République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : **Statistique**

Par

Baida Hadjira

Titre :

Modélisation des distributions à queues lourdes

Membres du Comité d'Examen :

Pr.	**Brahimi Brahim**	UMKB	Président
Pr.	**Touba Sonia**	UMKB	Encadreur
Dr.	**Kheireddine Souraya**	UMKB	Examinatrice

Juin 2023

Dédicace

Je dédicie ce travail

À ma mère pour sa générosité, et son soutien permanent,

À mon père bien-aimé pour sa sacrifié,

À mes frères et mes soeurs.

À tous mes neveux surtout anas,moade et farah

À mes amis et mes collègues de la promotion 2023

«Mathématiques»

REMERCIEMENTS

Avant tous, je remercie **ALLaH** qui ma donné la force, le courage et la patience pour réaliser ce mémoire.

Je remercie aussi mon encadreur **Dr.Touba Sonia** pour la direction de ce mémoire et pour ses conseils précieux qui m'ont aide dans ma travail.

J'adresse mes remerciements aux les membres du Jury qui d'examiner ce mémoire en lui apportant de l'intérêt.

Je tiens à remercier tous les enseignants de mon parcours scolaire.

Table des matières

Table des figures

Liste des tableaux

Introduction

La modélisation des distributions à queues lourdes est un sujet important en statistique et en probabilités. Les queues lourdes sont des propriétés de certaines distributions statistiques caractérisées par des fréquences de données extrêmes qui dépassent les prévisions de la distribution standard de Gauss.

Ces distributions sont importantes car elles sont souvent associées à des phénomènes naturels tels que les catastrophes naturelles, les pannes techniques et les événements extrêmes dans les marchés financiers. La modélisation de ces distributions implique l'utilisation des techniques mathématiques avancées pour comprendre la distribution des données à l'extrême fin de la courbe, où la densité de la distribution est faible et les événements rares ont une probabilité significative d'occurrence. Ce domaine de recherche est crucial pour comprendre et prédire les événements rares et extrêmes qui ont des conséquences importantes dans plusieurs domaines tels que l'économie, la finance, l'assurance et la gestion de la chaîne d'approvisionnement.

Ce mémoire est composé en deux parties clés :

La première partie du mémoire est un rappel de la théorie des probabilités et des statistiques telle que la fonction de répartition, la fonction quantile ainsi que les fonctions empiriques de ces dernières avec une représentation générale de la loi normale leurs définitions, leurs propriétés, ...

La deuxième partie du mémoire est consacré à la description des distributions à queues lourdes. En premier, nous introduisons les propriétés statistiques communes à la plupart de ces distributions. Puis, faire rebdre ces faits stylisés par un modèl convenable justifier par des causes possibles. Enfin, nous complétons cette partie par une étude des données réelles, le cas des rendements d'un actif financier.

Chapitre 1

Généralités

1.1 Fonction de répartition

Définition 1.1.1 *Soit X une variable aléatoire (continue ou discrète). On définit la fonction de répartition F de X par :*

$$F : \mathbb{R} \to [0, 1] \ \textit{telle que } F_x(x) = P(X < x).$$

Si X est une variable aléatoire (v.a) continue ayant une loi de probabilité $f_X(x)$:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$$

– Si X est une v.a discrète ayant une loi de probabilité $P(X = k)$, la fonction de répartition est une fonction en escalier définie par :

$$F_X(x) = \sum_{k<x} P(X = k).$$

– Dans les deux cas, F est une fonction monotone croissante, c-à-d :

$$F_X(a) \leq F_X(b) \text{ si } a \leq b.$$

– Elle prend ses valeurs entre 0 et 1 : $0 \leq F_X(x) \leq 1$ avec

$$\lim_{x \to -\infty} F_X(x) = 0 \text{ et } \lim_{x \to +\infty} F_X(x) = 1.$$

1.2 Fonction de répartion et quantiles

Nous considérons un échantillon aléatoire $X_1, ..., X_n$ tiré d'une loi dont la fonction de répartition est inconnue F. Nous pouvons estimer F en utilisant la fonction de répartition empirique,

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

Pour une valeur fixée $x \in \mathbb{R}$, $F_n(x)$ représente le nombre d'observations qui sont inférieures ou égales à x. Nous allons maintenant étudier le comportement asymptotique de F_n. Selon la loi des grands nombres, pour un x fixé, $F_n(x)$ converge en probabilité vers $F(x)$ lorsque n tend vers l'infini

Le Théorème de Glivenko-Cantelli étend largement ce résultat en démontrant la convergence uniforme de $F_n$vers F pour $x \in \mathbb{R}$. Ce résultat est énoncé dans la suite.

Théorème 1.2.1 *Soit $(X_n)_{n \geq 1}$, une suite de variables aléatoires de même fonction de répartition F. Alors*

$$\sup_{x \in \mathbb{R}} \mid F_n(x) - F(x) \mid,$$

converge en probabilité vers zéro. Considérons maintenant le processus empirique

$$F_n(x) = \sqrt{n}\{F_n(x) - F(x)\}.$$

Une application du Théorème central limite assure que pour un $x \in \mathbb{R}$ fixé, $F_n(x)$ converge vers une v.a Normale de moyenne nulle et de variance

$$\sigma_x^2 = F(x)(1 - F(x)).$$

Le Théorème de Donsker va beaucoup plus loin en obtenant le comportement limite en loi de $F_n(x)$ en tant que fonction aléatoire définie pour tout $x \in \mathbb{R}$.

Définition 1.2.1 *Les quantiles d'ordre α de la fonction de distribution F est défini par*

$$q(\alpha) = F^{-1}(\alpha) = \inf\{X : F(y) > \alpha\},$$

avec $\alpha \in \left]0, 1\right[$, où F^{-1}, est l'inverse généralisée de F. Par convention $\inf \emptyset = \infty$. *Notons que l'inverse généralisé d'une fonction coïncide avec l'inverse classique lorsque la fonction est continue. Ainsi, un quantile sera dit extrême il on remplace son ordre α par une suite $\alpha_n \to 0$ quand $n \to \infty$. Concrètement, un quantile extrême d'ordre* $1 - \alpha_n$ *de la fonction de distribution F est définie par :*

$$q(\alpha_n) = \overline{F}^{-1}(\alpha_n) = \inf\{y : \overline{F}(y) \leq \alpha_n\} \text{ avec } \alpha_n \to 0 \text{ quand } n \to \infty,$$

où $\overline{F} = 1 - F$ est la fonction de survie de F

Il faut dire que le fait que l'ordre $\alpha_n \to 0$ quand $n \to \infty$ indique que l'information la plus importante pour estimer des quantiles extrêmes est contenue dans la queue de distribution.

Si la taille de l'échantillon tend vers l'infini, nous avons

$$
\begin{aligned}
P(X_{(n)} < q(\alpha_n)) &= P\left(X_i < q(\alpha_n), \forall i = 1, .., n\right) \\
&= P\left(\cap_{i=1}^{n} \{X_i < q(\alpha_n)\}\right) \\
&= \Pi_{i=1}^{n} P(X_i < q(\alpha_n)) \\
&= F^n(q(\alpha_n)) \\
&= (1-\alpha_n)^n \\
&= \exp(n \log(1-\alpha_n)) \\
&= \exp(-n\alpha_n(1+0(1))) \ \textit{quand} \ \alpha_n \to 0.
\end{aligned}
$$

Cette probabilité dépond donc du comportement asymptotique de $n\alpha_n$*.Afin d'estimer les quantiles extrêmes, il est primordial d'étudier la limite de cette séquence lorsque* n *tend vers l'infini. Selon la vitesse de convergence de* α_n *vers* 0*, trois cas distincts se présentent.*

– **Première situation :** Si $n\alpha_n \to \infty$ alors

$$P(X_{(n)} \leq q(\alpha_n)) \to 0.$$

Dans cette situation, il est fort probable que le quantile à estimer se trouve dans l'échantillon disponible, étant donné que α_n converge lentement vers 0. Cela signifie que d'avantage d'observations sont nécessaires pour estimer ce quantile avec précision. En conséquence, le quantile à estimer converge lentement vers l'infini au fur et à mesure que n augmente. Pour estimer le quantile, on peut simplement interpoler à l'intérieur de l'échantillon, ce qui donne l'estimateur $X_{(n-\lfloor n\alpha_n \rfloor+1)}$.

– **Deuxième situation :** Si $n\alpha_n \to c \in [1, \infty[$ alors

$$P(X_{(n)} < q(\alpha_n)) \to e^{-c}.$$

Dans cette deuxième situation, il est peu probable que la valeur maximale de l'échantillon dépasse le quantile. Ainsi, l'estimation du quantile extrême doit se baser sur les grandes observations proches de la limite de l'échantillon, mais toujours incluses dans l'ensemble des données. Par conséquent, l'estimateur proposé dans le premier cas est le plus approprié.

– **Troisième situation :**Si $n\alpha_n \to c \in [0, 1[$ alors

$$P(X_n < q(\alpha_n)) \to e^{-c}.$$

Dans ce cas, il n'est pas possible de proposer un estimateur du quantile car le quantile à estimer est supérieur au maximum des observations disponibles avec la fonction de répartition empirique définie :

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leq x\}} = \begin{cases} 0 \text{ si } x < X_1 \\ \frac{i-1}{n} \text{ si } X_{(i-1)} \leq x \leq X_{(i)}, 2 \leq i \leq n \\ 1 \text{ si } x \geq X_{(n)} \end{cases}$$

Toute fois, pour estimer le quantile extrême, il est nécessaire d'extrapoler au-delà de la plus grande observation disponible car la fonction de répartition empirique, $F_n(y) = 1$ si y est supérieur ou égal à la plus grande observation de l'échantillon $X_{(n)}$. Cette extrapolation peut être représentée par le schéma suivant :

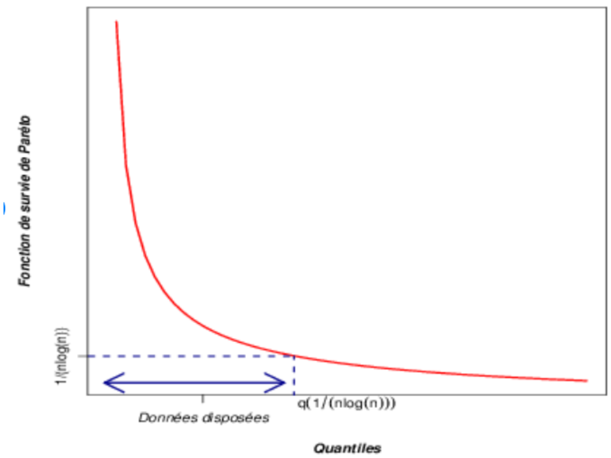


FIG. 1.1 – Explication graphique de quantiles extrêmes et de queue de distribution

La figure ci-dessus montre que le quantile d'ordre $1/(n\log(n))$, tend à 0 lorsque n tend vers l'infini et qu'il se situe au-delà de la valeur maximale des observations disponibles. Ainsi, pour le déterminer, il est nécessaire d'avoir une connaissance de la forme de la distribution en examinant la loi du maximum. Remarquons d'abord que

$$\lim_{n\to\infty} F_{X_{(n)}}(y) = \lim_{n\to\infty} [F(y)]^n I_{\{y\geq y_F\}} = \begin{cases} 1 \text{ si } y \geq y_f \\ 0 \text{ si } y < y_f \end{cases} \tag{1.1}$$

$$y_F = \sup\{y \in \mathbb{R}, F(y) < 1\}$$

est le point terminal de la loi F. l'équation 1.1 montre que la loi du maximum, $X_{(n)}$est dégénérée, donc fournit peu d'information.

Définition 1.2.2 (Fonction des quantiles empirique) *La fonction des quantiles ou l'inverse généralisé de la fonction de distribution F notée par Q, pour chaque entier $n \geq 1$*

$$Q(p) = F^{-1}(p) = \inf\{x : F(x) \geq p\}, 0 < p < 1.$$

La fonction des quantiles empiriques notée Q_n est définie par :

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

Où F^{-1} est l'inverse généralisée de la fonction de distribution F.

Définition 1.2.3 (Fonction des quantiles de queues) *La fonction des quantiles de queues (taille quantile fonction), notée U est définie par :*

$$U(p) = Q(1 - 1/p) = (1/\overline{F})^{-1}(p), 1 < p < \infty.$$

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

$$U_n(p) = Q_n(1 - 1/p), 1 < p < \infty.$$

1.3 Estimation empirique

L'estimation quantile joue un rôle important dans le contexte de la gestion des risques, où il est crucial d'évaluer correctement le risque d'une grosse perte qui survient très rarement. La principale difficulté de cette estimation est due au fait que lorsque p est très petit, le point x_p est au-delà de la plage de l'échantillon $(X_1, ..., X_n)$ tiré d'un cdf F inconnu.

Comme nous utilisons la théorie asymptotique, p doit dépendre de la taille de l'échantillon n, c.à-d. $p = p_n$.

Deux cas sont possibles pour x, à l'intérieur et à l'extérieur de l'échantillon.

– si $p_n \to 0$ avec $np_n \to c \in [1, \infty]$ comme $n \to \infty$, le $(1 - p)$-quantile est dans l'échantillon.

– si $p_n \to 0$ avec $np_n \to c[0, 1]$ comme $n \to \infty$, le $(1 - p)$-quantile est en dehors de l'échantillon.

En d'autres termes, l'estimation intra-échantillon est possible jusqu'au quantième $(1/n)$ alors que, pour $p < 1/n$, les estimations quantiles sont au-delà de la plage des données. Ce dernier cas est le plus pertinent pour les applications réelles. Pour la première situation, nous avons $Q_n(s) = X_{n-i+1,n}$, puis avec $s = 1 - p = 1 - (i-1)/n$ pour $i = 2, ..., n$, on a

$$Q_n\left(1 - \frac{(i-1)}{n}\right) = X_{n-i+1,n}, i = 2, ...n.$$

Ainsi, $X_{n-i+1,n}$ semble être un estimateur naturel pour le $(1 - \frac{(i-1)}{n})$-quantile. Dans le second cas, nous devons déduire au-delà des limites de l'échantillon en extrapolant à partir des quantiles intermédiaires. Évidemment, cela ne peut pas être fait sans une sorte d'information sur les queues est alors nécessaire.

1.4 Loi normale

On dit qu'une variable aléatoire continue X suit une loi normale de paramètres μ et σ^2 si sa fonction de densité est :

$$f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \text{pour tout } x.$$

On dénote ceci par :

$$X \rightsquigarrow N\left(\mu, \sigma^2\right).$$

1.4.1 Propriétés de la loi normale

Propriétés de f_x :

1. $\lim\limits_{x\to\infty} f_x(x) = 0.$

2. $f_x(\mu + x) = f_x(\mu - x)$ (symétrie par rapport à l'axe $x = \mu$).

3. f_x atteint son maximum en $x = \mu$ (μ est le mode de x).

4. Les points d'inflexion du graphe de f_x sont $x = \mu \neq \sigma$.

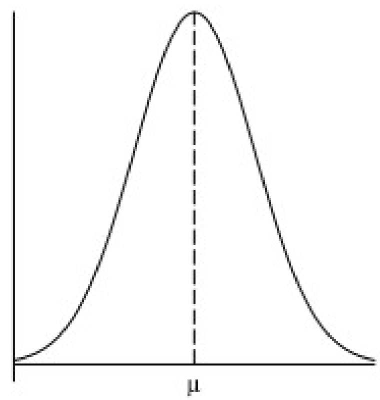


FIG. 1.2 – Fonction de densité de la loi normale

Si $X \rightsquigarrow N(\mu, \sigma^2)$ alors :

1. $P(X < \mu - x) = P(X > \mu + x)$.

2. $F_x(\mu - x) = 1 - F_x(\mu + x)$.

1.4.2 Moyenne et variance de la loi normale

Si $X \rightsquigarrow N(\mu, \sigma^2)$ alors :

1. $E(X) = \mu$

2. $V(X) = \sigma^2$

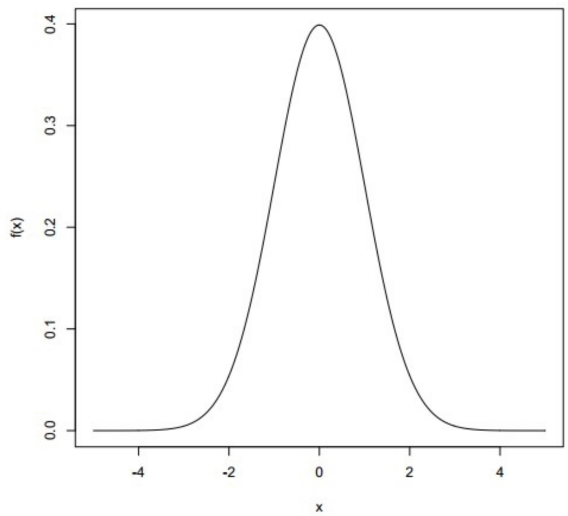


FIG. 1.3 – Fonction de densité de $X \rightsquigarrow N(0,1)$

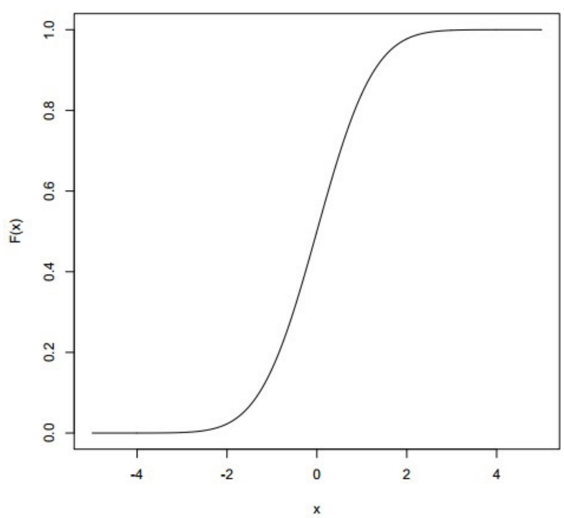


FIG. 1.4 – Fonction de répartition de $X \rightsquigarrow N(0,1)$

1.5 Théorème central limite(TCL)

Soit $X_1, X_2, ..., X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d) de moyenne, $E[X_i] = \mu_i$ et de variance $var(X_i) = \sigma^2$, pour $i = 1, 2, ..., n$.

$$Z = \frac{\sum_{i=1}^{n}(X_i - n\mu_i)}{\sqrt{\sum_{i=1}^{n} n\sigma^2}} \implies N(0,1)$$

suit approximativement une loi normale $N(0, 1)$ si n est grand.

Ou de manière équivalente

$$\frac{\sqrt{n}(\overline{X} - \mu_i)}{\sigma} \implies N(0,1)$$

avec $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

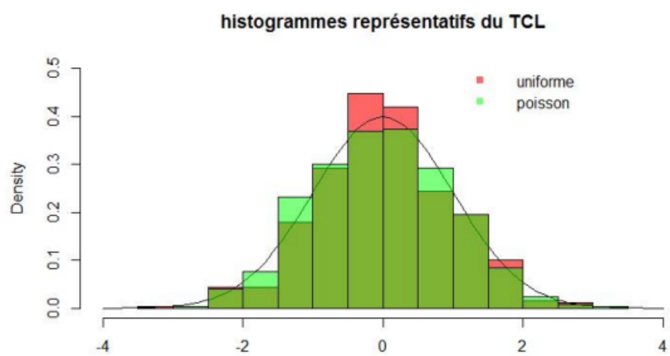


FIG. 1.5 – Histogrammes repésentatifs du TCL pour la loi uniforme et la loi de poisson

Sur la figure ci-dessus, on a illusté le TCL pour $n = 10000$ en représentant pour les deux lois l'histogramme d'un échantillon simulé suivant la loi de $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}}$.On observe ainsi la convergence vers loi normale.

Le logiciel R choisit le nombre de la classe de manière à optimiser la correspondance avec le TCL mais on constate que pour un nombre de classe fixe, on peut avoir parfois un histogramme peu ressemblant à la loi $N(0,1)$.

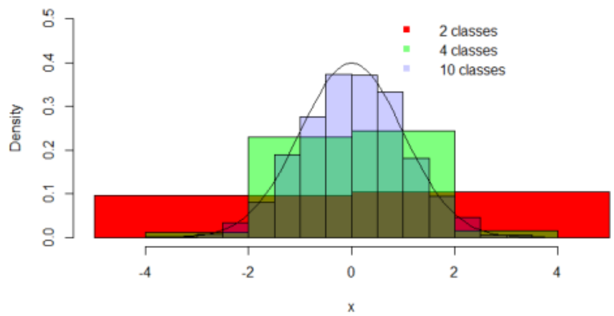


FIG. 1.6 – Histogramme de la loi N(0,1) avec différentes classe

1.5.1 Réciproque du théoreme central limite

Soit $(X_i)_{i\geq 1}$une suite de variables aléatoires i.i.d définies sur le même espace probabilisé et telle que $\frac{S_n}{\sqrt{n}}$ converge en loi vers $N(0,1)$.Alors X_1 est de carré intégrable, $E[X_1^2] = 1$ et $E[X] = 0$.

1.5.2 Utulisation du TCL

Le TCL est utilisé pour approximer des probabilités impliquant des sommes de variables aléatoires.

On a

$$E\left(\sum_{j=1}^{n} X_j\right) = n\mu, var\left(\sum_{j=1}^{n} X_j\right) = n\sigma^2,$$

donc

$$\frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n\sigma^2}} = \frac{n(\overline{X}-\mu)}{\sqrt{n\sigma^2}} = \frac{n^{1/2}(\overline{X}-\mu)}{\sigma} = Z_n$$

peut-être approximé par une variable normale :

$$P\left(\sum_{j=1}^{n} X_j \leq x\right) = P\left\{\frac{\sum_{j=1}^{n} X_j - n\mu}{\sqrt{n\sigma^2}} \leq \frac{x-n\mu}{(n\sigma^2)^{1/2}}\right\} = \Phi\left\{\frac{x-n\mu}{(n\sigma^2)^{1/2}}\right\}.$$

1.5.3 Lois des Grands Nombres

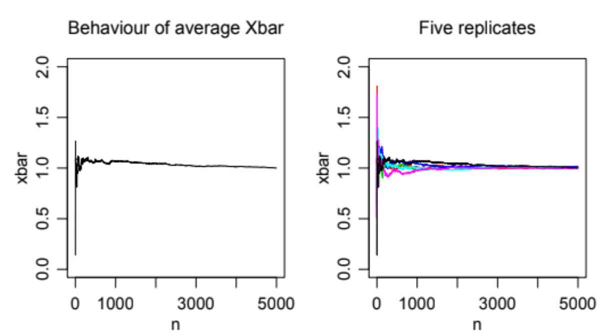
Théorème 1.5.1 (Loi faible des Grands Nombre) *Soit* $X_1, X_2, \ldots.., X_n$ *une suite de variables aléatoires indépendantes et identiquement distribuées, d'espérance finie* μ. *Notons leur moyenne par*

$$\overline{X} = n^{-1}(X_1 + \ldots + X_n).$$

Alors $\overline{X} \xrightarrow{p} \mu$;*c'est à dire,pour tout* $\varepsilon > 0$

$$P\left(\left|\overline{X} - \mu\right| > \varepsilon\right) \to 0 \textit{ quand } n \to \infty.$$

Ainsi, sous de légères conditions, les moyennes d'échantillons de taille importante convergente vers l'espérance de la distribution dont l'échantillon est issu .Le graphe ci-dessous illustre cela.



et effet est également valable pour de nombreuses autres statistique qui peuvent être exprimées sous forme moyenne, telle que les quantiles d'un échantillon : Soient $X_1, \ldots., X_n \overset{iid}{\sim} F$, où F est une fonction de répartition continue, et soit $x_p = F^{-1}(p)$ le p quantile de F. En notant que :

$$X_{(\lceil np \rceil)} \leq x_p \Leftrightarrow \sum_{j=1}^{n} I_{(X_j \leq x_p)} \geq \lceil np \rceil$$

et en appliquant la LfGN à la somme à droite, on a $X_{(\lceil np \rceil)} \xrightarrow{p} x_p$.

la variance

Exemple 1.5.1 *Soient* $X_1, ..., X_n \overset{iid}{\sim} (\mu, \sigma^2)$*,où* μ, σ^2 *sont finies. Monte que* $E(S^2) = \sigma^2$*,et indiquer pourquoi lorsque* $n \to \infty$

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{X})^2 \longrightarrow^p \sigma^2.$$

Théorème 1.5.2 (Loi Forte des Grands Nombres) *Sous les conditions du théorème précédent,* $\overline{X} \overset{p.s}{\to} \mu$*,c.à.d que,*

$$P(\lim_{n\to\infty} \overline{X} = \mu) = 1.$$

La loi faible stipule que pour tout $\varepsilon > 0$*, l'événement* $\mid X$ - $\mu \mid > \varepsilon$ *peut se produire un nombre infini de fois, avec toutefois des probabilités de moins en moins petites. En revanche, la loi forte exclut cette possibilité et impose que l'événement* $\mid X - \mu \mid > \varepsilon$ *ne se produit que finalement et avec une probabilité de* 0*.*
Il est à noter que ces deux lois peuvent également s'appliquer dans des cas de dépendance entre les variables aléatoires X_j*, sous certaines conditions.*

1.6 Statistique d'ordre

Définition 1.6.1 *La statistique d'ordre de l'échantillon* $(X_1, ..., X_n)$ *est le réarrangement croissant de* $(X_1, ..., X_n)$ *On le note par* $(X_{1,n}, ..., X_{n,n})$*.*
On a $X_{1,n} \leq ..... \leq X_{n,n}$*, et il existe une permutation aléatoire* $\sigma_n \in S_n$ *telle que*

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

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

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

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

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

Alors

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

Puis

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

D'où on déduit :

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

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

La fonction de répartition de X n'étant pas souvent connue, il n'est généralement pas possible de déterminer la distribution du maximum à partir de ce résultat.

On s'intéresse alors à la distribution asymptotique du maximum en faisant tendre n vers l'infini. On a :

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

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

Proposition 1.6.1 (Limite de $X_{n,n}$)

$$X_{n,n} \xrightarrow{p.s} x_F \ quand \ n \longrightarrow \infty.$$

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

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

Remarque 1.6.1 *x_F est le point extrême supérieur de la distribution F telle que :*

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

1.7 Lois dégénérées

1.7.1 Loi de Pareto

La loi de Pareto s'applique pour les distributions tronquées. Prenons un exemple de la vie courante, en France, la borne basse du salaire horaire est forcément le SMIG, il ne peut pas en être autrement. La loi de Pareto permet de tenir compte de cette contrainte en restreignant le domaine de définition de la v.a. X.

Définition 1.7.1 *La loi possède deux paramètres, $\alpha > 0$ et c qui introduit la contrainte $x > c$.*

Le domaine de définition de X est $]c, +\infty[$. La fonction de densité est monotone décroissante, elle s'écrit

$$f(x) = \frac{\alpha}{c}\left(\frac{c}{x}\right)^{\alpha+1}, x > c.$$

La fonction de répartition est directement obtenue avec

$$F(x) = 1 - \left(\frac{c}{x}\right)^{\alpha}$$

Caractéristiques de la loi

$$E[X] = \frac{\alpha}{\alpha - 1},\ pour\ \alpha > 1$$

$$E[X^k] = \frac{\alpha}{\alpha - c},\ pour\ \alpha > c$$

$$var(X) = \frac{\alpha}{(\alpha-1)^2(\alpha-2)}c^2\ pour\ \alpha > 2.$$

1.7.2 Loi de Burr

En théorie des probabilités, en statistique et en économétrie, la loi de Burr, loi de Burr de type XII, loi de Singh-Maddala, ou encore loi log-logistique généralisée est une loi de probabilité continue dépendant de deux paramètres réels positifs c et k. Elle est communément utilisée pour étudier les revenus des ménages

Si X suit une loi de Burr (ou Sigh-Maddala), on notera $X \rightsquigarrow SM(c,k)$.

Caractirisation

la densité de probabilité de la loi de Burr est donnée par :

$$f(x,c,k) = \begin{cases} ck\frac{x^{c-1}}{(1+x^c)^{k+1}} \text{ si } x > 0 \\ 0 \text{ sinon} \end{cases}$$

et sa fonction de répartition est :

$$F(x,c,k) = \begin{cases} 1-(1+x^c)^{-k} \text{ si } x > 0 \\ 0 \text{ sinon} \end{cases}$$

Si $c = 1$, la loi de Burr est la distribution de Pareto.

1.7.3 Loi de fréchet

La densité de probabilité de la distribution de paramètre $\alpha > 0$, peut être généralisée en introduisant un paramètre de position m du minimum et un paramètre d'échelle $s > 0$ prend la forme

$$f(x) = \begin{cases} \frac{\alpha}{s}\left(\frac{x-m}{s}\right)^{-1-\alpha} e^{-\left(\frac{x-m}{s}\right)^{-\alpha}} \text{ si } x > m \\ 0 \text{ sinon} \end{cases}$$

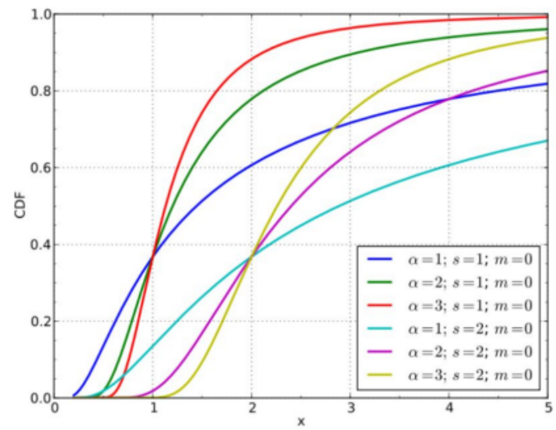


FIG. 1.7 – Fonction de densité de la loi de fréchet

et sa fonction de répartition est donnée par :

$$F_X(x) = P(X \leq x) = \begin{cases} e^{-x^{-\alpha}} \text{ si } x > 0 \\ 0 \text{ sinon} \end{cases}$$

où $\alpha > 0$ est un paramètre de forme. Cette loi peut être généralisée en introduisant un paramètre de position m du minimum et un paramètre d'échelle $s > 0$. La fonction de répartition est alors :

$$P(X \leq x) = \begin{cases} e^{-(\frac{x-m}{s})^{-\alpha}} \text{ si } x > m \\ 0 \text{ sinon} \end{cases} .$$

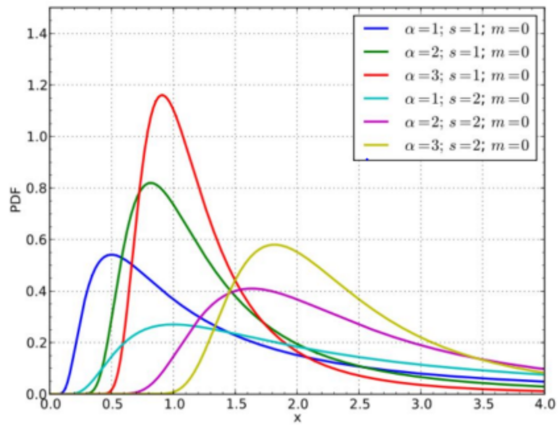


FIG. 1.8 – Fonction de répartion de la loi de fréchet.

La loi de Fréchet à un paramètre α à des moments standards :

$$\mu_k = \int_0^\infty x^k f(x)dx = \int_0^\infty t^{-\frac{k}{\alpha}} e^{-t} dt,$$

(avec $t = x^{-\alpha}$) définis pour $k < \alpha$:

$$\mu_k = \Gamma\left(1 - \frac{k}{\alpha}\right),$$

où $\Gamma(z)$ est la fonction Gamma.

En particulier :

– Pour $\alpha > 1$ l'espirance est :

$$E[X] = \Gamma\left(1 - \frac{1}{\alpha}\right).$$

– Pour $\alpha > 2$ la variance est :

$$var(X) = \Gamma\left(1 - \frac{2}{\alpha}\right) - \left(\Gamma\left(1 - \frac{1}{\alpha}\right)\right)^2.$$

Chapitre 2

Distributions à queues lourdes

Les distributions à queues lourdes jouent un rôle important dans la théorie des valeurs extrêmes. Elles ont été acceptées comme des modèles appropriés de divers phénomènes on peut citer dans le cas de montage de grand sinistre en assurance les fluctuations des prix en finance,... etc. Mathématiquement, ce type des distributions est défini ainsi :

2.1 Caractérisation des 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 exponentielle.

Soit X une variable aléatoire de fonction de répartition :

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

Noton

$$\overline{F}(x) = 1 - F(x)$$

$\overline{F}(x)$ est dite fonction des queue.

2.1.1 Definitions

Définition 2.1.1 *Une distribution F sur $\mathbb{R}$ est dite à support non borné si :*

$$\overline{F}(x) > 0, \forall x \in \mathbb{R}.$$

Définition 2.1.2 *Soit X une variable aléatoire de fonction de repartition à queue lourde, s'il existe une constante positive γ qui répartition l'indice de queue et prend la formule suivante :*

$$\overline{F}(x) \sim x^{-1/\gamma} I(x) \ ,pour\ x \to \infty.$$

où $l(x)$ la fonction à variation lente au voisinage de l'infini.

Définition 2.1.3 *Une distribution F (ou X) est dite à queue lourde si et seulement si :*

$$\int_{\mathbb{R}} e^{\lambda x} F(dx) = \infty \ ,\forall \lambda > 0. \tag{2.1}$$

(ce qui signifie que la fonction génératrice n'existe pas).

D'une manière équivalente

$$\lim_{x\to\infty} P[X > x + y \mid X > x] = \lim_{x\to\infty} \frac{\overline{F}(x+y)}{\overline{F}(x)} = 1,\ \forall x,\ \ y \geq 0,\ \ \overline{F}(x) > 0.$$

Cette dernière peut s'écrire comme suit :

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

2.2 Propriétés des distributions à queues lourdes

Proposition 2.2.1 *Soient $X \in L$ et $Y \geq 0$ une v.a indépendante de X, alors*

$$P[X - Y > x] \sim P[X > x]$$

Preuve.

$$\begin{aligned} P[X - Y > x] &= P[X - Y > x \mid X > x]\overline{F}(x) \\ &= \overline{F}(x) \int_0^\infty P[X > x + y \mid X > x] f_y(y) dy \\ &\to \overline{F}(x), \text{puisque } P[X > x + y \mid X > x] \to 1, x \to \infty. \end{aligned}$$

■

Lemme 2.2.1 *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 :

$$\lim_{n\to\infty} \sup p_n e^{\varepsilon n} = \infty, \, \forall \varepsilon > 0.$$

Proposition 2.2.2

$$X \in L \Longleftrightarrow \lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1, \forall y \geq 0. \tag{2.2}$$

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

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

par conséquent

$$\lim_{x\to\infty}\inf \frac{\overline{F}(x-y)}{\overline{F}(x)} \geq 1.$$

Alors, (2.2) *peut réécrire comme suit :*

$$X \in L \Leftrightarrow \lim_{x\to\infty}\sup \frac{\overline{F}(x-y)}{\overline{F}(x)} \leq 1,\ \forall y \geq 0.$$

2.3 Classe des distributions à queues loures sans comparaison à loi normale

Définition 2.3.1 *Une distribution F est dite à queue légère si et seulement si*

$$\int_{\mathbb{R}} e^{\lambda x} dF(x) < \infty,\ \textit{pour tout}\ \lambda > 0. \tag{2.3}$$

i.e :si et seulement si elle n'est pas à queue lourde.

Il est clair que pour toute distribution F à queue légère sur demi droite positive $\mathbb{R}^+ = [0, +\infty[$, tous les moments d'ordre k sont finis.

$$\int_{\mathbb{R}} x^K dF(x) < \infty,\ \textit{pour tout}\ K > 0.$$

On dira qu'une fonction non négative (tendant généralement vers zéro) est à queue lourde si elle n'est pas dominée par une fonction exponentielle décroissante. Plus précisément nous énonçons la définition suivante.

Définition 2.3.2 *On dira une fonction $f \geq 0$ qu'elle est à queue lourde si et seulement si*

$$\lim_{x\to\infty}\sup f(x)e^{\lambda x} = \infty\ \textit{,pour tout}\ \lambda > 0.$$

Définition 2.3.3 *Une distribution F est dite à queue lourde (à droite) si et seule-*

ment si

$$\int_{\mathbb{R}} e^{\lambda x} dF(x) = \infty \text{ , pour tout } \lambda > 0.$$

Théorème 2.3.1 *Pour toute distribution F, les assertions suivantes sont équivalentes :*

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

   Il est clair que la propriété de queue lourde d'une fonction revient à la forme de sa queue. Le théorème ci-dessus montre en particulier qu'une distribution est à queue lourde si et seulement si sa fonction de queue est une fonction à queue lourde. Nous introduisons la définition suivante l'extension de ce dernier théorème.

Définition 2.3.4 *Pour toute fonction F, la fonction*

$$R(x) = -\ln \overline{F}(x)$$

est dite la fonction hasard de la distribution .Si la fonction hasard (fonction de risque et danger) est différenciable alors sa dérivée

$$r(x) = R^{'}(x)$$

est dite taux de hasard (taux de risque).

Théorème 2.3.2 *Pour tout distribution F, les assertions suivantes sont équivalente :*

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

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

4. *Pour certains(tout)T fixe avec $T > 0$, la fonction $F(x, x+T]$ est à queue lourde.*

Preuve. 1)⇒4) Supposons que la fonction $F(x, x+T]$ n'est pas à queue lourde. Alors

$$c = \sup_{x\in\mathbb{R}} F(x, x+T]e^{\lambda' x} < \infty \text{ ,pour certains } \lambda' > 0,$$

et donc pour tous $\lambda < \lambda'$

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

Il s'ensuit que l'intégrale définie en (2.1) qu'il est finie pour tout $\lambda \in (0, \lambda')$,ce qui implique que la distribution F ne peut pas être à queue lourde d'où la contradiction, ce qui fait que l'implication est vraie.

4)⇒2).Cette implication découle de l'inégalité

$$\overline{F}(x) \geq F(x, x+T].$$

2)⇒3).Supposons qu'au contraire, la limite inférieure $\liminf$ dans 3) est strictement positive.

Alors il existe $x_0 > 0$ et $\varepsilon > 0$ telle que

$$R(x) \geq \varepsilon x$$

pour tous $x \geq x_0$ ce qui implique que

$$\overline{F} \leq e^{-\varepsilon x}$$

qui est contradiction avec 2).

3)⇒1). Supposons qu'au contraire, la fonction F est à queue légère. Il découle alors de (2.3) en utilisant par exemple l'inégalité de Tchebychev que pour certains $\lambda > 0$ et $c > 0$, nous avons

$$\overline{F}(x) \leq ce^{-\lambda x}$$

pour tous x. Ce qui implique que

$$\liminf_{x \to \infty} R(x)/x \geq \lambda$$

qui conterdit 3) ■

Lemme 2.3.1 *Soit F une distribution absolument continue de densité f .Supposons que la distribution F est à queue lourde . Alors la fonction f est aussi à queue lourde.*

Preuve. Supposons que f n'est pas à queue lourde, alors il existe $\lambda' > 0$. et x_0 telle que

$$c = \sup_{x > x_0} f(x) e^{\lambda' x} < \infty,$$

et donc pour tous $\lambda \in (0, \lambda')$

$$\int_{\mathbb{R}} e^{\lambda x} dF(x) \leq e^{\lambda x_0} + c \int_{x_0}^{\infty} e^{(\lambda - \lambda')x} dx < \infty.$$

Il s'ensuit que l'intégrale définie dans (2.1) est finie pour tout λ tel que $0 < \lambda < \lambda'$, ce qui contredit le fait que F est à queue lourde. ■

Remarque 2.3.1 *La réciproque est fausse, en général. Pour se convaincre, consi-*

dérons la fonction de densité continue par morceaux

$$f(x) = \sum_{n=1}^{\infty} I_{[n,n+2^{-n}]}(x).$$

nous avons d'une part

$$\limsup_{x\to\infty} f(x)e^{\lambda x} = \infty, \text{ pour tout } \lambda > 0,$$

ainsi la fonction f est à queue lourde. D'une autre part, pour tout $\lambda \in (0, \ln 2)$,

$$\int_0^{\infty} e^{\lambda x} f(x)dx < \sum_{n=1}^{\infty} e^{\lambda(n+2^{-n})} 2^{-n} = \sum_{n=1}^{\infty} e^{\lambda(n+2^{-n})-n\ln 2} < \infty,$$

donc F est à queue légère.

2.4 Classe des distribution à queue lourde avec comparaision à la loi normale

Définition 2.4.1 *Étant donnée une v.a réelle X d'espérance μ et d'écart type σ, on définit son coefficient d'aplatissement ou Kurtosis non normalisé $\beta_2$comme le moment d'ordre quatre de la variable centrée réduite :*

$$\beta_2 = E[(\frac{X-\mu}{\sigma})^4],$$

lorsque cette espérence existe.On a donc :

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{(\mu_2 - \mu_1^2)^2}$$

avec μ_i le moment d'ordre i.

Définition 2.4.2 *Une distribution est à queue lourde si et seulement si son coefficient d'aplatissement est supérieur à celui de la loi Normale (pour la quelle $\beta_2 = 3$).*

Remarque 2.4.1 *Cette caractérisation générale ne peut être appliquée que si le moment d'ordre 4 existe ce qui empêche toute discrimination entre les distributions qui ont un moment d'ordre 4 infini.*
On peut faire un classement selon la queue droite de certaines distributions comme suit, mais malheureusement en ne peut pas les classer toutes en se basant sur ce critère. Le classement suivant :

1. Les distributions avec des moments exponentielles inexistants(**E**).
2. Les distribution Sub exponentielles (**D**).
3. Les distribution à variation régulières(**C**).
4. Les distributions avec un comportement de Pareto(**B**).
5. Les distribution $\alpha - stable$ avec $\alpha < 2$(**A**).

   Ce sont classement emboîtées et on a :

$$\mathbf{A} \subset \mathbf{B} \subset \mathbf{C} \subset \mathbf{D} \subset \mathbf{E}.$$

La classe **B** et la classe **C** sont très importants du leur lien à la théorie des valeurs extrêmes

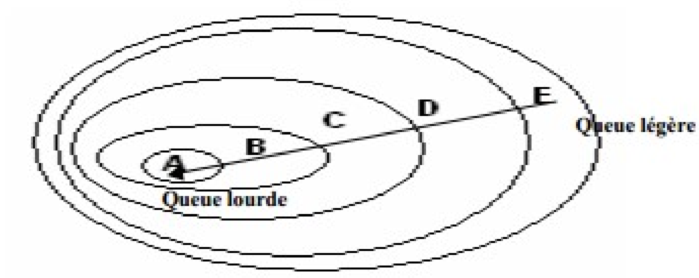


FIG. 2.1 – Classification des lois selon les queues

- La classe (**E**) contient toutes les lois tel que $E[e^X] = \infty$. La probabilité au dépassement, $\overline{F} = P(X \geq x) = 1 - F(x)$, pour les extrêmes de cette classe, décroît moins rapidement que celles des lois ayant une queue plus lourde que la lois normale.
- La classe (**D**) quant à elle, contient les lois tel que $\overline{F}$ décroît plus lentement que n'importe quelle loi exponentielle.
- Pour la classe (**C**) dite classe des lois à variation régulière, la probabilité au dépassement des extrêmes décroît vers une fonction puissance(appelée aussi décroissance géométrique).
- La classe (**B**) est celle des lois de Pareto.
- Enfin, la classe(**A**) est celle qui regroupe des lois à différentes asymétries, avec des queues très lourdes.

2.4.1 Distribution avec des moments exponentiels inexistant

La classe **E** contient toutes les distributions qui satisfont $E[e^x] = \infty$. Notez que la loi normale n'entre pas dans cette catégorie, car la probabilité de dépassement $\overline{F}$ dans ces cas extrêmes diminue plus rapidement que la loi normale. En ce sens, la classe **E** est une distribution avec des queues plus lourdes que la loi normale.

2.4.2 Distribution Subexpontielle

La classe **D** des distributions subexponentielles est caractérisée par la définition suivante (2.3) : On dit qu'une distribution est subexponentielle si :

$$\lim_{x\to\infty} \frac{p(X_1 + X_2 + ... + X_n)}{p(\max(X_1 + X_2 + ... + X_n) > x)} = 1 \tag{2.4}$$

Cela revient à dire que la somme de n distributions subexponentielles (indépendantes et iid) est extrême si et seulement si leurs maxima sont extrêmes. L'équation (2.4) peut être exprimée comme :

$$\lim_{x\to\infty} \frac{\overline{F}(x)}{e^{-\epsilon x}} = \infty, \forall \varepsilon > 0. \tag{2.5a}$$

Rappelons que e^{-tx} est la forme de la queue de la loi exponentielle. Comme son nom l'indique, la classe **D** contient les distributions telles que $\overline{F}$ décroît plus lentement que n'importe quelle loi exponentielle.

2.4.3 Distribution à variations réguliére d'indice $\alpha > 0$

La classe **C**, dont les distributions varient régulièrement, est une sous-classe des distributions sous-exponentielles (Goldie et Klüppelberg 1998). Ces distributions satisfont aux conditions suivantes (Embrechts, Klüppelberg et Mikosch 1997 ([13])).

$$\lim_{x\to\infty} \frac{\overline{F}(tx)}{\overline{F}(x)} = x^{-\alpha}, \tag{2.6}$$

Cela revient à dire que pour des valeurs extrêmes (t tend vers l'infini) la distribution se comporte de la même manière que la loi de Pareto. Par conséquent, la probabilité de dépasser la valeur extrême diminue selon une fonction puissance (également appelée descente géométrique). Le paramètre est appelé « exposant de queue » et peut être utilisé comme critère pour classer la distribution du comportement de la fonction de probabilité de dépassement F aux valeurs extrêmes.

2.4.4 Comportement de Pareto avec $\alpha > 0$

La classe **B** est une distribution de qui ont un comportement de Pareto dans la répartition est donnée par

$$F(x) = 1 - \left(\frac{u}{x}\right)^{\alpha}, x \geq u \text{ et } u > 0.$$

D'où

$$\overline{F}(x) = \left(\frac{u}{x}\right)^{\alpha}, x \geq u \text{ et } u > 0.$$

Pour ce type de distribution, les moments d'ordre k ne sont finis que si $k < \alpha$ car

$$\begin{aligned} E[X^k] &= \int_{\mathbb{R}} x^k dF(x) = \int_{\mathbb{R}} x^k d\left(1 - \left(\frac{u}{x}\right)^{\alpha} I_{[u,\infty[}(x)\right) \\ &= \alpha u^{\alpha} \int_u^{\infty} x^{k-\alpha-1} dx. \end{aligned}$$

2.4.5 Distributions $\alpha-$Stable

Les propriétés précédentes sont importantes pour définir la classe **A**, qui est la distribution $\alpha-$ stable (également appelée distribution stable). Une distribution stable est très des lois de probabilité riches peuvent représenter différentes asymétries et très Lourd. Lévy (1925)[25] dans son Variables indépendantes et identiquement distribuées (i.i.d). Manque de formule claire La densité de ces distributions limite leur utilisation. La distribution de cette classe à Comportement de Pareto asymptotique pour $0 < \alpha \leq 2$. Lorsque $\alpha = 2$, nous trouvons la distribution normale, cependant pour $\alpha < 2$ le moment d'ordre $r \geq \alpha$ n'est pas fini, ces distributions ont donc une variance infinie, et par conséquent une queue très lourde. Cette classe a une grande importance dans la théorie des valeurs extrêmes, puisque les distributions stables peuvent être caractérisées à partir du théorème de la Limite Centrale Généralisé. Une généralisation de ce théorème par Gnedenko et Kolmogorov (1954)[16].Récemment plusieurs logiciels ont été proposés pour permettre la résolution de ces problèmes, et on trouve des applications des distributions Stables dans plusieurs domaines tels que finances, physique et le trafic Internet.

2.5 La différence entre les distribution de la loi normale et les distributions à queues lourdes

En représente la différence dans le graphe ci-dessus

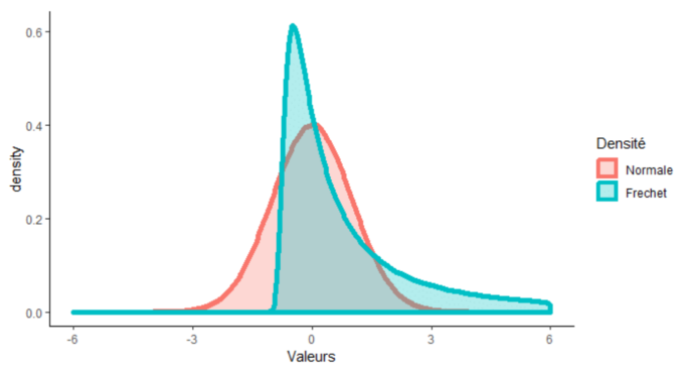


FIG. 2.2 – Comparaison d'une distribution normale à une distribution à queue lourde

D'après la figure ci-dessus on peut résumé la différence comme suite :

Les distributions à queue lourde sont celles dont les queues ne sont pas bornées de façon exponentielle. Contrairement à la courbe en cloche avec une distribution normale, les distributions à queue lourde s'approchent de zéro à un rythme plus lent et peuvent avoir des valeurs aberrantes avec des valeurs très élevées.(pour plus de détail voir [20]).

En termes de risque, les distributions à queue lourde ont une probabilité plus élevée qu'un événement important et imprévu se produise. Le modèle gaussien, ou courbe en cloche, de distribution normale

D'après l'équation suivante :

$$C_k = E\left(\frac{(X-\mu_x)^4}{\sigma_x}\right) = \frac{\delta_4}{\delta_2^2} > 3.$$

avec

δ_i :les moment centré d'ordre i.

C_k :coefficient d'aplatissement(Le kurtosis).

on peut conclure que la coefficient d'aplatissement (Le kurtosis) C_k de la distribution à queue lourde est supérieur à celui de la loi Normale (pour laquelle $C_k = 3$).

La différence entre la loi Normale et une loi avec une queue plus lourde a été illustrée par Hubert et Bendjoudi (1996)([21])comme. Dans cette figure on présente les fonctions de densité de probabilité de la loi normale et d'une distribution à queue plus lourde (la loi Halphen type B-1(HIB). On remarque que la fonction de répartition de la loi normale est presque nulle au niveau des extrêmes (queue droite), alors qu'elle ne l'est pas pour la loi HIB.

2.6 Estimation de la queue de la distribution

La distribution de Pareto généralisée est utilisée pour modéliser la queue (supérieur ou inférieur) d'une distribution présentant des valeurs extrêmes Pour pouvoir faire de l'estimation des quantiles, il faut avoir une formulation qui combine la distribution estimée dans la queue et la distribution centrale. Une telle formulation est donnée

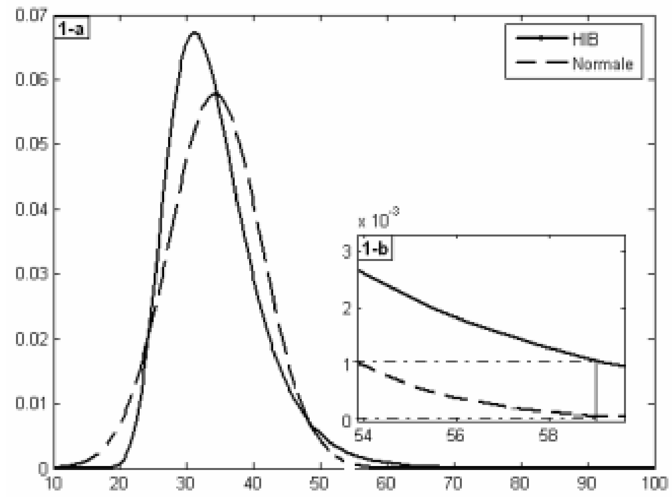


FIG. 2.3 – Illustration de la différance entre la loi normale et une loi à queue lourde(HIB)

par l'égalité suivante,$\forall u < x < x_F$

$$\overline{F}(x) = \overline{F}_u(x-u)\overline{F}(u), u < x < x_F.$$

où

$$F_u(y) = P(X - u \leq y \mid X > u).$$

Autrement dit, $\forall x > u$:

$$P(X > x) = P(X > u)P(X > x \mid X > u).$$

L'estimation se fait de la façon suivante :

$\overline{F}(x)$ est estimé avec la probabilité de dépassement empirique N_u/n.

$$\widehat{\overline{F}}(u) = \overline{F}_n(u) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X>u\}} = \frac{N_u}{n}, u < x_{F.}$$

la queue conditionnelle F_u de F est estimée par

$$\widehat{\overline{F}}(x-u) = 1 - G_{\widehat{\varepsilon},\widehat{\sigma}}(x-u) = \left(1 + \widehat{\varepsilon}_u \frac{x-u}{\widehat{\sigma}_u}\right)^{\frac{-1}{\widehat{\varepsilon}_u}}, u < x < x_F.$$

l'estimation de la queue de la distribution est donc :

$$\widehat{\overline{F}}(u) = \frac{N_u}{n}\left(1 + \widehat{\varepsilon}_u \frac{x-u}{\widehat{\sigma}_u}\right)^{\frac{-1}{\widehat{\varepsilon}_u}}, u < x < x_F.$$

2.7 Exemples des distributions à queues lourdes

2.7.1 Distribition de Cauchy (sur $\mathbb{R}$)

Sa fonction de densité f est donnée par :

$$f(x) = \frac{K}{\pi((x-\alpha)^2 + K^2)}; \text{tel que } K, x, \alpha \in \mathbb{R}$$

$$E(X^\alpha) < \infty, \forall \alpha < 1$$

$$E(X^\alpha) = \infty, \forall \alpha \geq 1$$

2.7.2 Distribution Weibull(sur $\mathbb{R}$)

Sa fonction de queue $\overline{F}$ donnée par :

$$\overline{F}(x) = e^{-(\frac{x}{K})^\alpha}, x \geq 0, 0 < \alpha < 1, \forall K > 0$$

$$E(X^\alpha) < \infty \text{ , } \forall \alpha.$$

Cette distribution est à queue lourde si et seulement si $\alpha < 1$. Tous les moments de cette distribution sont finis.

Remarque 2.7.1 *Si $\alpha = 1$,en retrouve la distribution exponentielle. Un exemple de distributions à queue non lourdes est la distribution exponentielle :*

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

2.7.3 Distribution de Burr

Sa fonction de queue est donnée par :

$$\overline{F}(x) = (\frac{K}{x^{\tau} + K})^{\alpha} \,;\, K, \alpha, \tau > 0, x \in \mathbb{R}^{+}.$$

On peut constater que cette distribution est une généralisation de la distribution de Pareto. Il est clair que

$$\overline{F}(x) \sim K^{\alpha} x^{-\alpha\tau} \text{ quand } x \to \infty,$$

donc la distribution de Burr a des moments finis pour tout ordre strictement inférieur à $\alpha\tau$ et des moment infinis pour les ordres qui sont supérieurs à $\alpha\tau$.
On a :

$$E(X^{\alpha}) < \infty, \forall \gamma < \alpha\tau.$$

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

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

2.7.4 Distribution de Pareto

Sa fonction de queue est donnée par :

$$\overline{F}(x) = (\frac{K}{x+K})^{\alpha} ; K \ , \alpha > 0 \ ,x \in \mathbb{R}^{+}$$

il est clair que

$$\overline{F}(x) \sim (\frac{x}{K})^{-\alpha} \text{ quand } x \to \infty,$$

la distribution de Pareto a des moment finie pour tout ordre strictement inférieur à α et des moments infinis pour les ordre qui sont supérieur à α .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)

2.7.5 Distribution lognormale(sur $\mathbb{R}^{+}$)

Sa fonction de densité f est donnée par :

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

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

Nous résumons ces exemples non exhaustifs sous forme de tableaux des distributions à queue légère et à queue lourde avec de nouveaux cas :

Nom	Paramétres	densité
Exponentielle	$\lambda > 0$	$\lambda e^{-\lambda x}$
Gamma	$\lambda > 0, \beta > 0$	$\frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
Weibull	$\lambda \geq 1, \beta > 0$	$\alpha\beta x^{\alpha-1} e^{-\beta x^{\alpha}}$

TAB. 2.1 – Distributions à queues légères

Nom	Paramétre	densité
Log-normale	$\mu \in \mathbb{R}, \sigma > 0$	$\frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma}}$
Pareto	$\alpha > 0, a > 0$	$\frac{\alpha}{a+x}\left(\frac{a}{a+x}\right)^{\alpha}$
Weibull	$0 < \alpha < 1, \beta > 0$	$\alpha\beta x^{\alpha-1} e^{-\beta x^{\alpha}}$
Burr	$\alpha > 0, \beta > 0, \gamma > 0$	$\frac{\alpha\gamma\beta x^{\gamma-1}}{(\beta+x^{\gamma})^{\alpha+1}}$

TAB. 2.2 – Distributions à queues lourdes

2.7.6 Etude des donnés réelles(Les rendements d'actifs)

Dans le cadre des distributions à variance infinie les lois stables sont les bonnes candidates en modélisation des rendements des actifs financiers dont le moment d'ordre 2 est infini dès que alpha est strictement inférieur à 2. Du fait qu'elles reflètent les propriétés statistiques communes à la plupart de ces derniers tels que les grandes à uctuations dont les accroissement ne sont pas réguliers au cours du temps (α < 2), une possibilité de dissymétrie dans la distribution, des queues épaisses, une distribution lebtokurtic ($kurtosis > 3$) et les moments d'ordre p ($p < \alpha$) sont infinis. Ce choix est justifié par au moins deux bonnes raisons :

1. Le théorème central limite généralisé, qui dit que les lois stables sont les seules distributions limites possibles pour des sommes, convenablement normalisées et centrées, des va $(i.i.d)$.

2. Les distributions stables peuvent être dissymétriques et permettent des queues épaisses, de telle sorte qu'elles ajustent les distributions empiriques beaucoup mieux que ne le font les distributions Gaussiennes.

Définition 2.7.1 *L'analyse statistique portée sur la distribution des rendements financiers est généralement basées sur la quantité :*

$$R_{t,t+\Delta t} = \log \frac{P_{t+\Delta t}}{P_t}, \tag{1}$$

où P_t est le prix d'un actif à l'instant t, qui peut être un stock, un indice de marché, ou un taux de change. On note que Δt désigne la rémunération générée par l'actif entre les dates t et $(t+T)$ dont les rendements sont calculés, par exemple, dans une période d'un jour, d'une semaine ou d'un mois.

– Les diverses propriétés statistique suivantes généralement observée dans les données financières, telles que le rendement des actifs, taux d'intérêt, taux de change, les actions :

– Les queues de distribution se comportent comme une loi puissance, pour x assez grand,

$$P(X > x) = 1 - F(x) \sim cx^{-\alpha}$$

Plus α est petit et plus la queue est épaisse,

α est appelé l'exposant de queue. $\gamma = 1/\alpha$ sera appelé l'exposant des valeurs extrêmes (Extrême Value Index). La plupart des actifs financiers possèdent des exposants de queues compris entre 3 et 5.

– Les moments d'ordres k : $E[x^k]$

ne sont définis que si $k < \alpha$,

– Les moments d'ordre $k \geq \alpha$ sont infinis :

– si $\alpha < 4$ la kurtosis est infinie

– si $\alpha < 2$ la variance est infinie

– si $\alpha < 1$ la moyenne est infinie

– Une possibilité de dissymétrie dans la distribution empirique des rendements, asymétrie négative : biais vers les pertes plus élevées.

– Des grandes fluctuations, où les accroissement ne sont pas régulières c'est à dire une variance infinie au cours du temps.

2.7.7 Modélisation

Certaines Distributions Spécifiques

- Le but dans cette section est de rendre en compte ces propriétés par un modèle.
- Les distributions $\alpha-$stables sont les bonnes candidates en modélisation des rendements des actifs financiers. Mandelbrot (1963), suivi par Fama (1965), ont proposé comme alternative pour modéliser les séries financières, la distribution Lévy-Stable (Lévy, 1925) ayant des propriétés très proches des distributions empiriques à queues lourdes. Ce choix est justifie par au moins deux bonnes raisons :

- Le théorème central limite généralisé, qui dit que les lois stables sont les seules distributions limites possibles pour des sommes, convenablement normalisées et centrées, de $v.a$ (iid).
- Les distributions stables peuvent être dissymétriques et permettent des queues épaisses, de telle sorte qu'elles ajustent les distributions empiriques beaucoup mieux que ne le font les distributions Gaussiennes.

Alors, une fonction de distribution est dans le domaine d'attraction de Lévy avec un indice de stabilité $0 < \alpha < 2$, noté $F \in D(\alpha)$, s'il existe deux séquences réelles $A_n > 0$ et C_n telle que :

$$A_n^{-1}\left(\sum_{i=1}^{n} X_i - C_n\right) \xrightarrow{D} S_\alpha(\gamma, \beta, \delta)\,, \text{ quand } n \to \infty$$

où D est la convergence en distribution et $S_\alpha(\sigma, \beta, \mu)$ est la distribution de Lévy avec $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma > 0$ et $\delta \in R$. On prend comme exemple des différentes indices boursiers Journaliers, hebdomadaires et mensuelles, le tableau suivant contient les valeurs estimées des paramètres de la loi stable via la méthode de McCuloch :

	Parameters	Nazdaq	Dow Jens	CAC40	Nekkei
Daily	$\widehat{\alpha}$	1.386	1.48	1.664	2
	$\widehat{\beta}$	-0.165	0	-0.36	0
	$\widehat{\gamma}$	0.005	0.005	0.007	0.007
	$\widehat{\delta}$	0.0011	0.0003	0.0003	0.0003
Weekly	$\widehat{\alpha}$	1.48	1.56	1.729	1.91
	$\widehat{\beta}$	-0.20	-0.25	-0.48	-1.59
	$\widehat{\gamma}$	0.015	0.012	0.017	0.017
	$\widehat{\delta}$	0.003	0.002	0.001	0.002
Monthly	$\widehat{\alpha}$	1.56	1.66	2	2
	$\widehat{\beta}$	-0.25	-0.36	-2.16	-2.16
	$\widehat{\gamma}$	0.036	0.028	0.04	0.04
	$\widehat{\delta}$	0.017	0.008	0.012	0.007

Conclusion

L'objectif de ce mémoire est présenter une catégorie de distributions à queue lourde qui s'appuie sur la théorie des valeurs extrêmes.

Cette approche nous permet d'évaluer les événements rares et les pertes qui en découlent. D'où en fait dans La première partie du mémoire porte sur la théorie des probabilités et des statistiques. Elle rappelle les concepts de base tels que la fonction de répartition, la fonction quantile et les fonctions empiriques, ainsi que la loi normale et ses propriétés.

La deuxième partie se concentre sur la description des distributions à queues lourdes. Nous commençons par introduire les propriétés statistiques communes à la plupart de ces distributions. Ensuite, nous expliquons comment modéliser ces phénomènes et justifier les causes possibles à cela. Enfin, nous illustrons cette partie par une étude des données réelles, en prenant comme exemple les rendements d'un actif financier.

Nous avons effectué une analyse des données réelles pour quelques indices boursiers afin d'illustrer l'application de notre méthode.

Bibliographie

[1] Alfonsi (2008), Fonction de R´epartition et Copules, Cours de Master Recherche : Probabilit´e et Application.

[2] Bailly, P., Carrère, C. (2015). Statistiques descriptives : Théorie et applications, PUG, coll., p. 165-167.

[3] Balakrishnan, N and Cohen, A. C. (1991). Order Statistics and Inference :EstimationMethods. Statist. Model. Decis. Sci. Academic Press

[4] Balkema, A., de haan, L. (1974). Residual life Time at Great Age. Annals of Probability,2 ; 792 ; 804.

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

[6] .Bernard, P.M., Lapointe, C. (1995). Mesures statistiques en épidémiologie, Presses de l'Université du Québec, Sainte-Foy, page 89.

[7] Charles Suquet, Th eor eme central limite.

[8] Chun Su, Qi-he Tang Acta Mathematicae Applicatae Sinica, English Series. Vol.19,No.1 (2003)135 - 137.

[9] David, H.A. (1970). Order Statistics. New York-London Sydney.

[10] Dress, François. "Les probabilités et la statistique de A à Z, dictionnaire." HAL 2007 (2007).

[11] Dufour A,Chessel D et Lobry J. (2018) : Lois de Probabilités. Univ-lyon1

[12] Dusart P. (2015). Cours de Statistiques inférentielles.

[13] Embrechts, P. Kluppelberg, C. et Mikosch, T. (2003). Modelling Extremal Events for Insurance and Finance. volume 33 of Applications of Mathematics Springer.

[14] Gabriela Ciuperca and Cécile Mercadier. Semi-parametric estimation for heavy tailed distributions. Extremes.

[15] Gardes, L., and Stuper, G. (2015) : Estimating extreme quantiles under random truncation. TEST, 24 ; 207 227.

[16] Gnedenko B. V. Kolmogorov A. N .(1954). Limit distribution for sums of random variable, revised engl transl. Reading Mass. Addison-Wesley.

[17] Haan, L., and Ferreira, A. (2006). Extreme value theory : an Introduction. Springer.

[18] Halphen, E. (1955). Les fonctions factorielles. Publications de l'Institut de Statistique de l'Université de Paris, Vol. IV, Fascicule I, 21-39.

[19] Hill, B. (1975). A Simple Approach to Inference About the Tail of a Distribution. Annals of Statistic.

[20] htts ://www.researchgate.net/figure/Comparaison-dune-distribution-normale-a-une-distribution-a-queue-lourde-fig10-334468946.

[21] Hubert P & Bendjoudi H. (1996). Introduction à l'étude des longues séries pluvoi-métriques. In : Journées hyrologiques de l'ORSTOM, Montpellier.pp 20.

[22] Jan Beirlant, Yuri Goegebeur, Johan Segers, and Jozef L Teugels. Statistics of extremes : theory and applications, volume 558. John Wiley & Sons, 2004

[23] J.Jiang, Large Sample Techniques for Statistics

[24] Laurent GARDES., 2003., Estimation d'une fonction quantile extrème.

[25] Lévy,P. (1925). Calcul des Probabilité. Gauthier-villars. Paris.

[26] Malamud and D. Turcotte. The applicability of power-law frequency statistics to floods. Journal of Hydrology, 322 : 168–180 ; 200.

[27] Nolan, J. P. (2001). Maximum likelihood estimation of stable parameters.

[28] Nolan, J. P. (2006). Stable Distributions - Models for Heavy Tailed Data. Boston : Birkhuser. In progress, Chapter 1 online at academic2.american.edu/~jpnolan.

[29] Ouarda, T. B.M.J., Ashkar, F., Bensaid, E. et I. Hourani (1994). Distributions statistiques utilisées en hydrologie : Transformations et propriétés asymptotiques. Département de mathématique, Université de Moncton, 31 p

[30] Perreault, L., Bobée, B. and Rasmussen, P.F. (1999a). Halphen Distribution System. I : Mathematical and Statistical Properties. Journal of Hydrologic Engineering,4 (3), 189- 199.

[31] Philippe Soulier., 2009., Some applications of regular variation in probability and statistics.

[32] Pickands, J. (1975). Statistical inference extreme order statistics. Annals of Statistics 3 : 119-130.

[33] P. Billingsley. Probability and measure. Wiley, third edition 1995.

[34] P. S. Toulouse. Th'emes de Probabilit´es et Statistique. Dunod, 1999

[35] Resnick, S.I. (1987). Extreme Values, Regular Variation, and point Processes. Springer,New York.

[36] Rosen, O. and Weissman, I. (1996). Comparison ofEstimation Methods in Extreme Value Theory. Communication in Statistics – Theory and Methods 25 (4), 759-773.

[37] Saporta, G. (2011). Probabilités, analyse des données et statistique. Technip, Paris.

[38] Scarf, P.A. (1992). Estimation for a four parameter generalized extreme value distribution, Comm. Stat. Theor. Meth., 21, 2185-2201.

[39] Teugels, J.L. (1975). The class of subexponential distributions. Ann. Proba .3 (6) 1001–1011.

[40] Werner, T. and C. Upper (2002). Time Variation in the tail behaviour of Bund Futures Returns.Working paper N 199. European Central Bank.

Annexe A : Logiciel R

2.8 Qu'est-ce-que le langage R ?

• Le langage R est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests d'hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.

• R a été créé par Ross Ihaka et Robert Gentleman en 1993 à l'Université d'Auckland, Nouvelle Zélande, et est maintenant développé par la R Developement Core Team.

L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (Ross Ihaka et Robert Gentleman) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparenté.

Annexe B : Abréviations et Notations

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

Notation		Significations
$c.à.d$	:	C'est à dire
$v.a$	:	Variable aléatoire
$i.i.d$	:	Indépondantes et identiquement distribuées
TCL	:	Théoreme centrale limite.
$\mathbb{R}$	:	Ensembles des valeurs réels
F	:	Fonction de répartition(f.d.r)
F_n	:	Fonction de répartition empirique
$var(X)$	:	Variance de X
$\mu = E(X)$	:	Espérance mathématique de$X$
(Ω, F, P)	:	Espace probabilisé
Q	:	Fonction de quantile
Q_n	:	Quantile empirique
$N(0,1)$	:	Loi normale standard

$\overline{X}$	:	Moyenne arétmitique
$\xrightarrow{p}$	:	Converge en probabilité
$\xrightarrow{p.s}$	:	Converge presque sûre
$X_{n,n}$	:	Maximum de $X_1, ..., X_n$.
$X_{1,n}$	:	Minimum de $X_1, ..., X_n$.
$X_1, ..., X_n$	:	Echantillon de taille n de X
$\sup(x)$	:	sup de l'ensemble x.
$\overline{F}$	:	Fonction suivede F
I_x	:	Indicatrice de x

Résumé

Notre tache dans ce travail est de présenter une classe de distributions à queue lourdes basée sur la théorie des valeurs extremes qui permet d'évaluer les évenements rareset les pertes associées à leur survenance. Une étude des données réelles a été réalisée pour quelques indices bourciers.

Mots clés : Loi normale, distribution à queue lourde, α-stable, indices bourciers.

Summary

Our task in this work is to present a class of heavy-tailed distributions based on extreme value theory that allows for the evaluation of rare events and the losses associated with their occurrence. A study of real data was carried out for some stock market indices.

Key words: Normal distribution, heavy-tailed , α- stable distribution, stock indices.distribution

ملخص

مهمتنا في هذا العمل هو تقديم فئة من التوزيعات ذات الذيول الثقيلة بناءً على نظرية القيم المتطرفة التي تسمح بتقييم الأحداث النادرة والخسائر المرتبطة بحدوثها. تم إجراء دراسة للبيانات الحقيقية لبعض المؤشرات البورصية.

الكلمات المفتاحية: التوزيع الطبيعي, التوزيع الثقيل الذيل, الثبات الفا, مؤشرالاسهم.