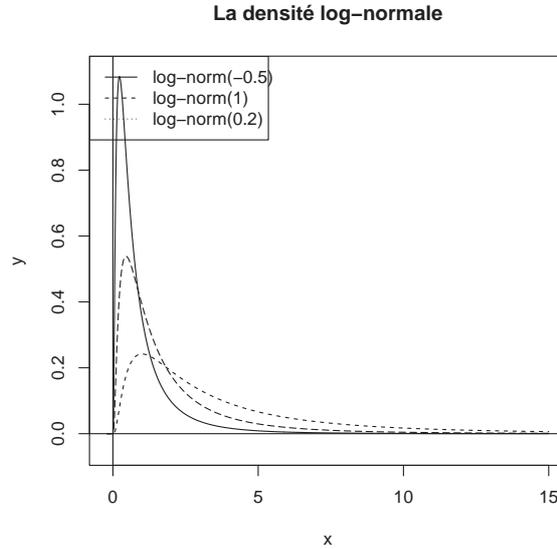


FIG. 1.3 – La densité log-normale

Pour simuler des lois usuelles avec R mais dont nous ne connaissons pas la commande, taper ? distribution Pour simuler des lois peu usuelles, aller sur <http://cran.r-project.org/>, cliquer sur "Task Views", puis sur "Distributions". Ensuite installer le package utile en tapant la commande `install.packages("nom du package")`. Nous pourrons alors utiliser ce package pour simuler la loi désirée.

Propriétés de distributions sous-exponentielles

Proposition 1.8.1. $S \subset l$, i.e. : toute distribution sous-exponentielle est à queue lourde.

preuve :

Soit $X \in S$, par la proposition 1.3.2, on montre :

$$\limsup_{x \rightarrow \infty} \frac{\bar{F}(x-y)}{\bar{F}(x)} \leq 1, y \geq 0.$$

Pour $y \leq x$, on considère

$$\begin{aligned} \mathbb{P}[X_1 + X_2 > x] &= \mathbb{P}[X_1 + X_2 > x \mathbb{I}_{\{X_1 > x\}}] + \mathbb{P}[X_1 + X_2 > x \mathbb{I}_{\{X_1 \leq y\}}] + \mathbb{P}[X_1 + X_2 > x \mathbb{I}_{\{y < X_1 \leq x\}}] \\ &= \mathbb{P}[X_1 > x] + \int_0^y \bar{F}(x-u) dF(u) + \int_y^x \bar{F}(x-u) dF(u) \end{aligned}$$

$$\begin{aligned}\frac{\mathbb{P}[X_1 + X_2 > x]}{\mathbb{P}[X_1 > x]} &= 1 + \int_0^y \frac{\overline{F}(x-u)}{\overline{F}(x)} dF(u) + \int_y^x \frac{\overline{F}(x-u)}{\overline{F}(x)} dF(u) \\ &\geq 1 + F(y) + \frac{\overline{F}(x-u)}{\overline{F}(x)} [F(x) - F(y)]\end{aligned}$$

Ce qui implique

$$\frac{\overline{F}(x-y)}{\overline{F}(x)} \leq \left[\frac{\overline{F^{*2}}(x)}{\overline{F}(x)} - 1 - F(y) \right] \frac{1}{F(x) - F(y)}$$

$\frac{\overline{F^{*2}}(x)}{\overline{F}(x)} \rightarrow 2$ quand $x \rightarrow \infty$ et

$$\limsup_{x \rightarrow \infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} \leq 1.$$

Remarque 1.8.2.

- Si (1.17) est vérifiée pour un n particulier alors elle est vérifiée $\forall n$.
- Soient X_1, \dots, X_n une suite de variables aléatoires i.i.d. (indépendantes et identiquement distribuées) de distribution F , on a :

$$\begin{aligned}\mathbb{P}[\max\{X_1, \dots, X_n\} < x] &= \mathbb{P}[X_1 < x, \dots, X_n < x] \\ &= (\mathbb{P}[X_1 < x])^n \\ &= [F(x)]^n\end{aligned}$$

$$\begin{aligned}\mathbb{P}[\max\{X_1, \dots, X_n\} > x] &= 1 - [F(x)]^n \\ &= [1 - F(x)](1 + F(x) + [F(x)]^2 + \dots + [F(x)]^{n-1}) \\ &\sim n\overline{F}(x), \quad x \rightarrow \infty.\end{aligned}$$

car $(1 + F(x) + [F(x)]^2 + \dots + [F(x)]^{n-1}) \rightarrow n, x \rightarrow \infty$.

Par définition, si X est sous-exponentielle, on a :

$$\mathbb{P}[S_n > x] = \overline{F^{*n}}(x) \sim n\overline{F}(x)$$

Ce qui veut dire que pour une distribution sous-exponentielle, la somme et le maximum des variables aléatoires se comportent de la même manière.

- Pour toute suite de variables aléatoires *i.i.d.*,

$$\mathbb{P}[X_1 + \dots + X_n > x] \geq \mathbb{P}[\max\{X_1, \dots, X_n\} > x],$$

En passant à la limite quand $x \rightarrow \infty$ on obtient :

$$\liminf_{x \rightarrow \infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \geq n.$$

Pour montrer que $X \in S$, il suffit que

$$\limsup_{x \rightarrow \infty} \frac{\overline{F^{*n}}(x)}{\overline{F}(x)} \leq n, n \geq 2.$$

Remarque 1.8.3. Si $X \in S$ alors $\mathbb{E}[e^{\epsilon X}] = \infty, \forall \epsilon > 0$, ce qui signifie que la fonction génératrice des moments n'existe pas.

Proposition 1.8.2. [19]

Si $F \in S$ et G est plus légère que F i.e.,

$$\lim_{x \rightarrow \infty} \frac{\overline{G}(x)}{\overline{F}(x)} = 0$$

alors $F * G \in S$ et $\overline{F * G}(x) \sim \overline{F}(x)$.

Définition 1.8.4 (Queues équivalentes). [5]

Deux distributions F et G de supports non bornés sont dites à queues équivalentes si

$$\lim_{x \rightarrow \infty} \overline{F}(x)/\overline{G}(x) = c \in (0, \infty).$$

Proposition 1.8.3. Si $F \in S, F$ et G sont des distributions à queues équivalentes, alors,

$$G \in S, F * G \in S \text{ et } \overline{F * G}(x) \sim (1 + c)\overline{F}(x).$$

Pour une suite de variables aléatoires *i.i.d.* sous-exponentielles, on sait que

$$\begin{aligned} \mathbb{P}[X_1 + \dots + X_n > x] &\sim n\overline{F}(x) \\ \mathbb{P}[\max\{X_1, \dots, X_n\} > x] &\sim n\overline{F}(x) \end{aligned}$$

Il s'ensuit que la somme et le maximum de variables aléatoires *i.i.d.* sous-exponentielles sont sous-exponentielles (par la proposition 1.8.3)

Proposition 1.8.4. [20]

Soit $Y = \sum_{i=1}^N X_i$ où $X = (X_i, i \geq 1)$ est une suite de variables aléatoires i.i.d., et N une variable aléatoire indépendante de X à valeurs entières et telle que $0 < \mathbb{E}[N] < \infty$.

- Si $F \in S$ et $\mathbb{E}[e^{\epsilon N}] < \infty, \epsilon > 0$ alors,

$$\mathbb{P}[Y > x] \sim \mathbb{E}[N]\bar{F}(x).$$

De plus $Y \in S$.

- Inversement, si $\mathbb{P}[Y > x] \sim \mathbb{E}[N]\bar{F}(x)$ et $\exists n \geq 2$ tel que $\mathbb{P}[N = n] > 0$ alors $F \in S$ et $Y \in S$.

Lemme 1.8.5. Si F est sous-exponentielle, alors $\forall t \in \mathbb{R}$

$$\lim_{x \rightarrow \infty} \frac{1 - F(x - t)}{1 - F(x)} = 1.$$

preuve :

Soit $t > 0$, si $x \rightarrow \infty$, on a :

$$\begin{aligned} \frac{1 - F^{*2}(x)}{1 - F(x)} - 1 &= \frac{F(x) - F^{*2}(x)}{1 - F(x)} \\ &= \int_{0-}^t \frac{1 - F(x - y)}{1 - F(x)} dF(y) + \int_t^x \frac{1 - F(x - y)}{1 - F(x)} \\ &\geq F(t) + \frac{1 - F(x - t)}{1 - F(x)} (F(x) - F(t)). \end{aligned}$$

Donc

$$1 \leq \frac{1 - F(x - t)}{1 - F(x)} \leq (F(x) - F(t))^{-1} \left(\frac{F(x) - F^{*2}(x)}{1 - F(x)} - 1 - F(t) \right)$$

Si $t < 0$

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{1 - F(x - t)}{1 - F(x)} &= \lim_{x \rightarrow \infty} \frac{1}{\frac{1 - F((x-t) - (-t))}{1 - F(x-t)}} \\ &= \lim_{y \rightarrow \infty} \frac{1}{\frac{1 - F(y - (-t))}{1 - F(y)}} \\ &= 1 \end{aligned}$$

Avec $y = x - t$

Lemme 1.8.6. Soit F une fonction de distribution sous-exponentielle et $r > 0$, alors

$$\lim_{t \rightarrow \infty} e^{rt}(1 - F(t)) = \infty$$

En particulier

$$\int_{0-}^{\infty} e^{rx} dF(x) = \infty.$$

Le lemme ci-dessous nous donne une condition pour qu'une distribution soit sous-exponentielle.

Lemme 1.8.7. Posons $\gamma(z) = \lim_{x \rightarrow \infty} \frac{1 - F(zx)}{1 - F(x)}$, $z \in (0, 1)$.

Si cette limite existe et $\gamma(z)$ est continue à gauche en 1, alors F est une fonction sous-exponentielle.

preuve :

$$F^{*2}(x) = P[X_1 + X_2 \leq x] \leq P[X_1 \leq x, X_2 \leq x] = F^2(x)$$

Pour simplifier, on pose $F(0)=0$, alors

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{1 - F^{*2}(x)}{1 - F(x)} &\geq \lim_{x \rightarrow \infty} \frac{1 - F^2(x)}{1 - F(x)} \\ &= \lim_{x \rightarrow \infty} (1 + F(x)) \\ &= 2. \end{aligned}$$

Soit $n \geq 1$ fixé, alors

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{1 - F^{*2}(x)}{1 - F(x)} &= 1 + \lim_{x \rightarrow \infty} \int_0^x \frac{1 - F(x-y)}{1 - F(x)} dF(y) \\ &\leq 1 + \sum_{k=1}^n \frac{1 - F(x - kx/n)}{1 - F(x)} (F(kx/n) - F((k-1)x/n)) \\ &= 1 + \gamma(1 - 1/n). \end{aligned}$$

car $\gamma(z)$ est continue à gauche en 1.

Exemple 1.8.8. *Considérons la distribution $Pa(\alpha, \beta)$*

$$\begin{aligned} \frac{1 - F(zx)}{1 - F(x)} &= \frac{\left(\frac{\beta}{\beta + zx}\right)^\alpha}{\left(\frac{\beta}{\beta + x}\right)^\alpha} \\ &= \left(\frac{\beta + x}{\beta + zx}\right)^\alpha \rightarrow z^{-\alpha}, \quad x \rightarrow \infty \end{aligned}$$

Ce qui montre que la distribution de Pareto est sous-exponentielle.

Lemme 1.8.9. *Soit F une distribution sous-exponentielle, alors $\forall \epsilon > 0, \exists d \in \mathbb{R}$ tel que :*

$$\frac{1 - F^{*n}(t)}{1 - F(t)} \leq d(1 + \epsilon)^n, \quad \forall t > 0, n \in \mathbb{N}.$$

preuve :

Soit $a_n = \sup_{t \geq 0} \frac{1 - F^{*n}(t)}{1 - F(t)}$, et on note :
 $1 - F^{n+1}(t) = 1 - F(t) + F * (1 - F^{*n})(t)$. Soit $T \geq 0$ tel que :

$$\sup_{t \geq T} \frac{F(t) - F^{*2}(t)}{1 - F(t)} < 1 + \epsilon/2.$$

Soit $A_T = (1 - F(t))^{-1}$. Pour $n \geq 1$

$$\begin{aligned} a_{n+1} &\leq 1 + \sup_{0 < t < T} \int_{0-}^t \frac{1 - F^{*n}(t-y)}{1 - F(t)} dF(y) + \sup_{t \geq T} \int_{0-}^t \frac{1 - F^{*n}(t-y)}{1 - F(t)} dF(y) \\ &\leq 1 + A_T + \sup_{t \geq T} \int_{0-}^t \frac{1 - F^{*n}(t-y)}{1 - F(t-y)} \frac{1 - F(t-y)}{1 - F(t)} dF(y) \\ &\leq 1 + A_T + \alpha_n \sup_{t \geq T} \frac{F(t) - F^{*2}(t)}{1 - F(t)} \leq 1 + A_T + \alpha_n(1 + \epsilon/2). \end{aligned}$$

On choisit $D = \max\left\{\frac{2(1+A_T)}{\epsilon}, 1\right\}$ et on note $\alpha_1 = 1 < D(1 + \epsilon)$. L'assertion est vérifiée par récurrence.

Lemme 1.8.10. [19] *Soit $F(x) = 0, \forall x < 0$, les assertions suivantes sont équivalentes :*

1. F est sous-exponentielle

2. Pour $n \in \mathbb{N}$

$$\lim_{x \rightarrow \infty} \frac{1 - F^{*2}(x)}{1 - F(x)} = n. \quad (1.18)$$

3. $\exists n \geq 2$ tel que (1.18) est vérifiée.

preuve :

1) \Rightarrow 2)

L'assertion est triviale pour $n \leq 2$.

Soit $\epsilon > 0$. On choisit t tel que $|\frac{1 - F^{*2}(x)}{1 - F(x)} - n| < \epsilon$

Pour $x \geq 2$

$$\frac{1 - F^{*(n+1)}(x)}{1 - F(x)} = 1 + \int_{0-}^{x-t} \frac{1 - F^{*n}(x-y)}{1 - F(x-y)} \frac{1 - F(x-y)}{1 - F(x)} dF(y) + \int_{x-t}^x \frac{1 - F^{*n}(x-y)}{1 - F(x)} dF(y).$$

La deuxième intégrale est bornée par

$$\int_{x-t}^x \frac{1 - F^{*n}(x-y)}{1 - F(x)} dF(y) \leq \frac{F(x) - F(x-t)}{1 - F(x)} = \frac{1 - F(x-t)}{1 - F(x)} - 1.$$

qui tend vers 0 quand $n \rightarrow \infty$ (grâce au lemme 1.8.5)

$$\int_{0-}^{x-t} n \frac{1 - F(x-y)}{1 - F(x)} dF(y) = n \left(\frac{F(x) - F^{*2}(x)}{1 - F(x)} - \int_{x-t}^x \frac{1 - F(x-y)}{1 - F(x)} dF(y) \right) \rightarrow n, \quad x \rightarrow \infty$$

Alors

$$\left| \int_{0-}^{x-t} \left(\frac{1 - F^{*n}(x-y)}{1 - F(x-y)} - n \right) \frac{1 - F(x-y)}{1 - F(x)} dF(y) \right| \leq \epsilon \left(\frac{F(x) - F^{*2}(x)}{1 - F(x)} - \int_{x-t}^x \frac{1 - F(x-y)}{1 - F(x)} dF(y) \right) \rightarrow \epsilon.$$

Alors

$$\overline{\lim}_{x \rightarrow \infty} \left| \frac{1 - F^{*(n+1)}(x)}{1 - F(x)} - (n+1) \right| \leq \epsilon.$$

Ce qui prouve l'assertion car ϵ est arbitraire.

2) \Rightarrow 3) est trivial.

3) \Rightarrow 1)

Supposons que (1.18) est vraie pour $n > 2$. Elle est donc vraie pour $n - 1$. puisque

$$\begin{aligned} \frac{1 - F^{*n}(x)}{1 - F(x)} - 1 &= \int_{0-}^x \frac{1 - F^{*(n-1)}(x-y)}{1 - F(x)} dF(y) \\ &\geq F(x) \frac{1 - F^{*(n-1)}(x)}{1 - F(x)} \end{aligned}$$

Alors

$$\overline{\lim}_{x \rightarrow \infty} \frac{1 - F^{*(n-1)}(x)}{1 - F(x)} \leq n - 1$$

Car $F^{*(n-1)}(x) \leq F(x)^{n-1}$ est triviale donc

$$\lim_{x \rightarrow \infty} \frac{1 - F^{*(n-1)}(x - y)}{1 - F(x)} \geq n - 1$$

Il s'ensuit que (1.18) est vérifiée pour $n = 2$ et F est une distribution sous-exponentielle.

Lemme 1.8.11. *Soit U et V deux distributions avec $U(x) = V(x) = 0, \forall x < 0$. Alors*

$$1 - V(x) \sim \alpha(1 - U(x)), x \rightarrow \infty, \alpha > 0.$$

Si U est sous-exponentielle alors V l'est aussi.

preuve :

On veut montrer que :

$$\lim_{x \rightarrow \infty} \int_{0-}^x \frac{1 - V(x - y)}{1 - V(x)} dV(y) \leq 1.$$

Soit $0 < \epsilon < 1, \exists y_0, \forall y \geq y_0$

$$(1 - \epsilon)\alpha \leq \frac{1 - V(y)}{1 - U(y)} \leq (1 + \epsilon)\alpha.$$

Soit $x > y_0$

$$\begin{aligned} \int_{x-y_0}^x \frac{1 - V(x - y)}{1 - V(x)} dV(y) &\leq \frac{V(x) - V(x - y_0)}{1 - V(x)} \\ &= \frac{1 - V(x - y_0)}{\frac{1 - U(x - y_0)}{1 - U(x)} \frac{1 - U(x - y_0)}{1 - U(x)} - 1} \rightarrow 0 \end{aligned}$$

par le lemme 1.8.5

$$\begin{aligned} \int_{0-}^{x-y_0} \frac{1 - V(x - y)}{1 - V(x)} dV(y) &\leq \frac{1 + \epsilon}{1 - \epsilon} \int_{0-}^{x-y_0} \frac{1 - U(x - y)}{1 - U(x)} dV(y) \\ &\leq \frac{1 + \epsilon}{1 - \epsilon} \int_{0-}^x \frac{1 - U(x - y)}{1 - U(x)} dV(y). \end{aligned}$$

$$\begin{aligned}
\frac{1+\epsilon}{1-\epsilon} \int_{0-}^x \frac{1-U(x-y)}{1-U(x)} dV(y) &= \frac{V(x) - U * V(x)}{1-U(x)} \\
&= 1 - \frac{1-V(x)}{1-U(x)} + \frac{U(x) - V * U(x)}{1-U(x)} \\
&= 1 - \frac{1-V(x)}{1-U(x)} + \int_{0-}^x \frac{1-V(x-y)}{1-U(x)} dU(y).
\end{aligned}$$

$1 - (1 - U(x))^{-1}(1 - V(x))$ tend vers $1 - \alpha$, $x \rightarrow \infty$, et

$$\int_{0-}^x \frac{1-V(x-y)}{1-U(x)} dU(y) \leq \frac{U(x) - U(x-y_0)}{1-U(x)} \text{ tend vers } 0.$$

Finalement,

$$\begin{aligned}
\int_{0-}^{x-y_0} \frac{1-V(x-y)}{1-V(x)} dU(y) &\leq (1+\epsilon)\alpha \int_{0-}^{x-y_0} \frac{1-U(x-y)}{1-U(x)} dU(y) \\
&\leq (1+\epsilon)\alpha \int_{0-}^x \frac{1-U(x-y)}{1-U(x)} dU(y)
\end{aligned}$$

Qui tend vers $(1+\epsilon)\alpha$. En additionnant ces limites, on obtient :

$$\overline{\lim}_{x \rightarrow \infty} \int_{0-}^x \frac{1-V(x-y)}{1-V(x)} dV(y) \leq \frac{(1+\epsilon)(1+\alpha\epsilon)}{1-\epsilon} \epsilon \text{ arbitraire.}$$

Car ϵ est arbitraire.

Le concept de la variation régulière est largement utilisé dans la théorie des valeurs extrêmes. Nous résumons ici quelques principaux résultats de la théorie de la variation régulière.

1.8.2 Fonctions à variation régulière

Définition 1.8.12. [3]

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

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

On note $g \in RV_\alpha$; α est appelé indice (ou exposant) de la fonction à variation régulière.

Remarque 1.8.13. Soit $\alpha > -1$. Une primitive de la fonction x^α sur \mathbb{R}_+ est la fonction $(1+\alpha)^{-1}x^{1+\alpha}$; plus généralement, toutes les primitives de x^α sont équivalentes à l'infini, et à variation régulière d'indice $(1+\alpha)$.

Exemples des fonctions à variations régulière

Distribution	la distribution F ou la queue \bar{F}	Indice de la variation régulière
Pareto	$\bar{F}(x) = x^{-\alpha}, F(x) = \begin{cases} 1 - cx^{-\gamma} & \text{si } x \geq x_m \\ 0 & \text{si } x < x_m \end{cases}$	$-\alpha$
Burr	$\bar{F}(x) = (\frac{1}{x^\tau} + 1)^\alpha$	$-\tau\alpha$

TAB. 1.2 – Exemples des fonctions à variation régulière

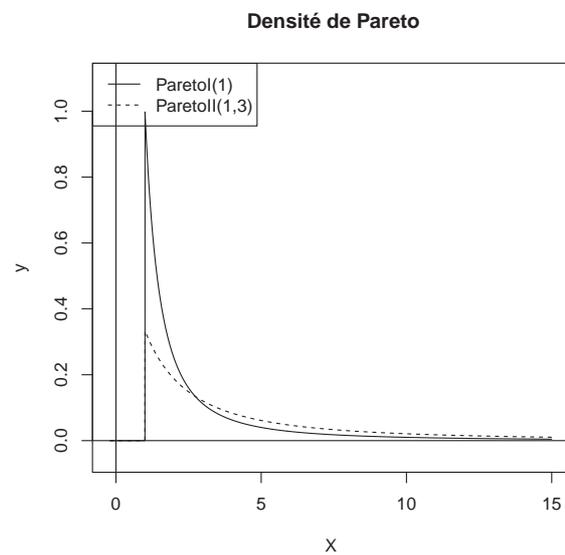


FIG. 1.4 – la densité de Pareto

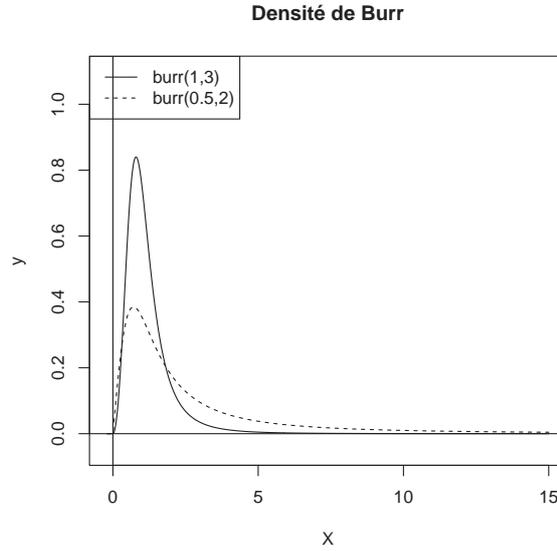


FIG. 1.5 – la densité de Burr

Remarque 1.8.14. Dans le cas particulier où $\alpha = 0$, on dit que g est à variation lente, i.e.,

$$\lim_{t \rightarrow \infty} \frac{g(tx)}{g(t)} = 1, \forall x > 0.$$

Les fonctions à variation lente sont notées par $\ell(x)$.

Conséquences

Une fonction $g : \mathbb{R}_+ \rightarrow \mathbb{R}$ mesurable est à variation régulière d'indice α au voisinage de 0 notée $g \in RV_\alpha^0$, si pour tout $x > 0$, $\lim_{t \rightarrow 0} \frac{g(tx)}{g(t)} = x^\alpha$. $g(1/x)$ est donc à variation régulière d'indice $(-\alpha)$ à l'infini.

Lemme 1.8.15. [3] [inverse généralisé d'une fonction à variation régulière]

- Si g est à variation régulière d'indice $\alpha > 0$, à l'infini, alors son inverse généralisée est à variation régulière d'indice $1/\alpha > 0$, (à l'infini).
- Si g est à variation régulière d'indice $\alpha < 0$, à l'infini, alors son inverse généralisée est à variation régulière d'indice $-1/\alpha > 0$, (à l'infini).

1.8.3 Les propriétés de base

En introduisant certaines des propriétés les plus importantes des fonctions à variation régulière, on a :

Proposition 1.8.5. Soient $a \in \mathbb{R}$ et $g \in RV_a$, alors il existe une fonction à variation lente ℓ à l'infini telle que :

$$\forall x > 0, g(x) = x^a \ell(x).$$

Ce résultat montre que l'étude des fonctions à variation régulière à l'infini se ramène à celle des fonctions à variation lente. Parmi les fonctions à variation lente, on peut citer :

1. Les fonctions possédant une limite strictement positive à l'infini.
2. Les fonctions de la forme $g : x \mapsto |\log x|^\beta$, $\beta \in \mathbb{R}$.
3. Les fonctions g telles que $\exists M > 0, \forall x \geq M, g(x) = c + dx^{-\beta}(1 + o(1))$ où $c, \beta > 0$ et $d \in \mathbb{R}$.

L'ensemble de ces fonctions g est appelé classe de Hall.

Le résultat suivant étend cette propriété à l'ensemble des fonctions à variation régulière.

Théorème 1.8.16 (Théorème de convergence uniforme). [3]

Soit α un nombre réel et g une fonction à variation régulière d'indice α à l'infini, alors :

$$\lim_{y \rightarrow +\infty} \frac{g(xy)}{g(y)} = x^\alpha,$$

et la convergence est uniforme en x sur tout compact de $(0, +\infty)$

Le théorème de Potter en découle :

Théorème 1.8.17 (Théorème de Potter). [3]

Soit α un nombre réel et g une fonction à variation régulière d'indice α en $+\infty$, alors pour tout $A > 1$, $\epsilon > 0$ et pour tous x, y suffisamment grands

$$\frac{g(y)}{g(x)} \leq A \max\left\{\frac{y^{\alpha+\epsilon}}{x}, \frac{y^{\alpha-\epsilon}}{x}\right\}$$

Théorème 1.8.18 (Représentation de Karamata).

1. l est à variation lente si et seulement si

$$l(x) = C(x) e^{\int_1^x t^{-1} \epsilon(t) dt}, \quad x > 0$$

$\lim_{x \rightarrow \infty} C(x) = c \in (0, +\infty)$ et $\lim_{t \rightarrow \infty} \epsilon(t) = 0$; $\epsilon, C : \mathbb{R}^+ \rightarrow \mathbb{R}^+$

2. Une fonction $U : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est à variation régulière d'indice ρ si et seulement si

$$U(x) = C(x)e^{\int_1^x t^{-1}\rho(t)dt}$$

$$\lim_{t \rightarrow +\infty} \rho(t) = \rho.$$

Théorème 1.8.19 (Théorème taubérien de Karamata). [12]

Soit $\alpha \geq 0$ et g une fonction positive, croissante sur \mathbb{R}_+ et nulle en 0, alors g est à variation régulière d'indice α à l'infini si et seulement si la fonction $\lambda \mapsto \int_{\mathbb{R}_+} e^{-st} d\gamma(t)$ est à variation régulière d'indice $(-\alpha)$ en 0. De plus,

$$\int_{\mathbb{R}_+} e^{-st} dg(t) \sim \gamma(1+\alpha)\gamma(1/s) \text{ en } 0.$$

Théorème 1.8.20 (Théorème taubérien de Karamata, $\alpha < 1$). [12]

Soit $\alpha < 1$ et g une fonction strictement positive et croissante sur \mathbb{R}_+ . Les propositions suivantes sont équivalentes

- g est à variation régulière d'indice α à l'infini.
- $\int_{\mathbb{R}_+} e^{-st} dg(t)$ est à variation régulière d'indice $\alpha - 1$ en 0

De plus, si ces propositions sont vérifiées, alors :

$$\int_{\mathbb{R}_+} \frac{1}{g(t)} e^{-st} dt \sim \frac{\gamma(1-\alpha)}{s\gamma(1/s)} \text{ en } 0$$

Théorème 1.8.21 (Théorème taubérien de Karamata, $\alpha = 1$). [12]

Soit g une fonction strictement positive, croissante sur \mathbb{R}_+ et à variation régulière d'indice 1 à l'infini, alors :

$$\int_{\mathbb{R}_+} \frac{1}{g(t)} e^{-st} dt \sim \int_0^{1/s} \frac{1}{g(t)} dt \text{ en } 0,$$

et le membre de droite est à variation lente en (0).

Proposition 1.8.6. Pour toute fonction à variation lente ℓ à l'infini, on a

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

1.8.4 Les distributions à variation dominée

On note D l'ensemble des fonctions à variation dominée tel que

$$D = \{F \text{ sur } (0, \infty) / \limsup_{x \rightarrow \infty} \frac{\overline{F}(\frac{x}{2})}{\overline{F}(x)} < \infty\}$$

Proposition 1.8.7. [9]

Si \overline{F} est à variation dominée alors, la distribution F est aussi sous-exponentielle.

On a les relations :

- $R \subset S \subset l$ et $R \subset D$
- $l \subset D \subset S$

Chapitre 2

Applications

2.1 Théorie de la ruine

L'étude de la probabilité de ruine dépend significativement du type de la loi de X (queue fine ou lourde).

2.1.1 Définitions et hypothèses simplificatrices

- On note Y_i le $i^{\text{ème}}$ sinistre pour l'assureur.
- Entre une date 0 et une date t , il ya $N(t)$ sinistres.

Définition 2.1.1 (Processus de renouvellement). *Un processus de renouvellement $N(t)$ est un processus de comptage où les inter-arrivées sont i.i.d. Soient Y_i , $i = 1, \dots, n$ les inter-arrivées entre le $(i - 1)$ et le $i^{\text{ème}}$ évènement.*

$$S_n = \sum_1^n Y_i, \quad S_0 = 0, \quad n \geq 1$$

$$N(t) = \sup_{n \geq 1} \{n/S_n \leq t\}, \quad t \geq 0$$

Lorsque $N(t)$ est un processus de renouvellement, le modèle est connu sous le nom de Sparre-Andersen.

2.1.2 Modèle de Sparre-Andersen

Ecriture du processus sous forme d'une marche aléatoire

$$R(t) = u + ct - S(t)$$

$$= u - \sum_{i=1}^{N(t)} Z_i, \quad Z_i = (Y_i - c\Delta T_i)$$

avec

- $R(t)$ est les réserves d'une compagnie d'assurance
- $\Delta T_i = T_i - T_{i-1}$, $(T_i)_i$ sont les temps d'inter-occurrence des sinistres indépendants et identiquement distribués
- u est la réserve initiale.

- c est les primes reçues par unités de temps.

Définition 2.1.2. [10] On appelle ruine l'évènement $\{\exists t \geq 0 : R(t) < 0\}$ et

- Le temps de ruine est alors $T_u = \inf\{t \geq 0 : R(t) < 0\} \in [0, \infty]$ avec la convention que $\inf \emptyset = \infty$
- La probabilité de ruine en un horizon infini $\psi(u) = P(\exists t \geq 0 : R(t) < 0 | R(0) = u) = P(T_u < \infty)$, $u > 0$.

Théorème 2.1.3. Dans le modèle de Sparre Andersen, notons Y les incréments de la perte agrégée $Y = X - cT$. x_0 est défini comme le supremum de l'ensemble $\{x, F_Y(x) < 1\}$. Pour $u \geq 0$, nous disposons de l'encadrement suivant

$$b_- e^{-\gamma u} \leq \psi(u) \leq b_+ e^{-\gamma u}$$

où γ est solution de l'équation $M_X(r)M_T(-rc) = 1$, et $M_X(r) = \mathbb{E}(e^{rX})$, $M_T(-rc) = \mathbb{E}(e^{-rcT})$.

Les constantes b_-, b_+ ont pour expression

$$b_- = \inf_{x \in [0, x_0[} \frac{e^{\gamma x} \bar{F}_Y(x)}{\int_x^{+\infty} e^{\gamma y} d\bar{F}_Y(y)} \quad \text{et} \quad b_+ = \sup_{x \in [0, x_0[} \frac{e^{\gamma x} \bar{F}_Y(x)}{\int_x^{+\infty} e^{\gamma y} d\bar{F}_Y(y)}$$

2.1.3 Ruines avec des grands sinistres

Théorème 2.1.4. [8] Dans le modèle de Sparre Andersen, où les espérances des montants $(X_i)_i$ et des temps d'attente des sinistres $(T_i)_i$ sont finis et

$\mathbb{E}(X) < c\mathbb{E}(T)$. Notons $F_{X,0} = \frac{1}{\mathbb{E}(X)} \int_0^x \bar{F}_X(y) d(y)$. Si F_X et $F_{X,0}$ appartiennent à la classe sous-exponentielle, on a alors

$$\psi(u) \sim \frac{1}{c\mathbb{E}(T) - \mathbb{E}(X)} \int_u^{+\infty} \bar{F}_X(y) d(y), \quad u \rightarrow +\infty$$

Ce théorème donne lieu aux cas particuliers suivants. Considérons des montants de sinistre Pareto $Pa(k, \alpha)$, c'est à dire $P(X > x) = (k/x)^\alpha$ avec $\alpha > 1$. On a alors

$$\psi(u) \sim \frac{k}{(\alpha - 1)(c\mathbb{E}(T) - \alpha k)} \left(\frac{k}{u}\right)^{\alpha-1}, u \rightarrow +\infty$$

De manière similaire, pour les lois à variations régulières dont les queues de distribution vérifient $P(X > x) \sim \ell(x)/x^\alpha$ pour des grandes valeurs de x et ℓ une fonction à variation lente, telle que $\ell(xt)/\ell(x) \rightarrow 1$ pour $t > 0$ et $x \rightarrow +\infty$, nous obtenons

$$\psi(u) \sim \frac{1}{c\mathbb{E}(T) - \alpha k} \frac{\ell(u)}{(\alpha - 1)u^{\alpha-1}}, u \rightarrow +\infty.$$

2.2 Application au calcul de la VaR

2.2.1 La VaR

La VaR est apparue (sous ce nom) dans les années 90, en réponse à de nombreux désastres qui ont touché les marchés de capitaux à cette période.

Définition 2.2.1 (L'inverse généralisé). [1]

Soit X une variable aléatoire de fonction de répartition F connue, on définit l'inverse généralisé de F , noté F^{-1} par :

$$\begin{aligned} F_X^{-1}(\alpha) &= \inf\{x \in \mathbb{R}/F(x) \geq \alpha\}; 0 < \alpha < 1 \\ &= \sup\{x \in \mathbb{R}/F(x) \leq \alpha\} \end{aligned}$$

Selon Esch et al en 1997 et Jorion en 2000 , la VaR est définie par

Définition 2.2.2 (La VaR). [4]

La VaR au niveau α est définie par le quantile de niveau α (α -quantile) :

$$P[X \leq VaR_\alpha] = \alpha \Leftrightarrow F(VaR_\alpha) = \alpha$$

D'où

$$VaR_\alpha(X) = F_X^{-1}(\alpha)$$

Définition 2.2.3. L'expected shortfall est une mesure de risque, elle mesure la taille moyenne des pertes sachant que celle-ci dépasse la VaR.

L'expected shortfall au seuil α d'une distribution F_X de X , notée $ES_\alpha(X)$ est définie par

$$ES_\alpha(X) = \frac{1}{1 - \alpha} \int_\alpha^1 VaR_\mu(X) d\mu$$

2.2.2 Distribution de Pareto généralisée (GPD)

Lorsque nous avons des données issues d'une distribution inconnue, il est possible d'approximer la distribution au-delà d'un certain seuil par une distribution de Pareto généralisée qui est une distribution à queue lourde.

Définition 2.2.4. Une distribution $G_{\xi,\beta}$ est dite de Pareto généralisée de paramètres $\xi \in \mathbb{R}$ et $\beta > 0$, si elle s'écrit

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

Cette distribution est définie pour

$$\begin{aligned} x &\geq 0, \text{ si } \xi \geq 0 \\ 0 < x < -\beta/\xi, \text{ si } \xi < 0. \end{aligned}$$

Le paramètre ξ a un effet important sur le caractère de la distribution

- Pour $\xi > 0$, on a une distribution de Pareto d'indice de queue $\gamma = 1/\xi$,
- Pour $\xi = 0$, on a la distribution exponentielle,
- Pour $\xi < 0$, on a une distribution sur l'intervalle borné $[0, -\beta/\xi]$ (appelée distribution de Pareto de type II.)

2.2.3 Estimation des paramètres de la GPD

La méthode du maximum de vraisemblance

Une fois le seuil optimal choisi, on construit une nouvelle série d'observations au-dessus de ce seuil et la distribution de ces données suit approximativement une distribution généralisée de Pareto

La fonction de densité de la distribution GPD est

$$g(x) = \begin{cases} \beta^{\frac{1}{\xi}} (\beta + \xi x)^{-\frac{1}{\xi}-1}, & \text{si } \xi \neq 0; \\ \beta^{-1} e^{-\frac{x}{\beta}} & \text{si } \xi = 0. \end{cases}$$

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

$$\log l(\xi, \beta; X_1, \dots, X_n) = -\log \beta - (1 + \frac{1}{\xi}) \sum_{i=1}^n \log(1 + \frac{\xi}{\beta} X_i)$$

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

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

L'estimateur de maximum de vraisemblance de (ξ, τ) est $(\widehat{\xi} = \widehat{\xi}(\widehat{\tau}), \widehat{\tau})$, où $\widehat{\tau}$ est solution de

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

2.2.4 Distribution de l'excès

Définition 2.2.5 ((Fonction de distribution de l'excès)). [17]

Considérons un échantillon X_1, \dots, X_n , de la variable aléatoire X de distribution F et de point terminal $x_F \leq \infty$. Alors, pour un seuil $u < x_F$ fixé, on s'intéresse aux excès au-dessus du seuil u , définies par $X_i - u$, $i = 1, \dots, n$. La fonction de distribution de l'excès de X au dessus du seuil u est

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

Pour $0 < x < x_F - u$

Définition 2.2.6 ((La fonction moyenne des excès)). [11]

La fonction moyenne des excès de X au dessus du seuil u est

$$e(u) = E(X - u / X > u).$$

Notons

$$N_u = \text{card}\{i \in \{1, \dots, n\}, X_i > u\}$$

L'estimateur empirique de la fonction moyenne des excès est défini comme :

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n I\{X_i > u\}} = \frac{1}{N_u} \sum_{i=1}^n (X_i - u)^+$$

C'est la somme des excès au-dessus du seuil u divisé par le nombre N_u de données qui excèdent u . La loi essentielle pour modéliser les excès est la GPD, sa fonction moyenne des excès est donnée par

$$e(u) = \frac{\beta + \xi u}{1 - \xi}.$$

Le théorème essentiel pour la modélisation des excès est celui de Balkema-de Haan- Pickands déterminé en 1974 – 1975.

Définition 2.2.7. *Le point terminal d'une distribution F est défini par :*

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

2.2.5 Théorème de Belkema-De Haan-Pickands

Le théorème suivant fait le lien entre le comportement asymptotique de la distribution des excès et la loi de Pareto généralisée.

Théorème 2.2.8. [2]

Soit F_u la distribution des excès; lorsque le seuil u tend vers le point terminal x_F , on a :

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

où $\beta(u)$ est une fonction mesurable positive.

2.2.6 Quantile extrême

On appelle quantile extrême d'ordre $(1 - p)$

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

où p est proche de 0.

2.2.7 Méthode POT (Peaks Over Threshold)

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

La méthode consiste à suivre les étapes suivantes :

1. Soit X_1, \dots, X_n , un échantillon; on choisit un certain seuil élevé u . Notons N_u le nombre d'observations qui dépassent le seuil u .

On définit les variables aléatoires des excès, $Y_i - u \geq 0$.

2. Soit Y_1, \dots, Y_{N_u} un échantillon des excès au dessus du seuil u , de distribution conditionnelle

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

Le théorème 2.2.8 nous donne l'approximation de cette distribution :

$$F_u(x) \approx G_{\xi, \beta}(x), \quad x > 0, \quad u \rightarrow +\infty$$

Ou encore

$$\bar{F}_u(x) \approx \bar{G}_{\xi, \beta}(x), \quad x > 0, \quad u \rightarrow +\infty.$$

3. La loi F peut se décomposer de la manière suivante :

$$F(x) = \mathbb{P}[X > u] \times F_u(x - u) + \mathbb{P}[X \leq u], \quad \text{pour } x > u.$$

Cette distribution peut être estimée par :

$$\hat{F}(x) = (1 - F_n(u)) \times G_{\hat{\xi}, \hat{\beta}} + F_n(u), \quad \text{pour } x > u. \quad (2.1)$$

où $F_n(u)$ est la fonction de répartition empirique au point u , tel que

$$F_n(u) = 1 - \frac{N_u}{n},$$

donc \hat{F} est la loi GPD de paramètres $\xi = \hat{\xi}$, $\beta = \hat{\beta}(1 - F_n(u))^{\hat{\xi}}$

et $\mu = u - \frac{\hat{\beta}}{\hat{\xi}}((1 - F_n(u))^{-\hat{\xi}} - 1)$.

On peut écrire aussi :

L'estimateur pour la queue $\bar{F}(u+x)$ prend la forme

$$\hat{\bar{F}}(u+x) = \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x}{\hat{\beta}}\right)^{-\frac{1}{\hat{\xi}}}$$

et

$$\hat{F}(x) = 1 - \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x}{\hat{\beta}}\right)^{-\frac{1}{\hat{\xi}}}$$

D'où $\hat{F}(x) = 1 - \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x - u}{\hat{\beta}}\right)^{-\frac{1}{\hat{\xi}}}$

4. Finalement on inverse la fonction définie dans (2.1) et l'estimateur du quantile extrême est donné par :

$$\hat{x}_p = G_{\hat{\xi}, \hat{\beta}}^{-1}\left(\frac{1 - p - F_n(u)}{1 - F_n(u)}\right)u + \frac{\hat{\beta}}{\hat{\xi}} \left[\left(\frac{n}{N_u} (1 - p)^{-\hat{\xi}} - 1\right)\right]$$

On obtient

$$\widehat{x}_p^{POT} = u + \frac{\widehat{\beta}}{\widehat{\xi}} \left[\left(\frac{n}{N_u} (1-p)^{-\widehat{\xi}} - 1 \right) \right].$$

2.2.8 Les données

On s'intéresse aux données qui représentent le nombre de réservations de billets internationaux (en milliers) par mois de la Pan Am pour la période 1949 – 1960 (AirPassengers). Ces données sont dans le logiciel *R*.

2.2.9 Analyse descriptive des données

On commence par donner les représentations graphiques : Le graphe ainsi que l'histogramme représentatif des données.

```
> AP<-AirPassengers  
> plot(AP,type='l',xlab="Time")
```

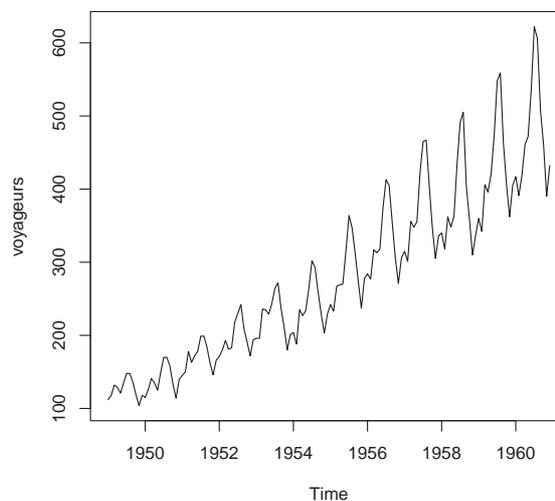


FIG. 2.1 – Graphique de AirPassengers entre 1949 et 1960.

On remarque que le nombre de voyageurs augmente avec le temps.

```
>hist(AP)
```

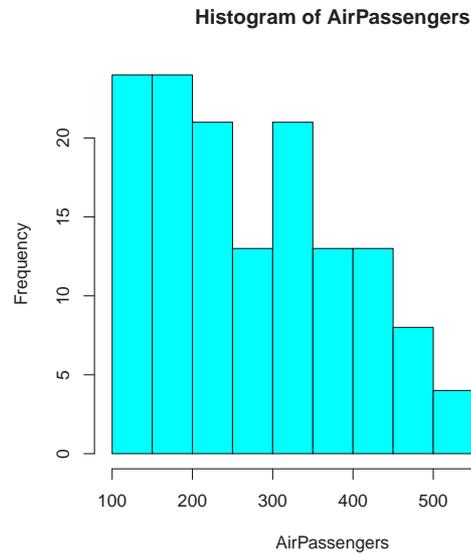


FIG. 2.2 – Histogramme de AirPassengers.

```
>summary(AP)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
104.0	180.0	265.5	280.3	360.5	622.0

2.2.10 Analyse descriptive des logarithmes des rendements de AirPassengers

L'objectif de la procédure suivante est d'obtenir une série des logarithmes des rendements puis calculer la VaR.

Soit P_t le nombre de réservations de billets à la date t :

Définition 2.2.9 (Les rendements).

Les rendements sont définis par l'expression suivante :

$$R_t = \frac{P_t}{P_{t-1}} - 1$$

Définition 2.2.10 (Les log-rendements).

Les log-rendements sont donnés par :

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

Comme on peut utiliser directement la fonction `log` et l'opérateur "`diff`" qui se trouve dans le package "`base`" :

```
log.rap <- diff(log(AP))
```

on affiche le graphe du log-rendement du AirPassengers par :

```
>plot(log.rap,type='l')
```

ainsi que l'histogramme des rendements du AirPassengers et la courbe de la loi normale $N(\mu, \sigma^2)$.

Où $\mu = \text{mean}(\text{log.rap})$ et $\sigma^2 = \text{sd}(\text{log.rap})$ sont la moyenne et la variance calculées à partir de la série des logarithmes des rendements.

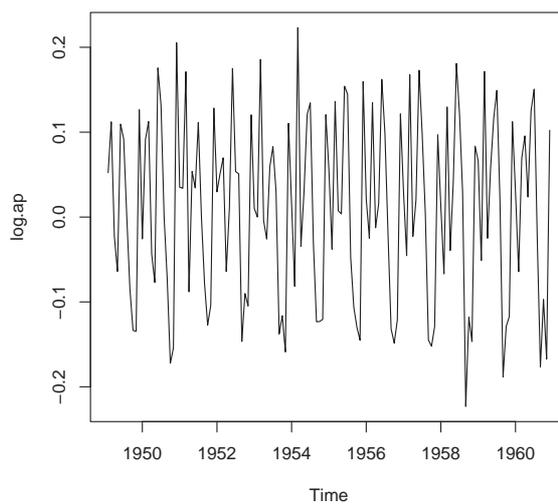


FIG. 2.3 – Logarithmes des rendements de AirPassengers.

```
>hist(log.rap,prob=T,breaks=100,col="light blue")
>curve(dnorm(x,mean(log.rap),sd(log.rap)),add=T, lwd=2)
```

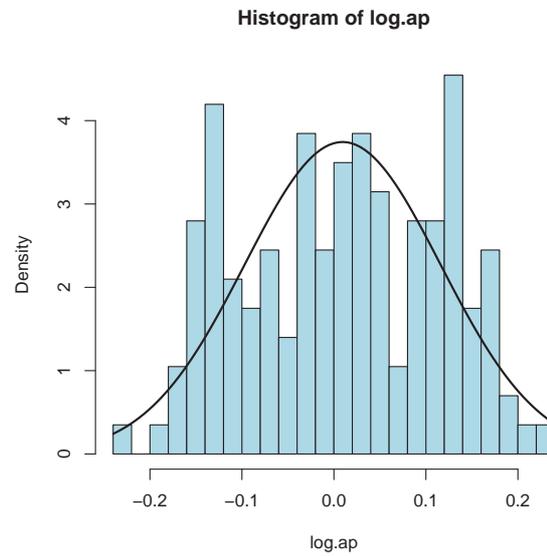


FIG. 2.4 – Histogramme des log-rendements.

```
>summary(log.rap)
  Min.   1st Qu.   Median     Mean   3rd Qu.   Max.
-0.22310 -0.08002  0.01482  0.00944  0.10590  0.22310
```

2.2.11 Estimation de la VaR

2.2.12 Modélisation par la méthode POT

On donne la représentation graphique de la GPD ;

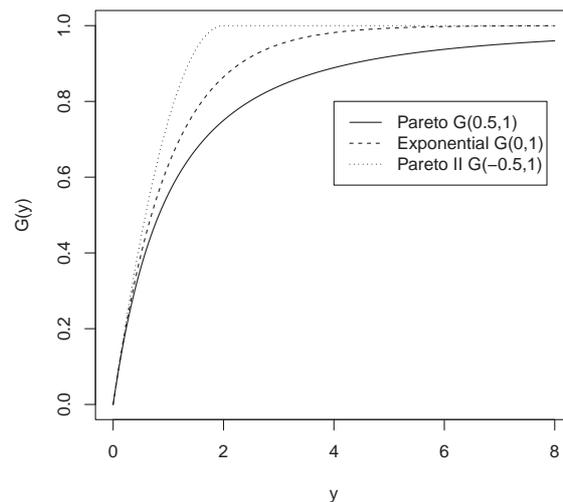


FIG. 2.5 – La distribution de Pareto généralisée $G_{\varepsilon,1}(x)$, pour Pareto ($\varepsilon = 0.5$), exponentielle ($\varepsilon = 0$) et Pareto de typeII ($\varepsilon = -0.5$).

Cette méthode consiste à sélectionner un échantillon des excès indépendants et à supposer que les valeurs des excès sont réparties selon une distribution généralisée de type Pareto(GPD).

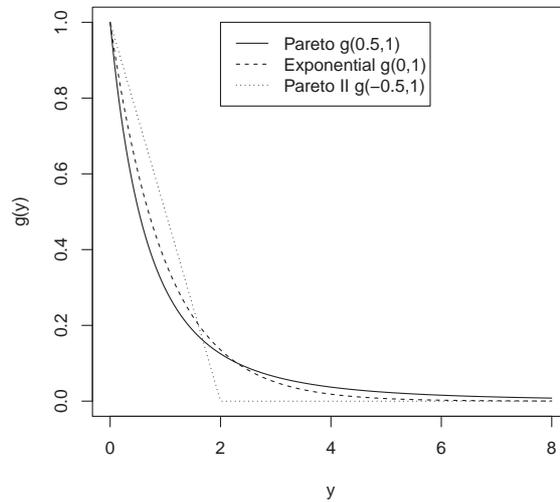


FIG. 2.6 – Densité de Pareto généralisée $g_{\varepsilon,1}(x)$, pour Pareto ($\varepsilon = 0.5$), exponentielle ($\varepsilon = 0$) et Pareto de typeII ($\varepsilon = -0.5$).

2.2.13 Sélection de seuil

Avant de pouvoir estimer la GPD, il nous faut trouver un seuil u au dessus duquel nous conservons assez de données pour des estimations précises. Pour cela, on fait appel à la courbe de la fonction moyenne des excès.

- Un des outils de choix du seuil est le graphe de la fonction moyenne des excès $e(u)$. On le trace à l'aide de R en utilisant la commande "meplot".

```
me.ap<-meplot(log.rap)
```

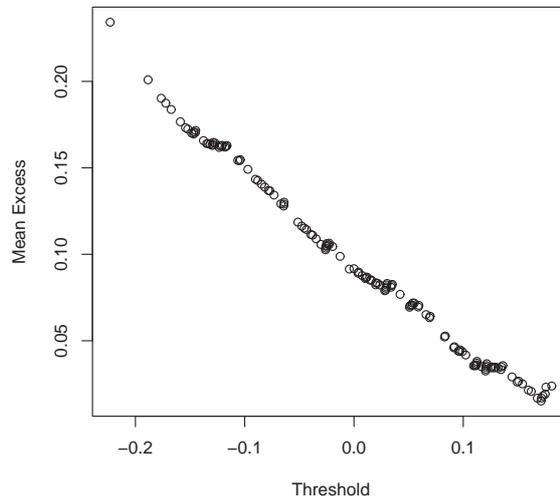


FIG. 2.7 – La distribution moyenne des excès.

le meplot semble avoir un comportement linéaire au-dessus d'une certaine valeur de u , cela signifie que les excès au-dessus de ce seuil suivent une GPD (comportement de type Pareto). Nous pouvons définir le seuil à -0.06 .

Après avoir obtenu le seuil u , nous estimons la GPD par la méthode de maximum de vraisemblance et nous obtenons le résultat suivant :

```
>t<-gpd(log.rap,-0.06,method="ml")
$n
[1] 143

$data
[1] 0.052185753 0.112117298 -0.022989518 0.109484233 0.091937495
.
.
.
[96] 0.023580943 0.125287761 0.150673346 -0.026060107 0.102278849
```

```
$threshold
```

```
[1] -0.06
```

```
$p.less.thresh
```

```
[1] 0.3006993
```

```
$n.exceed
```

```
[1] 100
```

```
$method
```

```
[1] "ml"
```

```
$par.ests
```

```
xi      beta  
-0.7438814 0.2123264
```

```
$par.ses
```

```
xi      beta  
1.026231e-02 1.999871e-06
```

```
$varcov
```

```
      [,1]      [,2]  
[1,] 1.053150e-04 -2.297986e-11  
[2,] -2.297986e-11 3.999485e-12
```

```
$information
```

```
[1] "observed"
```

```
$converged
```

```
[1] 0
```

```
$nllh.final
```

```
[1] -129.7186
```

```
attr(,"class")
```

```
[1] "gpd"
```

Pour obtenir les estimateurs :

```
>estim=t$par.ests[1]
```

```
>estim
xi
-0.7438814
```

```
>estim=t$par.ests[2]
>estim
beta
0.2123264
```

```
>riskmeasures(t,0.95)
      p      quantile      sfall
[1,] 0.95      0.1853216      0.2024307
```

```
VaR= 0.1853216
```

- Un autre outil pour le choix du seuil de la GPD est d'utiliser directement la commande "findthresh" qui se trouve dans le package "evir" qui consiste à trouver le seuil u en précisant la taille de l'échantillon des extrêmes par $N_u = 100$.

```
>findthresh(log.rap,100)
-0.06402186
```

Pour une taille de l'échantillon des excès $N_u = 100$, on trouve la valeur de u est approximativement égale à la valeur calculée à partir du graphe de la fonction moyenne des excès.

Conclusion et perspectives

Dans ce travail on s'est intéressé à l'étude des distributions à queues lourdes.

De nombreux phénomènes ne sont plus modélisés par des variables gaussiennes mais plutôt par des distributions à queues lourdes.

Ces distributions ont pris un essor considérable à cause des applications pratiques, comme c'est le cas en finance, l'actuariat, la météorologie, l'hydrologie, etc.

Dans ce travail, une étude exhaustive a été faite sur ce type de distributions et leurs propriétés ont été minutieusement détaillées. Nous avons donné quelques exemples pratiques de l'utilisation des distributions à queues lourdes en théorie de la ruine et dans la théorie des valeurs extrêmes.

Il serait intéressant d'élargir ces exemples aux files d'attente de type $G/G/1$ et à d'autres phénomènes.

Le cas multivarié serait aussi une perspective, du fait que dans la gestion des risques, ce cas multivarié est très répandu.

Bibliographie

- [1] Alfonsi (2008), Fonction de Répartition et Copules, Cours de Master Recherche : Probabilité et Application
(<http://cermics.enpc.fr/~alfonsi/mrf-copules-intro.pdf>).
- [2] Balkema, A., de haan, L. (1974). Residual life Time at Great Age. *Annals of Probability*, 2, 792, 804.
- [3] Bingham, N.H., Goldie, C.M., and Teugels, J.L., (1987. Regular variation, *Encyclopedia of Mathematics and its Applications* 27. Combridge University Press, Cambridge.
- [4] Carlo Acerbi Dirk Tasche, On the coherence of Expected Shortfall, *Journal Of Banking And Finance*, Volume 26, Pages 1487 – 1503.
- [5] Charles M. Goldie and Claudia Kluppelberg. Subexponential distributions. In Robert J. Adler, Raisa E. Feldman, and Murad S. Taqqu, editors, *A Practical Guide To Heavy Tails* (1998).
- [6] Chun Su, Qi-he Tang *Acta Mathematicae Applicatae Sinica, English Series*. Vol.19,No.1 (2003)135 – 137.
- [7] Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremeal Events for Insurance and Finance*, *journal stochastic Modelling and Applied Probability*. Springer, 1997.
- [8] Embrechts, P. et Veraverbeke, N. (1982). Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance : Mathematics and Economics*, 1(1) :55–72.
- [9] Gareth W.Peters, Pavel V.Shevchenko (2015), *Advances In Heavy Tailed Risk Modelling*. Wiley Handbooks in Finantial Enginnerig and Econometrics.

- [10] Gubinelli Massimiliano, Processus de Poisson et meth. actuariels M1 MMD poly4-v.420120514 <https://www.cermade.dauphine.fr/mgubi/e1112/pp-4.pdf>.
- [11] Imen Mathlouthi, Amel Zenaidi (2006), Theorie Des Valeurs Extremes, Vs, Methodes Classiques De Calcul De La VaR : Application au Tunindex, ECOFI, Institut des Hautes Etudes Commerciales, Carthage, 2016 Tunisie.
- [12] Korevaar, J., Tauberian theory, Grundlehren der Mathematischen Wissenschaften, 329, Springer-Verlag, Berlin (2004).
- [13] Latchman, F. Morgan, and W. Aspinall. Temporal changes in the cumulative piecewise gradient of a variant of the gutenbergrichter relationship, and the imminence of extreme events. *Earth-Science Reviews*, 87 : 94–112, 2008.
- [14] Malamud and D. Turcotte. The applicability of power-law frequency statistics to floods. *Journal of Hydrology*, 322 : 168–180, 2006.
- [15] Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36 : 394–419, 1963.
- [16] Mikosch, T. (1999), Regular Variation, Subexponentiality and Their Applications in Probability Theory, University of Groningen.
- [17] Pierre Clauss, (2008) Statistique Des Risques Extrêmes Application À La Mesure Du Risque De Marché, Ensaï Filière Gestion des Risques et Ingénierie Financière.
- [18] Rachev. Handbook of heavy-tailed distributions in finance. Elsevier, 2003.
- [19] Segey Foss, Dmitry Kurshonov et Stan Zachary, An introduction to heavy-tailed and subexponential distributions. Springer Series in Operations Research and Financial Engineering (2013).
- [20] Victor F. Araman, Theory and Applications of Stochastic Systems Lecture 11, November 21, 2003.