

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

**UNIVERSITE MOHAMED KHIDER, BISKRA**

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

**DEPARTEMENT DE MATHEMATIQUES**



Mémoire présenté par

**Azizi Soumia**

En vue de l'obtention du Diplôme de

**MASTER en Mathématiques**

Option : **Statistique**

Titre

**Généralités sur les distributions à queue lourde**

Membres du Comité d'Examen

<b>Pr.</b>	Meraghni Djamel	UMKB	Président
<b>Dr.</b>	Touba Sonia	UMKB	Encadreur
<b>Dr.</b>	Berkane Hassiba	UMKB	Examinatrice

**Soutenu publiquement le 26 Juin 2022**

## Dédicace

C'est avec une très grande émotion et un immense plaisir que je dédie

Ce modeste travail

À Mes chers père et mère, Pour tous les sacrifices qu'ils ont consentis pour me permettre de pour suivre mes études dans les meilleures conditions possibles et pour n'avoir jamais cessé de m'encourager tout au long de mes années d'étude.

À mes très chers frères et sœurs.

À tous mes amis qui m'ont toujours encourager.

À toute ma famille.

À tous ceux que j'aime et ceux qui m'aiment.

## REMERCIEMENTS

En premier lieu, je remercie *Dieu* le Tout Puissant de m'avoir donné le courage et la santé pour achever ce travail.

L'encadrement scientifique de ce travail a été assuré par Madame TOUBA SONIA, Dr à l'université de Biskra. Je tiens vivement à lui exprimer ma profonde reconnaissance et mes sincères pour m'avoir encadré, aider et orienter pour l'élaboration de ce mémoire. Je remercie également Monsieur MERAGHNI DJAMEL, Pf à l'université de Biskra, pour l'honneur qu'il nous fait en acceptant de présider le jury de ce mémoire. Ses grandes compétences dans le domaine de la recherche scientifique permettront assurément d'enrichir et d'améliorer ce travail. Veuillez trouver ici, Monsieur, le témoignage de mon profond respect.

Je remercie aussi, Madame BERKANE HASSIBA, Dr à l'université de Biskra, pour avoir accepté d'examiner et d'honorer de sa présence effective le jury de ce mémoire. Enfin, que toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce travail, trouvent ici ma sincère reconnaissance et mes remerciements.

# Table des matières

<b>Dédicace</b>	<b>i</b>
<b>Remerciements</b>	<b>ii</b>
<b>Table des matières</b>	<b>iii</b>
<b>Liste des figures</b>	<b>v</b>
<b>Liste des tableaux</b>	<b>vi</b>
<b>Introduction</b>	<b>1</b>
<b>1 La loi normale</b>	<b>3</b>
1.1 Historique . . . . .	3
1.2 Définitions . . . . .	4
1.3 Fonction de répartition . . . . .	5
1.4 Espérance et Variance . . . . .	5
1.5 Fonction caractéristique . . . . .	7
1.6 Fonction génératrice des moments . . . . .	8
1.7 Théorème central limite(TCL) . . . . .	8
1.8 Estimation . . . . .	10
1.8.1 Méthode du maximum de Vraisemblance (M.V) . . . . .	10
1.8.2 Estimation sous R ( $\mu$ et $\sigma^2$ ) . . . . .	12

1.9 Application de la loi normale . . . . .	13
1.9.1 Domaine d'utilisation . . . . .	13
1.9.2 Test de normalité . . . . .	13
<b>2 Distribution à queues lourdes</b>	<b>17</b>
2.1 Distribution à queues légères . . . . .	17
2.2 Distribution à queues lourdes . . . . .	18
2.2.1 Définitions . . . . .	18
2.2.2 Propriétés des distributions à queues lourdes . . . . .	19
2.2.3 Les sous-classes de distributions à queues lourdes . . . . .	21
2.2.4 Caractérisation des distributions à queues lourdes . . . . .	24
2.2.5 Exemples de distributions à queues lourdes . . . . .	26
2.3 La différence entre distribution de la loi normale et à queue lourde . . . . .	28
<b>Conclusion</b>	<b>31</b>
<b>Bibliographie</b>	<b>31</b>
<b>Annexe A : Quelques éléments du logiciel R</b>	<b>35</b>
<b>Annexe B : Abréviations et Notations</b>	<b>36</b>

# Table des figures

1.1 Abraham de Moivre . . . . .	4
1.2 La fonction de densité de la loi normale de paramètre $\mu$ et $\sigma^2$ . . . . .	4
1.3 PDF d'un échantillon normal (0, 1) et de taille 60 et la fonction de densité cumulée empirique (eCDF) de l'échantillon en bleu, la fonction de distribution cumulative (CDF) de la normale (0, 1) en noir et la statistique $D_n$ du test de Kolmogorov-Smirnov. . . . .	15
2.1 comparaison d'une distribution normale a une distribution à-queue lourde.	28
2.2 Illustration de la différence entre la loi normale et une loi à queue lourde (HIB) . . . . .	29
2.3 Comparaison du Comportement de la Queue. . . . .	30

# Liste des tableaux

2.1 Exemple de distribution à queue légère . . . . .	18
--	----

# Introduction

La théorie des probabilités et en statistique, les lois normales sont parmi les lois des probabilités les plus utilisées pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires. Elles sont également appelées lois Gaussiennes, loi de Gauss ou lois de Laplace-Gauss des noms de Laplace (1749-1827) et Gauss (1777-1855), deux mathématiciens, astronomes et physiciens qui l'ont étudiée.

Plus formellement, une loi normale est une loi de probabilité absolument continue qui dépend de deux paramètres : son espérance, un nombre réel noté  $\mu$  et son écart-type, un nombre réel positif noté  $\sigma$ .

Parmi les lois de probabilité, les lois normales prennent une place particulière grâce au théorème central limite. En effet, elles correspondent au comportement, sous certaines conditions, d'une suite d'expériences aléatoires similaires et indépendantes lorsque le nombre d'expériences est très élevé.

Grâce à cette propriété une loi normale permet d'approcher d'autres lois et ainsi de modéliser de nombreuses études scientifiques comme des tests statistiques en utilisant par exemple les tables de la loi normale centrée réduite par contre, il existe une loi.

Une loi de probabilité dont les queues ne sont pas exponentiellement bornées, ce qui signifie qu'elles ont des queues plus lourdes que la loi exponentielle.

Dans de nombreuses applications, c'est la queue droite de la distribution qui est intéressante, mais une distribution peut avoir une queue lourde à gauche, on les deux queues peuvent être lourdes.



Il ya trois sous classes importantes de distributions à queue lourde, les lois à queue épisse, les lois à langue queue et les lois sous-exponentielles. Dans la pratique toutes les lois à queue lourde couramment utilisées appartiennent à la classe sous-exponentielle

Ce travail se déroule autour de deux chapitres ,dans le premier chapitre on s'intéresse à l'étudié la loi normale, leur définition,propriété et caractéristique.

Le deuxième chapitre donne une définition et des exemples de distributions à queues, ainsi, on s'intéresse aux propriétés caractéristiques de ces dernières, i.e. des distributions dont trois les moments exponentiels sont infinis.

Ensuite, on présentera les différentes classes des distributions à queue lourdes : on parlera des distributions sous exponentielles et des distributions régulières qui sont les classes les plus importantes de distributions à queue lourde.

# Chapitre 1

## La loi normale

L'objectif du chapitre est de représenter la loi normale, le modèle probabiliste, le plus utilisé décrivant de nombreux phénomènes observés en pratique.

### 1.1 Historique

En théorie des probabilités et en statistique, la loi normale est l'une des lois de probabilité les mieux adaptées pour modéliser des phénomènes naturels provoqués par plusieurs événements aléatoires. La loi normale est apparue pour la première fois en 1733 par Abraham de Moivre en approfondissant l'étude de jeu de pile et de face. Il a en fait remarqué que lorsque le nombre de lancers devient très important, la probabilité d'obtenir  $K$  pile de  $n$  lancers se rapproche d'une courbe en cloche [14]

Elle est également connue sous le nom de la loi de Gauss ou la loi de Laplace-Gauss. Cette loi était alors considérée comme un idéal à réaliser par la nature, car elle est omniprésente dans un grand nombre de phénomènes physiques, biologiques et sociaux, d'où le nom de "normale".



FIG. 1.1 – Abraham de Moivre

## 1.2 Définitions

**Définition 1.2.1 (La loi normale générale)** Une  $v$  à réelle  $X$  suit une loi normale (où loi gaussienne, loi de Laplace-gauss) d'espérance  $\mu$  et de variance  $\sigma^2$  si cette  $v$  à réel  $X$  admet pour densité de probabilité la fonction  $f(x)$  définie, pour tout nombre réel  $x$ , par

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\},$$

en résumé cette loi par la notation  $\mathcal{N}(\mu, \sigma^2)$ .

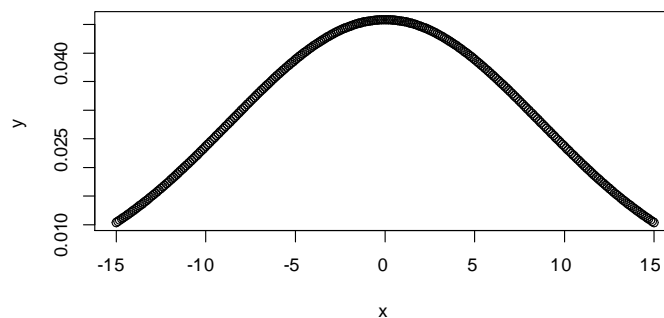


FIG. 1.2 – La fonction de densité de la loi normale de paramètre  $\mu$  et  $\sigma^2$

**Définition 1.2.2 (La loi normale standard)** la loi normale standard  $\mathcal{N}(0, 1)$  est celle

de densité

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{t^2}{2}\right\}.$$

### 1.3 Fonction de répartition

**Définition 1.3.1** On appelle fonction de répartition d'une v a  $X$  qui suit la loi  $\mathcal{N}(\mu, \sigma^2)$ , la fonction définie par

$$F_x(a) = P(X \leq a) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^a \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} dx.$$

**Remarque 1.3.1** L'intégrale de la densité d'une v a suit une loi normale, n'ayant pas d'expression mathématique explicite, résultant en l'émergence de tables de la fonction de répartition.

### 1.4 Espérance et Variance

une v a qui suit la loi normale de paramètre  $\mu$  et  $\sigma^2$ , admet une espérance et une variance telle que

$$E(X) = \mu \text{ et } V(X) = \sigma^2.$$

**Proof. 1/** L'espérance de  $X$  s'écrit

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} dx. \end{aligned}$$

En effet un changement de variable

$$z = \frac{x - \mu}{\sigma} \implies x = z\sigma + \mu \implies dx = \sigma dz.$$

En écrit alors

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (z\sigma + \mu) \exp\left(\frac{-(z)^2}{2}\right) \sigma dz \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \exp\left(\frac{-(z)^2}{2}\right) dz + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(\frac{-(z)^2}{2}\right) dz \\ &= \mu. \end{aligned}$$

**2/** La variance de  $X$  s'écrit

$$V(X) = E(X^2) - (E(X))^2,$$

et on a

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x^2 \exp\left\{-\frac{1}{2\sigma^2} (x - \mu)^2\right\} dx \end{aligned}$$

en posède un changement de variable  $z = \frac{x-\mu}{\sigma}$  on obtient

$$\begin{aligned}
 E(X^2) &= \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{+\infty} (z\sigma + \mu)^2 \exp\left(\frac{-(z)^2}{2}\right) \sigma dz \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 \sigma^2 \exp\left(\frac{-(z)^2}{2}\right) dz + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mu^2 \exp\left(\frac{-(z)^2}{2}\right) dz \\
 &\quad + \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z\mu\sigma \exp\left(\frac{-(z)^2}{2}\right) dz. \\
 &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 \exp\left(\frac{-(z)^2}{2}\right) dz + \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(\frac{-(z)^2}{2}\right) dz \\
 &\quad + \frac{2\mu\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \exp\left(\frac{-(z)^2}{2}\right) dz. \\
 &= \sigma^2 + \mu^2.
 \end{aligned}$$

donc

$$\begin{aligned}
 V(X) &= E(X^2) - (E(X))^2 \\
 &= \sigma^2 + \mu^2 - \mu^2 \\
 &= \sigma^2.
 \end{aligned}$$

d'où  $E(X) = \mu$  et  $V(X) = \sigma^2$ . ■

## 1.5 Fonction caractéristique

**Définition 1.5.1** On appelle fonction caractéristique de la v a  $X$  qui suit la loi normale  $\mathcal{N}(\mu, \sigma^2)$ . La fonction  $\varphi_X(t)$ , définie sur  $\mathbb{R}$  à valeur complexe par

$$\varphi_X(t) = \exp(it\mu) \exp\left(\frac{-(t\sigma)^2}{2}\right).$$

**Proof.** On a

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp(itx) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

On effectue un changement de variable

$$s = \frac{x - \mu}{\sigma} \implies x = s\sigma + \mu \implies dx = \sigma ds,$$

alors

$$\begin{aligned} \varphi_X(t) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{s^2}{2}\right) \exp(it(s\sigma + \mu)) \sigma ds \\ &= \frac{\exp(it\mu)}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}(s - \sigma t)^2\right) \exp\left(-\frac{1}{2}(\sigma t)^2\right) ds \\ &= \frac{\exp(it\mu) \exp(-\frac{1}{2}(\sigma t)^2)}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}(s - \sigma t)^2\right) ds \\ &= \exp(it\mu) \exp\left(-\frac{(\sigma t)^2}{2}\right), \end{aligned}$$

alors  $\varphi_X(t) = \exp(it\mu) \exp\left(-\frac{(\sigma t)^2}{2}\right)$ . ■

## 1.6 Fonction génératrice des moments

**Définition 1.6.1** La fonction génératrice des moments d'une loi  $\mathcal{N}(\mu, \sigma^2)$  est égale, pour tout  $t \in \mathbb{R}$  à :

$$g(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

## 1.7 Théorème central limite(TCL)

C'est ce théorème qui permet d'être sûr que les lois normales sont des lois de phénomènes naturels. Par exemple, si l'on regarde la taille des individus dans une population, celle-ci

suivra une distribution similaire à la loi normale.

**Définition 1.7.1** *Le théorème central limite établit la convergence vers la loi de Gauss sous des hypothèses peut contraignantes.*

**Théorème 1.7.1** *Soit  $(X_n)$  une suite de v.a indépendantes des mêmes loi d'espérance  $\mu$  et d'écart-type  $\sigma$ , alors*

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \quad (1.1)$$

où  $\xrightarrow{\mathcal{L}}$  signifie convergence en loi ou en distribution.

**Proof.** *Il existe une démonstration particulièrement simple utilisant les fonctions caractéristique cette démonstration ressemble à celle d'une des lois des grands nombres. Pour une v.a  $Y$  d'espérance 0 et variance 1. La fonction caractéristique admet un développement limité*

$$\varphi_Y(t) = 1 - \frac{t^2}{2} + o(t^2), \quad t \longrightarrow 0.$$

si  $Y$  vaut  $\frac{X_i - \mu}{\sigma}$ , il est facile de voir que la moyenne centrée réduite des observations  $X_1, \dots, X_n$  est simplement

$$Z_n = \frac{X_n - \mu}{\sigma/\sqrt{n}} = \sum_{i=1}^n \frac{Y_i}{\sqrt{n}}$$

D'après les propriétés élémentaires des fonction caractéristique, la fonction caractéristique de  $Z_n$  est

$$\left[ \varphi_Y \left( \frac{t}{\sqrt{n}} \right) \right]^n = \left[ 1 - \frac{t^2}{2n} + o \left( \frac{t^2}{n} \right) \right]^n \longrightarrow \exp \left( \frac{-t^2}{2} \right) \text{ lorsque } n \longrightarrow +\infty,$$

mais cette limite est la fonction caractéristique de la loi normale centrée réduite  $\mathcal{N}(0, 1)$ , d'où l'on déduit le théorème de la limite centrale grâce au théorème de continuité de Lévy, que la convergence des fonctions caractéristiques implique la convergence en loi.



■

**Remarque 1.7.1 (La continuité de Lévy)**

Posons une suite de v a  $(X_n)_{n \in \mathbb{N}}$  pas nécessairement définies sur le même espace de probabilité. Les fonctions  $\varphi$  et  $\varphi_n$  sont les fonctions caractéristiques respectives des variables aléatoires  $X_n$  et  $X$  définies par :

$$\varphi_n(t) = E(e^{itX_n}), \forall t \in \mathbb{R}, \forall n \in \mathbb{N} \text{ et } \varphi(t) = E(e^{itX}) \forall t \in \mathbb{R}.$$

la théorème de continuité de Lévy

$$\{\forall t \in \mathbb{R} \varphi_n(t) \longrightarrow \varphi(t)\} \iff \left\{ X_n \xrightarrow{\mathcal{L}} X \right\}.$$

## 1.8 Estimation

L'estimation comporté à donné des valeurs approximative aux paramètre d'une population à l'aide d'une échantion de  $n$  observations issues de cette population.

On peut se tromper sur la valeur precque, mais on donne la meilleure valeur possible que l'on peut supposer. Il existe plusieurs méthodes pour trouver l'estimateur d'une paramètre, Dans ce paragraphe, on s'intéressera aux deux premières méthodes.

### 1.8.1 Méthode du maximum de Vraisemblance (M.V)

C'est la plus répandu et la plus attrayant pour les raisons suivantes

1. Plus facile à mettre en œuvre et plus simple que les méthodes alternatives. elle Cela revient au problème classique de la résolution d'équations.
2. elle est optimal et asymptotiquement efficace (c'est-à-dire lorsque la taille de l'échantillon tend théoriquement vers l'infini). D'un point de vue pratique, pour un échantillon assez grand ( $n > 30$ ), elle fournit un estimateur de très bonne qualité.

3. elle donne l'estimateur sans biais avec une variance asymptotique minimale (voir [29]).

**Définition 1.8.1** Cette méthode consiste, étant donnée un échantillon de valeur  $X_1, \dots, X_n$  à prendre comme estimation de  $\theta$ , la valeur de  $\theta$  qui rend maximale la vraisemblance

$$L(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta),$$

en pratique, on prend comme estimation de  $\theta$  une solution à l'équation

$$\frac{\partial}{\partial \theta} \ln L(X, \theta) = 0,$$

dite l'équation de la vraisemblance.

**Exemple 1.8.1**  $(X_1, X_2, \dots, X_n)$  un échantillon aléatoire suit la loi  $\mathcal{N}(\mu, \sigma^2)$  et on peut estimer les paramètres de cette loi par (M.V). On a la fonction de densité de la loi normale

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} (x_i - \mu)^2\right\},$$

alors, l'équation de vraisemblance

$$\begin{aligned} L(x_1, \dots, x_n; \mu, \sigma^2) &= \prod_{i=1}^n f(x_i; \mu, \sigma^2) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \end{aligned}$$

$$\ln L(x_1, \dots, x_n; \mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

donc

$$(1) \frac{\partial \ln L(x_1, \dots, x_n; \mu, \sigma^2)}{\partial \mu} = \frac{2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

$$(2) \frac{\partial \ln L(x_1, \dots, x_n; \mu, \sigma^2)}{\partial \sigma^2} = -n + \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = 0$$

$$(1) \implies \sum_{i=1}^n x_i = n\mu \implies \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

$$(2) \implies \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = n$$

$$\implies \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

$$\implies \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Alors  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$ ,  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ .

### 1.8.2 Estimation sous $\mathbf{R}$ ( $\mu$ et $\sigma^2$ )

On peut estimer les paramètres de  $\mu$  et  $\sigma^2$  par logiciels, parmi lesquels on a choisi le logiciel **R**.

On suppose  $\theta(X_1, \dots, X_n)$  comme estimateur de (la valeur inconnue du paramètre).

-Sur un échantillon  $(X_1, \dots, X_n)$  on va estimer le paramètre  $\mu$  :

`theta <- 4` # valeur supposée inconnue.

`mean(rnorm(200, mean=theta))` # on se sert de l'échantillon observé  $(X_1, \dots, X_n)$  :

[1] 3.909667.

-Sur un échantillon  $(X_1, \dots, X_n)$  on va estimer le paramètre  $\sigma^2$  :

`theta <- 3` #  $V(X_1)$ , supposée inconnue.

`vecx <- rnorm(200, sd=sqrt(theta))`

`mean(vecx^2) - (mean(vecx))^2` # on se sert uniquement de l'échantillon observé.

[1] 3.187733.

## 1.9 Application de la loi normale

### 1.9.1 Domaine d'utilisation

La loi normale est la loi statistique la plus utilisée de toutes les distributions, en raison de la distributions de la loi normale se proche bien de nombreux phénomènes naturels, elle a évalué en une la norme de référence pour de nombreux problèmes de probabilité.

Elle représente la loi de distribution d'une  $v$  a  $X$  demandant d'un grand nombre de facteurs agissant sous forme additive ,chacun ayant une variable faible par rapport à la variance résultante.

Elle peut représenter la fin de vie des dispositifs subissant un phénomène de vieillissement, usure, corrosion...

### 1.9.2 Test de normalité

Le test de normalité est la partie importante requise pour observer la distribution des données. Des cas particuliers des tests (où tests d'ajustement, tests permettant de comparer des distributions), appliqués une loi normale.

Le test de normalité a deux conditions qui dépendent de la distribution des données, soit normale, soit pas normal. Si la distribution des données était normale, la méthode paramétrique pourrait être appliquée pour analyser les données. Sinon, si les données ne sont pas normales, la métrique non-paramétrique doit être appliquée pour analyser les données appliquées pour analyser les données.

Ces tests prennent une place importante en statistique. En effet, de nombreux tests supposent la normalité des distributions pour être applicables. En toute rigueur.

On cherche à se déterminer entre

$$\left\{ \begin{array}{l} H_0 : \text{les données suivent une loi normale} \\ H_1 : \text{les données ne suivent pas une loi normale} \end{array} \right. ,$$

on propose deux tests pour tester la normalité d'un échantillon.

### Test de kolmogorov-Smirnov

Le test de kolmogorov est un test non-paramétrique qui compare la distribution d'un échantillon statistique à une distribution fixée. Les distributions sont représentées par leur fonction de répartition, utilisées pour l'exécution du test. Étant donné un échantillon  $(X_1, X_2, \dots, X_n)$  de  $n$  valeurs observées d'une v a  $X$  de fonction de répartition  $F(x)$

Hypothèse testée

$$H_0 : F^\circ = F \text{ contre } H_1 : F^\circ \neq F.$$

Pour cela on définit la statistique de *Kolmogorov-Smirnov* par

$$D_n = \sup_{x \in \mathbb{R}} | F_n(x) - F^\circ(x) |$$

où  $F_n(x)$  est la proportion des observations dont la valeur est inférieure ou égale à  $x$  (fonction de répartition empirique).

La région critique : on rejette  $H_0$  si  $Dn > d(n)$  ou  $d(n)$  est le quantile théorique (si  $n \leq 80$  on lu à partir de la table uniforme, si  $n > 80$  on utilise la table de Kolmogorov-Smirnov).

### Test de Shapiro-Wilk

Le test de Shapiro-Wilk est très connue et populaire, il est basé sur la statistique  $W$ . En comparaison des autres tests, il est particulièrement puissant pour les petits effectifs ( $n \leq 50$ ) :

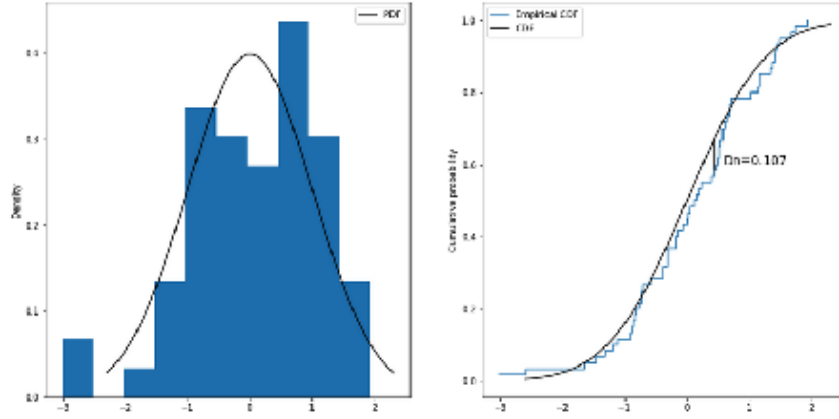


FIG. 1.3 – PDF d’un échantillon normal (0, 1) et de taille 60 et la fonction de densité cumulée empirique (eCDF) de l’échantillon en bleu, la fonction de distribution cumulée (CDF) de la normale (0, 1) en noir et la statistique  $D_n$  du test de Kolmogorov-Smirnov.

La statistique  $W$  est donnée par

$$W = \frac{\left[ \sum_{i=1}^n a_i x_{(i)} \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

où

$x_i$  : (avec des parenthèses entourant l’indice  $i$ ) désigne la  $i$ ème statistique d’ordre, i.e.le  $i$ ème plus petit nombre dans l’échantillon ;

$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$  est la moyenne de l’échantillon ;

$a_i$  : sont des constantes générées à partir de la moyenne et de la matrice de variance co-variance des quantiles d’un échantillon de taille  $n$  suivant la loi normale.

La statistique  $W$  peut donc être interprétée comme le coefficient de détermination (le carré du coefficient de corrélation) entre la série des quantiles générées à partir de la loi normale et les quantiles empiriques obtenues à partir des données.

La région critique

$$W < W_{crit}$$

Les valeurs seuils  $W_{crit}$  pour différents risques  $\alpha$  et effectif  $n$  sont lues dans la table de Shapiro-Wilk.

# Chapitre 2

## Distribution 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.

### 2.1 Distribution à queues légères

**Définition 2.1.1** Une distribution  $F$  est dite à queues légères si et seulement si

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

Ou d'une manière équivalente

$$E[e^{\epsilon x}] < \infty.$$

**Remarque 2.1.1** Si  $F$  est une distribution à queue légère, alors tous ses moments sont finis, i.e

$$\int_{\mathbb{R}} x^k F(x) dF(x) < \infty, \forall k > 0.$$



Nom	Paramètres	Densité
Exponentielle	$\beta > 0$	$f_x(x) = \beta e^{-\beta x}$
Gamma	$\alpha > 0, \beta > 0$	$f_x(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
Weibull	$\beta > 0, r > 0$	$f_x(x) = \beta r x^{r-1} e^{-\beta x^r}$
Mélange Exponentielles	$\beta_i > 0, \sum_{i=1}^n \alpha_i = 1$	$f_x(x) = \sum_{i=1}^n \{\alpha_i \beta_i e^{-\beta_i x}\}$

TAB. 2.1 – Exemple de distribution à queue légère

## 2.2 Distribution 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 v a de fonction de répartition

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

Notons

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

$\bar{F}(x)$  est dite fonction de queue.

### 2.2.1 Définitions

**Définition 2.2.1** Une distribution  $F$  sur  $\mathbb{R}$  est dite à support non borné si  $\bar{F}(x) > 0, \forall x \in \mathbb{R}$ .

**Définition 2.2.2** Une distribution  $F$  (ou  $X$ ) est dite à queue lourde si et seulement si

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

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

on note  $L$  l'ensemble des distribution à queues lourdes.

**Définition 2.2.3** 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 2.2.1** On dit qu'une fonction non négative est à queue lourde si elle est bornée par une fonction exponentielle décroissante.

**Définition 2.2.4** On dit qu'une distribution de  $X$  à la queue lourde si

$$C_k = E \left( \frac{(X - \mu_x)^4}{\sigma_x^4} \right) = \frac{\delta_4}{\delta_2^2} > 3. \quad (2.1)$$

Avec

$\delta_i$  : les moments centré d'ordre  $i$

$C_k$  : coefficient d'aplatissement (Le kurtosis)

Ce qui est équivalent à dire qu'une distribution a une queue lourde si et seulement si son coefficient d'aplatissement  $C_K$ , est supérieur à celui de la loi Normale (pour laquelle  $C_k = 3$ ).

## 2.2.2 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

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

**Proof.**

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

■

**Lemme 2.2.1** ([\[11\]](#))

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 2.2.2**

$$X \in L \iff \lim_{x \rightarrow \infty} \frac{\bar{F}(x - y)}{\bar{F}(x)} = 1, \forall y \geq 0. \quad (2.2)$$

Cette dernière peut s'écrire comme suit

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

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

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

Par conséquent

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

Alors, on peut se réécrire comme suit

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

### 2.2.3 Les sous-classes de distributions à queues lourdes

Il n'y a pas de critère pour classer toutes les distributions selon la queue droite . Cependant, pour certaines classes de distribution, nous obtenons un classement comme celui-ci

- les distributions avec des moments exponentielles inexistantes (**E**),
- les distributions Subexponentielles (**D**),
- les distributions à variations régulières (**C**),
- les distributions avec un comportement de Pareto (**B**)
- les distributions  $\alpha - Stables$  avec  $\alpha < 2$  (**A**).

Toutes ces classes sont emboîtées :

$$A \subset B \subset C \subset D \subset E.$$

voir([\[32\]](#)).

#### 1/Distributions avec des moments exponentiels inexistantes

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  $\bar{F}$  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/Distributions Subexponentielles

La classe **D** des distributions subexponentielles est caractérisée par la définition suivante ([2]) : On dit qu'une distribution est subexponentielle si :

$$\lim_{x \rightarrow \infty} \frac{p(X_1 + X_2 + \dots + X_n)}{p(\max(X_1 + X_2 + \dots + X_n) > x)} = 1, \quad (2.3)$$

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 [2.3] peut être exprimée comme :

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(x)}{e^{-\epsilon x}} = \infty, \forall \epsilon > 0. \quad (2.4)$$

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  $\bar{F}$  décroît plus lentement que n'importe quelle loi exponentielle.

## 3/Distributions à variations régulières 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 [10]).

$$\lim_{x \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = x^{-\alpha}, \quad (2.5)$$

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  $\alpha$  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.

#### 4/Comportement de Pareto avec $\alpha > 0$

La classe **B** est une distribution à comportement de Pareto. Une distribution de Pareto est définie par sa fonction de distribution

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

La fonction de probabilité de dépassement  $\bar{F}$  pour une distribution de classe **B** a la forme  $u^\alpha x^{-\alpha}$ . L'exposant de la valeur extrême peut être caractérisé à partir du moment de la distribution type Pareto. En fait, la fonction de densité de probabilité de la distribution de Pareto est

$$f_{\text{pareto}}(x) = \alpha u^\alpha x^{-\alpha-1},$$

et les moments d'ordre  $k$  sont donnés par :

$$E[X^k] = \alpha u^\alpha \int_u^\infty x^{k-\alpha-1} dx.$$

Par conséquent, seuls les moments d'ordre  $k$ , par exemple  $k < \alpha$ , pour ce type de distribué.

#### 5/Distributions $\alpha$ -Stables

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) [22] dans son Variables indépendantes et identiquement distribuées (iid). manque de formule claire La densité de ces distributions limite leur utilisation. La distribution de cette classe a 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) [21]. 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.

## 2.2.4 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 2.2.1** *Les assertions suivantes sont équivalentes*

1.  $F$  est une distribution à queue lourde.
2. La fonction  $F$  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

$$F[x, x + T] = F[x + T] - F[x].$$

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

$$c = \sup_{\varepsilon' > 0} F[x, x + T]e^{\varepsilon'x},$$

pour  $\varepsilon < \varepsilon'$

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

Cette intégrale est finie  $\forall \varepsilon \in (0, \varepsilon')$ , alors  $F$  n'est pas à queue lourde. pour plus de détails (voir : [11]). ■

### Remarque 2.2.2

$$\int_0^{\infty} x e^{(k-1) \ln x} dx = \infty \quad \text{et} \quad \int_0^{\infty} e^{(k-1) \ln x} dx < \infty, \quad (2.6)$$

(2.6) est un exemple particulier de théorème ci-dessus avec  $g(x) = \ln x$ . Si les moments de  $\varepsilon$  sont infinis, alors on peut prendre la fonction concave  $h(x) = (k-1) \ln x$ .

Ce théorème garantit l'existence d'une fonction concave  $h, \forall g$ .

**Théorème 2.2.2** : Soit  $\varepsilon \geq 0$  une v a de distribution à queue lourde et  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  une fonction concave telle que

$$E(e^{f(\varepsilon)}) = \infty.$$

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

$$h \leq f, E(e^{h(\varepsilon)}) < \infty, E(e^{h(\varepsilon)+g(\varepsilon)}) = \infty.$$

voir ([11]).



## 2.2.5 Exemples de distributions à queues lourdes

### 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

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

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

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

### 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

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

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

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

### Distribution de Cauchy (sur $\mathbb{R}$ )

Sa fonction de densité  $f$  est donnée par

$$f(x) = \frac{k}{\pi((x - \alpha)^2 + k^2)}, k, x, \alpha \in \mathbb{R},$$

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

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

### Distribution de Log-normale (sur $\mathbb{R}^+$ )

Sa fonction de densité  $f$  est donnée par

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

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

### 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 2.2.3** Si  $\alpha = 1$ , on retrouve la distribution exponentielle. Un exemple de dis-

tributions à queues non lourdes est la distribution exponentielle

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

## 2.3 La différence entre distribution de la loi normale et à queue lourde

**Question** : quelle est La différence entre distribution de la loi normale et à queue lourde ?.

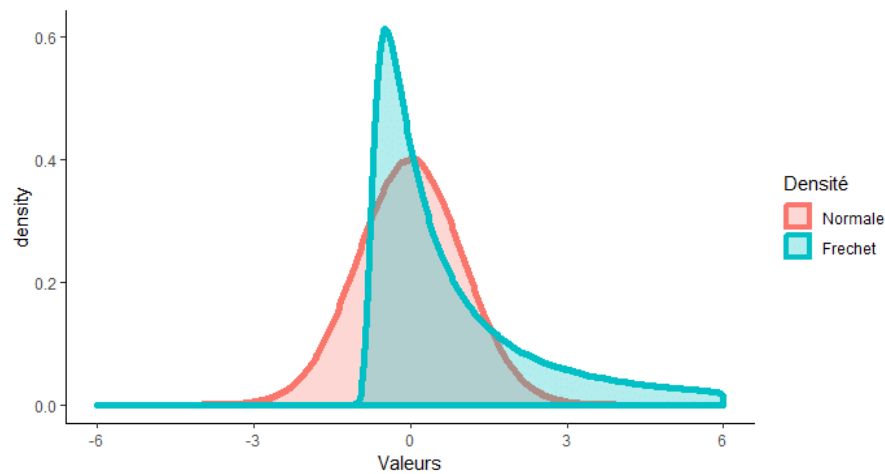


FIG. 2.1 – comparaison d’une distribution normale a une distribution à-queue lourde.

D’après la figure [2.1](#) 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 [\[19\]](#))

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 (2.1) on peut conclure que le 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) ([14]) comme dans (??). 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.

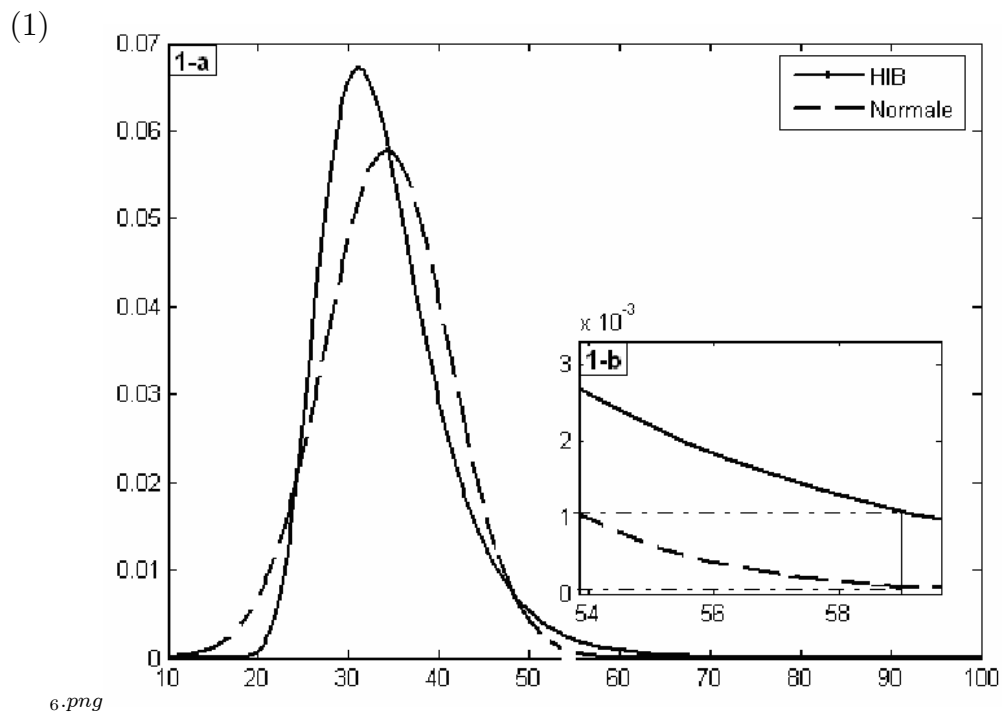


FIG. 2.2 – Illustration de la différence entre la loi normale et une loi à queue lourde (HIB)

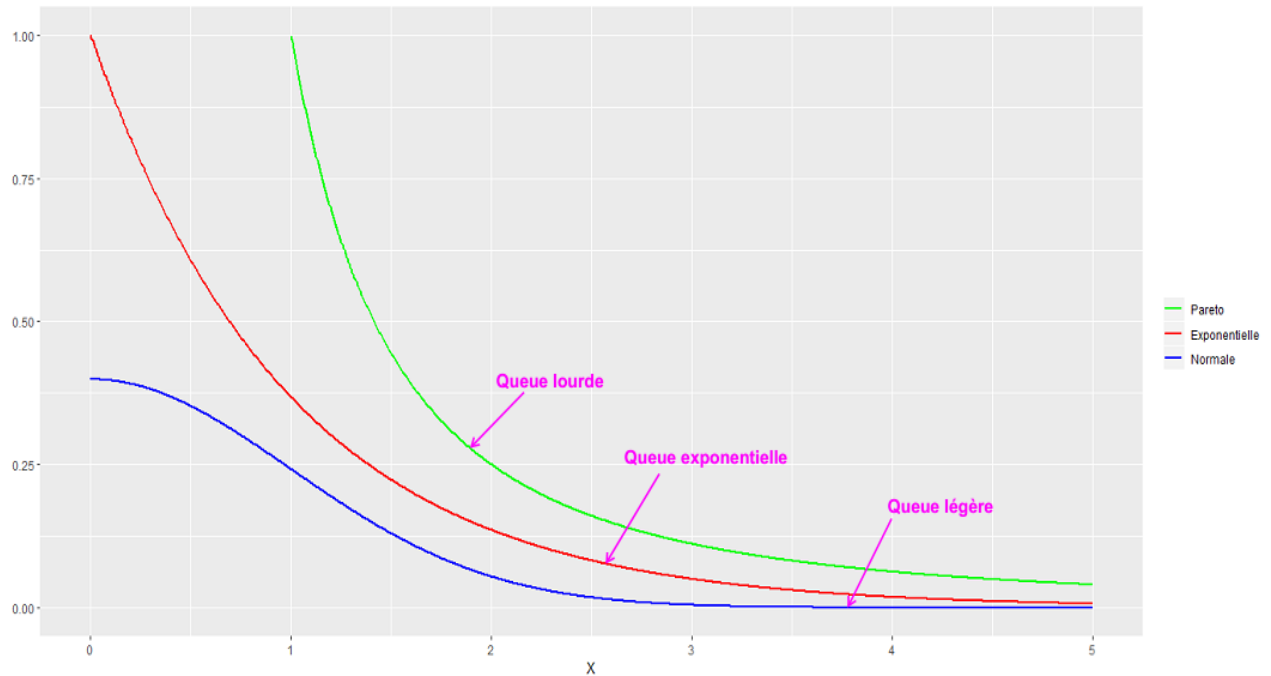


FIG. 2.3 – Comparaison du Comportement de la Queue.

La famille de distribution de queue lourde, elle se caractérise par une décroissance lente vers zéro par rapport à la distribution exponentielle, comme le montre la figure suivante. La figure ci-dessus, elles sont représentées par la courbe verte elle décroît lentement vers zéro par rapport la distribution exponentielle que représentée par la courbe rouge et la distribution à queue légère représentée par la courbe bleue.

## Conclusion

Dans ce mémoire 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.

Dans ce travail une étude approfondie à été faite sur ce type des distributions et leurs propriétés ont été détaillées. Nous avons donné quelque exemple des distributions à queues lourdes.

# Bibliographie

- [1] .Balakrishnan, N and Cohen, A. C. (1991). Order Statistics and Inference :Estimation Methods. Statist. Model. Decis. Sci. Academic Press
- [2] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004) Statistics of Extremes : Theory and Applications.Wiley.
- [3] Chun Su, Qi-he Tang Acta Mathematicae Applicatae Sinica, English Series. Vol.19,No.1 (2003)135 - 137.
- [4] Comets, Francis, and Thierry Meyre. Calcul stochastique et modèles de diffusions-3e éd. : Cours et exercices corrigés. Dunod, 2020.
- [5] De Micheaux, Pierre Lafaye, et al. Le logiciel R : maîtriser le langage, effectuer des analyses (bio) statistiques. Springer, 2014.
- [6] Dress, François. "Les probabilités et la statistique de A à Z, dictionnaire." HAL 2007 (2007).
- [7] Dufour A,Chessel D et Lobry J. (2018) : Lois de ProbabilitÈs. Univ-lyon1
- [8] Dusart P. (2015). Cours de Statistiques inférentielles.
- [9] El Adlouni, S., Ouarda, T. B., Zhang, X., Roy, R., & Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resources Research, 43(3).
- [10] Embrechts, P., Klüppelberg, C. et Mikosch, T. (2003). Modelling Extremal Events for Insurance and Finance, Volume 33 of Applications of Mathematics. Springer.

- [11] Foss, Sergey, Dmitry Korshunov, and Stan Zachary. An introduction to heavy-tailed and subexponential distributions. Vol. 6. New York : Springer, 2011.
- [12] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995). Continuous Univariate Distributions, Volume 2, 2nd Edition. Wiley, New York.
- [13] H. Queffelec, C. Zuily, *Éléments d'Analyse*, Dunod, 2002
- [14] Hubert P. & Bendjoudi H., (1996). Introduction à l'étude des longues séries pluviométriques. In : Journées hydrologiques de l'ORSTOM, , Montpellier, pp. 20.
- [15] [https://lecluseo.scenari-community.org/TES/Ch8%20LoisContinues/co/module\\_Ch\\_8\\_LoisContinues\\_2.html](https://lecluseo.scenari-community.org/TES/Ch8%20LoisContinues/co/module_Ch_8_LoisContinues_2.html).
- [16] <https://medium.com/@pabaldonado/kolmogorov-smirnov-test-may-not-be-doing-what-you-think-when-parameters-are-estimated-from-the-data-2d5c3303a020>
- [17] <https://time.graphics/fr/event/3387277>
- [18] <https://www.semanticscholar.org/paper/Caract%C3%A9risation-des-distributions-%C3%A0-queue-lourde-Adlouni-Bob%C3%A9/35785dfd9667ee9bfc7557e90a0119ff37b82c6e>.
- [19] [https://www.researchgate.net/figure/Comparaison-dune-distribution-normale-a-une-distribution-a-queue-lourde\\_fig10\\_334468946](https://www.researchgate.net/figure/Comparaison-dune-distribution-normale-a-une-distribution-a-queue-lourde_fig10_334468946).
- [20] Gareth W.Peters, Pavel V.Shevchenko (2015), *Advances In Heavy Tailed Risk Modelling*. Wiley Handbooks in Financial Engineering and Econometrics
- [21] Gnedenko B. V., Kolmogorov A. N. (1954), *Limit distributions for sums of random variables*, revised engl. transl., Reading, Mass., Addison-Wesley
- [22] Lévy, P. (1925). *Calcul des Probabilité*, Gauthier-villars, Paris
- [23] Mandelbrot. The variation of certain speculative prices. *The Journal of Business*,36 : 394–419, 1963
- [24] MOKRANI Tassadit (2016). *Distributions à Queues Lourdes et Applications* Mémoire de Master Université Mouloud Mammeri, Tizi-Ouzou



- [25] Nebbar, Belkis.Cherif, Hafsa(2020),Probabilité de Ruine du Processus de Risque avec des Arrivées Non Stationnaires et des Remboursements Sous-Exponentielles Mémoires de Master. Université Blida 1
- [26] Nolan, J. P. (2006). Stable Distributions - Models for Heavy Tailed Data. Boston : Birkhäuser. In progress, Chapter 1 online at [academic2.american.edu/~jpnolan](http://academic2.american.edu/~jpnolan).
- [27] Ouarda, T. B.M.J., Ashkar, F., Bensaïd, E. et I. Hourani (1994). Distributions statistiques utilisées en hydrologie : Transformations et propriétés asymptotiques. Département de mathématique, Université de Moncton, 31 p.
- [28] Rachev, Svetlozar Todorov, ed. Handbook of heavy tailed distributions in finance : Handbooks in finance, Book 1. Elsevier, 2003
- [29] Saporta, G. (2011). Probabilités, analyse des données et statistique. Technip, Paris.
- [30] Segey Foss, Dmitry Kurshonov et Stan Zachary (2013), An introduction to heavy-tailed and subexponential distributions. Springer Series in Operations Research and Financial Engineering.
- [31] Shuyan L. (2014) : Notes de cours Statistique avec le logiciel R. Univ-paris1.Von Mises R. (1954), La distribution de la plus grande de n valeurs. Amer. Math.Soc., 2 : 271-294.
- [32] 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 : Quelques éléments de R

Le logiciel **R** est un langage de programmation

## chapitre 01 :

Pour tester la normalité des variables  $X$  et  $Y$ , on utilise le test de Shapiro-Wilk qui vérifie la normalité avec la  $p$ -value.

```
#charger le package "nortest", library(nortest)
```

```
x<-rnorm(100)
```

```
y<-runif(100, -3, 3)
```

```
shapiro.test(x)
```

Shapiro-Wilk normality test

data : x

W = 0.99381, p-value = 0.9319

```
shapiro.test(y)
```

Shapiro-Wilk normality test

data : y

W = 0.94755, p-value = 0.0005744

**Comentaire** : On observe que :

-Pour la variable  $X$  :  $p$ -value = 0.9319 > 0.05, le test est non significatif. On ne peut pas rejeter l'hypothèse  $H_0$ .

-Pour la variable  $Y$  :  $p$ -value = 0.0005744 < 0.05, le test est significatif. On peut rejeter l'hypothèse  $H_0$ .

# Annexe B : Abréviations et Notations

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

<b>Notation</b>	<b>Signification</b>
$C_K$	Le kurtosis (ou le coefficient d'aplatissement).
<i>i.e.</i>	C'est à dire.Echantillon de taille $n$ .
$E(X)$	Espérance de $X$
$V(x)$	Variance de $X$ .
$\delta$	Variance de $X$ .L'ecart-type.
$v a$	Variable aléatoire.
$F_x$	Fonction de répartition.
$f(x)$	Fonction de densité.
$\varphi_X(t)$	Fonction caractéristique de $X$ .
$g(t)$	Fonction génératrice des moments.
$(X_1, X_2, \dots, X_n)$	Echantillon de taille $n$
$\bar{X}$	Moyenne empirique.
$S^2$	Variance empirique.
$\hat{\mu}$	Estimateur de l'espérance de la loi $\mathcal{N}(\mu, \delta^2)$ .
$L(x_1, x_2, \dots, x_n)$	Fonction de vraisemblance.

$\widehat{\delta^2}$	Estimateur de la variance de la loi $\mathcal{N}(\mu, \delta^2)$
$\theta$	Un paramètre inconnue.
$\widehat{\theta}$	Estimateur de la valeur $\theta$
$H_0$	Hypothèse nulle
$H_1$	Hypothèse alternative
$\mathcal{N}(\mu, \delta^2)$	Loi normale.
$\mathcal{N}(0, 1)$	Loi normale standard.
$\xrightarrow{\mathcal{L}}$	Convergence en loi.
$D_n$	La statistique de Kolmogorov-Smirnov.
$JB$	La statistique de Jarque-Bera.
$W$	La statistique de Shapiro-Wilk.
$TCL$	Téorème central limite.
$iid$	Indépendant identiquement distribuer.
$x \longrightarrow \infty$	$x$ tend vers l'infinie.
$F_n$	Fonction de répartition empirique.
$fdr$	Fonction de répartition.
$\xrightarrow{\mathbb{P}}$	Converge en probabilité.
$\xrightarrow{p.s.}$	Convegence presque sûre.
$\mathbb{R}$	Ensemble des valeurs réelles.
$\overline{F}(x)$	Fonction de queue.
$\limsup_{n \rightarrow \infty} A_n$	Un nombre infini $A_i$ d'événement a lieu.
$\liminf_{n \rightarrow \infty} A_n$	Seul un nombre infini $A_i$ d'événement a lieu.

## Résumé

Ces dernières années, les distributions à queue lourde ont attiré davantage l'attention des chercheurs dans les domaines liés à la théorie des probabilités. Les distributions à queue lourde ont des queues plus lourdes qu'une distribution exponentielle.

De nombreuses distributions utilisées en pratique sont des distributions à queue lourde, certains phénomènes aléatoires ne sont pas modélisés par de telles distributions mais par des distributions à queue lourde. L'utilisation du terme "distribution à queue lourde" varie selon le domaine d'intérêt, mais est souvent comprise comme correspondant à l'absence de sémantique exponentielle (positive). Par conséquent, nous étudions sa définition, ses caractéristiques et ses caractéristiques avec quelques exemples d'application.

**Mots clés :** distributions ; distribution exponentielle; sémantique exponentielle ; queue lourde ; aléatoire.

## Abstract

In recent years, heavy-tailed distributions have attracted more attention from researchers in fields related to probability theory. Heavy-tailed distributions have heavier tails than an exponential distribution.

Many of the distributions used in practice are heavy-tailed distributions, some random phenomena are not modeled by such distributions but by heavy-tailed distributions. The use of the term "heavy-tailed distribution" varies by field of interest, but is often understood to correspond to the absence of exponential (positive) semantics. Therefore, we study its definition, characteristics, and features with some examples of application.

**Keywords:** distributions; exponential distribution; exponential semantics; heavy tail; random.

## ملخص

في السنوات الأخيرة ، جذبت التوزيعات ذات الذيل الثقيل مزيداً من الاهتمام من الباحثين في المجالات المتعلقة بنظرية الاحتمالات. تمتلك التوزيعات ذات الذيل الثقيل ذيولاً أثقل من التوزيع الأسية.

العديد من التوزيعات المستخدمة في الممارسة العملية هي توزيعات ثقيلة الذيل ، بعض الظواهر العشوائية لا يتم نمذجتها من خلال مثل هذه التوزيعات ولكن من خلال توزيعات ثقيلة الذيل. فيختلف استخدام مصطلح "التوزيع الثقيل الذيل" حسب مجال الاهتمام ، ولكن غالباً ما يُفهم أنه يتوافق مع غياب الدلالات الأسية (الاجابية) . لذلك نقوم بدراسة تعريفها وخصائصها وميزاتها مع بعض الأمثلة للتطبيق.

**الكلمات المفتاحية:** توزيعات؛ التوزيع الاسي؛ الدلالات الاسية؛؛ ذيل ثقيل؛ عشوائية.