

# وزارة التعليم العالي والبحث العلمي

Université Badji Mokhtar  
Annaba

Badji Mokhtar University -  
Annaba



جامعة باجي مختار  
عنابة

Faculté des Sciences  
Département de Mathématiques

## THESE

Présentée en vue de l'obtention du diplôme de  
Doctorat en Mathématiques  
Option : Statistique

### *NOUVEAU MODELE DE SURVIE*

Par:

**Messaadia Hamouda**

Sous la direction de

**ZEGHDOUDI Halim**

MCA

U.B.M. Annaba

Et

**Chadli Assia**

Prof.

U.B.M. Annaba

Devant le jury

**PRESIDENT :**

**Hadji Med Lakhder**

MCA

U.B.M. Annaba

**EXAMINATEUR :**

**Brahimi Brahim**

Prof.

U. Biskra

**EXAMINATEUR :**

**Badreddine Mansouri**

MCA

U. Biskra

**EXAMINATEUR :**

**Tamer Lazhar**

MCA

U. Biskra

Année : 2018

## Remerciements

Mes sincères remerciements à Dieu le tout puissant, le miséricordieux qui m'a donnée la force, la volonté et le courage afin d'élaborer ce travail.

Je tiens également à exprimer ma reconnaissance et ma gratitude à mon mes encadreurs **Dr.Zeghdoudi Halim et Pr.Chadli Assia** qui a bien voulu accepter de m'accorder ce privilège ; et d'avoir consacré beaucoup de temps à me « former ».

Je salue en lui ses grandes compétences, sa qualité professionnelle et surtout sa gentillesse et son soutien dont il m'a gratifiée tout au long de ce travail.

Je remercie le **Dr. Hadji Med Lakhder** qui a accepté d'être le président de ce jury.

Je remercie vivement **Pr. Brahim Brahim, Dr. BadreddineMansouri, et Dr.Tamer Lazhar** de l'université de Biskra, pour l'honneur d'avoir accepté de faire partie du jury.

Je m'estime très honorée par l'intérêt qu'ils ont bien voulu accorder à mon travail et leur saurais gré pour toutes remarques, qui va m'aider à voir les ponts qui pouvaient exister entre mon travail et d'autres perspectives mathématiques.

Enfin je remercie tous ceux qui m'ont aidé de près ou de loin.

## **Dédicace**

Je dédie ce travail de longues années d'étude à

La lumière de ma vie, au cœur le plus tendre et le plus doux, à celle qui s'est  
tellement sacrifiée pour me voir toujours meilleure : ma très chère mère

A l'être le plus cher à mon cœur, à celui qui m'a toujours guidée par ses conseils  
et qui m'a encouragée à poursuivre mes études : Mon père ullah yerhmou

bonnes rails

A Mes chers frères et sœurs

**« Dr. Messaadia Hamouda »**

# Table des matières

Table des figures	3
<b>1 Généralités et quelques notions probabiliste</b>	<b>10</b>
1.1 Variables aléatoires continues . . . . .	12
1.2 Les différents types d'estimations . . . . .	16
1.2.1 L'estimation paramétrique . . . . .	16
1.2.2 L'estimation non paramétrique . . . . .	18
1.3 Ordre stochastique . . . . .	25
1.4 Théorie des valeurs extrêmes . . . . .	27
1.4.1 Théorie des valeurs extrêmes maximales . . . . .	27
1.4.2 Théorie des valeurs extrêmes minimales . . . . .	29
1.5 Courbe de <b>Lorenz</b> . . . . .	30
1.6 Entropie . . . . .	32
1.6.1 Entropie <b>Shannon</b> . . . . .	32
1.6.2 Entropie de <b>Rényi</b> . . . . .	32
<b>2 Distribution de Lindley et ses applications</b>	<b>34</b>
2.1 Distribution de <b>Lindley</b> . . . . .	34
2.2 Moments et mesures connexes . . . . .	36
2.3 Fonction de hasard et fonction de survie . . . . .	38
2.4 Ordre stochastique . . . . .	42
2.5 Courbe de Lorenz . . . . .	42
2.6 Statistiques d'ordre extrêmes . . . . .	43
2.7 Fonction Quantile de la distribution <b>Lindley</b> . . . . .	44
2.8 Estimation . . . . .	45
2.8.1 Estimation par la méthode des moments (MoM) . . . . .	45
2.8.2 Estimation par la méthode du maximum de vraisemblance . . . . .	47
2.9 Simulation . . . . .	47
2.10 Distributions à deux paramètres de <b>Lindley</b> . . . . .	49
2.10.1 Distribution Quasi <b>Lindley</b> . . . . .	50

2.10.2	Distribution de Pseudo <b>Lindley</b> ( <i>PsLD</i> ) . . . . .	51
2.10.3	Moments et Mesures connexes . . . . .	55
2.10.4	Fonction de hasard et Fonction de survie . . . . .	56
2.10.5	Ordres stochastiques . . . . .	57
2.10.6	Courbe de Lorenz . . . . .	57
2.10.7	Statistiques d'ordres extrêmes . . . . .	57
2.10.8	Fonction Quantile de la distribution de Pseudo <b>Lindley</b> . . .	58
2.10.9	Estimation . . . . .	59
2.10.10	Distribution de Gamma <b>Lindley</b> ( <i>Gal</i> ) . . . . .	60
2.10.11	Moments et mesures connexes . . . . .	61
2.10.12	Ordres Stochastiques . . . . .	62
2.10.13	Courbe de Lorenz . . . . .	63
2.10.14	Statistiques d'ordres extrêmes . . . . .	63
2.10.15	Fonction Quantile de la distribution de Gamma <b>Lindley</b> . . .	64
2.10.16	Estimation . . . . .	65
2.10.17	Autres distribution de plusieurs paramètres . . . . .	66
<b>3</b>	<b>Distribution de Zeghdoudi</b> . . . . .	<b>70</b>
3.1	Distribution de <b>Zeghdoudi</b> (ZD) et immédiate propriétés . . . . .	70
3.1.1	La fonction de Survie et taux de hasard . . . . .	72
3.2	Moments et mesures connexes . . . . .	72
3.3	Estimation du paramètre . . . . .	74
3.3.1	Ordres stochastiques . . . . .	76
3.4	Ecart moyen . . . . .	77
3.5	Statistiques de commande extrêmes . . . . .	77
3.6	Simulation et Ajustement . . . . .	78
3.6.1	Étude de simulation . . . . .	78
3.6.2	Applications . . . . .	81
3.7	Annexe . . . . .	84
	<b>Bibliographie</b> . . . . .	<b>86</b>

## ملخص

في هذه المذكرة ، نعرض توزيعا جديدا مستمرا بما يسمى بتوزيع زغدودي (ZD) ع يتميز هذا التوزيع ببعض الخواص كباقي التوزيعات الاخرى :تابع الصدفة , الفئات العشوائية , احصاء الفئة , تقدير الوسائط بطريقة المعقولية العظمى و طريقة العزوم و نختتم أطروحتنا بإعطاء بعض الامثلة و معطيات للمقارنة بين توزيع (ZD) و (LD) و توزيعات اخرى.

الكلمات المفتاحية: توزيع زغدودي، توزيع لندلي، تقدير، محاكاة

## Résumé

Dans cette thèse nous introduisons une nouvelle distribution nommée Zeghdoudi (ZD). Quelques propriétés de cette distribution ont été données à savoir : la fonction de hasard, ordres stochastiques, statistiques d'ordre, estimation des paramètres avec la méthode du maximum de vraisemblance et la méthode des moments. . Enfin, nous donnons des exemples illustratifs et une étude de simulation pour comparer ZD avec LD et d'autres distributions.

**Mots clés** : distribution de Lindley, distribution de Zeghdoudi, estimation, simulation

## **Abstract**

This thesis proposes a new distribution, called as Zeghdoudi distribution (ZD). Various statistical properties like stochastic ordering, moment method, maximum likelihood estimation, mean deviations and limiting distribution of extreme order statistics is established. An application of the model to a real data set is presented finally and compared with the  $t$  attained by some other well-known one and two parameters distributions.

**Keywords:** Lindley distribution, Zeghdoudi distribution, estimation, simulation

## Table des figures

2.1	Présentation graphique de la fonction de densité pour quelques valeurs de $\theta$ , noir( $\theta = 0.5$ ); rouge ( $\theta = 1$ ); bleu ( $\theta = 2$ ). . . . .	21
2.2	Présentation graphique du coefficient de variation $\gamma$ (noir), le coefficient de dissymétrie $\sqrt{\beta_1}$ ( <i>rouge</i> ) et le coefficient d'aplatissement $\beta_2$ ( <i>bleu</i> ) . . . . .	22
2.3	Présentation graphique de la fonction de taux de hasard pour quelques valeurs de $\theta$ , noir( $\theta = 0.5$ ); rouge ( $\theta = 1$ ); bleu ( $\theta = 2$ ). . . . .	24
2.4	Présentation graphique de la fonction de survie pour quelques valeurs de $\theta$ , noir( $\theta = 0.5$ ); rouge ( $\theta = 1$ ); bleu ( $\theta = 2$ ). . . . .	24
2.5	représentation graphique la fonction de densité de PsLD pour quelques valeurs de $(\theta, \beta)$ .noir(0.5,1.5);rouge(0.25,2);bleu(1,3);vert(3,3),jaune(0.1,4);gris(2,8).	36
2.6	représentation graphique la répartition de densité de PsLD pour quelques valeurs de $(\theta, \beta)$ . noir(0.5,1.5);rouge(0.25,2);bleu(1,3);vert(3,3),jaune(0.1,4);gris(2,8).	36

# Introduction

Nous allons étudier dans ce chapitre la distribution d'un seul paramètre de **Lindley** (voir **Lindley** 1958) qui définie sous desous par la fonction de densité :

$$f(x, \theta) = \begin{cases} \frac{\theta^2(1+x)e^{-\theta x}}{1+\theta} & \text{si } x, \theta > 0 \\ 0 & \text{ailleurs} \end{cases} \quad (1.1)$$

**Sankaran** (1970) utilise (1) la distribution mixte de **Poisson** qui s'appelle la distribution de **Poisson- Lindley** sachant que la fonction de densité est :

$$f_{PLD}(x, \theta) = \frac{\theta^2 (\theta + x + 2) e^{-\theta x}}{(1 + \theta)^x}, \quad x = 0, 1, \dots \quad (1.2)$$

où  $\theta$  est un paramètre positive.

**Asgharzadeh et al.** (2013), **Elbatal et al.** (2013), **Bakouch et al.**(2012), **Ghitany et al.** (2008a) et (2008b) ont redécouvert et étudié la nouvelle distribution limitée à (1), ce qu'ils ont dérivé est connu sous le nom de distributions **Poisson-Lindley** et **Pareto Poisson-Lindley** tronquées à zéro.

Récemment, **Zeghdoudi et Nedjar** (2016a), (2016b), (2016c) ont introduit une nouvelle distribution, nommée distribution de gamma **Lindley** et pseudo **Lindley**, basée sur des mélanges de gamma  $(2, \theta)$  et des distributions de **Lindley** à un paramètre pour la première distribution et sur des mélanges de distributions exponentielles de gamma  $(2, \theta)$  et d'un paramètre pour la deuxième distribution. L'idée d'utiliser un mélange de deux distributions connues pour générer une nouvelle distribution n'est

pas nouvelle. Par exemple, **Shanker, Sharma et Shanker** (2013) ont utilisé un mélange d'exponentielle ( $\theta$ ) et de gamma ( $2, \theta$ ) pour créer une distribution de **Lindley** à deux paramètres. Pour un autre exemple, **Zakerzadeh et Dolati** (2009) ont utilisé gamma ( $\alpha, \theta$ ) et gamma ( $\alpha + 1, \theta$ ) pour créer une distribution de **Lindley** généralisée. L'idée de ce travail est basée sur des mélanges des distributions gamma ( $2, \theta$ ) et Gamma ( $3, \theta$ ).

Ce projet de recherche est structuré de la manière suivante :

Le premier chapitre est consacré aux rappels des certaines définitions et certains résultats qui nous utilisons par la suite à savoir : généralités et quelques distributions de probabilités, fonction W de **Lambert**, fonction quantile, statistiques d'ordres extrêmes, estimation MM et MV, entropie.

Le chapitre II collecte quelques distributions à un et deux paramètres (**Lindley**, pseudo **Lindley**, gamma **Lindley**) dont on donne quelques propriétés à savoir : la fonction quantile, courbe de Lorenz, méthode des moments, estimation du maximum de vraisemblance et la distribution de limitation des statistiques d'ordre extrême. Plusieurs simulations sont établies pour examiner le biais et l'erreur quadratique moyenne des estimateurs des paramètres obtenus par la méthode du maximum de vraisemblance et leur application dans l'analyse de survie.

Enfin, le dernier chapitre contient l'essentiel des travaux regroupés dans les articles [41] en nous inspirant des travaux originaux de **Lindley** (1958), et **Zeghdoudi and Nedjar** (2016) en adaptant certains résultats. Par ailleurs, une simulation de biais et

l'erreur quadratique moyenne des estimateurs obtenus par la méthode du maximum de vraisemblance. Enfin, une comparaison des distributions est obtenue.

# Chapitre 1

## Généralités et quelques notions probabiliste

Dans cette introduction nous présentons les caractéristiques essentielles des données à analyser ainsi que les outils techniques permettant de décrire leur distribution.

Les temps de survie mesurée à partir d'une origine appropriée ont deux caractéristiques. La première est qu'ils sont non négatifs et tels qu'une hypothèse de normalité n'est généralement pas raisonnable en raison d'une asymétrie prononcée. La seconde est structurelle et tient au fait que pour certains individus l'évènement étudié ne se produit pas pendant la période d'observation et en conséquence certaines données sont censurées. Cette censure à droite est la plus courante mais n'est pas la seule censure que l'on peut rencontrer avec des données de survie.

Considérons une étude relative à la durée de survie de patients soumis à un trai-

tement particulier.

L'évènement d'intérêt est la mort de la personne.

Pour des données continues la durée de vie  $T$ , c'est-à-dire la durée observée d'un individu dans un état initial, est une variable aléatoire non positive représente la durées de survie d'un individu dans population homogène, il y'a trois fonctions particuliers très importants : la fonction de survie, la densité et la fonction de hazard en deux catégorie (models) discrete et continu distribution.

On considère une population de patients dont la durée de survie est décrite par une variable aléatoire  $T$ . Un échantillon de  $n$  individus est tiré aléatoirement de cette population. Les variables aléatoires  $T_i$  (avec  $i \in [1, n]$ ) décrivant la durée de survie des  $n$  patients sont supposées indépendantes et identiquement distribuées (selon la loi de  $T$ ). La durée de vie effective du patient d'ordre  $i$  est notée  $t_i$ . Elle est censurée si  $t_i > s_i$ . Posons

$$\tau_i = \min\{t_i, s_i\} \tag{1.3}$$

et

$$\delta_i = \mathbf{1}_{\{t_i \leq s_i\}} = \begin{cases} 1 & \text{si } t_i \leq s_i \text{ (détection)} \\ 0 & \text{si } t_i > s_i \text{ (donnée censurée),.} \end{cases} \tag{1.4}$$

Pour chaque patient, on connaît  $\tau_i$  et  $\delta_i$ . On connaît aussi le seuil  $s_i$  pour les données non censurées, mais cette information ne sera pas utilisée ici (elle pourrait l'être en revanche pour des simulations).

La probabilité de survie sera décrite par des fonctions.

## 1.1 Variables aléatoires continues

Supposons que la durée de survie  $T$  soit une variable positive ou nulle, et absolument continue, alors sa loi de probabilité peut être définie par l'une des cinq fonctions équivalentes suivantes (chacune des fonctions ci-dessous peut être obtenue à partir de l'une des autres fonctions).

### Fonction de densité

C'est la fonction positive  $f$  telle que :

$$\int_0^{\infty} f(x) dx = 1 \quad (1.5)$$

### Fonction de répartition

C'est la fonction  $F$  telle que :

$$F(x) = \mathbf{P}(T \leq x) = \int_0^x f(t) dt \quad (1.6)$$

### Fonction de survie

La fonction de survie est définie sur  $[0, +\infty[$  comme :

$$S(t) = \mathbf{P}(T > t), t \geq 0 \quad (1.7)$$

Clair que  $S$  décroissante, continue à droite avec  $S(0) = 1$  et  $\lim_{t \rightarrow \infty} S(t) = 0$ .

#### Remarque 1.1

On peut également pour préciser cette distribution recourir à la fonction de densité :

$$f(t) = \frac{dF(t)}{dt} = -\frac{dS(t)}{dt} \quad (1.8)$$

Un concept important dans ces analyses est celui de risque. Considérons la quantité

$$\mathbf{P}[t \leq T < t + \Delta t | T \geq t] \quad (1.9)$$

C'est la probabilité de survenue de l'évènement durant l'intervalle de temps  $[t, t + \Delta t[$  sachant qu'il ne s'était pas réalisé avant  $t$ . Naturellement, si l'intervalle de temps en question tend vers zéro alors avec une aléatoire continue la probabilité en question tend aussi vers zéro. Les choses changent si on la norme par la durée de l'intervalle lui-même :

$$\frac{\mathbf{P}[t \leq T < t + \Delta t | T \geq t]}{\Delta t} \quad (1.10)$$

On passe alors à une évaluation du risque de connaître l'évènement durant l'intervalle de temps considéré. La quantité obtenue mesure en effet le nombre moyen d'évènements que connaîtrait l'individu concerné au cours d'une unité de temps choisie (mois, année par exemple) si les conditions prévalant durant l'intervalle de temps considéré restaient inchangées tout au long de l'unité de temps choisie et pas seulement sur l'intervalle.

### Exemple 1.1

Supposons l'unité est l'année, si la durée de l'intervalle  $\Delta t$  correspond à un deux mois et si la probabilité de connaître l'évènement au cours de ce mois est de 30% alors l'expression ci-dessus vaut :

$$\frac{\mathbf{P}[t \leq T < t + \Delta t | T \geq t]}{\Delta t} = \frac{3}{10} \frac{12}{2} = 4.5$$

ce qui signifie qu'en moyenne si les conditions observées pendant le deux mois en ques-

tion se maintenaient toute l'année, l'individu connaîtrait en moyenne 4.5 évènements par an ce qui est bien l'évaluation d'un risque.

### Risque instantané $h$ (ou taux de hasard)

On va définir la fonction de risque (hasard fonction) qui apparaît comme une mesure du risque instantané (attention ainsi que nous l'avons vu ce n'est pas une probabilité. Elle peut en particulier prendre des valeurs supérieures à l'unité)

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}[t \leq T < t + \Delta t | T \geq t]}{\Delta t} \quad (1.11)$$

Cette fonction est liée aux précédents objets puisqu'en effet avec le théorème des probabilités conditionnelles il vient :

$$h(t) = \frac{f(t)}{S(t)} \quad (1.12)$$

### Taux de hasard cumulé $H$

Il est encore possible de définir le risque cumulé  $H(t)$  selon :

$$H(t) = \int_0^t h(s) ds \quad (1.13)$$

Avec l'égalité suivante entre fonction de survie et fonction de risque cumulé :

$$H(t) = -\ln S(t) \quad (1.14)$$

### Remarque 1.2

Toutes ces fonctions sont donc liées entre elles : la connaissance de  $S(t)$  permet celle de  $f(t)$  via (1.8) et donc celles de  $h(t)$  par (1.12) et  $H(t)$  par (1.14). De même,

la connaissance de  $h(t)$  permet celle de  $H(t)$  donc de  $S(t)$  et finalement de  $f(t)$ . En d'autres termes, si on se donne une seule de ces fonctions, alors les autres sont dans le même temps également définies. En particulier, un choix de spécification sur la fonction de risque instantané implique la sélection d'une certaine distribution des données de survie.

### Moyenne et variance de la durée de survie

#### Lemme 1.1

Le temps moyen de survie  $\mathbf{E}(T)$  et la variance de la durée de survie  $V(T)$  sont définis par les quantités suivantes :

$$\mathbf{E}(T) = \int_0^{\infty} S(t) dt \quad (1.15)$$

$$\mathbf{Var}(T) = 2 \int_0^{\infty} tS(t) dt - \left( \int_0^{\infty} S(t) dt \right)^2. \quad (1.16)$$

Ainsi on peut déduire l'espérance et la variance à partir de n'importe laquelle des fonctions  $F, S, f, H, h$  (mais pas l'inverse).

#### Démonstration

Nous savons que si le variable aléatoire  $T$  positif alors :

$$\mathbf{E}(T) = \int_0^{\infty} (1 - F(t)) dt = \int_0^{\infty} S(t) dt$$

Pour le variance  $V$  il faut calculer le terme  $\int_0^{\infty} t^2 f(t) dt$  par l'intégration par partie.

## 1.2 Les différents types d'estimations

### 1.2.1 L'estimation paramétrique

Une des caractéristiques des données de survie est l'existence d'observations incomplètes. En effet, les données sont souvent recueillies partiellement, notamment, à cause des processus de censure et de troncature.

Les données censurées ou tronquées proviennent du fait qu'on n'a pas accès à toute l'information : au lieu d'observer des réalisations indépendantes et identiquement distribuées (i.i.d.) de durées  $X$ , on observe la réalisation de la variable  $X$  soumise à diverses perturbations, indépendantes ou non du phénomène étudié.

#### Méthode du maximum de vraisemblance

Nous supposons ici la forme de la loi connue. Notons  $\vec{\theta} = (\theta_1, \dots, \theta_k)$  les paramètres de la loi et  $S(t; \vec{\theta})$  la probabilité  $\mathbf{P}(T > t | \vec{\theta})$ .

En l'absence de données censurées (et en négligeant les incertitudes observationnelles), la vraisemblance, c.-à-d. la probabilité que la situation observée se produise connaissant  $\vec{\theta}$ , donc d'obtenir simultanément  $T_i$  dans  $[t_i, t_i + \Delta t_i[$  pour tout  $i \in [1, n]$ , est la quantité

$$\mathcal{L}(\vec{\theta}) = \prod_{i=1}^n P[t_i \leq T_i < t_i + \Delta t_i | \vec{\theta}] = \prod_{i=1}^n f(t_i; \vec{\theta}) dt_i \quad (1.17)$$

car les  $T_i$  sont des variables aléatoires indépendantes.

La méthode du maximum de vraisemblance consiste à adopter comme estimateur

$\widehat{\vec{\theta}}_n$  le vecteur  $\vec{\theta}$  maximisant la vraisemblance ou, ce qui est plus commode numériquement, minimisant l'application  $\mathcal{L} : \Theta \subset \mathbb{R}^k \rightarrow \mathbb{R}$  où  $\Theta$  désigne l'espace des paramètres.

### Exemple 1.2

Dans le cas où  $f(t)$  est une loi normale de moyenne  $\mu = \theta_1$  et d'écart-type (c.-à-d. de dispersion intrinsèque)  $\sigma = \theta_2$

$$\mathcal{L}(\vec{\theta} = (\mu, \sigma)) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n \left(\frac{t_i - \mu}{\sigma}\right)^2}$$

alors l'estimateur est  $\widehat{\vec{\theta}}_n = (\theta_1, \theta_2) = (\bar{T}, S)$ .

### Méthode des moments

Dans cette section,  $T$  est le vecteur formé par un n-échantillon  $T_1, \dots, T_n$ . Les  $T_i$  sont à valeurs dans un ensemble  $\Omega$ . Soit  $f = (f_1, \dots, f_k)$  une application de  $\Omega$  dans  $\mathbb{R}^k$  telle que l'application

$$\Phi : \Theta \subset \mathbb{R}^k \rightarrow \mathbb{R} \tag{1.18}$$

$$\theta \rightarrow \mathbf{E}_\theta[f(X_i)]$$

soit injective. On définit l'estimateur  $\theta_n$  comme la solution dans  $\Theta$  (quand elle existe) de l'équation :

$$\Phi(\theta) = \frac{1}{n} \sum_{i=1}^n f(X_i) \tag{1.19}$$

Souvent, lorsque  $\Omega \subset R$ , la fonction on prend  $f_i(x) = x^i$  et  $\Phi$  correspond donc au  $i$ ème moment de la variables  $X_i$  sous  $P_\theta$ . Ce choix justifie le nom donné à la méthode.

**Exemple 1.3**

Dans le cas où  $k = 1$  et  $X$  est une v.a suivant la loi exponentiel de paramètre  $\theta$  et  $f(x) = e^x$  (*injective*). Comme pour tout  $\theta > 1$ ,  $\mathbf{E}_\theta[f(X_1)] = \Phi(\theta) = \frac{\theta}{1-\theta}$ .

L'estimateur obtenu par la méthode des moments est :

$$\theta_n = \sum_{i=1}^n e^{X_i} \left( n + \sum_{i=1}^n e^{X_i} \right)^{-1}$$

**Exemple 1.4**

Dans le cas où  $k = 2$  et  $X$  est une v.a suivant la loi normal de paramètre  $\theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$  pour tout  $\theta$ , et  $\mathbf{E}_\theta(X_1) = \mu$  et  $\mathbf{E}_\theta(X_1^2) = \mu^2 + \sigma$ , On peut donc prendre, par exemple,  $f_1(x) = x$  et  $f_2(x) = x^2$  ce qui donne  $\Phi(\mu, \sigma) = (\mu, \mu^2 + \sigma)$ .

L'estimateur obtenus par la méthode des moments vérifie

$$\hat{\mu} = \overline{X_n} \text{ et } \hat{\sigma} = \sum_{i=1}^n \frac{X_i^2}{n} - \overline{X_n}^2$$

c'est a dire

$$\theta_n = \left( \overline{X_n}, \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X_n})^2 \right)$$

L'estimateur est consistant mais l'estimateur de la variance est biaisé.

$$\mathbf{E}(\theta_n) = \left( \mu, \frac{n-1}{n} \sigma^2 \right)$$

**1.2.2 L'estimation non paramétrique****L'estimateur a noyau bidimensionnel**

Soit le couple aléatoire  $(X, Y) \sim F$  avec  $F(x, y) = \mathbf{P}(X \leq x, Y \leq y)$  la fonction de répartition de  $(X, Y)$ . Soit  $(X_1, Y_1), \dots, (X_n, Y_n)$  un échantillon id de  $F$  (id= iden-

tiquement distribuées).

Supposons que  $F$  soit complètement inconnue. Comment estimer  $F$ , en se basant sur les observations  $(X_1, Y_1), \dots, (X_n, Y_n)$  ?

Un bon estimateur pour  $F$  est la fonction de répartition empirique, notée  $F_n$  et définie par :

$$F_n(x, y) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{\{X_i \leq x, Y_i \leq y\}} \quad (1.20)$$

### Construction de l'estimateur

Rappelons que la densité de probabilité  $f$  est égale à la dérivée partielle par rapport à  $x$  et  $y$  de la fonction de répartition  $F$  (si cette dérivée existe).

#### Lemme 1.2

l'estimation non paramétrique de la densité de probabilité  $f$  est donnée par :

$$\hat{f}(x, y) = \frac{1}{4nh^2} \sum_{i=1}^n \mathbf{I}_{\{|\frac{X_i - x}{h}| \leq 1, |\frac{Y_i - y}{h}| \leq 1\}} \quad (1.21)$$

où  $\mathbf{I}_A$  désigne la fonction indicatrice sur l'ensemble  $A$ .

#### Preuve

l'estimateur empirique  $\hat{f}$  converge en moyen vers  $f$ , en effet :

$$\begin{aligned} \mathbf{E} \left( \hat{f}(x, y) \right) &= \frac{1}{4h^2} \mathbf{E} \left( \mathbf{I}_{\{|\frac{X-x}{h}| \leq 1, |\frac{Y-y}{h}| \leq 1\}} \right) \\ &= \frac{1}{4h^2} \mathbf{P} (|X - x| \leq h, |Y - y| \leq h) \\ &= \frac{1}{4h^2} \mathbf{P} (x - h \leq X \leq x + h, y - h \leq Y \leq y + h) \end{aligned}$$

et comme

$$\mathbf{P} (X \in [x_1, x_2], Y \in [y_1, y_2]) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$

Alors :

$$\mathbf{E} \left( \hat{f}(x, y) \right) = \frac{1}{4h^2} F(x-h, y-h) + F(x+h, y+h) - F(x-h, y+h) - F(x+h, y-h)$$

et nous savons :

$$\frac{\partial F(x, y)}{\partial x} = \lim_{h \rightarrow 0} \frac{F(x+h, y) - F(x-h, y)}{2h}$$

alors :

$$\begin{aligned} \frac{\partial^2 F(x, y)}{\partial y \partial x} &= \lim_{h \rightarrow 0} \frac{\frac{\partial F(x, y+h)}{\partial x} - \frac{\partial F(x, y-h)}{\partial x}}{2h} \\ &= \lim_{h \rightarrow 0} \frac{F(x-h, y-h) + F(x+h, y+h) - F(x-h, y+h) - F(x+h, y-h)}{4h^2} \end{aligned}$$

donc

$$E \left( \hat{f}(x, y) \right) \simeq f(x, y)$$

Notons que cet estimateur peut encore s'écrire comme

$$\hat{f}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n \mathbf{K} \left( \frac{X_i - x}{h}, \frac{Y_i - y}{h} \right) \quad (1.22)$$

Où

$$\mathbf{K}(u, v) = \begin{cases} \frac{1}{4} \text{ si } (u, v) \in [-1, 1]^2 \\ 0 \text{ si non} \end{cases} \quad (1.23)$$

### Conséquence

1-Le paramètre de lissage (fenêtre)  $h$  dépend de la taille de l'échantillon  $n$ .

2-Nous savons que

$$nF_n \rightsquigarrow \mathbf{Bin}(n, F)$$

3-On note la fonction de répartition théorique de  $\mathbf{P}(|X_i - x| \leq h, |Y_i - y| \leq h)$  par  $\Phi_h(x, y)$ . On a :

$$\mathbf{Var}(\hat{f}(x, y)) = \frac{1}{16nh_n^2} \Phi_h(x, y) (1 - \Phi_h(x, y)) \quad (1.24)$$

4-Si  $\lim_{n \rightarrow 0} h_n = 0$ . alors

$$\mathbf{E}(\hat{f}(x, y)) \rightarrow f(x, y) \quad \text{et} \quad nh_n^2 \mathbf{Var}(\hat{f}(x, y)) \rightarrow \frac{1}{4} f(x, y) \quad (1.25)$$

5-Le risque quadratique moyen de l'estimateur  $\hat{f}(x, y)$  de  $f(x, y)$  est donnée par

$$\mathbf{E}(\hat{f}(x, y) - f(x, y))^2 = \mathbf{Var}(\hat{f}(x, y)) + Bias^2(\hat{f}(x, y)) \quad (1.26)$$

Donc si  $h_n \rightarrow 0, \sqrt{nh_n} \rightarrow 0$ , quand  $n \rightarrow \infty$ , on a que

$$\mathbf{E}(\hat{f}(x, y) - f(x, y))^2 \rightarrow 0 \quad (1.27)$$

pour tout point  $(x, y)$  l'estimateur simple  $\hat{f}(x, y)$  est alors un estimateur consistant de  $f(x, y)$ .

Rappelons l'estimateur simple

$$\hat{f}(x, y) = \frac{1}{nh_n^2} \sum_{i=1}^n \mathbf{W}\left(\frac{X_i - x}{h}, \frac{Y_i - y}{h}\right)$$

avec

$$\mathbf{W}(a, \beta) = \begin{cases} \frac{1}{4} & \text{si } (a, \beta) \in [-1, 1]^2 \\ 0 & \text{si non} \end{cases}$$

la densité de probabilité uniforme sur le carré  $[-1, 1]^2$ , cet estimateur peut être généralisé en remplaçant la fonction de poids  $\mathbf{W}(\cdot, \cdot)$  (la densité de probabilité uniforme) par une fonction de poids plus générale  $\mathbf{K}$  (par exemple une densité de probabilité quelconque). Ceci résulte en l'estimateur

$$\hat{f}(x, y) = \frac{1}{nh_n^2} \sum_{i=1}^n \mathbf{K} \left( \frac{X_i - x}{h}, \frac{Y_i - y}{h} \right)$$

$\mathbf{K}$  : le noyau.

$h$  : la fenêtre ou le paramètre de lissage.

Souvent on prend  $\mathbf{K}$  une densité de probabilité symétrique c-à-d  $\mathbf{K}(z, t) = \mathbf{K}(t, z)$ .

### Expression du biais et de la variance

Considérons l'estimateur à noyau

$$\hat{f}(x, y) = \frac{1}{nh_n^2} \sum_{i=1}^n \mathbf{K} \left( \frac{X_i - x}{h}, \frac{Y_i - y}{h} \right)$$

pour calculer le biais de l'estimateur à noyau, remarquons d'abord que

$$\begin{aligned} \mathbf{E}(\hat{f}(x, y)) &= \frac{1}{h^2} \mathbf{E} \left( \mathbf{K} \left( \frac{X_i - x}{h}, \frac{Y_i - y}{h} \right) \right) \\ &= \frac{1}{h^2} \int_{\mathbb{R}^2} \mathbf{K} \left( \frac{\alpha - x}{h}, \frac{\beta - y}{h} \right) f(\alpha, \beta) d\alpha d\beta \end{aligned} \quad (1.28)$$

car les  $(X_i, Y_i)$  sont identiquement distribuées.

Dés lors, nous avons :

$$\mathbf{E} \left( \hat{f}(x, y) - f(x, y) \right) = \frac{1}{h^2} \left( \int_{\mathbb{R}^2} \left[ \mathbf{K} \left( \frac{\alpha - x}{h}, \frac{\beta - y}{h} \right) f(\alpha, \beta) - h^2 \mathbf{K}(\alpha, \beta) f(x, y) \right] d\alpha d\beta \right)$$

Pour la variance :

$$\begin{aligned}
\mathbf{Var} \left( \hat{f}(x, y) \right) &= \mathbf{E} \left( \hat{f}(x, y)^2 \right) - \mathbf{E}^2 \left( \hat{f}(x, y) \right) \\
&= \frac{1}{n^2 h^4} \left[ n \mathbf{E} \left( \mathbf{K}^2 \left( \frac{X_i - x}{h}, \frac{Y_i - y}{h} \right) \right) - n^2 \mathbf{E}^2 \left( \mathbf{K} \left( \frac{X_i - x}{h}, \frac{Y_i - y}{h} \right) \right) \right. \\
&\quad \left. + 2 \sum_{1 \leq i < j \leq n} \mathbf{E} \left( \mathbf{K} \left( \frac{X_i - x}{h}, \frac{Y_i - y}{h} \right) \mathbf{K} \left( \frac{X_j - x}{h}, \frac{Y_j - y}{h} \right) \right) \right]
\end{aligned}$$

L'erreur quadratique moyenne (en anglais : "Mean squared error", MSE) de l'estimateur à noyau est donnée par :

$$\mathbf{MSE} \left( \hat{f}(x, y) \right) = \mathbf{E} \left( \left( \hat{f}(x, y) - f(x, y) \right)^2 \right) \quad (1.29)$$

L'expression exacte de l'erreur quadratique moyenne intégrée (en anglais : "Mean Integrated Squared Error", MISE) peut être obtenue à partir de

$$\mathbf{MISE} = \int \mathbf{MSE} \left( \hat{f}(x, y) \right) dx dy \quad (1.30)$$

et égale à

$$\mathbf{MISE} = \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \mathbf{K}(u, v) [f(uh + x, vh + y) - f(x, y)] dudv dx dy \quad (1.31)$$

### Expressions asymptotiques du biais et de la variance

Une approximation asymptotique de l'espérance de l'estimateur  $\hat{f}(x, y)$  est donnée (sous certaines conditions sur  $f$  et  $\mathbf{K}$ ) par :

$$\begin{aligned}
\mathbf{E} \left( \hat{f}(x, y) \right) &= \int_{\mathbb{R}^2} \mathbf{K}(u, v) f(uh + x, vh + y) dudv \\
&= f \int \mathbf{K}(u, v) dudv + h \left[ \frac{\partial f}{\partial x} \int u \mathbf{K}(u, v) dudv + \frac{\partial f}{\partial y} \int v \mathbf{K}(u, v) dudv \right] \\
&\quad + h^2 \left[ \frac{\partial^2 f}{2 \partial x^2} \int u^2 \mathbf{K}(u, v) dudv + \frac{\partial^2 f}{2 \partial y^2} \int v^2 \mathbf{K}(u, v) dudv + \frac{\partial^2 f}{\partial x \partial y} \int uv \mathbf{K}(u, v) dudv \right]
\end{aligned}$$

**Notation**

$\langle \cdot, \cdot \rangle$  : le produit scalaire.

$\nabla f(x, y)$  : le gradient de  $f$  au point  $(x, y)$ .

$\Delta f(x, y)$  : le laplacien de  $f$  au point  $(x, y)$ .

nous avons introduit la notation  $\int \mathbf{K}(u, v) \phi(u, v) dudv = \Gamma_\phi$ .

Alors on peut écrire l'espérance de l'estimateur comme suite :

$$\mathbf{E} \left( \hat{f}(x, y) \right) = f(x, y) \Gamma_1 + h \langle \nabla f(x, y), (1, 1) \rangle \Gamma_u + h^2 \frac{\partial^2 f(x, y)}{\partial x \partial y} \Gamma_{uv} + \frac{h^2}{2} \Delta f(x, y) \Gamma_{u^2} + o(h^2) \quad (1.32)$$

Supposons maintenant que le noyau  $\mathbf{K}$  satisfait :

- 1)  $\mathbf{K} > 0$ .
- 2)  $\Gamma_1 = 1$ .
- 3)  $\Gamma_u = \Gamma_{uv} = 0$ .
- 4)  $\Gamma_{u^2} < \infty$ .

Alors :

$$\mathbf{E} \left( \hat{f}(x, y) - f(x, y) \right) = \frac{h^2}{2} \Delta f(x, y) \Gamma_{u^2} + o(h^2)$$

Remarquons que

Si  $h$  décroît alors le  $(bias)^2 \searrow$  et la variance  $\nearrow$

Si  $h$  augmente alors le  $(bias)^2 \nearrow$  et la variance  $\searrow$

Il faut donc essayer de choisir un  $h$  qui fasse un compromis entre le  $(bias)^2$  et la variance.

Les expressions asymptotiques du biais et de la variance de  $\hat{f} = \hat{f}_n$  nous permettent

de trouver des expressions asymptotiques pour la **MSE** et la **MISE**. Rappelons ces expressions asymptotiques du biais et de la variance :

$$\mathbf{Bias} \left\{ \hat{f}(x, y) \right\} = \frac{h^2}{2} \Delta f(x, y) \Gamma_{u^2} + o(h^2) \quad (1.33)$$

$$\begin{aligned} \mathbf{Var} \left( \hat{f}(x, y) \right) &= \frac{1}{nh^2} f(x, y) \int \mathbf{K}^2(u, v) dudv + o\left(\frac{1}{nh^2}\right) \\ &= \frac{1}{nh^2} f(x, y) \Gamma_{\mathbf{K}} + o\left(\frac{1}{nh^2}\right) \end{aligned} \quad (1.34)$$

Apartir de (1.33) et (1.34) on peut obtenir facilement l'approximation asymptotique pour la **MSE** et la **MISE**.

$$\mathbf{MSE} \left( \hat{f}(x, y) \right) = \frac{h^4}{4} \Delta^2 f(x, y) \Gamma_{u^2}^2 + \frac{1}{nh^2} f(x, y) \Gamma_{\mathbf{K}} + o\left(h^4 + \frac{1}{nh^2}\right) \quad (1.35)$$

$$\mathbf{MISE} \left( \hat{f}(x, y) \right) = \frac{h^4}{4} \Gamma_{u^2}^2 \int \Delta^2 f(x, y) dx dy + \frac{1}{nh^2} \Gamma_{\mathbf{K}} + o\left(h^4 + \frac{1}{nh^2}\right) \quad (1.36)$$

Sous des conditions appropriées d'intégrabilité de  $f$  et ses dérivées.

On note l'approximation asymptotique de la **MSE** par

$$\mathbf{AMSE} \left( \hat{f}(x, y) \right) = \frac{h^4}{4} \Delta^2 f(x, y) \Gamma_{u^2}^2 + \frac{1}{nh^2} f(x, y) \Gamma_{\mathbf{K}} \quad (1.37)$$

et l'approximation asymptotique de la **MISE** par

$$\mathbf{AMISE} \left( \hat{f}(x, y) \right) = \frac{h^4}{4} \Delta^2 f(x, y) \Gamma_{u^2}^2 + \frac{1}{nh^2} \Gamma_{\mathbf{K}} \quad (1.38)$$

### 1.3 Ordre stochastique

Un des principaux objectifs des statistiques est la comparaison de variables aléatoires. Bien que très populaire, le critère probabiliste « moyenne variance » ne suffit

pas toujours à comparer deux variables aléatoires et peut même conduire à des aberrations. Cependant, il arrive souvent qu'on possède des informations plus détaillées concernant les variables aléatoires à comparer décrites par leurs fonctions de répartition. Une comparaison basée sur les distributions est plus informative que celle basée uniquement sur deux critères. La méthode utilisée pour comparer deux distributions est nommée « ordre stochastique ».

L'ordre stochastique usuel est l'ordre le plus naturel pour comparer deux variables aléatoires réelles (comparer des risques). Il consiste à comparer leurs fonctions de répartition ou leurs fonctions de survie. Cet ordre est souvent appelé ordre stochastique usuel selon Shaked et Shanthikumar (2007) [4]. Ceci conduit à la définition suivante.

**Définition 1.1** Une variable aléatoire  $X$  est dite inférieure ou égale à une variable aléatoire  $Y$  dans :

- a) L'ordre stochastique ( $X \preceq_s Y$ ) (Stochastic order) si  $F_X(t) \geq F_Y(t), \forall t$ ;
- b) L'ordre de taux de risque (Hazard rate order) ( $X \preceq_{hr} Y$ ), si  $h_X(t) \geq h_Y(t), \forall t$ ;
- c) L'ordre de rapport de vraisemblance (Likelihood ratio order) ( $X \preceq_{lr} Y$ ), si  $\frac{f_X(t)}{f_Y(t)}$  diminue en  $t$ .
- d) L'ordre convexe (Convex order) ( $X \preceq_{cx} Y$ ), si, pour toute fonction convexe  $\phi$  on a,  $\mathbf{E}[\phi(X)] \leq \mathbf{E}[\phi(Y)]$ .

**Remarque 1.3** Les implications suivantes (Shaked et Shanthikumar [4]) sont bien

connues :

$$X \preceq_{lr} Y \Rightarrow X \preceq_{hr} Y \Rightarrow X \preceq_s Y. \quad (1.39)$$

$$\text{Si } E[X] = E[Y], \text{ alors } X \preceq_{cx} Y \Leftrightarrow X \preceq_s Y. \quad (1.40)$$

## 1.4 Théorie des valeurs extrêmes

La théorie des valeurs extrêmes est une branche des statistiques qui s'intéresse aux valeurs extrêmes (minimum ou maximum) des distributions de probabilité. Elle a été développée par **Émil Julius Gumbel** [26].

### 1.4.1 Théorie des valeurs extrêmes maximales

L'étude des extrêmes d'un processus passe naturellement par l'analyse du maximum d'un échantillon de taille  $n$  donnée  $M_n = \max\{X_1, X_2, \dots, X_n\}$ , où  $X_1, X_2, \dots, X_n$  est un échantillon i.i.d. de loi  $F$ . La distribution de  $M_n$  est connue exactement :

$$\mathbf{P}(M_n \leq x) = \mathbf{P}(X_1 \leq x, \dots, X_n \leq x) = F^n(x),$$

En pratique,  $F$  est inconnue, et la relation (1.3) n'est donc pas utilisable directement. De façon analogue au théorème central limite, la théorie des valeurs extrêmes montre qu'il existe des suites  $a_n, b_n$  (constante de normalisation) et une distribution non dégénérée, telles que

$$\mathbf{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} G(x). \quad (1.41)$$

Le théorème suivant spécifie la forme de la loi limite de valeurs extrêmes du maximum  $G(x)$ .

**Théorème 1.1 (de Fisher-Tippett ou théorème des 3 types extrêmes)** La loi limite  $G(x)$  peut prendre trois formes possibles avec  $a_n > 0$  et  $b_n \in \mathbb{R}$  comme suit :

-La loi de **Gumbel** (type**I**)

$$G_X(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R} \quad (1.42)$$

-La loi de **Fréchet** (type**II**)

$$G_X(x) = \begin{cases} \exp(-x^{-\alpha}), & \forall x > 0, \alpha > 0 \\ 0 & \forall x \leq 0 \end{cases} \quad (1.43)$$

-La loi de **Weibull** (type**III**)

$$G_X(x) = \begin{cases} \exp(-(-x)^\alpha), & \forall x < 0, \alpha > 0 \\ 1 & \forall x \geq 0 \end{cases} \quad (1.44)$$

Le théorème suivant nous permet de déterminer le domaine d'attraction et de son type pour notre fonction de distribution commune.

**Théorème 1.2 ( Leadbetter et al. [7] ).** Soit  $(X_n)$  une suite de variables aléatoires indépendantes de la distribution commune  $F(x)$ , et  $x_F = \sup\{x \mid F(x) < 1\}$ . Les conditions nécessaires et suffisantes pour que la fonction  $F$  appartienne au domaine d'attraction des types possibles sont :

**Type I.** Il existe une fonction strictement positive  $g(t) > 0$  définie sur l'ensemble  $] -\infty, x_F]$ ,

telle que  $\lim_{t \rightarrow x_F} \frac{1-F(t+g(t)x)}{1-F(t)} = \exp(-x), \forall x \in \mathbb{R}$ .

**Type II.**  $x_F = \infty$  et  $\lim_{t \rightarrow \infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha}, \forall x > 0$  et  $\alpha > 0$ .

**Type III.**  $x_F < \infty$  et  $\lim_{t \rightarrow 0} \frac{1-F(x_F-tx)}{1-F(x_F-t)} = x^\alpha, \forall x > 0$  et  $\alpha > 0$ . ■

Le corollaire suivant permet de trouver les constantes de normalisation.

**Corollaire 1.1 ( Leadbetter et al. [7] )** Les constantes de normalisation  $a_n$  et  $b_n$

correspondantes aux différents types de loi limite sont :

**Type I.**  $a_n = g\left(F^{-1}\left(1 - \frac{1}{n}\right)\right)$  et  $b_n = F^{-1}\left(1 - \frac{1}{n}\right)$ .

**Type II.**  $a_n = F^{-1}\left(1 - \frac{1}{n}\right)$  et  $b_n = 0$ .

**Type III.**  $a_n = x_F - F^{-1}\left(1 - \frac{1}{n}\right)$  et  $b_n = x_F$ . ■

### 1.4.2 Théorie des valeurs extrêmes minimales

De façon similaire on note  $m_n = \min\{X_1, X_2, \dots, X_n\}$ , où  $X_1, X_2, \dots, X_n$  est un échantillon i.i.d. de loi  $F$ . La distribution de  $m_n$  est :

$$\mathbf{P}(m_n \leq x) = 1 - [1 - F(x)]^n,$$

Au lieu de traiter l'échantillon minimal de  $X$  on peut le voir comme le négatif du maximum de  $(-X)$ .

$$\min(X_1, X_2, \dots, X_n) = -\max(-X_1, -X_2, \dots, -X_n).$$

La distribution asymptotique minimale peut-être déduite de la distribution limite maximale.

La théorie des valeurs extrêmes montre qu'il existe des suites  $a_n, b_n$  (constante de

normalisation) , avec  $a_n > 0, b_n \in \mathbb{R}$  telles que

$$\mathbf{P}\left(\frac{m_n - a_n}{b_n} \leq x\right) \xrightarrow{n \rightarrow \infty} G^\star(x) = 1 - G(-x).$$

où  $G(x)$  est la loi limite des valeurs extrêmes du maximum définie dans le théorème 1.3.

Alors la loi limite des valeurs extrêmes du minimum  $G^\star(x)$  doit être l'un des trois types :

-La loi de **Gumbel** (type**I**)

$$G_X^\star(x) = 1 - \exp(-\exp(x)), x \in \mathbb{R}. \quad (1.45)$$

-La loi de **Fréchet** (type**II**)

$$G_X^\star(x) = \begin{cases} 1 - \exp(-(-x)^{-\alpha}), & \forall x \leq 0, \alpha > 0. \\ 1 & \forall x > 0. \end{cases}$$

-La loi de **Weibull** (type**III**)

$$G_X^\star(x) = \begin{cases} 1 - \exp(-(x)^\alpha), & \forall x > 0, \alpha > 0 \\ 0 & \forall x \leq 0. \end{cases}$$

**Remarque 1.4** Les constantes de normalisation peuvent être obtenues à l'aide de résultats de corollaire 1.1 avec les modifications appropriées.

## 1.5 Courbe de Lorenz

La courbe de **Lorenz** est une des mesures d'inégalité la plus largement utilisée dans plusieurs domaines (épidémiologie, traitement du signal, psychologie expérimentale...etc). Elle peut être facilement transposée, notamment à la répartition d'une

donnée statistique quelconque, comme les inégalités de répartition d'un actif ou de toute autre distribution de richesse, l'état de la répartition des clients au sein d'une clientèle, le revenu et la répartition des richesses. Dans le cas de l'analyse des revenus, la courbe de Lorenz  $L(P)$  représente la part du revenu total détenue par la proportion  $P \in [0, 1]$  d'individus les plus pauvres :

$$L(P) = \frac{\text{revenu total des plus pauvres}}{\text{revenu total}}.$$

La courbe de **Lorenz** pour une variable aléatoire  $X$  positive est définie comme le graphe du rapport

$$L(P) = \frac{\int_0^x tf(t)}{\int_0^{+\infty} tf(t)} \quad 0 \leq P \leq 1$$

$$\text{Où } P = F(x)$$

$$L(P) = \frac{\mathbf{E}(X|X \leq x)P(X \leq x)}{\mathbf{E}(X)}.$$

$$L(P) = \frac{\mathbf{E}(X|X \leq x)P}{\mathbf{E}(X)}. \quad (1.46)$$

Avec les propriétés  $L(p) \leq p$ ,  $L(0) = 0$  et  $L(1) = 1$ . Si  $X$  représente le revenu annuel,  $L(p)$  est la proportion du revenu total qui revient aux personnes ayant les revenus les plus faibles de  $100p\%$ . Si tous les individus gagnent le même revenu alors  $L(p) = p$  pour tout  $p$ . La zone située entre la ligne  $L(p) = p$  et la courbe de **Lorenz** peut être considérée comme une mesure de l'inégalité des revenus, ou plus généralement, de la variabilité de  $X$ , voir **Gail** et **Gastwirth** [5] et **Dagum** [6] pour une vaste discussion des courbes de **Lorenz**.

## 1.6 Entropie

L'entropie est la mesure du désordre dans les systèmes physiques, ou une quantité d'information qui peut être acquise par les observations des systèmes désordonnés.

**Claude Shannon** [6] définit une mesure formelle de l'entropie, appelée entropie de **Shannon**

### 1.6.1 Entropie Shannon

L'entropie de **Shannon** est une fonction décroissante d'une dispersion de variable aléatoire, et est maximale lorsque tous les résultats sont également probables.

L'entropie de **Shannon** est comme :

$$\eta = \mathbf{E} \{-\ln(g(x))\} \quad (1.47)$$

### 1.6.2 Entropie de Rényi

Alfred **Rényi** [3] est proposé d'entropie généralisée qui, pour  $s = 1$ , se réduit à l'entropie de **Shannon**. L'entropie de **Rényi** est définie comme :

$$I_R(s) = \frac{1}{1-s} \left\{ \ln \int_1^\infty g^s(x) dx \right\} \quad s > 0, s \neq 1 \quad (1.48)$$

Il a des propriétés similaires à celles de l'entropie de **Shannon** :

- Elle est additive
- Il a maximum =  $\ln(n)$  pour  $p_i = \frac{1}{n}$

Mais elle contient un paramètre supplémentaire  $s$  qui peut être utilisé pour le rendre plus ou moins sensible à la forme des distributions de probabilité.

## Chapitre 2

# Distribution de Lindley et ses applications

Une distribution de **Lindley** d'un seul paramètre a attiré les chercheurs pour son usage en modélant des données de vie, et on l'a observé en plusieurs articles que cette distribution a exécuté excellemment. Cette distribution est introduit par **Lindley** en 1958 comme mélange d' $\text{Exp}(\theta)$  et de  $\text{Gamma}(2, \theta)$ . plus de détails sur la distribution de **Lindley** peut être trouvés en Ghitany et autres ( [1], [23] ).

### 2.1 Distribution de Lindley

Soient  $Y_1 \sim \text{exp}(\theta)$  et  $Y_2 \sim \text{Gamma}(2, \theta)$  deux variables aléatoires indépendantes. Pour  $\theta > 0$ , on considère la variable aléatoire  $X = Y_1$  et  $X = Y_2$  avec les probabilités respectivement  $P_1 = \frac{\theta}{1+\theta}$  et  $P_2 = \frac{1}{1+\theta}$ .

La fonction de densité de  $X$  est donnée par :

$$f(x; \theta) = \frac{\theta^2(1+x)e^{-\theta x}}{1+\theta}; \quad x, \theta > 0 \quad (2.1)$$

La fonction de répartition correspondante est :

$$F(x) = 1 - \frac{\theta + 1 + \theta x}{1 + \theta} e^{-\theta x}; \quad x > 0, \theta > 0 \quad (2.2)$$

La première dérivée de (2.1) est :

$$\frac{d}{dx} f(x) = \frac{\theta^2}{1+\theta} (1 - \theta - \theta x) e^{-\theta x}.$$

Il en résulte que

(i) pour  $\theta < 1$ ,  $\frac{d}{dx} f(x) = 0$  implique que  $x_0 = \frac{1-\theta}{\theta}$  est le point critique unique à laquelle  $f(x)$  est maximisée.

(ii) pour  $\theta \geq 1$ ,  $\frac{d}{dx} f(x) \leq 0$ , c-à-d  $f(x)$  diminue en  $x$ .

Par conséquent, le mode de cette distribution est :

$$\text{Mode}(X) = \begin{cases} \frac{1-\theta}{\theta}, & 0 < \theta < 1 \\ 0, & \text{ailleurs} \end{cases}$$

La figure 2.1 représente la fonction de densité de la distribution de **Lindley** pour quelques valeurs de  $\theta$ .

## 2.2 Moments et mesures connexes

Le moment d'ordre  $k$  de la distribution de **Lindley** est :

$$\mu_k^{\prime} = E(X^k) = \frac{k!(\theta + k + 1)}{\theta^k(\theta + 1)}, k = 1, 2, \dots$$

d'où, on a :

$$\mu_1^{\prime} = \frac{(\theta + 2)}{\theta(\theta + 1)}, \mu_2^{\prime} = \frac{2(\theta + 3)}{\theta^2(\theta + 1)}, \mu_3^{\prime} = \frac{6(\theta + 4)}{\theta^3(\theta + 1)}, \mu_4^{\prime} = \frac{24(\theta + 5)}{\theta^4(\theta + 1)}.$$

Le moment centré d'ordre  $k$  de la distribution de **Lindley** est défini par :

$$\mu_k = \mathbf{E}\{(X - \mu)^r\} = \sum_{r=0}^k \binom{k}{r} \mu_k^{\prime} (-\mu)^{k-r}.$$

D'où, on a

$$\mu_2 = \frac{\theta^2 + 4\theta + 2}{\theta^2(\theta + 1)} = \sigma^2, \mu_3 = \frac{2(\theta^3 + 6\theta^2 + 6\theta + 2)}{\theta^3(\theta + 1)^3}, \mu_4 = \frac{3(3\theta^4 + 24\theta^3 + 44\theta^2 + 32\theta + 8)}{\theta^4(\theta + 1)^4}.$$

Le coefficient de variation ( $\gamma$ ), le coefficient de dissymétrie ( $\sqrt{\beta_1}$ ) et le coefficient d'aplatissement ( $\beta_2$ ) sont :

$$\begin{aligned} \gamma &= \frac{\sqrt{\theta^2 + 4\theta + 2}}{\theta + 2}, \\ \sqrt{\beta_1} &= \frac{2(\theta^3 + 6\theta^2 + 6\theta + 2)}{(\theta^2 + 4\theta + 2)^{\frac{3}{2}}}, \\ \beta_2 &= \frac{3(3\theta^4 + 24\theta^3 + 44\theta^2 + 32\theta + 8)}{(\theta^2 + 4\theta + 2)^2}. \end{aligned}$$

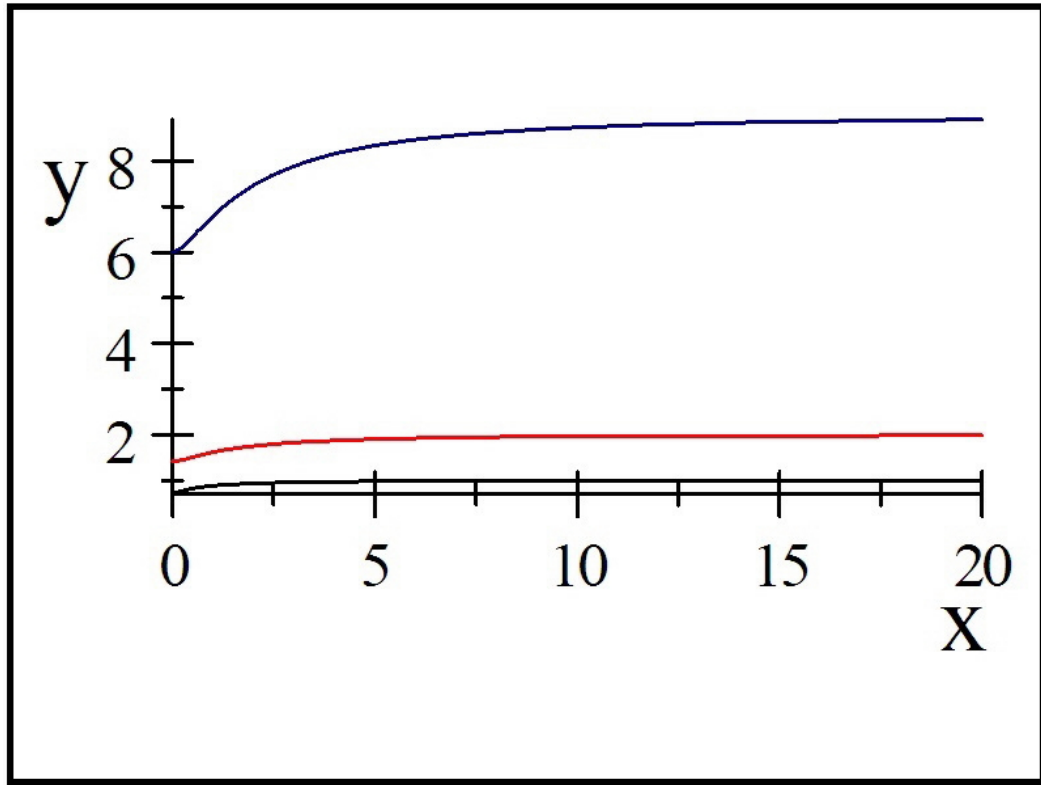


FIG. 2.1: Présentation graphique du coefficient de variation  $\gamma$  (noir), le coefficient de dissymétrie  $\sqrt{\beta_1}$  (rouge) et le coefficient d'aplatissement  $\beta_2$  (bleu) .

**Remarque 2.1 :**

- (i)  $\gamma$  est une fonction croissante en  $\theta$  et  $(1/\sqrt{2}) < \gamma < 1$ , voir Fig2.2.
- (ii)  $\sqrt{\beta_1}$  est une fonction croissante en  $\theta$  et  $\sqrt{2} < \sqrt{\beta_1} < 2$ , voir Fig2.2.
- (iii)  $\beta_2$  est une fonction croissante en  $\theta$  et  $6 < \beta_2 < 9$ , voir Fig2.2.

**Théorème 2.1** Soit  $X \sim \text{Lindley}(\theta)$ . Alors

$$\text{Mode}(X) < \text{Median}(X) < E(X).$$

**Preuve.** Soient  $M = \text{Mode}(X)$ ,  $m = \text{Median}(X)$  et  $\mu = \mathbf{E}(X) = \frac{\theta+2}{\theta(\theta+1)}$ . Depuis la

fonction de répartition de la distribution de **Lindley**, il en résulte que :

$$F(M) = \begin{cases} 1 - \frac{2}{1+\theta}e^{-(1-\theta)}, & 0 < \theta < 1 \\ 0 & \text{ailleurs} \end{cases}, \quad F(me) = \frac{1}{2}$$

et

$$F(\mu) = 1 - \frac{\theta^2 + 3\theta + 3}{(1+\theta)^2}e^{-\frac{\theta+2}{\theta+1}}.$$

Notons que  $F(M)$  est une fonction décroissante en  $\theta \in (0, 1)$  et, pour tout  $\theta > 0$ ,

$$0 \leq F(M) < 1 - 2e^{-1} < (1/2).$$

De même,  $F(\mu)$  est une fonction croissante en  $\theta > 0$  et  $(1/2) < 1 - 3e^{-2} < F(\mu) < 1$ .

Enfin, étant donné que  $F(x)$  est une fonction croissante en  $x > 0$ . Il est facile de vérifier que  $F(M) < F(me) = \frac{1}{2} < F(\mu)$ , alors on a  $M < m < \mu$ . ■

## 2.3 Fonction de hasard et fonction de survie

Soit

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{P(X < x + \Delta x \mid X > x)}{\Delta x} = \frac{f(x)}{1 - F(x)} = \frac{\theta^2(1+x)}{\theta + 1 + \theta x}.$$

et

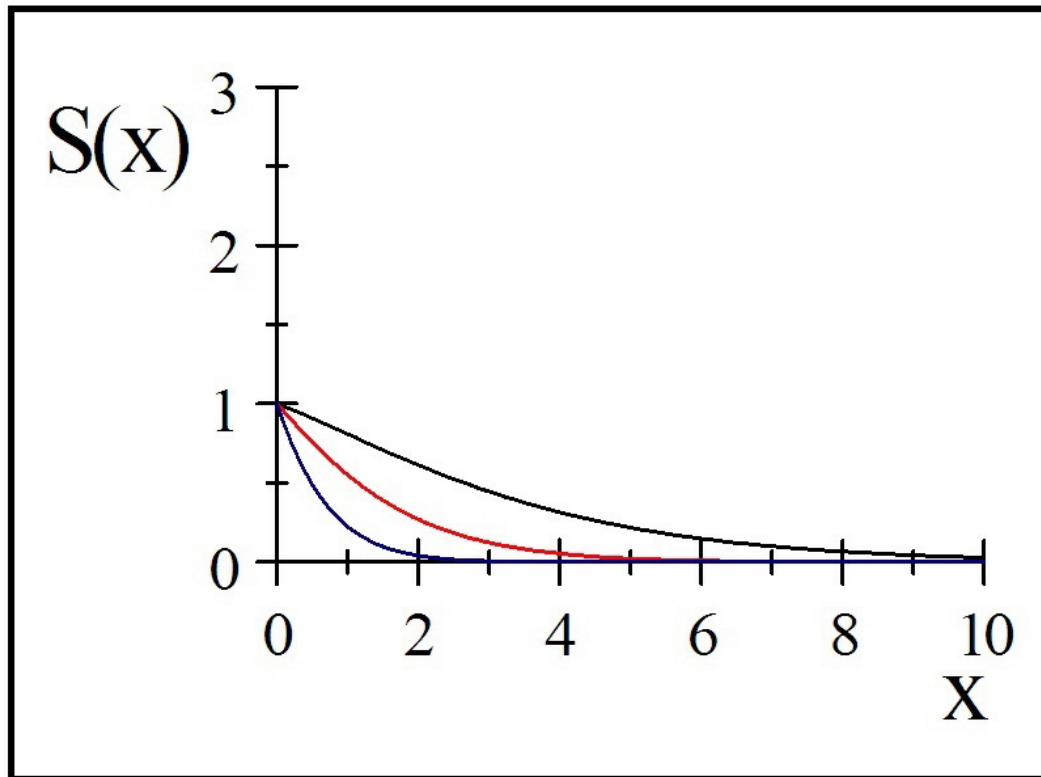
$$S(x) = 1 - F(x) = \frac{\theta + 1 + \theta x}{1 + \theta}e^{-\theta x}.$$

La fonction de taux de hasard et la fonction de survie de la distribution de **Lindley**, respectivement.

**Remarque 2.2**

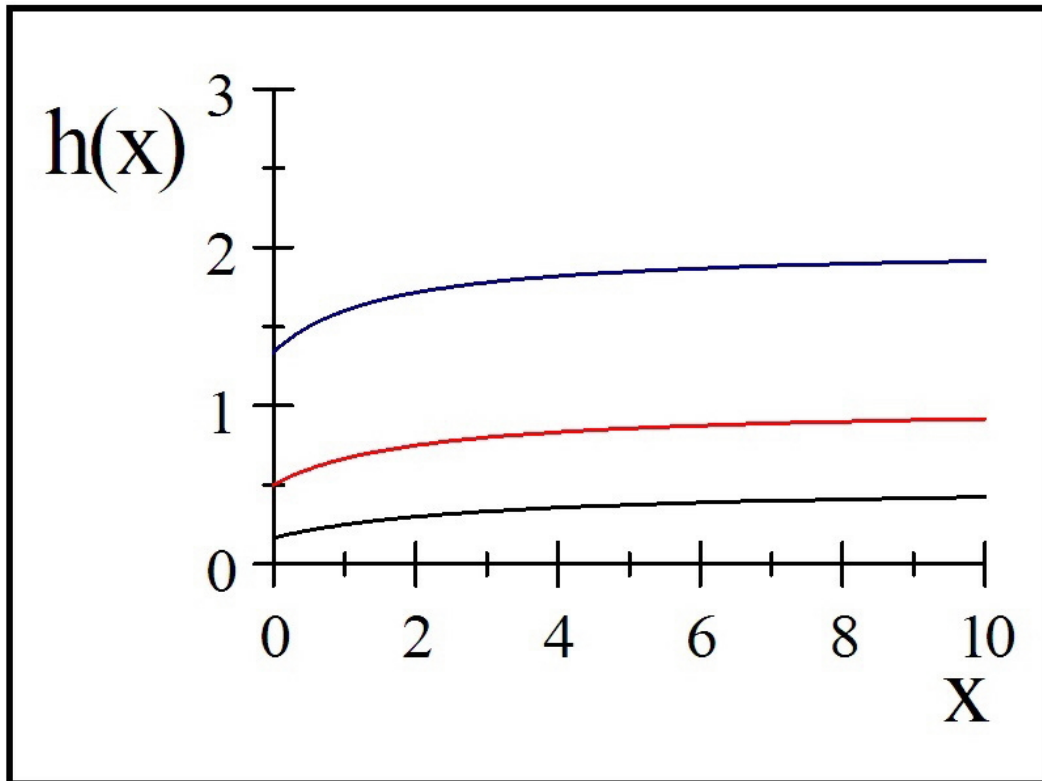
(i)  $h(0) = f(0) = \frac{\theta^2}{\theta+1}$ .

(ii) Comme  $\frac{d}{dx}h(x) = \frac{\theta^2(1+x)}{\theta+1+\theta x}$ ,  $h(x)$  est une fonction croissante en  $x$  et  $\theta$  en plus  $\frac{\theta^2}{\theta+1} < h(x) < \theta$ .



Présentation graphique de la fonction de taux de hasard pour quelques valeurs de  $\theta$

noir ( $\theta=0.5$ ) ; rouge ( $\theta=1$ ) ; bleu ( $\theta=2$ )



Présentation graphique de la fonction de survie pour quelques valeurs de  $\theta$ ,

noir ( $\theta=0.5$ ); rouge ( $\theta=1$ ); bleu ( $\theta=2$ ).

## 2.4 Ordre stochastique

**Théorème 2.2** Soient  $X \sim \mathbf{Lindley}(\theta_1)$  et  $Y \sim \mathbf{Lindley}(\theta_2)$  : Si  $\theta_1 > \theta_2$  alors  $X \preceq_{lr} Y$  et donc  $X \preceq_{hr} Y$  et  $X \preceq_s Y$ .

**Preuve.** Notons d'abord que

$$\frac{f_X(t)}{f_Y(t)} = \frac{\theta_1^2 (1 + \theta_2)}{\theta_2^2 (1 + \theta_1)} e^{-(\theta_1 - \theta_2)t}. \quad t > 0$$

On a, pour  $\theta_1 > \theta_2$ ,

$$\frac{d}{dt} \frac{f_X(t)}{f_Y(t)} = (\theta_2 - \theta_1) \frac{f_X(t)}{f_Y(t)} < 0,$$

$\frac{f_X(t)}{f_Y(t)}$  est décroissante en  $X$ . Alors  $X \preceq_{lr} Y$ . Les états restants découlent des implications dans (1.39).

## 2.5 Courbe de Lorenz

La courbe de Lorenz pour une variable aléatoire  $X$  positif est défini comme le graphe du rapport

$$L(F(x)) = \frac{\mathbf{E}(X|X \leq x)F(x)}{\mathbf{E}(X)} \quad (2.3)$$

Pour la distribution de **Lindley** (2.1) on a,

$$\mathbf{E}(X|X \leq x)F(x) = \frac{2 + \theta}{\theta(1 + \theta)} - \frac{e^{-\theta x}}{1 + \theta} \left[ \frac{2}{\theta} + 1 + \theta x^2 + 2x + x\theta \right].$$

Ainsi, à partir de (2.3), on obtient la courbe de Lorenz pour la distribution de **Lindley** comme suit :

$$L(p) = 1 - \frac{\theta(1 + \theta)(1 - p)}{(2 + \theta)(1 + \theta + \theta x)} \left[ \frac{2}{\theta} + 1 + \theta x^2 + 2x + x\theta \right].$$

Où  $x = F^{-1}(p)$  avec  $F(\cdot)$  donnée par (2.2).

## 2.6 Statistiques d'ordre extrêmes

Si  $X_1, \dots, X_n$  un échantillon de  $n$  variables aléatoires qui suivent la distribution de **Lindley** et si  $\bar{X} = (X_1 + \dots + X_n)/n$  représente la moyenne d'échantillon alors par le théorème central limite  $\frac{\sqrt{n}(\bar{X} - \mathbf{E}(X))}{\sqrt{\text{Var}(X)}}$  se rapproche de la distribution normale standard quand  $n \rightarrow \infty$ .

**Théorème 2.3** Parfois, on serait intéressé à étudier la loi asymptotique des valeurs extrêmes  $M_n = \max(X_1, \dots, X_n)$  et  $m_n = \min(X_1, \dots, X_n)$ , Pour la fonction de répartition définie dans (2.2), on constate que :

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

et

$$\lim_{t \rightarrow 0} \frac{F(tx)}{F(t)} = x.$$

Ainsi, il en résulte du *Théorème 1.2* ( **Leadbetter** et **al.** [7] ) qu'il doit y avoir les constantes de normalisation  $a_n > 0, b_n, c_n > 0$  et  $d_n$  de telle sorte que :

$$\mathbf{P}\{a_n(M_n - b_n) \leq x\} \rightarrow -\exp\{(-\theta x)\}.$$

et

$$\mathbf{P}\{c_n(m_n - d_n) \leq x\} \rightarrow 1 - \exp(-x). \quad (2.4)$$

Comme  $n \rightarrow \infty$ . La forme des constantes de normalisation peut également être déterminée. Par exemple, en utilisant le *Corollaire 1.1* ( **Leadbetter et al.** [7] ), on peut voir que  $a_n = 1$  et  $b_n = F^{-1}(1 - 1/n)$  avec  $F(\cdot)$  donnée par (2.2)

## 2.7 Fonction Quantile de la distribution Lindley

D'après la fonction de répartition de la distribution **Lindley** définie en (2.2). Il convient de noter qu'est continue et strictement croissante de sorte que la fonction de quantile  $X$  est  $Q_X(u) = F_X^{-1}(u)$ ,  $0 < u < 1$ . Dans le résultat suivant, on donne une expression explicite de  $Q_X$  en fonction de la fonction  $W$  de **Lambert**.

**Théorème 2.4** Pour tout  $\theta > 0$ , la fonction quantile de la distribution **Lindley**  $X$  est :

$$Q_X(u) = -1 - \frac{1}{\theta} - \frac{1}{\theta} W_{-1} \left( \frac{\theta + 1}{\exp(\theta + 1)} (u - 1) \right), \quad 0 < u < 1, \quad (2.5)$$

Où  $W_{-1}$  désigne la branche négative de la fonction  $W$  de **Lambert**.

**Preuve.** Pour tout  $\theta$  fixe,  $\theta > 0$ , soit  $u \in (0, 1)$ . On doit résoudre l'équation  $F_X(x) = u$  par rapport à  $x$ , pour tous  $X > 0$  comme suit :

$$(\theta + 1 + \theta x) e^{-\theta x} = (\theta + 1) (1 - u). \quad (2.6)$$

En multipliant par  $-\exp(-\theta - 1)$  l'équation (2.6), on obtient :

$$-(\theta + 1 + \theta x) \exp(-\theta - 1 - \theta x) = (\theta + 1) (u - 1) \exp(-\theta - 1). \quad (2.7)$$

D'après l'équation (2.7), conjointement avec l'équation (2.5), on voit que  $-(\theta + 1 + \theta x)$  est la fonction  $W$  de **Lambert** de l'argument réel  $(\theta + 1) (u - 1) \exp(-\theta - 1)$ . Alors,

on a

$$W\left(\frac{\theta+1}{\exp(\theta+1)}(u-1)\right) = -(\theta+1+\theta x), 0 < u < 1. \quad (2.8)$$

Toujours, pour tout  $\theta > 0$  et  $x > 0$  il est immédiat que  $\theta+1+\theta x > 1$  et il peut également être vérifié que puisque  $u \in (0, 1)$ . Il pour, en prenant en compte les propriétés de la branche négative de la fonction  $W$  de **Lambert** a présenté en premier chapitre, l'équation (2.8) devient

$$W_{-1}\left(\frac{\theta+1}{\exp(\theta+1)}(u-1)\right) = -(\theta+1+\theta x). \quad (2.9)$$

Ce qui implique le résultat. la preuve du théorème est terminée ■.

## 2.8 Estimation

### 2.8.1 Estimation par la méthode des moments (MoM)

Étant donné un échantillon aléatoire  $X_1, \dots, X_n$ , de la distribution de **Lindley** (2.1), l'estimateur des moments (*MoM*) de  $\theta$  est :

$$\hat{\theta}_{MoM} = \frac{-(\bar{X}-1) + \sqrt{(\bar{X}-1)^2 + 8\bar{X}}}{2\bar{X}}, \bar{X} > 0. \quad (2.10)$$

Le théorème suivant montre que l'estimateur  $\hat{\theta}_{MoM}$  de  $\theta$  est biaisé.

**Théorème 2.5** L'estimateur  $\hat{\theta}_{MoM}$  de  $\theta$  est positivement biaisée, i.e.  $\mathbf{E}(\hat{\theta}) - \theta > 0$ .

**Preuve.** Soient  $\hat{\theta}_{MoM} = g(\bar{X})$  et  $g(t) = \frac{-(t-1) + \sqrt{(t-1)^2 + 8t}}{2t}$ ,  $\forall t > 0$ .

Comme  $g''(t) = \frac{1}{t^3} \left[ 1 + \frac{3t^3 + 15t^2 + 9t + 1}{[(t-1)^2 + 8t]^{\frac{3}{2}}} \right] > 0$ ,  $g(t)$  est strictement convexe.

Ainsi, par l'inégalité de **Jensen**, on a  $\mathbf{E}(g(\bar{X})) > g[\mathbf{E}(\bar{X})]$ . Enfin, étant donné que

$$\mathbf{E}(g(\bar{X})) = g(\mu) = g\left(\frac{\theta + 2}{\theta(\theta + 1)}\right) = \theta, \quad (2.11)$$

On obtient

$$\mathbf{E}(\hat{\theta}_{MoM}) > \theta.$$

Le théorème suivant donne la loi limite de  $\hat{\theta}_{MoM}$ .

**Théorème 2.6.** L'estimateur  $\hat{\theta}_{MoM}$  de  $\theta$  est convergent et asymptotiquement normal :

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{P} N\left(0, \frac{1}{\sigma^2}\right). \quad (2.12)$$

L'intervalle de confiance de  $\theta$  pour un seuil de confiance  $100(1 - \alpha)\%$  est donné par :

$$\hat{\theta} \pm z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n\hat{\sigma}^2}}. \quad (2.13)$$

Où  $z_{\frac{\alpha}{2}}$  est le  $(1 - \frac{\alpha}{2})$  percentile de la distribution normale standard.

**Preuve.** Étant donné  $\mu$  est finie,  $\bar{X} \xrightarrow{P} \mu$ .  $g(t)$  est une fonction continue à  $t = \mu$ ,  $g(\bar{X}) \xrightarrow{P} g(\mu)$ , c'est-à-dire  $\hat{\theta} \xrightarrow{P} \theta$ . Comme  $\sigma^2 < \infty$ , par le théorème central limite, on a

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{P} N(0, \sigma^2). \quad (2.14)$$

En outre, puisque  $g(\mu)$  est différentiable et  $g'(\mu) \neq 0$ , par la méthode Delta, on a :

$$\sqrt{n}(g(\bar{X}) - g(\mu)) \xrightarrow{P} N\left(0, [g'(\mu)]^2 \sigma^2\right). \quad (2.15)$$

Enfin, étant donné que

$$g(\bar{X}) = \hat{\theta}_{MoM}, g(\mu) = \theta, \quad \text{et} \quad g'(\mu) = \frac{-1}{2\mu^2} \left[ 1 + \frac{1 + 3\mu}{\sqrt{(\mu - 1)^2 + 8\mu}} \right] = -\frac{1}{\sigma^2}.$$

## 2.8.2 Estimation par la méthode du maximum de vraisemblance

Soient  $X_i \sim LD(\theta)$ ,  $i = \overline{1, n}$   $n$  variables aléatoires. La fonction de logvraisemblance est :

$$\ln l(x_i; \theta) = 2n \ln \theta - n \ln(\theta + 1) + \sum_{i=1}^n \ln(x_i + 1) - n\theta\bar{X}. \quad (2.16)$$

L'estimateurs de la méthode du maximum de vraisemblance  $\hat{\theta}_{MV}$  de  $\theta$  est une solution de l'équation :

$$\frac{\partial \ln l(x_i; \beta, \theta)}{\partial \theta} = \frac{2n}{\theta} - \bar{X} - \frac{n}{(\theta + 1)} = 0.$$

On obtient

$$\hat{\theta}_{MV} = \frac{-(\bar{X} - 1) + \sqrt{(\bar{X} - 1)^2 + 8\bar{X}}}{2\bar{X}}, \bar{X} > 0.$$

Avec

$$\frac{\partial^2 \ln l(x_i; \beta, \theta)}{\partial \theta^2} = -\frac{2n}{\theta^2} - \sum_{i=1}^n \frac{x_i^2}{(\theta x_i + \theta)^2} < 0.$$

**Remarque 2.3** L'estimateur de la méthode des moments  $\hat{\theta}_{MoM}$  et l'estimateur du maximum de vraisemblance  $\hat{\theta}_{MV}$  du paramètre  $\theta$  sont les mêmes.

## 2.9 Simulation

Cette section étudie le comportement des estimateurs du maximum de vraisemblance  $\hat{\theta}_{MV}$  pour un échantillon de taille finie ( $n$ ). La simulation est réalisée pour chaque couple

$(\theta; n)$ , où  $\theta = 0.1, 1, 9$  et  $n = 20, 40, 60, 80, 100$ . Alors on a l'algorithme suivant :

- Choisir les valeurs initiales de  $\theta_0$  pour spécifier la distribution de **Lindley** ;
- Choisir la taille de l'échantillon  $n$  ;
- Générer  $N$  échantillons indépendants de taille  $n$  de  $LD(\theta)$  ;
- Calculer les estimations  $\hat{\theta}_{MV}$  de  $\theta$  pour chacun des  $N$  échantillons ;
- Calculer :

(i) La moyenne des estimateurs obtenus sur tous les  $N$  échantillons

$$\text{biais moyen}(\theta) = \frac{1}{N} \sum_{i=1}^N (\hat{\Theta}_i - \Theta_0).$$

(ii) L'erreur quadratique moyenne  $EQM$  des estimations simulées

$$EQM(\theta) = \frac{1}{N} \sum_{i=1}^N (\hat{\Theta}_i - \Theta_0)^2. \quad (2.17)$$

$n$	$\theta = 0.1$	$\theta = 1$	$\theta = 9$
20	0.00265	0.03068	0.41459
40	0.00118	0.01699	0.18019
60	0.00090	0.01060	0.13665
80	0.00061	0.00746	0.09073
100	0.00058	0.00570	0.07728

**Tableau 2.1** : les biais moyens de l'estimateur  $\hat{\theta}$

$n$	$\theta = 0.1$	$\theta = 1$	$\theta = 9$
20	0.00028	0.03335	4.25425
40	0.00013	0.01543	1.91203
60	0.00009	0.01024	0.22017
80	0.00007	0.00752	0.90044
100	0.00005	0.00590	0.70490

**Tableau 2.2** : la moyenne EQM de l'estimateur  $\hat{\theta}$

**Remarque 2.4**

(i) Le tableau 2.1 présente un biais positif, comme indiqué dans le *théorème 2.11*.

Le tableau montre également que le biais diminue (augmente) quand  $n(\theta)$  augmente respectivement.

(ii) Le tableau 2.2 montre que l'erreur quadratique moyenne diminue (augmente) lorsque  $n(\theta)$  augmente respectivement .

## 2.10 Distributions à deux paramètres de Lindley

Dans cette section on va déterminer quelques distributions de deux et trois paramètres dont la distribution de **Lindley** est un cas particulier

### 2.10.1 Distribution Quasi Lindley

La distribution Quasi **Lindley**(*QLD*) avec deux paramètres  $\alpha$  et  $\theta$  est définie par sa fonction de densité de probabilité

$$f(x; \alpha, \theta) = \frac{\theta^2(\alpha + x\theta)e^{-\theta x}}{\alpha + 1}; \quad x > 0, \theta > 0, \alpha > -1. \quad (2.18)$$

Il est facile de voir que si  $\alpha = \theta$ , l'équation (2.22) de *QLD* se réduit à la fonction de distribution de **Lindley** (2.1) et si  $\alpha = 0$ , elle réduit à la distribution *Gamma* (2,  $\theta$ ).

La fonction de densité (2.22) de *QLD* peut être montré sous forme de mélange des distributions Exponentielle ( $\theta$ ) et *Gamma* (2,  $\theta$ ) comme suit :

$$f(x; \alpha, \theta) = pf_1(x) + (1 - p) f_2(x)$$

Où  $p = \frac{\alpha}{\alpha+1}$ ,  $f_1(x) = \theta e^{-\theta x}$  et  $f_2(x) = \theta^2 x e^{-\theta x}$ . (pour plus de détails voir. **Shanker** et **al.**(2013) [8]).

La fonction de répartition correspondante est :

$$F(x) = 1 - \frac{1 + \alpha + \theta x}{\alpha + 1} e^{-\theta x}; \quad x > 0, \theta > 0, \alpha > -1. \quad (2.19)$$

Le moment d'ordre k de la distribution Quasi **Lindley** est :

$$\mu_k^1 = E(X^k) = \frac{\Gamma(k+1)(\alpha+k+1)}{\theta^k(\alpha+1)}, \quad k = 1, 2, \dots$$

d'où, on a

$$\mu_1^1 = \frac{(\alpha+2)}{\theta(\alpha+1)}, \mu_2^1 = \frac{2(\alpha+3)}{\theta^2(\alpha+1)}, \mu_3^1 = \frac{6(\alpha+4)}{\theta^3(\alpha+1)}, \mu_4^1 = \frac{24(\alpha+5)}{\theta^4(\alpha+1)}.$$

La fonction de logvraisemblance de la distribution Quasi **Lindley** est :

$$\log L(x; \alpha, \theta) = n \log \theta - n \log (1 + \alpha) + \sum_{i=0}^n \log (\alpha + x_i \theta) - n \theta \bar{X}.$$

### Estimation par la méthode des moments

Étant donné un échantillon aléatoire  $X_1, \dots, X_n$ , de  $QLD$ , les estimateurs des moments

$\hat{\theta}$  de  $\theta$  et  $\hat{\alpha}$  de  $\alpha$  peuvent être obtenus comme suit :

$$\hat{\theta} = \left( \frac{\hat{\alpha} + 2}{\hat{\alpha} + 1} \right) \frac{1}{\bar{X}}, \quad \bar{X} > 0. \quad (2.20)$$

$$\hat{\alpha} = \frac{-1}{k-2} \left[ 2k + \sqrt{2} \sqrt{2-k} - 4 \right] \quad (2.21)$$

Où

$$\frac{\mu_2^1}{\mu_1^1} = k = \frac{2(\alpha + 3)(\alpha + 1)}{(\alpha + 2)^2}.$$

### 2.10.2 Distribution de Pseudo Lindley ( $PsLD$ )

Dans cette subsection, on introduit la distribution de Pseudo **Lindley** [18] et étudier ses propriétés de base. Cette distribution de deux paramètres  $\theta > 0$  et  $\beta \geq 1$ , le premier est appelé paramètre d'échelle alors que  $\beta$  est le paramètre de mélange est défini comme un mélange d'Exp( $\theta$ ) et de Gamma(2,  $\theta$ )

Soit  $Y_1 \sim Exp(\theta)$  et  $Y_2 \sim Gamma(2, \theta)$  deux variables aléatoires indépendantes.

Pour  $\beta \geq 1$ , on considère la variable aléatoire  $X = Y_1$  et  $X = Y_2$  avec les probabilités respectivement  $P_1 = \frac{\beta-1}{\beta}$  et  $P_2 = \frac{1}{\beta}$ .

La fonction de densité de  $X$  est donnée par :

$$f_{PsLD}(x; \theta, \beta) = \frac{\theta (\beta - 1 + \theta x) e^{-\theta x}}{\beta}, \quad x > 0, \theta > 0, \beta \geq 1. \quad (2.22)$$

La fonction de répartition correspondante est :

$$F_{PsLD}(x) = 1 - \frac{(\beta + \theta x) e^{-\theta x}}{\beta}; x > 0, \beta \geq 1, \theta > 0. \quad (2.23)$$

**Remarque 2.5**

1) Si  $\beta = \theta + 1$ , cette distribution est la distribution de **Lindley** LD ( $\theta$ ).

2) Si  $\beta = 1$ , cette distribution est la distribution de Gamma( $2, \theta$ ).

La première dérivée de  $f_{PsLD}(x)$  est :

$$\frac{d}{dx} f_{PsLD}(x) = \frac{\theta^2 (2 - \beta - \theta x) e^{-\theta x}}{\beta} = 0 \text{ donne } x = \frac{2 - \beta}{\theta}.$$

Il en résulte que :

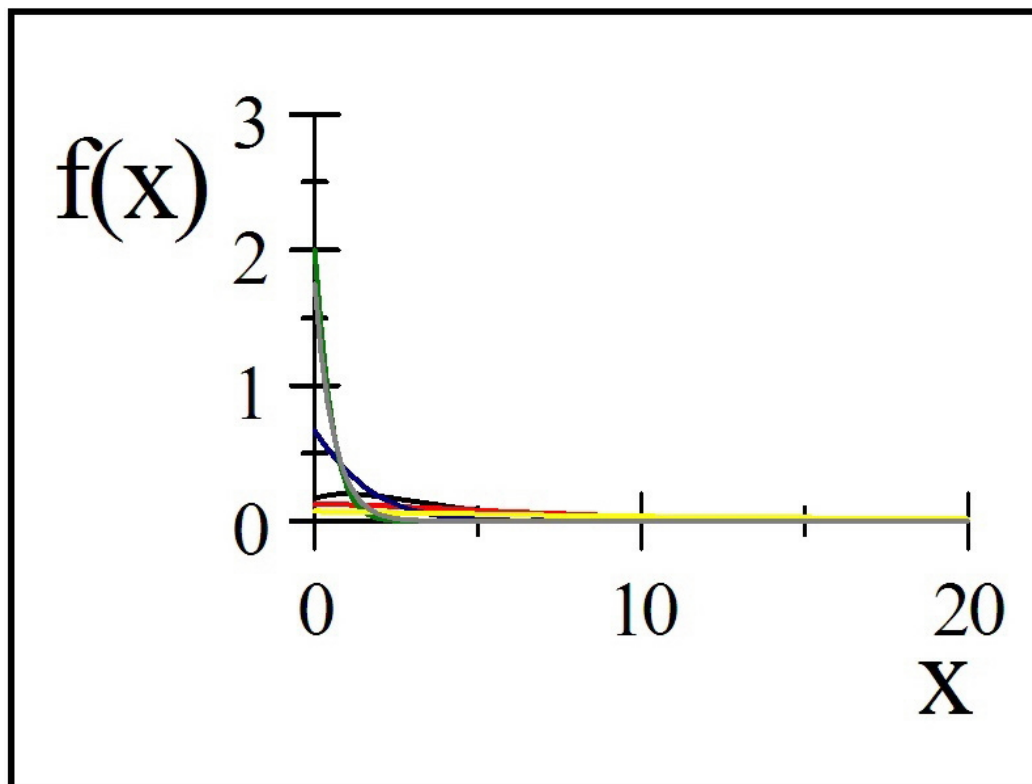
(i) pour  $1 \leq \beta < 2$ ,  $\frac{d^2}{dx^2} f_{PsLD}(\hat{x}) < 0$  implique que  $\hat{x} = \frac{2-\beta}{\theta}$  est l'unique point critique unique à laquelle  $f_{PsLD}(x; \theta, \beta)$  est maximisée.

(ii) pour  $\beta \geq 2$ ,  $\frac{d}{dx} f_{PsLD}(x; \theta, \beta) \leq 0$ , c-à-d  $f_{PsLD}(x; \theta, \beta)$  est décroissante en  $x$ .

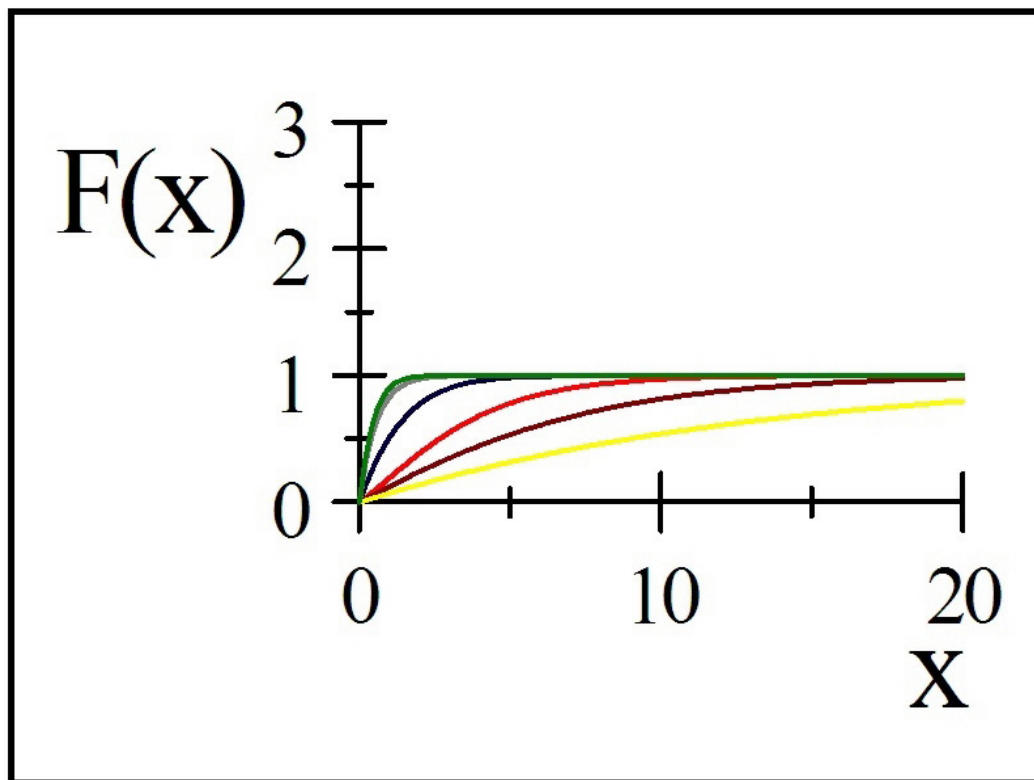
Par conséquent, le mode de cette distribution est :

$$Mode(X) = \begin{cases} \frac{2-\beta}{\theta} & \forall 1 \leq \beta < 2 \\ 0 & \text{ailleurs.} \end{cases} \quad (2.24)$$

Les figures 3.1 et 3.2 représentent la fonction de densité et la fonction de répartition de la distribution de Pseudo **Lindley** pour quelques valeurs de  $(\theta, \beta)$ .



Représentation graphique la fonction de densité de PsLD pour quelques valeurs de  $(\theta, \beta)$ . noir(0.5,1.5);rouge(0.25,2);bleu(1,3);vert(3,3),jaune(0.1,4);gris(2,8).



Représentation graphique la rpartition de densite de PsLD pour quelques valeurs de  $(\theta, \beta)$  noir(0.5,1.5);rouge(0.25,2);bleu(1,3);vert(3,3),jaune(0.1,4);gris(2,8).

### 2.10.3 Moments et Mesures connexes

Le moment d'ordre  $k$  de la distribution de Pseudo **Lindley** est :

$$\mathbf{E}(X^k) = \frac{k! (\beta + k)}{\theta^k \beta}, k = 1, 2, \dots$$

**Proposition 2.1** Soient  $X_1, X_2, \dots, X_n$   $n$  variables aléatoires indépendantes de  $PsLD(\theta, \beta)$ .

La fonction génératrice de  $S = \sum_{i=1}^n X_i$ , est donnée par

$$M_S(t) = \frac{\theta^n ((1 - \beta)t + \theta\beta)^n}{\beta^n (t - \theta)^{2n}}.$$

d'où

$$M_X(t) = \mathbf{E}(e^{tX}) = \frac{\theta ((1 - \beta)t + \theta\beta)}{\beta (t - \theta)^2}. \quad (2.25)$$

**Corollaire 2.1** Soit  $X \sim PsLD(\theta, \beta)$ , le moment d'ordre 1, 2 et la variance de  $X$  sont :

$$\mathbf{E}(X) = \frac{\beta + 1}{\theta\beta}, \mathbf{E}(X^2) = \frac{\beta + 2}{\theta^2\beta}, \mathbf{Var}(X) = \frac{\beta^2 + 2\beta - 1}{\beta^2\theta^2}. \quad (2.26)$$

**Théorème 2.7** Soient  $X \sim PsLD(\theta, \beta)$ ,  $M = Mode(X)$ ,  $me = Median(X)$  et  $Moyenne(X) = \mu = \mathbf{E}(X)$ . Alors

$$M < me < \mu. \quad (2.27)$$

**Preuve.** Selon La fonction de répartition de  $PsLD(\theta, \beta)$  pour tout  $x, \theta$  et  $\beta$

$$F(M) = \begin{cases} 1 - \frac{2e^{-(2-\beta)}}{\beta} & \forall 1 \leq \beta < 2 \\ 0 & \text{ailleurs} \end{cases}, F(me) = \frac{1}{2}.$$

et

$$F(\mu) = 1 - \frac{(\beta^2 + \beta + 1) e^{-\left(\frac{\beta+1}{\beta}\right)}}{\beta^2}.$$

Notons que  $F(M)$  est une fonction croissante pour tous  $\beta \succeq 1$ . Il est facile de vérifier que

$F(M) < F(me) < F(\mu)$ , alors on a  $M < me < \mu$ . ■

Le coefficient de variation  $\gamma$ , le coefficient de dissymétrie  $\sqrt{\beta_1}$  et le coefficient d'aplatissement  $\beta_2$  sont :

$$\begin{aligned} \gamma &= \frac{\sqrt{Var(X)}}{\mathbf{E}(X)} = \frac{\sqrt{\beta^2 + 2\beta - 1}}{\beta + 1}, \\ \sqrt{\beta_1} &= \frac{\mathbf{E}(X^3)}{(Var(X))^{\frac{3}{2}}} = \frac{6\beta^2(\beta + 3)}{(\beta^2 + 2\beta - 1)^{\frac{3}{2}}}, \\ \beta_2 &= \frac{\mathbf{E}(X^4)}{(Var(X))^2} = \frac{24\beta^3(\beta + 4)}{(\beta^2 + 2\beta - 1)^2}. \end{aligned}$$

**Remarque 2.6** Toutes ces expressions sont indépendantes du paramètre  $\theta$  et dépendantes uniquement du paramètre  $\beta$ , avec  $6\sqrt{2} < \sqrt{\beta_1} < 6$  et  $\beta_2 > 30$ .

#### 2.10.4 Fonction de hasard et Fonction de survie

Soit

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{P(X < x + \Delta x \mid X > x)}{\Delta x} = \frac{f(x)}{1 - F(x)} = \frac{\theta(\beta + \theta x - 1)}{\beta + \theta x}.$$

et

$$S(x) = 1 - F(x) = \frac{(\beta + \theta x) e^{-\theta x}}{\beta}.$$

La fonction de taux de hasard et la fonction de survie de la distribution de Pseudo **Lindley** respectivement.

**Proposition 2.2** Soit  $h_{PsLD}(x)$  la fonction de taux hasard de  $X$ . Alors  $h_{PsL}(x)$  est croissante.

**Preuve.** Il est facile de vérifier que  $\frac{d}{dx}h_{PsLD}(x) = \frac{\theta^2}{(\beta+\theta x)^2} > 0$ .

### 2.10.5 Ordres stochastiques

**Théorème 2.8.** Soient  $X_i \sim PsL(\theta_i, \beta_i), i = 1, 2$  deux variables aléatoires. Si  $\theta_1 = \theta_2$  et  $\beta_1 \geq \beta_2$ , alors  $X_1 \prec_{lr} X_2, X_1 \prec_{hr} X_2, X_1 \prec_s X_2$  et  $X_1 \leq_{cx} X_2$ .

### 2.10.6 Courbe de Lorenz

La courbe de Lorenz pour la distribution de Pseudo **Lindley** est :

$$L(p) = 1 - e^{-\theta x} \frac{[\beta(x\theta + 1) + x^2\theta^2 + x\theta + 1]}{\beta + 1}.$$

Où  $x = F^{-1}(p)$  avec  $F(\cdot)$  donnée par (2.2).

### 2.10.7 Statistiques d'ordres extrêmes

Soient  $X_1, \dots, X_n$   $n$  variables aléatoires qui suivent la distribution de Pseudo **Lindley**. Dans le théorème suivant on va étudier la loi asymptotique des valeurs extrêmes

$M_n = \max(X_1, \dots, X_n)$  et  $m_n = \min(X_1, \dots, X_n)$ .

**Théorème 2.9** Pour la fonction de répartition définie dans (2.2), on constate que :

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

et

$$\lim_{t \rightarrow 0} \frac{F(tx)}{F(t)} = x.$$

Ainsi, il résulte du **théorème 1.2 ( Leadbetter et al. [7] )** qu'il doit y avoir constante de normalisation  $a_n > 0, b_n, c_n > 0$  et  $d_n$  de telle sorte que :

$$Pr\{a_n(M_n - b_n) \leq x\} \rightarrow \exp(-\exp(-\theta x)).$$

et

$$Pr\{c_n(m_n - d_n) \leq x\} \rightarrow 1 - \exp(-x). \quad (2.28)$$

Comme  $n \rightarrow \infty$ . La forme des constantes de normalisation peut également être déterminée. Par exemple, en utilisant le **Corollaire 2.3 ( Leadbetter et al. [7] )**, on peut voir que  $a_n = 1$  et  $b_n = F^{-1}(1 - 1/n)$  avec  $F(\cdot)$  donnée par (2.2).

### 2.10.8 Fonction Quantile de la distribution de Pseudo Lindley

D'après la fonction de répartition de la distribution de Pseudo **Lindley** définie en (2.2), il convient de noter qu'est continue et strictement croissante de sorte que la fonction de quantile  $X$  est  $Q_X(u) = F_X^{-1}(u)$ ,  $0 < u < 1$ . Dans le résultat suivant, on donne une expression explicite de  $Q_X$  en fonction de la fonction  $W$  de **Lambert**.

**Théorème 2.8** Pour tout  $\theta > 0$  et  $\beta > 1$ , la fonction quantile de  $X$  qui suit la distribution de Pseudo **Lindley** est

$$Q_X(u) = -\frac{\beta}{\theta} - \frac{1}{\theta} W_{-1}(\beta e^{-\beta}(u-1)), \quad 0 < u < 1, \quad (2.29)$$

Où  $W_{-1}$  désigne la branche négative de la fonction  $W$  de **Lambert**.

**Preuve.** Voir ( **Zeghdoudi et Nedjar (2017)**)[21].

### 2.10.9 Estimation

#### Estimation par la méthode des moments (MoM)

Étant donné un échantillon aléatoire  $X_1, \dots, X_n$ , de la distribution de Pseudo **Lindley** (2.23), en utilisant le premier moment  $m$  et la variance  $s^2$  on a :

$$m = \frac{\beta + 1}{\theta\beta}, s^2 = \frac{\beta^2 + 2\beta - 1}{\beta^2\theta^2}.$$

On résout ce système non linéaire pour tout  $s > 0, m > 0$ , pour trouver les estimateurs des moments  $\hat{\theta}_{MoM}$  et  $\hat{\beta}_{MoM}$  de  $\theta$  et  $\beta$  respectivement comme suit :

$$\hat{\theta}_{MoM} = \frac{2m + \sqrt{2}\sqrt{m^2 - s^2}}{m^2 + s^2} \quad \text{et} \quad \hat{\beta}_{MoM} = \frac{m^2 + s^2}{m^2 - s^2 + \sqrt{2}m\sqrt{m^2 - s^2}}. \quad (2.30)$$

Le théorème suivant montre que l'estimateur  $\hat{\theta}_{MoM}$  de  $\theta$  est positivement biaisé.

**Théorème 2.11** L'estimateur des moments  $\hat{\theta}_{MoM}$  de  $\theta$  est positivement biaisée , c'est-à-dire  $E(\hat{\theta}_{MoM}) - \theta > 0$ .

### 2.10.10 Distribution de Gamma Lindley (*GaL*)

Dans cette subsection, on introduit la distribution de Gamma **Lindley** ([17], [19]) et étudier ses propriétés de base. Cette distribution est défini comme un mélange de Gamma  $(2, \theta)$  et de **Lindley** $(\theta)$ .

Soit  $Y_1 \sim \text{Gamma}(2, \theta)$  et  $Y_2 \sim \text{LD}(\theta)$  deux variables aléatoires indépendantes. Pour  $\theta > 0, \beta > \frac{\theta}{1+\theta}$  les paramètres d'échelle et de mélange respectivement, on considère la variable aléatoire  $X = Y_1$  et  $X = Y_2$  avec les probabilités respectivement  $P_1 = \frac{\beta-1}{\beta}$  et  $P_2 = \frac{1}{\beta}$ .

La fonction de densité de  $X$  est donnée par :

$$f_{\text{GaL}}(x; \theta, \beta) = \frac{\theta^2((\beta + \beta\theta - \theta)x + 1)e^{-\theta x}}{\beta(1 + \theta)} \quad x > 0, \theta > 0, \beta > \frac{\theta}{1+\theta}. \quad (2.31)$$

**Remarque 2.7** Si  $\beta = 1$ , cette distribution est la distribution de **Lindley**  $\text{LD}(\theta)$ .

Par conséquent, le mode de cette distribution est :

$$\text{Mode}(X) = \begin{cases} \frac{\beta\theta + \beta - 2\theta}{\theta(\beta + \beta\theta - \theta)}, & \forall \beta \in \left[ \frac{2\theta}{\theta + 1}, \infty \right] \\ 0, & \text{ailleurs.} \end{cases} \quad (2.32)$$

On peut facilement trouver la fonction de répartition de *GaL* :

$$F_{\text{GaL}}(x) = 1 - \frac{((\theta\beta + \beta - \theta)(\theta x + 1) + \theta)e^{-\theta x}}{\beta(1 + \theta)}; x > 0, \theta > 0, \beta > \frac{\theta}{1 + \theta}. \quad (2.33)$$

Pour plus détail voir(**Zeghdoudi et Nedjar** [17]).

### 2.10.11 Moments et mesures connexes

**Proposition 2.3** Soient  $X_1, X_2, \dots, X_n$   $n$  variables aléatoires indépendantes de  $GaL(\theta, \beta)$ . La fonction génératrice de  $S = \sum_{i=1}^n X_i$ , est donnée par :

$$M_S(t) = \left( \frac{\theta}{\theta - t} \right)^{2n} \left( \frac{\theta\beta + \beta - t}{\beta(1 + \theta)} \right)^n. \quad (2.34)$$

d'où

$$M_X(t) = \mathbf{E}(e^{tX}) = \left( \frac{\theta}{\theta - t} \right)^2 \left( \frac{\theta\beta + \beta - t}{\beta(\theta + 1)} \right). \quad (2.35)$$

**Remarque 2.8** La fonction génératrice des moments pour  $X$  et  $S$  existent

( $\mathbf{E}(e^{tX}) < \infty$ ) si seulement si  $t < \theta$ .

**Corollaire 2.2** Soit  $X \sim GaL(\theta, \beta)$  le moment d'ordre 1, 2 et la variance de  $X$  sont :

$$\begin{aligned} \mathbf{E}(X) &= \frac{2\beta(1 + \theta) - \theta}{\theta\beta(1 + \theta)}, \quad \mathbf{E}(X^2) = \frac{6\beta\theta + 6\beta - 4\theta}{\beta\theta^2 + \beta\theta^3}, \\ \mathbf{Var}(X) &= \frac{-(-2\beta\theta + \theta)^2 + (2 + 6\theta)\beta^2 - 2\beta(\beta\theta - 3\beta\theta^2 + 2\theta^2)}{\beta^2\theta^2(1 + \theta)^2}. \end{aligned} \quad (2.36)$$

**Théorème 2.12** Soit  $X \sim GaLD(\theta, \beta)$ ,  $M = Mode(X)$ ,  $me = Median(X)$  et

$Moyenne(X) = \mu = \mathbf{E}(X)$  Alors

$$M < me < \mu. \quad (2.37)$$

**Preuve.** Voir Théorème 1 (Zeghdoudi et Nedjar (2016))[17]. ■

### Fonction de hasard et la La fonction de survie

Soit

$$h_{GaL}(x) = \frac{f_{GaL}(x)}{1 - F_{GaL}(x)} = \frac{((\beta + \theta\beta - \theta)x + 1)\theta^2}{\theta(\beta + \theta\beta - \theta)x + \beta + \theta\beta}.$$

et

$$S_{GaL}(x) = 1 - F_{GaL}(x) = \frac{((\theta\beta + \beta - \theta)(\theta x + 1) + \theta)e^{-\theta x}}{\beta(1 + \theta)}$$

La fonction de taux de hasard et la fonction de survie de la distribution de Gamma **Lindley**, respectivement.

**Proposition 2.4** Soit  $h_{GaL}(x)$  la fonction de taux de risque de  $X$ . Alors  $h_{GaL}(x)$  est croissante.

**Preuve.** D'après *Glaser* (1980) et à partir de la fonction de densité (2.31)

$$\rho(x) = -\frac{f'_{GaL}(x)}{f_{GaL}(x)} = -\frac{(\beta - 2\theta + \theta\beta + x\theta^2 - x\theta\beta - x\theta^2\beta)}{x(\beta - \theta + \theta\beta) + 1}.$$

Comme  $\rho'(x) = \frac{(\beta - \theta + \theta\beta)^2}{(x\beta - x\theta + x\theta\beta + 1)^2} \geq 0, \forall x, \beta, \theta$  alors  $h_{GaL}(x)$  est croissante. ■

## 2.10.12 Ordres Stochastiques

**Théorème 2.14** Soient  $X_i \sim GaL(\theta_i, \beta_i)$ ,  $i = 1, 2$  deux variables aléatoires.

Si  $\theta_1 = \theta_2$  et  $\beta_1 \geq \beta_2$ , alors  $X_1 \prec_{lr} X_2, X_1 \prec_{hr} X_2, X_1 \prec_s X_2$  et  $X_1 \leq_{cx} X_2$ .

**Preuve.** On a

$$\frac{f_{X_1}(t)}{f_{X_2}(t)} = \frac{\theta_1^2 \beta_2 (1 + \theta_2) ((\beta_1 + \beta_1 \theta_1 - \theta_1) t + 1)}{\theta_2^2 \beta_1 (1 + \theta_1) ((\beta_2 + \beta_2 \theta_2 - \theta_2) t + 1)} e^{-(\theta_1 - \theta_2)t}.$$

Pour simplifier, on utilise  $\ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right)$ , on peut trouver

$$\frac{d}{dt} \ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right) = -(\theta_1 - \theta_2) + \frac{\beta_1 \theta_1 - \beta_2 \theta_2 + (\beta_1 - \beta_2) + (\theta_2 - \theta_1)}{(t\beta_1 \theta_1 + t\beta_1 - t\theta_1 + 1)(t\beta_2 \theta_2 + t\beta_2 - t\theta_2 + 1)}.$$

À cet effet, si  $\theta_1 \succeq \theta_2$  et  $\beta_1 \leq \beta_2$ , on a  $\frac{d}{dt} \ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right) \leq 0$ . Les états restants découlent des implications dans (1.39), (1.40).

### 2.10.13 Courbe de Lorenz

La courbe de Lorenz pour la distribution de Gamma **Lindley**

$$L(p) = 1 - \frac{\theta^3 e^{-\theta x}}{2\beta(1+\theta) - \theta} \left[ (\beta + \beta\theta - \theta) \frac{(x^2\theta^2 + 2\theta x + 2)}{\theta^3} + \frac{x\theta + 1}{\theta^2} \right]$$

Où  $x = F^{-1}(p)$  avec  $F(\cdot)$  donnée par (2.33).

### 2.10.14 Statistiques d'ordres extrêmes

Si  $X_1, \dots, X_n$  un échantillon de  $n$  variables aléatoires qui suivent la distribution de Gamma **Lindley** et si  $\bar{X} = (X_1 + \dots + X_n)/n$  représente la moyenne d'échantillon alors par le théorème central limite  $\frac{\sqrt{n}(\bar{X} - \mathbf{E}(X))}{\sqrt{\mathbf{Var}(X)}}$  se rapproche de la distribution normale standard quand  $n \rightarrow \infty$ .

**Théorème 2.15** On va étudier la loi asymptotique des valeurs extrêmes

$M_n = \max(X_1, \dots, X_n)$  et  $m_n = \min(X_1, \dots, X_n)$ , Pour la fonction de répartition définie dans (2.33), on constate que

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

et

$$\lim_{t \rightarrow 0} \frac{F(tx)}{F(t)} = x.$$

Ainsi, il résulte du **théorème 1.2** (théorème 1.6.2 dans **Leadbetter et al.** [7]) qu'il doit y avoir normalisation constantes  $a_n > 0, b_n, c_n > 0$  et  $d_n$  de telle sorte que :

$$Pr\{a_n(M_n - b_n) \leq x\} \rightarrow \exp(-\exp(-\theta x)).$$

et

$$Pr\{c_n(m_n - d_n) \leq x\} \rightarrow 1 - \exp(-x). \quad (2.38)$$

Comme  $n \rightarrow \infty$ . La forme des constantes de normalisation peut également être déterminée. Par exemple, en utilisant le **Corollaire 2.2** (Corollaire 1.6.3 de **Leadbetter** et al. [7].), on peut voir que  $a_n = 1$  et  $b_n = F^{-1}(1 - 1/n)$  avec  $F(\cdot)$  donnée par (3.32).

### 2.10.15 Fonction Quantile de la distribution de Gamma Lindley

Dans le théorème suivant, on donne une expression explicite de  $Q_X$  en fonction de la fonction  $W$  de **Lambert**.

**Théorème 2.16** Pour tout  $\theta > 0, \beta > \frac{\theta}{1+\theta}$ , la fonction quantile de  $X$  qui suit la distribution de Gamma **Lindley** est :

$$Q_X(u) = -\frac{\beta(1+\theta)}{\theta(\beta(1+\theta) - \theta)} - \frac{1}{\theta} W_{-1} \left( \frac{\beta(1+\theta)(y-1)}{\beta(1+\theta) - \theta} e^{-\frac{\beta(1+\theta)}{\beta(1+\theta) - \theta}} \right), \quad 0 < u < 1. \quad (2.39)$$

Où  $W_{-1}$  désigne la branche négative de la fonction  $W$  de **Lambert**.

**Preuve.** Voir Théorème 1 ( **Zeghdoudi et Nedjar (2016)**)[19] ■.

### 2.10.16 Estimation

#### Estimation par la méthode des moments (MoM)

Étant donné un échantillon aléatoire  $X_1, \dots, X_n$ , de la distribution de Gamma Lindley (2.3), en utilisant le premier moment  $m$  et la variance  $s^2$  on a :

$$m = \frac{2\beta(1+\theta) - \theta}{\theta\beta(1+\theta)}, s^2 = \frac{-(-2\beta\theta + \theta)^2 + (2 + 6\theta)\beta^2 - 2\beta(\beta\theta - 3\beta\theta^2 + 2\theta^2)}{\beta^2\theta^2(1+\theta)^2}.$$

On résout ce système non linéaire pour tous  $s > 0, m > 0$ , pour trouver les estimateurs des moments  $\hat{\theta}_{MoM}$  et  $\hat{\beta}_{MoM}$  de  $\theta$  et  $\beta$  respectivement comme suit :

$$\hat{\theta}_{MoM} = \frac{1}{s^2 + m^2} \left( 2m + \sqrt{2}\sqrt{-s^2 + m^2} \right) \quad \text{et} \quad \hat{\beta}_{MoM} = \frac{\hat{\theta}}{(1 + \hat{\theta})(2 - \hat{\theta}m)}. \quad (2.40)$$

#### L'estimation par la méthode du maximum de vraisemblance

Soient  $X_i \sim GaL(\theta, \beta), i = \overline{1, n}$ ,  $n$  variables aléatoires. La fonction de logvraisemblance est :

$$\ln l(x_i; \theta, \beta) = 2n \ln \theta - n \ln \beta - n \ln(\theta + 1) + \sum_{i=1}^n \ln((\beta + \theta\beta - \theta)x_i + 1) - \theta \sum_{i=1}^n x_i. \quad (2.41)$$

Les dérivées de  $\ln l(x_i; \theta, \beta)$  par rapport à  $\theta$  et  $\beta$  sont :

$$\frac{\partial \ln l(x_i; \theta, \beta)}{\partial \theta} = \frac{2n}{\theta} - \frac{n}{1 + \theta} - \sum_{i=1}^n x_i + (\beta - 1) \sum_{i=1}^n \frac{x_i}{(\beta + \theta\beta - \theta)x_i + 1}. \quad (2.42)$$

$$\frac{\partial \ln l(x_i; \theta, \beta)}{\partial \beta} = \frac{-n}{\beta} + (1 + \theta) \sum_{i=1}^n \frac{x_i}{(\beta + \theta\beta - \theta)x_i + 1}. \quad (2.43)$$

Les deux équations (2.42) et (2.43) ne peut pas résoudre directement, on doit utiliser la méthode de **Fisher Scoring**, on a :

$$\begin{bmatrix} \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta^2} & \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta \partial \beta} \\ \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta \partial \theta} & \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta^2} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial Lnl(x_i; \beta, \theta)}{\partial \theta} \\ \frac{\partial Lnl(x_i; \beta, \theta)}{\partial \beta} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \quad (2.44)$$

L'équation (2.44) peut se résoudre de façon itérative où  $\theta_0, \beta_0$  sont les valeurs initiales de  $\theta, \beta$ .

### 2.10.17 Autres distribution de plusieurs paramètres

#### Distribution de Deux Paramètres de Lindley (*Two Parameter Lindley* (*TwoPLD*))

La distribution de Deux Paramètres de **Lindley** avec les paramètres  $\alpha$  et  $\theta$  est définie par sa fonction de densité de probabilité

$$f(x; \alpha, \theta) = \frac{\theta^2}{\theta + \alpha} (1 + \alpha x) e^{-\theta x}; \quad x > 0, \theta > 0, \alpha > -\theta. \quad (2.45)$$

Il est facile de voir que si  $\alpha = 1$ , la fonction (2.45) de *TwoPLD* réduit à la fonction de distribution de LD (2.1) et si  $\alpha = 0$ , elle se réduit à la distribution Exponentielle ( $\theta$ ).

La fonction de densité (2.45) de *TwoPLD* peut être montré sous forme de mélange des distributions Exponentielle ( $\theta$ ) et Gamma ( $2, \theta$ ) comme suit :

$$f(x; \alpha, \theta) = p f_1(x) + (1 - p) f_2(x)$$

Où  $p = \frac{\theta}{\theta + \alpha}$ ,  $f_1(x) = \theta e^{-\theta x}$  et  $f_2(x) = \theta^2 x e^{-\theta x}$ . (pour plus de détails voir. **S.Sharma** et al.(2013) [9]).

La fonction de répartition correspondante est :

$$F(x) = 1 - \frac{\theta + \alpha + \alpha\theta x}{\theta + \alpha} e^{-\theta x}; x > 0, \theta > 0, \alpha > -\theta. \quad (2.46)$$

Le moment d'ordre  $k$  de la distribution de Deux-Paramètres de **Lindley** est :

$$\mu_k^l = E(X^k) = \frac{\Gamma(k+1)(\theta + \alpha + \alpha k)}{\theta^k(\theta + \alpha)}, k = 1, 2, \dots$$

d'où, on a

$$\mu_1^l = \frac{(\alpha + 2\alpha)}{\theta(\theta + \alpha)}, \mu_2^l = \frac{2(\alpha + 3\alpha)}{\theta^2(\theta + \alpha)}, \mu_3^l = \frac{6(\alpha + 4\alpha)}{\theta^3(\theta + \alpha)}, \mu_4^l = \frac{24(\alpha + 5\alpha)}{\theta^4(\theta + \alpha)}.$$

La fonction de logvraisemblance de la distribution de deux-Paramètres de **Lindley**

est:

$$\log L(x; \alpha, \theta) = n \log \theta^2 - n \log(\alpha + \theta) + \sum_{i=0}^n \log(1 + \alpha\theta x_i) - n\theta \bar{X}.$$

### Estimation par la méthode des moments

Étant donné un échantillon aléatoire  $X_1, \dots, X_n$ , de Deux-Paramètre de **Lindley**, les

estimateurs des moments  $\hat{\theta}_{MoM}$  de  $\theta$  et  $\hat{\alpha}$  de  $\alpha$  peuvent être obtenus comme suit :

$$\hat{\theta} = \frac{1}{\mu_2^l} \left( 2\bar{X} + \sqrt{2} \sqrt{2\bar{X}^2 - \mu_2^l} \right), \quad \bar{X} > 0. \quad (2.47)$$

$$\hat{\alpha} = \frac{\hat{\theta} - \bar{X}\hat{\theta}^2}{\bar{X}\hat{\theta} - 2}. \quad (2.48)$$

### Distribution de Lindley Généralisée

La distribution de **Lindley** Généralisée (Generalized **Lindley** Distribution (*GLD*))

avec les paramètres  $\alpha$  et  $\theta$  est définie par sa fonction de densité de probabilité

$$f(x; \alpha, \theta, \gamma) = \frac{\theta^2 (\theta x)^{\alpha-1} (\alpha + \gamma x) e^{-\theta x}}{(\theta + \gamma) \Gamma(\alpha + 1)}; x, \theta, \alpha, \gamma > 0. \quad (2.49)$$

Il est facile de voir que si  $\alpha = \gamma = 1$ , la fonction (2.49) de *GLD* se réduit à la fonction de distribution de *LD* (2.1), si  $\gamma = 0$ , elle se réduit à la distribution *Gamma* ( $\alpha, \theta$ ) et si  $(\alpha, \gamma) = (1, 0)$ , elle se réduit à la distribution Exponentielle ( $\theta$ ). La fonction de densité (2.49) de *GLD* peut être montrée sous forme de mélange des distributions *Gamma* ( $\alpha, \theta$ ) et *Gamma* ( $\alpha + 1, \theta$ ) comme suit :

$$f(x; \alpha, \theta, \gamma) = p_1 f_1(x) + p_2 f_2(x).$$

Où  $p_1 = \frac{\theta}{\theta + \gamma}$ ,  $p_2 = \frac{\gamma}{\theta + \gamma}$ ,  $f_1(x) = \frac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x}$  et  $f_2(x) = \frac{\theta^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\theta x}$ . (pour plus de détails voir. Dolati et al. (2010) [10]).

La fonction génératrice correspondante est :

$$M(t) = \left( \frac{\theta}{\theta - t} \right)^{\alpha+1} \left( \frac{\theta - t + \gamma}{\theta + \gamma} \right).$$

Les moments d'ordre 1,2 et 3 de la distribution de **Lindley** Généralisée sont :

$$\begin{aligned} \mu_1^l &= -\frac{1}{\theta + \gamma} + \frac{\alpha + 1}{\theta}, \mu_2^l = -\frac{2(\alpha + 1)}{\theta(\theta + \gamma)} + \frac{\alpha^2 + 3\alpha + 2}{\theta^2}, \\ \mu_3^l &= \frac{\alpha + 1}{\theta^2} (-3\alpha + 6) + \frac{(\alpha + 1)(\theta + \gamma)}{\theta} \left[ (\alpha + 1)^2 + \frac{3\alpha + 5}{\theta^2} \right]. \end{aligned}$$

La fonction de logvraisemblance de la distribution de **Lindley** Généralisée est :

$$\begin{aligned} \log L(x; \alpha, \theta, \gamma) &= n(\alpha + 1) \log \theta - n \log(\gamma + \theta) - n \log \Gamma(\alpha + 1) + \\ &+ (\alpha - 1) \sum_{i=0}^n \log(x_i) + \sum_{i=0}^n \log(\alpha + \gamma x_i) - n\theta \bar{X}. \end{aligned}$$

### Estimation par la méthode des moments

Étant donné un échantillon aléatoire  $X_1, \dots, X_n$  de taille  $n$  de la distribution *GLD*,

les estimateurs des moments (*MoM*),  $\hat{\gamma}$  de  $\gamma$ ,  $\hat{\theta}$  de  $\theta$  et  $\hat{\alpha}$  de  $\alpha$  sont la solution des équations des moments :

$$\alpha(\theta + \gamma) + \gamma - \bar{X}\theta(\theta + \gamma) = 0$$

$$-2(\alpha + 1)\theta + (\alpha^2 + 3\alpha + 2)(\theta + \gamma) - n\theta^2(\theta + \gamma) = 0$$

$$\frac{\alpha + 1}{\theta^2}(-3\alpha + 6) + \frac{(\alpha + 1)(\theta + \gamma)}{\theta} \left[ (\alpha + 1)^2 + \frac{3\alpha + 5}{\theta^2} \right] - K = 0$$

Il est possible de calculer  $\hat{\gamma}$ ,  $\hat{\theta}$  et  $\hat{\alpha}$  numériquement. mais en utilisant  $\bar{X}$ ,  $n = \mu_2$  et  $k = \mu_3$ .

## Chapitre 3

# Distribution de Zeghdoudi

Dans ce chapitre, on introduit une nouvelle distribution nomée **Zeghdoudi**.

### 3.1 Distribution de Zeghdoudi (ZD) et immédiate propriétés

Dans cette section, nous donne la distribution de **Zeghdoudi** et étudions son propriétés.

Soit  $Y_1 \sim \text{gamma}(2, \theta)$  et  $Y_2 \sim \text{gamma}(3, \theta)$  deux variables aléatoires indépendantes. Nous considérons variable aléatoire  $X = Y_1$  et  $X = Y_2$  avec probabilité  $\frac{\theta}{1+\theta}$  et  $\frac{2}{2+\theta}$  respectivement. Maintenant la fonction de densité de  $X$  est donnés par :

$$f_{NZD}(x; \theta) = \begin{cases} \frac{\theta^3 x(1+x)e^{-\theta x}}{2+\theta} & \text{si } x, \theta > 0 \\ 0 & \text{ailleurs} \end{cases} \quad (3-1)$$

**Remarque 3.1.**

Nous pouvons trouver une même distribution en utilisant la technique de la taille biaisée sous la distribution de **Lindley**, voir M. E. **Ghitany** et D. K. **Al-Mutairi** (2008).

Les dérivées première et seconde de  $f_{ZD}$ .

$$\frac{d}{dx}f_{ZD}(x) = -\frac{\theta^3(x\theta - 2x + x^2\theta - 1)e^{-x\theta}}{\theta + 2} = 0$$

donne :

$$x = -\frac{1}{2\theta}(-\theta + 2 + \sqrt{\theta^2 + 4}) \quad (3.2)$$

$$\frac{d^2}{dx^2}f_{ZD}(x) = \frac{\theta^3(x^2\theta^2 + x\theta^2 - 4x\theta - 2\theta + 2)e^{-x\theta}}{\theta + 2}$$

et

$$\frac{d^2}{dx^2}f_{ZD}(x) < 0 \quad (3.3)$$

pour  $\theta > 0$ ,  $\hat{x} = \frac{1}{2\theta}(-\theta + 2 + \sqrt{\theta^2 + 4})$  est le point critique unique que  $f_{NZD}(\hat{x}; \theta)$  est le maximum.

Par conséquent, le mode de ZD est donné par

$$mode(X) = \frac{1}{2\theta}(-\theta + 2 + \sqrt{\theta^2 + 4}) \text{ pour } \theta > 0 \quad (3.4)$$

Nous pouvons trouver facilement la fonction de distribution cumulative (c.d.f) de la ZD

$$f_{ZD}(x) = 1 - \left( \frac{x^2\theta^2 + \theta(\theta + 2)x + \theta + 2}{\theta + 2} \right) e^{-x\theta} \quad x > 0, \theta > 0 \quad (3.5)$$

### 3.1.1 La fonction de Survie et taux de hasard

Soit :

$$\begin{aligned} S_{ZD}(x) &= 1 - F_{ZD}(x) \\ &= \left( \frac{x^2\theta^2 + \theta(\theta + 2)x + \theta + 2}{\theta + 2} \right) e^{-x\theta} ; x, \theta > 0 \end{aligned} \quad (3.6)$$

et

$$h_{ZD}(x) = \frac{f_{ZD}(x)}{1 - F_{ZD}(x)} = \frac{\theta^3 x(x + 1)}{x^2\theta^2 + \theta(\theta + 2)x + \theta + 2} \quad (3.7)$$

les fonction de survie et de taux de risque, respectivement.

**Proposition 3.1.** Soit  $h_{ZD}(x)$  la fonction de taux de risque de  $X$ , alors  $h_{ZD}(x)$  augmente.

**Preuve.** C'est facile de vérifier ça

$$\frac{d}{dx} h_{ZD}(x) = \theta^3 \frac{4x + \theta + 2x\theta + 2x^2\theta + 2}{(x^2\theta^2 + \theta(\theta + 2)x + \theta + 2)^2} > 0$$

## 3.2 Moments et mesures connexes

Le kième moment sur l'origine du ZD est :

$$\mathbf{E}(X^k) = \frac{(k + 1)!(\theta + k + 2)}{\theta^k(\theta + 2)} ; k = 1, 2, \dots \quad (3.8)$$

**Remarque 3.2 :** Le k-ième moment sur l'origine de la distribution de **Lindley** est

$$\mathbf{E}(X^k) = \frac{(k)!(\theta + k + 1)}{\theta^k(\theta + 1)} ; k = 1, 2, \dots \quad (3.9)$$

**Corollaire 3.1**

Soit  $X \sim ZD(\theta)$ , la moyenne et la variance de  $X$  respectivement sont :

$$\mathbf{E}(X) = \frac{2(\theta + 3)}{\theta(\theta + 2)}, \quad \mathbf{Var}(X) = \frac{2(\theta^2 + 6\theta + 6)}{\theta^2(\theta + 2)^2} \quad (3.10)$$

### Théorème 3.1

Soit  $X \sim ZD(\beta, \theta)$ ,  $M = mode(X)$ ,  $me = median(X)$  et  $\mu = \mathbf{E}(X)$ . Alors

$$M < me < \mu \quad (3.11)$$

### Epreuve

$$F(M) = 1 - \frac{2(\sqrt{\theta^2 + 4} + 6)\exp(\frac{1}{2}\theta - \frac{1}{2}\sqrt{\theta^2 + 4} - 1)}{\theta + 2}, \quad F(me) = \frac{1}{2}$$

et

$$F(\mu) = 1 - \frac{3\theta^3 + 24\theta^2 + 68\theta + 68}{(\theta + 2)^3} \exp\left(-\frac{2\theta + 6}{\theta + 2}\right)$$

Notez que :

$$0 < F(M) < 1 - 5e^{-2} < \frac{1}{2}$$

De même  $F(\mu)$  et

$$\frac{1}{2} < 1 - \frac{17}{2}e^{-3} < F(\mu) < 1$$

Il est facile de vérifier que

$$F(M) < F(me) < F(\mu)$$

et finalement, on obtient

$$M < me < \mu$$

Les coefficients de variation  $\gamma$ , l'asymétrie et l'aplatissement du ZD ont été obtenu comme :

$$\begin{aligned}\gamma &= \frac{\sqrt{\mathbf{Var}(X)}}{\mathbf{E}(X)} = \frac{\sqrt{2(\theta^2 + 6\theta + 6)}}{2\theta + 6} \\ l'asymétrie &= \frac{\mathbf{E}(X^3)}{(\mathbf{Var}(X))^{\frac{3}{2}}} = \frac{24(\theta + 2)^2(\theta + 5)}{(2\theta^2 + 12\theta + 12)^{\frac{3}{2}}} \\ l'aplatissement &= \frac{\mathbf{E}(X^4)}{(\mathbf{Var}(X))^2} = \frac{30(\theta + 2)^3(\theta + 6)}{(\theta^2 + 6\theta + 6)}\end{aligned}\quad (3.12)$$

### 3.3 Estimation du paramètre

Soit  $X_i \sim Z(\theta)$ ,  $i = 1 \dots n$ ,  $n$  variables aléatoires. La fonction ln-vraisemblance,  $\ln l(x_i; \theta)$  est :

$$\ln l(x_i; \theta) = 3n \ln \theta - n \ln(\theta + 2) + \sum_{i=1}^n \ln(x_i + x_i^2) - \theta \sum_{i=1}^n x_i \quad (3.13)$$

Les dérivées de  $\ln l(x_i; \theta)$  par rapport à  $\theta$  est :

$$\frac{d \ln l(x_i; \theta)}{d\theta} = \frac{3n}{\theta} - \frac{n}{\theta + 2} - \sum_{i=1}^n x_i$$

A partir de la distribution de **Zeghdoudi** (2), la méthode des moments (MoM) et les estimateurs du maximum de vraisemblance (ML) du paramètre  $\theta$  sont les mêmes et peuvent être obtenus en résolvant l'équation non linéaire suivante :

$$\frac{3}{\theta} - \frac{1}{\theta + 2} - \bar{x} = 0 \Rightarrow \hat{\theta}_M = \hat{\theta}_{ML} = \left\{ \frac{1}{\bar{x}}(-\bar{x} + \sqrt{4\bar{x} + \bar{x}^2 + 1} + 1) \right\} \quad (3.14)$$

Le théorème suivant montre que l'estimateur de  $\theta$  est biaisé positivement.

**Théorème 3.2.** L'estimateur  $\hat{\theta}$  de  $\theta$  est polarisé positivement, c'est-à-dire :

$$E(\hat{\theta}) - \theta > 0 \quad (3.15)$$

### Epreuve

Soit

$$g(\bar{x}) = \hat{\theta}, g(t) = \frac{1}{t} \left( -t + \sqrt{4t + t^2 + 1} + 1 \right)$$

et

$$\frac{d^2g(t)}{dt^2} = \frac{\left( 12t + 2(4t + t^2 + 1)^{\frac{3}{2}} + 15t^2 + 4t^3 + 2 \right)}{t^3(4t + t^2 + 1)^{\frac{3}{2}}} > 0, t > 0$$

$g$  est strictement convexe. Ainsi, par l'inégalité de **Jensen**, nous avons  $\mathbf{E}(g(\bar{x})) > g(\mathbf{E}(\bar{x}))$ . Enfin, puisque

$$g(\mathbf{E}(\bar{x})) = g(\mu) = g\left(\frac{2(\theta + 3)}{\theta(\theta + 2)}\right) = 0$$

on obtient  $\mathbf{E}(\hat{\theta}) > 0$ .

**Théorème 3.3** L'estimateur  $\hat{\theta}$  de  $\theta$  est constant et asymptotiquement normal :

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N\left(0, \frac{1}{\sigma^2}\right) \quad (3.16)$$

L'intervalle de confiance de  $100(1 - \alpha)\%$  de grand échantillon pour  $\theta$  est donné par  $\hat{\theta} \pm$

$Z_{\frac{\alpha}{2}} \sqrt{\frac{1}{n\sigma^2}}$  c-à-d

$$I_{\alpha} = \left[ \hat{\theta} - Z_{\frac{\alpha}{2}} \sqrt{\frac{1}{n\sigma^2}}, \hat{\theta} + Z_{\frac{\alpha}{2}} \sqrt{\frac{1}{n\sigma^2}} \right]$$

La preuve est omise car elle est très similaire à la preuve du théorème 4, **Ghitany** (2008b).

**Remarque 3.3.** Pour la distribution de **Lindley**,  $\hat{\theta} = \frac{1}{2\bar{x}}(-\bar{x} + \sqrt{6\bar{x} + \bar{x}^2 + 1} + 1)$  est la vraisemblance maximale et l'estimateur de la méthode des moments de  $\theta$ . Il est également biaisé positivement, cohérent et asymptotiquement normal.

### 3.3.1 Ordres stochastiques

**Définition 3.1.** Considérons deux variables aléatoires  $X$  et  $Y$ . Alors  $X$  est dit plus petit que  $Y$  dans le :

- a) Ordre stochastique ( $X \prec_s Y$ ), si  $F_X(t) \geq F_Y(t), \forall t$
- b) Ordre convexe ( $X \prec_{cx} Y$ ), si pour toutes les fonctions convexes  $\phi$  et fourni l'attente existe,  $\mathbf{E}[\phi(X)] \leq \mathbf{E}[\phi(Y)]$ .
- c) Ordre du taux de risque (Haz)  $X \prec_{cx} Y$ , si  $h_X(t) \leq h_Y(t) \forall t$
- d) Ordre du rapport de vraisemblance ( $X \prec_{lr} Y$ ), si  $\frac{f_X(t)}{f_Y(t)}$  est décroissant dans  $t$ .

**Remarque 3.4** Ordre du rapport de vraisemblance  $\Rightarrow$  Ordre du taux de risque  $\Rightarrow$  Ordre stochastique.

Si  $E[X] = E[Y]$  alors l'ordre convexe  $\Leftrightarrow$  ordre stochastique.

**Théorème 3.4.** Soit  $X_i \sim Z(\theta_i), i = 1, 2$  deux variables aléatoires. Si  $\theta_1 \geq \theta_2$ , alors

$$X_1 \prec_{lr} X_2, X_1 \prec_{hr} X_2, X_1 \prec_s X_2 \text{ et } X_1 \prec_{cx} X_2 \quad (3.17)$$

**Preuve.** Nous avons

$$\frac{f_{X_1}(t)}{f_{X_2}(t)} = \frac{\theta_1^3(\theta_2 + 2)}{\theta_2^3(\theta_1 + 2)} e^{-(\theta_1 - \theta_2)t}$$

Pour simplifier, nous utilisons  $\ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right)$ , maintenant, nous pouvons trouver

$$\frac{d}{dt} \ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right) = -(\theta_1 - \theta_2)$$

A cette fin, si  $\theta_1 \geq \theta_2$ , on a  $\frac{d}{dt} \ln\left(\frac{f_{X_1}(t)}{f_{X_2}(t)}\right) \leq 0$ .

Cela signifie que  $X_1 \prec_{lr} X_2$  Aussi, selon la remarque 3, le théorème est prouvé.

### 3.4 Ecart moyen

Il s'agit de deux écarts moyens : à propos de la moyenne et de la médiane, définis comme  $MD_1 = \int_0^\infty |x - \mu| f(x) dx$  et  $MD_2 = \int_0^\infty |x - me| f(x) dx$  respectivement, où  $\mu = \mathbf{E}(X)$  et  $me = Median(X)$ . Les mesures  $MD_1$  et  $MD_2$  peuvent être calculées en utilisant les formules simplifiées suivantes :

$$\begin{aligned} MD_1 &= 2\mu F(\mu) - 2 \int_0^\infty x f(x) dx \\ MD_2 &= \mu - 2 \int_0^{me} x f(x) dx \end{aligned} \quad (3.18)$$

### 3.5 Statistiques de commande extrêmes

Si  $X_1, \dots, X_n$  est un échantillon aléatoire de (3.1) et si  $\bar{X} = \frac{X_1 + \dots + X_n}{n}$  désigne l'échantillon signifie alors par le théorème de la limite centrale habituelle  $\frac{\sqrt{n}(\bar{X} - \mathbf{E}(X))}{\sqrt{\mathbf{Var}(X)}}$  se rapproche de la distribution normale standard comme  $n \rightarrow \infty$ . Parfois on serait intéressé par l'asymptotique des valeurs extrêmes  $M_n = \max(X_1, \dots, X_n)$  et  $m_n = \min(X_1, \dots, X_n)$ . Pour le c.d.f. en (3.5), on peut voir que

$$\lim_{t \rightarrow \infty} \frac{1 - F(t+x)}{1 - F(t)} = e^{-\theta x} \quad (3.19)$$

et

$$\lim_{t \rightarrow 0} \frac{F(tx)}{F(t)} = x \quad (3.20)$$

Ainsi, il ressort du théorème 1.6.2 de **Leadbetter et al.** (1987) doit être des constantes de normalisation  $a_n, b_n, c_n > 0$  et  $d_n$  telque :

$$P(a_n(M_n - b_n) \leq x) \xrightarrow[n \rightarrow \infty]{} e^{-\theta x} \quad (3.21)$$

La forme des constantes normatives peut également être déterminée. Par exemple, en utilisant le corollaire 1.6.3 dans **Leadbetter et al** (1987), on peut voir que  $a_n = 1$  et  $b_n = F^{-1}(1 - \frac{1}{n})$  avec  $F(\cdot)$  donné par (3.5).

## 3.6 Simulation et Ajustement

### 3.6.1 Étude de simulation

Nous pouvons voir que l'équation  $F(x) = u$ , où  $u$  est une observation de la distribution uniforme sur  $(0, 1)$ , ne peut pas être résolu explicitement en  $x$  (ne peut pas utiliser la fonction **Lambert** (W), la méthode d'inversion pour générer des données aléatoires de la distribution **Zeghdoudi** échoue. Cependant, nous pouvons utiliser le fait que la distribution de **Zeghdoudi** est un mélange de distributions gamma  $(3, \theta)$  et gamma  $(2, \theta)$  :

$$f_{ZD}(x; \theta) = \frac{\theta}{\theta + 2} \text{gamma}(2, \theta) + \frac{2}{\theta + 2} \text{gamma}(3, \theta) \quad (3.22)$$

Dans cette sous-section, nous étudions le comportement des estimateurs de la MV pour une taille d'échantillon finie (n). Une étude de simulation comprenant les étapes

suivantes est effectuée  $N = 10000$  fois pour les valeurs sélectionnées de  $(\theta; n)$  où  $\theta = 0.01, 0.1, 0.5, 1, 3, 10$  et  $n = 10, 30, 50$

- Produire  $U_i \sim Uniform(0, 1), i = 1, \dots, n$

- Produire  $Y_i \sim Gamma(2, \theta), i = 1, \dots, n$

- Produire  $Z_i \sim Gamma(3, \theta), i = 1, \dots, n$

Si  $U_i \leq p = \frac{\theta}{\theta+2}$ , alors mettre  $X_i = Y_i$ , sinon, mettre  $X_i = Z_i, i = 1, \dots, n$

$$Moyenne\ bias(\theta) = \frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta) \quad (3.24)$$

et l'erreur quadratique moyenne (voir le tableau 1 et le tableau 2)

$$MSE(\theta) = \frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta)^2 \quad (3.25)$$

	$\theta = 0.01$	$\theta = 0.1$	$\theta = 0.5$
	$bias(\theta)$	$bias(\theta)$	$bias(\theta)$
$n = 10$	0.0015	0.0021	0.0147
$n = 30$	0.0008	0.0016	0.0103
$n = 50$	0.0005	0.0010	0.0081
	$\theta = 1$	$\theta = 3$	$\theta = 10$
	$bias(\theta)$	$bias(\theta)$	$bias(\theta)$
$n = 10$	0.0275	0.283	0.358
$n = 30$	0.0133	0.094	0.186
$n = 50$	0.0079	0.056	0.084

Table1. Average bias of the simulated estimates

	$\theta = 0.01$	$\theta = 0.1$	$\theta = 0.5$
	$MSE(\theta)$	$MSE(\theta)$	$MSE(\theta)$
$n = 10$	0.00014	0.00047	0.011
$n = 30$	0.00007	0.00024	0.003
$n = 50$	0.00003	0.00013	0.002
	$\theta = 1$	$\theta = 3$	$\theta = 10$
	$MSE(\theta)$	$MSE(\theta)$	$MSE(\theta)$
$n = 10$	0.020	0.802	3.667
$n = 30$	0.011	0.267	1.222
$n = 50$	0.007	0.132	0.133

Table2. Average MSE of the simulated estimates

Les tableaux 3 et 4 affichent le biais et l'erreur quadratique moyenne du maximum de vraisemblance estimateurs des paramètres.

### 3.6.2 Applications

#### Exemple 3.1

Les tableaux 3 et 4 représentent les données de survie (en mois) de (94,91) des cobayes infectés par le virus Ebola, que l'on compare à la distribution **Zeghdoudi** et distribution de **Lindley**.

durée de survie $m = 3.17, s = 2.095$	Obs freq	LD $\hat{\theta} = 0.522$	ZD $\hat{\theta} = 0.852$
[0, 2[	<b>32</b>	<b>38.217</b>	<b>33.252</b>
[2, 4[	<b>35</b>	<b>28.16</b>	<b>32.366</b>
[4, 6[	<b>17</b>	<b>15.089</b>	<b>17.799</b>
[6, 8[	<b>7</b>	<b>7.33</b>	<b>7.133</b>
[8, 10[	<b>3</b>	<b>3.152</b>	<b>2.418</b>
Total	<b>94</b>	<b>94</b>	<b>94</b>
$\chi^2$	—	<b>2.9244</b>	<b>0.40020</b>

Table3

durée de survie $m = 3$	Obs freq	LD $\hat{\theta} = 0.54858$	ZD $\hat{\theta} = 0.852$
[0, 2[	<b>35</b>	<b>39.100</b>	<b>34.56</b>
[2, 4[	<b>32</b>	<b>27.390</b>	<b>31.497366</b>
[4, 6[	<b>16</b>	<b>13.92</b>	<b>16.206</b>
[6, 8[	<b>6</b>	<b>6.2475</b>	<b>6.0697</b>
[8, 10[	<b>2</b>	<b>2.6189</b>	1.9218
Total	<b>91</b>	<b>91</b>	<b>91</b>
$\chi^2$	—	<b>1.6165</b>	<b>0.01995</b>

Table 4.

Données	Distribution	$\theta$	Log-likelihood	Kolmogrov-Smirnov
Données 1	<b>ZD</b>	<b>1.365</b>	<b>-52.4</b>	<b>0.292</b>
$n = 20$	<b>Aradhana</b>	<b>1.123</b>	<b>-56.4</b>	<b>0.302</b>
$m = 1.9$	<b>Akash</b>	<b>1.157</b>	<b>-59.5</b>	<b>0.320</b>
$s = 0.704$	<b>Shanker</b>	<b>0.804</b>	<b>-59.7</b>	<b>0.315</b>
Données 2	<b>ZD</b>	<b>0.095</b>	<b>-239.7</b>	<b>0.251</b>
$n = 25$	<b>Aradhana</b>	<b>0.094</b>	<b>-242.2</b>	<b>0.274</b>
$m = 30.811$	<b>Akash</b>	<b>0.097</b>	<b>-240.7</b>	<b>0.266</b>
$s = 7.253$	<b>Shanker</b>	<b>0.063</b>	<b>-252.3</b>	<b>0.326</b>

Table 5.

**Exemple 3.2**

Maintenant, nous comparons un paramètre (**Aradhana, Akash, Shanker**) voir **Shanker** (2015a, 2015b, 2016) avec la distribution de **Zeghdoudi** (voir tableaux 5 et 6).

Données	Distribution	$\beta$	$\theta$	Log-likelihood	Kolmogrov-Smirnov
Données 3	<b>ZD</b>	–	<b>1.365</b>	<b>58.67</b>	<b>0.081</b>
$n = 15$	<b>Gamma</b>	<b>1.442</b>	<b>1.123</b>	<b>–64.197</b>	<b>0.102</b>
$m = 27.546$	<b>Weibull</b>	<b>1.306</b>	<b>1.157</b>	<b>–64.026</b>	<b>0.450</b>
$s = 20.059$	<b>Lognormal</b>	<b>1.061</b>	<b>0.804</b>	<b>–65.626</b>	<b>0.163</b>
Données 4	<b>ZD</b>	–	<b>0.095</b>	<b>–142.32</b>	<b>0.108</b>
$n = 25$	<b>Gamma</b>	<b>1.794</b>	<b>0.094</b>	<b>–152.371</b>	<b>0.135</b>
$m = 178.32$	<b>Weibull</b>	<b>1.414</b>	<b>0.097</b>	<b>–152.440</b>	<b>0.697</b>
$s = 131.097$	<b>Lognormal</b>	<b>0.891</b>	<b>0.063</b>	<b>–154.092</b>	<b>0.155</b>

Table 6.

## Conclusion et Perspectives

Ainsi, nous avons réussi à introduire une nouvelle distribution nommée "distribution de **Zeghdoudi**", ensuite, à étudier quelques propriétés à savoir : la fonction quantile, la courbe de Lorenz, l'estimation par la méthode des moments, l'estimation du maximum de vraisemblance et les distributions limites des statistiques d'ordre. La méthode des moments et la méthode d'estimation du maximum de vraisemblance ont également été discutées.

Enfin, le test d'adéquation utilisant les statistiques K-S (statistiques de **Kolmogorov-Smirnov**) pour quatre séries de données réelles a été présenté pour démontrer l'applicabilité et comparabilité des distributions **Zeghdoudi**, Gamma, **Weibull**, lognormales, **Aradhana**, **Akash**, **Shanker**, **Lindley** et exponentielles pour la modélisation des données de durée de vie.

Pour les études futures, nous pouvons expliquer la dérivation des distributions postérieures pour la distribution de **Zeghdoudi** sous les fonctions de perte de Linex et erreur au carré en utilisant des a priori non informatifs et informatifs (l'extension de **Jeffreys** et priorités gamma inversées) respectivement.

### 3.7 Annexe

Dans cet article nous avons utilisé quatre données, la première représente les données de la vie concernant les temps de soulagement (en minutes) de 20 patients re-

cevant un analgésique et rapportés par **Gross et Clark** (1975, P. 105). Le deuxième est la force données de verre de la fenêtre de l'avion rapportées par **Fuller et al** (1994). le troisième représente les temps d'échec (mm) pour un échantillon de .fifteen électronique composants dans un test de durée d'accélération (**Lawless** (2003), pp 204).

Le dernier, sont le nombre de cycles à l'échec pour 25 échantillons de 100 cm de fil, testés à un niveau de contrainte particulier (**Lawless** (2003), pp. 263).

Ensemble de données 1 : 1.1, 1.4, 1.3, 1.7, 1.9, 1.8, 1.6, 2.2, 1.7, 2.7, 4.1, 1.8, 1.5, 1.2,1.4, 3.0, 1.7, 2.3, 1.6, 2.0.

Ensemble de données 2 : 18.83, 20.80, 21.657, 23.03, 23.23, 24.05, 24.321, 25.50, 25.52, 25.80, 26.69, 26.77, 26.78, 27.05,27.67, 29.90, 31.11, 33.20, 33.73, 33.76, 33.89, 34.76, 35.75, 35.91, 36.98, 37.08, 37.09, 39.58, 44.045, 45.29, 45.381

Ensemble de données 3 : 1.4, 5.1, 6.3, 10.8, 12.1, 18.5, 19.7, 22.2, 23, 30.6, 37.3, 46.3,53.9, 59.8, 66.2.

Ensemble de données 4 : 15, 20, 38, 42, 61, 76, 86, 98, 121, 146, 149, 157, 175, 176,180, 180, 198, 220, 224, 251, 264, 282, 321, 325, 653.

## Bibliographie

- [1] Asgharzadeh Hassan A., Bakouch,S.,& Esmaeili ,L. (2013). Pareto poisson-  
**Lindley** distribution and its application, *Journal of Applied Statistics*, pp. 1-18.
- [2] C. Dagum, Lorenz curve, in : S. Kotz, N.L. Johnson, C.B. Read (Eds.), in :  
*Encyclopedia of Statistical Sciences*, vol. 5,Wiley, New York, 1985,pp. 156–161.
- [3] E.M. L emeray, Racines de quelques  equations transcendantes. Int egration d’une  
 equation aux diff erences m el ees. Racines imaginaires,Nouvelles Ann. Math. 16  
(1897) 540–546.
- [4] F. Lawless, *Statistical Models and Methods for Lifetime Data*, Wiley, New York,  
200
- [5] Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *J. Amer.*  
*Statist. Assoc.*, 75, 667-672
- [6] H. S. Bakouch, B. M. Al-Zahrani, A. A. Al-Shomrani, V. A. Marchi and F. Lou-  
zada (2012). An extended **Lindley** distribution, *Journal of the Korean Statistical*  
*Society*, Vol 41, 75-85.

- [7] Hussain, E. The Non-Linear Functions of Order Statistics and their Properties in Selected Probability Models (Doctoral dissertation, University of Karachi, Karachi) (2008).
- [8] H. Zakerzadah, A. Dolati (2010). Generalized **Lindley** distribution. J. Math. Ext. 3(2),pp. 13-25..
- [9] H. **Zeghdoudi**, S. **Nedjar** (2016a). Gamma **Lindley** distribution and its application. *Journal of Applied Probability and Statistics* Vol. **11**, N° **1**, 129-138
- [10] H. **Zeghdoudi**, S. **Nedjar** (2016b) A Pseudo **Lindley** distribution and its application., J. Afrika Statistika Vol 11 (1) page 923\_932.
- [11] H. **Zeghdoudi**, S. **Nedjar** (2016c). On Gamma **Lindley** distribution :properties and simulation. *Journal of Computational and Applied Mathematics* 298, pp167-174.
- [12] H. **Zeghdoudi**, S. **Nedjar** (2017). A Poisson pseudo **Lindley** distribution and its application . *Journal of Probability and Statistical Sciences* . Vol. 15, N° 1.
- [13] H. **Zeghdoudi**, S. **Nedjar** (2017). On Pseudo **Lindley** Distribution : Properties and Applications . *Journal of New Trends in Mathematical Sciences*
- [14] J. F. Lawless , 2003. Statistical models and methods for lifetime data. Wiley, New York pp 204 - 263.
- [15] Jodrá, P. (2010). Computer generation of random variables with **Lindley** or

- Poisson **Lindley** distribution via the **Lambert** W function. Mathematics and Computers in Simulation (MATCOM), Vol. 81, issue 4, pages 851-859.
- [16] **Lindley**, D. V. (1958). Fiducial distributions and bayes theorem. Journal of the Royal Society, series B, 20, 102-107
- [17] M. E. Ghitany, B. Atieh, S. Nadarajah (2008a). **Lindley** distribution and its applications. *Math. Comput. Simulation* , **78**, pp. 493-506 .
- [18] M..E. Ghitany, D.K. Al-Mutairi, S. Nadarajah (2008b). Zero-truncated Poisson–**Lindley** distribution and its application, *Math. Comput.*
- [19] M..E.Ghitany.D.K Almutairi(2009) : Estimation methods for the discrete Poisson–**Lindley** distribution, Journal of Statistical Computation and Simulation, 79 :1, 1-9
- [20] M. H. Gail and J. L.Gastwirth (1978). A scale-free goodness of fit test for the exponential distribution based on the Lorenz curve. *Journal of the American Statistical Association* **73** 787–793.
- [21] M..R. Leadbetter, G. Lindgren, H. Rootz´en, Extremes and Related Properties of Random Sequences and Processes, Springer Verlag, New York,1987.
- [22] M. Sankaran (1970). The discrete Poisson-**Lindley** distribution. *Biometrics*, **26**, pp. 145-149.
- [23] M. Shaked, J.G. Shanthikumar, Stochastic Orders and Their Applications, Academic Press, New York, 1994..

- [24] R.Shanker and A. Mishra (2013). A quasi **Lindley** distribution. African Journal of Mathematics and Computer Science Research, Vol. 6(4), pp. 64-71.
- [25] R.Shanker, S.Sharma, R.Shanker. (2013). A two-parameter **Lindley** distribution for modeling waiting and survival times data. *Applied Mathematics*, vol. 4, pp. 363-368.
- [26] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth (1996). On the **Lambert** W function, *Adv. Comput. Math.* 5 329–359.