Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE SAAD DAHLEB DE BLIDA
Faculté des sciences

Département de Mathématiques

Mémoire de Master
Spécialité : Mathématiques
Option : Modèles stochastique et statistique (MSS)

Par

Bouzar Nawal & Bouziani Fatima

Caractérisation des modèles composés et ses applications pour des données actuarielles

---

Devant le jury composé de :

TAMI Omar	Université Blida 1	Président
FRIHI Redhouane	Université Blida 1	Examinateur
RASSOUL Abdelaziz	ENSH de Blida	Promoteur

Mars 2023

REMERCIEMENTS

Ce travail est le résultat d'un dur labeur et de beaucoup de sacrifices, nos remerciements vont d'abord à ALLAH, créateur de l'univers qui nous a doté d'intelligence, le courage la volonté et nous a maintenu en santé mentale et physique pour mener à bien réaliser ce mémoire.

Nous aimerions exprimer nos profonde gratitude à notre promoteur RASSOUL Abdelaziz avec qui nous avons eu le plaisir de travailler sous sa direction et qui a guidé notre mémoire pour avoir accepté de nous guider par sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion, et ses critiques constructives.

Nous remercions également tous les membres du jury pour nous avoir honorés par leur présence et pour avoir accepté d'évaluer notre travail.

DÉDICACE 1

Je dédie ce modeste travail Mes très chers parent sans leur amours, leur sacrifies et leurs encouragements je ne serais jamais arrivée à réussir dans mes études. Le sais bien quel que soit les remerciements que je leurs adresse c'est peu, que Dieu les protège et leur donne la santé et une longue vie.

Une spéciale dédicace à une persenne qui compte beaucoup pour moi mon mari.

Mes copines GACEM Halima, HAMMADI Khadidja , TREA Fatma ,

"B. Fatima"

DÉDICACE 2

Je tiens c'est avec grande plaisir que je dédie ce modeste travail :
A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs peiéres tout au long de mes études
A mon cher mari,"Djamel" .
Je tiens á remercier tout particuliérement mon promoteur, RASSOUL Abdelaziz, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contri-bué á alimenter ma réflexion.
A toute personne qui accupe une place dans mon coeur.

"B. Nawal"

Table des matières

TABLE DES FIGURES

LISTE DES TABLEAUX

Abréviations et Notations

Abréviations

E(.) :espérance mathématique .
V(.) : variance.
f :densité de probabilité.
F :fonction de répartition
$X_1, X_2, .., X_n$: échantillon de taille n .
v. a : Variable aléatoire.
$B(n, \theta)$:le biais de l'estimateur T_n.
SLD : Soins Longue Durée .
SAP : Systems, Applications and Products in data processing (Le système logiciel SAP a été développé en 1971 par cinq ingénieurs d'IBM soient, Hopp, Wellenreuther, Hector, Tschira et Plattner, qui travaillaient ensemble sur un projet interne. En juin 1972, ils quittent IBM et fondent SAP) .
AIC : Akaike's Information Criterion.
K-S :Kolmogorov-Smirnov .
CMS :Cramer-von Mises statistic .
ADS :Anderson-Darling statistic .

ملخص

في هذه المذكرة قمنا بدراسة و توصيف النماذج المركبة من عدة متغيرات عشوائية مستمرة ، خاصة بالنسبة للظواهر المركبة التى تعد اكثر انتشارا في عدة مجالات تطبيقية . و قد بينا ان هذه النماذج المقترحة تعمل عتى تحسين و نمذجة الظواهر بطريقة فعالة و مرنة، و قد بينا من خلال التطبيقات المدروسة مدى فعالية هذه النماذج .

Résumé :

Dans ce mémoire, nous avons étudié et décrit les modèles composés de plusieurs variables aléatoires continues, en particulier pour les phénomènes complexes qui sont largement répandus dans plusieurs domaines d'application. Nous avons démontré que ces modèles proposés permettent d'améliorer et de modéliser efficacement les phénomènes de manière flexible. À travers les applications étudiées, nous avons mis en évidence l'efficacité de ces modèles.

Abstract :

In this paper, we have studied and described composite models composed of several continuous random variables, especially for complex phenomena that are widely prevalent in various application domains. We have demonstrated that these proposed models work to improve and model phenomena in an effective and flexible manner. Through the studied applications, we have shown the effectiveness of these models.

INTRODUCTION GÉNÉRALE

Dans la modélisation statistique, l'objectif principale est de rechercher la loi de probabilité qui décrit le mieux les observations issues d'un ensemble de donné et qui doit représenter le processus de génération de données sous-jacent. Les distributions de probabilité obtenues doivent posséder des propriétés souhaitables telles que la flexibilité de la modélisation de différentes formes et rester sous une forme traitable.
Les distributions courantes dans la littérature telles que : exponentielle, normale, Gamma, weibull,...etc. n'ont pas la capacité d'incorporer toutes les caractéristiques d'un ensemble de données sur la taille des sinistres. Par conséquent, le concept de distribution composite a été introduit pour modéliser les données sur la taille des sinistres. Avec un tel concept, de nombreux modèles composites différents ont été développés, notamment lognormal-Pareto, exponentiel-Pareto, Weibull-Pareto, etc.
L'idée du modèle composite a été introduite plus tard par Bakar et al. [1] pour construire des nouveaux modèles composites basés sur la distribution de Weibull pour les données de pertes d'assurance à queue lourde. Avec une telle idée, un grand nombre de modèles composites possibles ont été explorés par Michael Mitzenmacher (2004) [2] pour proposer des modèles par une distribution en loi de puissance ou une distribution log-normale. En général, la distribution de Pareto est considérée comme bonne pour modéliser les réclamations de grande taille. Cependant, pour modéliser les sinistres de petite taille, il existe de nombreuses variantes dans la littérature.
Les modèles composés, également appelés modèles de mélange ou modèles hiérarchiques, sont des modèles statistiques qui combinent différents composants pour capturer diverses sources de variabilité dans les données. Ces modèles sont particulièrement utiles en actuariat pour analyser et prévoir les données liées à l'assurance.

Les modèles composites fournissent un cadre souple pour intégrer diverses sources de variabilité dans l'analyse des données actuarielles. Ils permettent aux actuaires de prendre en compte différents facteurs de risque, de capturer des interactions complexes et de prendre des décisions plus éclairées dans les applications liées à l'assu-

rance. De nombreuses méthodes peuvent être mises en œuvre pour exploiter des modèles mathématiques de situations réelles et étudier leurs propriétés. Pour des valeurs de paramètres fixes, si le problème de précision finie est exclu des calculs informatiques, il est parfois possible de calculer les grandeurs d'intérêt explicitement, ou du moins de manière exacte, via des formules fermées. Malheureusement, ce n'est pas possible, la plupart temps, plutôt que des modèles très simplifiés tels que des propriétés avec de nombreux invariants et symétries. On peut obtenir des résultats du premier type, mais dans le cadre asymptotique, caractérisant le comportement de la grandeur étudiée lorsqu'un paramètre tend vers une valeur limite (Par exemple, lorsque la taille du système modélisé ou l'échelle de temps considérée tend vers l'infini, voire lorsque certains paramètres correspondant à des interactions tendent vers zéro, etc.). Alternativement, on peut étudier des approximations (plus ou moins bien contrôlées) du modèle initial, plus adaptées au premier type d'approche.

Enfin, des résultats plus qualitatifs sur le comportement du modèle peuvent parfois être obtenus. La simulation consiste à reproduire artificiellement les fonctions du modèle étudié et constitue l'un des moyens importants de l'exploiter. En particulier, il peut tester ou réfuter des hypothèses, obtenir des informations quantitatives (qui peuvent être utilisées pour affiner dès le cas échéant), de vérifier certaines approximations, d'évaluer la sensibilité du modèle à certaines hypothèses ou à certains paramètres, ou tout simplement d'explorer comment le modèle se comporte lorsqu'il est peu connu ou mal compris. Ce mémoire est composé de trois chapitre qui

- Le premier chapitre présente des généralités sur les lois de probabilités premièrement par variable aléatoire qui est une variable peut prendre différentes valeurs (attribuées au hasard), chacune ayant une probabilité associée, et ce dernier peut être discrète (le support est fini ou dénombrable) ou continue (le support est un intervalle des nombres réels), après on a déterminer la relation entre la fonction de probabilité et la variable aléatoire est examinée puis on a parlé sur la théorie de l'estimation est l'une des branches les plus fondamentales de la statistique. Cette théorie est généralement divisée en deux composantes principales : estimation paramétrique et non paramétrique. Les approches non paramétriques estiment diverses fonctions directement à partir des informations disponibles sur l'ensemble d'observations. Avec cette approche, on dit souvent que les donnèrent d'elles-mêmes. Dans le cadre de cette recherche, nous nous intéressons aux approches paramétriques.
- Dans le Deuxième chapitre on a présenté des variables aléatoires composées, notamment les modèles log-normale composé avec pareto, lognormale-loglogistique et lognormale-Stoppa, modèle Weibull-pareto, Pareto-Stoppa, et Weibull-Stoppa.
- Le Troisième chapitre consacré pour la simulation de quelques modèles composite, notamment, lognormal-Fréchet, et le modèle Weibull-Gamma. Une autre

partie est pour l'ajustement des données réelles avec des modèles composé, on considère le jeu des données danish fire et on a montré que ce jeu est bien modélisé par lognormale-loglogistique, un autre exemple sur des données de l'indice financier SAP qui est modélisé par un modèle Pareto-Normal-Pareto.

CHAPITRE 1

GÉNÉRALITÉS SUR LES LOIS DE PROBABILITÉS

1.1 Définition des Variables aléatoires

Soit $(\Omega, \mathcal{T})$ un espace de probabilisable, on appelle **variable aléatoire** sur $(\Omega, \mathcal{T})$, toute **application** $X : \Omega \longrightarrow \mathbb{R}$ telle que

$$\forall a \in \mathbb{R}, \{\omega \in \Omega; X(\omega) \leq a\} \in \mathcal{T}$$

si $\mathcal{T} = P(\Omega)$ (ce qui sera le cas pour nous si Ω est fini ou dénombrable) alors toute application de Ω dans $\mathbb{R}$ est une variable aléatoire .

Définition 1.1 *Soit X une variable aléatoire sur l'espace probabilité $(\Omega, \mathcal{T}, P)$. On appelle* ***Fonction de répartition*** *de X l'application définie par :*

$$F_X : \mathbb{R} \longrightarrow [0,1]$$
$$x \longmapsto F_X(x) = F(x) = P(X \leq x)$$

et qui vérifié les propriétés suivantes :

1. $0 \leq F(x) \leq 1$
2. $\lim_{x \to -\infty} F(x) = 0$ *et* $\lim_{x \to +\infty} F(x) = 1$
3. $F(x)$ *est une fonction croissante*
4. F_X *est continue a droite en tout point de* $\mathbb{R}$, *i.e.*

$$\forall a \in \mathbb{R}, \lim_{x \to a^+} F_X(x) = F_X(a)$$

1.1.1 Variables aléatoires discrètes

Définition 1.2 *Une variable aléatoire est dite discrète si elle ne prend que des valeurs discontinues dans un intervalle donné (borné ou non borné).L'ensemble des nombres entiers*

et discret.En règle générale,toutes les variables qui résultent d'un dénombrement ou d'une numération sont de type discret.

Définition 1.3 *Soit X une v.a.r discret prenant ses valeurs dans un ensemble* $\{x_1, x_2, .., x_n\}$ *éventuellement infini.Alors la loi de X est caractérisée par l'ensemble des probabilités* $P(X = x_i)$*, c'est-à-dire les nombres réels positifs* P_i *tel que* $P = (X = x_i) = p_i$ *avec* $0 \leq P_i \leq 1$ *et* $\sum_{i=1}^{n} P_i = 1$.

1.1.2 Exemples des v.a. discrètes

1.1.2.1 loi uniforme sur [1 ;n]

X suit la loi uniforme sur [1 ;n] si :

1. $X(\Omega)$ =[1 ;n]
2. quelque soit appartient [1 ;n],

$$P(X = i) = \frac{1}{n}$$

On note alors : $X \rightsquigarrow u([1; n])$
Il s'agit simplement, comme son nom l'indique, d'une loi dont tous les poids de probabilité sont identiques.

- Son espérance est :

$$E(X) = \frac{n+1}{2}$$

- Sa variance est :

$$V(X) = \frac{n^2 - 1}{12}$$

1.1.2.2 Loi de Bernoulli

Définition 1.4 *La variable aléatoire X définie sur un espace de probabilité* (Ω, A, P)*; où est l'ensemble fondamental des évènements , A la* σ*-algèbre des évènements etP la probabilités définie sur l'espace probabilisable* (Ω, A)*, est dite suivre une loi de Bernoulli de paramètre P si* $X(\Omega) = [0, 1]$ *ces deux valeurs sont dites aussi échec et succès. On a :*

$$P(X = 0) = q, P(X = 1) = p \text{ avec } p + q = 1$$

- *Son espérance est* $E(X) = p$

- *Sa variance est* $V(X) = pq$ *est appelée loi de bernoulli notée* $X \rightsquigarrow \mathcal{B}(1; p)$

1.1.2.3 Loi Binomiale $\mathcal{B}(n,p)$

Définition 1.5 *soit la répétition de n épreuves indépendantes de BERNOULLI de paramètre p .la loi de la variable aléatoire X égale au nombre de succès dans la répétition de ces n épreuves est dite Binomiale ; On a :*

$$P(X=K)=C_n^K(1-p)^{n-K}, K=0,1,2,..n.$$

- *Son espérance est* $E(X)=np$
- *Sa variance est* $V(X)=npq$, *avec* $p=q=1$

1.1.2.4 Loi de Poisson

Définition 1.6 *On dit qu'une v.a X a valeurs dans* $\mathbb{N}$ *suit une loi de Poisson de paramètre* $\lambda>0$, *notée* $\mathcal{P}(\lambda)$, *si*

$$P(X=K)=e^{-\lambda}(\frac{\lambda^K}{K!}), K\in\mathbb{N}.$$

- *Son espérance est* $E(X)=\lambda$
- *Sa variance est* $V(X)=\lambda$

1.1.3 Variables aléatoires Continues

Définition 1.7 *une variable aléatoire est dite continue si elle peut prendre toutes les valeurs dans un intervalle donné (dorne ou non borné). En règle générale, toutes les variables qui résultent d'une mesure sont de type continu.*

Définition 1.8 *Soit X une v.a.r. qui prend un nombre infini non dénombrable de valeurs. Si* F_X *est une fonction continue , on dit que X est une v.a.r.Continue. Dans ce cas, la loi de X est déterminée par l'ensemble des probabilités* $P(a<X<b)$.

1.1.4 Exemples des v.a. continues

1.1.4.1 Loi Uniforme

Définition 1.9 *une variable aléatoire X absolument continue est dite distribuée uniformément sur l'intervalle [a,b] si sa densité de probabilité est :*

$$f(x)=\begin{cases}\frac{1}{b-a}, & six\in[a,b]\\ 0, & ailleurs\end{cases}$$

◊ *Sa fonction de répartition :*

$$F_X(x) = \begin{cases} 0, & si \quad x < a \\ \frac{x-a}{b-a}, & si \quad x \in [a,b] \\ 1, & si \quad x \geq b \end{cases}$$

◊ *l'espérance d'une variable aléatoire uniforme est :*

$$E(X) = \frac{a+b}{2}$$

◊ *la variance d'une variable aléatoire uniforme est :*

$$Var(X) = \frac{(a-b)^2}{12}$$

1.1.4.2 Loi exponentielle

Définition 1.10 *La loi exponentielle décrit la durée de vie d'un phénomène sans vieillissement (particule radioactive,temps d'attente,.....). La densité de probabilité d'une variable aléatoire continue suivante une loi exponentielle $exp(a)$ est :*

$$f(x) = ae^{-ax}, x \in R,$$

où a est un nombre réel strictement positif, d'espérance :

$$E(X) = \frac{1}{a}$$

et de variance :

$$V(X) = \frac{1}{a^2}$$

1.1.4.3 loi normale

Définition 1.11 *Une variable aléatoire X absolument continue est dite normal (ou suit une loi normale, ou gaussienne) si elle admet pour densité de probabilité la fonction :*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{(x-m)^2}{2\sigma^2}), x \in R$$

ou m et σ sont respectivement la moyenne et l'écart-type. On dit aussi que X suit une loi normal de paramètre $\mathcal{N}(m,\sigma)$.

si

$$X^* = \frac{x-m}{\sigma}$$

est la variable aléatoire réduite correspondante, alors $E(X^*) = 0$ et $V(X^*) = 1$
et la densité de X^* s'exprime par :

$$g(u) = \frac{1}{\sqrt{2\pi}} * exp(-\frac{u^2}{2}),$$

si X suit une loi normale $\mathcal{N}(m, \sigma)$, alors $E(X) = m$,et $V(X) = \sigma^2$.
On désigné par ϕ cette fonction :

$$P(X \leq a) = P(\frac{X-m}{\sigma} \leq \frac{a-m}{\sigma}) = P(X^* \leq \frac{a-m}{\sigma}) = \Phi(\frac{a-m}{\sigma})$$

-les valeurs de Q sont données par les tables . Elles permettent de faire les calculs sur une loi normale de paramètres (m, σ) quelconques.
Cette loi est très importante tant du point de vue théorique que pratique.

1.1.4.4 Loi de Gamma

Définition 1.12 *On présent la famille de lois Gamma où d'Euler très utiles pour les propriétés de décroissance rapide de leur fonction de survie. Une variable aléatoire continue suit une loi Gamma , de paramètres $\alpha, \beta \in \mathbb{R}_+^*$*
le premier est appelé paramètre d'échelle alors que β est le paramètre de forme, si elle admet pour densité de probabilité la fonction :

$$f_X(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} x^{-\beta x}, & \forall x > 0, \\ 0, & ailleurs \end{cases}$$

où

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

On note $X \rightsquigarrow GA(\alpha, \beta)$.

- ◊ *L'espérance mathématique de X est :*
$$E(X) = \frac{\alpha}{\beta}$$
- ◊ *la variance de X est :*
$$V(X) = \frac{\alpha}{\beta^2}$$

1.1.4.5 Loi de Fisher-Snedecor

Définition 1.13 *Soit x une v.a. suit une loi Fisher-Snedecor á n_1 et n_2 degrés de liberté si :*

1. $X(\Omega) = \mathbb{R}_+$.

2. *la fonction de densité de probabilité f est définie pour tout x positif par :*

$$f(X) = n_1^{\frac{n_1}{2}} . n_2^{\frac{n_2}{2}} \cdot \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{n_2}).\Gamma(\frac{n_2}{n_1})} \cdot \frac{x^{\frac{n_2-1}{2}}}{(n_2 x+n_1)^{(\frac{n_1+n_2}{2})}}$$

avec :

$$\Gamma(n) = \int_0^\infty u^{n-1} exp(-u) du.$$

Notation : $X \rightsquigarrow \mathbf{F}(n_1, n_2)$
Les moments de la loi de Fisher-Snedecor font l'objet de la proposition suivante :

Proposition 1.1 *Si $X \rightsquigarrow \mathbf{F}(n_1; n_2)$, alors*

$$E(X) = \frac{n_2}{n_2-2} \ \ pour \ \ n_2 > 2$$

et

$$V(X) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-4)(n_{22})^2} \ \ pour \ \ n_2 > 4$$

L'espérance de X ne dépend pas de n_1 et lorsque n_2 tend vers 1 par valeur supérieure, par ailleurs, lorsque n_1 et n_2 tendent vers l'infinie , la variance de X tend vers 0.

1.1.4.6 Loi du Khi-Deux

Définition 1.14 *Soient K variable aléatoires $X_1, X_2, ...X_K$ indépendantes suivant une loi normale*
$\mathcal{N}(m_i; \sigma_i), i = 1, ..., K.$
la variable aléatoire

$$Z = \sum_{i=1}^K (\frac{X_i - m_i}{\sigma_i})^2$$

Suit une loi de Khi-deux (χ_K^2) a K degrés de liberté.
La densité de probabilité de la variable aléatoire Z est :

$$f(X) = \begin{cases} \frac{1}{2^{\frac{K}{2}}\Gamma(\frac{K}{2})} e^{-\frac{x}{2}} x^{\frac{K}{2}-1}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

oú Γ est la fonction gamma, On a :

$$E(Z) = K \ \ et \ \ V(Z) = 2K.$$

Cette loi est surtout utile dans les tests statistiques.

1.1.4.7 Loi de Student

Définition 1.15 *Soient Z et Q deux variable aléatoires indépendantes telles que Z suit $\mathcal{N}(0;1)$ et Q suit $\chi^2(v)$. Alors la variable aléatoire*

$$T = \frac{Z}{\sqrt{\frac{Q}{v}}}$$

*suit une loi appelée **loi de Student** a v degrés de liberté , notée $St(v)$.*

1. *La densité de la loi de Student á v degrés de liberté est*

$$f(x) = \frac{1}{\sqrt{\pi v}} \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})} \frac{1}{(1+\frac{x^2}{v})^{\frac{v+1}{2}}}$$

2. *L'espérance n'est pas définie pour $v = 1$ et vaut 0 si $v \geq 2$. Sa variance n'existe pas pour $v \leq 2$ et vaut $v/(v-2)$ pour $v \geq 3$.*
   La loi de Student converge en loi vers la loi normale centrée réduite .

1.1.5 La Fonction de densité

Définition 1.16 *Soit X une v.a.r **continu** : la fonction de densité f_X qui caractérise entièrement la loi de X (elle est dite loi de probabilité de X), est une fonction $f : \mathbb{R} \longrightarrow \mathbb{R}$ telle que :*

1. $\forall x \in \mathbb{R} : f(x) \geq 0.$

2. $\int_{\mathbb{R}} f(x).dx = 1$

1.1.6 Fonction caractéristique

Soit X une variable aléatoire , on appelle fonction caractéristique de X,la fonction de la variable réelle t définie par :

$$\varphi_X(t) = E[e^{ixt}].$$

Soient X une variable aléatoire $\varphi(t)$ sa fonction caractéristique , on appelle seconde fonction caractéristique de X la fonction

$$\Psi(t) = \log \varphi(t), t \in \mathbb{R}.$$

1. Si X est une variable aléatoire discrète alors

$$\varphi(t) = \sum_K P_K e^{itx}$$

2. Si X est une variable aléatoire continue de densité f , alors

$$\varphi(t) = \int_{\mathbb{R}} e^{itx} f(x)dx.$$

1.1.7 Fonction génératrices

Soit X une variable aléatoire a valeurs dans $\mathcal{N}$, pour tout lois de probabilité $P_K = P(X = K)$ en désignons par $G(S)$ la fonction génératives de X :

$$G_X(s) = \sum_{K\geq 0} P_K s^K.$$

Proposition 1.2 *La fonction génératrice $G_X(s)$ admet dérivée gauche G' en S=1, si seulement si $E(x)$ exist et est fini*

$$E(X) = G'(1).$$

Corollaire 1.1

1. *Si $X_1, X_2, ...X - n$ sont des variables aléatoires mutuellement indépendantes et de meme loi , dont la fonction génératrice est $G(s)$ alors la fonction génératrice est $S_n = X_1 + X_2 + ... + X_n$ est donnée par :*

$$G_{Sn}(s) = (G(s))^n$$

2. *La fonction génératrice de la somme $S_n = X_1 + X_2 + ...X_n$ où $\mathbb{N}$ est une variable aléatoire á valeurs entières indépendantes de la suit X_n, est la fonction composée*

$$\begin{aligned} G_{SN}(s) &= G_N(G_X(s)) \\ &= G_N \circ G_X(s) \end{aligned}$$

1.1.8 Fonctions génératrice des moments

Soit X une variable aléatoire réelle telle que la fonction

$$f_X(u) = E[e^u t]$$

cette fonction appelé fonction génératrice des moments de X.

Définition 1.17 *Soit (X, Y) un couple de variable aléatoire indépendantes dont chacune admet une fonction génératrice des moments, alors la somme $X + Y$ admet une fonction génératrice des moments et l'on a*

$$g_{X+Y} = g_X g_Y$$

Proposition 1.3 *1. Pour a, b réelles , on a*

$$g_{aX+b}(u) = e^{bu} g_X(au)$$

2. La fonction $g_x(-u)$ est aussi une fonction génératrice des moments.

3. la fonction g_x est convexe.

1.2 Éspérance

L'ésperance d'une variable aléatoire X, notée $E(X)$, représente la valeur moyenne pondérée prise par la variable X, plus précisément, c'est le barycentre du système $(x_i, P_i)(i = 1, ..., n)$, le "point" x_i étant effectué a la "masse" $P_i, P_i = P_X(x_i)$.

1. Si X v.a.r **discrète** :

$$E(X) = \sum_i x_i.P_X(x_i)$$

où P_X est la fonction de masse .

• Si $X(\Omega)$ est finie ie $X(\Omega) = \{x_1, ..., x_n\}$

$$\begin{aligned} E(X) &= x_1.P_X(x_1) + x_2.P_X(x_2) + ... + x_n.P_X(x_n) \\ &= \sum_{i=1}^{n} x_i.P_X(x_i) \end{aligned}$$

• Si $X(\Omega)$ est infini ie $X(\Omega) = \{x_1, x_2, ...\}$

$$E(X) = \sum_{i \geq 1} x_i.P_X(x_i)$$

2. Si X v.a.r **continue**

$$E(X) = \int_{\mathbb{R}} x.f(x).dx$$

♣ **Propriétés** :

◊ X est dite v.a.r **centrée** si $E(X) = 0$

◊ L'espérance est **linéaire** : en effet, soient $a, b \in \mathbb{R}$

$$E(aX + b) = aE(X) + b$$

$E(b) = b$ pour tout réel b

◊ Si $X \geq 0$, alors $E(X) \geq 0$

◊ Si $X \geq Y$, alors $E(X) \geq E(Y)$

◊ (L'espérance d un produit de v.a. indépendantes) : si le v.a X et Y ont un moment d ordre un et sont indépendantes , alors la v.a XY a un moment d'ordre un et on

a :

$$E(XY) = E(X).E(Y)$$

1.3 La variance et écart-type

La variance d'une variable aléatoire X , notée $V(X)$, est définie par :

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

souvent dans les calculs on utilisera la formule de Koenig

$$V(X) = E(X - E(X))^2 = E(X^2) - [E(X)]^2.$$

l'écart-type d'une variable aléatoire X , noté $\sigma(X)$, est définie par :

$$\sigma(X) = \sqrt{V(X)}.$$

Propriétés :

◊ X est dite une v.a.r **réduite** si

$$\sigma(X) = 1, V(X) = 1$$

◊ Soient $a, b \in \mathbb{R}$

$$V(aX + b) = a^2.V(X)$$

$$V(b) = 0, \quad V(aX) = a^2.V(X)$$

◊ Pour toutes les v.a.r

$$V(X) \geq 0 \quad et \quad \sigma(X) \geq 0$$

◊ (La variance de v.a **indépendances**) : Si les v.a. X et Y ont un moment d'ordre un et deux et sont indépendantes, alors , on a :

$$V(X + Y) = V(X) + V(Y).$$

1.4 Estimations

1.4.1 Généralités

La théorie de l'estimation étudie les propriétés des estimateurs et des méthodes générales d'estimation. L'objectif est de comparer les lois d'échantillonage des esti-

mateurs. Elle consiste à approximer les valeurs exacte et inconnues des paramètres d'une population statistique considéré ou d'un modèle mathématique à partir d'observation d'individus s'appelle échantillon. Le paramètre de la population est estimé à partir d'une statistique calculée sur la base d'un échantillon. L'estimation ponctuelle d'un paramètre consiste à évaluer la valeur du paramètre de la population à l'aide d'une valeur unique prise dans un échantillon. Pour évaluer la précision d'un estimateur, il est d'usage de construire un intervalle de confiance autour de cet estimateur. Soit X une variable aléatoire associée à un certain phénomène aléatoire observable de facon répétée. Notre objectif est d'estimer certaines caractéristiques d'intérêt de sa loi (la moyenne, la variance, ...) sur la base d'une série d'observations $x_1, x_2, ..., x_n$. Considérerons toujours, même si des développements analogues sont possibles dans d'autre circonstances, que$x_1, x_2, ..., x_n$ sont des réalisations d'un n échantillon aléatoire $X_1, X_2, ..., X_n$ Cette hypothèse sur nos observations qui peut être plus ou moins réaliste est nécessaire pour étudier de façons simple, en termes probabilistes, la qualité des estimations que l'on cherche à produire.

1.4.2 Définition d'un estimateur

Soit $(X_1, X_2, ..., X_n)$ un n-échantillon d'une loi $\mathbb{P}_\theta$ dépendant d'un paramètre inconnu $\theta \in \Theta \subset \mathbb{R}$ (ouvert de $\mathbb{R}^d; d \geq 1$) . On appelle **estimateur de** θ une variable aléatoire T_n obtenue comme fonction du n-échantillon aléatoire $(X_1, X_2, ..., X_n)$; autrement dit :

$$T_n = (X_1, X_2, .., X_n)$$

oú

$$f : \mathbb{R}^n \to \mathbb{R}^d; d \geq 1.$$

T_θ fournit une réalisation $(x_1, x_2, ..., x_n)$ qui est une estimation ponctuelle de θ

$$\theta_n = f(x_1, x_2, ..., x_n)$$

l'estimation est dite ponctuelle si on estime un paramètre de la population avec un seul nombre.
Pour un même paramètre, il peut y avoir plusieurs estimateurs possibles .

1.4.3 Propriétés générales d'un estimateur :

T_n désigne l'estimateur du paramètre θ.

1. Tout estimateur peut donner lieu á l'écriture :

$$B_n(\theta) = E_\theta(T_n)\text{-}\theta$$

où $B_n(\theta)$ est le biais de T_n.

2. Qn dite que T_n est un estimateur sans biais de θ si $E_\theta(T_n) = \theta$ oú E_θ désigne l'espérance sous la loi $\mathbb{P}_\theta$.
3. Si $B_n(\theta)$ tend vers 0 quand n tend vers l'infini, alors T_n est un estimateur de θ **asymptotique ment sans biais**.

1.4.4 Qualité d un estimateur :

1.4.4.1 Estimateur efficace

Un estimateur sans biais est efficace si sa variance est la plus faible parmi les variances autres estimateurs sans biais . Ainsi si $\hat{\theta}_1$ et $\hat{\theta}_2$ sont deux estimateurs sans biais du paramètre θ, l'estimateur $\hat{\theta}_1$ est efficace si :

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

et

$$E(\hat{\theta}_1) = E(\hat{\theta}_2)$$

1.4.4.2 Estimateur convergent

Un estimateur $\hat{\theta}$ est convergent si sa distribution tend á se concentrée autour de la valeur inconnue á estimerθ , á mesure que la taille d'échantillon augmente, c est-á-dire si :

$$\lim_{n\to\infty} V(\hat{\theta}) = 0.$$

Un estimateur sans biais et convergent est dit **absolument correct**.

1.4.5 Les méthodes d'estimation :

1.4.5.1 la méthode du maximum de vraisemblance

Soit un échantillon $(X_1, X_2, ..., X_n)$ dont la loi mère appartient á la famille paramétrique de densités $f(x_i, \theta)$ (ou de fonction de probabilité $\mathbb{P}(x_i, \theta)$) $\theta \in \Theta, \Theta \subset \mathbb{R}^d$.
La vraisemblance $L(x_1, x_2, ..., x_n; \theta)$ représente la probabilité d'observer n uplet $(x_1, ..., x_n)$ pour une valeur fixée du paramètre θ. Dans la situation inverse ici oú on a observé $(x_1, ..., x_n)$ sans connaître la valeur de θ , on va attribuer á θla valeur qui parait la plus vraisemblance , complet tenu de l'observation dont on dis-ose , c-á-d celle qui va lui attribuer la plus forte probabilité.
I : la fonction de vraisemblance :

Définition 1.18 *On appelle fonction vraisemblance de θ pour une réalisation $x = (x_1, ..., x_n)$ de l'échantillon $X = (X_1, ..., X_n)$ la fonction de θ :*

$$L(x_1, ..., x_n; \theta) = \begin{cases} \prod_{i=1}^{n} \mathbb{P}(x_i, \theta), & si \quad les \quad X_i \text{ sont discrètes.} \\ \prod_{i=1}^{n} f(x_i; \theta), & si \quad les \quad X_i \text{ sont continues.} \end{cases}$$

II : l'estimateur de maximum de vraisemblance(EMV) :

Définition 1.19 *On appelle estimateur du maximum de vraisemblance de θ la v.a qui maximise $L(x; \theta)$ et on a le notée $\hat{\theta}$:*

$$\hat{\theta} = arg_{\theta \in \Theta} max L(x; \theta).$$

en d'autre termes :

$$L(x; \theta) \geq L(x; \theta) \quad \forall \theta \in \Theta.$$

III : Équation de vraisemblance :
Soit un n échantillon $X = (X_1, ..., X_n)$ de valeur $(x_1, ..., x_n)$, cette méthode permet de prendre comme estimateur de $\theta \in \Theta$ la valeur $\hat{\theta}$ qui rend maximale la vraisemblance.
Cas d'un seul paramètre
l'estimateur du maximum de vraisemblance est une solution du système :
le système (I) est donc équivalent au système (I).

$$(I) \begin{cases} \frac{\partial L(x_1, .., x_n; \theta)}{\partial \theta} = 0 \\ \frac{\partial^2 L(x_1, ..., x_n; \theta)}{\partial \theta^2}(\hat{\theta}) < 0 \end{cases}$$

ou bien le système :

$$(II) \begin{cases} \frac{\partial ln L(x_1, .., x_n; \theta)}{\partial \theta} = 0 \\ \frac{\partial^2 ln L(x_1, .., x_n; \theta)}{\partial \theta^2}(\hat{\theta}) < 0 \end{cases}$$

le système (I) est donc équivalent au système (II).

1.4.5.2 La méthode des moments :

Cette méthode peut être la méthode la plus ancienne de trouver des estimateurs ponctuels et la plus naturelle.
L'idée de base est d'écrire les moments théoriques $E(X), E(X^2), ..., E(X^d)$ en fonction des paramètres.
Soit X une v.a qui suit une certaine loi de probabilité qui dépend d'un paramètre $\theta = (\theta_1, ..., \theta_d)$.
L'estimateur de θ par la méthode des moments (EMM) est obtenu en remplaçant les moments théoriques par les moments empiriques c-à-d :

$$\begin{cases} E(X) = m_1 = f_1(\theta_1,..,\theta_d), \\ \theta_1 = E(X) = m_2 = f_2(\theta_1,..,\theta_d), \\ . \\ . \\ E(X^d) = m_d = f_d(\theta_1,...,\theta_d), \end{cases} \Longleftrightarrow \begin{cases} \theta_1 = g_1(m_1,..,m_d), \\ \theta_2 = g_2(m_1,..,m_d), \\ . \\ . \\ \theta_d = g_d(m_1,..,m_d), \end{cases}$$

1.4.5.3 La méthode d'estimation par intervalle de confiance :

L'estimation ponctuelle d'un paramètre θ donne une valeur numérique unique à ce paramètre, mais, n'apporte aucune information sur la précision des résultats, c-à-d qu'elle ne tient pas compte des erreurs ducs aux fluctuations d'échantillonnage, par exemple pour evaluer la confiance que l'on peut avoir en une estimation, il est nécessaire de lui associer un intervalle qui contient avec une certaine probabilité la vraie valeur du paramètre, c'est l'estimation par intervalle de confiance.

Définition 1.20 *Soit*$(X_1,..,X_n)$ *un échantillon de loi . On appelle intervalle de confiance (IC) de niveau de confiance* $1-\alpha$ *telles que* $\alpha \in [0,1]$ *donné , un intervalle aléatoire* $[\theta_1,\theta_2]$ *oú* $\theta_1 \leq \theta_2$ *sont deux statistique fonction de l'échantillon, telles que :*

$$\mathbb{P} = (\theta_1 \leq \theta \leq \theta_2) = 1-\alpha$$

α est donc la probabilité que $[\theta_1;\theta_2] nerecouvrepaslavraievaleurduparamtre.$
La détermination des bornes d'un intervalle de confiance dépend de la coupure de α en α_1 et α_2 ; telles que θ_1 est le fractile d'ordre α_1 et θ_2 est le fractile d'ordre $(1-\alpha_2)$ d'une certaine loi. Cependant deux cas sont possibles Recherche d'un intervalle bilatéral. Correspondant à $\alpha_1 \neq 0$ et $\alpha_2 \neq 0$. Il n'y a aucune raison d'avoir $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ sauf pour certaines lois symétriques .
Recherche d'un intervalle unilatéral de la forme $[\theta_1;+\infty[$ associé à $\alpha_1 = \alpha$ et $\alpha_2 = 0$ ou de forme $]-\infty;\theta_2]$ associé à $\alpha_1 = 0$ et $\alpha_2 = \alpha$.
les valeurs usuelles de α sont 0.1,0.05 et 0.01 .

1.4.5.4 Estimation par intervalle de confiance pour les paramètres de la loi normale

Soit $(X_1,...,X_n)$ un échantillon d'une v.a X qui suit une loi normale $\mathcal{N}(m,\sigma^2)$,$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ sa moyenne empirique et $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ sa variance empirique. Pour construire un intervalle de confiance relatif à l'un des deux paramètres de cette loi, l'autre paramètre étant connu ou non. Ceci correspond aux di¤érentes situations que nous allons étudier maintenant. Nous rassemblons ci dessous et nous admettrons, les trois résultats permettant de calculer les intervalles de confiance de la moyenne m et de la variance σ^2.

Théorème 1.1 *Si $(X_1,..,X_n)$ est un échantillon de la loi $\mathcal{N}\left(m.\sigma^2\right)$, alors :*

1. $\sqrt{n}\frac{\overline{X}-m}{\sigma}$ *suit la loi normal $\mathcal{N}(m,\sigma^2)$ (d'après le théorème centrale limite (TCL)).*
2. $\frac{\overline{X}-m}{S}$ *suit une loi de student $\mathcal{T}_{n-1}$ de degré de liberté (n-1).*
3. $\frac{nS^2}{\sigma^2}$ *suit la loi du Khi-deux χ^2_{n-1}, degré de liberté (n-1).*

► **La moyenne**

Si la variance est connue :

$\overline{X} \sim \mathcal{N}(m, \frac{\sigma^2}{n})$, on a aussi $\overline{X}$ est le meilleur estimateur de m, et on a

$$\frac{\sqrt{n}(\bar{X}-m)}{\sigma} \sim \mathcal{N}(0,1)$$

alors

$$\mathbb{P}(-u_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{X}-m)}{\sigma} \leq u_{1-\frac{\alpha}{2}}) = 1-\alpha,$$

$$\mathbb{P}(-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq (\overline{X}-m) \leq u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

$$\mathbb{P}(\bar{X}-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq m \leq \overline{X}+u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) = 1-\alpha,$$

donc l'intervalle de confiance est :

$$IC(m) = [\overline{X}-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X}+u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}].$$

oú $u_{1-\frac{\alpha}{2}}$ est le fractille d'ordre $(1-\frac{\alpha}{2})$ de la loi $\mathcal{N}(0,1)$.

Si $n \geq 30$ et σest inconnu en remplace σ par S' telles que$S' = \sqrt{\frac{n}{n-1}}S$.

Si la variance est inconnue (n<30)

On utilise la fait que

$$T = \frac{\overline{X}-m}{S}\sqrt{n-1} \sim \mathcal{T}_{n-1}$$

$$\mathbb{P}(-t_{1-\frac{\alpha}{2}} \leq \frac{\overline{X-m}}{S}\sqrt{n-1} \leq t_{1-\frac{\alpha}{2}}) = 1-\alpha,$$

$$\mathbb{P}(-\frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}} \leq \overline{X}-m \leq \frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}}) = 1-\alpha,$$

$$\mathbb{P}(\overline{X}-\frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}} \leq m \leq \overline{X}+\frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}}) = 1-\alpha$$

donc l'intervalle de confiance est :

$$IC(m) = \left[\overline{X}-t_{1-\frac{\alpha}{2}\frac{S}{\sqrt{n-1}}}; \overline{X}+t_{1-\frac{S}{\sqrt{n-1}}}\right],$$

oú $t_{1-\frac{\alpha}{2}}$ est le quantille d'ordre $(1-\frac{\alpha}{2})$ de la loi student á (n-1) degré de liberté.

► **La variance**

Si la moyenne est connue

$T = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$ est le meilleur estimateur de la varaince σ^2 et on a $\frac{nT}{\sigma^2} \sim \chi^2_n$ comme somme de n carrées de $\mathcal{N}(0,1)$ indépendantes, alors :

$$\mathbb{P}\left(k_1 \leq \frac{nT}{\sigma^2} \leq k_2\right) = 1 - \alpha,$$

$$\mathbb{P}\left(\frac{k_1}{nT} \leq \frac{1}{\sigma^2} \leq \frac{k_2}{nT}\right) = 1 - \alpha,$$

$$\mathbb{P}\left(\frac{nT}{k_2} \leq \sigma^2 \leq \frac{nT}{k_1}\right) = 1 - \alpha,$$

donc l'intervalle de confiance de σ^2 est :

$$IC(m) = \left[\frac{nt}{k_2}, \frac{nt}{k_1}\right],$$

oú k_1 est le quantile d'ordre $\frac{\alpha}{2}$ et k_2 est le quantile d'ordre $(1 - \frac{\alpha}{2})$ de la loi χ^2_n.

Si la moyenne est inconnue

On utilise $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ et on sait que $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$,

$$\mathbb{P}\left(l_1 \leq \frac{nS^2}{\sigma^2} \leq l_2\right) = 1 - \alpha,$$

$$\mathbb{P}\left(\frac{l_1}{nS^2} \leq \frac{1}{\sigma^2} \leq \frac{l_2}{nS^2}\right) = 1 - \alpha,$$

$$\mathbb{P}\left(\frac{nS^2}{l_2} \leq \sigma^2 \leq \frac{nS^2}{l_1}\right) - 1 - \alpha,$$

donc l'intervalle de confiance de σ^2 est :

$$IC(\sigma^2) = [\frac{nS^2}{l_2}, \frac{nS^2}{l_1}].$$

oú l_1 est le quantille d'ordre $\frac{\alpha}{2}$ et l_2 d'ordre $1 - \frac{\alpha}{2}$ de la loi χ^2_{n-1}.

Si $n > 30$ donc on a les deux approximations suivantes :

$$\sqrt{2}\chi^2_p - \sqrt{2p-1} \to \mathcal{N}(0,1) si \quad p > 30$$

approximation de Fisher et

$$\chi_p^2 = p\left(u\sqrt{\frac{2}{9p}} + 1 - \frac{2}{9p}\right)^3$$

approximation de wilson Hilferty valable même pour les valeurs faible de p.

1.4.6 Estimateur des paramètres de distribution normale et exponentielle

1.4.6.1 Distribution normale

Estimateur de la méthode des moments :
Soit $X \sim N(\mu, \sigma^2)$, où et σ^2 sont inconnus. les deux premiers moments autour de l'origine sont donnés par :

$$\mu_1' = E(X) = \mu,$$

$$\mu_2' = E(X^2) = \sigma^2 + \mu^2$$

et les moments d'échantillonnage sont donnés par :

$$m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$m_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

ainsi, en utilisant la méthode des moments, nous avons :

$$m_1' = \mu_1' \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i = \mu \Rightarrow \hat{\mu} = \bar{X}.$$

De plus nous avons

$$m_1' = \mu_1' \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2 \Rightarrow \bar{X}^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Après avoir résolu pour σ^2 nous obtenons,

$$\begin{aligned} \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\ &= \frac{S^2}{n} \\ &= \hat{\sigma}^2 \end{aligned}$$

où

$$S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

la méthode des moments estimateur de μ et σ^2 sont $\hat{\mu} = \bar{X}$ et $\hat{\sigma}^2 = \frac{s^2}{n}$ respectivement .
Estimateur du maximum de variaseblance : Soit $X_1, X_2, ..., X_n$ des échantillons aléatoires distribués de manière identique et indépendante tirés de la distribution normale $X \sim N(\mu, \sigma^2)$. le pdf de la variable aléatoire X est donnée par

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

$$-\infty < x < +\infty; \sigma > 0; -\infty < \mu < +\infty$$

la fonction de vraisemblance est donnée par

$$\begin{aligned} L(x, \mu, \sigma^2) &= \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right) \\ &= \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} e^{\frac{-1}{2\sigma^2} \sum i=1^n (x_i-\mu)^2} \end{aligned}$$

CHAPITRE 2

VARIABLES ALÉATOIRES COMPOSÉES

Introduction

La modélisation des données de perte de type uni-modal avec une queue lourde a été un sujet intéressant pour les actuaires. Les distributions qui peuvent imiter la queue lourde des données sur les pertes d'assurance sont cruciales pour fournir suffisamment une bonne estimation du niveau de risque commercial associé. Plusieurs modèles à queue lourde ont été discutés dans la littérature, notamment les modèles de Pareto, log-normal, Weibull et gamma. Outre la modélisation des sinistres en assurance, l'application de ces modèles est diverse, par exemple, pour modéliser les rendements financiers et la taille des fichiers sur les serveurs de réseau, voir Resnick (2007)[3]. L'accent mis récemment sur la modélisation de ces données a été orienté vers des modèles composites suite à un article de Cooray et Ananda (2005) [4]. Ces modèles sont constitués en assemblant deux distributions pondérées à un seuil spécifié.
Pour plus de simplicité, nous nous référons à la distribution jusqu'au seuil et celle au-delà comme la tête et la queue de la distribution, respectivement. En général, le modèle composite discuté ci-dessus prend la forme suivante :

$$f(x) = \begin{cases} a_1 f_1^*(x) & if \quad x \leq \theta \\ a_2 f_2^*(x) & if \quad x \geq \theta \end{cases} \tag{2.1}$$

où $a_i, i = 1, 2$ sont les poids et $f_i^*(x), i = 1, 2$ sont les fonctions de densité de probabilité tronquées (pdf) du modèle composite. Notez cependant que la forme (2.1) n'est pas continue et lisse en général. Cooray et Ananda (2005)[5] modélisent la tète et la queue par des distribution lognormales et de Pareto pondérées, respectivement, et ont montré un meilleur ajustement á des donnée de perte réelles asymétriques que plusieurs modèles standard.Ils utilisent également les condition de continuité et de dérivabili-

téau seul,θ, pour s'assurer que les modèles résultants sont á la fois continus et lisses , Scollnik (2007)[6] a amélioré le modèle composite Lognormal-Pareto en permettant des poids de mélange flexibles,remplaçant un poids constant appliqué par Cooray et Ananda (2005) [5], résultant en un meilleur ajustement aux données sur les pertes . Il réitères ce fait en utilisant le modèle composite de Weibull-Pareto dans Scollnik et Sun (2012) [7].les poids de mélange des modèles développés dans Cooray et Ananda (2005) [5] peuvent être restrictifs, nous ne poursuivrons donc pas une telle approche. les découvertes les plus récentes de Nadarajah et Bakar (2014) [8] suggèrent que les données de perte peuvent être mieux modélisées par un modèle lognormal composite avec la queue d'une distribution de Bure que par un modèle avec la queue d'une distribution de Pareto. Tous ces auteurs illustrent leurs conclusions á l'aide des données bien connues des assurances incendie danoises. le modèle proposé par Scollnik et Sun (2012)[7] permettant un mélange flexible poids,r, peut être exprimé comme suit

$$f(x) = \begin{cases} r f_1^*(x), & if \quad 0 < x \leq \theta \\ (1-r) f_2^*(x), & if \quad \theta < x < \infty \end{cases} \tag{2.2}$$

où $f_1^*(x) = \frac{f_1(x)}{F_1(\theta)}$ et $f_1^*(x) = \frac{f_2(x)}{1-F_2(\theta)}$ sont respectivement les fdp de Weibull et de Pareto tronquées et de Pareto tronquées de deuxième espèce, où $f_1(x)$ et $f_2(x)$ sont les fdp de Weibull et de Pareto de deuxième espèce (cette dernière également connu sous le nom de lomax pdf) spécifié par

$$f_1(x) = \frac{\alpha}{\lambda}(\frac{x}{\lambda})^{\alpha-1} exp[-(\frac{x}{\lambda})^{\alpha}], \quad x > 0 \tag{2.3}$$

et

$$f_2(x) = \frac{\beta \sigma^{\beta}}{(\sigma + x)^{\beta+1}}, \quad x > 0 \tag{2.4}$$

respectivement , et $F_i(x), i = 1, 2$ sont les fonctions de distribution cumulées (cdfs) correspondantes.En appliquant la condition de continuité,$f(\theta-) = f(\theta+)$,r a la forme général suivante, comme indiqué dans Nadarajah et Bakar (2014)

$$r = \frac{f_1(\theta)[1 - F_2(\theta)]}{f_1(\theta)[1 - F_2(\theta)] + f_2(\theta)F_1(\theta)} \tag{2.5}$$

oú $0 \leq r \leq 1$. Le poids de mélange du modèle composite de Weibull-Pareto proposition Scollnik et Sun(2012) est donné par

$$r = \frac{\frac{\beta}{\alpha}}{\frac{\sigma+\theta}{\theta}\frac{\frac{\theta}{\lambda}}{exp(\frac{\theta}{\lambda})-1} + \frac{\beta}{\alpha}} \tag{2.6}$$

En appliquant la condition de dérivabilité en θ ,$f'(\theta-) = f'(\theta+)$,Scollnik et Sun(2012)[7] ont montré que les paramètres du modèle peuvent être réduits,c'est -á -dire en exprimant l'un des paramètres en fonction des autres

$$\lambda = \frac{\theta}{[\frac{\beta\theta-\sigma}{\alpha(\sigma+\theta)} - 1]^{\frac{1}{\alpha}}} \tag{2.7}$$

Le modèle composite résultant est donc continu et lisse sur l'espace $x > 0$. En bref ,le modèle composite de Weibull-Pareto est défini en spécifiant le poids de mélange,r, et le paramètre d'échelle de la distribution de Weibull,λ.
Dans cet article,nous proposons une nouvelle approche pour développer un modèle composite pour deux distribution quelconques quelconques.Plusieurs nouveaux modèles composites avec la tète basée sur la distribution de Weibull et la queue appartenant à une famille de distribution bêta transformées sont proposés.Dans la section 2 ,nous décrivons la méthode de construction du modèle composite en spécifiant les poids de mélange et le seuil en termes d'autres paramètres du modèle.Tous les nouveaux modèles ainsi que certaines distribution standard sont la section 3.En autre ,certaines applications des modèles aux mesures de risque sont présentées dans la section 4.Enfin,certaines conclusions sont tirées dans la section5.

2.1 Modèles composites de Pareto avec poids de mélange il imités

,Scollnik(2007)[6] a amélioré le modèle composite donné dans Cooray et Ananda(2005)[9] en incorporant évaluation des poids de mélange sans restriction.La fonction de densité du modèle composite peut s'écrire

$$f(x) = \begin{cases} r f_1^*(x) & 0 < x \leq \theta \\ (1-r) f_2^*(x) & \theta < x < \infty \end{cases} \tag{2.8}$$

avec $0 \leq r \leq 1, f_1^*(x) = \frac{f_1(x)}{F_1(\theta)}$ et $f_2^* = \frac{f_2(x)}{1-F_2(\theta)}$ des fonctions de densité de probabilité (pdf) f_1 et f_2 jusqu'à et après une valeur de seuil inconnue θ où $F_1(\theta)$ et $F_2(\theta)$ dénotent les cdf de f_1 et f_2 en θ , respectivement. Ensuite, (1) peut être vu comme une somme convexe de deux fonctions de densité et se présente donc sous la forme d'un modèle de mélange . Après avoir imposé la condition de continuité (i.e $f(\theta^-) = f(\theta^+)$), on a

$$r = \frac{f_2(\theta)F_1(\theta)}{f_2(\theta)F_1(\theta) + (1 - F_2(\theta))} \tag{2.9}$$

Ensuite, une condition de dérivabilité en θ également été imposée afin de rendre (1) lisse et de réduire le nombre de paramètres.

2.1.1 Modèles log-normal-Pareto

Soit X une variable aléatoire avec le pdf

$$f(x) = \begin{cases} cf_1(x) & si \quad 0 < x \leq \theta \\ cf_2(x) & si \quad \theta \leq x < \infty \end{cases} \tag{2.10}$$

où c est la constante de normalisation ,$f_1(x)$ a la forme de la densité log-normal á deux paramètres et $f_2(x)$ a la forme de la densité de Pareto á deux paramètres , c'est-á-dire .

$$f_1(x) = \frac{1}{\sqrt{2\pi}x\sigma} \exp\left(-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right), \quad x > 0 \tag{2.11}$$

$$f_2(x) = \frac{\alpha\theta^\alpha}{x^{\alpha+1}} \quad x > \theta \tag{2.12}$$

où , $\theta, \mu, \sigma, \alpha$ sont des paramètres inconnus tels que $\theta > 0, \sigma > 0, \alpha > 0$.
Imposons les condition de continuité et dérivabilité en θ

$$f_1(\theta) = f_2(\theta), \quad f_1^{'}(\theta) = f_2^{'}(\theta), \tag{2.13}$$

où $f^{'}(x)$ est la dérivée première de $f(x)$ évaluée á θ . Ces conditions garantissent que l'on a une fonction de densité de probabilité lisse . Ces deux restrictions réduisent le nombre total de paramètres inconnus de quatre á deux . On peut montrer que (les détails sont á la fin de la section 2) cette densité composite peut être reparamétrée et réécrite comme

$$f(x) = \begin{cases} \frac{\alpha\theta^\alpha}{(1+\Phi(k))x^{\alpha+1}} \exp\left(-\frac{\alpha^2}{2k^2}\ln^2\left(\frac{x}{\theta}\right)\right) & si \quad 0 < x \leq \theta \\ \frac{\alpha\theta^\alpha}{(1+\Phi(k))x^{\alpha+1}} & si \quad \theta \leq x < \infty \end{cases} \tag{2.14}$$

où$\Phi(.)$ est la fonction de distribution cumulative de la distribution normale standard et$k$ est une constante connue qui est donnée par la solution positive de l'équation $exp(-k^2) = 2\pi k^2$. Cette valeur est $k = 0.372238898$. Ici $\alpha\sigma = k$ et $c = 1/(1 + \Phi(k)))$. Par conséquent, cette densité de probabilité composite n'a que deux paramètres inconnus $\theta > 0$, et $\alpha > 0$. la fonction de distribution cumulative du modèle composite susmentionné est

$$F(x) = \begin{cases} \frac{1}{(1+\Phi(k))}\Phi((\alpha/k)ln(x/\theta) + k), & si \quad 0 < x \leq \theta \\ 1 - \frac{1}{(1+\Phi(k))}(\theta/x)^\alpha, & si \quad \theta < \infty \end{cases} \tag{2.15}$$

soit la pdf d'une distribution de Pareto à deux paramètres α et $\theta > 0$.
La fonction de densité de ce modèle composite est donnée par

$$f(x) = \begin{cases} r\frac{f_1(x)}{\phi(\alpha\sigma)}, & 0 < x \leq \theta \\ (1-r)f_2(x), & \theta < x < \infty \end{cases} \tag{2.16}$$

avec $0 \leq r \leq 1$ et $\Phi(.)$ désigne de cdf de la distribution normal standard .En tenant compte de la continuité et de la dérivabilité en θ , nous avons que

$$r = \frac{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)\exp(\frac{1}{2}(\alpha\sigma)^2)}{\sqrt{2\pi}\alpha\sigma\Phi(\alpha\sigma)\exp(\frac{1}{2}(\alpha\sigma)^2)+1} \quad et \quad \alpha\sigma = \frac{\ln\theta - \mu}{\sigma}. \tag{2.17}$$

Ensuite,(5) est défini au moyen du seuil θ, d'une indice de queue α et d'un petit paramètre de perte σ .
Scollnik(2007)[6]a également considéré un autre modèle composite ,le lognormal type Pareto (Lomax) avec pdf

$$f(x) = \begin{cases} r\frac{f_1(x)}{\Phi(\mathcal{A})}, & 0 < x \leq \theta \\ (1-r)f_2(x), & \theta < x < \infty \end{cases} \tag{2.18}$$

où

$$\mathcal{A} = \frac{\ln\theta - \mu}{\sigma} = (\frac{\alpha\theta - \lambda}{\lambda + \theta})\sigma$$

et $f_2(x)$ est la pdf du type II Pareto donnée par

$$f_2(x) = \frac{\alpha(\lambda+\theta)^\alpha}{(\lambda+x)^{\alpha+1}}, \quad x > \theta \tag{2.19}$$

où les paramètres sont $\theta > 0$, $\alpha > 0$, et $\lambda > -\theta$.
Après avoir imposé les exigences des continuité et de dérivabilite en θ , un modèle lisse á quatre paramètres densité est obtenue , r est maintenant fourmi par

$$r = \frac{\sqrt{2\pi}\alpha\theta\sigma\Phi(\mathcal{A})exp(\frac{1}{2}\mathcal{A}^2)}{\sqrt{2\pi}\alpha\theta\sigma\Phi(\mathcal{A})exp(\frac{1}{2}\mathcal{A}^2)+\lambda+\theta} \tag{2.20}$$

Notez que ce modèle imbrique le modèle composite lognormal-Pareto si $\lambda = 0$.

2.1.1.1 Estimation des paramètres

I : Estimation du maximum de variasrmblance pour des donnée complètes Soit $X_1, X_1, ..X_n$, un échantillon aléatoire du modèle composite lognormal-Pareto á deux paramètres décrit dans l'équation (2.14) . Sans perte de généralité (avec notre travail), nous pouvons supposer qu'il s'agit d'un échantillon ordonné , c'est-á-dire $x_1 \leq x_2 \leq$

$x_3 \leq ... \leq x_n$ .Supposons que paramètre inconnu θ se situe entre la m^{th} observation et $m+1^{th}$ observation , c'est-á-dire, $x_m \leq \theta \leq x_{m+1}$. Alors la fonction de vraisemblance est donné par

$$L(\alpha,\theta) = C_0 \alpha^n \theta^{n\alpha} \left(\prod_{i=1}^{n} x_i^{-\alpha} \right) \exp\left[-\frac{\alpha^2}{2k^2} \sum_{i=1}^{m} \ln^2(x_i/\theta) \right] \tag{2.21}$$

avec $C_0 = 1/[(\prod_{i=1}^{n} x_i)(1+\Phi(k))^n]$.
estimateurs du maximum de vraisemblance (ML) de θ et α, $\hat{\theta}_{ML}$ et $\hat{\alpha}_{ML}$ respectivement, peuvent être obtenus numériquement comme suit. D'abord pour θ donné , trouvez numériquement le valeur de α qui maximise $l(\alpha,\theta)$. Ensuite, en changeant θ sur l'intervalle $(0,\infty)$, trouvez les valeurs de θ et α qui maximise $L(\alpha,\theta$. Il est important de noter ici que lorsque θ change, puisque $x_m \leq \hat{\theta} \leq x_{m+1}$ la somme dans $L(\alpha,\theta$ devrait changer en conséquence. L'algorithme suivant fournit un moyen simple et direct de calculer les estimateurs du maximum de vraisemblance .

Un algorithme pour évaluer les estimateurs ML

Étape 1, Pour chaque $m(m = 1,2,..,n-1)$. calculez $\hat{\alpha_{tem}} et \hat{\theta_{tem}}$ comme suit :
pour m=1

$$\hat{\alpha}_{tem} = (\sum_{i=1}^{n} ln(x_i/x_1)^{-1} \tag{2.22}$$

$$\hat{\theta}_{tem} = x_i \prod_{i=1}^{n} (x_i/x_1)^{k^2} \tag{2.23}$$

sinon

$$\begin{aligned} \hat{\alpha}_{tem} &= \frac{k^2(n\sum_{i=1}^{m} lnx_i - m\sum_{i=1}^{n} lnx_i)}{2(m\sum_{i=1}^{\infty}(lnx_i)^2 - (\sum_{i=1}^{\infty} lnx_i)^2)} \\ &+ \frac{\sqrt{k^4(n\sum_{i=1}^{m} lnx_i - m\sum_{i=1}^{n} lnx_i)^2 + 4mnk^2(m\sum_{i=1}^{m}(lnx_i)^2 - (\sum_{i=1}^{m} lnx_i)^2)}}{2(m\sum_{i=1}^{\infty}(lnx_i)^2 - \sum_{i=1}^{\infty}(lnx_i)^2)} \end{aligned} \tag{2.24}$$

$$\hat{\theta}_{tem} = \left(\prod_{i=1}^{m} x_i \right)^{1/m} \exp\left(\frac{nk^2}{m\hat{\alpha}^{tem}} \right) \tag{2.25}$$

S'il est entre $x_m \leq \hat{\theta}_{tem} \leq x_{m+1}$, alors les estimateurs ML de α et θ sont

$$\hat{\alpha}_{ML} = \hat{\alpha}_{tem}, \hat{\theta}_{ML} = \hat{\theta}_{tem} \tag{2.26}$$

Réécrivons l'équation (2.25) réelle que

$$\frac{nk^2}{\hat{\alpha}_{tem}} + \sum_{i=1}^{m} ln(x_i/\hat{\theta}_{tem}) = 0 \tag{2.27}$$

D'après l'équation (2.26), il doit x_i avoir au moins un $\hat{\theta}_{tem}$, valeur inférieure á ma puisque n , k et $\hat{\alpha}_{tem}$ en sont des valeurs positives , Par conséquent, l'estimation ML de ne peut pas se produire a x_i.
Étape 2, S'il n'y a pas de solution pour θ (c'est-á-dire $x_n \leq \hat{\theta}_{tem}$ avec les conditions données á l'étape 1 , l'estimation ML de α et θ est

$$\hat{\alpha}_{ML} = nk/\sqrt{n\sum_{i=1}^{n}(lnx_i)^2 - \left(\sum_{i=1}^{n} lnx - i\right)^2} \tag{2.28}$$

$$\hat{\theta}_{ML} = \left(\prod_{i=1}^{n} x_i\right)^{i/n} \exp\left(\frac{k^2}{\hat{\alpha}_{ML}}\right). \tag{2.29}$$

Notez que si est plus proche de $x_1 ou x_n$, Pareto ou lognormal sera respectivement un modèle supérieur au modèle composite lognormal-Pareto. Afin de trouver les estimateurs ML, il suffi de vérifier n-1 intervalles et ces calculs peuvent être effectués même avec une simple calculatrice .
II :une procédure ad-hoc pour estimer les paramètres des données complètes Dans cette section , nous décrivons une procédure ad hoc pour estimer les paramètres du modèle lognormal-Pareto. la procédure ad hoc fournit des expressions de forme fermée pour les paramètres qui peuvent être facilement estimés . Cependant , un paramètre est estimé á l'aide de centiles et l'autre paramètre est estimé á l'aide du principe du maximum de vraisemblance. Pour décrire la procédure, soit $x_1 \leq x_2 \leq x_3 \leq ... \leq x_n$ un échantillon ordonné comme précédemment et la paramètre inconnu θ se situe entre la même observation et la $m+1^{th}$ observation , c'est-á-dire $x_m \leq \theta \leq x_{m+1}$.
En utilisant des centiles , le paramètre θ peut être estimé en utilisant l'estimation empirique lisse du centile p^{th} :

$$\hat{\theta} = (1-h)x_m + hx_{m+1} \tag{2.30}$$

où $m = [(n+1)p], h = (n+1)p - m, etp = \Phi(k)(1+\Phi(k))$.Ici [.] indique la plus grande fonction entière . Dans estimations de centiles similaires sont souvent utilisées dans les études actuarielles[1].
lorsque θ est connu (en utilisant $\bar{\theta}$pour θ) . la fonction de log vraisemblance est donnée par

$$lnL(\alpha/\theta = \bar{\theta}) = -nln(1+\Phi(k)) + nln(\alpha/\bar{\theta}) - (1+\alpha)\sum_{i=1}^{n} ln(x_i/\bar{\theta}) - \frac{\alpha^2}{2k^2}\sum_{i=1}^{m} ln^2(\alpha/\bar{\theta}) \tag{2.31}$$

En maximisant $lnL(\alpha/\theta = \bar{\theta}$ par rapport á α, l'estimation du maximum de vraisemblance de α est donné par

$$\hat{\hat{\alpha}} = \frac{\sqrt{k^4(\sum_{i=1}^{n} ln(x_i/\bar{\theta}))^2 + 4nk^2 \sum_{i=1}^{m} ln^2(x_i/\bar{\theta})} - k^2 \sum_{i=1}^{n} ln(x_i/\bar{\theta})}{2\sum_{i=1}^{m} ln^2(x_i/\bar{\theta})} \tag{2.32}$$

Un intervalle de confiance approximatif de $100(1-\alpha_0)\%$ pour le paramètres α peut être trouvé á partir des éléments suivants :

$$\hat{\hat{\alpha}} \pm Z_{\alpha_0/2}\{n + mk^2 + m\Phi(k)/(1+\Phi(k))\}^{-1/2}\hat{\hat{\alpha}} \tag{2.33}$$

III : Estimation du maximum de vraisemblance pour les donnée censurées
Dans la plupart des cas avec les paiements d'assurance, il y a une limite pour le montant maximum de paiement, c'est-á-dire les données sont censurées á droite de type II. Dans cette section, nous examinons l'estimation des paramètres du modèle de la distribution composite log-normale-Pareto pour telles données. Supposons que nous ayons $n+f$ valeurs d'échantillons et que f de ces valeurs soient censurées á u et comme précédemment, les n valeurs ordonnées non censurées restantes sont $X_1, X_2, ..., X_n$. Si le paramètre inconnu θ se situe entre lam^{th} observation et m_+1^{th} observation , la fonction de vraisemblance est donnée par

$$L(\alpha,\theta) = C_1\alpha^n\theta^{(n+f)\alpha}\left(\prod_{i=1}^{n} x_i^{-\alpha}\right)\exp\left(-\frac{\alpha^2}{2k^2}\sum_{i=1}^{m}\ln^2(x_i/\theta\right) \tag{2.34}$$

où $C_1 = 1/\left[(\prod_{i=1}^{n} x_i)(1+\Phi(k))^{n+f}\right]$
Un algorithme pour évaluer les estimateurs ML
Étape 1. pour chaque $m(m = 1, 2, ..., n-1)$ calculez $\hat{\alpha}_{tem}$ et $\hat{\theta}_{tem}$ comme suit :
Pour m=1

$$\hat{\alpha}_{tem} = n(f\ln(u/x_i) + \sum_{i=1}^{n} ln(x_i/x_1)^{-1} \tag{2.35}$$

$$\hat{\theta}_{tem} = x_1(u/x_1)^{(1+f/n)fk^2}\prod_{i=1}^{n}(x_i/x_1)^{(1+f/n)k^2} \tag{2.36}$$

Sinon

$$\begin{aligned}\hat{\alpha}_{tem} &= \frac{k^2(mf\ln u - (n+f)\sum_{i=1}^{m}\ln x_i + m\sum_{i=1}^{n}\ln x_i)}{2(m\sum_{i=1}^{m}(\ln x_i)^2 - (\sum_{i=1}^{m} lnx_i)^2)} \\ &+ \frac{\sqrt{k^4(mf\ln u - (n+f)\sum_{i=1}^{m}\ln x_i + m\sum_{i=1}^{n}\ln x_i)^2 + 4mnk^2(m\sum_{i=1}^{n}(\ln x_i)^2 - (\sum_{i=1}^{m}\ln x_i)^2)}}{2(m\sum_{i=1}^{m}(\ln x_i)^2 - (\sum_{i=1}^{m}\ln x_i)^2)}\end{aligned} \tag{2.37}$$

$$\hat{\theta}_{tem} = \left(\prod_{i=1}^{m} x_i \right)^{1/m} \exp\left(\frac{(n+f)k^2}{m\hat{\alpha}_{tem}} \right) \tag{2.38}$$

Si $\hat{\theta}_{tem}$ est entre $x_m \leq \hat{\theta}_{tem} \leq x_{m+1}$ alors les estimateurs ML de $\alpha et \theta$sont

$$\hat{\alpha}_{ML} = \hat{\alpha}_{tem}, \quad \hat{\theta}_{ML} = \hat{\theta}_{tem} \tag{2.39}$$

Réécrivons l'équation (2.39) telle que

$$\frac{(n+f)k^2}{\hat{\alpha}_{tem}} + \sum_{i+1}^{m} \ln(x_i/\hat{\theta}_{tem}) = 0 \tag{2.40}$$

2.1.2 Modèles de Weiull-Pareto

Soit

$$f_1(x) = \frac{\tau}{x}\left(\frac{x}{\Phi}\right)^{\tau} \exp\left(-\left(\frac{x}{\Phi}\right)^{\tau}\right), \quad x > 0 \tag{2.41}$$

la pdf d'une distribution de Pareto à deux paramètres, où ϕ,$\tau > 0$, et

$$f_2(x) = \frac{\alpha\theta^{\alpha}}{x^{\alpha+1}}, \quad x > \theta \tag{2.42}$$

Soit la pdf d'une distribution de Pareto a deux paramètres , où $\theta, \alpha > 0$
Scollnik Sun (2012)[7] ont construit un modèle composite avec pdf

$$f(x) = \begin{cases} r\frac{f_1(x)}{F_1(\theta)}, & 0 < x \leq \theta \\ (1-r)f_2(x), & \theta < x < \infty \end{cases} \tag{2.43}$$

avec $0 \leq r \leq 1$ et $F_1(\theta)$ est la cdf de la distribution de Weibull en θ. En permettant la continuité et dérivabilité en θ, on obtient un modèle à trois paramètres. Maintenant r est fourni par

$$r = \frac{\alpha \quad \exp(\mathcal{B}) - \alpha}{\alpha \quad exp(\mathcal{B}) + \tau} \quad et \quad \mathcal{B} = (\frac{\theta}{\phi})^{\tau} = \frac{\alpha}{\tau} + 1. \tag{2.44}$$

De la même manière que dans la famille composite log-normal-Pareto. Scollnik Sun(2012)[7] ont également considéré le composite Weibull-Type II Pareto (lomax) avec pdf

$$f(x) = \begin{cases} r\frac{f_1(x)}{F_1(\theta)}, & 0 < x \leq \theta \\ (1-r)f_2(x), & \theta < x < \infty \end{cases} \tag{2.45}$$

où $f_2(x)$ est la pdf du type II Pareto et elle est donnée par

$$f_2(x) = \frac{\alpha(\lambda+\theta)^\alpha}{(\lambda+x)^{\alpha+1}}, \quad x > \theta, \theta > 0, \alpha > 0, \quad et \quad \lambda > -\theta \tag{2.46}$$

Puis en imposant les conditions de continuité et de dérivabilité á θ, un lissage à quatre paramètres fonction de densité est dérivée, Ici, r est fourni par

$$r = \frac{\alpha}{\tau}\left(\frac{\lambda+\theta}{\theta}\frac{\mathcal{C}}{exp(\mathcal{C})-1} + \frac{\alpha}{\tau}\right)^{-1} \tag{2.47}$$

oú

$$\mathcal{C} = (\frac{\theta}{\phi})^\tau = \frac{\alpha\theta-\lambda}{(\lambda+\theta)\tau} + 1,$$

clairement, ce modèle imbrique le composite Weibull-Pareto modèle si $\lambda = 0$.

2.1.3 Modèle Pareto-Stoppa

Bien qu'elle ne soit pas suffisamment bien décrite dans la littérature anglophone, une généralisation de la distribution de Pareto a été proposée par Stoppa(1990)[10]. La méthodologie pour dériver cette famille de distributions consiste à appliquer une transformation de puissance à la cdf de Pareto. La cdf de la distribution Stoppa est donnée par

$$F(x) = \left(1 - \left(\frac{x}{x_0}\right)^{-\sigma}\right)^\gamma, \quad 0 < x_0 \leq x, \tag{2.48}$$

oú $\sigma, \gamma > 0$ spécifient la forme de la distribution et x_0 est la valeur minimale possible. La distribution de Pareto classique est obtenue lorsque $\gamma = 1$. La pdf de la distribution de Stoppa est fourni par

$$f(x) = \gamma\sigma x_0^\sigma x^{-(\sigma+1)}\left(1 - \left(\frac{x}{x_0}\right)^{-\sigma}\right)^{\gamma-1}, \quad 0 < x_0 \leq x, \tag{2.49}$$

Certaines propriétés de cette distribution peuvent être trouvées dans Kleiber Kotz(2003) [11] .Á cet égard ,le moment d'ordre K existe pour $K < \sigma$ est donné par

$$E(X^k) = \gamma x_0^k Be\left(1 - \frac{k}{\sigma}, \gamma\right) \tag{2.50}$$

où $Be(.,.)$ représente la fonction Bêta définie par

$$Be(a,b) = \int_0^1 z^{a-1}(1-z)^{b-1}dz \quad avec \quad a, b > 0 \tag{2.51}$$

De plus , la fonction quantité peut être facilement dérivée.

$$F^{-1}(u) = x_0\left(1-u^{\frac{1}{\gamma}}\right)^{-\frac{1}{\sigma}}, \quad 0<u<1 \tag{2.52}$$

Par rapport à la distribution de Pareto , la distribution de Stoppa est plus flexible car elle inclut un paramètre de forme supplémentaire γ qui permet l'unimodalité pour $\gamma > 1$ et la modalité nulle lorsque $\gamma \leq 1$. Pour le premier cas , le mode est situé à

$$x_{Mode} = x_0\left(\frac{1+\gamma\sigma}{1+\sigma}\right)^{\frac{1}{\sigma}}, \quad \gamma > 1, \tag{2.53}$$

alors que pour ce dernier cas ,c'est à x_0 .la figure 1 montre l'effet de l'augmentation du paramètre de forme γ sur la pdf de la distribution de Stoppa tout en maintenant x_0 et σ fixes .Lorsque $\gamma \leq 1$ augmente, le mode se déplace vers la droite produisant une queue plus épaisse.

2.2 Les modèles composite Stoppa

Les modèles composites de Pareto décrits utilisent la distribution de Pareto au-dessus de la valeur seuil. Comme les distributions de Pareto classiques est de type décroissant et de manière monotone, l'estimation du seuil est toujours supérieure à la valeur modale du modèle composite. Cette formulation est naturelle car l'intrention est d'utiliser la distribution de Pareto pour modéliser les grands sinistres.
Pourtant,un autre point de vue qui pourrait être utile est de considérer un modèle composite où être le point de jonction est au mode des données. Selon cette construction, l'utilisation des deux distributions pour les différents côtés de la distribution des sinistres modélise en fait à quelle vitesse la probabilité diminue du mode aux deux extrémités de la distribution.Cela résout la nature asymétrique de la distribution des réclamations résultant des tailles asymétrique de l'espace d'états des deux cotés du mode des données .En conséquence ,le support de données des deux cotés du modèle épissé est plus équilibré,par rapport,par exemple,à un modèle composite de Pareto avec une estimation large. Cela présente l'avantage que l'ajustement du modèle est plus robuste aux nouvelles données des sinistres importants et que les informations contenues dans les sinistres moyens à grands peuvent être capturées , en modélisant la transition entre les deux classes de sinistres.En utilisant une procédure d'appariement de mode ,la construction du modèle composite Stoppa est maintenant donné.Notez que cette procédure incorpore des poids de mélange sans restriction dans le modèle.Le première composante du modèle épissé est utilisée jusqu'à la valeur modale (qui doit être estimée à partir de les donnée) et la troncature alequate du fils une distribution donnée en par la suite .Ensuite , fonction de densité du modèle composite Stoppa peut

être écrite comme

$$f(x) = \begin{cases} r f_1^*(x), & 0 < x \leq x_m \\ (1-r) f_2^*(x), & x_m < x < \infty \end{cases} \tag{2.54}$$

avec $0 \leq r \leq f_1^*(x) = \frac{f_1(x)}{F_1(x_m)}$ une troncature adéquate des pdfs f_1 jusqu'à la valeur modale ,où$F_1(x_m)$ est la cdf de f_1 évaluée á x_m et $f_1^*(x) = \frac{f_2(x)}{1-F_2(x_m)}$ une troncature appropriée de la distribution de Stoppa , où $1 - F_2(x_m)$ est la fonction de survie évaluée á Again, (2.54) est sous la forme d'un modèle de mélange. A la place des condition habituelles de continuité de dérivabilité, une procédure d'appariement de mode est utilisée.Cette procédure permet de s'assurer que les conditions de continuité et de dérivabilité sont satisfaites .De plus, cela donne une dérivation plus simple du modèle composé avec toutes les distribution avec un mode qui a une expression de forme fermée .Les conditions d'adaptation de mode sont donnée comme suit.Dénotons les modes des distribution utilisées par les première et deuxième composantes du modèle composite par $x_m^{first}, x_m^{second}$ respectivement. Alors, les conditions d'appariement de mode sont :

$$x_m^{first} = x_m^{second} \tag{2.55}$$

$$r \quad f_1^*(x_m^{first}) = (1-r) f_2^*(x_m^{second}) \tag{2.56}$$

De toute évidence, (2.56) implique que la condition de continuité est satisfaite, et puisque l'égalité dans (2.55) nous permet de supprimer les étiquettes "première" et "seconde", l'expression simple du poids de mélange, comme on le voit dans les autres modèles composites existants, est conservé est donné par

$$r = \frac{f_2(x_m)F_1(x_m)}{f_2(x_m)F_1(x_m) + f_1(x_m)(1-F_2(x_m))} \tag{2.57}$$

Ensuite note que pour la distribution uni-modale , la dérivée au mode est nulle , il est donc clair que la condition de dérivabilité est également satisfaite .

2.2.1 Modèle log-normal-Stoppa

Le modèle composite log-normal-Stoppa sera dérivé en termes de modèle de mélange (2.54). Sa fonction de densité est donnée par

$$f(x) = \begin{cases} r \frac{\frac{1}{\sqrt{2\pi}x\sigma} exp(-\frac{1}{2}(\frac{\ln x - \mu}{\sigma})^2)}{\Phi(\frac{\ln x_m - \mu}{\sigma})}, & 0 < x \leq x_m \\ (1-r) \frac{\gamma\delta x_0^{\sigma} x^{-(\delta+1)}[1-(\frac{x}{x_0})^{-\delta}]^{\gamma-1}}{1-[1-(\frac{x_m}{x_0})^{-\delta}]^{\gamma}}, & x_m < x < \infty \end{cases} \tag{2.58}$$

avec$\mu \in \mathrm{R}, \sigma > 0, \gamma > 1, \delta > 0, 0 \leq r \leq 1 \quad et \quad \Phi(.)$ désignera la cdf de la distribution

normale standard. En utilisant la procédure d'appariement de mode (2.55) donne

$$\sigma = \sqrt{\mu - \ln[x_0(\frac{1+\gamma\delta}{1+\delta})^{\frac{1}{\delta}}]} \tag{2.59}$$

noter que ce résultat implique une contrainte supplémentaire qui $\mu > \ln[x_0(\frac{1+\gamma\delta}{1+\delta})^{\frac{1}{\delta}}]$

En substituant les densités et fonctions de distributions correspondantes dans (2.57) donne

$$\begin{aligned} r &= \gamma\delta x_0^{\delta} x_m^{-(\delta+1)}[1-(\frac{x_m}{x_0})^{-\delta}]^{\gamma-1}\Phi(\frac{\ln x_m-\mu}{\sigma}) * (\gamma\delta x_0^{\delta} x_m^{\delta+1}[1-(\frac{x_m}{x_0})^{-\delta}]^{\gamma-1}\Phi(\frac{\ln x_m-\mu}{\sigma}) \\ &+ \frac{1}{\sqrt{2\pi}x_m\sigma} exp(-\frac{1}{2}(\frac{\ln x_m-\mu}{\sigma})^2)(1-[1-(\frac{x_m}{x_0})^{-\gamma})^{-1} \end{aligned}$$

Il garantit que (2.58) est continu et lisse .A noter que le nombre de paramètres est réduit à quatre Le cdf de la distribution composite lognormal-Stoppa est fourni par

$$F(x) = \begin{cases} r\frac{\Phi(\frac{\ln x-\mu}{\sigma})}{\Phi(\frac{\ln x_m-\mu}{\sigma})}, & 0 < x \le x_m \\ r + (1-r)\frac{[1-(\frac{x}{x_0})^{-\delta}]^{\gamma}-[1-(\frac{x_m}{x_0})^{-\delta}]^{\gamma}}{1-[1-(\frac{x_m}{x_0})^{-\delta}]^{\gamma}}, & x_m < x < \infty \end{cases} \tag{2.60}$$

De plus ,le moment d'ordre k de la distribution composite Lognormal-Stoppa existe lorsque $\delta > k$ Son expression analytique donné par

$$E(X^k) = r\frac{\Phi(\frac{\ln x_m-\mu-k\sigma^2}{\sigma})}{\Phi(\frac{\ln x_m-\mu}{\sigma})}e^{k\mu+\frac{k^2\sigma^2}{2}} + (1-r)\frac{1}{1-[1-(\frac{x_m}{x_0})^{-\delta}]}\gamma x_0^k Be((\frac{x_m}{x_0})^{-\delta}; 1-\frac{k}{\delta}, \gamma) \tag{2.61}$$

avec $Be(.;.,.)$ ou représente la fonction bêta incomplète définie par

$$Be(x;a,b) = \int_0^x z^{a-1}(1-z)^{b-1}dz \quad avec \quad a,b>0 \tag{2.62}$$

En préparation du schéma de ré-échantillonnage dans la section des application numériques suivantes, une procédure de génération de génération de variables aléatoire à partir de la distribution composite Lognormale-Stoppa est présentée.Comme la cdf des distribution lognormal et Stoppa peut être inversée , la méthode de simulation par transformation inverse peyt être utilisée pour cette famille composite .Si est une valeur générée à partir de la distribution uniforme $u(0,1)$,alors une valeur générée à partir (2.58) est obtenue comme suit
Si $u \le r$ alors

$$x = \exp\left(\mu + \sigma.\Phi^{-1}\left(\frac{u}{r}\Phi\left(\frac{\ln x_m-\mu}{\sigma}\right)\right)\right)$$

Si $u > r$ alors

$$x = x_0\left(1-\left(\frac{u-r}{1-r}\left[1-\left(1-(\frac{x_m}{x_0})^{-\delta}\right)^{\gamma}\right]+\left(1-(\frac{x_m}{x_0})^{-\delta}\right)^{\gamma}\right)^{1/\gamma}\right)^{-1/\delta}$$

2.2.2 Modèle Weibull-Stoppa

Le modèle composite de Weibull-Stoppa sera également obtenu en termes de modèles de mélange(2.54). Sa fonction de densité est donné par

$$f(x) = \begin{cases} r\frac{1}{1-\exp(-(\frac{x_m}{\Phi})^{\tau})}(\frac{\tau}{x})(\frac{x}{\phi})^{\tau}\exp(-(\frac{x}{\phi})^{\tau}). & 0 < x \leq x_m \\ (1-r)\frac{\gamma\delta x_0^{\delta}x^{-(\delta+1)}[1-(\frac{x}{x_0})^{-\delta}]^{\gamma-1}}{1-[1-(\frac{x_m}{x_0})^{-\delta}]^{\gamma}}, & x_m < x < \infty \end{cases} \tag{2.63}$$

avec $\Phi > 0, \gamma > 1, \delta > 0, 0 \leq r \leq 1$ et $\tau > 1$ pour définir une valeur modale positive.Maintenant,n appliquant à nouveau les conditions d'adaptation de mode. (2.25) donne

$$\phi = \left[x_0\left(\frac{1+\gamma\tau}{1+\tau}\right)^{1/\tau}\right]\left(\frac{\tau}{\tau-1}\right)^{1/\tau} \tag{2.64}$$

De même la subsituation des densités et des fonctions de distribution corespondantes dans (2.25) donne

$$\begin{aligned} r &= \gamma\delta x_0^{\delta}x_m^{-(\delta+1)}\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma-1}\left(1-\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)\right) \\ &\times \left(\gamma\delta x_0^{\delta}x_m^{-(\delta+1)}\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma-1}\left(1-\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)\right)\right) \\ &+ \left(\frac{\tau}{x_m}\right)\left(\frac{x_m}{\phi}\right)^{\tau}\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)\left(\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}\right)^{-1} \end{aligned} \tag{2.66}$$

La cdf de la distribution composite de Weibull-Stoppa est obtenu par

$$F(x) = \begin{cases} r\frac{1-\exp\left(-\left(\frac{x}{\phi}\right)\tau\right)}{1-\exp\left(-\left(\frac{x_m}{\phi}\right)^{\tau}\right)}, & 0 < x \leq x_m \\ r+(1-r)\frac{\left[1-\left(\frac{x}{x_0}\right)^{-\delta}\right]^{\gamma}-\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}}{1-\left[1-\left(\frac{x_m}{x_0}\right)^{-\delta}\right]^{\gamma}}, & x_m < x < \infty \end{cases} \tag{2.67}$$

Maintenant, le moment d'ordre de la distribution composite de Weibull-Stoppa existe à nouveau si $\delta > k$. Son expression analytique est donnée par

$$\begin{aligned} E(X^k) &= r\frac{\phi^k\Gamma(1+\frac{k}{\tau}0-\Gamma(1+\frac{k}{\tau};(\frac{x_m}{\Phi})^\tau)}{(1-\exp(-(\frac{x_m}{\Phi})^\tau))} \\ &+ (1-r)\frac{1}{1-[1-(\frac{x_m}{x_0})^\delta]^\gamma}{}_k{}_0 Be\left((\frac{x_m}{x_0})^{-\delta};1-\frac{k}{\delta},\gamma\right). \end{aligned}$$

Ou $\Gamma(.)$ et $\Gamma(.;.)$ sont les fonctions gamma complètes et incomplètes définies par

$$\begin{aligned} \Gamma(a) &= \int_0^\infty z^{a-1}e^{-z}dz \\ \Gamma(a;x) &= \int_0^x z^{a-1}e^{-z}dz \quad avec \quad a,x>0 \end{aligned}$$

respectivement et $Be(.;,.)$ est la fonction bêta incomplète. La procédure de génération de variables aléatoires à partir de la distribution de Weibull-Stoppa est également présentée.Comme dans la section précédente,la méthode de simulation par transformation inverse peut être appliquée car les cdf des distributions de Weiull-Stoppa sont inversibles.Si est une valeur générée à partir de la distribution uniforme $U(0,1)$,alors une valeur générée à partir (2.63) oeut être obtenue comme suit
si $u \leq r$ alors

$$x = -\phi\left(\ln\left[1-\frac{u}{r}\left(1-\exp\left[-\left(\frac{\theta}{\phi}\right)^\tau\right]\right)\right]\right)^{1/\tau}$$

si $u > r$ alors

$$x = x_0\left(1-\left(\frac{u-r}{1-r}\left[1-\left(1-(\frac{x_m}{x_0})^{-\delta}\right)^\gamma\right]+\left(1-(\frac{x_m}{x_0})^{-\delta}\right)^\gamma\right)^{1/\gamma}\right)^{-1/\delta}$$

CHAPITRE 3

SIMULATIONS ET APPLICATIONS SUR DES DONNÉES RÉELLES

Introduction

Les simulations et les applications sur des données réelles sont des outils puissants utilisés dans de nombreux domaines pour étudier, analyser et prédire des phénomènes complexes. Ces méthodes permettent de reproduire virtuellement des situations réelles en utilisant des modèles mathématiques et des données empiriques.

Les simulations sur des données réelles impliquent la création d'un modèle informatique qui représente un système ou un processus réel. Ce modèle est alimenté en données réelles collectées à partir d'observations ou d'expérimentations. En utilisant ces données, la simulation permet de comprendre et de prévoir le comportement du système dans différentes conditions.

Les applications sur des données réelles vont au-delà de la simple simulation en utilisant des techniques d'analyse de données pour extraire des informations utiles et prendre des décisions éclairées. Ces applications peuvent être utilisées dans de nombreux domaines, tels que la recherche scientifique, l'ingénierie, la finance, la santé, la météorologie, l'urbanisme, etc.

L'avantage des simulations et des applications sur des données réelles est qu'elles permettent d'explorer des scénarios virtuels sans avoir à expérimenter directement sur le système réel, ce qui peut être coûteux, risqué ou difficile à mettre en œuvre. De plus, elles offrent la possibilité de tester différentes hypothèses, de prendre en compte des facteurs multiples et de quantifier les incertitudes associées aux résultats.

3.1 Simulation des modèles composé :

3.1.1 Simulation de la densité de la loi composé Weibull-Gamma

La loi composée Weibull-Gamma est une distribution statistique qui combine une distribution de Weibull avec une distribution de Gamma. La densité de cette loi peut être obtenue en utilisant des techniques de convolution. Cependant, la formulation exacte de la densité dépend des paramètres spécifiques de la loi composée Weibull-Gamma.

La distribution de Weibull est souvent utilisée pour modéliser des durées de vie, tandis que la distribution de Gamma est utilisée pour modéliser des décomptes ou des occurrences d'événements. En combinant ces deux distributions, on peut modéliser des données qui représentent des durées de vie avec des occurrences d'événements.

La densité de la loi composée Weibull-Gamma peut être exprimée mathématiquement comme suit :

$$f(x) = \int_0^{+\infty} f_{weibull}(x|\lambda, k) f_{gamma}(t|\alpha, \beta) dt$$

où $f_{Weibull}$ est la densité de la distribution de Weibull et f_{gamma} est la densité de la distribution de Gamma. Les paramètres λ et k sont spécifiques à la distribution de Weibull, tandis que les paramètres α et β sont spécifiques à la distribution de Gamma.

La forme exacte de la densité dépendra des valeurs spécifiques de ces paramètres. Il n'existe pas de formule générale unique pour la densité de la loi composée Weibull-Gamma. Afin de calculer la densité pour des valeurs spécifiques des paramètres, il est généralement nécessaire d'utiliser des méthodes numériques ou des logiciels de statistiques qui prennent en charge cette distribution.

Si vous avez des valeurs spécifiques pour les paramètres de la distribution de Weibull-Gamma, je peux vous aider à calculer la densité à l'aide d'un logiciel de statistiques ou de techniques numériques appropriées

3.1 La loi composée Weibull-Gamma est une distribution de probabilité qui combine les distributions Weibull et Gamma. La densité de cette distribution dépend des paramètres α, σ, β et λ. Cependant, la forme exacte de la densité dépendra également des conventions spécifiques utilisées pour définir la loi composée Weibull-Gamma.

3.1.2 Simulation de la densité de log-normale-loglogistique

Les courbes composites de densité log-normal log-logistique sont des courbes qui combinent les caractéristiques des distributions log-normal et log-logistique. Ces courbes sont utilisées pour modéliser des données qui ont des comportements similaires à la fois à la distribution log-normal et à la distribution log-logistique.

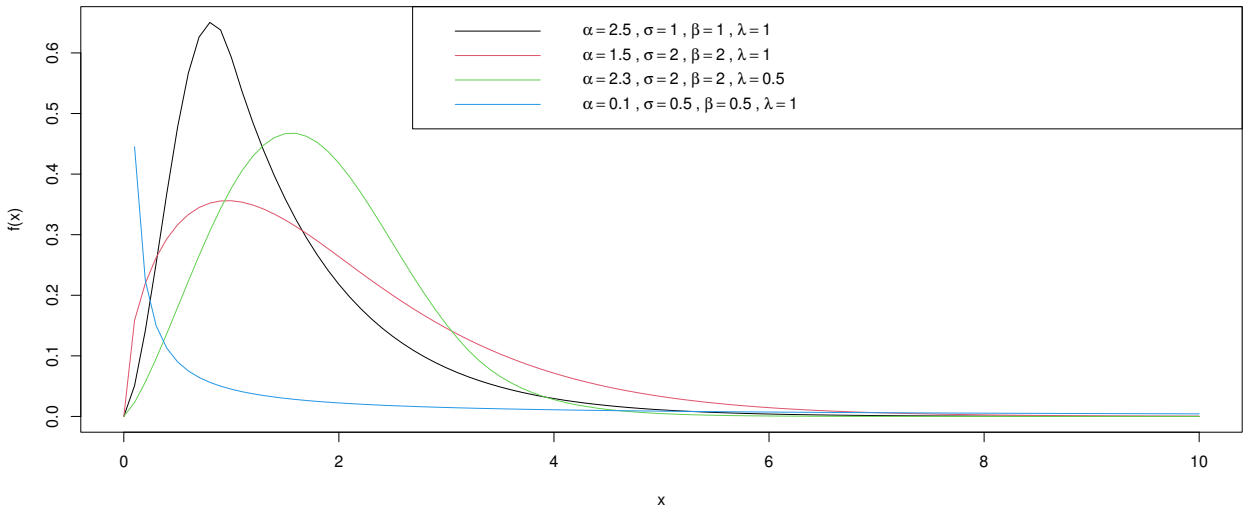


FIGURE 3.1 – Figure de la densité de la loi composé Weibull-Gamma pour différentes valeurs de α, σ, β et λ.

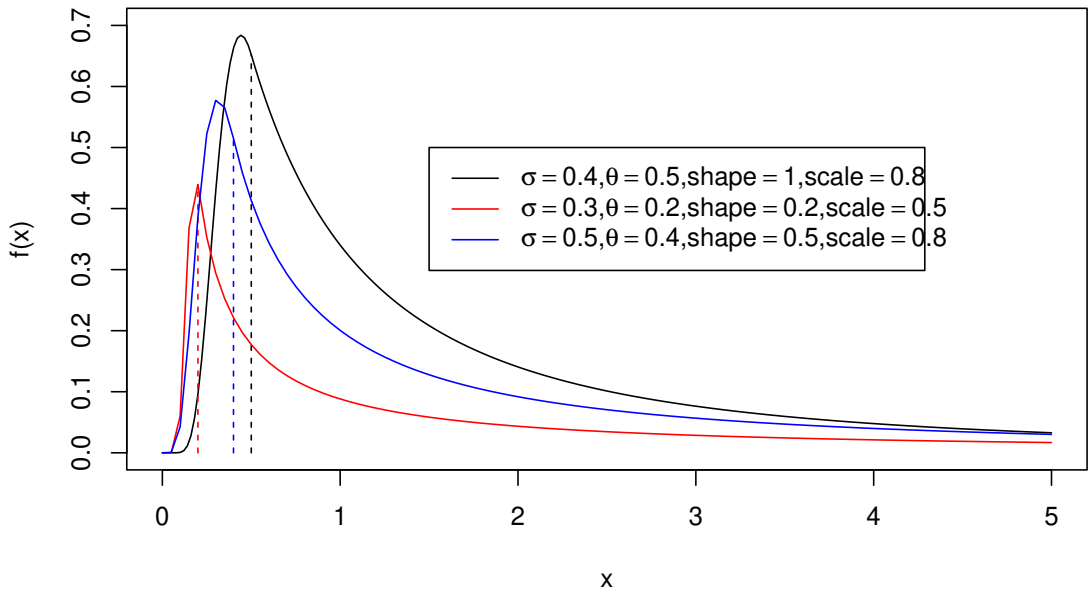
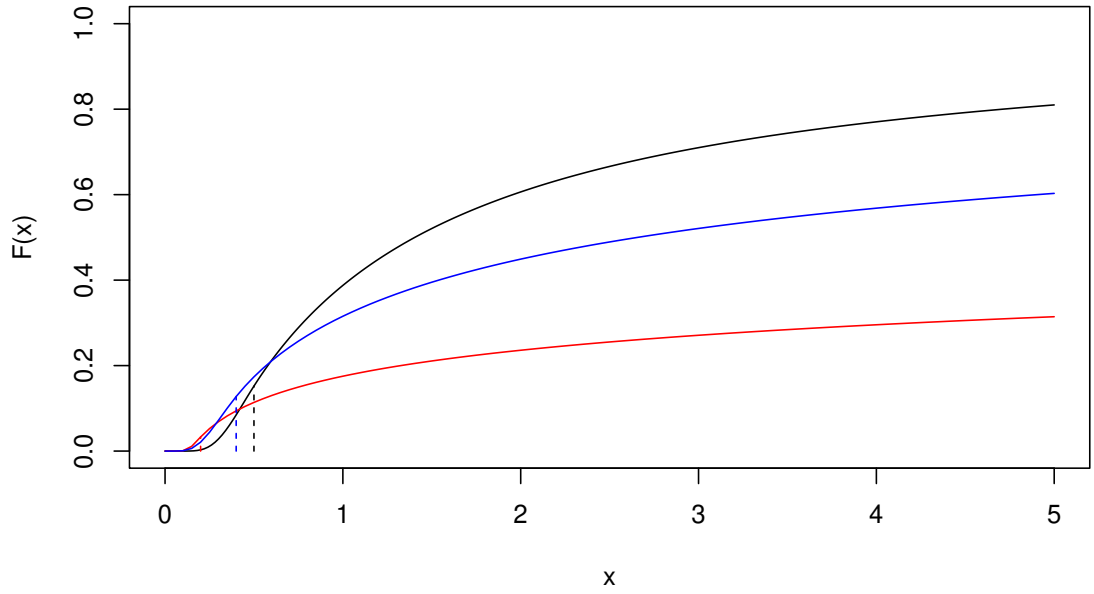


FIGURE 3.2 – Les courbes composites de densité log-normal log-logitique



C

La distribution log-normal est souvent utilisée pour modéliser des données positives qui ont une distribution asymétrique. Elle est caractérisée par des valeurs qui suivent une distribution log-normale, c'est-à-dire que le logarithme des valeurs suit une distribution normale.

D'autre part, la distribution log-logistique est également utilisée pour modéliser des données positives, mais elle est plus flexible en termes de forme. Elle peut représenter à la fois des queues épaisses (taux de décroissance plus lent) et des queues minces (taux de décroissance plus rapide) par rapport à la distribution log-normal.

En combinant ces deux distributions, les courbes composites de densité log-normal log-logistique peuvent capturer une plus grande variété de formes de données et offrir une meilleure adaptation aux données observées.

3.2 Application sur les donnes réels (Danish fire dataset)

(Analyse des données danoises sur les sinistres de l'assurance incendie) Dans cet exemple, nous analysons ensemble complet de données danoises, qui consiste en 2492 pertes d'assurance incendie en danois Couronne (DKK) des années 1980 à 1990 inclus. Le chiffre de perte est un chiffre de perte totale pour les événements concernés et comprend les dommages aux bâtiments, au mobilier et aux personnes propriété ainsi qu'un manque à gagner. Les données enregistrées ont été convenablement ajustées pour refléter valeurs de 1985. Les valeurs des pertes ajustées en couronnes danoises vont de

(en millions) 0,3134041 au 263.2503660. McNeil [7] a analysé la partie supérieure de ces données, qui consiste en 2156 pertes supérieures à un million de couronnes danoises, comme exemple de l'utilisation de la théorie des valeurs extrêmes par estimer les queues des distributions de gravité des pertes. Pour la partie supérieure des données, il utilisé le modèle de Pareto décalé à deux paramètres comme modèle paramétrique et a conclu que le Le modèle de Pareto décalé à deux paramètres est un modèle utile pour estimer les queues de perte distributions de gravité. Resnick [5] a analysé l'ensemble complet des données danoises pour démontrer plusieurs des techniques statistiques alternatives et des dispositifs de traçage qui peuvent être utilisés pour évaluer la

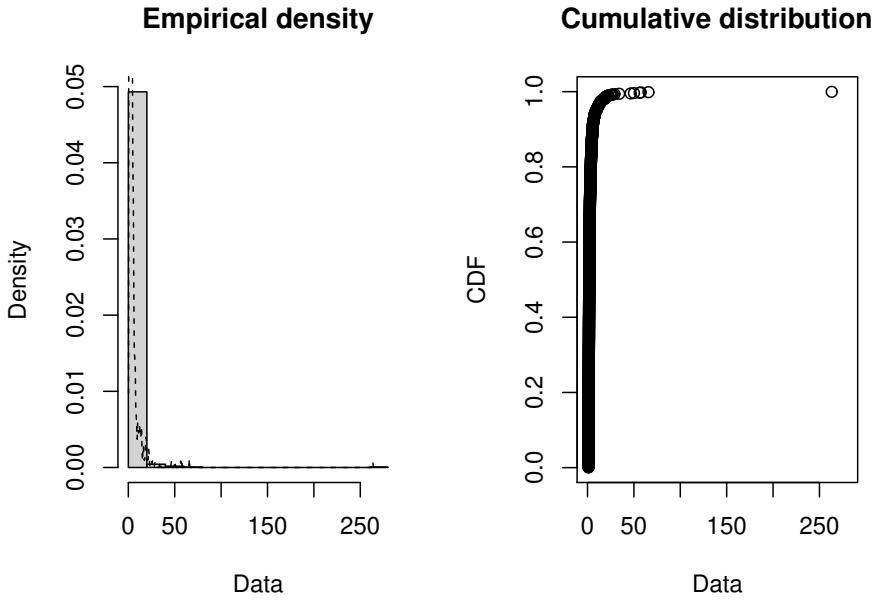


FIGURE 3.3 – Graphe des données danish fire

Les résultats d'ajustement des données réelles (danish fire) sont présentées dans le tableau 3.1.

TABLE 3.1 – Valeurs estimées des modèles ajustés pour les données danoises sur les sinistres de l'assurance incendie.

Distributions	Loglik	AIC	K-S	C M S	A D S
Lognormal	-4893.944	9791.887	0.4367645	88.9503140	416.2567545
Pateto	-4622.833	9249.666	0.312342	37.711019	208.291065
Gamma	-4767.096	9538.191	0.2018827	37.0636521	inf
Weibull	-4803.621	9611.243	0.2732043	36.2608757	inf
loglogistique	-3859.657	7727.314	0.01908345	0.18375167	1.81546259
lognrmlfrechet	-3859.293	7726.586	0.01908854	0.17271519	1.77348924
lognrmlpareto	-3865.864	7739.728	0.0322770	0.4774124	3.1565166
lognrmlweibull	-3859.657	7727.314	0.5158309	308.0674026	inf

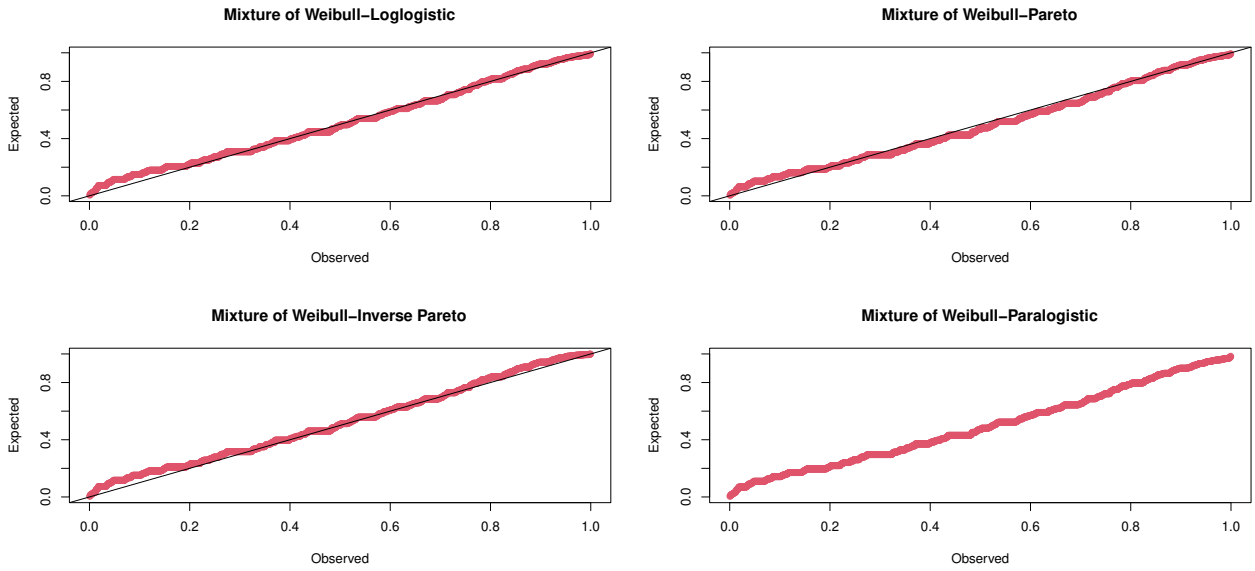


FIGURE 3.4 – pp-plot des données observées et des lois composées

À partir du tableau donné, nous extrapolons une meilleure distribution, qui est la distribution lognormale-Fréchet, basée sur les critères suivants :

Loglik (-3859.293)

AIC (7726.586)

K-S (0.01908854)

C M S (0.17271519)

A D S (1.77348924)

3.3 Application sur des données de log-rendement des indice boursiers

Un ensemble de données contenant les log-rendements des cours de clôture ajustés du 04.01.2007 au 30.10.2017. L'ensemble de données contient des données de Microsoft, SAP, Adidas, SP 500 (indice) et Dow Jones Industrial Average (indice).
Dans ce travail, on s'intéresse par l'indice SAP. SAP SE est une société européenne, dont le siège se trouve en Allemagne, qui conçoit et vend des logiciels, notamment des systèmes de gestion et de maintenance, principalement à destination des entreprises et des institutions dans le monde entier. Son produit le plus connu est le progiciel de gestion intégré SAP ERP. La base de donnée de l'indice SAP sont disponible dans le package **mistr** du softwre R.
La description de l'ensemble des données est présenter dans la figure de la boite à moustaches suivantes : La figure 3.7 indiquent que le modèle le plus adéquate est le

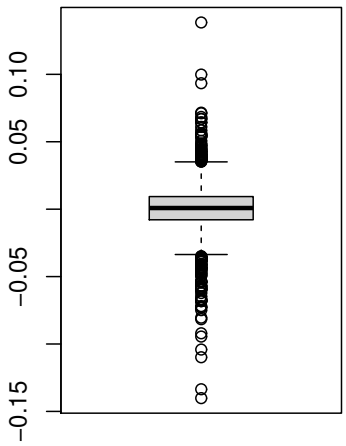


FIGURE 3.5 – Résumé de la série des données de l'indice SAP

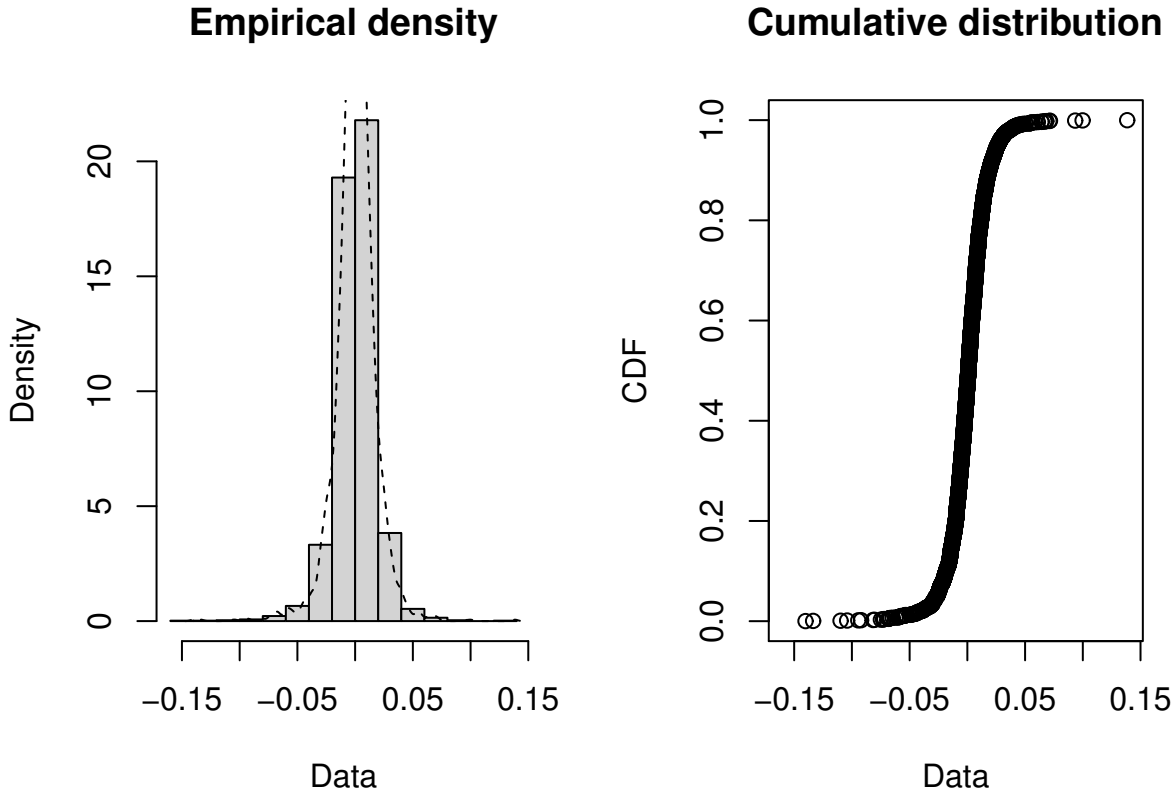


FIGURE 3.6 – Histograme, polygône et CDF empirique d'indice SAP

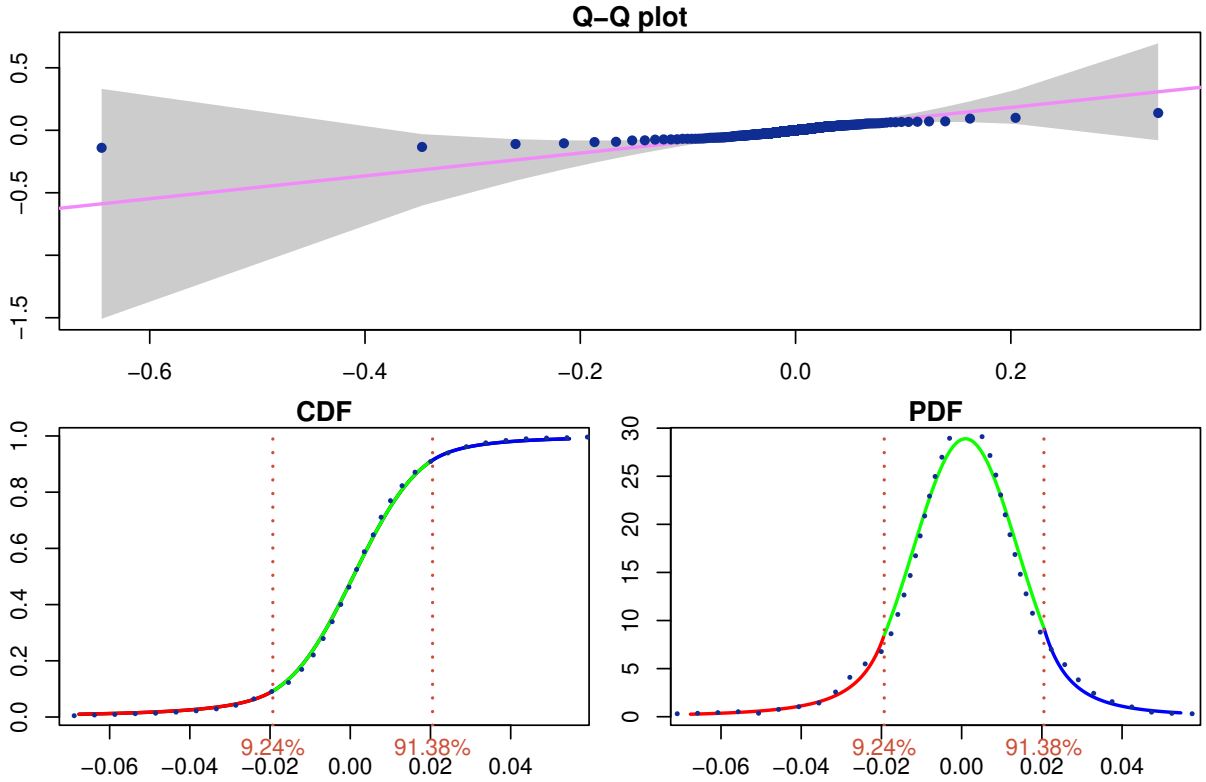


FIGURE 3.7 – Ajustement par PNP du données stocks market

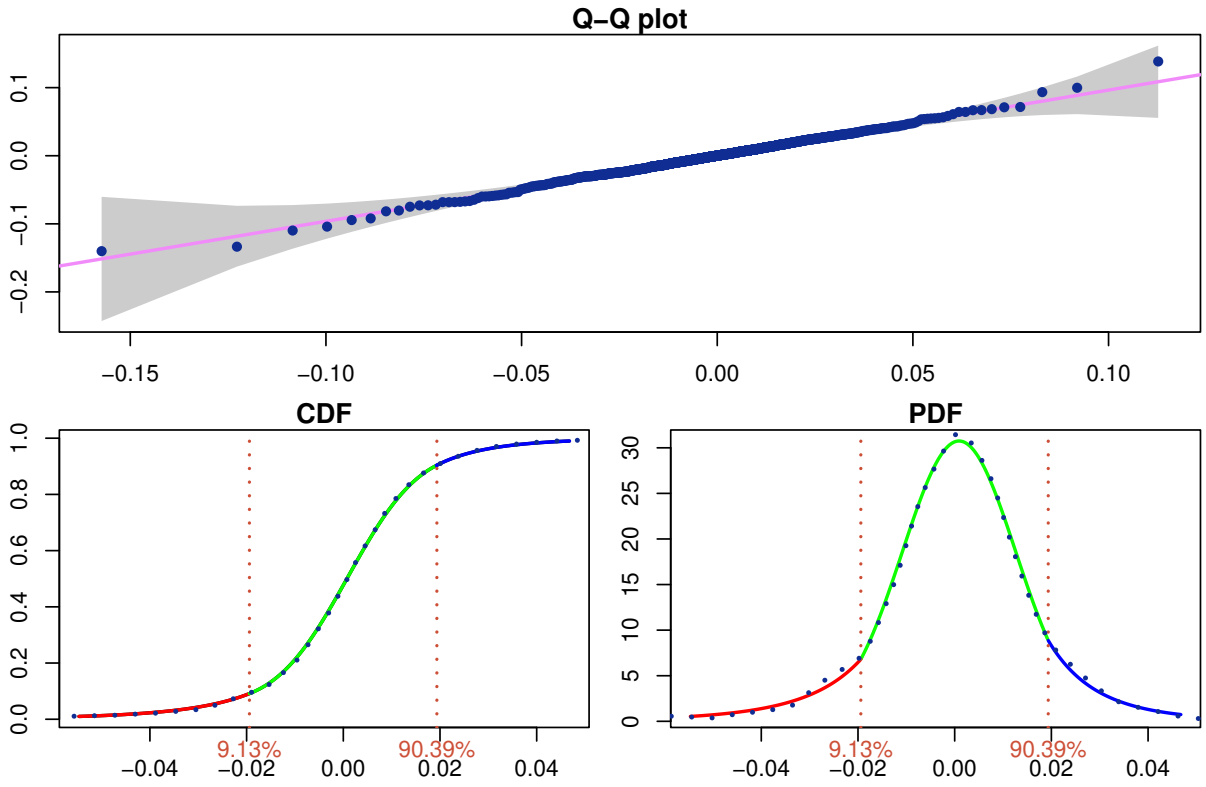


FIGURE 3.8 – Ajustement par GPD-N-GPD du données de stocks market

modèle Pareto-Normal-Pareto (PNP) car tous les quantiles sont dans le confiance liée. Le deuxième modèle proposé est une distribution similaire au précédent, sauf que nous allons remplacer les distributions de Pareto

Parameters	loc1	scale1	shape1	mean	sd	loc2	scale2	shape2
PNP		0.0193	1.7736	0.0009	0.0121		0.0205	2.1981
GPD-N-GPD	0.0194	0.0134	0.1501	0.0009	0.0117	0.0194	0.0108	0.0968

TABLE 3.2 – Paramètres des modèles composées

Modèle	P-N-P			GPD-N-GPD		
Breakpoints	-0.0193	0.0205		-0.0194	0.0194	
Weights	0.0924	0.821	0.0862	0.0913	0.8126	0.0961

TABLE 3.3 – Points de ruptures et poids de chaque modèle composé

CONCLUSION

La caractérisation des modèles composés est une approche utilisée en actuariat pour analyser et modéliser des données complexes. Elle consiste à décomposer les données en différentes composantes afin de comprendre les facteurs qui influencent les phénomènes actuariels tels que les sinistres d'assurance ou la mortalité.

Les modèles composés sont appliqués dans divers domaines de l'actuariat, tels que l'assurance automobile et vie. Ils permettent de modéliser les sinistres en fonction de variables telles que l'âge, le sexe et l'historique des sinistres, ou d'estimer la mortalité en se basant sur des facteurs tels que l'âge, le sexe et les antécédents médicaux.

Ces modèles ont de nombreuses applications pratiques, notamment la tarification des produits d'assurance, l'estimation des réserves techniques, la gestion des risques et la planification financière. Ils aident les actuaires à évaluer et à gérer les risques associés aux assurances, en fournissant des informations précieuses pour la prise de décisions stratégiques.

En résumé, la caractérisation des modèles composés est une approche clé en actuariat pour l'analyse des données. Ces modèles permettent de mieux comprendre les relations entre les différentes composantes des phénomènes actuariels et offrent des applications pratiques dans la tarification, la gestion des risques et la planification financière.

Bibliographie

[1] SA Abu Bakar, Nor A Hamzah, Mastoureh Maghsoudi, and Saralees Nadarajah. Modeling loss data using composite models. *Insurance : Mathematics and Economics*, 61 :146–154, 2015.

[2] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. draft manuscript, 2005.

[3] Sidney I Resnick. *Heavy-tail phenomena : probabilistic and statistical modeling*. Springer Science Business Media, 2007.

[4] Kahadawala Cooray and Malwane MA Ananda. Modeling actuarial data with a composite lognormal-pareto model. *Scandinavian Actuarial Journal*, 2005(5) :321–334, 2005.

[5] Vasile Preda, Roxana Ciumara, et al. On composite models : Weibull-pareto and lognormal-pareto. a comparative study. *Romanian Journal of Economic Forecasting*, 3(2) :32–46, 2006.

[6] David PM Scollnik. On composite lognormal-pareto models. *Scandinavian Actuarial Journal*, 2007(1) :20–33, 2007.

[7] David PM Scollnik and Chenchen Sun. Modeling with weibull-pareto models. *North American Actuarial Journal*, 16(2) :260–272, 2012.

[8] Saralees Nadarajah and SA Abu Bakar. New composite models for the danish fire insurance data. *Scandinavian Actuarial Journal*, 2014(2) :180–187, 2014.

[9] Kahadawala Cooray and Chin-I Cheng. Bayesian estimators of the lognormal–pareto composite distribution. *Scandinavian Actuarial Journal*, 2015(6) :500–515, 2015.

[10] Gabriele Stoppa. A new model for income size distributions. In *Income and Wealth Distribution, Inequality and Poverty : Proceedings of the Second International Conference on Income Distribution by Size : Generation, Distribution, Measurement and Applications, Held at the University of Pavia, Italy, September 28–30, 1989*, pages 33–41. Springer, 1990.

[11] Christian Kleiber and Samuel Kotz. *Statistical size distributions in economics and actuarial sciences*. John Wiley Sons, 2003.