



## THESE

Présentée en vue de l'obtention du diplôme de

Doctorat en Sciences

Spécialité : Mathématiques

Option : Probabilité et Statistique

**Une distribution de gamma Lindley généralisée:  
Propriétés et applications**

Par:

***Benatmane Chaabane***

Sous la direction de

***ZEGHDOUDI Halim***

Prof. U.B.M. Annaba

Devant le jury

**PRESIDENT :** Hacem Boutabia

Prof. U.B.M. Annaba

**EXAMINATEUR :** Brahimi Brahim

Prof. U. Biskra

**EXAMINATEUR :** Yahia Djabrane

Prof. U. Biskra

# TABLE DES MATIÈRES

<b>Résumé en Anglais</b>	iv
<b>Résumé en Français</b>	v
<b>Résumé en Arabe</b>	vi
<b>Remerciements</b>	vii
<b>Introduction générale</b>	1
<b>1 Généralités sur quelques distributions</b>	1
<b>1.1 Distribution de Gamma</b> . . . . .	2
<b>1.1.1 Distribution de Gamma standard</b> . . . . .	2
<b>1.1.2 Distribution gamma de Pearson</b> . . . . .	4
<b>1.1.3 Distribution Gamma de Stacy</b> . . . . .	6
<b>1.2 Distribution de Pareto</b> . . . . .	8
<b>1.3 Distribution de Rayleigh (Lord Rayleigh)</b> . . . . .	10
<b>1.4 Distribution de Lindley (1958)</b> . . . . .	11
<b>1.5 Distribution de Lindley avec deux paramètres</b> . . . . .	12

<b>2 Modèles combinés, T-X et Mélange</b>	<b>13</b>
2.1 Modèles combinés . . . . .	14
2.2 Modèle $T - X$ . . . . .	17
2.3 Modèle de Mélange . . . . .	20
2.4 Distribution de Lindley Généralisée . . . . .	20
2.4.1 Estimation par la méthode des moments . . . . .	21
2.5 Distribution Pseudo Lindley . . . . .	22
2.5.1 Moments et mesures connexes . . . . .	23
2.5.2 Fonction de hasard et Fonction de survie . . . . .	24
2.5.3 Fonction Quantile de la distribution de Pseudo Lindley . . . . .	25
2.5.4 Estimation par la méthode des moments (MoM) . . . . .	25
2.6 Distribution Quasi Lindley . . . . .	28
2.6.1 Fonction de hasard et fonction de survie . . . . .	30
2.6.2 Fonction Quantile de la distribution Quasi Lindley . . . . .	30
2.6.3 Estimation par la méthode des moments (MoM) . . . . .	32
2.7 Distribution de Gamma Lindley . . . . .	32
2.7.1 Fonction Quantile de la distribution de Gamma Lindley . . . . .	34
2.7.2 Estimation par la méthode des moments (MoM) . . . . .	35
2.8 Comparaison des distributions . . . . .	35
<b>3 Distribution composite Rayleigh-Pareto et ses applications</b>	<b>38</b>
3.1 Le modèle composite Pareto-Rayleigh . . . . .	39
3.2 Génération de valeurs aléatoires à partir de la distribution composite de Rayleigh-Pareto . . . . .	41
3.2.1 Estimation des paramètres . . . . .	41
3.2.2 Le premier algorithme Une procédure ad-hoc basée sur les centiles . . . . .	41
3.2.3 Le deuxième algorithme . . . . .	42

Table des matières

---

<b>3.3 Exemples numériques</b> . . . . .	43
<b>3.3.1 Exemple</b> . . . . .	44
<b>3.3.2 Qualité de l'ajustement</b> . . . . .	46
<b>Conclusion générale et perspectives</b>	49
<b>3.4 Annexe</b> . . . . .	50
<b>3.4.1 Sélection de model avec différents critères</b> . . . . .	50
<b>3.4.2 Fonction W de Lambert</b> . . . . .	52
<b>Bibliographie</b>	53

This thesis introduces a new probability densitie called" **composite Rayleith-Pareto distribution**", which will add some value to the existing literature on the probability theory and the modeling of the data for life and actuarial sciences. Various statistical properties, 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.

---

**Keywords** : Gamma distribution, Lindley distribution, Pareto distribution, Rayleith distribution.

Nous nous proposons, dans cette thèse, d'introduire une nouvelle densité de probabilité appelée distribution **composite de Rayleith-Pareto**, qui ajouteront de la valeur à la littérature existant sur la théorie des probabilités et la modélisation des données à la vie et la science actuarielle. Diverses propriétés statistiques ont été obtenues. Une application sur des ensembles de données d'assurance incendie (les pertes d'assurance incendie algériennes et danoises) a été établie pour montrer l'importance et l'applicabilité de cette nouvelle distribution.

---

**Mots clés :** Distribution de Gamma, Distribution de Lindley, Distribution de Pareto, Distribution de Rayleith.

ملخص

نقدم في هذه الأطروحة كثافة احتمالية جديدة تسمى توزيع رايليث-باريتو المركب والتي ستضيف إلى الأبجديات الحالية حول نظرية الاحتمالات و نمذجة البيانات في الحياة و العلوم الإكتوارية. كذلك تم الحصول على خصائص احصائية مختلفة. و اكثر من هذا قمنا بتطبيق هذا النموذج على مجموعة بيانات التأمين ضد الحرائق ، خسائر التأمين ضد الحرائق الجزائرية و الدنماركية، لإظهار أهمية وإمكانية تطبيق هذا النموذج الجديد.

الكلمات المفتاحية: توزيع غاما. توزيع ليندلي. توزيع باريتو. توزيع رايت-لايث

---

## 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 encadreur **Pr. Zeghdoudi Halim** 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 **Pr Boutabia Hacene** de l'université **Badji Mokhtar Annaba** qui a accepté d'être le président de ce jury.

Je remercie vivement **Pr. Brahim Brahimi** et **Pr. Yahia Djabrane** de l'université de **Mouhamed Khidhar Biskra**, pour l'honneur d'avoir accepté de faire partie du jury. Je m'estime très honoré par l'intérêt de qu'ils ont bien voulu accorder à mon travail et leur sauraient gré pour toutes remarques, qui va m'aider pour voir les ponts qui pouvaient exister entre mon travail et d'autres perspectives mathématiques.

Je remercie mon père, ma mère, ma femme, mes fils et mes frères.

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



---

## Introduction générale

Les distributions statistiques sont couramment utilisées pour décrire les phénomènes du monde réel. En raison de l'utilité des distributions statistiques, leur théorie est largement étudiée et de nouvelles distributions sont développées. L'intérêt pour le développement de distributions statistiques plus souples reste important dans la profession statistique. De nombreuses classes de distributions généralisées ont été développées et appliquées pour décrire divers phénomènes.

Par exemple, la distribution Gamma est la distribution la plus courante pour l'analyse de données de durée de vie. Elle a de nombreuses applications dans différents domaines autres que les distributions de durée de vie. Les deux paramètres d'une distribution gamma représentent les paramètres d'échelle et de forme et il est assez flexible pour analyser des données réelles positives. Le taux d'échec augmente et diminue en fonction du paramètre de forme, ce qui donne un avantage supplémentaire sur la distribution exponentielle, qui a un taux d'échec constant. Puisque la somme des variables aléatoires indépendantes et identiquement distribuées a une distribution gamma, elle a également une belle interprétation physique. Une autre propriété intéressante de la famille des distributions gamma à deux paramètres est qu'elle dispose

d'un ordre de rapport de vraisemblance, par rapport au paramètre de forme, lorsque le paramètre d'échelle reste constant. Cela implique naturellement la commande en taux de risque ainsi qu'en distribution.

La distribution gamma généralisée avec le paramètre de forme  $\alpha$  et le paramètre d'échelle  $\beta$  spécifiée par la fonction de densité de probabilité (p.d.f.)

$$f_{GG(\alpha,\beta,\gamma)}(x) = \frac{\gamma\beta^{\gamma\alpha}}{\Gamma(\alpha)} x^{\gamma\alpha-1} \exp(-(\beta x)^\gamma), \quad x, \alpha, \beta, \gamma > 0, \quad (1)$$

où

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} \exp(-\alpha x) dx \quad (2)$$

est la fonction gamma.

La distribution gamma [\(1,2\)](#) avec le paramètre de forme  $\alpha$  et le paramètre d'échelle  $\beta$  est un cas particulier de la distribution gamma généralisée si  $\gamma = 1$

$$f_{G(\alpha,\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad x, \alpha, \beta > 0. \quad (3)$$

La distribution de Lindley spécifiée par la fonction de densité de probabilité (p.d.f) donnée par

$$f_{L(\beta)}(x) = \frac{\beta^2}{1+\beta} (1+x) \exp(-\beta x), \quad x, \beta > 0, \quad (4)$$

a été présenté par Lindley (1958).

En raison de l'existence d'un seul paramètre, la distribution de Lindley ne fournit pas suffisamment de flexibilité pour analyser différents types de données de survie. Pour augmenter la flexibilité à des fins de modélisation, il serait utile d'envisager d'autres alternatives de cette distribution.

Ghitany et al (2011) et Ghitany et Al-Mutairi (2013) introduisent deux nouvelles classes de distributions de Lindley pondérée, ce qui donne plus de flexibilité à la distribution de Lindley. Les fonctions de densités sont données

par

$$f_{WL(\alpha,\beta)}(x) = \frac{\beta^{\alpha+1}}{(\alpha + \beta) \Gamma(\alpha)} x^{\alpha-1} (1 + x) \exp(-\beta x); x, \alpha, \beta > 0 \quad (5)$$

$$f_{PL(\alpha,\beta)}(x) = \frac{\gamma \beta^{2\gamma}}{(1 + \beta^\gamma)} x^{\gamma-1} (1 + x^\gamma) \exp(-\beta x^\gamma), x, \alpha, \beta, \gamma > 0. \quad (6)$$

Ramos and Louzada (2016) a proposé une nouvelle distribution de Lindley pondérée à trois paramètres, nommée la distribution généralisée de Lindley pondérée (GLW). La fonction de densité de probabilité est donnée par

$$f_{GWL(\alpha,\beta,\gamma)}(x) = \frac{\gamma \beta^{\gamma\alpha}}{(\alpha + \beta) \Gamma(\alpha)} x^{\gamma\alpha-1} (\beta + (\beta x)^\gamma) e^{-(\beta x)^\gamma}, x, \alpha, \beta, \gamma > 0. \quad (7)$$

Bouchahed et Zeghdoudi (2018) a proposé une nouvelle distribution de Lindley à un paramètre, nommée la distribution généralisée de Zeghdoudi dont la fonction de densité est donnée par

$$f_{GZ}(x) = \frac{\sum_{k=0}^n a_k x^k}{\sum_{k=0}^n a_k \frac{k!}{\theta^{k+1}}} \exp(-\theta x), x, \theta > 0. \quad (8)$$

Dans ce travail, des nouvelles distributions de durée à la vie ont été proposées à savoir : Lindley Généralisée, Pseudo Lindley, Quasi Lindley et Gamma Lindley. Par ailleurs, une comparaison des distributions est obtenue.

D'autre part, le domaine de la science actuarielle utilise fréquemment les distributions de Rayleigh et de Pareto pour modéliser des données de paiement. La distribution Rayleigh a de nombreuses applications dans la vie réelle pour tester la durée de vie d'un objet que la durée de vie dépend de son âge. La distribution de Rayleigh est souvent utilisée dans différents domaines de la physique pour modéliser la hauteur des vagues, le rayonnement son et lumière, les signaux radio et l'énergie éolienne, l'image par ultrasons ; durée de vie des tubes, résistances, réseaux, cristaux, boutons, transformateurs, relais et condensateurs des radars d'aéronefs, vitesse du vent sur un an sur les

sites d'éoliennes, vitesse moyenne quotidienne du vent, etc. La distribution de Rayleigh a également fait l'objet d'une attention précieuse le domaine de la fiabilité, de l'analyse de survie et de la science actuarielle. Mais pour la modélisation des données en actuariat, nous ne pouvons pas utiliser une seule distribution, en particulier si nous étudions les sinistres affectant un portefeuille d'assurance tel que l'assurance automobile, les données de paiement, etc. La combinaison de densités est très utile pour modéliser cette répartition des sinistres.

Plusieurs chercheurs utilisent la distribution généralisée de Pareto pour modéliser les données de paiement (données de perte importante ou paiements de réassurance), voir Cooray et M. Ananda (2005). En effet, les données de paiement des assurances sont généralement très asymétriques et distribuées avec de grandes queues hautes. L'inconvénient de l'utilisation du modèle de Pareto pour les données actuarielles est qu'il couvre bien le comportement des pertes importantes, mais pas les pertes mineures, qui sont assez bien modélisées par les distributions de Rayleigh, Log-normal, Gamma ou Weibull. Dans certains cas, les problèmes de la science, nous ne pouvons pas utiliser une seule distribution. Par exemple, si nous étudions les sinistres affectant un certain portefeuille d'assurance (assurance auto, donnée de paiement, ...). Pour cela, la répartition des revendications peut être modélisée comme une combinaison de densités.

Le présent travail est basé sur les idées de Cooray et Ananda (2005), Teodorescu et Vernic (2006), Preda et Ciumara (2006) et Scollnik (2007). Un composite de la distribution de Rayleigh et de Pareto nommé modèle de Rayleigh-Pareto a été proposé et étudié. En général, lorsqu'on considère les distributions composites, la première distribution est la distribution à queue légère et la seconde distribution est la distribution à queue épaisse. Dans le

modèle de Rayleigh-Pareto, la densité de Rayleigh a été prise comme  $f$ , car il s'agit d'une distribution à queue claire et la distribution de Pareto a été prise comme  $f_2$  parce qu'il s'agit d'une distribution à queue épaisse. Certaines propriétés du modèle composite de Rayleigh-Pareto pouvant être appliquées à des applications pratiques dans des ensembles de données du monde réel ont été examinées.

Ce projet de recherche est structuré de la manière suivante :

Le premier chapitre est consacré aux rappels des certaines définitions et certains résultats sur les distributions de probabilités qui nous utilisons par la suite (Gamma, Rayleigh, Pareto, Lindley, Lindley de deux paramètres ...).

Dans le deuxième chapitre, nous donnons une bref rappelle sur les modèles combinés, transformé-transformer et mélange dont nous introduisons des nouvelles distributions de probabilités, ces distributions résultent d'un mélange linéaire convexe de distributions Gamma et Lindley. En effet, nous donnons quelques propriétés tel que la fonction de masse, les moments. Nous nous intéressons aussi à l'estimation des paramètres de ces distributions utilisant la méthode des moments. Nous finalisons ce chapitre avec une étude comparative par des données réelles.

Enfin, le dernier chapitre contient l'essentiel des travaux regroupés dans l'article (*Composite Rayleigh-Pareto distribution : Application to real fire insurance losses data set. Journal of Statistics and Management Systems.*) en nous inspirant des travaux originaux de Cooray et Ananda (2005), Teodorescu et Vernic (2006), Preda et Ciumara (2006) et Scollnik (2007) en adaptant certains résultats. Par ailleurs, une simulation et une étude comparative entre plusieurs distributions ont été données.

# CHAPITRE 1

## Généralités sur quelques distributions

### Sommaire

---

<a href="#">1.1 Distribution de Gamma</a> . . . . .	2
<a href="#">1.2 Distribution de Pareto</a> . . . . .	8
<a href="#">1.3 Distribution de Rayleigh (Lord Rayleigh)</a> . . . . .	10
<a href="#">1.4 Distribution de Lindley (1958)</a> . . . . .	11
<a href="#">1.5 Distribution de Lindley avec deux paramètres</a> . . . . .	12

---

**D** De nombreuses situations pratiques peuvent être modélisées à l'aide de variables aléatoires qui sont régies par des lois spécifiques. Il importe donc d'étudier ces modèles probabilistes qui pourront nous permettre par la suite d'analyser les fluctuations de certains phénomènes en évaluant, par exemple, les probabilités que tel événement ou tel résultat soit observé. La connaissance de ces lois théoriques possède plusieurs avantages sur le plan pratique :

- Les observations d'un phénomène particulier peuvent être remplacées par l'expression analytique de la loi où figure un nombre restreint de paramètres (1 ou 2, rarement plus).

- La loi théorique agit comme modèle (idéalisation) et permet ainsi de réduire les irrégularités de la distribution empirique. Ces irrégularités sont souvent inexplicables et proviennent de fluctuations d'échantillonnage, d'imprécision d'appareils de mesure ou de tout autre facteur incontrôlé ou incontrôlable.

- Des tables de probabilités ont été élaborées pour les lois les plus importantes. Elles simplifient considérablement les calculs.

Ce chapitre contient quelques lois de probabilité et certains résultats que nous utiliserons par la suite. En effet, des généralités sur quelques distributions de probabilités ont été élaborées à savoir : Gamma, distribution de Pareto, distribution de Rayleigh, et distribution de Lindley.

## 1.1 Distribution de Gamma

### 1.1.1 Distribution de Gamma standard

**Définition 1.1.1** Une variable aléatoire  $X$  suit une loi de Gamma de paramètre  $\alpha \in \mathbb{R}_+$ , si elle admet pour densité de probabilité *pdf* la fonction :

$$f_{G(\alpha)}(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x); x > 0 \quad (1.1)$$

ceci est noté  $X \sim \text{Gamma}(\alpha)$ .

- Si  $\alpha = 1$  la distribution  $\text{Gamma}(1)$  est la distribution exponentielle de paramètre 1.
- la somme de deux variables aléatoires indépendantes suit la même loi de Gamma  $\text{Gamma}(\alpha_1)$  et  $\text{Gamma}(\alpha_2)$  est une variable aléatoire de

$Gamma(\alpha_1 + \alpha_2)$ .

- Le produit de deux variables aléatoires de Gamma  $Gamma(\alpha_1)$  et  $Gamma(\alpha_2)$  est une variable aléatoire de Bessel  $K_\alpha(x)$

$$f_{X_1 X_2}(y) = \frac{y^{\frac{\alpha_1 + \alpha_2}{2} - 1}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} K_\alpha(2\sqrt{y}), \text{ avec } \alpha = \alpha_1 - \alpha_2$$

où  $K$  est la fonction de Bessel modifiée du troisième type.

$$K_\alpha(x) = \int_0^{+\infty} t^{\alpha-1} \exp\left(\frac{t^2 + 1}{t} \sqrt{x}\right) dt$$

- Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes selon des distributions  $Gamma(\alpha_1)$  et  $Gamma(\alpha_2)$  respectivement, alors  $X/(X+Y)$  a une distribution  $Beta(\alpha_1, \alpha_2)$ .

$$f_{Beta(\alpha_1, \alpha_2)}(x) = \frac{y^{\frac{\alpha_1 + \alpha_2}{2} - 1}}{\Gamma(\alpha_1) \Gamma(\alpha_2)} K_\alpha(2\sqrt{y})$$

- Le paramètre  $\alpha$  est appelé le paramètre de forme, car il influence le plus le pic de la distribution.
- Les moments de  $X$  sont donnés par :

$$E(X^r) = \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)}$$

- L'espérance et la variance de  $X$  sont les suivantes :

$$E(X) = Var(X) = \alpha$$

- Le coefficient d'asymétrie de  $X$  est

$$\gamma_1 = \frac{2}{\sqrt{\alpha}}$$

- La kurtosis de  $X$  est

$$\gamma_2 = 3 + \frac{6}{\alpha}$$

- La transformation de Laplace de est

$$\mathcal{L}(s) = \left(\frac{1}{1+s}\right)^\alpha$$



### 1.1.2 Distribution gamma de Pearson

Les distributions de Pearson sont les solutions d'une équation différentielle de Karl Pearson (1894), afin que les chercheurs obtiennent des propriétés statistiques souhaitables et suffisantes. Par la suite; ces solutions ont reçu des noms. L'une des plus importantes et des plus utilisées est la distribution suivante.

**Définition 1.1.2** On dit qu'une variable aléatoire a la distribution gamma à deux paramètres  $\alpha > 0$  et  $\beta > 0$  si sa distribution est donnée par

$$f_{G(\alpha,\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x); x > 0 \quad (1.2)$$

Le paramètre  $\alpha$  est appelé paramètre de forme, car il influe le plus sur le pic de la distribution. Le paramètre  $\beta$  est appelé paramètre d'intensité ou paramètre d'échelle inverse, car son influence est principalement sur l'étalement de la distribution, c'est-à-dire que l'écart type de la distribution gamma est proportionnel à  $\frac{1}{\beta}$ .

- Si  $\beta = 1$  la distribution  $Gamma(\alpha, 1)$  est la distribution  $Gamma(\alpha)$ .

$$f_{G(\alpha,1)}(x) = f_{G(\alpha)}(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x)$$

- la somme de deux distributions de  $Gamma(\alpha_1, \beta)$  et  $Gamma(\alpha_2, \beta)$  est une distribution de  $Gamma(\alpha_1 + \alpha_2; \beta)$ .
- Le produit d'une distribution de  $Gamma(\alpha, \beta)$  avec une constante positive  $a$  est une distribution de  $Gamma(\alpha, \frac{\beta}{a})$ .
- Si  $\alpha = 1$  et  $\beta = \frac{1}{\lambda}$  la distribution  $Gamma(1, \frac{1}{\lambda})$  est la distribution exponentielle  $\mathcal{E}(\lambda)$ .

$$f_{G(1,\frac{1}{\lambda})}(x) = f_{\mathcal{E}(\lambda)}(x) = \lambda \exp(-\lambda x); x > 0$$

- Si  $\alpha = \frac{v}{2}$  et  $\beta = \frac{1}{2}$  la distribution  $Gamma\left(\frac{v}{2}, \frac{1}{2}\right)$  est la distribution de.chi-deux avec  $v$  degrés de liberté

$$f_{G\left(\frac{v}{2}, \frac{1}{2}\right)}(x) = f_{\chi^2(v)}(x) = \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} x^{\frac{v}{2}-1} \exp\left(-\frac{x}{2}\right); x > 0$$

- Si  $\alpha$  est un entier, la distribution Gamma est une distribution d'Erlang.
- Si  $X$  une variable aléatoire a une distribution  $Gamma(\alpha, \beta)$ , alors  $1/X$  a une distribution Gamma inverse.

$$f(x) = \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} x^{\frac{v}{2}-1} \exp\left(-\frac{x}{2}\right); x > 0, v = \alpha\beta$$

- Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes selon des distributions  $Gamma(\alpha; \theta)$  et  $Gamma(\beta; \theta)$  respectivement, alors  $X/(X + Y)$  a une distribution  $beta(\alpha, \beta)$ .
- Le produit de deux variables aléatoires de Gamma  $Gamma(\alpha_1, \beta)$  est  $Gamma(\alpha_2, \beta)$  est une variable aléatoire de Bessel  $K(\alpha_1, \alpha_2)$

$$f_{X_1 X_2}(y) = \frac{2y^{\frac{\alpha_1 + \alpha_2}{2} - 1}}{\beta^{\alpha_1 + \alpha_2 + 1} \Gamma(\alpha_1) \Gamma(\alpha_2)} K_{\alpha_1 - \alpha_2}(2\sqrt{y})$$

où  $K$  est la fonction de Bessel modifiée du troisième type.

$$K_\alpha(x) = \int_0^{+\infty} t^{\alpha-1} \exp\left(\frac{t^2 + 1}{t} \sqrt{x}\right) dt$$

- Si  $(X_i)_{i=1, n}$  sont des variables aléatoires indépendantes selon des distributions  $Gamma(\alpha_i; \beta)$   $i = 1, 2, \dots, n$  respectivement, alors le vecteur  $(X_1/S, \dots, X_n/S)$ , où  $S = \sum_{i=1}^n X_i$  à une distribution de Dirichlet de  $n$  paramètres  $(\alpha_1, \alpha_2, \dots, \alpha_n)$ .
- Les moments sont donnés par :

$$E(X^r) = \frac{\Gamma(\alpha + r)}{\beta^r \Gamma(\alpha)}$$

- L'espérance et la variance d'une variable aléatoire de Gamma sont les suivantes :

$$\begin{aligned}E(X) &= \frac{\alpha}{\beta} \\ \text{Var}(X) &= \frac{\alpha}{\beta^2}\end{aligned}$$

- Le coefficient d'asymétrie est

$$\gamma_1 = \frac{2}{\sqrt{\alpha}}$$

- La kurtosis est

$$\gamma_2 = 3 + \frac{6}{\alpha}$$

- Transformation de Laplace  $\mathcal{L}(s) = \left(\frac{1}{1+\frac{s}{\beta}}\right)^\alpha$

### 1.1.3 Distribution Gamma de Stacy

La distribution gamma à trois paramètres ayant, entre autres, Weibull, gamma, exponentiel et Raleigh comme sous-distributions spéciales, est une distribution très populaire pour la modélisation des données de durée de vie et pour le phénomène de modélisation avec des taux d'échec monotones.

**Définition 1.1.3** On dit qu'une variable aléatoire a la distribution gamma à trois paramètres  $\alpha > 0$ ,  $\beta > 0$  et  $\gamma > 0$  si sa distribution est donnée par

$$f_{G(\alpha,\beta,\gamma)}(x) = \frac{\beta^\gamma}{\Gamma(\alpha)} x^{\gamma\alpha-1} \exp(-(\beta x)^\gamma); x > 0 \quad (1.3)$$

- Si  $\gamma = 1$  la distribution Gamma  $Gamma(\alpha, \beta, 1)$  est la distribution  $Gamma(\alpha, \beta)$ .

$$f_{G(\alpha,\beta,1)}(x) = f_{G(\alpha,\beta)}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), x > 0$$

- Si  $\alpha = \frac{v}{2}$ ,  $\beta = \frac{1}{2}$  et  $\gamma = 1$  la distribution  $Gamma\left(\frac{v}{2}, \frac{1}{2}, 1\right)$  est la distribution Chi-deux avec  $v$  degrés de liberté.

$$f_{G\left(\frac{v}{2}, \frac{1}{2}\right)}(x) = f_{\chi^2(v)}(x) = \frac{1}{2^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right)} x^{\frac{v}{2}-1} \exp\left(-\frac{x}{2}\right); x > 0$$

- Si  $\alpha = \frac{v}{2}$ ,  $\beta = \frac{1}{\sqrt{2}}$  et  $\gamma = 2$  la distribution  $Gamma\left(\frac{v}{2}, \frac{1}{\sqrt{2}}, 2\right)$  est la distribution Chi-deux avec  $v$  degrés de liberté.

$$f(x) = \frac{1}{\sqrt{2} \Gamma\left(\frac{v}{2}\right)} \left(\sqrt{2}x\right)^{v-1} \exp(-2x^2); x > 0$$

- Si  $\alpha = 1$ ,  $\beta = \frac{1}{\sqrt{2}\sigma^2}$  et  $\gamma = 2$  la distribution  $Gamma\left(1, \frac{1}{\sqrt{2}\sigma^2}, 2\right)$  est la distribution de Rayleigh

$$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), x > 0$$

- Les moments sont donnés par :

$$E(X^r) = \frac{\Gamma\left(\alpha + \frac{r}{\gamma}\right)}{\beta^r \Gamma(\alpha)}$$

- L'espérance et la variance d'une variable aléatoire de Gamma sont les suivantes :

$$E(X) = \frac{\Gamma\left(\alpha + \frac{1}{\gamma}\right)}{\beta}$$
$$Var(X) = \frac{\Gamma\left(\alpha + \frac{2}{\gamma}\right)}{\beta^2} - \left(\frac{\Gamma\left(\alpha + \frac{1}{\gamma}\right)}{\beta}\right)^2$$

- Le coefficient d'asymétrie est

$$\gamma_1 = \frac{\Gamma\left(\alpha + \frac{1}{\gamma}\right)}{\beta}$$

- La kurtosis est

$$\gamma_2 = \frac{\Gamma\left(\alpha + \frac{1}{2\gamma}\right)}{\beta}$$

## 1.2 Distribution de Pareto

L'économiste italien Vilfredo Fr d rico Damaso surnomm  par ses  tudiants Marquis de Pareto est le premier   mettre en  vidence l'application d'une loi puissance   un ph nom ne  conomique. En 1906 il  tudie la r partition des revenus des habitants de divers pays industrialis s et constate que 80% des richesses appartiennent   20% de la population et ce, quel que soit le pays observ .

En effet, par projection, ce principe se v rifie dans de nombreux domaines. La majorit  des r sultats ou des impacts, sont d s   une minorit  des causes. 20% des bugs sont responsables de 80% des plantages de logiciels, 20% des articles en stock repr sentent 80% du co t de stockage, etc. D'o  la seconde appellation : la loi des 20/80 (ou 80/20).

- En gestion des stocks, sous sa forme de loi ABC de Joseph Juran qui  nonce que les ressources repr sentant 70   80% du CA sont rassembl s dans la classe A, les ressources contribuant entre 10   15% au CA sont dans la classe C et la classe B rassemble les ressources interm diaires. Juran d clare que : le principe de Pareto permet seulement de s parer les choses en deux parts. En r alit , il existe 3 parties. La troisi me est un « **r sidu** » qui prend place entre les composantes prioritaires et les composantes secondaires. Ce « r sidu » peut  tre d nomm  « **zone   risques** » (awkward-zone). Chaque  l ment de cette zone   risques n'est pas assez important pour justifier un lourd investissement dans l'analyse, mais leur regroupement d passe les capacit s d'analyse.

**D finition 1.2.1** Une variable al atoire continue  $X$  suit une loi de Pareto

de paramètres  $\alpha, k \in \mathbb{R}_+^*$ , si elle admet pour densité :

$$f(x) = \begin{cases} \frac{k\alpha^k}{x^{k+1}} & \text{si } x \geq \alpha \\ 0 & \text{si } x < \alpha \end{cases}$$

Ceci est noté  $X \sim Pa(\alpha, k)$ .

**Proposition 1.1** *Si  $X \sim Pa(\alpha, k)$  alors la fonction de répartition  $F(x)$  et la fonction de survie  $S(x)$  de  $X$  sont données par :*

$$F(x) = \begin{cases} 1 - \left(\frac{\alpha}{x}\right)^k & \text{si } x \geq \alpha \\ 0 & \text{si } x < \alpha \end{cases} \quad \text{et} \quad S(x) = \begin{cases} \left(\frac{\alpha}{x}\right)^k & \text{si } x \geq \alpha \\ 0 & \text{si } x < \alpha \end{cases}$$

**Proposition 1.2** *Si  $X \sim Pa(\alpha, k)$  alors la moyenne théorique et la variance théorique de  $X$  sont données respectivement par :*

$$E(X) = \frac{k\alpha}{\alpha - 1} \quad (\text{n'existe que si } k > 1),$$
$$V(X) = \frac{k\alpha^2}{(\alpha - 1)^2(\alpha - 2)} \quad (\text{n'existe que si } k > 2).$$

**Proposition 1.3** *Si  $X \sim Pa(\alpha, k)$  alors le coefficient d'asymétrie théorique (skewness) et le coefficient d'aplatissement théorique (kurtosis) de  $X$  sont donnés respectivement par :*

$$\gamma_1 = \frac{2(k+1)}{k-3} \sqrt{\frac{k-2}{k}} \quad \text{si } k > 3 \quad \text{et} \quad \gamma_2 = \frac{9k^3 - 15k^2 - 12}{k^3 - 7k^2 + 12k} \quad \text{si } k > 4$$

### 1.3 Distribution de Rayleigh (Lord Rayleigh).

**Définition 1.3.1** Une variable aléatoire continue  $X$  suit une loi de Rayleigh de paramètre  $\sigma > 0$ , si elle admet pour densité :

$$f_{R(\sigma)}(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), x > 0$$

- La distribution de Rayleigh est un cas particulier de la distribution de Weibull avec un paramètre d'échelle égal à 2.
- Lorsqu'un Rayleigh est défini avec un paramètre de forme  $\sigma$  égal à 1, il est égal à une distribution de khi carré à 2 degrés de liberté.
- $X$  suit la loi de Rayleigh de paramètre  $\sigma$  si  $X = \sqrt{A^2 + B^2}$  où  $A$  et  $B$  sont deux variables gaussiennes indépendantes, ce qui explique le choix du symbole  $\sigma$  pour paramétriser la loi de Rayleigh.
- Si  $X$  suit la loi de Rayleigh de paramètre  $\sigma$  alors  $X^2$  suit la loi du qui-deux avec deux degrés de liberté, qui est une loi exponentielle de paramètre  $\frac{1}{2}$ .
- Si  $X$  suit une loi exponentielle de paramètre  $\lambda$  alors  $\sqrt{2\sigma^2\lambda X}$  suit une loi de Rayleigh de paramètre  $\sigma$ .
- Si  $(X_i)_{i=1,n}$  une suite de variables indépendantes suivent la même loi de Rayleigh de paramètre  $\sigma$  alors  $\sum_{i=1}^n X_i^2$  suit une loi gamma de paramètres  $n$  et  $2\sigma^2$
- Les moments sont donnés par :

$$E(X^r) = \sigma^r 2^{\frac{r}{2}} \Gamma\left(1 + \frac{r}{2}\right)$$

- L'espérance et la variance d'une variable aléatoire de Gamma sont les suivantes :

$$\begin{aligned} E(X) &= \sigma \sqrt{\frac{\pi}{2}} \\ \text{Var}(X) &= \frac{4 - \pi}{2} \sigma^2 \end{aligned}$$

- Le coefficient d’asymétrie est

$$\gamma_1 = \frac{2\sqrt{\pi}(\pi - 3)}{(4 - \pi)^{\frac{3}{2}}}$$

- La kurtosis est

$$\gamma_2 = \frac{6\pi^2 - 24\pi + 16}{(4 - \pi)^2}$$

- Transformation de Laplace  $\mathcal{L}(s) = 1 + \sigma s \exp\left(\frac{\sigma^2 t^2}{2}\right) \sqrt{\frac{\pi}{2}} \left(\operatorname{erf}\left(\frac{\sigma t}{\sqrt{2}}\right) + 1\right)$   
où erf est la fonction d’erreur.

## 1.4 Distribution de Lindley (1958)

Une distribution de Lindley d’un seul paramètre a attiré les chercheurs pour son usage en modelant des données de vie, et on l’a observé en plusieurs articles que cette distribution a exécutés excellentment . Cette distribution est introduite par Lindley en 1958 comme mélange d’ $Exp(\beta)$  et de  $Gamma(2; \beta)$ .

**Définition 1.4.1** I Une variable aléatoire  $X$  suit une loi de Lindley de paramètre  $\beta \in \mathbb{R}_+$ , si elle admet pour densité :

$$f_{L(\beta)}(x) = \frac{\beta^2}{1 + \beta} (1 + x) \exp(-\beta x), x > 0 \quad (1.4)$$

ceci est noté  $X \sim L(\beta)$

- La fonction de répartition correspondante est :

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

- La première dérivée de (1.4) est

$$f'(x) = \frac{\beta^2}{1 + \beta} (1 - \beta - \beta x) e^{-\beta x}$$



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

$$E(X^k) = \frac{k! (\beta + k + 1)}{\beta^k (1 + \beta)}$$

- Le coefficient de variation

$$\gamma = \frac{\sqrt{\beta^2 + 4\beta + 2}}{\beta + 2}$$

## 1.5 Distribution de Lindley avec deux paramètres

**Définition 1.5.1** Une variable aléatoire  $X$  suit une loi de Lindley de deux paramètres  $(\alpha, \beta) \in \mathbb{R}_+^2$ , si elle admet pour densité :

$$f_{L(\alpha, \beta)}(x) = \frac{\beta^{\alpha+1}}{(\alpha + \beta) \Gamma(\alpha)} x^{\alpha-1} (1+x) e^{-\beta x}, x > 0 \quad (1.5)$$

ceci est noté  $X \sim L(\alpha, \beta)$

- Si  $\alpha = 1$ ,  $L(1, \beta) = L(\beta)$ .
- La fonction de répartition correspondante est :

$$F(x) = \frac{(\beta + \alpha) \mathcal{G}(\alpha, \beta x) - (\beta x)^{\alpha} e^{-\beta x}}{(\alpha + \beta) \Gamma(\alpha)},$$

avec  $\mathcal{G}(\cdot, \cdot)$  la fonction gamma inverse.

- La première dérivée de (1.5) est

$$f'(x) = \frac{g(x)}{x(1+x)} f(x)$$

où

$$g(x) = \beta x^2 + (\beta - \alpha)x + (1 - \alpha).$$

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

$$E(X^k) = \frac{\Gamma(\alpha + k) (\beta + k + \alpha)}{\beta^k (\alpha + \beta) \Gamma(\alpha)}.$$

## CHAPITRE 2

### Modèles combinés, T-X et Mélange

#### Sommaire

---

<b>2.1 Modèles combinés</b> . . . . .	14
<b>2.2 Modèle <math>T - X</math></b> . . . . .	17
<b>2.3 Modèle de Mélange</b> . . . . .	20
<b>2.4 Distribution de Lindley Généralisée</b> . . . . .	20
<b>2.5 Distribution Pseudo Lindley</b> . . . . .	22
<b>2.6 Distribution Quasi Lindley</b> . . . . .	28
<b>2.7 Distribution de Gamma Lindley</b> . . . . .	32
<b>2.8 Comparaison des distributions</b> . . . . .	35

---

Des nombreuses classes de distributions généralisées ont été développées et appliquées pour décrire divers phénomènes. Il existe plusieurs méthodes pour introduire des nouvelles distributions à savoir : Modèles combinés ; transformed-transformer ( $T - X$ ), Mélange.

## 2.1 Modèles combinés

Dans certaines situations, les statisticiens rencontrent des données qui proviennent évidemment de deux modèles différents. C'est souvent le cas avec par exemple : Les données sur les paiements d'assurance, lorsque les actuaires doivent traiter des données plus petites avec des fréquences élevées et des données plus grandes occasionnelles avec des fréquences plus basses. La distribution correspondante peut alors être modélisée sous la forme d'une combinaison de deux distributions, consistant en une distribution moins lourde jusqu'à une certaine valeur de seuil et une répartition à queue grossie à partir de ce seuil. Les cas particuliers de cette distribution sont des modèles combinés qui sont définis en matière de deux fonctions de densité de probabilité tronquée, des poids de mélange et de la limite de porter du domaine. qui est suggéré par Cooray et Ananda [23]. Ils ont construit un modèle combiné ayant la fonction de densité de probabilité

$$f(x) = \begin{cases} c_1 f_1(x) & -\infty < x \leq \theta \\ c_2 f_2(x) & \theta \leq x < \infty \end{cases} \quad (2.1)$$

où  $f_1$  et  $f_2$  sont des fonctions de densité de probabilité, tandis que  $c_1$  et  $c_2$  sont des constantes de normalisation qui résultent en imposant des conditions de continuité et de différentiabilité, (2.1) est appelé un modèle épissé à deux composantes des petites pertes. Par conséquent, le rôle du seuil  $\theta$  ( $0 < \theta < \infty$ ) (limite supérieure du contrat) est très important, étant le point où sur la base de l'analyse des données, nous changeons le modèle.

D'après Cooray et Ananda (2005),  $f_1$  est pris comme étant la densité Lognormal, alors que  $f_2$  est la densité de Pareto. Sur la base des mêmes distributions, Scollnik (2007) a proposé deux modèles combinés différents, plus généraux que ceux étudiés par Cooray et Ananda (2005). De la même

manière, Teodorescu et Vernic (2006) ont proposé des modèles Exponentiel-Pareto combinés. De plus, Preda et Ciumara (2006) ont réalisé une étude comparative entre les distributions Weibull-Pareto et Lognormal-Pareto.

### Construction et caractéristiques du modèle combiné

Suivant Cooray et Ananda (2005), une densité composite est dérivée de

$$f(x) = \begin{cases} c_1 f_1(x) & -\infty < x \leq \theta \\ c_2 f_2(x) & \theta \leq x < \infty \end{cases}$$

afin d'obtenir une fonction de densité combiné lisse, nous imposons des conditions de continuité et de différentiabilité au seuil  $\theta$ , c'est-à-dire

$$\int_0^\infty f(x)dx = 1, \quad f_1(\theta) = f_2(\theta) \text{ et } \frac{d}{dx} f_1(\theta) = \frac{d}{dx} f_2(\theta).$$

Scollnik (2007) a remarqué que la fonction de densité du modèle combiné (2.1) pourrait également être écrite comme

$$f(x) = \begin{cases} r f_1^*(x) & -\infty < x \leq \theta \\ (1-r) f_2^*(x) & \theta \leq x < \infty \end{cases} \quad (2.2)$$

où  $0 \leq r \leq 1$ , tandis que  $f_1^*(x)$  et  $f_2^*(x)$  sont des troncatures adéquates des fonctions de densité  $f_1$  et  $f_2$ . Plus précisément, si  $F_i$  désigne la distribution de répartition de  $f_i$ , nous avons

$$\begin{aligned} f_1^*(x) &= \frac{f_1(x)}{\int_{-\infty}^{\theta} f_1(y) dy} = \frac{f_1(x)}{F_1(\theta)}, & -\infty < x \leq \theta \\ f_2^*(x) &= \frac{f_2(x)}{\int_{\theta}^{\infty} f_2(y) dy} = \frac{f_2(x)}{1 - F_2(\theta)}, & \theta \leq x < \infty \end{aligned} \quad (2.3)$$

Il est facile de voir que la fonction de densité (2.2) peut être interprétée comme un modèle de mélange à deux composants avec des poids de mélange  $r$  et  $1-r$  (c'est-à-dire une combinaison convexe de deux fonctions de densité).

$$f(x) = r f_1^*(x) + (1-r) f_2^*(x), \quad r \in [0, 1]. \quad (2.4)$$

Teodorescu et Vernic(2006) ont déduit certaines propriétés du modèle (2.2) concernant sa fonction de répartition, les moments initiaux et la fonction caractéristique. Ces propriétés sont présentées ci-dessous. Si  $X$  est une variable aléatoire avec fonction de densité  $f$ , on note son moment initial de la  $n^{\text{ième}}$  entrée par

$$E_n(f) = E(X^n)$$

et sa fonction caractéristique par

$$\varphi_f(t) = \varphi_X(t) = E(e^{itX}).$$

**Proposition 2.1.1.**

a) Soit  $F$  la fonction de répartition complète de la fonction de densité donnée dans (2.2). alors

$$F(x) = \begin{cases} r \frac{F_1(x)}{F_1(\theta)} & -\infty < x \leq \theta \\ r + (1-r) \frac{F_2(x) - F_2(\theta)}{1 - F_2(\theta)} & \theta \leq x < \infty \end{cases} \quad (2.5)$$

b) En supposant que toutes les quantités impliquées existent, le  $n^{\text{ième}}$  ordre moment initial de la fonction de densité (2.2) est

$$E_n(f) = r E_n(f_1^*) + (1-r) E_n(f_2^*),$$

alors que sa fonction caractéristique est

$$\varphi_f(t) = r \varphi_{f_1^*}(t) + (1+r) \varphi_{f_2^*}(t), \quad t \in \mathbb{R}. \quad (2.6)$$

**Preuve.** (voir Teodorescu S, Vernic R (2009)). ■

## 2.2 Modèle $T - X$

Eugene et al. (2002) a introduit la famille de distributions bêta-généralisée avec fonction de répartition

$$G(x) = \int_a^{F(x)} b(t) dt, \quad (2.7)$$

où  $b(t)$  est la fonction de densité de probabilité du variable aléatoire bêta et  $F(x)$  est la fonction de répartition pour un autre variable aléatoire. La fonction de densité de probabilité correspondant à (2.7) est donné par

$$g(x) = \frac{1}{B(\alpha, \beta)} f(x) F^{\alpha-1}(x) (1 - F(x))^{\beta-1}.$$

La variable aléatoire bêta se situe entre 0 et 1, de même que la fonction de répartition  $F(x)$  d'un autre variable aléatoire. La limitation de l'utilisation d'un générateur avec un support compris entre 0 et 1 soulève une question intéressante : «*Peut-on utiliser d'autres distributions avec un support différent en tant que générateur pour dériver différentes classes de distributions ?*» Cette section abordera cette question et introduira une nouvelle technique pour dériver des familles de distributions en utilisant n'importe quelle fonction de densité de probabilité en tant que générateur.

Alzaatreh et al. (2013) a étendu la famille bêta généralisée et a défini la famille  $T - X(W)$ . La fonction de distribution cumulative de la distribution  $T - X(W)$  est

$$G(x) = \int_a^{W(F(x))} r(t) dt,$$

où  $r(t)$  la densité de probabilité d'une variable aléatoire  $T \in [a, b]$ , pour  $-\infty \leq a < b \leq \infty$ . Soit  $W(F(x))$  une fonction de la fonction de répartition  $F(x)$  de toute variable aléatoire  $X$  pour que  $W(F(x))$  satisfait aux conditions suivantes :

$$\left\{ \begin{array}{l} W(F(x)) \in [a, b] \\ W(F(x)) \text{ est différentiable et monotoniquement non décroissante} \\ \text{si } x \rightarrow -\infty, W(F(x)) \rightarrow a \text{ et si } x \rightarrow \infty, W(F(x)) \rightarrow b. \end{array} \right. \quad (2.8)$$

La méthode pour générer des nouvelles familles de distribution est présentée à la suivante définition.

**Définition 2.2.1** Soit  $X$  une variable aléatoire avec fonction de densité de probabilité  $f(x)$  et fonction de répartition cumulative  $F(x)$ . Soit  $T$  une variable aléatoire continue avec fonction de densité de probabilité  $r(t)$  définie sur  $[a, b]$ . La fonction de distribution cumulative d'une nouvelle famille de distributions est définie comme

$$G(x) = \int_a^{W(F(x))} r(t) dt \quad (2.9)$$

où  $W(F(x))$  satisfait aux conditions de (2.8). La fonction de distribution cumulative  $G(x)$  dans (2.9) peut être écrit comme  $G(x) = R\{W(F(x))\}$ , où  $R(t)$  est la fonction de distribution cumulative de la variable aléatoire  $T$ . La fonction de distribution cumulative associée à (2.8) est

$$g(x) = \left\{ \frac{d}{dx} W(F(x)) \right\} r\{W(F(x))\} \quad (2.10)$$

Notez que :

- La fonction de distribution cumulative dans (2.9) est une fonction composite de  $(R.W.F)(x)$ .
- La fonction de densité de probabilité  $r(t)$  dans (2.9) est "transformé" en un nouveau fonction de distribution cumulative  $G(x)$  à travers la fonction  $W(F(x))$  qui agit comme un «transformateur». Par conséquent, nous nous

référerons à la distribution  $g(x)$  dans (2.10) transformée à partir de la variable aléatoire  $T$  par la variable aléatoire du transformateur  $X$  et l'appeler "Transformé-Transforme" ou distribution  $T - X$ .

- La variable aléatoire  $X$  peut être discrète et dans ce cas  $G(x)$  est le fonction de distribution cumulative d'une famille de distributions discrètes.

Des différentes  $W(F(x))$  donneront une nouvelle famille de distributions. La définition de  $W(F(x))$  dépend du support de la variable aléatoire  $T$ . Voici quelques exemples de  $W(\cdot)$ .

1. Lorsque le support de  $T$  est borné : Sans perte de généralité, on suppose que le support de  $T$  est  $[0, 1]$ . Les distributions pour ce type de  $T$  incluent des distributions uniformes  $(0, 1)$ , bêta, Kumaraswamy et d'autres types de distribution bêta généralisée.  $W(F(x))$  peut être défini comme  $F(x)$  ou  $F^\alpha(x)$ . Il s'agit de la famille de distributions générées par le bêta qui ont été bien étudiées au cours de la dernière décennie.

2. Si le support de  $T$  est  $[a, \infty)$ ,  $a \geq 0$  : Sans perte de généralité, on suppose que  $a = 0$ .  $W(F(x))$  peut être défini comme  $-\log(1 - F(x))$ ,  $\frac{F(x)}{1-F(x)}$ ,  $-\log(1 - F^\alpha(x))$  et  $\frac{F^\alpha(x)}{1-F^\alpha(x)}$ .

3. Lorsque le support de  $T$  est  $(-\infty, \infty)$  :  $W(F(x))$  peut être défini comme  $\log[-\log(1 - F(x))]$ ,  $\log\left[\frac{F(x)}{1-F(x)}\right]$ ,  $\log[-\log(1 - F^\alpha(x))]$  et  $\log\left[\frac{F^\alpha(x)}{1-F^\alpha(x)}\right]$ .



**Tableau 2.1** Fonctions de densité de certaines familles  $T - X$  basées sur différentes fonctions  $W(.)$

$T$	$W(F(x))$	$g(x)$
$[0, \infty)$	$\frac{F(x)}{1-F(x)}$	$\frac{f(x)}{(1-F(x))^2} r \left[ \frac{F(x)}{1-F(x)} \right]$
$[0, \infty)$	$-\log(1 - F(x))$	$\frac{f(x)}{1-F(x)} r [-\log(1 - F(x))]$
$[0, \infty)$	$\frac{F^\alpha(x)}{1-F^\alpha(x)}$	$\frac{\alpha f(x) F^{\alpha-1}(x)}{(1-F^\alpha(x))^2} r \left[ \frac{F^\alpha(x)}{1-F^\alpha(x)} \right]$
$[0, \infty)$	$-\log(1 - F^\alpha(x))$	$\frac{\alpha f(x) F^{\alpha-1}(x)}{1-F^\alpha(x)} r [-\log(1 - F^\alpha(x))]$
$(-\infty, +\infty)$	$\log \left[ \frac{F(x)}{1-F(x)} \right]$	$\frac{f(x)}{F(x)(1-F(x))} r \left[ \log \left( \frac{F(x)}{1-F(x)} \right) \right]$
$(-\infty, +\infty)$	$\log [-\log(1 - F(x))]$	$\frac{f(x)}{(F(x)-1) \log(1-F(x))} r [\log(-\log(1 - F(x)))]$
$(-\infty, +\infty)$	$\log \left[ \frac{F^\alpha(x)}{1-F^\alpha(x)} \right]$	$\frac{\alpha f(x)}{F(x)(1-F^\alpha(x))} r \left[ \log \left( \frac{F^\alpha(x)}{1-F^\alpha(x)} \right) \right]$
$(-\infty, +\infty)$	$\log [-\log(1 - F^\alpha(x))]$	$\frac{\alpha f(x) F^{\alpha-1}(x)}{(F^\alpha(x)-1) \log(1-F^\alpha(x))} r [\log(-\log(1 - F^\alpha(x)))]$

## 2.3 Modèle de Mélange

Soient  $Y_i \sim \text{loi}(I_i)$   $n$  variables aléatoires indépendantes de densités  $f_i$ . On considère la variable aléatoire  $X = Y_1$ ,  $X = Y_2$ , ...,  $X = Y_n$  avec les probabilités  $P_i$ . La fonction de densité de  $X$  est donnée par :

$$f(x) = \sum_{i=1}^n p_i f_i(x) \text{ avec } 0 \leq p_i \leq 1 \text{ et } \sum_{i=1}^n p_i = 1.$$

Dans le reste de ce chapitre, nous introduisons des nouvelles distributions à savoir : Lindley Généralisée, Pseudo Lindley, Quasi Lindley et Gamma Lindley utilisant le modèle de mélange.

## 2.4 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

$$f(x; \alpha, \theta, \gamma) = \frac{\theta^2 (\theta x)^{\alpha-1} (\alpha + \gamma x) e^{-\theta x}}{(\theta + \gamma) \Gamma(\alpha + 1)}; \quad x, \theta, \alpha, \gamma > 0. \quad (2.11)$$

Il est facile de voir que si  $\alpha = \gamma = 1$ , la fonction (2.11) de *GLD* se réduit à la fonction de distribution de *Lindley*, si  $\gamma = 0$ , elle se réduit à la distribution de *Gamma*( $\alpha, \theta$ ) et si  $(\alpha, \gamma) = (1, 0)$ , elle se réduit à la distribution Exponentielle ( $\theta$ ). La fonction de densité (2.11) 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. Zakerzadeh and Dolati (2010)). 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 :

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

### 2.4.1 Estimation par la méthode des moments

Étant donné un échantillon aléatoire  $X_1, \dots, X_n$  de taille  $n$  de la distribution *GLD* (2.11), 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 :

$$\begin{aligned} \alpha(\theta + \gamma) + \gamma - \bar{X}\theta(\theta + \gamma) &= 0 \\ -2(\alpha + 1)\theta + (\alpha^2 + 3\alpha + 2)(\theta + \gamma) - n\theta^2(\theta + \gamma) &= 0 \\ \frac{\alpha + 1}{\theta^2} (-3\alpha + 6) + \frac{(\alpha + 1)(\theta + \gamma)}{\theta} \left[ (\alpha + 1)^2 + \frac{3\alpha + 5}{\theta^2} \right] - K &= 0 \end{aligned}$$

Il est possible de calculer  $\hat{\gamma}$ ,  $\hat{\theta}$  et  $\hat{\alpha}$  numériquement.

## 2.5 Distribution Pseudo Lindley

Dans cette section, on introduit la distribution de Pseudo Lindley ( $PsLD$ ) ([34], [37]) 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 :

$$f_{PsLD}(x; \theta, \beta) = \frac{\theta(\beta - 1 + \theta x) e^{-\theta x}}{\beta}, \quad x > 0, \theta > 0, \beta \geq 1. \quad (2.12)$$

La fonction de répartition correspondante est :

$$F_{PsLD}(x) = 1 - \frac{(\beta + \theta x) e^{-\theta x}}{\beta}; \quad x > 0, \beta \geq 1, \theta > 0. \quad (2.13)$$

**Remarque 2.5.1** 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 :

$$Mode(X) = \begin{cases} \frac{2-\beta}{\theta} & \forall 1 \leq \beta < 2 \\ 0 & \text{ailleurs.} \end{cases} \quad (2.14)$$

### 2.5.1 Moments et mesures connexes

**Proposition 2.1** *Le moment d'ordre  $k$  de la distribution de Pseudo Lindley est :*

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

*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*

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

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

$$\mathbb{E}(X) = \frac{\beta + 1}{\theta\beta}, \mathbb{E}(X^2) = \frac{\beta + 2}{\theta^2\beta}, \text{Var}(X) = \frac{\beta^2 + 2\beta - 1}{\beta^2\theta^2}. \quad (2.15)$$

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

$$M < me < \mu.$$

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  $\beta \geq 1$ . Il est facile de vérifier que

$$F(M) < F(me) < F(\mu), \text{ alors on a } M < me < \mu.$$

■

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

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

**Remarque 2.5.2** Toutes ces expressions sont indépendantes du paramètre  $\theta$  et dépendantes uniquement du paramètre  $\beta$ , avec  $6\sqrt{2} < \sqrt{\beta_1} < 6$  et  $\beta_2 > 30$ .

## 2.5.2 Fonction de hasard et Fonction de survie

Soit

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

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  $\frac{d}{dx}h_{P_sLD}(x) = \frac{\theta^2}{(\beta+\theta x)^2} > 0$ . ■

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

D'après la fonction de répartition de la distribution de Pseudo Lindley définie en (3.2), 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 2.5.1** *Pour tout  $\theta > 0$  et  $\beta > 1$ , la fonction quantile de  $X$  qui suit la distribution de Pseudo Lindley est*

$$Q_X(u) = -\frac{\beta}{\theta} - \frac{1}{\theta}W_{-1}(\beta e^{-\beta}(u-1)), \quad 0 < u < 1, \quad (2.16)$$

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

*Preuve.* Voir ( Zeghdoudi et Nedjar (2017)). ■

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

Étant donné un échantillon aléatoire  $X_1, \dots, X_n$ , de la distribution de Pseudo Lindley (3.1), 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 :

$$\hat{\theta}_{MoM} = \frac{2m + \sqrt{2}\sqrt{m^2 - s^2}}{m^2 + s^2} \quad \text{et} \quad \hat{\beta}_{MoM} = \frac{m^2 + s^2}{m^2 - s^2 + \sqrt{2}m\sqrt{m^2 - s^2}}. \quad (2.17)$$

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

**Théorème 2.5.2** *L'estimateur des moments  $\hat{\theta}_{MoM}$  de  $\theta$  est positivement biaisée, c'est-à-dire*

$$E(\hat{\theta}_{MoM}) - \theta > 0.$$

**Théorème 2.5.3** *Soient  $\hat{\theta}_{MoM} = N(m)$  et  $N(t) = \frac{\beta + 1}{t\beta}$ , pour  $t > 0$ , on a*

$$\frac{d^2}{dt^2}N(t) = \frac{2(\beta + 1)}{\beta t^3} > 0.$$

$N(t)$  est strictement convexe. Maintenant, par l'inégalité de Jensen, on a  $\mathbb{E}(N(m)) > N(\mathbb{E}(m))$ . Ainsi,  $\mathbb{E}(m) = N(\mu) = N\left(\frac{\beta+1}{\theta\beta}\right) = \theta$ ,  $E(\hat{\theta}) > \theta$ . Alors  $\hat{\theta}_{MoM}$  est positivement biaisé.

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

**Théorème 2.5.4** *L'estimateur  $\hat{\theta}_{MoM}$  de  $\theta$  est convergent et asymptotiquement normal :*

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

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

$$\hat{\theta}_{MoM} \pm z_{\frac{\alpha}{2}} \frac{1}{\sqrt{n\hat{\sigma}^2}}.$$

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

**Preuve.**

**Théorème 2.5.5 Preuve.** Soient  $\hat{\theta}_{MoM} = g(m)$  et  $g(t) = \frac{\beta+1}{t^\beta}$ , étant donné aussi  $m$  est finie,  $m \xrightarrow{P} \mu$ ,  $g(t)$  est une fonction continue à  $t = \mu$ ,  $g(m) \xrightarrow{P} g(\mu)$ , c'est-à-dire  $\hat{\theta}_{MoM} \xrightarrow{P} \theta$ . Comme  $\sigma^2 < \infty$ , par le théorème central limite, on a

$$\sqrt{n}(m - \mu) \xrightarrow{P} N(0, \sigma^2).$$

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

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

Enfin, étant donné que

$$g(m) = \hat{\theta}_{MoM}, \quad g(\mu) = \theta, \quad \text{et} \quad g'(\mu) = -\frac{\beta+1}{\mu\beta}.$$

■  
■

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

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

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

Les équations des dérivées de  $\ln l(x_i; \theta, \beta)$  par rapport à  $\theta$  et  $\beta$  sont ainsi obtenus sous la forme :

$$\frac{\partial \ln l(x_i; \theta, \beta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n x_i + \sum_{i=1}^n \left( \frac{x_i}{\beta - 1 + \theta x_i} \right). \quad (2.18)$$

$$\frac{\partial \ln l(x_i; \theta, \beta)}{\partial \beta} = \frac{-n}{\beta} + \sum_{i=1}^n \left( \frac{1}{\beta - 1 + \theta x_i} \right). \quad (2.19)$$



Les deux équations (2.18) et (2.19) ne peut pas résoudre directement, on doit utiliser la méthode Fisher Scoring, on a :

$$\begin{bmatrix} \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta^2} & \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta \partial \beta} \\ \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta \partial \theta} & \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta^2} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial Lnl(x_i; \beta, \theta)}{\partial \theta} \\ \frac{\partial Lnl(x_i; \beta, \theta)}{\partial \beta} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \quad (2.20)$$

Où,  $\frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta^2} = -\frac{n}{\theta^2} - \sum_{i=1}^n \left( \frac{x_i^2}{(\beta-1+\theta x_i)^2} \right)$ ;  $\frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta^2} = \frac{n}{\beta^2} - \sum_{i=1}^n \left( \frac{1}{(\beta-1+\theta x_i)^2} \right)$   
 et  $\frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta \partial \theta} = \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta \partial \beta} = -\sum_{i=1}^n \left( \frac{x_i}{\beta-1+\theta x_i} \right)$ .

L'équation (2.19) peut se résoudre de façon itérative où  $\theta_0$  et  $\beta_0$  sont les valeurs initiales de  $\theta$  et  $\beta$  respectivement.

## 2.6 Distribution Quasi Lindley

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

$$f(x; \alpha, \theta) = \frac{\theta^2(\alpha + x\theta)e^{-\theta x}}{\alpha + 1}; \quad x, \theta, \alpha > 0. \quad (2.22)$$

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

La dérivée première de (2.22) est :

$$f'(x) = \frac{\theta^2 e^{-\theta x}}{\alpha + 1} (1 - \alpha - \theta x)$$

Il en résulte que

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

(ii) pour  $\alpha \geq 1$ ,  $\frac{d}{dx}f(x) \leq 0$ , c-à-d  $f(x)$  diminue en  $x_0$  , dont le mode de la distribution QL est donnée par :

$$Mode = \begin{cases} \frac{1-\alpha}{\theta}, & \alpha < 1 \\ 0 & \alpha \geq 1 \end{cases}$$

La fonction de répartition correspondante est :

$$F(x) = 1 - \frac{1 + \alpha + \theta x}{\alpha + 1} e^{-\theta x}; x > 0, \theta, \alpha > 0. \quad (2.23)$$

Le  $r^{ieme}$  moment de la distribution de quasi Lindley est :

$$\mu_r^{\downarrow} = \frac{\Gamma(r+1)(\alpha+r+1)}{\theta^r(\alpha+1)}, r = 1, 2, \dots$$

En particulier on a

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

Le moment centré de la distribution de Lindley est donné par :

$$\mu_k = E \left\{ (X - \mu)^k \right\} = \sum_{r=0}^k \binom{k}{r} \mu_r^{\downarrow} (-\mu)^{k-r}.$$

En particulier on a

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

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

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

### 2.6.1 Fonction de hasard et fonction de survie

Pour la distribution Quasi Lindley la fonction de taux de hasard est :

$$h(x) = \lim_{\Delta x \rightarrow 0} \frac{P(X < x + \Delta x \mid X > x)}{\Delta x} = \frac{f(x)}{1 - F(x)} = \frac{\theta(\alpha + \theta x)}{\alpha + 1 + \theta x}. \quad (2.24)$$

et la fonction de survie est :

$$S(x) = 1 - F(x) = \frac{\alpha + 1 + \theta x}{1 + \alpha} e^{-\theta x}. \quad (2.25)$$

### 2.6.2 Fonction Quantile de la distribution Quasi Lindley

La distribution Quasi Lindley de paramètre  $\theta$  et  $\alpha$  est spécifiée par sa fonction de répartition définie en (2.22). 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 2.6.1** *Pour tout  $\theta > 0$ , la fonction quantile de la distribution Lindley  $X$  est :*

$$Q_X(u) = -\frac{1}{\theta} - \frac{\alpha}{\theta} - \frac{1}{\theta} W_{-1} \left( \frac{\alpha + 1}{\exp(\alpha + 1)} (u - 1) \right), \quad 0 < u < 1, \quad (2.26)$$

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 :

$$(\alpha + 1 + \theta x) e^{-\theta x} = (\alpha + 1) (1 - u). \quad (2.27)$$

En multipliant par  $-\exp(-\alpha - 1)$  l'équation (2.27), on obtient :

$$-(\alpha + 1 + \theta x) \exp(-\alpha - 1 - \theta x) = (\alpha + 1) (u - 1) \exp(-\alpha - 1). \quad (2.28)$$

D'après l'équation (2.28), conjointement avec l'équation (1.3), on voit que  $-(\alpha + 1 + \theta x)$  est la fonction  $W$  de Lambert de l'argument réel  $(\alpha + 1) (u - 1) \exp(-\alpha - 1)$ . Alors, on a

$$W \left( \frac{\alpha + 1}{\exp(\alpha + 1)} (u - 1) \right) = -(\alpha + 1 + \theta x), \quad 0 < u < 1. \quad (2.29)$$

Toujours, pour tout  $\theta > 0$  et  $x > 0$  il est immédiat que  $(\alpha + 1 + \theta x) > 1$  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 première chapitre, l'équation (2.29) devient

$$W_{-1} \left( \frac{\alpha + 1}{\exp(\alpha + 1)} (u - 1) \right) = -(\alpha + 1 + \theta x). \quad (2.30)$$

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

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

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 :

$$\hat{\theta} = \left( \frac{\hat{\alpha} + 2}{\hat{\alpha} + 1} \right) \frac{1}{\bar{X}}, \quad \bar{X} > 0. \quad (2.31)$$

$$\hat{\alpha} = \frac{-1}{k-2} \left[ 2k + \sqrt{2\sqrt{2-k} - 4} \right] \quad (2.32)$$

où

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

## 2.7 Distribution de Gamma Lindley

Dans cette section, on introduit la distribution de Gamma Lindley ( $Gal$ ) ( [35], [36] ) et étudier ses propriétés de base. Cette distribution est défini comme un mélange de  $Gamma(2, \theta)$  et de  $Lindley(\theta)$ . Soit  $Y_1 \sim Gamma(2, \theta)$  et  $Y_2 \sim LD(\theta)$  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  $P_1$ ,  $P_2$  et  $P_1 + P_2 = 1$ .

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

$$f_{Gal}(x; \theta, \beta) = \frac{\theta^2((\beta + \beta\theta - \theta)x + 1)e^{-\theta x}}{\beta(1 + \theta)} \quad x > 0, \theta > 0, \beta > \frac{\theta}{1+\theta}. \quad (2.33)$$

**Remarque 2.7.1** Si  $\beta = 1$ , cette distribution est la distribution de Lindley  $LD(\theta)$ .

Par conséquent, le mode de cette distribution est :

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

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

$$F_{GaL}(x) = 1 - \frac{((\theta\beta + \beta - \theta)(\theta x + 1) + \theta)e^{-\theta x}}{\beta(1 + \theta)}; x > 0, \theta > 0, \beta > \frac{\theta}{1 + \theta}. \quad (2.34)$$

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

### Moments et mesures connexes

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 :

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

d'où

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

**Remarque 2.7.2** La fonction génératrice des moments pour  $X$  et  $S$  existent ( $\mathbb{E}(e^{tX}) < \infty$ ) 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 :

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

**Théorème 2.7.1** Soit  $X \sim GaLD(\theta, \beta)$ ,  $M = Mode(X)$ ,  $me = Median(X)$  et  $Moyenne(X) = \mu = E(X)$ . Alors

$$M < me < \mu.$$

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

### 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 2.3** 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 (3.30)

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

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. ■

### 2.7.1 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 2.7.2** *Pour tout  $\theta > 0, \beta > \frac{\theta}{1+\theta}$ , la fonction quantile de  $X$  qui suit la distribution de Gamma Lindley est :*

$$Q_X(u) = -\frac{\beta(1+\theta)}{\theta(\beta(1+\theta)-\theta)} - \frac{1}{\theta} W_{-1} \left( \frac{\beta(1+\theta)(y-1)}{\beta(1+\theta)-\theta} e^{-\frac{\beta(1+\theta)}{\beta(1+\theta)-\theta}} \right), \quad 0 < u < 1. \quad (3.38)$$

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

**Preuve.** Voir Théorème1( Zeghdoudi et Nedjar (2016)). ■

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

Étant donné un échantillon aléatoire  $X_1, \dots, X_n$ , de la distribution de Gamma Lindley (3.34), en utilisant le premier moment  $m$  et la variance  $s^2$  on a :

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

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 :

$$\hat{\theta}_{MoM} = \frac{1}{s^2 + m^2} \left( 2m + \sqrt{2\sqrt{-s^2 + m^2}} \right) \quad \text{et} \quad \hat{\beta}_{MoM} = \frac{\hat{\theta}}{(1+\hat{\theta})(2-\hat{\theta}m)}. \quad (3.35)$$

## 2.8 Comparaison des distributions

Dans cette section on confirme la flexibilité des distributions  $Gal$  et  $PsLD$  en considérant deux ensembles de données différentes utilisées par différents chercheurs (données de durée de survie). On étudie également les



distributions GaL, PsLD, Quasi Lindley, Lindley Généralisée et gamma. Dans chacune de ces distributions, les paramètres sont estimés en utilisant la méthode des moments.

Pour donner une signification à cette comparaison, on utilise les valeurs de log-vraisemblance négatives ( $-LL$ ), le critère d'information Akaike ( $AIC$ ), le critère d'information Bayésienne ( $BIC$ ) et le test statistique de Kolmogorov-Smirnov ( $K - S$ ).

### Application

On considère de Lawless (2003) deux séries de données réelles. La première, représente les temps de défaillances ( $mm$ ) pour un échantillon de quinze composants électroniques dans un test de la survie. 1.4, 5.1, 6.3, 10.8, 12.1, 18.5, 19.7, 22.2, 23.1, 24.5, 25.8, 27.1, 28.4, 29.7, 31.0, 32.3, 33.6, 34.9, 36.2, 37.5, 38.8, 40.1, 41.4, 42.7, 44.0, 45.3, 46.6, 47.9, 49.2, 50.5, 51.8, 53.1, 54.4, 55.7, 57.0, 58.3, 59.6, 60.9, 62.2, 63.5, 64.8, 66.1, 67.4, 68.7, 70.0, 71.3, 72.6, 73.9, 75.2, 76.5, 77.8, 79.1, 80.4, 81.7, 83.0, 84.3, 85.6, 86.9, 88.2, 89.5, 90.8, 92.1, 93.4, 94.7, 96.0, 97.3, 98.6, 99.9, 101.2, 102.5, 103.8, 105.1, 106.4, 107.7, 109.0, 110.3, 111.6, 112.9, 114.2, 115.5, 116.8, 118.1, 119.4, 120.7, 122.0, 123.3, 124.6, 125.9, 127.2, 128.5, 129.8, 131.1, 132.4, 133.7, 135.0, 136.3, 137.6, 138.9, 140.2, 141.5, 142.8, 144.1, 145.4, 146.7, 148.0, 149.3, 150.6, 151.9, 153.2, 154.5, 155.8, 157.1, 158.4, 159.7, 161.0, 162.3, 163.6, 164.9, 166.2, 167.5, 168.8, 170.1, 171.4, 172.7, 174.0, 175.3, 176.6, 177.9, 179.2, 180.5, 181.8, 183.1, 184.4, 185.7, 187.0, 188.3, 189.6, 190.9, 192.2, 193.5, 194.8, 196.1, 197.4, 198.7, 200.0, 201.3, 202.6, 203.9, 205.2, 206.5, 207.8, 209.1, 210.4, 211.7, 213.0, 214.3, 215.6, 216.9, 218.2, 219.5, 220.8, 222.1, 223.4, 224.7, 226.0, 227.3, 228.6, 229.9, 231.2, 232.5, 233.8, 235.1, 236.4, 237.7, 239.0, 240.3, 241.6, 242.9, 244.2, 245.5, 246.8, 248.1, 249.4, 250.7, 252.0, 253.3, 254.6, 255.9, 257.2, 258.5, 259.8, 261.1, 262.4, 263.7, 265.0, 266.3, 267.6, 268.9, 270.2, 271.5, 272.8, 274.1, 275.4, 276.7, 278.0, 279.3, 280.6, 281.9, 283.2, 284.5, 285.8, 287.1, 288.4, 289.7, 291.0, 292.3, 293.6, 294.9, 296.2, 297.5, 298.8, 300.1, 301.4, 302.7, 304.0, 305.3, 306.6, 307.9, 309.2, 310.5, 311.8, 313.1, 314.4, 315.7, 317.0, 318.3, 319.6, 320.9, 322.2, 323.5, 324.8, 326.1, 327.4, 328.7, 330.0, 331.3, 332.6, 333.9, 335.2, 336.5, 337.8, 339.1, 340.4, 341.7, 343.0, 344.3, 345.6, 346.9, 348.2, 349.5, 350.8, 352.1, 353.4, 354.7, 356.0, 357.3, 358.6, 359.9, 361.2, 362.5, 363.8, 365.1, 366.4, 367.7, 369.0, 370.3, 371.6, 372.9, 374.2, 375.5, 376.8, 378.1, 379.4, 380.7, 382.0, 383.3, 384.6, 385.9, 387.2, 388.5, 389.8, 391.1, 392.4, 393.7, 395.0, 396.3, 397.6, 398.9, 400.2, 401.5, 402.8, 404.1, 405.4, 406.7, 408.0, 409.3, 410.6, 411.9, 413.2, 414.5, 415.8, 417.1, 418.4, 419.7, 421.0, 422.3, 423.6, 424.9, 426.2, 427.5, 428.8, 430.1, 431.4, 432.7, 434.0, 435.3, 436.6, 437.9, 439.2, 440.5, 441.8, 443.1, 444.4, 445.7, 447.0, 448.3, 449.6, 450.9, 452.2, 453.5, 454.8, 456.1, 457.4, 458.7, 460.0, 461.3, 462.6, 463.9, 465.2, 466.5, 467.8, 469.1, 470.4, 471.7, 473.0, 474.3, 475.6, 476.9, 478.2, 479.5, 480.8, 482.1, 483.4, 484.7, 486.0, 487.3, 488.6, 489.9, 491.2, 492.5, 493.8, 495.1, 496.4, 497.7, 499.0, 500.3, 501.6, 502.9, 504.2, 505.5, 506.8, 508.1, 509.4, 510.7, 512.0, 513.3, 514.6, 515.9, 517.2, 518.5, 519.8, 521.1, 522.4, 523.7, 525.0, 526.3, 527.6, 528.9, 530.2, 531.5, 532.8, 534.1, 535.4, 536.7, 538.0, 539.3, 540.6, 541.9, 543.2, 544.5, 545.8, 547.1, 548.4, 549.7, 551.0, 552.3, 553.6, 554.9, 556.2, 557.5, 558.8, 560.1, 561.4, 562.7, 564.0, 565.3, 566.6, 567.9, 569.2, 570.5, 571.8, 573.1, 574.4, 575.7, 577.0, 578.3, 579.6, 580.9, 582.2, 583.5, 584.8, 586.1, 587.4, 588.7, 590.0, 591.3, 592.6, 593.9, 595.2, 596.5, 597.8, 599.1, 600.4, 601.7, 603.0, 604.3, 605.6, 606.9, 608.2, 609.5, 610.8, 612.1, 613.4, 614.7, 616.0, 617.3, 618.6, 619.9, 621.2, 622.5, 623.8, 625.1, 626.4, 627.7, 629.0, 630.3, 631.6, 632.9, 634.2, 635.5, 636.8, 638.1, 639.4, 640.7, 642.0, 643.3, 644.6, 645.9, 647.2, 648.5, 649.8, 651.1, 652.4, 653.7, 655.0, 656.3, 657.6, 658.9, 660.2, 661.5, 662.8, 664.1, 665.4, 666.7, 668.0, 669.3, 670.6, 671.9, 673.2, 674.5, 675.8, 677.1, 678.4, 679.7, 681.0, 682.3, 683.6, 684.9, 686.2, 687.5, 688.8, 690.1, 691.4, 692.7, 694.0, 695.3, 696.6, 697.9, 699.2, 700.5, 701.8, 703.1, 704.4, 705.7, 707.0, 708.3, 709.6, 710.9, 712.2, 713.5, 714.8, 716.1, 717.4, 718.7, 720.0, 721.3, 722.6, 723.9, 725.2, 726.5, 727.8, 729.1, 730.4, 731.7, 733.0, 734.3, 735.6, 736.9, 738.2, 739.5, 740.8, 742.1, 743.4, 744.7, 746.0, 747.3, 748.6, 749.9, 751.2, 752.5, 753.8, 755.1, 756.4, 757.7, 759.0, 760.3, 761.6, 762.9, 764.2, 765.5, 766.8, 768.1, 769.4, 770.7, 772.0, 773.3, 774.6, 775.9, 777.2, 778.5, 779.8, 781.1, 782.4, 783.7, 785.0, 786.3, 787.6, 788.9, 790.2, 791.5, 792.8, 794.1, 795.4, 796.7, 798.0, 799.3, 800.6, 801.9, 803.2, 804.5, 805.8, 807.1, 808.4, 809.7, 811.0, 812.3, 813.6, 814.9, 816.2, 817.5, 818.8, 820.1, 821.4, 822.7, 824.0, 825.3, 826.6, 827.9, 829.2, 830.5, 831.8, 833.1, 834.4, 835.7, 837.0, 838.3, 839.6, 840.9, 842.2, 843.5, 844.8, 846.1, 847.4, 848.7, 850.0, 851.3, 852.6, 853.9, 855.2, 856.5, 857.8, 859.1, 860.4, 861.7, 863.0, 864.3, 865.6, 866.9, 868.2, 869.5, 870.8, 872.1, 873.4, 874.7, 876.0, 877.3, 878.6, 879.9, 881.2, 882.5, 883.8, 885.1, 886.4, 887.7, 889.0, 890.3, 891.6, 892.9, 894.2, 895.5, 896.8, 898.1, 899.4, 900.7, 902.0, 903.3, 904.6, 905.9, 907.2, 908.5, 909.8, 911.1, 912.4, 913.7, 915.0, 916.3, 917.6, 918.9, 920.2, 921.5, 922.8, 924.1, 925.4, 926.7, 928.0, 929.3, 930.6, 931.9, 933.2, 934.5, 935.8, 937.1, 938.4, 939.7, 941.0, 942.3, 943.6, 944.9, 946.2, 947.5, 948.8, 950.1, 951.4, 952.7, 954.0, 955.3, 956.6, 957.9, 959.2, 960.5, 961.8, 963.1, 964.4, 965.7, 967.0, 968.3, 969.6, 970.9, 972.2, 973.5, 974.8, 976.1, 977.4, 978.7, 980.0, 981.3, 982.6, 983.9, 985.2, 986.5, 987.8, 989.1, 990.4, 991.7, 993.0, 994.3, 995.6, 996.9, 998.2, 999.5, 1000.8, 1002.1, 1003.4, 1004.7, 1006.0, 1007.3, 1008.6, 1009.9, 1011.2, 1012.5, 1013.8, 1015.1, 1016.4, 1017.7, 1019.0, 1020.3, 1021.6, 1022.9, 1024.2, 1025.5, 1026.8, 1028.1, 1029.4, 1030.7, 1032.0, 1033.3, 1034.6, 1035.9, 1037.2, 1038.5, 1039.8, 1041.1, 1042.4, 1043.7, 1045.0, 1046.3, 1047.6, 1048.9, 1050.2, 1051.5, 1052.8, 1054.1, 1055.4, 1056.7, 1058.0, 1059.3, 1060.6, 1061.9, 1063.2, 1064.5, 1065.8, 1067.1, 1068.4, 1069.7, 1071.0, 1072.3, 1073.6, 1074.9, 1076.2, 1077.5, 1078.8, 1080.1, 1081.4, 1082.7, 1084.0, 1085.3, 1086.6, 1087.9, 1089.2, 1090.5, 1091.8, 1093.1, 1094.4, 1095.7, 1097.0, 1098.3, 1099.6, 1100.9, 1102.2, 1103.5, 1104.8, 1106.1, 1107.4, 1108.7, 1110.0, 1111.3, 1112.6, 1113.9, 1115.2, 1116.5, 1117.8, 1119.1, 1120.4, 1121.7, 1123.0, 1124.3, 1125.6, 1126.9, 1128.2, 1129.5, 1130.8, 1132.1, 1133.4, 1134.7, 1136.0, 1137.3, 1138.6, 1139.9, 1141.2, 1142.5, 1143.8, 1145.1, 1146.4, 1147.7, 1149.0, 1150.3, 1151.6, 1152.9, 1154.2, 1155.5, 1156.8, 1158.1, 1159.4, 1160.7, 1162.0, 1163.3, 1164.6, 1165.9, 1167.2, 1168.5, 1169.8, 1171.1, 1172.4, 1173.7, 1175.0, 1176.3, 1177.6, 1178.9, 1180.2, 1181.5, 1182.8, 1184.1, 1185.4, 1186.7, 1188.0, 1189.3, 1190.6, 1191.9, 1193.2, 1194.5, 1195.8, 1197.1, 1198.4, 1199.7, 1201.0, 1202.3, 1203.6, 1204.9, 1206.2, 1207.5, 1208.8, 1210.1, 1211.4, 1212.7, 1214.0, 1215.3, 1216.6, 1217.9, 1219.2, 1220.5, 1221.8, 1223.1, 1224.4, 1225.7, 1227.0, 1228.3, 1229.6, 1230.9, 1232.2, 1233.5, 1234.8, 1236.1, 1237.4, 1238.7, 1240.0, 1241.3, 1242.6, 1243.9, 1245.2, 1246.5, 1247.8, 1249.1, 1250.4, 1251.7, 1253.0, 1254.3, 1255.6, 1256.9, 1258.2, 1259.5, 1260.8, 1262.1, 1263.4, 1264.7, 1266.0, 1267.3, 1268.6, 1269.9, 1271.2, 1272.5, 1273.8, 1275.1, 1276.4, 1277.7, 1279.0, 1280.3, 1281.6, 1282.9, 1284.2, 1285.5, 1286.8, 1288.1, 1289.4, 1290.7, 1292.0, 1293.3, 1294.6, 1295.9, 1297.2, 1298.5, 1299.8, 1301.1, 1302.4, 1303.7, 1305.0, 1306.3, 1307.6, 1308.9, 1310.2, 1311.5, 1312.8, 1314.1, 1315.4, 1316.7, 1318.0, 1319.3, 1320.6, 1321.9, 1323.2, 1324.5, 1325.8, 1327.1, 1328.4, 1329.7, 1331.0, 1332.3, 1333.6, 1334.9, 1336.2, 1337.5, 1338.8, 1340.1, 1341.4, 1342.7, 1344.0, 1345.3, 1346.6, 1347.9, 1349.2, 1350.5, 1351.8, 1353.1, 1354.4, 1355.7, 1357.0, 1358.3, 1359.6, 1360.9, 1362.2, 1363.5, 1364.8, 1366.1, 1367.4, 1368.7, 1370.0, 1371.3, 1372.6, 1373.9, 1375.2, 1376.5, 1377.8, 1379.1, 1380.4, 1381.7, 1383.0, 1384.3, 1385.6, 1386.9, 1388.2, 1389.5, 1390.8, 1392.1, 1393.4, 1394.7, 1396.0, 1397.3, 1398.6, 1399.9, 1401.2, 1402.5, 1403.8, 1405.1, 1406.4, 1407.7, 1409.0, 1410.3, 1411.6, 1412.9, 1414.2, 1415.5, 1416.8, 1418.1, 1419.4, 1420.7, 1422.0, 1423.3, 1424.6, 1425.9, 1427.2, 1428.5, 1429.8, 1431.1, 1432.4, 1433.7, 1435.0, 1436.3, 1437.6, 1438.9, 1440.2, 1441.5, 1442.8, 1444.1, 1445.4, 1446.7, 1448.0, 1449.3, 1450.6, 1451.9, 1453.2, 1454.5, 1455.8, 1457.1, 1458.4, 1459.7, 1461.0, 1462.3, 1463.6, 1464.9, 1466.2, 1467.5, 1468.8, 1470.1, 1471.4, 1472.7, 1474.0, 1475.3, 1476.6, 1477.9, 1479.2, 1480.5, 1481.8, 1483.1, 1484.4, 1485.7, 1487.0, 1488.3, 1489.6, 1490.9, 1492.2, 1493.5, 1494.8, 1496.1, 1497.4, 1498.7, 1499.9, 1501.2, 1502.5, 1503.8, 1505.1, 1506.4, 1507.7, 1509.0, 1510.3, 1511.6, 1512.9, 1514.2, 1515.5, 1516.8, 1518.1, 1519.4, 1520.7, 1522.0, 1523.3, 1524.6, 1525.9, 1527.2, 1528.5, 1529.8, 1531.1, 1532.4, 1533.7, 1535.0, 1536.3, 1537.6, 1538.9, 1540.2, 1541.5, 1542.8, 1544.1, 1545.4, 1546.7, 1548.0, 1549.3, 1550.6, 1551.9, 1553.2, 1554.5, 1555.8, 1557.1, 1558.4, 1559.7, 1561.0, 1562.3, 1563.6, 1564.9, 1566.2, 1567.5, 1568.8, 1570.1, 1571.4, 1572.7, 1574.0, 1575.3, 1576.6, 1577.9, 1579.2, 1580.5, 1581.8, 1583.1, 1584.4, 1585.7, 1587.0, 1588.3, 1589.6, 1590.9, 1592.2, 1593.5, 1594.8, 1596.1, 1597.4, 1598.7, 1599.9, 1601.2, 1602.5, 1603.8, 1605.1, 1606.4, 1607.7, 1609.0, 1610.3, 1611.6, 1612.9, 1614.2, 1615.5, 1616.8, 1618.1, 1619.4, 1620.7, 1622.0, 1623.3, 1624.6, 1625.9, 1627.2, 1628.5, 1629.8, 1631.1, 1632.4, 1633.7, 1635.0, 1636.3, 1637.6, 1638.9, 1640.2, 1641.5, 1642.8, 1644.1, 1645.4, 1646.7, 1648.0, 1649.3, 1650.6, 1651.9, 1653.2, 1654.5, 1655.8, 1657.1, 1658.4, 1659.7, 1661.0, 1662.3, 1663.6, 1664.9, 1666.2, 1667.5, 1668.8, 1670.1, 1671.4, 1672.7, 1674.0, 1675.3, 1676.6, 1677.9, 1679.2, 1680.5, 1681.8, 1683.1, 1684.4, 1685.7, 1687.0, 1688.3, 1689.6, 1690.9, 1692.2, 1693.5, 1694.8, 1696.1, 1697.4, 1698.7, 1699.9, 1701.2, 1702.5, 1703.8, 1705.1, 1706.4, 1707.7, 1709.0, 1710.3, 1711.6, 1712.9, 1714.2, 1715.5, 1716.8, 1718.1, 1719.4, 1720.7, 1722.0, 1723.3, 1724.6, 1725.9, 1727.2, 1728.5, 1729.8, 1731.1, 1732.4, 1733.7, 1735.0, 1736.3, 1737.6, 1738.9, 1740.2, 1741.5, 1742.8, 1744.1, 1745.4, 1746.7, 1748.0, 1749.3, 1750.6, 1751.9, 1753.2, 1754.5, 1755.8, 1757.1, 1758.4, 1759.7, 1761.0, 1762.3, 1763.6, 1764.9, 1766.2, 1767.5, 1768.8, 1770.1, 1771.4, 1772.7, 1774.0, 1775.3, 1776.6, 1777.9, 1779.2, 1780.5, 1781.8, 1783.1, 1784.4, 1785.7, 1787.0, 1788.3, 1789.6, 1790.9, 1792.2, 1793.5, 1794.8, 1796.1, 1797.4, 1798.7, 1799.9, 1801.2, 1802.5, 1803.8, 1805.1, 1806.4, 1807.7, 1809.0, 1810.3, 1811.6, 1812.9, 1814.2, 1815.5, 1816.8, 1818.1, 1819.4, 1820.7, 1822.0, 1823.3, 1824.6, 1825.9, 1827.2, 1828.5, 1829.8, 1831.1, 1832.4, 1833.7, 1835.0, 1836.3, 1837.6, 1838.9, 1840.2, 1841.5, 1842.8, 1844.1, 1845.4, 1846.7, 1848.0, 1849.3, 1850.6, 1851.9, 1853.2, 1854.5, 1855.8, 1857.1, 1858.4, 1859.7, 1861.0, 1862.3, 1863.6, 1864.9, 1866.2, 1867.5, 1868.8, 1870.1, 1871.4, 1872.7, 1874.0, 1875.3, 1876.6, 1877.9, 1879.2, 1880.5, 1881.8, 1883.1, 1884.4, 1885.7, 1887.0, 1888.3, 1889.6, 1890.9, 1892.2, 1893.5, 1894.8, 1896.1, 1897.4, 1898.7, 1899.9, 1901.2, 1902.5, 1903.8, 1905.1, 1906.4, 1907.7, 1909.0, 1910.3, 1911.6, 1912.9, 1914.2, 1915.5, 1916.8, 1918.1, 1919.4, 1920.7, 1922.0, 1923.3, 1924.6, 1925.9, 1927.2, 1928.5, 1929.8, 1931.1, 1932.4, 1933.7, 1935.0, 1936.3, 1937.6, 1938.9, 1940.2, 1941.5, 1942.8, 1944.1, 1945.4, 1946.7, 1948.0, 1949.3, 1950.6, 1951.9, 1953.2, 1954.5, 1955.8, 1957.1, 1958.4, 1959.7, 1961.0, 1962.3, 1963.6, 1964.9, 1966.2, 1967.5, 1968.8, 1970.1, 1971.4, 1972.7, 1974.0, 1975.3, 1976.6, 1977.9, 1979.2, 1980.5, 1981.8, 1983.1, 1984.4, 1985.7, 1987.0, 1988.3, 1989.6, 1990.9, 1992.2, 1993.5, 1994.8, 1996.1, 1997.4, 1998.7, 1999.9, 2001.2, 2002.5, 2003.8, 2005.1, 2006.4, 2007.7, 2009.0, 2010.3, 2011.6, 2012.9, 2014.2, 2015.5, 2016.8, 2018.1, 2019.4, 2020.7, 2022.0, 2023.3, 2024.6, 2025.9, 2027.2, 2028.5, 2029.8, 2031.1, 2032.4, 2033.7, 2035.0, 2036.3, 2037.6, 2038.9, 2040.2, 2041.5, 2042.8, 2044.1, 2045.4, 2046.7, 2048.0, 2049.3, 2050.6, 2051.9, 2053.2, 2054.5, 2055.8, 2057.1, 2058.4, 2059.7, 2061.0, 2062.3, 2063.6, 2064.9, 2066.2, 2067.5, 2068.8, 2070.1, 2071.4, 2072.7, 2074.0, 2075.3, 2076.6, 2077.9, 2079.2, 2080.5, 2081.8, 2083.1, 2084.4, 2085.7, 2087.0, 2088.3, 2089.6, 2090.9, 2092.2, 2093.5, 2094.8, 2096.1, 209

que les distributions GaLD et PsLD sont beaucoup mieux que les distributions quasi Lindley, Lindley Généralisée et gamma pour les deux séries.

- D'après les valeurs BIC et AIC les meilleures distributions sont GaLD et PsLD. Aussi, on peut observer que les distributions Gamma Lindley et PsLD fournissent plus petit  $k - S$  par rapport aux distributions quasi Lindley, Lindley Généralisée et gamma.

## CHAPITRE 3

# Distribution composite Rayleigh-Pareto et ses applications

### Sommaire

---

<b>3.1 Le modèle composite Pareto-Rayleigh</b> . . . . .	<b>39</b>
<b>3.2 Génération de valeurs aléatoires à partir de la</b> <b>distribution composite de Rayleigh-Pareto</b> . . .	<b>41</b>
<b>3.3 Exemples numériques</b> . . . . .	<b>43</b>

---

Ce chapitre propose une nouvelle distribution, appelée distribution **composite de Rayleigh-Pareto** (CRP). Diverses propriétés statistiques telles que la méthode du moment et l'estimation du maximum de vraisemblance sont établies. Comparée à la distribution composite log-normale-Pareto, la densité résultante a une forme similaire et une queue plus grande. C'est pourquoi nous nous attendons à ce qu'un tel modèle convienne mieux que le modèle composite Log-normal-Pareto pour certaines données relatives

aux demandes d'indemnisation. Des exemples simulés et des ensembles de données d'assurance incendie (données réelles sur les pertes d'assurance incendie algériennes et danoises) ont été analysés pour montrer l'importance et l'applicabilité de cette nouvelle distribution.

### 3.1 Le modèle composite Pareto-Rayleigh

Soit  $X$  une variable aléatoire avec fonction de densité

$$f(x) = \begin{cases} cf_1(x) & \text{si } 0 \leq x \leq \theta \\ cf_2(x) & \text{si } \theta \leq x \leq +\infty \end{cases}$$

où  $f_1$  est une densité de Rayleigh à un paramètre,  $f_2$  est une densité de Pareto à deux paramètres et  $c$  est la constante de normalisation. C'est

$$f_1(x) = \frac{x}{\lambda^2} \exp\left(-\frac{x^2}{2\lambda^2}\right); x > 0$$

et

$$f_2(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}}; x > \theta$$

où  $\lambda > 0$ ,  $\alpha > 0$ ,  $\theta > 0$  sont les paramètres inconnus.

Pour obtenir une fonction de densité lisse combiné, nous imposons des conditions de continuité et différentiable au point de seuil  $\theta$ , c'est-à-dire

$$\begin{cases} f_1(\theta) = f_2(\theta) \\ \frac{d}{dx}f_1(\theta) = \frac{d}{dx}f_2(\theta) \end{cases}$$

Ces deux restrictions donnent

$$\begin{cases} \frac{\theta}{\lambda^2} \exp\left(-\frac{\theta^2}{2\lambda^2}\right) = \frac{\alpha}{\theta} \\ \frac{(\theta^2 - \lambda^2)}{\lambda^4} \exp\left(-\frac{\theta^2}{2\lambda^2}\right) = \frac{\alpha(\alpha+1)}{\theta^2} \end{cases}$$

et, après calcul, on obtient

$$\begin{cases} \beta \exp(-\beta) - \beta + 1 = 0 \\ \alpha = 2\beta \exp(-\beta) \end{cases} \quad \text{avec } \beta = \frac{\theta^2}{2\lambda^2}$$

Résoudre la seconde équation par des méthodes numériques, il en résulte la solution

$$\begin{cases} \beta = 1.35 \\ \alpha = 0.69995 \end{cases}$$

Afin de trouver la constante de normalisation, nous imposons la condition  $\int_0^{+\infty} f(x) = 1$ , qui donne

$$\begin{aligned} c &= \frac{1}{2 - \exp(-\beta)} \\ &= 0.57446 \end{aligned}$$

Ainsi,  $f(x)$  peut être exprimé par

$$f(x) = \begin{cases} 1.551 \frac{x}{\theta^2} \exp\left(-1.35 \frac{x^2}{\theta^2}\right) & \text{si } 0 \leq x \leq \theta \\ \frac{0.40209}{x^{1.69995}} \theta^{0.69995} & \text{si } \theta \leq x \leq +\infty \end{cases} \quad (3.1)$$

**Proposition 3.1** *La fonction de distribution cumulative (c.d.f.) de ce modèle composite est*

$$F(x) = \begin{cases} 0.57446 \left(1 - \exp\left(-1.35 \frac{x^2}{2\theta^2}\right)\right) & \text{si } 0 \leq x \leq \theta \\ 1 - 0.57446 \left(\frac{\theta}{x}\right)^{0.69995} & \text{si } \theta \leq x \leq +\infty \end{cases} \quad (3.2)$$

**Proposition 3.2**

$$x = F^{-1}(u) = \begin{cases} \sqrt{-\frac{\theta^2}{1.35} \ln(1 - 1.7408u)} & \text{si } 0 \leq u \leq p \\ \theta \left(\frac{0.57446}{1-u}\right)^{0.69995} & \text{si } p \leq u \leq +\infty \end{cases}$$

**Proposition 3.3** *Le mode unique de la distribution composite Rayleigh-Pareto est*

$$x_{\text{mod}} = 0.60858\theta.$$

**Proposition 3.4** *Le 1er moment initial de la distribution composite de Rayleigh-Pareto est donné par*

$$E(X) = 0.24528\theta$$

## 3.2 Génération de valeurs aléatoires à partir de la distribution composite de Rayleigh-Pareto

### 3.2.1 Estimation des paramètres

Dans cette section, nous présenterons deux algorithmes pour l'estimation du paramètre inconnu  $\theta$ .

Soit  $x_1; x_2; \dots; x_n$  est un échantillon aléatoire du modèle composite de Rayleigh-Pareto. En supposant que  $x_1 \leq x_2 \leq \dots \leq x_n$  et nous supposons que le paramètre inconnu  $\theta$  se situe entre la  $m^{\text{ième}}$  observation et la  $m+1^{\text{ième}}$  observation, c.-à-d.

$$x_m \leq \theta \leq x_{m+1}$$

### 3.2.2 Le premier algorithme Une procédure ad-hoc basée sur les centiles

La procédure ad-hoc suivante fournit un formulaire fermé pour le paramètre  $\theta$ , estimé à l'aide de centiles. Sur la base des centiles, le paramètre  $\theta$  peut être estimé, en  $p^{\text{th}}$  percentile, où  $p = F(\theta)$

$$p = 0.57446 \left( 1 - \exp \left( -\frac{1.35}{2} \right) \right) = 0.42554$$

De Klugman et al. (1998), nous avons une estimation empirique lisse du  $p^{\text{th}}$  percentile donné par

$$\hat{\theta} = hx_m + (1 - h)x_{m+1}$$

avec

$$\begin{aligned} m &= [(n+1)h] \\ h &= (n+1)p - m \end{aligned}$$

Ici  $[a]$  indique le plus grand entier plus petit ou égal à  $a$ .

La distribution de Pareto ou la distribution de Rayleigh sera un modèle supérieur à la distribution composite de Pareto-Rayleigh selon laquelle  $\theta$  est plus proche de  $x_1$  et  $x_n$ .

### 3.2.3 Le deuxième algorithme

Dans Teodorescu et Vernic [3], l'algorithme suivant basé sur la méthode du maximum de vraisemblance (ML) a été suggéré.

La fonction de vraisemblance est

$$\begin{aligned} L(x_1; x_2; \dots; x_n, \theta) &= \prod_{i=1}^n f(x_i) \\ &= \prod_{i=1}^m f(x_i) \prod_{i=m+1}^n f(x_i) \\ &= \prod_{i=1}^m 1.551 \frac{x_i}{\theta^2} \exp\left(-1.35 \frac{x_i^2}{\theta^2}\right) \prod_{i=m+1}^n \frac{0.40209}{x_i^{1.69995}} \theta^{0.69995} \\ &= (1.551)^m (0.40209)^{n-m} \theta^{0.69995(n-m)-2m} \\ &\quad \frac{\prod_{i=1}^m x_i}{\prod_{i=m+1}^n x_i^{1.69995}} \exp\left(-\frac{1.35}{\theta^2} \sum_{i=1}^m x_i^2\right) \\ &= k \theta^{0.69995n-2.69995m} \exp\left(\frac{-1.35 \sum_{i=1}^m x_i^2}{\theta^2}\right) \end{aligned}$$

avec

$$k = (1.551)^m (0.40209)^{n-m} \frac{\prod_{i=1}^m x_i}{\prod_{i=m+1}^n x_i^{1.69995}}$$

Étape 1 : Différencier  $\ln L$  par rapport à  $\theta$  donne

$$\frac{d \ln L}{d\theta} = \frac{0.699\,95n - 2.700\,0m}{\theta} + \frac{2.7}{\theta^3} \sum_{i=1}^m x_i^2 \quad (3.3)$$

D'où la solution de l'équation de vraisemblance  $\frac{d \ln L}{d\theta} = 0$  est

$$\hat{\theta}^2 = \frac{2.7 \sum_{i=1}^m x_i^2}{2.700\,0m - 0.699\,95n} \quad (3.4)$$

$$\hat{\theta} = \sqrt{\frac{2.7 \sum_{i=1}^m x_i^2}{2.700\,0m - 0.699\,95n}}$$

avec

$$m > 0 : 25925n$$

si  $x_m \leq \hat{\theta} \leq x_{m+1}$  alors l'estimateur de ML est

$$\hat{\theta}^{ML} = \hat{\theta}$$

Étape 2 : Si l'étape 1 ne donne aucune solution pour  $t$ , vous pouvez utiliser la procédure suivante.

Pour chaque  $m$  ( $m = 1; 2; \dots; n - 1$ ), évaluez  $\theta$  à partir de l'équation (4).

Si  $x_m \leq \hat{\theta} \leq x_{m+1}$  alors l'estimateur de ML est

$$\hat{\theta}^{ML} = \hat{\theta}$$

Si ceci n'est pas satisfait, alors nous devons passer à la valeur suivante de  $m$ .

### 3.3 Exemples numériques

Dans cette section, la procédure d'estimation décrite dans la section 3 a été expliquée à l'aide de deux échantillons de données générés à partir du modèle Rayleigh pareto. L'algorithme de génération utilisé est basé sur l'inversion de c.d.f. (2) (méthode d'inversion présentée à la sous-section 2.1).



### 3.3.1 Exemple

L'ensemble de données composé de 100 valeurs a été échantillonné à partir d'une population de Rayleigh-Pareto avec le paramètre  $\theta = 5$  (voir tableau 1).

Tableau 1. 100 valeurs de Rayleigh-Pareto pour  $\theta = 5$

0.062227	0.08189	0.08360	0.118 93	0.120 32
0.96591	1.033 3	1.043 2	1.488 4	1.673
2.4295	2.525 6	2.620 9	2.927	3.049 5
4.118 4	4.224 4	4.885 9	4.889 1	4.934 8
4.9010	4.9853	6.494 4	6.750 6	7.311 1
8.712 3	8.822 0	9.273 6	9.446 6	9.715 2
12.525	12.833	13.705	14.018	16.975
25.354	26.153	36.598	39.859	40.844
83.511	93.847	179.06	197.63	202.50
607.06	807.75	986.43	1409.1	1647.0
0.231 5	0.372 07	0.393 59	0.557 29	0.718 96
1.701 3	1.952 9	2.036 5	2.175 9	2.325 6
3.324 5	3.434 3	3.569 6	3.697 2	3.841 0
4.967 8	4.984 5	5.068 3	5.225 8	5.603 6
7.649 9	7.978 9	8.116 5	8.293 7	8.665 9
9.9481	9.995 7	10.440	10.543	11.080
17.790	17.845	18.315	19.836	22.665
49.691	50.788	55.562	63.398	71.889
242.33	374.76	363.58	439.32	503.91
1783.8	3157.8	6380.8	8319.4	8439.3

Les valeurs estimées du paramètre sont :

- par l'algorithme 1 : For  $m = 42$ ,  $\hat{\theta}_1 = 5.1012$
- par l'algorithme 2 : MLE Step 1 :  $\hat{\theta}_2 = 4.9812$
- par l'algorithme 3 : MLE Step 2 :  $\hat{\theta}_3 = 4.9810$ .

Nous remarquons que cette fois, l'algorithme 2, étape 1, donne une valeur plus précise.

Nous avons également appliqué le test  $\chi^2$  pour vérifier l'ajustement de distribution, et les résultats pour  $\hat{\theta}_3$  sont présentés dans le tableau 2.

Tableau 2. Données groupées et test  $\chi^2$ . (les colonnes 2 et 3 résultent de l'échantillon de données, tandis que la colonne 4 est calculée à l'aide de la fonction de distribution de Rayleigh-Pareto)

<i>Classes</i>	<i>Frequencies, <math>n_i</math></i>	<i>Theoretical freq., <math>p_i</math></i>	$\frac{n(f_i - p_i)^2}{p_i}$ (CRP)
[0,2)	11	0.107 65	$5.130 1 \times 10^{-3}$
[2,4)	23	0.216 33	$8.638 1 \times 10^{-2}$
[4,6)	22	0.270 91	0.956 71
[6,10)	19	0.154 14	0.834 27
[10,60)	23	0.256 3	0.269 88
[60,300)	2	0.069 15	3.493 5
$\Sigma$	100	$\chi^2 distance:$	5.6459

The  $\chi^2$  distances calculated for the three estimated values of the parameters are

$$\begin{aligned} d^2(\hat{\theta}_1) &= 4.60 \\ d^2(\hat{\theta}_2) &= 5.56 \\ d^2(\hat{\theta}_3) &= 5.64 \end{aligned}$$

le test  $\chi^2$  accepte le modèle de Rayleigh-Pareto pour les trois valeurs du paramètre comme prévu. La chose intéressante est que, contrairement aux

attentes,  $d^2(\hat{\theta}_2)$  est minimale.

### 3.3.2 Qualité de l'ajustement

Dans cette sous-section, nous appliquons le modèle composite de Rayleigh-Pareto à un jeu de données d'assurance réel. Les deux ensembles de données, les 100 premiers dommages d'assurance incendie algériens (société SAA) exprimés en millions de dinars algériens pour la période allant de 2010 à 2016 (voir <http://www.saa.dz/>).

Number of obs	Mean	Standard Deviation	First Quartile
100	1.273	4.2467	0.1823
Median	Third Quartile	Skewness	Kurtosis
0.6188	1.7432	14.215	194.65

La deuxième 2156 pertes d'assurance incendie danoise se trouve dans le module complémentaire R *SMPracticals* disponible sur le site Web de CRAN <http://cran.r-project.org/>. Les pertes sont exprimées en millions de couronnes danoises (DKK) des années 1980 à 1990 inclusivement et ont été ajustées pour refléter les valeurs de 1985. McNeil (1997) et Resnick (1997), entre autres, ont analysé la partie supérieure de ces données.

Number of obs	Mean	Standard Deviation		
2156	2.3973	8.5274		
First Quartile	Median	Third Quartile	Skewness	Kurtosis
0.33081	0.78178	1.97249	18.722	484.58

#### **Pour le premier ensemble de données (algérien)**

Nous fournissons dans le tableau 3 les valeurs estimées des modèles ajustés et les valeurs de  $-LL$  (log-vraisemblance négatif (les valeurs les plus petites sont bonnes)) et  $AIC$  (critère d'information d'Akaike (les valeurs les plus élevées sont bonnes)) évaluées par les estimateurs de vraisemblance

maximale.

Distributions	Parametres	$-LL$	$AIC$
Lognormal	$\hat{\mu} = 0.567, \hat{\sigma} = 0.623$	4.56	8.96
Pareto	$\hat{\theta} = 0.234, \hat{\alpha} = 0.460$	6.27	12.85
Gamma	$\hat{\lambda} = 0.351, \hat{\alpha} = 1.157$	5.98	10.54
Weibull	$\hat{\theta} = 2.752, \hat{\tau} = 0.817$	6.02	10.78
c Rayleigh-P	$\hat{\theta} = 2.654$	4.08	8.76
c Exponential-P	$\hat{\theta} = 1.867$	4.58	9.75
c Lognormal-P	$\hat{\theta} = 1.241, \hat{\alpha} = 1.267$	4.28	8.92

Tableau 3 Valeurs estimées des modèles ajustés et  $-LL$ ,  $AIC$  pour les données de sinistres algériens dues aux incendies.

**Pour le deuxième ensemble de données (danois)**

We use a same analysis , we find

Distributions	Parameters	$-LL$	$AIC$
Lognormal	$\hat{\mu} = 0.6718, \hat{\sigma} = 0.7323$	4,434	8,872
Pareto	$\hat{\theta} = 0.3134, \hat{\alpha} = 0.5460$	5,675	11,354
Gamma	$\hat{\lambda} = 0.4107, \hat{\alpha} = 1.2578$	5,243	10,490
Weibull	$\hat{\theta} = 2.9531, \hat{\tau} = 0.9476$	5,270	10,544
c Rayleigh-P	$\hat{\theta} = 2.848$	3.568	7.718
c Exponential-P	$\hat{\theta} = 2.477$	6.585	12.56
c Lognormal-P	$\hat{\theta} = 1.3851, \hat{\alpha} = 1.4363$	3,878	7,760

Tableau 4 Valeurs estimées des modèles ajustés et  $-LL$ ,  $AIC$  pour les données danoises d'incendie en cas d'incendie.

**Remarque 3.3.1** Les paramètres sont estimés par maximum de vraisemblance (ML).

**Remarque 3.3.2** Les valeurs de  $-LL$  et  $AIC$  des deux ensembles de données montrent que le modèle composite de Rayleigh-Pareto offre un meilleur

ajustement que les distributions classiques et le composite composite Exponential-Pareto, composite Lognormal-Pareto.

## Conclusion générale et perspectives

**A**u cours de notre recherche, nous avons adopté deux méthodes pour créer de nouvelles distributions de probabilités basées sur des distributions de probabilités connues. Surtout la famille des distributions de probabilités exponentielles, pour son importance dans l'étude des phénomènes de la durée de vie et d'actuariat. Le premier est par mélange convexe linéaire et le second est en installant deux distributions aux deux extrémités du domaine de définition. Dans les deux méthodes, le respect des conditions de distribution de probabilité de la condition de continuité et d'ajustement est le principal facteur pour déterminer les nouveaux coefficients de distribution de probabilité. À travers les transactions obtenues et en étudiant leurs évolutions, nous avons remarqué qu'ils sont plus flexibles pour traiter différents phénomènes statistiques par rapport aux distributions de probabilité précédentes exponentielle, Gamma, Weibul, Pareto .....

Dans le premier cas, nous donnons des distributions de probabilité avec un mélange linéaire convexe de différentes distributions de probabilités gamma et Lindley. notre travail s'est appuyée sur une étude comparative entre les nouvelles différentes distributions en utilisant des données réelles.

Dans le second cas, nous introduisons une nouvelle distribution de pro-

babilité en combinons des distributions de probabilité de Rayleith et Pareto. Une simulation et une étude comparative entre plusieurs distributions ont été données.

## 3.4 Annexe

### 3.4.1 Sélection de model avec différents critères

Dans le cas où on à des modèles estimés par une méthode du maximum de vraisemblance, Le test du rapport des vraisemblances est souvent utilisé pour comparer ces modèles deux à deux. Il ne s'applique qu'à des modèles emboîtés (dérivant l'un de l'autre par ajout ou suppression de termes) et il est supposé que les deux modèles comparés ajustent correctement les données. Quand de nombreux modèles doivent être comparés entre eux, le risque de rejeter l'hypothèse nulle alors qu'elle est vraie augmente.

Pour résoudre ce problème, une solution possible (il y en a d'autres) consiste à comparer les modèles en utilisant le critère d'information d'Akaike (Pour plus de détails voir Akaike,H.,1974)

#### **Le critère d'information d'Akaike (AIC) et (AICc)**

Le critère d'information d'Akaike est donnée par :

$$AIC = -2 \log(L) + 2k,$$

où  $L$  est la vraisemblance maximisée et  $k$  le nombre de paramètres dans la modél. Avec ce critère, la déviance du modèle ( $-2 \log(L)$ ) est pénalisée par 2 fois le nombre de paramètres. L'AIC représente donc un compromis entre le biais (qui diminue avec le nombre de paramètres) et la parcimonie (nécessité de décrire les données avec le plus petit nombre de paramètres possible).

-La rigueur voudrait que tous les modèles comparés dérivent tous d'un même « complet » inclus dans la liste des modèles comparés.

-Il est nécessaire de vérifier que ce modèle « complet » ajuste correctement les données.

-Le meilleur modèle est celui possédant l'AIC le plus faible.

Lorsque le nombre de paramètres  $k$  est grand par rapport au nombre d'observations  $n$ , c'est-à-dire si  $n/k < 40$ , il est recommandé d'utiliser l'AIC corrigé.

Le critère d'information d'Akaike corrigé AICc, est défini par :

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}.$$

### **Le critère BIC**

Dans le cas où on a plusieurs modèles sont très proches l'un de l'autre, et il est délicat de décider lequel d'entre eux est réellement le meilleur. Une approche possible est d'utiliser l'ensemble de ces modèles pour réaliser les inférences.

A cet effet, la tendance actuelle est plutôt de se baser sur le BIC.

Le Bayesian information criterion BIC est défini par :

$$BIC = -2 \log(L) + k \log(n).$$

Le BIC a été initialement proposé (Schwartz, 1978) pour sélectionner les modèles dans le cas de grands échantillons (plusieurs milliers d'observations) pour lesquels l'AIC et l'AICc ont tendance à sélectionner des modèles comportant de nombreuses variables explicatives, le BIC aboutit à des modèles plus parcimonieux. Cependant, les bases théoriques sous-tendant les deux approches (AIC, BIC) sont différentes, l'utilisation de l'AIC étant en premier lieu dans un objectif de prédiction, et nom de décision vis-à-vis de la



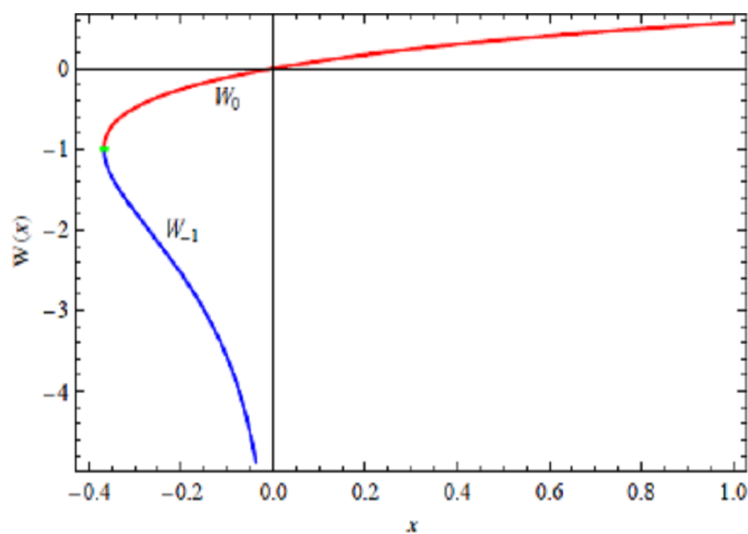


FIG. 3.1 – Les deux branches de la fonction de Lambert

signification statistique des paramètres retenus dans le modèle.

### 3.4.2 Fonction W de Lambert

Lambert en 1758 définit une fonction complexe multivaluée appelé fonction de Lambert ( $W$ ) qui est la solution de l'équation suivante

$$W(z) \exp(W(z)) = z. \quad (3.5)$$

Où  $z$  est un nombre complexe. Si  $z$  est un nombre réel sachant que  $z \geq -1/e$  alors  $W(z)$  devient une fonction réelle et elle a deux branches réelles possibles. La première branche réelle qui a des valeurs dans l'intervalle  $]-\infty, -1]$  c'est la branche négative et notée par  $W_{-1}$ . et la deuxième branche réelle qui a des valeurs dans l'intervalle  $[-1, \infty[$  est appelée la branche principale et notée  $W_0$ . Les deux réelles branches de  $W$  sont représentés par la figure.1.1. Lémeray (1897) montre qu'il y a d'autres équations peuvent être résolues en termes de la fonction (3.5)  $W$  de Lambert.

## BIBLIOGRAPHIE

- [1] Akaike H.(1974), A new look at statistical model identification, IEEE Transactions on Automatic Control AU-19 : 716-722.
- [2] Alzaatreh A, Famoye F, Lee C (2012). Gamma-Pareto distribution and its applications. Journal of Modern Applied Statistical Methods. 11(1), 78–94
- [3] Alzaatreh A, Famoye F, Lee C (2013). A new method for generating families of continuous distributions. Metron 71 :63–79.
- [4] Cebrian A, Denuit M, Lambert Ph (2003). Generalized Pareto fit to the society of Actuaries' large claims database. North American Actuarial Journal 7, 18-36.
- [5] Cooray K, Ananda M.A (2005). Modeling actuarial data with a composite Lognormal-Pareto model. Scand. Actuar. J. 5 , 321–334.
- [6] Cordeiro G.M, de Castro M (2011). A new family of generalized distributions. Journal of Statistical Computation and Simulation. 81(7), 883–898.

## Bibliographie

---

- [7] Eugene N, Lee C, Famoye F (2002), Beta-normal distribution and its applications, *Commun Stat Theory* 31,497-512.
- [8] Ghitany M.E, Atieh B, Nadarajah S. (2008). *Lindley distribution and its application*. *Mathematics and Computers in Simulation* 78 493–506
- [9] Ghitany M.E, Alqallaf F, Al-Mutairi D.K, Husain H.A (2011). *A two-parameter weighted Lindley distribution and its applications to survival data*. *Mathematics and Computers in Simulation* 81 (2011) 1190–1201.
- [10] Ghitany M.E, Al-Mutairi D.K, Balakrishnan N, Al-Enezi L.J (2013). *Power Lindley distribution and associated inference*. *Computational Statistics and Data Analysis* 64 (2013) 20–33.
- [11] Glaser R. E (1980). *Bathtub and related failure rate characterizations*. *J. Amer. Statist. Assoc.*, 75, pp. 667-672.
- [12] Hassan S. Bakoucha, Bander M. Al-Zahrana , Ali A. Al-Shomrania , Vitor A.A. Marchib , Francisco Louzadac (2012). *An extended Lindley distribution*. *Journal of the Korean Statistical Society* 41, 75–85.
- [13] Ibrahim Elbatal, Gokarna Aryal (2017). A new generalization of the exponential Pareto distribution, *Journal of Information and Optimization Sciences*, 38 :5, 675-697.
- [14] Klugman S.A, Panjer H.H, Willmot G.E (1998). *Loss Models : From Data to Decisions*. Wiley, New York.
- [15] Lee C, Famoye F, Alzaatreh A (2013). Methods for generating families of univariate continuous distributions in the recent decades, *WIREs. Computational Statistics*. 5(3), 219–238.
- [16] Lémeray E.M. Racines de quelques équations transcendentes. Intégration d'une équation aux différences mêlées. Racines imaginaires, *Nouvelles Ann. Math.* 16 (1897) 540–546.

## Bibliographie

---

- [17] Lindley D.V (1958). *Fiducial distributions and Bayes theorem*, *Journal of the Royal Statistical Society*, Series B 20 102–107.
- [18] Mojtaba Alizadeh, Seyyed Fazel Bagheri, Mohammad Alizadeh & Sara-lees Nadarajah (2017) A new four-parameter lifetime distribution, *Journal of Applied Statistics*, 44 :5, 767-797.
- [19] Pearson K. (1893). *Contributions to the mathematical theory of evolution*. *Philos. Trans. R. Soc. A* 54, 329-333.
- [20] Pearson, K. (1895). *Contributions to the mathematical theory of evolution - ii. skew variation in homogeneous material*. *Philos. Trans. R. Soc. A* 186, 343-414.
- [21] Preda V, Ciumara R (2006). “On composite models : Weibull-Pareto and Lognormal-Pareto”. A comparative study. *Rom. J. Econ. Forecast.* 3 , 32–46.
- [22] Ramos P.L, Louzada F (2016). *The generalized weighted Lindley distribution : properties, estimation, and applications*. *Cogent Mathematics*,3 :1, DOI : 10.1080/23311835.2016.1256022.
- [23] Rashad M. EL-Sagheer, Mohamed A. W. Mahmoud & Samah H. M. Abdallah (2018). Statistical inferences for new Weibull-Pareto distribution under an adaptive type-ii progressive censored data, *Journal of Statistics and Management Systems*, 21 :6, 1021-1057.
- [24] Shanker R, Sharma S (2013). *A tow-parameter Lindley distribution for modeling waiting and survival times data*. *Applied Mathematics*, vol 4, pp. 363-368.
- [25] Stacy E. W. (1962). *A generalization of the gamma distribution*. *Ann.Math. Statist.* 33,3 (Sep), 1187-1192

- [26] Scollnik, D.P.M. (2007). On composite Lognormal-Pareto models. *Scandinavian Actuarial Journal*, 20–33.
- [27] Teodorescu, S., Vernic, R. (2006). A composite Exponential–Pareto distribution. *The Annals of the “ Ovidius” University of Constanta, Mathematics Series*, vol.XIV (1) : 99-108.
- [28] Teodorescu S, Vernic R (2009). Some composite Exponential-Pareto models for actuarial prediction. *Rom. J. Econ. Forecast.* 12, 82–100.
- [29] Teodorescu, S., Vernic (2013). On composite Pareto models”. *MATH. REPORTS* 15(65), 1, 11–29.
- [30] Zakerzadeh H, Dolati A (2009). *Generalized Linley Distribution*. *Journal of Mathematical extension*. Vol. 3, N° 2 , 13-25.
- [31] Zeghdoudi. H, Bouchahed.L (2018). A new and unified approach in generalizing the Lindley’s distribution with applications. *Statistics in Transition new series*. Vol. 19, No. 1, pp. 61–74, DOI 10.21307/stattrans-2018-004.
- [32] Zeghdoudi H, Lazri N, Yahia D (2018). Lindley Pareto Distribution. *Statistics in Transition New Series*. Volume 19, Issue 4, Pages 671-692, DOI : <https://doi.org/10.21307/stattrans-2018-035>.
- [33] Zeghdoudi H , Messaadia H (2018). Zeghdoudi distribution and its applications. *International Journal of Computing Science and Mathematics* 9(1) : 58 - 65.
- [34] Zeghdoudi H, Nedjar S (2016b). A pseudo Lindley distribution and its application’, *AfrStat.*, Vol. 11, No. 1, pp.923–932.
- [35] Zeghdoudi H, Nedjar S (2016). *On gamma Lindley distribution : properties and simulations*. *Journal of Computational and Applied Mathematics* 298,167–174.

## Bibliographie

---

- [36] Zeghdoudi H, Nedjar S (2016a). Gamma Lindley distribution and its application. *J. Appl. Prob.Stat.* 11(1) : 129-138.
- [37] Zeghdoudi H, Nedjar S (2017). On Pseudo Lindley Distribution : Properties and Applications. *New Trends in Mathematical Sciences. NTMSCI* 5, No. 1, 59-65 (2017).