

# وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR –ANNABA  
UNIVERSITY  
UNIVERSITE BADJI MOKHTAR  
ANNABA



جامعة باجي مختار  
- عنابة -

Faculté des Sciences

Année : 2023 /2024

Département de Mathématiques



## THÈSE

Présenté en vue de l'obtention du diplôme de Doctorat en mathématiques

**SUR LES DISTRIBUTIONS COMPOSEES ET LEUR  
UTILITE EN SCIENCE ACTUARIELLE ET AUTRES  
DOMAINES D'APPLICATIONS**

**Filière**

Mathématiques Appliquées

**Spécialité**

Actuariat

Par

**BENCHETTAH Moulouk Halima**

**DIRECTEUR DE THÈSE :** ZEGHDOUDI Halim Prof U.B.M. ANNABA  
**CO- DIRECTEUR DE THÈSE :** SENOUSSI Rachid Prof PACA, INRA, France

**Devant le jury**

**PRESIDENT :** BENMOSTEFA Fatima Zohra Prof U.B.M. ANNABA  
**EXAMINATEUR :** BELHAMRA Thara M.C.A U.B.M. ANNABA  
**EXAMINATEUR :** BEGHRICHE Abdelfateh M.C.A U. Frères Mentouri  
CONSTANTINE 1  
**EXAMINATEUR :** SEGHIER Fatma Zohra M.C.A ENSET. SKIKDA

# TABLE DES MATIÈRES

|                                                                                                                                |      |
|--------------------------------------------------------------------------------------------------------------------------------|------|
| <b>Abstract</b>                                                                                                                | iii  |
| <b>Résumé</b>                                                                                                                  | iv   |
| <b>Remerciements</b>                                                                                                           | vi   |
| <b>Abréviations et Notations</b>                                                                                               | viii |
| <b>Introduction générale</b>                                                                                                   | 1    |
| <b>1 Rappels de probabilité et d'actuariat</b>                                                                                 | 1    |
| 1.1 Lois de probabilités discrètes : binomiale, binomiale négative<br>et de Poisson . . . . .                                  | 2    |
| 1.2 Lois de probabilité continues : distributions de Pareto, Gamma,<br>XGamma, XLindley, NewXLindley, Zeghdoudi et de Nakagami | 3    |
| 1.3 Notions d'actuariat : loi de Poisson composée, probabilité de<br>ruine, prime et prime stop-loss . . . . .                 | 10   |
| <b>2 Modèles composées et théorie de la ruine</b>                                                                              | 13   |

Table des matières

---

|                                                                      |           |
|----------------------------------------------------------------------|-----------|
| 2.1 Définitions et notions sur les modèles composées . . . . .       | 13        |
| 2.1.1 Définitions et propriétés . . . . .                            | 13        |
| 2.2 Théorie de la ruine . . . . .                                    | 19        |
| 2.2.1 Définitions et propriétés . . . . .                            | 19        |
| <b>3 Approximation de la densité de probabilité en utilisant une</b> |           |
| <b>représentation polynomiale</b>                                    | <b>23</b> |
| 3.1 Rappel sur les polynômes orthogonaux . . . . .                   | 23        |
| 3.2 Familles exponentielles naturelles et leurs polynômes orthogo-   |           |
| naux . . . . .                                                       | 26        |
| 3.3 Représentation polynomiale de la densité dans la dimension un    | 27        |
| 3.3.1 Application aux distributions composées dans la di-            |           |
| mension un . . . . .                                                 | 30        |
| 3.4 Application numérique : distribution de Poisson composée . . .   | 33        |
| 3.4.1 Cas des sinistres de loi gamma $\Gamma(3, 1)$ . . . . .        | 34        |
| <b>4 Article de thèse</b>                                            | <b>36</b> |
| <b>Bibliographie</b>                                                 | <b>44</b> |

In this thesis, we introduce a new distribution called the "composite Nakagami-Pareto» and its applications by combining the distribution Nakagami, and Pareto. This new distribution will add some value to existing literature on the modeling of natural and actuarial sciences. In addition, we provide a synthesis of a numerical method of approximation of probability densities in the context of non-life insurance risk models using orthogonal polynomials. Also, a simulation and comparative study of several traditional and new distributions were presented.

---

**Keywords** : Composite Nakagami-Pareto distribution, Non-life insurance, Orthogonal polynomials.

Dans cette thèse, nous introduisons une nouvelle distribution appelée « composite Nakagami Pareto distribution » et ses applications en combinant les distributions Nakagami et Pareto. Grâce à cette nouvelle distribution, la littérature existante sur la modélisation des sciences naturelles et actuarielles sera enrichie. En outre, nous donnons une synthèse d'une méthode numérique d'approximation des densités de probabilités dans le contexte des modèles de risque en assurance non-vie en utilisant les polynômes orthogonaux. Par ailleurs, une simulation et une étude comparative entre plusieurs distributions classiques et nouvelles ont été présentées.

---

**Mots clés :** Composite Nakagami-Pareto distribution, Assurance non-vie, polynômes orthogonaux.

---

## ملخص

في هذه الأطروحة، نقدم توزيعًا جديدًا مركب يُسمى "ناكاغامي باريتو" وتطبيقاته من خلال الجمع بين توزيع ناكاغامي وباريتو. وسيضيف هذا التوزيع الجديد قيمة إلى الأعمال الموجودة حول نمذجة العلوم الطبيعية والإكتوارية. بالإضافة إلى ذلك، نقدم توليفًا بالطريقة العددية لتقريب الكثافات الاحتمالية في سياق نماذج مخاطر التأمين باستخدام كثيرات الحدود المتعامدة. كما تم تقديم محاكاة ودراسة مقارنة للعديد من التوزيعات التقليدية والجديدة.

---

## Remerciements

Avant tout, je tiens à exprimer ma profonde gratitude envers Allah le Tout-Puissant, qui m'a accordé la volonté, la patience, la force et la santé nécessaires à l'accomplissement de ce travail modeste. Je tiens également à exprimer ma profonde gratitude envers mes parents pour leur patience, leur générosité et leur merveilleuse présence. Ils ont constamment été une source de motivation et d'encouragement pour moi. Je tiens à exprimer ma gratitude particulière envers Monsieur ZEGHDOUDI Halim, Professeur au département de mathématiques de l'Université Badji-Mokhtar à Annaba, pour sa direction tout au long de mes années de thèse. Sa disponibilité constante, son écoute attentive à mes nombreuses questions et son intérêt constant pour l'avancement de mes travaux, ainsi que ses conseils visés, qui ont largement contribué au résultat final de ce travail. Je tiens également à exprimer ma gratitude envers mon père "mon héros", BENCHETTAH Azzedine, Professeur au département de mathématiques de l'Université Badji-Mokhtar à Annaba, pour son soutien, ses encouragements, son aide et ses sacrifices pour ma réussite. Je le remercie d'avoir cru en mes capacités, d'avoir été toujours à mes côtés et pour ses nombreuses relectures et corrections apportées à cette thèse. Aussi un spécial remerciement à ma mère BELABED Assia, ma confi-

## Remerciements

---

dente qui a été toujours là pour me soutenir et me donner du courage et de l'amour pour pouvoir poursuivre ce chemin, que dieu vous garde pour moi mes chers parents, vous êtes ma force. Je tiens à exprimer mes respectueux remerciements à Madame BENMOSTEFA Fatima Zohra, Professeur à l'Université Badji Mokhtar-Annaba, pour m'avoir fait l'honneur de présider le jury de soutenance de cette thèse. Mes vifs remerciements s'adressent également aux membres du jury, Madame BELHAMRA Thara, Maitre de conférence A à l'Université Badji Mokhtar-Annaba, Monsieur BEGHRICHE Abdelfateh, Maitre de conférence A à l'Université Frères Mentouri de Constantine et Madame SEGHIER Fatma Zohra, Maitre de conférence A à l'ENEST de Skikda. Je leur suis également reconnaissant pour l'intérêt qu'ils ont porté à mes recherches, en acceptant d'examiner mon travail et de l'enrichir avec leurs précieuses suggestions.

J'adresse mes plus sincères remerciements à tous mes proches, à ma famille, mes petits frères et ma soeur et à mes amis, qui m'ont constamment soutenu et encouragés tout au long de la réalisation de ce travail.

---

## Abréviations et Notations

|                     |                                                           |
|---------------------|-----------------------------------------------------------|
| $\mathbb{E}(\cdot)$ | Espérance mathématique                                    |
| $\mathbb{V}(\cdot)$ | Variance mathématique                                     |
| $\gamma_1$          | coefficient d'asymétrie théorique ( <i>skewness</i> )     |
| $\gamma_2$          | coefficient d'aplatissement théorique ( <i>kurtosis</i> ) |
| <i>f.d.r.</i>       | fonction de répartition                                   |
| <i>f.d.s.</i>       | fonction de survie                                        |
| <i>f.g.m.</i>       | fonction génératrice des moments                          |
| <i>f.g.p.</i>       | fonction génératrice des probabilités                     |
| <i>f.g.c.</i>       | fonction génératrice des cumulants                        |
| <i>VaR</i>          | Value at Risk                                             |
| <i>FEN</i>          | Famille Exponentielle Naturelle                           |
| <i>FENQ</i>         | Famille Exponentielle Naturelle Quadratique               |
| <i>EQI</i>          | Erreur quadratique intégrée                               |
| <i>CNPD</i>         | Composite Nakagami-Pareto Distribution                    |

**L**e travail a été réalisé dans le cadre d'une thèse LMD inscrit dans un projet de recherche PFRU sous le titre : LES MODELES COMPOSES EN STATISTIQUE : APPLICATIONS EN ÉPIDEMIOLOGIE ET EN ASSURANCE, code C00L03UN230120220003, utilisant les pertes d'assurance incendie Danoise et des données réelles du Marburg virus..

La section de recherche académique de ce travail présente deux orientations.

Dans un premier temps, il est nécessaire de créer une méthode approximative pour évaluer les distributions dans le cadre des modèles de risque en assurance actuarielle. En particulier, au cours d'une période d'exercice donnée, pour un ensemble de contrats d'assurance non-vie, le nombre de sinistres est une variable aléatoire discrète et le montant des sinistres est une variable aléatoire continue positive. Tous les services à verser par l'assureur pendant la période d'exercice examinée, connue sous le nom de charge totale du portefeuille, joue un rôle crucial dans la gestion des risques de l'entreprise, car elle évalue les engagements pris par l'entreprise d'assurance envers ses clients. L'actuaire doit examiner la répartition du cumul des prestations afin de calculer la prime et les marges de solvabilité, ainsi que l'évaluation

de son espérance, de sa variance, de sa fonction de survie et d'autres caractéristiques. La charge totale du portefeuille doit être répartie de manière composée pour d'améliorer son efficacité. Mais, étant donné que ces distributions de probabilité présentent des problèmes de calcul car leur densité de probabilité n'est accessible qu'à un nombre très restreint de cas, les calculs de probabilité peuvent être effectués à l'aide de méthodes de simulation de Monte-Carlo. De plus, comme les calculs sont complexes et nécessitent un temps considérable pour obtenir une précision adéquate, cela nous a amené à utiliser des méthodes numériques pour estimer la densité de probabilité.

La deuxième, elle consiste à introduire un nouveau modèle composé qui remédie aux inconvénients du modèle de Pareto en science actuarielle notamment l'assurance incendie. La distribution « Nakagami », qui trouve de nombreuses applications dans la réalité pour évaluer la durée de vie d'un objet en fonction de son âge, est fréquemment employée dans divers domaines de la physique : à savoir modéliser la hauteur des vagues, le rayonnement sonore et lumineux, les signaux radio et l'énergie éolienne, l'image par ultrasons, la durée de vie des tubes, des résistances, des réseaux, des cristaux, des boutons, des transformateurs, des relais et des condensateurs des radars d'aéronefs, la vitesse du vent sur un an sur les sites d'éoliennes, la vitesse moyenne quotidienne du vent, etc. Cette distribution a une grande importance dans les domaines de la fiabilité, de l'analyse de survie et de la science actualisée. Cependant, en actuariat, il est impossible d'utiliser une seule distribution pour modéliser les données, surtout si nous examinons les sinistres qui affectent un portefeuille d'assurance tels que l'assurance automobile, les données de paiement, etc. Dans cette optique, une combinaison de densités s'avère extrêmement bénéfique pour représenter cette répartition des accidents. Pour modéliser les données de perte importante ou de paiement

de réassurance, de nombreux chercheurs font recours au modèle de Pareto, comme le soulignent Cooray et M. Ananda (2005). En effet, les données de paiement des assurances sont souvent très asymétriques et distribuées avec de grandes queues. Toutefois, même si le modèle de Pareto aborde de manière satisfaisante le comportement des pertes importantes pour les données actuarielles, il présente un inconvénient en écartant les pertes mineures qui, en revanche, sont assez bien modélisées par les distributions de Length-biased exponential, Rayleigh, Log-normal, Gamma ou Weibull. Dans certaines situations, afin d'effectuer une étude plus efficace et performante, il est préférable d'éviter d'utiliser une seule distribution. Effectivement, si nous examinons les sinistres qui affectent un certain portefeuille d'assurance (assurance auto, données de paiement,...), il est préférable et même nécessaire de représenter la répartition des revendications comme une combinaison de densités.

Les idées de Cooray et Ananda (2005), Teodorescu et Vernic (2006), Preda et Ciumara (2006), Scollnik (2007) et Benatmane (2020) sont à l'origine de ce travail. Dans ce travail, nous avons introduit un modèle composé Nakagami-Pareto et nous avons donné ses principales propriétés statistiques. Comme généralement, lorsqu'il s'agit de distributions composées, la première est choisie comme distribution à queue légère et la deuxième comme distribution à queue épaisse. Pour notre modèle de Nakagami-Pareto étudié, nous avons pris comme  $h_1$ , la densité de la distribution de Nakagami qui est à queue légère, et  $h_2$  la densité de la distribution de Pareto qui est à queue épaisse. Quelques caractéristiques du modèle de Length-biased exponential Pareto qui peuvent être utilisées dans des ensembles de données du monde réel ont été étudiées.

La structure de ce projet de thèse est la suivante :

Dans le premier chapitre, nous rappelons les concepts généraux de l'actuariat et les lois de probabilités, dont nous fournissons les définitions ainsi

que leurs propriétés. Plus précisément, nous rappelons quelques résultats concernant les distributions de probabilités que nous utiliserons ultérieurement (Gamma, Pareto, XLindley, Size biased, Poisson composée, Length-biased exponential, probabilité de ruine, prime stop-loss ...).

Le deuxième chapitre se focalise sur les distributions composées et à leur utilisation en science actuarielle et en théorie de la ruine. On montre que le cumul du montant des prestations sur une période d'exercice spécifique pour un portefeuille de contrats peut être modélisé par la distribution composée en dimension 1. Ensuite, nous élargissons le concept de distribution composée à une dimension supérieure à 1 afin d'obtenir une modélisation qui intègre des risques auxquels plusieurs portefeuilles sont confrontés et auxquels les charges totales sont associées. Enfin, nous introduisons la théorie de la ruine, en présentant le modèle de ruine de Poisson composé et en rappelant la notion de la probabilité de ruine ultime.

Le troisième chapitre consiste à l'introduction d'une méthode numérique d'approximation et à rappeler certains résultats théoriques mathématiques utiles dans notre étude. Nous projettons la densité de probabilité sur un espace de Hilbert particulier. L'approximation est obtenue en tronquant à un certain ordre la représentation polynomiale de la densité de probabilité, nous obtenons l'approximation. Il est possible d'approcher la f.d.r. et la f.d.s. en intégrant l'approximation polynomiale de la densité de probabilité. La méthode d'approximation, en utilisant la technique d'approximation pour les distributions composées en dimension 1, ce qui nous amène à choisir une mesure de probabilité gamma liée aux polynômes de Laguerre généralisés. Aussi, nous utilisons ce résultat, condition d'intégrabilité, qui nous permet de valider la méthode de l'approximation polynomiale. En outre, nous utilisons des résultats mathématiques théoriques pour gérer la densité de probabi-

lité, afin de pouvoir sélectionner correctement les paramètres de la mesure de référence pour garantir l'efficacité de l'approximation polynomiale. Nous montrons que cette procédure a une motivation certaine pour l'obtention d'une approximation la plus efficace. En effet, nous avons pu montrer l'efficacité de l'approximation polynomiale par rapport à celles de Panjer et de X-Lindley.

Enfin, le dernier chapitre correspond à l'article intitulé : "Novel composite model : application in epidemic and actuarial science", paru dans le journal "Studies in engineering and exact sciences". Ici, nous avons introduit une nouvelle distribution composée et étudié ses propriétés théoriques statistiques. Nous avons effectué une étude comparative avec d'autres distributions classiques et composées en utilisant des données réelles en épidémie et en assurance et nous avons montré la puissance de notre distribution.

# CHAPITRE 1

## Rappels de probabilité et d'actuariat

### Sommaire

|                                                                                                                                       |    |
|---------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 Lois de probabilités discrètes : binomiale, binomiale négative et de Poisson . . . . .                                            | 2  |
| 1.2 Lois de probabilité continues : distributions de Pareto, Gamma, XGamma, XLindley, NewXLindley, Zeghdoudi et de Nakagami . . . . . | 3  |
| 1.3 Notions d'actuariat : loi de Poisson composée, probabilité de ruine, prime et prime stop-loss . . . . .                           | 10 |

Dans ce chapitre, nous rappelons quelques notions générales d'actuariat et des lois de probabilités que nous utilisons dans cette thèse. Ainsi, nous commençons par rappeler les définitions et les propriétés de certaines lois de probabilités à savoir : loi binomiale, loi de Poisson, loi de Pareto, loi binomiale négative, loi de Gamma, loi de XGamma, loi de XLindley, loi de New XLindley, loi de length biased exponential, loi de Nakagami et loi

de Zeghdoudi. Ensuite, nous donnons quelques concepts utiles en science actuarielle, telles que la loi de Poisson composée, la probabilité de ruine, la prime et la prime stop-loss.

## 1.1 Lois de probabilités discrètes : binomiale, binomiale négative et de Poisson

**Définition 1.1.1** Une variable aléatoire discrète  $X$  est dite de loi binomiale  $\mathcal{B}(n, p)$ ,  $0 \leq p \leq 1$ ,  $n \in \mathbb{N}^*$ , si sa densité  $f_X(k)$  est donnée par

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = \overline{1, n}.$$

Sa f.g.p. :  $\varphi_X(x) = (1 - p + px)^n$ ,

$$\mathbb{E}(X) = np,$$

$$\mathbb{V}(X) = np(1-p).$$

Si  $n = 1$ , elle correspond à la loi de Bernoulli.

**Définition 1.1.2** Une variable aléatoire discrète  $X$  est dite de loi binomiale négative  $\mathcal{BN}(n, p)$  si sa densité est donnée par

$$f_X(k) = \binom{k+n-1}{k} p^n (1-p)^k, k \in \mathbb{N}.$$

Sa f.g.p. :  $\varphi_X(x) = \left(\frac{p}{1-(1-p)x}\right)^n$ ,

$$\mathbb{E}(X) = n(1-p)/p,$$

$$\mathbb{V}(X) = n(1-p)/p^2.$$

Si  $n = 1$ , elle correspond à la loi géométrique.

**Définition 1.1.3** Une variable aléatoire discrète  $X$  est dite de loi de Poisson  $\mathcal{P}(\lambda)$ ,  $\lambda \geq 0$ , si sa densité est donnée par

$$f_X(k) = \frac{e^{-\lambda} \lambda^k}{k!}, k \in \mathbb{N}.$$

Sa f.g.p. :  $\varphi_X(x) = e^{\lambda(x-1)}$ ,

$\mathbb{E}(X) = \lambda$ ,

$\mathbb{V}(X) = \lambda$ .

## 1.2 Lois de probabilité continues : distributions de Pareto, Gamma, XGamma, XLindley, NewXLindley, Zeghdoudi et de Nakagami

La distribution de Pareto repose sur le principe de la règle des 20/80 (ou 80/20). En effet, Pareto en 1906, établit que 80% du revenu de plusieurs pays industrialisés est détenu par 20% des citoyens. Dans de nombreux domaines, ce principe est également vrai lorsque la plupart des résultats ou des impacts sont attribuables à une minorité des causes.

**Définition 1.2.1** Une variable aléatoire continue  $X$  suit une loi de Pareto  $P_a(\alpha, k)$ , de paramètres  $\alpha, k \in \mathbb{R}_+^*$ , si sa densité est donnée par

$$f_X(x) = \begin{cases} \frac{k}{\alpha} \left(\frac{\alpha}{x}\right)^{k+1} & \text{si } x > \alpha \\ 0 & \text{si } x \leq \alpha. \end{cases}$$

Rappelons quelques propriétés de la distribution de Pareto.

**Proposition 1.1** Si  $X$  suit la loi de Pareto  $P_a(\alpha, k)$  alors sa f.d.r.  $F_X(x)$  et sa f.d.s.  $S_X(x)$  sont données par

$$F_X(x) = \begin{cases} 1 - \left(\frac{\alpha}{x}\right)^k & \text{si } x > \alpha \\ 0 & \text{si } x \leq \alpha \end{cases} \quad \text{et} \quad S_X(x) = \begin{cases} \left(\frac{\alpha}{x}\right)^k & \text{si } x > \alpha \\ 0 & \text{si } x \leq \alpha. \end{cases}$$

**Proposition 1.2** Si  $X$  suit la loi de Pareto  $P_a(\alpha, k)$  alors  $\mathbb{E}(X)$  et  $\mathbb{V}(X)$  sont données par

$$\begin{aligned} \mathbb{E}(X) &= \frac{k\alpha}{\alpha - 1} \quad (\text{pour } \alpha > 1), \\ \mathbb{V}(X) &= \frac{k\alpha^2}{(\alpha - 1)^2(\alpha - 2)} \quad (\text{pour } \alpha > 2). \end{aligned}$$

**Proposition 1.3** Si  $X$  suit la loi de Pareto  $P_a(\alpha, k)$  alors  $\gamma_1$  (skewness) et  $\gamma_2$  (kurtosis) sont donnés par

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

Soit la fonction gamma  $\Gamma$  définie par

$$\Gamma(\lambda) = \int_0^{\infty} x^{\lambda-1} \exp(-x) dx, \quad \lambda > 0,$$

vérifiant

$$\Gamma(\lambda + 1) = \lambda\Gamma(\lambda), \quad \lambda > 0, \quad \Gamma(n + 1) = n!, \quad \forall n \in \mathbb{N}$$

**Définition 1.2.2** Une variable aléatoire continue  $X$  suit une loi Gamma  $Ga(\lambda, \beta)$  de paramètres  $\lambda$  et  $\beta$  (strictement positifs), si sa densité est donnée par

$$f_X(x) = \frac{x^{\lambda-1} \exp\left(-\frac{x}{\beta}\right)}{\beta^\lambda \Gamma(\lambda)}, \quad x > 0.$$

**Proposition 1.4** Si  $X \sim Ga(\lambda, \beta)$  alors sa f.d.r.  $F_X(x)$  et sa f.d.s.  $S_X(x)$  sont données par

$$F(x) = \frac{\gamma\left(\lambda, \frac{x}{\beta}\right)}{\Gamma(\lambda)} \quad , \quad S(x) = 1 - \frac{\gamma\left(\lambda, \frac{x}{\beta}\right)}{\Gamma(\lambda)}$$

**Proposition 1.5** Si  $X \sim Ga(\lambda, \beta)$  alors  $\mathbb{E}(X)$  et  $\mathbb{V}(X)$  sont données par

$$\begin{aligned}\mathbb{E}(X) &= \lambda\beta \ , \\ \mathbb{V}(X) &= \lambda\beta^2.\end{aligned}$$

**Proposition 1.6** Si  $X \sim Ga(\lambda, \beta)$  alors  $\gamma_1$  (skewness) et  $\gamma_2$  (kurtosis) sont données par

$$\gamma_1 = \frac{2}{\sqrt{\lambda}} \quad , \quad \gamma_2 = \frac{3(\lambda + 2)}{\lambda}.$$

Il est important de souligner que la loi Gamma constitue un modèle de probabilité efficace pour étudier la durée de vie des appareils qui sont soumis à une usure (véhicules, appareils ménagers, ordinateurs). Nous utilisons aussi cette loi en météorologie et en biostatistique.

**Définition 1.2.3** Une variable aléatoire continue  $Y$  suit une loi XGamma  $XGa(\theta)$ , si sa densité est donnée par

$$f_Y(y) = \frac{\theta^2}{(1 + \theta)} \left(1 + \frac{\theta}{2}y^2\right) e^{-\theta y}, \quad y > 0, \theta > 0.$$

**Proposition 1.7** Si  $Y \sim XGa(\xi)$  alors sa f.d.r.  $F_Y(y)$  et sa f.d.s.  $S_Y(y)$  sont données par

$$\begin{aligned}F(y) &= 1 - \frac{\left(1 + \xi + \xi y + \frac{\xi^2 y^2}{2}\right)}{(1 + \xi)} e^{-\xi y}, \quad y > 0, \xi > 0, \\ S(y) &= -\frac{\left(1 + \xi + \xi y + \frac{\xi^2 y^2}{2}\right)}{(1 + \xi)} e^{-\xi y}, \quad y > 0, \xi > 0.\end{aligned}$$

**Proposition 1.8** Si  $Y \sim Ga(\xi)$  alors  $\mathbb{E}(Y)$  et  $\mathbb{V}(Y)$  sont données par

$$\begin{aligned}\mathbb{E}(Y) &= \frac{\xi}{\xi(1+\xi)}, \\ \mathbb{V}(Y) &= \end{aligned}$$

**Définition 1.2.4** Une variable aléatoire continue  $X$  suit une loi Zeghdoudi  $ZD(x, \rho)$ , si sa densité est donnée par

$$f_{ZD}(x) = \begin{cases} \frac{\rho^3 x(1+x)e^{-\rho x}}{2+\rho}, & x > 0, \rho > 0, \\ 0, & \text{sinon.} \end{cases}$$

**Proposition 1.9** Si  $X \sim ZD(x, \rho)$  alors sa f.d.r.  $F_{ZD}(x)$  et sa f.d.s.  $S_{ZD}(x)$  sont données par

$$F_{ZD}(x) = 1 - \left( \frac{x^2 \rho^2 + \rho(\rho+2)x + \rho+2}{\rho+2} \right) e^{-x\rho}, \quad x > 0, \rho > 0.$$

$$S_{ZD}(x) = \left( \frac{x^2 \rho^2 + \rho(\rho+2)x + \rho+2}{\rho+2} \right) e^{-x\rho}, \quad x > 0, \rho > 0.$$

**Proposition 1.10** Si  $X \sim ZD(x, \rho)$  alors  $\mathbb{E}(X)$  et  $\mathbb{V}(X)$  sont données par

$$\begin{aligned}\mathbb{E}(X) &= \frac{2(\rho+3)}{\rho(\rho+2)}, \\ \mathbb{V}(X) &= \frac{2(\rho^2+6\rho+6)}{\rho^2(\rho+2)}.\end{aligned}$$

**Proposition 1.11** Si  $X \sim ZD(x, \rho)$  alors  $\gamma_1$  (skewness) et  $\gamma_2$  (kurtosis) sont donnés par

$$\begin{aligned}\gamma_1 &= \frac{24(\rho+2)^2(\rho+5)}{(2\rho^2+12\rho+12)^{\frac{3}{2}}}, \\ \gamma_2 &= \frac{30(\rho+2)^3(\rho+6)}{(\rho^2+6\rho+6)^2}.\end{aligned}$$

En s'inspirant de la distribution de Lindley, Chouia et Zeghdoudi (2021) ont introduit la distribution XLindley avec un paramètre  $\theta$  (réel positif) pour analyser et modéliser des données liés aux épidémiologies (Ebola virus). En effet, cette distribution est caractérisée par sa capacité à modéliser les données de survie avec des taux de hasard croissants. Notons que cette distribution est une combinaison de distributions Exponentielle et Lindley dont les proportions de mélange sont respectivement  $\frac{\theta}{(1+\theta)}$  et  $\frac{1}{(1+\theta)}$ .

**Définition 1.2.5** Une variable aléatoire continue  $X$  suit une loi de XLindley  $XL(\theta)$  de paramètres  $\theta \in \mathbb{R}_+^*$ , si sa densité est donnée par

$$f_{XL}(x) = \begin{cases} \frac{\theta^2(2+\theta+x)}{(1+\theta)^2} e^{-\theta x}, & x > 0, \theta > 0 \\ 0 & \text{sinon.} \end{cases}$$

**Proposition 1.12** Si  $X \sim XL(\theta)$  alors sa f.d.r.  $F_{XL}(x)$  et sa f.d.s.  $S_{XL}(x)$  sont données par

$$F_{XL}(x) = \begin{cases} 1 - \left(1 + \frac{\theta x}{(1+\theta)^2}\right) e^{-\theta x}, & x > 0, \theta > 0 \\ 0 & \text{sinon} \end{cases},$$

$$S_{XL}(x) = \begin{cases} \left(1 + \frac{\theta x}{(1+\theta)^2}\right) e^{-\theta x}, & x > 0, \theta > 0 \\ 0 & \text{sinon.} \end{cases}$$

**Proposition 1.13** Si  $X \sim XL(\theta)$  alors on a

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

$$\mathbb{V}(X) = \frac{(1+\theta)^4 + 4\theta^2 + 6\theta + 1}{(1+\theta)^4 \theta^2}.$$

$$Skewness = \sqrt{\beta_1} = \frac{\mathbb{E}(X^3)}{(\mathbb{V}(X))^{\frac{3}{2}}} = \frac{6(\theta^2 + 2\theta + 4)(1+\theta)^4}{[(1+\theta)^4 + 4\theta^2 + 6\theta + 1]^{\frac{3}{2}}}$$

$$Kurtosis = \beta_2 = \frac{\mathbb{E}(X^4)}{(\text{Var}(X))^2} = \frac{24(\theta^2 + 2\theta + 5)(1 + \theta)^6}{[(1 + \theta)^4 + 4\theta^2 + 6\theta + 1]^2}$$

$$C.V = \gamma = \frac{\sqrt{\text{Var}(X)}}{\mathbb{E}(X)} = \frac{\sqrt{(1 + \theta)^4 + 4\theta^2 + 6\theta + 1}}{(1 + \theta)^2 + 1}$$

Notons que ces coefficients sont croissants en  $\theta$ .

**Définition 1.2.6** Une variable aléatoire continue  $X$  suit une loi de NewX-Lindley  $NXL(x, \theta)$  de paramètre  $\theta \in \mathbb{R}_+^*$ , si sa densité est donnée par

$$f_{NXL}(x) = \frac{\theta}{2} (1 + \theta x) \exp(-\theta x), \quad x > 0, \theta > 0.$$

**Proposition 1.14** Si  $X \sim NXL(x, \theta)$  alors sa f.d.r.  $F_{NXL}(x)$  et sa f.d.s.  $S_{NXL}(x)$  sont données par

$$F_{NXL}(x) = 1 - \left(\frac{1}{2}\theta x + 1\right) e^{-x\theta},$$

$$S_{NXL}(x) = \left(\frac{1}{2}\theta x + 1\right) e^{-x\theta}.$$

**Proposition 1.15** Si  $X \sim NXL(x, \theta)$  alors on a

$$\begin{aligned} \mathbb{E}(X) &= \frac{3}{2\theta}, \\ \mathbb{V}(X) &= \frac{7}{4\theta^2}. \end{aligned}$$

$$Skewness = \sqrt{\beta_1} = \frac{\mathbb{E}(X^3)}{(\text{Var}(X))^{\frac{3}{2}}} = \frac{\frac{15}{\theta^3}}{\left(\frac{7}{4\theta^2}\right)^{\frac{3}{2}}}.$$

$$Kurtosis = \beta_2 = \frac{\mathbb{E}(X^4)}{(\text{Var}(X))^2} = \frac{\frac{72}{\theta^4}}{\left(\frac{7}{4\theta^2}\right)^2}.$$

$$C.V = \gamma = \frac{\sqrt{\text{Var}(X)}}{\mathbb{E}(X)} = \frac{\sqrt{\frac{7}{4\theta^2}}}{\frac{3}{2\theta}}.$$

**Définition 1.2.7** Une variable aléatoire continue  $X$  suit une loi Nakagami  $Naka(x; m, k)$  de paramètre  $m$  et  $k$  (strictement positifs), si sa densité est donnée par

$$f(x) = \begin{cases} \frac{2m^m}{\Gamma(m)k^m} x^{2m-1} \exp\left(-\frac{m}{k}x^2\right) & \text{pour } \begin{cases} m > \frac{1}{2} \\ x > 0 \\ k > 0 \end{cases} , \\ 0 & \text{sinon.} \end{cases}$$

**Proposition 1.16** Si  $X \sim Naka(x; m, k)$ , alors sa f.d.r.  $F_X(x)$  et sa f.d.s.  $S_X(x)$  sont données par

$$F_X(x) = \begin{cases} \frac{\gamma(m, \frac{m}{k}x^2)}{\Gamma(m)} & \text{pour } x > 0, \\ 0 & \text{sinon.} \end{cases}$$

$$S_X(x) = 1 - \frac{\gamma(m, \frac{m}{k}x^2)}{\Gamma(m)}$$

où  $\gamma(.,.)$  est la fonction Gamma incomplète.

**Proposition 1.17** Si  $X$  suit la loi de Nakagami  $Naka(x; m, k)$ , alors  $\mathbb{E}(X)$  et  $\mathbb{V}(X)$  sont données par

$$\begin{aligned} \mathbb{E}(X) &= \frac{\Gamma(m + \frac{1}{2})}{\Gamma(m)} \left(\frac{k}{m}\right)^{\frac{1}{2}}, \\ \mathbb{V}(X) &= k \left(1 - \frac{1}{m} \left(\frac{\Gamma(m + \frac{1}{2})}{\Gamma(m)}\right)^2\right). \end{aligned}$$

### 1.3 Notions d'actuariat : loi de Poisson composée, probabilité de ruine, prime et prime stop-loss

**Définition 1.3.1** Un processus aléatoire  $(X_t)_{t \geq 0}$  est dit suivre une loi de Poisson composée s'il s'écrit sous la forme suivante

$$X_t = \sum_{i=1}^{N_t} U_i,$$

où pour tout  $t$  fixé,  $N_t \sim \mathcal{P}(\lambda)$ , et  $(U_i)_{i \in \mathbb{N}}$  est une suite de variables aléatoires (i.i.d), i.e., indépendantes et identiquement distribuées, et indépendantes de  $N_t$ .

Par hypothèse, on pose  $X_t = 0$  si  $N_t = 0$ . Notons que, puisque la f.g.m. de  $N_t$  est donnée par

$$\mathbb{M}_{N_t}(s) = e^{\lambda t[\exp(s)-1]},$$

et puisque  $\mathbb{M}_{X_t}(s) = \mathbb{M}_{N_t} \{ \ln [\mathbb{M}_U(s)] \}$ , on en déduit la f.g.m. de  $X_t$  donnée par

$$\mathbb{M}_{X_t}(s) = e^{\lambda t[M_U(s)-1]}.$$

Par développement limité en  $s$ , on obtient

$$\mathbb{E}(X_t) = \lambda \mathbb{E}(U) \quad \text{et} \quad \mathbb{V}(X_t) = \mathbb{E}(U^2) \lambda t.$$

Introduisons maintenant le concept de la théorie de ruine.

Soit  $(R_t)_{t \geq 0}$  le processus de réserve, avec réserve initiale  $R_0 = u$ .

Soient

-  $(T_i)_{i \in \mathbb{N}^*}$  suite de variables aléatoire positives i.i.d. qui correspond aux instants d'arrivée des sinistres,

- $\sigma_n = \sum_{i=1}^n T_i$  instant d'occurrence du  $n^{\text{ième}}$  sinistre,
- $N_t = \max \{n \in \mathbb{N} ; \sigma_n \leq t\}$  processus de comptage donnant le nombre de sinistres jusqu'à l'instant  $t$ ,
- $(U_i)_{i \in \mathbb{N}^*}$  suite de variables aléatoire positives i.i.d. égales aux montants des sinistres,
- $c$ , flux de prime générée par le portefeuille par unité de temps.

Ainsi, le processus de réserve  $R_t$  s'écrit

$$R_t = u + ct - \sum_{i=1}^{N_t} U_i,$$

d'où l'on déduit le processus de surplus  $(S_t)_{t \geq 0}$  donné par

$$S_t = u - R_t.$$

Définissons maintenant les notions de probabilité de ruine.

La probabilité de ruine ultime ou probabilité de ruine à horizon de temps infini, notée  $\Psi$ , est définie par

$$\Psi(u) = \mathbb{P}(\inf_{t \geq 0} R_t < 0 : R_0 = u).$$

La probabilité de ruine avant l'instant  $T$  ou probabilité de ruine à horizon de temps fini est définie par

$$\Psi(u, T) = \mathbb{P}(\inf_{t \in [0, T]} R(t) < 0 : R_0 = u).$$

D'où, la probabilité complémentaire ou probabilité de non ruine, notée  $\Phi$ , est donnée par

$$\Phi(u) = 1 - \Psi(u), \quad \text{pour temps infini}$$

$$\Phi(u, T) = 1 - \Psi(u, T) \quad \text{pour temps fini.}$$

L'instant de ruine, noté  $\tau_u$ , correspondant à une réserve initiale  $u$ , est défini par

$$\tau_u = \inf \{t \geq 0 : R_t < 0\} = \inf \{t \geq 0 : S_t > u\}.$$

Les maxima du processus de surplus en temps fini et infini, respectivement notés  $M_T$  et  $M$ , sont définis par

$$M = \sup_{t \geq 0} S_t,$$

$$M_T = \sup_{t \in [0, T]} S_t;$$

ce qui nous permet de redéfinir les probabilités de ruine comme suit

$$\Psi(u) = \mathbb{P}(\tau_u < \infty) = \mathbb{P}(M > u),$$

$$\Psi(u, T) = \mathbb{P}(\tau_u < T) = \mathbb{P}(M_T > u).$$

**Définition 1.3.2** La prime est la somme réglée par la personne qui souscrit à un contrat d'assurance, en échange de la protection permise pour un certain nombre de garanties, i.e., elle est la somme totale qu'une personne verse annuellement pour bénéficier d'une assurance.

**Définition 1.3.3** La prime stop-loss est une option pour la gestion des contrats d'assurance vie qui limite les pertes.

Elle permet des arbitrages automatiques sur les investissements effectués. Le but de cette opération consiste à réduire les pertes d'investissements, ainsi, la prime stop-loss est utilisée lors d'un excédent de sinistres.

## CHAPITRE 2

### Modèles composées et théorie de la ruine

## 2.1 Définitions et notions sur les modèles composés

### 2.1.1 Définitions et propriétés

La principale question concernant la gestion des risques d'une société d'assurance est d'évaluer ses obligations envers les assurés. Ainsi, l'engagement d'une compagnie d'assurance non vie est égale au montant cumulé des prestations distribué aux assurés pour une période d'exercice.

Notons  $N$  la variable aléatoire de comptage définissant le nombre de sinistres subis par les assurés et  $U$  la variable aléatoire positive et continue modélisant le coût unitaire des sinistres. Ainsi, la charge totale du portefeuille est une variable aléatoire  $X$  de distribution composée qui, d'après le chapitre 1, suit une loi de probabilité composée  $(\mathbb{P}_N, \mathbb{P}_U)$  et s'écrit

$$X = \sum_{i=1}^N U_i, \quad (2.1)$$

où  $\mathbb{P}_N$  est la loi image de  $\mathbb{P}$  relativement à  $N$ , de densité  $f_N(k) = \mathbb{P}(N = k)$ ,  $k \in \mathbb{N}$  et  $(U_i)_{i \in \mathbb{N}}$  est une suite de variables aléatoires réelles, positives, i.i.d., de loi de probabilité  $\mathbb{P}_U$ , et de densité  $f_U$  par rapport à la mesure de Lebesgue. Aussi, on suppose que  $N$  est indépendante de la suite  $(U_i)_{i \in \mathbb{N}}$  et  $X = 0$  si  $N = 0$ .

Le modèle collectif en théorie de risque, connu sous le nom de portefeuille, est étudié dans [5], [18], [15] et [13].

Ainsi, on montre que la variable aléatoire  $X$  donnée par [2.1] à une distribution mixte

$$d\mathbb{P}_X(x) = f_N(0)\delta_0(x) + d\mathbb{Q}_X(x),$$

où  $\delta_0$  est la mesure de Dirac et  $\mathbb{Q}_X$  est une mesure de probabilité défailante, i.e.,  $\int d\mathbb{Q}_X(x) = 1 - f_N(0) \leq 1$ , de densité  $q$  donnée par

$$q(x) = \sum_{k=1}^{+\infty} f_N(k) f_U^{(*k)}(x), \quad (2.2)$$

où  $f_U^{(*k)}$  est la densité de la somme de  $k$  variables aléatoires i.i.d. de loi  $\mathbb{P}_U$ . Comme elle s'exprime comme convolutions successives de  $f_U$ , il est difficile de l'utiliser de par sa nature comme série. Toutefois, sa connaissance est nécessaire puisqu'elle nous permet d'évaluer la f.d.r. et la f.d.s. très utiles pour déterminer les primes stop-loss généralisées.

**Définition 2.1.1** La prime stop-loss généralisée est définie par

$$\Pi_{c,d}(X) = \mathbb{E} [(X - c)_+^d]$$

où  $c \in \mathbb{R}_+$  et  $d \in \mathbb{N}^*$ .

En science actuarielle, certaines configurations de la prime stop-loss généralisée sont compatibles avec des quantités standard. Si  $c = 0$  et  $d = 1$ , on a la prime pure

$$\Pi_{0,1}(X) = \mathbb{E}(X).$$

Si  $d = 1$ , on a la prime stop-loss usuelle du seuil de rétention

$$\Pi_{1,c}(X) = \mathbb{E}[(X - c)_+]. \quad (2.3)$$

**Proposition 2.1** *La forme générale de la prime stop-loss est donnée par*

$$\Pi_{c,d} = \int_0^\infty d \times y^{d-1} \Pi_{c+y,0}(X) dy. \quad (2.4)$$

Notons que la connaissance de la f.d.s. permet de calculer toutes les primes stop-loss.

Soit  $P$  le montant cumulé des primes que la compagnie d'assurance reçoit de la part des preneurs d'assurance sur une période donnée, et  $X$  le montant des prestations qu'elle doit déboursier aux assurés suite aux sinistres. Comme le principe de tarification consiste à évaluer le niveau de prime par rapport à de la prime pure, le cumul des primes est donné par

$$P = \Pi_{0,1}(X)(1 + \eta), \quad (2.5)$$

où  $\eta$  désigne un chargement de sécurité.

Ceci nous permet alors d'obtenir la probabilité de ruine donnée par

$$\begin{aligned} \Psi(u) &= \mathbb{P}(X > P + u) \\ &= 1 - \mathbb{P}(X \leq P + u), \end{aligned}$$

où  $P - X$  correspond au résultat annuel et  $u$  est le capital qui amortit un résultat annuel négatif.

La résolution de l'équation

$$\Psi(u) = \alpha,$$

où  $\alpha \in [0.01; 0.5]$  donne la valeur du risque prise par la société d'assurance qui est le montant du capital  $u$ .

Aussi,  $u$  peut être déterminé à partir de la f.d.s. comme suit

$$u_\alpha = VaR_X(1 - \alpha) - P,$$

où  $VaR_X(1 - \alpha)$  désigne la *Value\_at\_Risk* de niveau  $1 - \alpha$  de  $X$ .

Comme la densité de probabilité défaillante [2.2](#) ne peut être donnée en général sous forme explicite, ceci nous amène à l'approximer numériquement, approximation donnant une bonne approche de la f.d.r. et de la f.d.s.

Rappelons le résultat qui nous donne la f.g.m. de la variable aléatoire  $X$  définie par

$$\mathcal{L}_X = \mathbb{E}(e^{sX}),$$

**Proposition 2.2** *La f.g.m. de la variable aléatoire  $X$  est donnée par*

$$\mathcal{L}_X(s) = \varphi_N[\mathcal{L}_U(s)],$$

où  $\varphi_N(s) = \mathbb{E}(s^N)$  désigne la f.g.p. de la variable aléatoire  $N$ .

On a aussi

$$\mathbb{E}(X) = \mathbb{E}(N)\mathbb{E}(U),$$

et

$$\mathbb{V}(X) = \mathbb{E}(N)\mathbb{V}(U) + \mathbb{E}(U^2)\mathbb{V}(N).$$

Nous allons maintenant introduire l'algorithme de Panjer qui est utile à l'approximation des distributions composées.

### Algorithme de Panjer

#### Définition et mise en œuvre de l'algorithme de Panjer

---

Rappelons que l'algorithme de Panjer est une méthode d'approximation assez fine pour déterminer les probabilités d'une distribution composée  $(\mathbb{P}_N, \mathbb{P}_U)$ . Cependant, notons que cette approximation ne peut être utilisée que dans le cas où le coût de sinistres  $U$  est discret, et par conséquent  $X$  est de loi discrète. De plus, cet algorithme de Panjer n'est valable que lorsque  $\mathbb{P}_N$  appartient à une certaine famille dite de Panjer

$$f_X(k) = \mathbb{P}(X = k), k \in \mathbb{N}. \quad (2.6)$$

**Définition 2.1.2** [16]. On dit que  $\mathbb{P}_N$  appartient à la famille de Panjer si ses probabilités admettent la formule de récurrence suivante

$$f_N(k) = \left(a + \frac{b}{k}\right) f_N(k-1), k = 1, 2, \dots, \quad (2.7)$$

où  $a < 1$  et  $b \in \mathbb{R}$  avec  $b \leq k$ .

La caractérisation de cette famille est le résultat suivant

**Théorème 2.1.1** [21]. *Considérons  $N$  une variable aléatoire discrète avec des valeurs entières. Lorsque la relation [2.7] est vérifiée par la loi de probabilité de  $N$ , cela signifie que  $N$  admet une distribution binomiale, binomiale négative ou Poisson. En outre,,*

- (i) Si  $a = 0$ , donc  $b > 0$  et  $N \sim \mathcal{P}(b)$ .
- (ii) Si  $0 < a < 1$ , donc  $a + b > 0$ , et  $N \sim \mathcal{NB}(\alpha, p)$  ainsi que  $p = a$  et  $\alpha = 1 + bp^{-1}$
- (iii) Si  $a < 0$ , donc  $b = -a(n+1)$  pour  $n \in \mathbb{N}$ , et  $N \sim \mathcal{B}(n, p)$  ainsi que  $p = a(a-1)^{-1}$  et  $n = -1 - ba^{-1}$ .

**Théorème 2.1.2** [16]. *Considérons  $X$  une variable aléatoire de distribution composée  $(\mathbb{P}_N, \mathbb{P}_U)$ . Si  $U$  est une variable aléatoire discrète à valeurs entières*

et que  $N$  est une variable aléatoire de comptage dont les probabilités respectent l'équation de récurrence [2.7](#), alors

$$f_X(k) = \begin{cases} \varphi_N [f_U(0)], & \text{pour } k = 0, \\ \frac{1}{1-af_U(0)} \sum_{j=1}^k (a + \frac{bj}{k}) f_U(j) f_X(k-j), & \text{pour } k \geq 1. \end{cases} \quad (2.8)$$

### Remarques

- Une généralisation de la formule de récurrence [2.7](#) est donnée dans [20](#) comme suit

$$f_N(k) = \sum_{i=1}^k \left[ a(i) + \frac{b(i)}{k} \right] f_N(k-i), \quad (2.9)$$

- Comme cet algorithme n'est valable que dans le cas où le coût des sinistres est discret, plusieurs auteurs ont essayé de traiter le cas continue dont les calculs sont laborieux. Afin de surmonter ce problème, ils ont été obligés de procéder à une discrétisation des formules pour pouvoir utiliser l'algorithme de Panjer. Notons que cette procédure demande beaucoup de calculs, mais elle est nécessaire puisque elle nous permet d'obtenir certains résultats.

Rappelons un résultat permettant de déterminer les moments de la distribution composée.

**Proposition 2.3** [8](#). Soit  $X$  une variable aléatoire  $\sim (\mathbb{P}_N, \mathbb{P}_U)$ . Si la loi de probabilité satisfait la relation [2.7](#), alors

$$\mathbb{E}(X^{k+1}) = \frac{1}{1-a} \sum_{i=0}^k \binom{k}{i} \left( \frac{k+1}{i+1} a + b \right) \mathbb{E}(U^{i+1}) \mathbb{E}(X^{k-1}).$$

## 2.2 Théorie de la ruine

### 2.2.1 Définitions et propriétés

Le modèle de la ruine étant déjà défini dans le premier chapitre, nous allons maintenant donner quelques résultats concernant le modèle de ruine de Poisson composé. Dans ce cas, d'après la loi des grands nombres, les assurés coûtent en moyenne

$$\lim_{t \rightarrow +\infty} \frac{1}{t} \sum_{i=1}^{N_t} U_i = \lambda \mathbb{E}(U).$$

Le niveau de prime  $c$  est donné par

$$c = (1 + \eta) \lambda \mathbb{E}(U),$$

où  $\eta$ , exprimé en pourcentage, nous fournit le montant de la prime qui doit excéder le coût moyen des sinistres par unité de temps.

Notons que pour avoir une activité rentable, nous devons imposer la condition  $\eta > 0$ , appelée **Net Benefit Condition**, décrite par le résultat suivant.

**Proposition 2.4** *Si  $\eta > 0$ , alors*

$$\lim_{t \rightarrow +\infty} R_t = +\infty,$$

*$\psi(u) < 1$ , On appelle cette activité rentable.*

*Si  $\eta < 0$ , alors*

$$\lim_{t \rightarrow +\infty} R_t = -\infty,$$

*$\psi(u) = 1$ , On appelle cette activité non rentable.*

Poursuivons un résultat qui met en relation les distributions composées avec la probabilité de ruine ultime.

**Théorème 2.2.1** (Pollaczek et Khinchine). *La probabilité de ruine ultime dans le modèle de ruine de Poisson composé est égale à la f.d.s. d'une distribution géométrique composée.  $(\mathbb{P}_N, \mathbb{P}_V)$ . Soit  $X = \sup_{t \geq 0} S_t$ . Alors*

$$\psi(u) = \mathbb{P}(X > u) = (1-p) \sum_{i=1}^{+\infty} p^i \bar{F}_V^{(*i)}(u), \quad (2.10)$$

avec

$$X = \sum_{i=1}^N V_i, \quad (2.11)$$

où  $N$  est une variable aléatoire de comptage suivant la loi géométrique  $\mathcal{NB}(1, p)$  de paramètre  $p = \frac{\lambda \mathbb{E}(U)}{c}$ , et  $(V_i)_{i \in \mathbb{N}}$  est une suite de variables aléatoires, positives, (i.i.d.) d'une loi  $\mathbb{P}_V$  et indépendantes de  $N$ . Nous désignons la distribution de probabilité  $\mathbb{P}_V$  comme **integrated tail distribution** de la distribution  $\mathbb{P}_U$ . La densité de probabilité de  $\mathbb{P}_V$  est exprimée par

$$f_V(x) = \frac{\bar{F}_U(x)}{\mathbb{E}(U)}.$$

La transformée de Laplace de la probabilité de ruine correspond à la transformée de Laplace de la f.d.s. de la variable aléatoire de  $X$ , donnée par

$$\mathcal{L}_\Psi(s) = \frac{1}{s} (\mathcal{L}_X(s) - 1), \quad (2.12)$$

avec

$$\mathcal{L}_X(s) = \frac{1-p}{1-p\mathcal{L}_V(s)},$$

et

$$\mathcal{L}_V(s) = \frac{1}{s\mathbb{E}(U)} (\mathcal{L}_U(s) - 1).$$

Étant donné que la présence de la série infinie dans [2.10](#) ne permet pas de déterminer la probabilité de ruine ultime, nous allons nous restreindre à un cas spécifique où nous pourrions encadrer la probabilité de ruine à l'aide de fonctions exponentielles décroissantes.

**Théorème 2.2.2** [18]. Si  $X$  suit une distribution géométrique composée  $[\mathcal{NB}(1, p), \mathbb{P}_U]$  telle que l'équation

$$\mathcal{L}_U(s) = \frac{1}{p}$$

admette une et une seule solution positive, notée  $\gamma$ , alors

$$a_- e^{-\gamma x} < \bar{F}_X(x) < a_+ e^{-\gamma x},$$

avec

$$a_- = \inf_{x \in \text{Supp}U} \frac{e^{\gamma x} \bar{F}_U(x)}{\int_x^{+\infty} e^{\gamma x} \bar{F}_U(x) dx}, \quad a_+ = \sup_{x \in \text{Supp}U} \frac{e^{\gamma x} \bar{F}_U(x)}{\int_x^{+\infty} e^{\gamma x} \bar{F}_U(x) dx}.$$

Ainsi, dans notre cas, nous déduisons le résultat suivant.

**Corollaire 2.1** La probabilité de ruine pour le modèle poisson composé satisfait

$$a_- e^{-\gamma x} < \Psi(u) < a_+ e^{-\gamma x}, \quad (2.13)$$

où

$$a_- = \inf_{x \in \text{Supp}V} \frac{e^{\gamma x} \bar{F}_V(x)}{\int_x^{+\infty} e^{\gamma x} \bar{F}_U(x) dx}, \quad a_+ = \sup_{x \in \text{Supp}V} \frac{e^{\gamma x} \bar{F}_V(x)}{\int_x^{+\infty} e^{\gamma x} \bar{F}_U(x) dx}.$$

et  $\gamma$  est l'unique solution positive de l'équation

$$\lambda(\mathcal{L}_U(s) - 1) = cs, \quad (2.14)$$

appelée équation fondamentale de Cramer-Lundberg.

Maintenant, dans le cas général, nous utilisons certaines approximations.

- La première, de Cramer-Lundberg, repose sur l'encadrement [2.13]. Dans ce cas, la probabilité de ruine ultime est approximée par

$$\Psi(u) = \frac{c - \mathbb{E}(u)}{\lambda \mathcal{L}_U(\gamma) - c} e^{-\gamma u},$$

qui se déduit à partir de l'égalité suivante

$$\lim_{u \rightarrow +\infty} e^{\gamma u} \Psi(u) = \frac{c - \mathbb{E}(u)}{\lambda \mathcal{L}_U(\gamma) - c}.$$

Notons que cette approximation est efficace pour des grandes valeurs de réserves initiales.

- La deuxième, de Beekman-Bowers, consiste à approximer la densité de probabilité par des sommes de densités gamma, i.e., une variable aléatoire  $Z$  de f.d.r.

$$F_Z(u) = 1 - \frac{\Psi(u)}{p},$$

est approximée par une loi gamma  $\Gamma(\alpha, \beta)$ .

## CHAPITRE 3

### Approximation de la densité de probabilité en utilisant une représentation polynomiale

#### 3.1 Rappel sur les polynômes orthogonaux

Soit une fonction  $\rho \in L^1(]a, b[, \mathbb{R}_+)$ , continue par morceaux. On suppose que  $\int_a^b \rho(x)x^m dx < \infty$ , pour tous  $m \in \mathbb{N}$ . Etant donné  $P(x)$  et  $Q(x)$  deux polynômes réels quelconques, on définit l'application  $\langle \cdot, \cdot \rangle$  par

$$\langle P, Q \rangle = \int_a^b P(x) Q(x) \rho(x) dx.$$

Il est clair que cette application est un produit scalaire.

Ainsi, on dit que les polynômes  $P$  et  $Q$  sont orthogonaux si  $\langle P, Q \rangle = 0$ .

**Définition 3.1.1** Une suite  $\{P_n(x)\}$ , où les  $P_n(x)$  sont des polynômes de degré  $n$  est dite orthogonale par rapport au poids  $\rho(x)$  sur l'intervalle  $]a, b[$  si

$$\langle P_n, P_m \rangle = 0, \text{ si } n \neq m.$$

### Chapitre 3. Approximation de la densité de probabilité en utilisant une représentation polynomiale

---

De plus, elle est dite orthonormale si

$$\langle P_n, P_m \rangle = \delta_{mn},$$

où  $\delta_{mn}$  est le symbole de Kronecker.

Notons que, pour tout polynôme  $P(x)$  de degré  $m \leq n$ , on a

$$P(x) = \sum_{n=0}^m c_n P_n(x), \quad \text{avec} \quad c_n = \frac{\langle P, P_n \rangle}{\langle P_n, P_n \rangle}, \quad (3.1)$$

où  $\{P_n(x)\}$  est une suite de polyômes orthogonaux.

Aussi, on établit la formule de récurrence suivante : il existe  $A_n, B_n$ , et  $C_n$ , réels tels que, pour tout  $n \geq 1$ ,

$$xP_n(x) = A_n P_{n+1}(x) + B_n P_n(x) + C_n P_{n-1}(x), \quad (3.2)$$

avec  $A_n, C_n \neq 0$ .

En effet, comme  $xP_n$  est un polynôme de degré  $n + 1$ , on a

$$xP_n = \sum_{k=0}^{n+1} c_k P_k(x), \quad \text{avec} \quad c_k = \frac{\langle xP_n, P_k \rangle}{\langle P_k, P_k \rangle}, \quad (3.3)$$

or  $\langle xP_n, P_k \rangle = \langle xP_k, P_n \rangle = 0$  si  $k > n + 1$  et  $n > k + 1$ , i.e.,  $k < n - 1$  et  $k > n + 1$ .

Ainsi, d'après [3.2](#) et [3.3](#), les coefficients  $c_{n+1}, c_n$  et  $c_{n-1}$  s'identifient à  $A_n, B_n$ , et  $C_n$ .

Posons

$$P_n(x) = \sum_{k=0}^n a_k^{(n)} x^k.$$

Par identification, on a

$$A_n = a_n^{(n)} / a_{n+1}^{(n+1)}, \quad \text{et} \quad A_n a_n^{(n+1)} + B_n a_n^{(n)} = a_{n-1}^{(n)},$$

d'où

$$B_n = \frac{a_{n-1}^{(n)}}{a_n^{(n)}} - \frac{a_n^{(n+1)}}{a_{n+1}^{(n+1)}},$$

et comme

$$xP_{n-1}(x) = \frac{a_{n-1}^{(n-1)}}{a_n^{(n)}} P_n(x) + xP_n = \sum_{k=0}^{n+1} c_k P_k(x),$$

on trouve

$$C_n = \frac{\langle xP_n, P_{n-1} \rangle}{\langle P_{n-1}, P_{n-1} \rangle} = \frac{\langle P_n, xP_{n-1} \rangle}{\langle P_{n-1}, P_{n-1} \rangle} = \frac{a_{n-1}^{(n-1)}}{a_n^{(n)}} \frac{\langle P_n, P_n \rangle}{\langle P_{n-1}, P_{n-1} \rangle},$$

si de plus, la suite  $\{P_n(x)\}$  est orthonormale, on trouve  $C_n = A_{n-1} \neq 0$ .

**Définition 3.1.2** Soit  $\{\phi_n(x)\}$  une suite de fonctions. La fonction  $G(x, t)$  est dite fonction génératrice des fonctions  $\phi_n(x)$  si elle est développable en série entière de  $t$ , sous la forme suivante

$$G(x, t) = \sum_{n=0}^{\infty} \phi_n(x) t^n.$$

Dans le cas où les fonctions  $\phi_n(x)$  correspondent à des polynômes  $P_n(x)$  de degré  $n$ , on a

$$G(x, t) = \sum_{n=0}^{\infty} P_n(x) t^n. \quad (3.4)$$

**Théorème 3.1.1** La condition nécessaire et suffisante pour que les polynômes  $P_n(x)$  définis par le développement [3.4](#) soient orthogonaux sur un intervalle  $]a, b[$  relativement au poids  $\rho(x)$  est que l'intégrale

$$I = \int_a^b G(x, t) G(x, t') \rho(x) dx,$$

ne dépende que du produit  $tt'$ .

**Preuve.** Comme on a

$$I = \int_a^b \sum_{n,m=0}^{\infty} P_n(x) P_m(x) t^n t'^m \rho(x) dx = \sum_{n,m=0}^{\infty} t^n t'^m \langle P_n, P_m \rangle.$$

Si l'on suppose que la suite  $\{P_n(x)\}$  est orthogonale, alors on a bien

$$I = \sum_{n,m=0}^{\infty} (tt')^n \langle P_n, P_n \rangle = I(tt').$$

Réciproquement, si

$$I = \sum_{n,m=0}^{\infty} t^n t^m \langle P_n, P_m \rangle = \sum_{n,m=0}^{\infty} c_n (tt)^n,$$

alors on a bien  $c_n = \langle P_n, P_n \rangle$ . ■

## 3.2 Familles exponentielles naturelles et leurs polynômes orthogonaux

**Définition 3.2.1** Une famille  $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ ,  $\Theta \subset \mathbb{R}$ , est appelé (FEN) générée par une mesure de probabilité  $\nu$  sur  $\mathcal{B}(\mathbb{R})$ , si, pour toute variable aléatoire  $X$  de loi de probabilité  $\mathbb{P}_\theta$ , on a

$$\begin{aligned} \mathbb{P}_\theta(X \in A) &= \int_A \exp[x\theta - \psi(\theta)] d\nu(x) \\ &= \int_A f(x, \theta) d\nu(x), \quad A \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

### Remarques

- Notons que  $f(x, \theta)$  correspond à la dérivée de Radon-Nikodym  $\frac{d\mathbb{P}_\theta}{d\nu}$  et l'application  $s \rightarrow \psi(s) = \log[\mathcal{L}_X(s)]$  est dite Fonction Génératrice des Cumulants (f.g.c.) et on montre que l'espérance et la variance de  $X$  sont données respectivement par

$$\begin{aligned} \mathbb{E}(X) &= \psi'(\theta) \triangleq m, \\ \mathbb{V}(X) &= \psi''(\theta) \triangleq \mathbb{V}(m). \end{aligned}$$

- L'application  $\theta \rightarrow \psi'(\theta)$  étant bijective, admet une inverse  $m \rightarrow h(m)$  définie sur  $\mathcal{M} = \psi'(\Theta)$ , et donc on peut reparamétriser la famille  $\mathcal{P}$  comme suit

$\mathcal{P} = \{\mathbb{P}_m, m \in \mathcal{M}\}$ , où chaque  $\mathbb{P}_m$  est de moyenne  $m$  et  $\frac{d\mathbb{P}_m}{d\nu} = \exp [h(m) - \psi [h(m)]] \stackrel{\Delta}{=} f(x, m)$ . voir [2].

Dans le cas où  $\mathbb{V}(m) = a_2 m^2 + a_1 m + a_0$ ,  $a_2, a_1, a_0 \in \mathbb{R}$ , la FEN est notée FENQ (quadratique) et comprend les distributions normales, gammas, hyperboliques, Poisson, binomiales et négatives binomiales.

- Soient les polynômes

$$Q_\alpha(x, m) = \mathbb{V}^\alpha(m) \left[ \frac{\partial^\alpha}{\partial m^\alpha} f(x, m) \right] / f(x, m), \quad \forall \alpha \in \mathbb{N},$$

de degré  $\alpha$  en  $m$  et en  $x$ .

Si une FENQ est générée par  $\nu$ , alors la suite  $\{Q_\alpha\}_{\alpha \in \mathbb{N}}$  vérifie

$$\langle Q_\alpha, Q_\beta \rangle = \int Q_\alpha(x, m) Q_\beta(x, m) f(x, m) d\nu(x) = \delta_{\alpha\beta} \|Q_\alpha(\cdot, m)\|^2, \quad \alpha, \beta \in \mathbb{N}.$$

Dans le cas où  $\nu$  est de moyenne  $m_0$ , on obtient  $f(x, m_0) = 1$ , et les polynômes orthogonaux liés à  $\nu$  sont donnés par

$$Q_\alpha(x, m_0) = \mathbb{V}^\alpha(m) \left[ \frac{\partial^\alpha}{\partial m^\alpha} f(x, m) \right]_{m=m_0}, \quad \forall \alpha \in \mathbb{N}. \quad (3.5)$$

Dans toute la suite, on va utiliser comme polynômes orthonormaux relativement à  $\nu$ , les  $Q_\alpha(x)$ , donnés par

$$Q_\alpha(x) = \frac{Q_\alpha(x, m_0)}{\|Q_\alpha(\cdot, m_0)\|^2}, \quad (3.6)$$

obtenus par normalisation. Pour plus de détails voir [15].

### 3.3 Représentation polynomiale de la densité dans la dimension un

Soit  $\nu$  une mesure de probabilité génératrice d'une FENQ. On considère l'espace  $L^2(\nu)$  muni de la norme

$$\|f\|^2 = \int f^2(x) d\nu(x).$$

### Chapitre 3. Approximation de la densité de probabilité en utilisant une représentation polynomiale

---

Il est clair que l'ensemble des polynômes est dense dans  $L^2(\nu)$ , ainsi,  $\{Q_\alpha\}_{\alpha \in \mathbb{R}}$  définie par [3.6](#) constitue un système orthonormal complet de  $L^2(\nu)$ . Le résultat suivant nous permet d'obtenir la formule d'approximation de la densité de probabilité.

**Théorème 3.3.1** *Soit une FENQ générée par une mesure de probabilité  $\nu$  sur  $\mathcal{B}(\mathbb{R})$  et  $\{Q_\alpha\}_{\alpha \in \mathbb{R}}$  la suite de polynômes orthonormaux par rapport à  $\nu$ . Soit  $X$  une variable aléatoire telle que  $\mathbb{P}_X \prec\prec \nu$ . Si  $\frac{d\mathbb{P}_X}{d\nu} \in L^2(\nu)$  alors*

$$\frac{d\mathbb{P}_X}{d\nu}(x) = \sum_{\alpha=0}^{+\infty} a_\alpha Q_\alpha(x), \quad x \in \mathbb{R}, \quad (3.7)$$

où

$$a_\alpha = \mathbb{E}[Q_\alpha(X)], \quad \forall \alpha \in \mathbb{N}, \quad (3.8)$$

et on a l'identité de Parseval

$$\left\| \frac{d\mathbb{P}_X}{d\nu} \right\|^2 = \sum_{\alpha=0}^{+\infty} a_\alpha^2. \quad (3.9)$$

#### Remarques

- Les coefficients dans la formule [3.8](#) sont les moments de la distribution de  $X$ .

- Comme  $\nu \prec\prec \lambda$  et  $\mathbb{P}_X \prec\prec \lambda$ , en notant  $f_\nu = \frac{d\nu}{d\lambda}$  et  $f_X = \frac{d\mathbb{P}_X}{d\lambda}$ ,  $\lambda$  étant la mesure de Lebesgue, d'après le Théorème précédent, on a

$$\begin{aligned} f_X(x) &= \sum_{\alpha=0}^{+\infty} a_\alpha Q_\alpha(x) f_\nu(x) \\ &= f_{X,\nu}(x) f_\nu(x), \quad x \in \mathbb{R}, \end{aligned}$$

avec  $f_{X,\nu}(x) = \frac{d\mathbb{P}_X}{d\nu}(x)$ , et

$$\lim_{N \rightarrow +\infty} \int [f_{X,\nu}(x) - f_{X,\nu}^N(x)]^2 d\nu(x) = 0, \quad (3.10)$$

où

$$f_{X,\nu}^N(x) = \sum_{\alpha=0}^N a_\alpha Q_\alpha(x). \quad (3.11)$$

D'après l'identité de Parseval, on a

$$\lim_{\alpha \rightarrow +\infty} a_\alpha = 0,$$

et donc l'approximation [3.11](#) nous donne la vitesse de convergence de la suite  $\{a_\alpha\}_{\alpha \in \mathbb{N}}$ .

- La définition de l'Erreur Quadratique Intégrée (EQI) de l'approximation de  $f_{X,\nu}$  par  $f_{X,\nu}^N$  est donnée par

$$L(f_{X,\nu}, f_{X,\nu}^N) = \int (f_{X,\nu}(x) - f_{X,\nu}^N(x))^2 d\nu(x),$$

ou bien

$$L(f_{X,\nu}, f_{X,\nu}^N) = \sum_{\alpha=N+1}^{+\infty} a_\alpha^2.$$

Comme l'approximation d'ordre  $N$  de la densité de  $X$  est donnée par

$$f_X^N(x) = f_{X,\nu}^N(x) f_\nu(x), \quad x \in \mathbb{R}, \quad (3.12)$$

et puisque  $Q_0(x) = 1$ ,  $x \in \mathbb{R}$  et  $a_0 = 1$ , alors on trouve

$$\begin{aligned} \int f_X^N(x) dx &= \sum_{\alpha=0}^{+\infty} a_\alpha \int Q_0(x) \times Q_\alpha(x) d\nu(x) \\ &= \sum_{\alpha=0}^{+\infty} a_\alpha \delta_{0\alpha} = 1. \end{aligned}$$

Comme il est difficile de garantir la positivité de l'approximation [3.12](#) pour un ordre de troncature  $N$  pas assez grand, nous surmontons cette difficulté en introduisant la modification suivante

$$\tilde{f}_X^N(x) = \frac{\max[f_X^N(x), 0]}{\lambda_N}, \quad x \in \mathbb{R}, \quad (3.13)$$

$\lambda_N$  étant une constante de normalisation définie par

$$\lambda_N = \int \max[f_X^N(x), 0] d\nu(x).$$

### 3.3.1 Application aux distributions composées dans la dimension un

Soit  $X$  une variable aléatoire définie comme suit

$$X = \sum_{i=1}^S U_i, \quad (3.14)$$

où  $S$  est une variable aléatoire de comptage de loi de probabilité  $\mathbb{S}$ ,  $\{U_i\}_{i \in \mathbb{N}}$  est une suite de variables aléatoires réelles, continues, positives et i.i.d. de loi de probabilité  $\mathbb{P}_U$ . Ainsi, la variable aléatoire  $X$  a une distribution composée  $(\mathbb{P}_S, \mathbb{P}_U)$  et sa loi de probabilité est donnée comme suit

$$d\mathbb{P}_X(x) = f_S(0) \delta_0(x) + d\mathbb{G}_X(x).$$

#### Choix de la mesure de référence

D'après le Théorème 3.3.1, il est nécessaire de choisir une mesure  $\nu$  qui appartient aux FENQ, de telle sorte que la mesure de probabilité  $\mathbb{P}_X$  soit absolument continue par rapport à  $\nu$ , condition irréalisable à cause de la masse de probabilité en 0 de la loi de  $X$ .

La mesure de probabilité défailante  $\mathbb{G}_X \prec\prec \lambda$ , nous sert pour déterminer la loi de probabilité de  $X$ . Ainsi, nous utilisons la méthode d'approximation polynomiale pour retrouver l'expression de la densité défailante  $g_X$  suivante

$$g_X(x) = \sum_{\alpha=1}^{+\infty} f_S(\alpha) f_U^{(*\alpha)}(x).$$

Comme l'application du Théorème 3.3.1 nécessite que  $Supp(\mathbb{G}_X) \subset Supp(\nu)$ , et puisque on a  $Supp(\mathbb{G}_X) = \mathbb{R}_+$ , il est donc raisonnable de choisir une loi gamma  $\Gamma(r, m)$  comme mesure de référence  $\nu$  de densité

$$f_\nu(x) = \frac{e^{-x/m} x^{r-1}}{m^r \Gamma(r)}, \quad x \in \mathbb{R}_+, \quad (3.15)$$

et dont les polynômes orthogonaux associés sont les polynômes généralisés de Laguerre donnés par

$$Q_\alpha(x) = (-1)^n \binom{n+r-1}{n}^{-\frac{1}{2}} L_n^{r-1}(x/m), \quad (3.16)$$

où

$$L_n^{r-1}(x) = \sum_{k=0}^n \binom{n+r-1}{n} \frac{(-x)^k}{k!}.$$

Maintenant, si  $\frac{d\mathbb{G}_X}{d\nu} = g_{X,\nu} \in L^2(\nu)$  alors, d'après le Théorème 3.3.1, on trouve que la densité  $g_X$  est donnée par

$$\begin{aligned} g_X(x) &= \sum_{\alpha=0}^{+\infty} a_\alpha Q_\alpha(x) f_\nu(x) \\ &= g_{X,\nu}(x) f_\nu(x), \end{aligned} \quad (3.17)$$

et par troncature à l'ordre  $N \geq 0$  de la série infinie dans [3.17](#), on obtient l'approximation polynomiale suivante

$$\begin{aligned} g_X^N(x) &= \sum_{\alpha=0}^N a_\alpha Q_\alpha(x) f_\nu(x) \\ &= g_{X,\nu}^N(x) f_\nu(x), \end{aligned} \quad (3.18)$$

Ainsi, l'intégration de [3.18](#) va nous permettre d'obtenir la f.d.r., la f.d.s. et la prime stop loss généralisée.

### Vérification de la condition d'intégrabilité

Comme la validation de l'approximation repose sur la condition d'intégrabilité suivante

$$\int_0^{+\infty} \left( \frac{d\mathbb{G}_X}{d\nu}(x) \right)^2 d\nu(x) < +\infty \Leftrightarrow \int_0^{+\infty} g_X(x)^2 e^{-\frac{x}{m}} x^{1-r} dx < +\infty. \quad (3.19)$$

Nous allons donc énoncer le résultat qui nous donne une majoration de  $g_X$  qui nous permet de choisir convenablement les paramètres de  $\nu$  et assurer la validité des approximations en remplaçant  $f_X(x)$  par  $g_X(x)$ .

**Proposition 3.1** *Soit  $X$  une variable aléatoire réelle, positive et continue, de loi de probabilité  $\mathbb{P}_X$  et de densité  $f_X(x)$  continuellement dérivable. Soit  $\gamma_X = \inf \{s > 0 : \mathcal{L}_X(s) = +\infty\}$ . S'il existe  $a \in \mathbb{R}^+$  tel que  $f_X(\cdot)$  soit strictement décroissante pour  $x > a$ , alors on a*

$$f_X(x) < A(s_0) e^{-s_0 x}, \quad \forall x > a,$$

où  $s_0 \in [0, \gamma_X[$ , avec  $A(s_0)$  indépendante de  $x$ .

Nous donnons également un résultat permettant de choisir les paramètres  $m$  et  $r$  de la mesure de référence [3.15](#) afin de garantir la validité des représentations polynomiales.

**Corollaire 3.1** *Si la densité défaillante  $g_X(x)$  vérifie les hypothèses de la Proposition 3.1, alors la paramétrisation suivante*

$$r \leq 1 ; m \in \left] \frac{1}{2\gamma}, +\infty \right[ ,$$

*assure la condition d'intégrabilité [3.19](#).*

### Calcul des coefficients de la représentation polynomiale et vitesse de convergence de l'approximation

Sachant qu'une approximation est adéquate si l'erreur est assez petite et l'ordre de troncature est petit, critères déterminés par la vitesse de la convergence vers 0 de la suite  $\{a_\alpha\}_{\alpha \in \mathbb{N}}$ . Nous avons le résultat correspondant au choix où gamma est la mesure de référence associée aux polynômes de Laguerre généralisés.

**Proposition 3.2** *Considère  $\nu$  la mesure de probabilité d'une loi gamma  $\Gamma(m, r)$  liée à la suite de polynômes orthonormaux  $\{Q_\alpha\}_{\alpha \in \mathbb{N}}$ . Supposons que  $X$  une variable aléatoire continue, réelle et positive avec une loi  $\mathbb{P}_X$  et une densité  $f_{X,\nu}(x)$  par rapport à  $\nu$ . Si  $f_{X,\nu}(x)$  est une fonction continue et admet ses deux premières dérivées, toutes  $\in L^2(\nu)$ , alors*

$$a_\alpha \underset{\alpha \rightarrow +\infty}{=} o\left(\frac{1}{\alpha}\right).$$

### 3.4 Application numérique : distribution de Poisson composée

Dans cette section, nous allons montrer que l'approximation utilisant les polynômes orthogonaux est performante.

Considérons que  $X$  une variable aléatoire de loi de probabilité composée  $(\mathbb{P}_S, \mathbb{P}_U)$ . Pour évaluer la densité défaillante  $g_X(x)$ , la f.d.s.,  $\tilde{F}_X$  et la prime stop-loss généralisée  $\Pi_{c,d}(X)$ , nous employons la méthode d'approximation. Il est possible de mesurer la précision de l'approximation en utilisant un critère d'erreur relative par rapport à la valeur exacte, ou à défaut en utilisant une approximation benchmark. Supposons que  $\tilde{f}(x)$  soit une estimation de la fonction  $f(x)$ , on définit la valeur de l'erreur relative de  $\tilde{f}(x)$  par

$$\Delta \tilde{f}(x) = \frac{\tilde{f}(x) - f(x)}{f(x)}.$$

L'approximation de la densité défaillante de  $X$  est donnée par

$$g_X^N(x) = \sum_{\alpha=0}^N a_\alpha Q_\alpha(x) f_\nu(x), \quad x \geq 0, \quad (3.20)$$

où  $N$  est l'ordre de troncature,  $a_\alpha = \mathbb{E}[Q_\alpha(X)]$ , pour  $\alpha = 1, \dots, N$  et  $\{Q_\alpha\}_{\alpha \in \mathbb{N}}$  est la suite orthonormale de polynômes définie dans [3.17](#).

Comme l'approximation de la f.d.s. est donnée par

$$\tilde{F}_X^N(x) = \int_x^{+\infty} g_X^N(y) dy, \quad x \geq 0, \quad (3.21)$$

alors, l'approximation de la prime stop-loss généralisée devient

$$\Pi_{c,d}(X) = \int_0^{+\infty} y^{d-1} \tilde{F}_X^N(y+c) dy \quad c \geq 0, \quad d \geq 0.$$

Ici, nous considérons uniquement le cas  $d = 1$ , d'après les résultats précédents, les paramètres  $m$  et  $r$  sont déterminés comme suit

$$m = \frac{\mathbb{V}(X)}{\mathbb{E}(X)}; \quad r = \frac{\mathbb{E}(X)^2}{\mathbb{V}(X)}. \quad (3.22)$$

Si de plus, on choisit  $r = 1$ , on obtient

$$m = \mathbb{E}(X) ; \quad r = 1.$$

Dans le cas où la variable aléatoire de comptage  $S$  est de loi de Poisson  $\mathcal{P}(\lambda)$ , on montre que la f.g.m. de  $X \sim (\mathcal{P}(\lambda), \mathbb{P}_U)$  est donnée par

$$\mathcal{L}_X(s) = \exp[\lambda(\mathcal{L}_U(s) - 1)]. \quad (3.23)$$

### 3.4.1 Cas des sinistres de loi gamma $\Gamma(3, 1)$

Lorsque la suite  $\{U_i\}_{i \in \mathbb{N}}$  suit une distribution  $\Gamma(3, 1)$ , dans ce cas, l'approximation de la densité défailante [3.20](#) devient

$$g_X(x) = \sum_{l=0}^{+\infty} \frac{e^{-\lambda}}{\lambda^l l!} \times \frac{e^{-x} x^{2l}}{\Gamma(3l)},$$

et la f.g.m. [3.23](#) s'écrit

$$\mathcal{L}_X(s) = \exp \left[ \lambda \left( \frac{1}{1-s} \right)^3 - 1 \right],$$

d'où  $\gamma_X = 1$ .

Chapitre 3. Approximation de la densité de probabilité en utilisant une représentation polynomiale

---

| $x$ | Exact       | Polynomial  | Panjer      | XLindley |
|-----|-------------|-------------|-------------|----------|
| 3   | 0.685132    | 0.685132    | 0.688771    | 0.68594  |
| 6   | 0.4313      | 0.4313      | 0.433313    | 0.4196   |
| 9   | 0.238763    | 0.238763    | 0.24049     | 0.24106  |
| 12  | 0.118895    | 0.118895    | 0.120076    | 0.1330   |
| 15  | 0.0542376   | 0.0542376   | 0.0549268   | 0.0713   |
| 18  | 0.0229767   | 0.0229767   | 0.0233334   | 0.0375   |
| 21  | 0.00913388  | 0.00913388  | 0.00930159  | 0.0194   |
| 24  | 0.00343513  | 0.00343513  | 0.003508    | 0.0099   |
| 27  | 0.00123021  | 0.00123021  | 0.00125982  | 0.0050   |
| 30  | 0.000421752 | 0.000421752 | 0.000433105 | 0.0025   |

TAB. 3.1 – Etude comparative entre plusieurs méthodes

La table 3.1 et figure 3.1 donnent les valeurs de la f.d.s. de  $X \sim (\mathcal{P}(\lambda), \Gamma(3, 1))$  par différentes méthodes pour  $m = r = 1$

**Table 3.1** : Etude comparative entre plusieurs méthodes

On voit que l'approximation par la méthode polynomiale donne les meilleurs résultats.

## CHAPITRE 4

Article de thèse

Ce chapitre correspond à l'article publié intitulé : « Novel composite model : application in epidemic and actuarial science ».



## Novel composite model: application in epidemic and actuarial science

## Novo modelo composto: aplicação em ciência epidemiológica e actuarial

DOI: 10.54021/seesv5n1-095

Recebimento dos originais: 12/04/2024  
Aceitação para publicação: 03/05/2024

### Moulouk Halima Benchettah

PhD Student in Mathematics  
Institution: Badji Mokhtar-Annaba University  
Address: P. O. Box 12, Annaba, Algeria  
E-mail : benmlk067@gmail.com

### Halim Zeghdoudi

PhD in Mathematics  
Institution: Badji Mokhtar-Annaba University  
Address: P. O. Box 12, Annaba, Algeria  
E-mail: halim.zeghdoudi@univ-annaba.dz

### Raman Vinoth

PhD in Mathematics  
Institution: University of Imam Abdulrahman Bin Faisal  
Address: P. O. Box 1982, Dammam, Saudi Arabia  
E-mail : vrrangan@iau.edu.sa

### ABSTRACT

Actuarial sciences frequently use Nakagami and Pareto distributions to model their payment data. The Nakagami distribution is frequently employed for modeling the lifespans of things that are affected by their age, and it possesses a multitude of practical applications. Alternatively, academics sometimes utilize the Pareto distribution to model payments data, especially for instances of substantial loss data or reinsurance payments. The present study presents and investigates a new composite model known as the composite Nakagami-Pareto distribution (CNPD), which integrates components from both the Nakagami and Pareto distributions. When examining composite distributions, the first distribution is usually characterized by a thin tail, whereas the next distribution has a thick tail. In the Nakagami-Pareto model, the Nakagami density is chosen as  $h_1$  due to its characteristic of being a light-tailed distribution. Similarly, the Pareto distribution is selected as  $h_2$  because it exhibits a heavy-tailed distribution. An investigation has been conducted to assess the practical usefulness of the composite Nakagami-Pareto model when applied to real-world data sets. This analysis has emphasized the significant components of the model. Statistical qualities such as cumulative distribution function, quantile function, mode, first moment, ad-hoc procedure and



maximum likelihood estimation have been established. Furthermore, a method for estimating has been described utilizing a data sample derived from the Composite Nakagami-Pareto model. The composite exponential-Pareto and the composite lognormal-Pareto distributions' resultant densities have similar shapes, but their tails are more noticeable. Therefore, we anticipate that our model will be more suitable than the composite exponential-Pareto distribution, composite lognormal-Pareto distribution, and other traditional one (or two)-parameter distributions. The significance and practicality of this novel approach were demonstrated by analyzing simulated cases, data sets are available for the recovery periods (measured in weeks) of 75 persons from Angola who were infected with the Marburg virus, as well as for 2156 fire insurance losses in Denmark.

**Keywords:** composite distribution, Pareto distribution, Nakagami distribution, maximum likelihood estimation, Marburg virus, fire insurance losses.

## RESUMO

As ciências atuariais frequentemente usam distribuições de Nakagami e Pareto para modelar seus dados de pagamento. A distribuição Nakagami é frequentemente empregada para modelar a vida útil de coisas que são afetadas pela sua idade, e possui uma infinidade de aplicações práticas. Alternativamente, os acadêmicos às vezes utilizam a distribuição de Pareto para modelar dados de pagamentos, especialmente para exemplos de dados de perdas substanciais ou pagamentos de resseguros. O presente estudo apresenta e investiga um novo modelo composto conhecido como a distribuição Nakagami-Pareto (CNPD), que integra componentes das distribuições de Nakagami e Pareto. Ao examinar distribuições compostas, a primeira distribuição é geralmente caracterizada por uma cauda fina, enquanto que a próxima distribuição tem uma espessa cauda. No modelo de Nakagami-Pareto, a densidade de Nakagami é escolhida como  $h_1$  devido à sua característica de ser uma distribuição de cauda leve. Da mesma forma, a distribuição de Pareto é selecionada como  $h_2$  porque exibe uma distribuição pesada. Uma investigação foi conduzida para avaliar a utilidade prática do modelo composto Nakagami-Pareto quando aplicado a conjuntos de dados do mundo real. Esta análise enfatizou os componentes significativos do modelo. Foram estabelecidas qualidades estatísticas como função de distribuição cumulativa, função quântica, modo, primeiro momento, procedimento ad hoc e estimativa de probabilidade máxima. Além disso, foi descrito um método de estimativa utilizando uma amostra de dados derivada do modelo Nakagami-Pareto composto. As densidades resultantes das distribuições composta exponencial-Pareto e composta lognormal-Pareto têm formas semelhantes, mas suas caudas são mais perceptíveis. Portanto, antecipamos que o nosso modelo será mais adequado do que a distribuição composta exponencial-Pareto, distribuição lognormal-pareto composta, e outras distribuições tradicionais de um (ou dois) parâmetros. A importância e a praticidade desta nova abordagem foram demonstradas através da análise de casos simulados, conjuntos de dados disponíveis para os períodos de recuperação (medidos em semanas) de 75 pessoas de Angola que foram infectadas com o vírus Marburg, bem como para 2156 perdas de seguro de incêndio na Dinamarca.

**Palavras-chave:** distribuição composta, distribuição de Pareto, distribuição de Nakagami, estimativa de probabilidade máxima, vírus de Marburg, perdas de seguro de incêndio.



## 1 INTRODUCTION

Contemporary numerical techniques find practical use in various sectors such as medical, economics, biological engineering, physical sciences, and statistics, among others. Statistics is essential in practical applications since the analysis heavily relies on the assumed probability model or distribution. Generally, there exist numerous statistical problems that do not adhere to classical or traditional probability models.

Currently, we are encountering significant outbreaks of viruses, including Marburg viruses, which have garnered much interest in the past two years. As statisticians, it is crucial for us to play a significant role in comprehending and formulating statistical models for Marburg infections. These models should be capable of accurately representing and analyzing both continuous and discrete random variables associated with Marburg infections. In this paper, we exerted much effort to select a suitable model that demonstrates higher accuracy in fitting the Marburg infections.

Other hand, the actuarial science often employs Nakagami and Pareto distributions to represent their payment data. The Nakagami distribution is commonly used to model the lifetimes of objects that are influenced by their age, and it has numerous practical uses. The Nakagami distribution is frequently employed in various branches of physics to represent phenomena such as wave heights, sound and light radiation, radio signals, and wind power and ultrasound images. It is also used to model the lifetimes of components such as tubes, resistors, networks, crystals, knobs, transformers, relays, and capacitors in aircraft radar sets. Additionally, it can be used to analyze wind speeds over a year at wind turbine sites and the daily average wind speed. The Nakagami distribution has received significant attention in the fields of dependability, survival analysis, and actuarial science. However, when it comes to modeling data in actuarial science, relying on a single distribution is insufficient, especially when studying the claims that impact a specific insurance portfolio, such as vehicle insurance or payment data. The utilization of two densities in combination is highly beneficial for estimating the distribution of these claims. Cooray and Ananda (2005) demonstrate that the generalized Pareto distribution is commonly employed by academics to represent payments data, particularly for significant loss data or



reinsurance payments. This is due to the fact that insurance payment data typically exhibit a high degree of skewness and are distributed with significant outliers at the upper end. An inherent drawback of employing the Pareto model for actuarial data is its limited ability to accurately represent tiny losses, as it primarily focuses on effectively capturing large losses. Conversely, the Nakagami, Lognormal, Gamma, and Weibull distributions offer more suitable modeling options for smaller losses. In certain modeling problems in actuarial science, it is not sufficient to rely on a single distribution. For instance, we can scrutinize the declarations associated with a specific insurance portfolio, like car insurance and payment details. To do this, the distribution of claims can be represented as a composite of two densities. To access a comprehensive presentation on the topic, please consult the study conducted by Elbatal and Aryal (2017).

The previous study primarily examines the Kumaraswamy Exponential Pareto distribution and conducts an examination of its characteristics, simulation, and practical applications. The study by EL-Sagheer et al. (2018) introduces a new Weibull-Pareto distribution for censored type II progressive adaptive data.

The current study is grounded on the concepts proposed by Cooray and Ananda (2005), Teodorescu and Vernic (2006), Preda and Ciumara (2006), and Scollnik (2007).

The current work introduces and examines a composite model called the Nakagami-Pareto model, which combines elements of the Nakagami and Pareto distributions. Typically, when analyzing composite distributions, the initial distribution is characterized by a light tail, while the subsequent distribution has a heavy tail. In the Nakagami-Pareto model, Nakagami density has been taken as  $h_1$  because it is light-tailed distribution and Pareto distribution has been taken as  $h_2$  because it is heavy-tailed distribution. The practical applicability of the composite Nakagami-Pareto model to real world datasets has been examined, highlighting its relevant aspects.

The current paper has been structured in the following manner: Section 2 focuses on the development of the composite Nakagami-Pareto model, its characteristics, and the generation of random values. Section 3 focuses on parameter estimate, whereas section 4 presents numerical examples. Section 5 presents an application study that utilizes two data sets.



## 2 ON COMPOSITE NAKAGAMI PARETO DISTRIBUTION (CNPD)

Consider the random variable  $X$ , which has a probability density function

$$h(x; \lambda) = \begin{cases} ch_1(x; \lambda) & 0 < x \leq \lambda \\ ch_2(x; \lambda) & \lambda < x \leq \infty \end{cases} \quad (1)$$

Where:

$h_1$  is a Nakagami density,  
 $h_2$  a two-parameter Pareto density, and  $c$  the normalizing constant. Hence,

$$h_1(x; m; k) = \frac{2m^m x^{2m-1} \exp\left(-\frac{mx^2}{k}\right)}{\Gamma(m)k^m}, \quad m > 0.5, k > 0, x > 0 \quad (2)$$

To simplify the problem, we put  $m = 1$ ,

$$h_1(x; k) = \frac{2x}{k} \exp\left(-\frac{x^2}{k}\right) \quad (3)$$

$$h_2(x; \alpha; \lambda) = \frac{\alpha \lambda^\alpha}{x^{\alpha+1}}, \quad x > \lambda \quad (4)$$

To achieve a composite smooth density function, we enforce continuity and differentiability constraints at the threshold point  $\lambda$ , i.e.

$$\begin{cases} h_1(\lambda) = h_2(\lambda) \\ \frac{d}{dx} h_1(\lambda) = \frac{d}{dx} h_2(\lambda) \end{cases} \quad (5)$$

These two restrictions give

$$\begin{cases} \frac{2\lambda^2}{k} \exp\left(-\frac{\lambda^2}{k}\right) = \alpha \\ \frac{(2\lambda^4 - k\lambda^2)}{k^2} \exp\left(-\frac{\lambda^2}{k}\right) = \alpha(\alpha + 1) \end{cases} \quad (6)$$

and, after some calculation, we obtain



$$\begin{cases} 2\gamma \exp(-\gamma) = \alpha \\ 2\gamma^2 - \gamma \exp(-\gamma) = \alpha(\alpha + 1) \end{cases} \text{with } \gamma = \frac{\lambda^2}{k} \quad (7)$$

Solving the second equation by numerical methods, it results the solution

$$\begin{cases} \gamma = 0.90395 \\ \alpha = 0.73214 \end{cases}$$

In order to find the normalizing constant, we impose the condition  $\int_0^\infty h(x; \theta) dx = 1$ , which gives

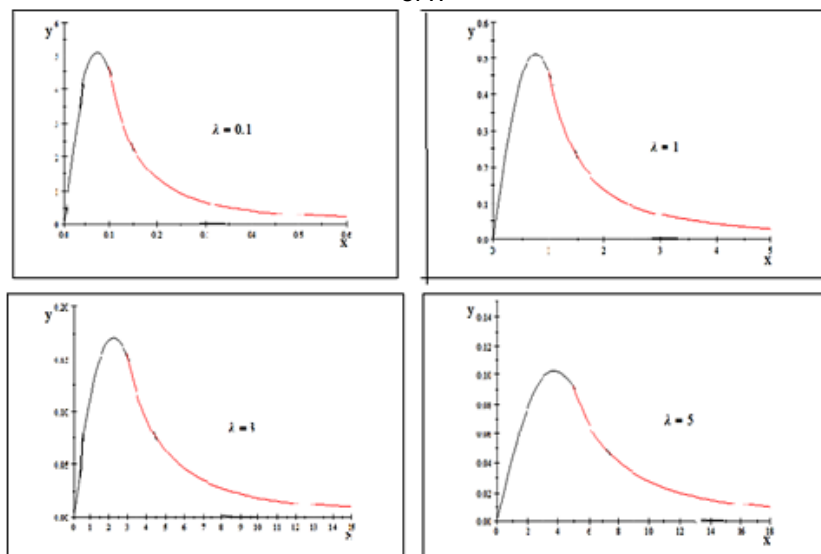
$$c = \frac{1}{2 - \exp\left(-\frac{\lambda^2}{k}\right)} = 0.62695$$

Thus,  $h(x; \lambda)$  can be expressed as

$$h(x; \lambda) = \begin{cases} \frac{(1.335)x}{\lambda^2} \exp\left(-\frac{0.90395x^2}{\lambda^2}\right) & 0 < x \leq \lambda \\ \frac{(0.45902) \lambda^{0.73214}}{x^{1.73214}} & \lambda \leq x < \infty \end{cases} \quad (8)$$

The various shapes that the CNPD PDF may take are seen in Fig 1.

Figure 1: The plots of PDF of composite Nakagami-Pareto distribution for some parameter value of  $\lambda$



Source : The authors.



**Proposition 1:** The cumulative distribution function (c.d.f.) of composite Nakagami distribution (CNPDP) is

$$H(x; \lambda) = \begin{cases} 0.62697(1 - \exp\left(0.90395 \frac{x^2}{\lambda^2}\right)) & 0 < x \leq \lambda \\ 0.62697(1 - \left(\frac{\lambda}{x}\right)^{0.73214}) & \lambda \leq x < \infty \end{cases} \quad (9)$$

**Proposition 2:** The quantile function of the CNPD is

$$H^{-1}(u; \lambda) = \begin{cases} \sqrt{-1.1063\lambda^2 \log(1 - 1.595u)} & \text{if } 0 < u < u_0 \\ \frac{\lambda}{(1-1.595u)^{1.3659}} & \text{if } u_0 < u < 1 \end{cases} \quad (10)$$

**Proposition 3:** The unique mode of CNPD is

$$x_{mod} = (0.74373)\lambda, \text{ where } 0 < x \leq \lambda$$

**Proposition 4:** The initial 1st moment of the CNPD is given by

$$E(x) = (0.22606)\lambda$$

The proofs of proposition 1-4 are omitted due to their simplicity.

### 3 GENERATING RANDOM VALUES OF THE CNPD

#### 3.1 PARAMETER ESTIMATION

In this section, we will present two algorithms for the estimation of the unknown parameter  $\lambda$ . Let  $x_1, x_2, \dots, x_n$  be a random sample from the CNPD.

Assuming that  $x_1 \leq x_2 \leq \dots \leq x_n$  and we assume that the unknown parameter  $\lambda$  is in between the  $m^{th}$  observation and  $(m + 1)^{th}$  observation, i.e.  $x_m \leq \lambda \leq x_{m+1}$ .

##### 3.1.1 A First algorithm

An ad-hoc procedure based on percentiles.



The following ad-hoc procedure provides a closed form for the parameter  $\lambda$ , estimated using percentiles. Based on percentiles, the parameter  $\lambda$  can be estimated, as the  $p^{th}$  percentile, where  $p = H(\lambda)$ .

$$p = 0.62697(1 - \exp(-0.90395)) = 0.37307$$

From Klugman et al. (1998), we have a smooth empirical estimate of the  $p^{th}$  percentile given by

$$\hat{\lambda} = (1 - h)x_m + hx_{m+1} \tag{11}$$

With

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

Here  $[\alpha]$  indicates the greatest integer smaller or equal with  $\alpha$ . The Pareto distribution or the Nakagami distribution will be a superior model than the composite Pareto- Nakagami distribution according as  $\hat{\lambda}$  is closer to  $x_1$  or  $x_n$ .

In Teodorescu and Vernic [3], the following algorithm based on the maximum likelihood (ML) method was suggested.

### 3.2 MAXIMUM LIKELIHOOD ESTIMATION (MLE)

The likelihood function is

$$L(x_1; \dots; x_n; \lambda) = \prod_{i=1}^n h(x_i) = \prod_{i=1}^m h(x_i) \prod_{i=m+1}^n h(x_i)$$

$$\prod_{i=1}^m (1.1335) \frac{x_i}{\lambda^2} \exp\left(- (0.90395) \frac{x_i^2}{\lambda^2}\right) \prod_{i=m+1}^n \frac{(0.45902) \lambda^{0.73214}}{x_i^{1.73214}}$$

$$= (1.1335)^m ((0.45902)^{n-m} \frac{\prod_{i=1}^m x_i}{\prod_{i=m+1}^n x_i^{1.73214}} \lambda^{0.73214(n-m) - 2m} \exp\left(\sum_{i=1}^m - (0.90395) \frac{x_i^2}{\lambda^2}\right))$$



$$= k\lambda^{0.73214n-2.73214m} \exp(-(0.90395)\lambda^{-2} \sum_{i=1}^m x_i^2) \quad (12)$$

With

$$k = (0.45902)^{n-m} (1.1335)^m \frac{\prod_{i=1}^m x_i}{\prod_{i=m+1}^n x_i^{1.73214}}$$

Step 1. Differentiating  $\ln(L)$  with respect to  $\lambda$  gives

$$\frac{d \ln(L)}{d\lambda} = \frac{0.73214n-2.73214m}{\lambda} + \frac{1.8079 \sum_{i=1}^m x_i^2}{\lambda^3} \quad (13)$$

Hence the solution of the likelihood equation  $\frac{d \ln(L)}{d\lambda} = 0$  is

$$\hat{\lambda}^2 = \frac{1.8079 \sum_{i=1}^m x_i^2}{0.73214n-2.73214m} \quad (14)$$

$$\hat{\lambda} = \sqrt{\frac{1.8079 \sum_{i=1}^m x_i^2}{0.73214n-2.73214m}}, \quad m > 0.26797n \quad (15)$$

If  $x_m \leq \hat{\lambda} \leq x_{m+1}$  then the ML estimator is

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

If this is not satisfied, then we have to go for next value of  $m$ .

#### 4 SIMULATION EXAMPLE

In this section, the estimation procedure described in section 3, has been explained using data sample generated from the Composite Nakagami-Pareto model. The generating algorithm used is based on the inversion of the c.d.f. (2).

The data set consisting of 100 values was sampled from a Nakagami-Pareto population with parameter  $\lambda = 5$  (see Table 1)



Table 1: Generated values of Nakagami-Pareto model for  $\lambda = 5$

|        |        |        |        |        |        |        |        |        |        |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.062  | 0.082  | 0.083  | 0.119  | 0.120  | 0.231  | 0.372  | 0.396  | 0.554  | 0.792  |
| 0.965  | 1.033  | 1.043  | 1.488  | 1.673  | 1.701  | 1.953  | 2.036  | 2.176  | 2.325  |
| 2.42   | 2.525  | 2.621  | 2.927  | 3.0495 | 3.3245 | 3.4343 | 3.5696 | 3.6972 | 3.841  |
| 4.118  | 4.224  | 4.886  | 4.889  | 4.934  | 4.967  | 4.984  | 5.068  | 5.225  | 5.603  |
| 5.902  | 5.986  | 6.494  | 6.750  | 7.311  | 7.65   | 7.98   | 8.116  | 8.293  | 8.666  |
| 8.712  | 8.822  | 9.273  | 9.446  | 9.715  | 9.948  | 9.995  | 10.440 | 10.543 | 11.080 |
| 12.52  | 12.833 | 13.705 | 14.018 | 16.975 | 17.790 | 17.845 | 18.315 | 19.836 | 22.665 |
| 25.35  | 26.153 | 36.624 | 39.59  | 40.844 | 49.691 | 50.788 | 55.562 | 63.398 | 71.889 |
| 83.51  | 93.84  | 179.06 | 197.63 | 202.50 | 242.33 | 374.76 | 363.58 | 439.32 | 503.91 |
| 607.06 | 807.75 | 986.46 | 1409.1 | 1647.3 | 1783.8 | 3157.5 | 6380.6 | 8319.4 | 8439.3 |

Source: The authors.

The estimated values of the parameter are:

- by algorithm 1: For  $m = 37, \hat{\lambda}_1 = 5.0415$ ;
- by algorithm 2 MLE Step 1:  $\hat{\lambda}_2 = 4.9901$ ;
- by algorithm 2 MLE Step 2:  $\hat{\lambda}_3 = 4.9884$ ;

We notice that this time, algorithm 2 step 1 gives a more accurate value.

We also applied the  $\chi^2$  test to check the distribution fitting, and the results for  $\hat{\lambda}_3$  are given in Table 2.

Table 2: Grouped data of Table 1 and  $\chi^2$  test

| Classes  | Frequencies, $n_i$ | Theoretical freq., $p_i$ | $\frac{n(f_i - p_i)}{p_i}$ (CNP) |
|----------|--------------------|--------------------------|----------------------------------|
| [0,2)    | 11                 | 0.108                    | $5.13(10^{-3})$                  |
| [2,4)    | 23                 | 0.216                    | $8.64(10^{-2})$                  |
| [4,6)    | 22                 | 0.271                    | 0.95                             |
| [6,10)   | 19                 | 0.154                    | 0.83                             |
| [10,60)  | 23                 | 0.256                    | 0.27                             |
| [60,300) | 2                  | 0.069                    | 3.49                             |
| $\Sigma$ | 100                | $\chi^2$ distance:       | 5.65                             |

(Columns 2 and 3 results from the data sample, while column 4 is calculated using the Nakagami-Pareto model distribution function)

Source: The authors

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

$$d^2(\hat{\lambda}_1) = 4.56$$

$$d^2(\hat{\lambda}_2) = 3.87$$

$$d^2(\hat{\lambda}_3) = 4.08$$





Table 4: Estimated values of fitted models and  $-2LL$ ;  $AIC$ ,  $AICc$  and  $BIC$  for data set I

| Model        | Density                                                                                             | $\theta$ | $-2LL$   | AIC     | AICc     | BIC     |
|--------------|-----------------------------------------------------------------------------------------------------|----------|----------|---------|----------|---------|
| Exponential  | $\theta \exp(-\theta x)$                                                                            | 0.199    | 386.995  | 388.995 | 389.049  | 391.312 |
| Lindley      | $\frac{\theta^2}{\theta + 1} (x + 1) \exp(-\theta x)$                                               | 0.346    | 365.986  | 367.98  | 368.041  | 370.304 |
| XLindley     | $\frac{\theta^2(\theta + 2 + x) \exp(-\theta x)}{(\theta + 1)^2}$                                   | 0.313    | 372.8971 | 374.897 | 374.9518 | 377.214 |
| New-XLindley | $\frac{\theta}{2} (\theta x + 1) \exp(-\theta x)$                                                   | 0.311    | 375.825  | 377.825 | 378.384  | 381.646 |
| Xgamma       | $\frac{\theta^2}{\theta + 1} \left( x^2 \frac{\theta}{2} + 1 \right) \exp(-\theta x)$               | 0.483    | 368.3286 | 370.328 | 370.832  | 373.015 |
| Zeghdoudi    | $\frac{\theta^3 x(1 + x) \exp(-\theta x)}{\theta + 2}$                                              | 0.553    | 349.3284 | 351.328 | 351.836  | 353.646 |
| C Nakagami-P | $\frac{(1.1335)x}{\theta^2} \exp\left(-\frac{(0.90395)x^2}{\theta^2}\right)$<br>$0 < x \leq \theta$ | 0.386    | 334.544  | 336.544 | 336.599  | 338.862 |
|              | $\frac{(0.45902)\theta^{0.73214}}{x^{1.73214}}$<br>$\theta \leq x < \infty$                         |          |          |         |          |         |

Source: The autores.

**For Second Data Set.** We provide in Table 5 Estimated values of fitted models and the values of the  $-2LL$ ,  $AIC$ ,  $AICc$  and  $BIC$  evaluated at the maximum likelihood estimators.

Table 5: Estimated values of fitted models and  $-2LL$ ;  $AIC$ ,  $AICc$  and  $BIC$  for data set II.

| Distributions   | Parameters                                         | $-2LL$ | AIC    | AICc    | BIC     |
|-----------------|----------------------------------------------------|--------|--------|---------|---------|
| Lognormal       | $\hat{\mu} = 0.6718$ ; $\hat{\sigma} = 0.7323$     | 8.86   | 8.872  | 8.877   | 24.212  |
| Pareto          | $\hat{\theta} = 0.3134$ ; $\hat{\alpha} = 0.5460$  | 11.35  | 11.354 | 11.359  | 26.702  |
| Gamma           | $\hat{\lambda} = 0.4107$ ; $\hat{\alpha} = 1.2578$ | 10.48  | 10.490 | 10.495  | 25.832  |
| Weibull         | $\hat{\theta} = 2.9531$ ; $\hat{t} = 0.9476$       | 10.54  | 10.544 | 10.549  | 25.892  |
| c Nakagami-P    | $\hat{\theta} = 2.532$                             | 7.04   | 7.667  | 7.668   | 14.716  |
| c Exponential-P | $\hat{\theta} = 2.477$                             | 13.17  | 13.565 | 13.5668 | 20.8460 |
| c Lognormal-P   | $\hat{\theta} = 1.3851$ ; $\hat{\alpha} = 1.4363$  | 7.756  | 7.982  | 7.9875  | 23.1080 |

Source: The authors.

**Remark 1.** Parameters are estimated by maximum likelihood (ML).

**Remark 2.** According to Tables 3 and 4, the values of 2LL , AIC, AICc and BIC of the two data sets show that the composite Nakagami-Pareto model provides a better fit than classical distributions and composite Exponential-Pareto, composite Lognormal-Pareto.



## 6 CONCLUSION

We have proposed a new distribution a composite Nakagami-Pareto generated by Pareto distribution. Its pdf, cdf, mode and the mean have been given. The estimation of parameter using both the ad-hoc procedure and the method of maximum likelihood has been discussed. Applications of the proposed model have been discussed with two examples-the first with samples data generated from the composite Nakagami-Pareto distribution and the second with the individuals infected with the Marburg virus and the fire insurance losses data set. The goodness of fit of the composite Nakagami-Pareto distribution shows much closer fit than composite exponential- Pareto distribution, composite lognormal-Pareto distribution and other classic one(two)-parameter distributions. We hope that the composite Nakagami-Pareto distribution will attract attention of researchers working in reliability analysis, epidemic and actuarial sciences to model data in their respective fields. In future research, we can employ three density functions to introduce a new composite model.



## REFERENCES

- [1] CEBRIAN, A.; DENUIT, M.; LAMBERT, P. Generalized Pareto fit to the society of Actuaries' large claims database. **North American Actuarial Journal**, v. 7, p. 18-36, 2003.
- [2] ELBATAL, I.; ARYAL, G. A new generalization of the exponential Pareto distribution. **Journal of Information and Optimization Sciences**, v. 38, n. 5, p. 675-697, 2017. DOI:10.1080/02522667.2016.1220079
- [3] Cooray, K.; Ananda. M. A. Modeling actuarial data with a composite Lognormal-Pareto model. **Scand. Actuar. J.**, v. 5, p. 321–334, 2005.
- [4] KLUGMAN, S. A.; PANJER, H. H.; WILLMOT, G. E. **Loss Models: From Data to Decisions**. New York: Wiley, 1998.
- [5] MCNEIL, A. J. Estimating the Tails of Loss Severity Distributions Using Extreme Value Theory. **Astin Bulletin**, v. 27, n. 1, p. 1-21, 1997.
- [6] Rashad, M. E.-S., Mahmoud, M. A. W.; & Abdallah, S. H. M. Statistical inferences for new Weibull-Pareto distribution under an adaptive type-ii progressive censored data. **Journal of Statistics and Management Systems**, v. 21, n. 6, p. 1021-1057, 2018. DOI:10.1080/09720510.2018.1467628
- [7] SCOLLNIK, D. P. M. On composite Lognormal-Pareto models. **Scan-dinavian Actuarial Journal**, p. 20–33, 2007.
- [8] TEODORESCU, S.; VERNIC, R. Some composite Exponential-Pareto models for actuarial prediction. **Rom. J. Econ. Forecast.**, v. 12, p. 82–100, 2009.
- [9] TEODORESCU, S.; VERNIC, R. A composite Exponential-Pareto distribution. **The Annals of the “Ovidius” University of Constanta, Mathematics Series**, v. XIV, n. 1, p. 99-108, 2006.
- [10] TEODORESCU, S.; VERNIC, R. On composite Pareto models. **Math.Reports**, v. 15, n. 65, 1, p. 11–29, 2013.
- [11] PREDA, V.; CIUMARA, R. On composite models: Weibull-Pareto and Lognormal-Pareto. A comparative study. **Rom. J. Econ. Forecast.**, v. 3, p. 32–46, 2006.

## BIBLIOGRAPHIE

- [1] AMINZADEH, Mostafa S. et DENG, Min. Bayesian predictive modeling for Inverse Gamma-Pareto composite distribution. *Communications in Statistics-Theory and Methods*, 2019, vol. 48, no 8, p. 1938-1954.
- [2] BARNDORFF, Nielsen. Information and exponential families ; in statistical theory. 1978.
- [3] BENATMANE, Chaabane, ZEGHDOUDI, Halim, SHANKER, Rama, et al. Composite Rayleigh-Pareto distribution : Application to real fire insurance losses data set. *Journal of Statistics and Management Systems*, 2021, vol. 24, no 3, p. 545-557.
- [4] BOWERS, Newton L. et NEWTON, L. Expansion of probability density functions as a sum of gamma densities with applications in risk theory. *Trans. Soc. Actuaries*, 1966, vol. 18, no 52, p. 125-147.
- [5] BOWERS, N. L., H. U. GERBER, J. C. HICKMAN, D. A. JONES et C. J. NESBITT. (1997), *Actuarial Mathematics*, 2e éd., The Society of Actuaries, Schaumburg, Illinois.

## Bibliographie

---

- [6] CEBRIÁN, Ana C., DENUIT, Michel, et LAMBERT, Philippe. Generalized Pareto fit to the society of actuaries' large claims database. *North American Actuarial Journal*, 2003, vol. 7, no 3, p. 18-36.
- [7] COORAY, Kahadawala et ANANDA, Malwane MA. Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, 2005, vol. 2005, no 5, p. 321-334.
- [8] DE PRIL, Nelson. Moments of a class of compound distributions. *Scandinavian Actuarial Journal*, 1986, vol. 1986, no 2, p. 117-120.
- [9] EL-SAGHEER, Rashad M., MAHMOUD, Mohamed AW, et ABDALLAH, Samah HM. Statistical inferences for new Weibull-Pareto distribution under an adaptive type-ii progressive censored data. *Journal of Statistics and Management Systems*, 2018, vol. 21, no 6, p. 1021-1057.
- [10] ELBATAL, Ibrahim et ARYAL, Gokarna. A new generalization of the exponential Pareto distribution. *Journal of Information and Optimization Sciences*, 2017, vol. 38, no 5, p. 675-697.
- [11] KAAS, Rob, GOOVAERTS, Marc, DHAENE, Jan, et al. *Modern actuarial risk theory : using R*. Springer Science & Business Media, 2008.
- [12] KLUGMAN, Stuart A., PANJER, Harry H., et WILLMOT, Gordon E. *Loss models : from data to decisions*. John Wiley & Sons, 2012.
- [13] KLUGMAN, Stuart A., PANJER, Harry H., et WILLMOT, Gordon E. *Loss models : from data to decisions*. John Wiley & Sons, 2012.
- [14] LIU, Bowen et ANANDA, Malwane MA. Analyzing insurance data with an exponentiated composite inverse Gamma-Pareto model. *Communications in Statistics-Theory and Methods*, 2023, vol. 52, no 21, p. 7618-7631.

## Bibliographie

---

- [15] MORRIS, Carl N. Natural exponential families with quadratic variance functions. *The Annals of Statistics*, 1982, p. 65-80.
- [16] PANJER, Harry H. Recursive evaluation of a family of compound distributions. *ASTIN Bulletin : The Journal of the IAA*, 1981, vol. 12, no 1, p. 22-26.
- [17] PREDA, Vasile et CIUMARA, Roxana. On composite models : Weibull-Pareto and Lognormal-Pareto. A comparative study. *Romanian Journal of Economic Forecasting*, 2006, vol. 3, no 2, p. 32-46.
- [18] ROLSKI, Tomasz, SCHMIDLI, Hanspeter, SCHMIDT, Volker, et al. *Stochastic processes for insurance and finance*. John Wiley & Sons, 2009.
- [19] SCOLLNIK, David PM. On composite lognormal-Pareto models. *Scandinavian Actuarial Journal*, 2007, vol. 2007, no 1, p. 20-33.
- [20] SUNDT, Bjørn. On some Extensions of Panjer's Class of Counting Distributions1. *ASTIN Bulletin : The Journal of the IAA*, 1992, vol. 22, no 1, p. 61-80.
- [21] SUNDT, Bjørn et JEWELL, William S. Further results on recursive evaluation of compound distributions. *ASTIN Bulletin : The Journal of the IAA*, 1981, vol. 12, no 1, p. 27-39.
- [22] TEODORESCU, Sandra, VERNIC, Raluca, et al. Some composite Exponential-Pareto models for actuarial prediction. *Romanian Journal of Economic Forecasting*, 2009, vol. 12, no 4, p. 82-100.
- [23] TEODORESCU, Sandra, VERNIC, Raluca, et al. Some composite Exponential-Pareto models for actuarial prediction. *Romanian Journal of Economic Forecasting*, 2009, vol. 12, no 4, p. 82-100.
- [24] TEODORESCU, Sandra et VERNIC, Raluca. On composite Pareto models. *Mathematical Reports*, 2013, vol. 15, no 65, p. 11-29.