



جامعة باجي مختار - عنابة

BADJI MOKHTAR - ANNABA UNIVERSITY
UNIVERSITÉ BADJI MOKHTAR - ANNABA

كلية العلوم
FACULTY OF SCIENCES
FACULTE DES SCIENCES



قسم الرياضيات
DEPARTMENT OF MATHEMATICS
DEPARTEMENT DE MATHEMATIQUES

THESE

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

La nouvelle distribution de Poisson-gamma Lindley : Simulation et Applications

Par:

SEGHIER Fatma Zohra

Sous la direction de

ZEGHDOUDI Halim

Prof.

U.B.M. Annaba

Devant le jury

PRESIDENT :

Benchettah Azzedine

Prof.

U.B.M. Annaba

EXAMINATEUR :

Benchaabane Abbas

MCA

U. Guelma

EXAMINATEUR :

Ezzebsa Abdelali

MCA

U. Guelma

EXAMINATEUR :

Yahia Djabrane

Prof.

U. Biskra

Année : 2020

Table des matières

Table des figures	4
1 Généralités et quelques notions probabiliste	14
1.1 variables aléatoires discrètes	15
1.2 Fonction de répartition	15
1.3 Moments d'une variable aléatoire discrète	16
1.3.1 Moment ordinaire	16
1.3.2 Moment factoriel	17
1.3.3 Moment centré	18
1.4 Fonction génératrice des moments	20
1.5 Distribution d'une durée de survie	21
1.5.1 Fonction de survie $S(t)$	23
1.5.2 Fonction de risque $h(t)$ (taux de hasard)	23
1.5.3 Le concept de censure	23
1.6 Estimation	24
1.6.1 Définition d'un estimateur	25
1.6.2 Propriétés d'un estimateur	25
1.6.3 Normalité asymptotique	28
1.7 Méthodes de construction d'un estimateur	29
1.7.1 Méthode du maximum de vraisemblance	29
1.7.2 Méthode des Moments	30
1.8 Quelques lois de probabilités discrètes usuelles	31
1.8.1 Loi binômiale	31
1.8.2 Loi géométrique ou de Pascal	32
1.8.3 Loi de Poisson	33
1.8.4 Loi binômiale négative	34
1.9 Distributions de Size biased	35
1.9.1 Quelques Size-biased distributions discrètes	36

2	Quelques nouvelles distribution discrètes et ses applications	39
2.1	Distribution Poisson Lindley	40
2.1.1	Moments et Mesures Connexes	41
2.1.2	Fonction génératrice	43
2.1.3	Estimation	44
2.1.4	Simulation	47
2.2	Distribution de Size Baised poisson Lindley	49
2.2.1	Moments et mesures connexes	51
2.2.2	Fonction génératrice	54
2.2.3	Estimation du paramètre	54
2.2.4	Simulation	58
2.3	Distribution de Poisson Pseudo Lindley	60
2.3.1	Moments et Mesures connexes	61
2.3.2	Fonction génératrice	63
2.3.3	Estimation	63
2.3.4	Simulations	65
2.4	Distribution Poisson Quasi Lindley	67
2.4.1	Moments et Mesures connexes	68
2.4.2	Fonction génératrice	70
2.4.3	Fonction de survie et Fonction de hasard	71
2.4.4	Estimations	71
2.4.5	Simulation	73
3	Distribution size biased Poisson-Gamma Lindley et ses applications	75
3.1	Distribution Poisson Gamma Lindley	76
3.1.1	Fonction de survie et Fonction de hasard	77
3.1.2	Moments et mesures connexes	78
3.1.3	Estimations	78
3.2	Distribution de Size-Biased de Poisson-Gamma Lindley	80
3.2.1	Moments et mesures connexes	83
3.2.2	Propriétés statistiques de SBPGaLD	86
3.2.3	Estimation du paramètre	87
3.2.4	Simulation	90
3.3	Application aux ensembles de données réelles	92
4	Nouvelle distribution tronquée à zéro et ses applications	96
4.1	Introduction	96
4.2	Modèles zero tronquées discrètes	98
4.2.1	Distribution zero tronquée Poisson	99
4.2.2	Distribution zéro tronquée Poisson Lindley	101
4.3	Distribution Poisson Quasi-Lindley zéro tronquée	103
4.3.1	Moments et mesures connexes	106

4.3.2	Propriétés statistiques de ZTPQLD	109
4.3.3	Estimation du paramètre	110
4.4	Application aux ensembles de données réelles	113
Bibliographie		117

Table des Figures

Figure 3.1.	La représentation graphique de la distribution de Size biased Poisson Gamma Lindley pour différentes valeurs de θ et β .	Page 81
Figure 3.2.	Comparaison des différents modèles OBS, PD, PQLD, PPsLD et SBPGaLD.	Page 94
Figure 4.1.	La représentation graphique de la distribution zéro tronquée Poisson Quasi Lindley pour différentes valeurs de θ et β .	Page 105
Figure 4.2.	Comparaison des différents modèles OBS, ZTPD, ZTPLD et ZTPQLD.	Page 115

Remerciment

*En premier lieu et avant tout, je remercie énormément **ALLAH**, qui m'a aidé à réaliser ce travail.*

*Je suis très heureuse d'exprimer ici mes respectueux remerciements et ma profonde gratitude à mon directeur de thèse **H. ZEGHDOUDI**, Professeur à l'université d'Annaba pour le sujet qu'il m'a proposé aussi pour ses précieux conseils, ses encouragements, ses suggestions valeureuses, sa patience et sa bonne humeur, sa grande érudition, disponibilité et son sens de la recherche qui ont été déterminants pour la réalisation de ce travail.*

*Mes respectueux remerciements vont à monsieur **A. BENCHETTAH** Professeur à l'université d'Annaba de me faire l'honneur de présider le jury.*

*Je remercie vivement Monsieur **Y. Djabrane**, Professeur à l'université de Biskra, également je remercie Monsieur **A. Benchabene**, Maître de conférences à l'université de Guelma et je remercie Monsieur **A. Ezzebsa**, Maître de conférences à l'université de Guelma pour l'honneur d'avoir accepté de faire partie du jury.*

Mes vifs remerciements vont également à tous les enseignants du département de mathématique.

*Du fond du cœur, un grand merci à mon marie **Brahim** pour ses encouragements et pour son aide très précieuse.*

Merci à toute ma famille.

Abstract

This thesis proposes two new distributions, called as **A Size Baised Piosson Gamma Lindley** (SBPGaL) and **zero truncated Poisson Quasi-Lindley**. Various statistical properties, moment method, maximum likelihood estimation, mean deviations and limiting distribution of extreme order statistics is established. An application of the model to a real data set is presented finally and compared with the fit attained by some other well-known one and two parameters distributions.

Keywords : Size biased distributions, Poisson Gamma Lindley distribution, Poisson Lindley distribution, Zero-Truncated Poisson Quasi-Lindley distribution.

Résumé

Dans cette thèse, nous nous introduisons deux nouvelles distributions nommées **Size Baised Poisson Gamma Lindley (SBPGaL)** et **Poisson Quasi-Lindley zéro tronquée**. Diverses propriétés statistiques ont été données à savoir : la méthode du moment, l'estimation du maximum de vraisemblance. Une simulation de biais et l'erreur quadratique moyenne des estimateurs sont obtenus par la méthode du maximum de vraisemblance. Enfin, une étude comparative entre les distributions à un et deux paramètres a été discutée.

Mots-clés : distributions Size biased, Distribution de Poisson Gamma Lindley, Distribution de Poisson Lindley, Poisson Quasi-Lindley zéro tronquée.

ملخص

في هذه الأطروحة نقدم توزيعين متقطعين جديدين يسميان توزيع

(SBPGaL) Size biased Poisson Gamma Lindly

وتوزيع

Poisson Quasi-Lindley zéro tronquée

يعتمد هذان التوزيعان على وسيطين، حيث يتم تحديد الخواص الإحصائية المختلفة مثل التقدير بطريقة العزوم، والتقدير بطريقة المعقولية العظمى. أكثر من هذا أجريت دراسة محاكاة وتحليل بيانات حقيقية. بالإضافة إلى ذلك، هناك تطبيق للنماذج الجديدة على مجموعة بيانات حقيقية ومقارنتها مع نماذج أخرى متقطعة ذات وسيط ووسيطين.

الكلمات المفتاحية: توزيع Size biased ، توزيع Poisson Gamma Lindly ، توزيع Poisson Quasi-Lindley ، توزيع Poisson Lindly ، توزيع Poisson Quasi-Lindley zéro tronquée.

Introduction

De nombreux domaines de la connaissance pratique s'appuient sur l'étude de collections homogènes d'objets ou de personnes. La statistique est un ensemble de méthodes permettant de dégager les caractéristiques ou la répartition de ces objets en fonction de critères d'étude déterminés.

L'analyse de survie est usuellement définie comme l'étude du délai de la survenue au cours du temps d'un événement d'intérêt, comme une panne de machine dans le domaine de la fiabilité, ou un décès, une rechute ou une rémission dans le domaine médical. Dans le cadre de cette thèse, nous considérons que l'évènement d'intérêt est le décès. La caractéristique principale des données de survie est la présence de « données incomplètes » lorsque l'évènement d'intérêt n'a pas encore été observé à la fin de l'analyse. Ces données sont dites censurées.

Soit X une variable aléatoire suivant la distribution à un paramètre avec la fonction de densité

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

Introduit par Lindley (1958). Cette distribution a suscité l'intérêt de nombreux chercheurs et a été généralisée plusieurs fois par divers auteurs.

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

alternatives de cette distribution.

En premier lieu, Sankaran (1970) utilisé (1) lorsque le paramètre suit une loi de Poisson pour dériver leur distribution de Poisson Lindley (PLD). Par ailleurs, certains chercheurs ont proposé de nouvelles classes de distributions basées sur des modifications de la distribution de Lindley, y compris leurs propriétés. L'idée principale est toujours dirigée en intégrant les anciennes distributions à des structures plus flexibles voir (Ghitany et al. (2008, a), Mahmoudi and Zakerzadeh (2010), Dolati (2010), Asgharzadeh et al. (2013).

Récemment, Zeghdoudi et al. (2017,2018,...) ont introduit des nouvelles distributions de durée de vie en utilisant les modèles de mélange et en combinant la distribution de Poisson et autres distributions à deux paramètres (Pseudo Lindley, Quasi Lindley et Gamma Lindley) qui apportera un plus à la littérature existant sur la modélisation des données de survie, des sciences biologiques et des sciences actuarielles.

Ainsi, ils ont réussi à trouver des nouvelles distributions discrètes nommées respectivement les distributions Poisson Pseudo Lindley, Poisson Quasi Lindley et Poisson Gamma Lindley.

Pour augmenter la flexibilité à des fins de modélisation, il sera utile d'envisager d'autres alternatives de ces distributions.

Les distributions *Size Biased*, constituent un cas particulier de la forme plus générale connue sous le nom de distributions pondérées. Introduites pour la première fois par Fisher (1934) pour modéliser la détermination biais, les distributions pondérées

ont ensuite été formalisées dans une théorie unificatrice par Rao (1965). De telles distributions apparaissent naturellement dans la pratique lorsque les observations d'un échantillon sont enregistrées avec une probabilité inégale, telle que la probabilité proportionnelle à dessins de taille (PPT). En bref, si la variable aléatoire X a la distribution $f(x; \theta)$, avec paramètres inconnus θ , la distribution pondérée correspondante est de la forme

$$f^\alpha(x, \theta) = \frac{w(x)f(x; \theta)}{\mathbb{E}(w(X))}$$

où $w(x)$ est une fonction de pondération non négative telle que $\mathbb{E}(w(X))$ existe. Un cas particulier des distributions pondérées, les distributions biaisées par la taille est proposé par Rao (1965) lorsque la fonction pondérée est de la forme $w(x) = x^\beta$, appelée distribution d'ordre biaisé par la taille β , où $\beta = 1$ ou $\beta = 2$, appelé longueur respectivement à base de zones et à zones biaisées. Par conséquent, la densité de la distribution biaisée en longueur est définie par

$$f_L = \frac{x f(x, \theta)}{\mathbb{E}(X)}, \quad -\infty < x < \infty. \quad (2)$$

Les distributions pondérées ont de nombreuses applications, nous tenons à citer quelques-uns des résultats les plus récents sur les size biased.

Warren (1975) a été le premier à les appliquer dans le cadre de l'échantillonnage de cellules de bois. Van Deusen (1986) est arrivé indépendamment à la théorie de la distribution biaisée par la taille et l'appliquer à la distribution appropriée des don-

nées de diamètre à hauteur d'homme (DHP) issues d'inventaires par échantillonnage horizontal (HPS) (Grosenbaugh, 1958).

Par la suite, Lappi et Bailey (1987) ont utilisé des distributions pondérées pour analyser les données d'incrément de diamètre HPS. Plus récemment, des distributions pondérées ont été utilisées par Magnussen et Al. (1999) pour récupérer la distribution des hauteurs de la canopée à partir de mesures au scanner laser. En écologie, Dennis et Patil (1984) utilisent des méthodes stochastiques. équations différentielles pour arriver à une distribution gamma pondérée en tant que fonction de densité de probabilité stationnaire (PDF) pour un modèle de population stochastique avec des effets de prédation. Dans les pêcheries, Taillie et Al. (1995) ont modélisé des populations de stocks de poissons utilisant des distributions pondérées. Récemment, Ghitany et al (2008) ont introduit une distribution de durée de vie nommée distribution Size biased Poisson Lindley, où la fonction de densité de la variable aléatoire X est donnée par :

$$f_{SBPL}(x, \theta) = \frac{\theta^3}{\theta + 2} \frac{x(x + \theta + 2)}{(1 + \theta)^{x+2}}, \quad x = 1, 2, \dots, \theta > 0.$$

Dans cette thèse, nous introduisons une nouvelle distribution de durée de vie en utilisant les modèles pondérés (Size Baised) donné par la relation (2).

Notre projet de recherche est structuré de la manière suivante :

Le premier chapitre est consacré aux rappels des certaines définitions et certains résultats sur les probabilités qui nous utilisons par la suite.

Dans le chapitre 2, nous faisons une synthèse des résultats sur les nouvelles dis-

tributions : Poisson Lindley, Size biased Poisson Lindley, Poisson Pseudo Lindley et Poisson Quasi Lindley dont on donne quelques propriétés.

Le chapitre 3, se focalise sur la nouvelle distribution à deux paramètres (size biased Poisson Gamma Lindley) dont on donne quelques propriétés à savoir : Moments, Estimation, méthode des moments, estimation du maximum de vraisemblance et la distribution de limitation des statistiques d'ordre extrême. Plusieurs simulations sont établies pour examiner le biais et l'erreur quadratique moyenne des estimateurs des paramètres obtenus par la méthode du maximum de vraisemblance et leur application dans l'analyse de survie. Par ailleurs, une comparaison des distributions est obtenue.

Dans le chapitre 4, nous nous intéressons à la nouvelle distribution tronqué nommé la distribution de Poisson Quasi-lindley zéro tronquée qui est une extension de l'article de Seghier F.Z. et Zaghdoudi. H.(2020) dont nous donnons quelques propriétés tel que la fonction de masse, les moments. Nous nous intéressons aussi à l'estimation des paramètres de cette distribution nous utilisons deux méthodes, la méthode des moments et la méthode du maximum de vraisemblance. Nous finalisons ce chapitre avec une application par des données réelles.

Chapitre 1

Généralités et quelques notions probabiliste

Dans ce chapitre, nous rappelons quelques notions et certaines propriétés de la théorie des probabilités que nous utiliserons dans cette thèse. Nous commençons tout d'abord par rappeler quelques définitions de base (variables aléatoires discrètes, Fonction de répartition, moments, fonction génératrice). Ensuite nous donnons quelques concepts nécessaires à l'étude de l'analyse de survie, telles que les fonctions de survie et de risque, Comme illustration, nous introduisons la notion d'estimateur et nous définissons les propriétés essentielles que doit vérifier un estimateur. Enfin, dans les situations où il n'y a pas d'estimateur évident, on est amené à recourir à une méthode de construction d'un estimateur, les deux méthodes que nous présenterons ici étant celles du maximum de vraisemblance et des moments.

Nous supposons tout au long de cette thèse que X est une variable aléatoire positive ou nulle discrète.

1.1 variables aléatoires discrètes

La théorie des probabilités a pour objet l'étude des phénomènes aléatoires ou du moins considérés comme tels par l'observateur. Pour cela on introduit le concept d'expérience aléatoire dont l'ensemble des résultats possibles constitue l'ensemble fondamental, noté habituellement Ω . On parle de variable aléatoire (abréviation : v.a.) lorsque les résultats sont numériques, c'est-à-dire que Ω est identique à tout ou partie de l'ensemble des nombres réels \mathbb{R} .

Définition 1.1 *les variables aléatoires discrètes pour lesquelles l'ensemble Ω des résultats possibles est un ensemble discret de valeurs numériques $x_1, x_2, \dots, x_n, \dots$ fini ou infini (typiquement : l'ensemble des entiers naturels).*

1.2 Fonction de répartition

La fonction de répartition est l'instrument de référence pour définir de façon unifiée la loi de probabilité d'une variable aléatoire qu'elle soit discrète ou continue. Si cette fonction est connue, il est possible de calculer la probabilité de tout intervalle et donc,

en pratique, de tout évènement. C'est pourquoi c'est elle qui est donnée dans les tables des lois de probabilité.

Définition 1.2 *Soit X une variable aléatoire, on appelle fonction de répartition de X , que l'on note F_X , la fonction définie sur \mathbb{R} par :*

$$F_X(x) = \mathbb{P}(X \leq x).$$

La valeur prise par la fonction de répartition au point x est donc la probabilité de l'évènement $] -\infty, x]$. En anglais on l'appelle «cumulative distribution function» par analogie avec la notion de fréquence cumulée en statistique descriptive.

1.3 Moments d'une variable aléatoire discrète

En plus des moyenne, qui sont les caractéristiques de position, on utilise également certaines autres caractéristiques des répartitions. les moments sont les plus courantes de ces caractéristiques.

Dans les applications pratiques le plus souvent on utilise les moments initiaux et les moments centrés.

1.3.1 Moment ordinaire

Le moment ordinaire (initial) d'ordre $r \in \mathbb{N}$ de X est défini, s'il existe, par :

$$\mu_r^l = \mathbb{E}(X^r) = \sum_{x \in I} x^r \mathbb{P}[X = x].$$

1.3.2 Moment factoriel

Un moment factoriel est un moment particulier, utilisé notamment dans l'étude d'une loi discrète

On appelle moment factoriel d'ordre r de la variable aléatoire X , où r est un entier positif, la valeur (si elle existe)

$$\mu'_{(r)} = \mathbb{E}(X_{(r)}),$$

avec $X_{(r)} = X(X-1)(X-2)\dots(X-r+1)$, où $r = 1, 2, 3, \dots$ est l'ordre du moment.

Pour la conversion d'un moment factoriel en un moment ordinaire, la relation est

$$\mu_r^{\prime} = \sum_{j=1}^r S(r, j) \mu'_{(j)}, \quad \text{pour } r = 1, 2, \dots$$

où les coefficients $S(r, j)$ sont les nombres de Stirling de deuxième espèce [30]

Pour des moments dont l'ordre $r = 1, 2, 3, 4$ on peut déduire les relations suivantes :

$$\mu_1^{\prime} = \mu'_{(1)},$$

$$\mu_2^{\prime} = \mu'_{(2)} + \mu'_{(1)},$$

$$\mu_3^{\prime} = \mu'_{(3)} + 3\mu'_{(2)} + \mu'_{(1)},$$

$$\mu_4^{\prime} = \mu'_{(4)} + 6\mu'_{(3)} + 7\mu'_{(2)} + \mu'_{(1)}.$$

Avant de donner la définition du moment centré introduisons la notion de variable aléatoire centrée.

Soit une variable aléatoire X d'espérance mathématique $\mu_1^{\prime} = \mathbb{E}(X)$. On appelle

variable aléatoire centrée associée à X la différence

$$X - \mathbb{E}(X).$$

Il est facile de voir que l'espérance mathématique d'une variable aléatoire centrée est nulle. En effet pour une variable discrète on a :

$$\begin{aligned} \mathbb{E}([X - \mathbb{E}(X)]) &= \sum_{i=1}^n (x_i - \mathbb{E}(X)) p_i \\ &= \sum_{i=1}^n x_i p_i - \mathbb{E}(X) \sum_{i=1}^n p_i \\ &= \mathbb{E}(X) - \mathbb{E}(X) = 0. \end{aligned}$$

Dans le cas continu on a le même résultat.

1.3.3 Moment centré

Le moment centré d'ordre $r \in \mathbb{N}$ de X est défini, s'il existe, par :

$$\mu_r = \mathbb{E}([X - \mathbb{E}(X)]^r) = \sum_{i=1}^n [x_i - \mathbb{E}(X)]^r p_i.$$

Il est évident que pour toute variable aléatoire X , le moment centré d'ordre 1 est nul :

$$\mu_1 = \mathbb{E}([X - \mathbb{E}(X)]) = 0,$$

car l'espérance mathématique d'une variable centrée est toujours nulle.

Cherchons les relations entre les moments centrés et initiaux de différents ordres.

Considérons maintenant le moment centrés deux :

$$\begin{aligned}
 \mu_2 &= \mathbb{E}([X - \mathbb{E}(X)]^2) = \sum_{i=1}^n (x_i - \mu'_1)^2 p_i \\
 &= \sum_{i=1}^n x_i^2 p_i - 2\mu'_1 \sum_{i=1}^n x_i p_i + (\mu'_1)^2 \sum_{i=1}^n p_i \\
 &= \mu'_2 - 2(\mu'_1)^2 + (\mu'_1)^2 \\
 &= \mu'_2 - (\mu'_1)^2 .
 \end{aligned}$$

D'une manière analogue pour le moments centré trois on obtient :

$$\begin{aligned}
 \mu_3 &= \mathbb{E}([X - \mathbb{E}(X)]^3) = \sum_{i=1}^n (x_i - \mu'_1)^3 p_i \\
 &= \sum_{i=1}^n x_i^3 p_i - 3\mu'_1 \sum_{i=1}^n x_i^2 p_i + 3(\mu'_1)^2 \sum_{i=1}^n x_i p_i - (\mu'_1)^3 \sum_{i=1}^n p_i \\
 &= \mu'_3 - 3\mu'_2 \cdot \mu'_1 + 2(\mu'_1)^3 .
 \end{aligned}$$

Les expressions pour $\mu_4, \mu_5 \dots$ etc. peuvent être obtenues d'une manière analogue.

Pour les moments centrés d'une variable aléatoire quelconque X on a donc :

$$\begin{aligned}
 \mu_1 &= 0, \\
 \mu_2 &= \mu'_2 - (\mu'_1)^2, \\
 \mu_3 &= \mu'_3 - 3\mu'_2 \cdot \mu'_1 + 2(\mu'_1)^3, \\
 \mu_4 &= \mu'_4 - 4\mu'_3 \cdot \mu'_1 + 6\mu'_2 \cdot (\mu'_1)^2 - 3(\mu'_1)^4, \\
 &\dots
 \end{aligned}$$

L'espérance mathématique μ'_1 et la variance μ_2 sont les caractéristiques les plus souvent utilisées d'une répartition. Elles caractérisent ses traits essentiels, à savoir sa

position et son degré de dispersion. pour en donner une description plus détaillée, on fait appel à des moments d'ordre supérieur.

Le moment centré trois caractérise l'asymétrie d'une répartition. on appelle alors *coefficient d'asymétrie* (skewness) la quantité :

$$\sqrt{\beta_1} = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}}.$$

Le moment centré quatre sera à caractériser l'aplatissement de la courbe de répartition. cette propriété de la répartition est donnée par *le coefficient d'aplatissement*. on appelle alors *coefficient d'aplatissement* (kurtosis) la quantité :

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2}.$$

Le *coefficient de variation* (C.V.) est :

$$C.V. = \frac{\sqrt{\mu_2}}{\mu_1'}.$$

1.4 Fonction génératrice des moments

La fonction génératrice des moments nous intéresse dans la mesure où elle peut faciliter le calcul des moments d'une loi. Cependant son existence -et donc son usage- sera limitée aux lois dont la densité (éventuellement la fonction de probabilité) décroît plus vite qu'une exponentielle à l'infini. Nous supposons ci-après qu'elle existe au moins au voisinage de 0 et que la loi admet des moments de tous ordres.

Définition 1.3 *On appelle fonction génératrice des moments de la v.a. X , si elle existe, la fonction :*

$$G_X(t) = \mathbb{E}(e^{tX}).$$

C'est une fonction de t par la variable X introduite dans la fonction aléatoire e^{tX} .

Proposition 1.4 *Le moment d'ordre r de la v.a. X est donné par :*

$$\mu_r = G_X^{(r)}(0),$$

où $G_X^{(r)}$ est la dérivée d'ordre r de G_X .

En particulier l'espérance mathématique (μ_1) de X est la valeur de la dérivée première G_X' pour $t = 0$.

1.5 Distribution d'une durée de survie

Pour mener à bien une analyse de survie sur population, il est nécessaire de connaître pour chaque patient quelques données.

- La date d'origine est la date à partir de laquelle le patient est observé. Dans notre cas, cela correspondra à la date de diagnostic de cancer. Dans le cadre d'essais cliniques, il peut s'agir de la date d'entrée dans l'étude.

- La date des dernières nouvelles est la date la plus récente à laquelle on a recueilli des informations sur le patient. À cette date, nous disposons du statut aux dernières nouvelles. Dans notre cas, cela signifie que l'on sait si le patient est vivant ou décédé.

– La date de point est commune à tous les individus de la cohorte puisqu’il s’agit de la date d’arrêt de l’étude. À partir de cette date, on ne tient plus compte des informations dont on peut éventuellement disposer sur certains patients.

Ces données nous permettent de calculer le recul, qui est le délai entre la date d’origine et la date de point.

Enfin, il est possible de calculer le temps de suivi qui est la durée entre la date d’origine et

- la date des dernières nouvelles si celle-ci est antérieure à la date de point ;
- la date de point sinon.

La durée de survie est la durée entre la date d’origine et la survenue de l’évènement d’intérêt, c’est-à-dire du décès. Elle correspond au temps de suivi lorsque le décès est observé avant la date de point.

Soit X une variable aléatoire positive ou nulle représentant une durée de survie. Définissons les fonctions qui caractérisent la loi de probabilité de X . Notons que chacune d’entre elles peut être obtenue à partir de l’une des autres.

Deux fonctions décrivant la distribution des temps de survie, qui sont d’importance centrale dans l’analyse des données de survie : la fonction de survie et la fonction de risque.

1.5.1 Fonction de survie $S(t)$

Définition 1.5 *La fonction de survie est, pour t fixé, la probabilité de survivre jusqu'à l'instant t , c'est à dire :*

$$S(t) = \mathbb{P}(X > t); \quad t \geq 0.$$

Définition 1.6 *La fonction de répartition F fait correspondre à un temps t la probabilité de décéder avant le temps t :*

$$F_X(t) = \mathbb{P}(X \leq t) = 1 - S(t).$$

1.5.2 Fonction de risque $h(t)$ (taux de hasard)

Le risque instantané $h(x)$ (ou taux d'incidence), pour t fixé caractérise la probabilité de mourir dans un petit intervalle de temps après t , conditionnellement au fait d'avoir survécu jusqu'au temps t (c'est-à-dire le risque de mort instantané pour ceux qui ont survécu) :

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{\mathbb{P}[t < X \leq t + \Delta t, X > t]}{\mathbb{P}[X > t]} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{\mathbb{P}[t < X \leq t + \Delta t, X > t]}{1 - \mathbb{P}[X \leq t]} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{F_X[t + \Delta t] - F_X[t]}{1 - F_X[t]}. \end{aligned}$$

1.5.3 Le concept de censure

La spécificité des données de survie est que l'on dispose de « données incomplètes » dans le sens où le décès n'est pas forcément observé durant le temps de suivi.

Plusieurs types de censure existent et nous nous limiterons, dans le cadre de cette thèse, à la censure dite censure aléatoire à droite.

Soit C_i une variable aléatoire positive ou nulle représentant le temps de censure pour le patient i et T_i sa durée de survie. La durée T_i est dite censurée à droite si $C_i < T_i$.

Le temps d'observation du patient i est alors $U_i := \min(T_i, C_i)$ et son statut est donné par $\delta_i := 1_{(T_i \leq C_i)}$ qui vaut 1 si le patient est décédé et 0 s'il est censuré.

1.6 Estimation

Dans cette section nous considérerons toujours, même si des développements analogues sont possibles dans d'autres circonstances, que x_1, x_2, \dots, x_n sont des réalisations d'un n échantillon aléatoire X_1, X_2, \dots, X_n .

La théorie de l'estimation étudie les propriétés des estimations et des méthodes générales d'estimation comme celle dite du «maximum de vraisemblance». L'objectif est de comparer les lois d'échantillonnage des «estimateurs» pour établir des préférences lorsque plusieurs choix se présentent. La notion d'estimateur est la notion centrale de ce chapitre alors même qu'elle ne se définit pas formellement en termes mathématiques.

1.6.1 Définition d'un estimateur

On a donc construit un modèle statistique où la v.a. X suit une loi \mathbb{P}_θ et pour se faire une idée de la valeur inconnue du paramètre θ qui détermine la loi de probabilité, on utilise un échantillon de cette loi. À partir des valeurs observées x_1, \dots, x_n on calcule ensuite une certaine valeur numérique que l'on considèrera comme une valeur approchée de θ et qu'on appellera une estimation de θ . La règle qui permettra d'effectuer ce calcul est un estimateur, dont la définition précise est la suivante.

Définition 1.7 *Un estimateur de θ est une application T_n de E^n ($E \subset \mathbb{R}$ ou \mathbb{R}^k) dans F ($F \subset \mathbb{R}$ ou \mathbb{R}^k) qui à un échantillon (X_1, \dots, X_n) de la loi \mathbb{P}_θ associe une variable aléatoire réelle (ou plusieurs dans le cas d'un paramètre multidimensionnel) dont on peut déterminer la loi de probabilité.*

Nous allons maintenant définir les propriétés essentielles que doit vérifier un estimateur.

1.6.2 Propriétés d'un estimateur

Biais d'un estimateur

Définition 1.8 *Soit une v. a. X de loi de fonction de probabilité $P(x; \theta)$ où $\theta \in \Theta \subseteq \mathbb{R}$. Soit X_1, X_2, \dots, X_n un n échantillon issu de cette loi et T_n un estimateur de θ . On appelle biais de T_n pour θ la valeur :*

$$b_\theta(T_n) = \mathbb{E}_\theta(T_n) - \theta.$$

Si $b_\theta(T_n) = 0$ quel que soit $\theta \in \Theta$, on dit que T_n est sans biais pour θ .

Cependant, cette propriété peut ne pas être strictement vérifiée, le biais diminuant seulement quand la taille d'échantillon augmente. Ceci correspond à la définition suivante.

Définition 1.9 *Un estimateur T_n de θ est dit asymptotiquement sans biais si pour tout θ de :*

$$\mathbb{E}_\theta(T_n) \rightarrow \theta \quad \text{quand } n \rightarrow \infty.$$

Convergence d'un estimateur

Intuitivement, on est amené à penser que si la taille de l'échantillon augmente, l'information sur le paramètre θ va augmenter et donc l'estimateur devrait d'une certaine manière se «rapprocher» de la valeur de θ . Cet estimateur étant une v.a., la formulation mathématique de cette notion intuitive va faire appel à la convergence en probabilité d'une suite de v.a. : T_n prendra des valeurs proches de θ avec une probabilité d'autant plus proche de un que la taille d'échantillon n sera grande. Ceci conduit à la définition suivante.

Définition 1.10 *Un estimateur T_n est convergent si la suite de v.a. (T_n) converge en probabilité vers la valeur du paramètre, c'est-à-dire pour tout θ de :*

$$\begin{aligned} T_n \xrightarrow{\mathbb{P}} \theta &\Leftrightarrow \mathbb{P}_\theta(|T_n - \theta| < \varepsilon) \rightarrow 1, \forall \varepsilon > 0, n \rightarrow \infty \\ &\Leftrightarrow \mathbb{P}_\theta(|T_n - \theta| > \varepsilon) \rightarrow 0. \end{aligned}$$

Il existe un moyen en général simple de vérifier cette propriété de convergence, à partir de conditions suffisantes faisant intervenir les deux premiers moments de l'estimateur, et que nous énonçons dans le théorème ci-après.

Théorème 1.11 [20] *Tout estimateur sans biais dont la variance tend vers zéro est convergent :*

$$\begin{cases} \mathbb{E}_\theta(T_n) = \theta \\ V_\theta(T_n) \rightarrow 0 \end{cases} \Rightarrow T_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \theta.$$

Variance et erreur quadratique moyenne d'un estimateur

La variance $V_\theta(T_n)$ de l'estimateur est un critère important dans la mesure où elle caractérise la dispersion des valeurs de T_n dans l'univers des échantillons possibles. Toutefois il s'agit de la dispersion autour de $\mathbb{E}_\theta(T_n)$ et non pas autour de θ . Pour prendre en compte l'écart par rapport à θ on introduit le critère d'erreur quadratique moyenne.

Définition 1.12 *On appelle erreur quadratique moyenne de T_n par rapport à θ , la valeur, notée $EQM_\theta(T_n)$, définie par :*

$$EQM_\theta(T_n) = \mathbb{E}_\theta[(T_n - \theta)^2],$$

et l'on a :

$$EQM_\theta(T_n) = [b_\theta(T_n)]^2 + V_\theta(T_n).$$

En effet :

$$\begin{aligned}
\mathbb{E}_\theta[(T_n - \theta)^2] &= \mathbb{E}_\theta[\{T_n - \mathbb{E}_\theta(T_n) + \mathbb{E}_\theta(T_n) - \theta\}^2] \\
&= \mathbb{E}_\theta[\{T_n - \mathbb{E}_\theta(T_n)\}^2] + [\mathbb{E}_\theta(T_n) - \theta]^2 + 2\mathbb{E}_\theta[T_n - \mathbb{E}_\theta(T_n)][\mathbb{E}_\theta(T_n) - \theta] \\
&= V_\theta(T_n) + [b_\theta(T_n)]^2 \\
\text{car } \mathbb{E}_\theta[T_n - \mathbb{E}_\theta(T_n)] &= 0.
\end{aligned}$$

1.6.3 Normalité asymptotique

On admettra le théorème central limite (TCL) et la méthode Delta (D) suivants :

Théorème 1.13 (*Théorème central limite*)

Si (X_n) est une suite de v.a. indépendantes et de même loi, admettant des moments d'ordres un et deux notés $\mu = \mathbb{E}(X_n)$ et $\sigma^2 = V(X_n)$, alors :

$$\sqrt{n} \frac{\overline{X_n} - \mu}{\sigma} \rightsquigarrow \mathcal{N}(0, 1).$$

La suite $\sqrt{n} \frac{\overline{X_n} - \mu}{\sigma}$ converge en loi vers la loi normale centrée réduite.

Méthode Delta

Soient une suite de variables aléatoires X_1, X_2, \dots, X_n d'espérance θ et d'écart-type σ . Si $\sqrt{n}(X_n - \theta) \xrightarrow{L} \mathcal{N}(0, \sigma^2)$, et une fonction g dérivable telle que $g'(\theta) \neq 0$. Dans ce cas la méthode delta donne :

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{L} \mathcal{N}\left(0, \sigma^2 [g'(\theta)]\right).$$

Définition 1.14 Soit $\hat{\theta}$ un estimateur du paramètre θ de la loi \mathbb{P}_θ d'une v. a. observée X . On suppose qu'il existe deux suites de fonctions réelles strictement positives, $a =$

$a_n(\theta)$ et $b = b_n(\theta)$ telles que :

$$\frac{\hat{\theta} - a}{b} \xrightarrow{L} \mathcal{N}(0, 1).$$

On dit alors que $\hat{\theta}$ est un estimateur asymptotiquement normal.

1.7 Méthodes de construction d'un estimateur

Dans les situations où il n'y a pas d'estimateur évident, on est amené à recourir à une méthode de construction d'un estimateur, les deux méthodes que nous présenterons ici étant celles du maximum de vraisemblance et des moments.

1.7.1 Méthode du maximum de vraisemblance

Définition 1.15 On appelle *vraisemblance (likelihood)* de l'échantillon (X_1, \dots, X_n)

la loi de probabilité de ce n -uplet, notée $L(x_1, \dots, x_n; \theta)$, et définie par :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n P(X_i = x_i),$$

où X est une v.a. discrète.

La vraisemblance $L(x_1, \dots, x_n; \theta)$ représente la probabilité d'observer le n -uplet (x_1, \dots, x_n) pour une valeur fixée de θ . Dans la situation inverse ici où on a observé (x_1, \dots, x_n) sans connaître la valeur de θ , on va attribuer à θ la valeur qui paraît la plus vraisemblable, compte tenu de l'observation dont on dispose, c'est-à-dire celle qui va

lui attribuer la plus forte probabilité. On se fixe donc la règle suivante : à (x_1, \dots, x_n) fixé, on considère la vraisemblance L comme une fonction de θ et on attribue à θ la valeur qui maximise cette fonction. D'où la définition suivante :

Définition 1.16 *On appelle estimateur du maximum de vraisemblance (EMV) toute fonction $\hat{\theta}_n$ de (x_1, \dots, x_n) qui vérifie :*

$$L(x_1, \dots, x_n; \hat{\theta}_n) = \max_{\theta \in \Theta} L(x_1, \dots, x_n; \theta).$$

Quand $\theta = (\theta_1, \dots, \theta_d) \in \Theta$ et que toutes les dérivées partielles ci-dessous existent, $\hat{\theta}_n$ est solution du système d'équations appelées équations de vraisemblance :

$$\forall j \in \{1, \dots, d\}, \quad \frac{\partial}{\partial \theta_j} \ln L(\theta; x_1, \dots, x_n) = 0,$$

Où

$$\frac{\partial^2}{\partial \theta_j^2} \ln L(\theta; x_1, \dots, x_n) < 0.$$

Dans ce cas, on le résout par des méthodes numériques, comme la méthode de Fisher Scoring.

1.7.2 Méthode des Moments

L'idée de base est d'estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

Autrement dit, si $\theta = \mathbb{E}(X)$, alors l'estimateur de θ par la méthode des moments est

$$\hat{\theta}_n = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Plus généralement, pour $\theta \in \Theta$, si $\mathbb{E}(X) = \varphi(\theta)$, où φ est une fonction inversible, alors l'estimateur de θ par la méthode des moments est :

$$\hat{\theta}_n = \varphi^{-1}(\bar{X}).$$

De la même manière, on estime la variance de la loi des X_i par la variance empirique de l'échantillon $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

1.8 Quelques lois de probabilités discrètes usuelles

Si on améliore la compréhension et l'analyse d'un phénomène complexe par l'introduction d'un modèle qui la simplifie, celui-ci ne doit cependant pas être trop éloigné de la réalité. Nous allons présenter ici les principaux modèles qui peuvent être retenus pour une modélisation aléatoire.

1.8.1 Loi binômiale

Si on effectue n épreuves successives indépendantes où on note à chaque fois la réalisation ou non d'un certain événement A , on obtient une suite de la forme $AA\bar{A}\bar{A}\bar{A}\dots\bar{A}AA$. À cet événement élémentaire ω on associe le nombre $X(\omega)$ de réalisations de A .

On définit ainsi une v.a. X qui suit une *loi binômiale* de paramètres n et $p = \mathbb{P}(A)$, caractérisée par $X(\Omega) = \{0, 1, \dots, n\}$ et pour $k \in X(\Omega)$:

$$\mathbb{P}_X(X = k) = C_n^k p^k (1 - p)^{n-k}, \quad (1.1)$$

car C_n^k est le nombre d'échantillons de taille n comportant exactement k événements A , de probabilité p^k , indépendamment de l'ordre, et donc $n - k$ événements \bar{A} , de probabilité $(1 - p)^{n-k}$. On écrit $X \rightsquigarrow \mathcal{B}(n, p)$. Pour calculer facilement les moments de cette loi, nous allons associer à chaque épreuve i , $1 \leq i \leq n$, une v.a. de Bernoulli :

$$X_i = \begin{cases} 1 & \text{si } A \text{ est réalisé} \\ 0 & \text{si } \bar{A} \text{ est réalisé} \end{cases}$$

On peut alors écrire :

$$X = X_1 + X_2 + \dots + X_n,$$

d'où on déduit aisément :

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathbb{E}(X_i) = np,$$

et

$$Var(X) = Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) = np(1 - p).$$

1.8.2 Loi géométrique ou de Pascal

On effectue des épreuves successives indépendantes jusqu'à la réalisation d'un événement particulier A et on note X le nombre (aléatoire) d'épreuves effectuées. On définit ainsi une v.a. à valeurs entières de *loi géométrique*, ou de *Pascal*. À chaque épreuve est associé l'ensemble fondamental $\Omega = \{A, \bar{A}\}$ et l'événement $\{X = k\}$ pour $k \in \mathbb{N}^*$ est représenté par une suite de $k - 1$ événements \bar{A} , terminée par l'événement

A :

$$\underbrace{\overline{AA\dots AA}}_{k-1}$$

Si on pose $p = \mathbb{P}(A)$, la probabilité de cet événement est :

$$\mathbb{P}_X(X = k) = p(1 - p)^{k-1}, \quad (1.2)$$

d'où on déduit :

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} kp(1 - p)^{k-1} = \frac{1}{p},$$

et

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{(1 - p)}{p^2}.$$

1.8.3 Loi de Poisson

Une v. a. X suit une *loi de Poisson* de paramètre $\lambda > 0$ si c'est une variable à valeurs entières, $X(\Omega) = \mathbb{N}$, donc avec une infinité de valeurs possibles, de probabilité :

$$\mathbb{P}_X(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k \in \mathbb{N}^*, \quad (1.3)$$

loi qui ne dépend que d'un seul paramètre réel positif, avec l'écriture symbolique $X \rightsquigarrow \mathcal{P}(\lambda)$.

On déduit aisément :

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k \mathbb{P}_X(X = k) = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda,$$

et

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \lambda.$$

1.8.4 Loi binômiale négative

On effectue cette fois des épreuves successives indépendantes jusqu'à ce que n événements A soient réalisés et on note Y le nombre (aléatoire) d'épreuves effectuées. L'événement $\{Y = y\}$, pour tout entier $y \geq n$, est représenté par une suite de la forme :

$$\underbrace{\overline{A}A\overline{A}\dots\overline{A}A\overline{A}\dots\overline{A}A}_{y-1}$$

qui comporte $n-1$ réalisations de l'événement A au cours des $y-1$ premières épreuves et qui se conclut par un événement A . On en déduit la probabilité individuelle :

$$\mathbb{P}_Y(Y = y) = C_{y-1}^{n-1} p^n (1-p)^{y-n}, \quad y \geq n. \quad (1.4)$$

Pour obtenir sans calculs les moments de Y , nous allons décomposer la suite des épreuves en n séquences se terminant toutes par un événement A , associant à chacune de ces séquences une v. a. de Pascal X_i , $1 \leq i \leq n$, qui représente le nombre d'épreuves nécessaires pour que le i -ème événement A soit réalisé, en comptant à partir de la réalisation du précédent A :

$$\underbrace{\overline{A}\overline{A}\dots\overline{A}A}_{X_1} \underbrace{\overline{A}\overline{A}\overline{A}\overline{A}\dots\overline{A}A}_{X_2} \dots \underbrace{\overline{A}\overline{A}\dots\overline{A}A}_{X_n}$$

Ceci permet de définir la loi de Y , dite *loi binômiale négative*, comme somme de lois de Pascal indépendantes et de même paramètre p :

$$Y = X_1 + X_2 + \dots + X_n.$$

On en déduit alors facilement :

$$\mathbb{E}(Y) = n\mathbb{E}(X_1) = \frac{n}{p},$$

et

$$\text{Var}(Y) = n\text{Var}(X_1) = \frac{n(1-p)}{p^2}.$$

1.9 Distributions de Size biased

Les distributions biaisées par la taille *Size Biased* constituent un cas particulier de la forme plus générale connue sous le nom de distributions pondérées introduit pour la première fois par Fisher (1934) pour modéliser la détermination biais, ensuite les distributions pondérées ont été formalisées dans une théorie unificatrice par Rao (1965). De telles distributions apparaissent naturellement dans la pratique lorsque les observations d'un échantillon sont enregistrées avec une probabilité inégale, telle que de probabilité proportionnelle à dessins de taille (PPT). En bref, si la variable aléatoire X a une densité $P_0(x; \theta)$, avec paramètres inconnus θ , la densité pondérée correspondante est de la forme Size-biased distributions de forme :

$$P_1(x, \theta) = \frac{x^\alpha P(x; \theta)}{\mu'_\alpha},$$

où x^α est une fonction de pondération non négative telle que μ'_α existe. Un cas particulier des distributions pondérées, les distributions biaisées par la taille est proposé par Rao (1965) lorsque la fonction pondérée est de la forme x^α , appelée distribution

d'ordre biaisé par la taille α , où $\alpha = 1$ ou $\alpha = 2$, appelé longueur respectivement à base de zones et à zones biaisées. Par conséquent, la densité de la distribution biaisée en longueur est défini par :

$$P_1(x, \theta) = \frac{xP_0(x; \theta)}{\mu'_1}, \quad (1.5)$$

où

$$\mu'_1 = \mathbb{E}(X) = \sum_{x=0}^{\infty} xP_0(x; \theta), \quad (1.6)$$

est la moyenne de la distribution original.

1.9.1 Quelques Size-biased distributions discrètes

Dans cette section, nous avons obtenu quelques distributions de size-biased (biaisées par la taille) discrètes, en utilisant les équations (1.5) et (1.6).

Size-biased Binomial Distribution (SBBD)

La fonction de masse de la distribution de size biased binomiale (SBBD) peut être obtenue comme suit :

$$\begin{aligned} \mathbb{P}(X = x) &= \frac{x C_n^x p^x (1-p)^{n-x}}{np} \\ &= C_{n-1}^{x-1} p^{x-1} (1-p)^{n-x}, \quad x = 1, 2, \dots \end{aligned}$$

La moyenne et la variance de la distribution (SBBD) peuvent être obtenues et

sont données comme suit :

$$\begin{aligned}\mu_1' &= \frac{np + 1}{1 - p}, \\ \mu_2 &= \frac{np + p}{(1 - p)^2}.\end{aligned}$$

La fonction génératrice de moment de la distribution SBB est donnée par :

$$G_X(t) = \mathbb{E}(e^{tX}) = e^t (1 - p)^{n+1} (1 - pe^t)^{-(n+1)}.$$

Size- biased Poisson Distribution (SBPD)

La fonction de masse de la distribution de size biased binomiale (SBPD) peut être obtenue comme suit :

$$\mathbb{P}(X = x) = \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!}, \quad x = 1, 2, \dots$$

La moyenne et la variance de la distribution (SBPD) peuvent être obtenues et sont données comme suit :

$$\begin{aligned}\mu_1' &= \frac{np + 1}{1 - p}, \\ \mu_2 &= \frac{np + p}{(1 - p)^2}.\end{aligned}$$

La fonction génératrice de moment de la distribution SBP est donnée par :

$$G_X(t) = \mathbb{E}(e^{tX}) = e^t (1 - p)^{n+1} (1 - pe^t)^{-(n+1)}.$$

Size-biased Negative Binomial Distribution (SBNBD)

La fonction de masse de la distribution de size biased binomiale (SBNBD) peut être obtenue comme suit :

$$\mathbb{P}(X = x) = C_{n+x-1}^{x-1} p^{x-1} (1-p)^{n+1}, \quad x = 1, 2, \dots$$

La moyenne et la variance de la distribution (SBNBD) peuvent être obtenues et sont données comme suit :

$$\begin{aligned} \mu_1' &= \frac{np + 1}{1 - p}, \\ \mu_2 &= \frac{np + p}{(1 - p)^2}. \end{aligned}$$

La fonction génératrice de moment de la distribution SBNB est donnée par :

$$G_X(t) = \mathbb{E}(e^{tX}) = e^t (1-p)^{n+1} (1-pe^t)^{-(n+1)}.$$

Chapitre 2

Quelques nouvelles distribution discrètes et ses applications

Dans ce chapitre on collecte des nouvelles distributions discrètes dont on donne quelques propriétés à savoir : moments et mesure connexe, estimation du méthode des moments, estimation du maximum de vraisemblance. Plusieurs simulations sont établies pour examiner le biais et l'erreur quadratique moyenne des estimateurs des paramètres obtenus par la méthode du maximum de vraisemblance et leur application dans l'analyse de survie.

2.1 Distribution Poisson Lindley

Une distribution composée de Poisson peut être obtenue en composant la distribution de Poisson et une distribution due à Lindley. Cette distribution a été introduite par Sankaran (1970) [32] pour modéliser des données de comptage.

Supposons que le paramètre λ de la distribution de Poisson à une distribution appartenant à la famille exponentielle de distribution donnée par :

$$dF(\lambda) = e^{\lambda\Phi} h(\lambda) B(\Phi) d\lambda,$$

où $h(\lambda) = 1 + \lambda$ et $B(\Phi) = \frac{[-\Phi]^2}{(1-\Phi)}$.

Alors la distribution de Poisson composée est :

$$\begin{aligned} P_x(\Phi) &= \int_0^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} dF(\lambda) \\ &= \frac{B(\Phi)}{x!} \left[\int_0^{\infty} e^{(\Phi-1)\lambda} \lambda^x d\lambda + \beta \int_0^{\infty} e^{(\Phi-1)\lambda} \lambda^{x+1} d\lambda \right] \\ &= \frac{\Phi^2}{(1-\Phi)} \left(\frac{1-\Phi+x+1}{(1-\Phi)^{x+2}} \right). \end{aligned}$$

Alors on remplace Φ par $-\theta$ on trouve :

$$\mathbb{P}_x(\theta) = \theta^2 \frac{(x+2+\theta)}{(\theta+1)^{x+3}}.$$

La fonction de masse de Poisson Lindley (*PLD*) est :

$$f_{PLD}(x; \theta) = P_x(\theta) = \theta^2 \frac{(x+2+\theta)}{(\theta+1)^{x+3}}, \quad x = 0, 1, \dots, \theta > 0. \quad (2.1)$$

La fonction de répartition correspondante est :

$$F_{PLD}(x) = 1 - \frac{\theta^2 + 3\theta + 1 - \theta x}{(\theta+1)^{x+3}}, \quad x = 0, 1, \dots, \theta > 0. \quad (2.2)$$

2.1.1 Moments et Mesures Connexes

Le moment factoriel d'ordre r de la distribution Poisson-Lindley peut être obtenu en tant que :

$$\begin{aligned}
 \mu'_{(r)} &= \mathbb{E}(\mathbb{E}(X^{(r)} | \lambda)) = \int_0^{\infty} \left[\sum_{x=0}^{\infty} x^r e^{-\lambda} \frac{\lambda^x}{x!} \right] \cdot h(\lambda; \theta) d\lambda \\
 &= \int_0^{\infty} \left[\sum_{x=0}^{\infty} x^r e^{-\lambda} \frac{\lambda^x}{x!} \right] \cdot \frac{\theta^2}{\theta + 1} (1 + \lambda) e^{-\theta\lambda} d\lambda \\
 &= \frac{\theta^2}{\theta + 1} \int_0^{\infty} \lambda^r \left[\sum_{x=r}^{\infty} e^{-\lambda} \frac{\lambda^{x-r}}{(x-r)!} \right] \cdot (1 + \lambda) e^{-\theta\lambda} d\lambda.
 \end{aligned}$$

En prenant $y = x - r$, on a

$$\begin{aligned}
 \mu'_{(r)} &= \frac{\theta^2}{\theta + 1} \int_0^{\infty} \lambda^r \left[\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} \right] \cdot (\lambda + 1) e^{-\theta\lambda} d\lambda \\
 &= \frac{\theta^2}{\theta + 1} \int_0^{\infty} \lambda^r (\lambda + 1) e^{-\theta\lambda} d\lambda,
 \end{aligned}$$

nous obtenons finalement une expression générale du moment factoriel du distribution PL comme :

$$\mu'_{(r)} = \frac{r!(\theta + r + 1)}{(\theta + 1)\theta^r}; \quad r = 1, 2, 3, \dots \quad (2.3)$$

En substituant $r = 1, 2, 3$ et 4 dans (2.3), les quatre premiers moments factoriels

de PLD peuvent être obtenus comme :

$$\begin{aligned}\mu'_{(1)} &= \frac{\theta + 2}{\theta(\theta + 1)}, \\ \mu'_{(2)} &= \frac{2(\theta + 3)}{\theta^2(\theta + 1)}, \\ \mu'_{(3)} &= \frac{6(\theta + 4)}{\theta^3(\theta + 1)}, \\ \mu'_{(4)} &= \frac{24(\theta + 5)}{\theta^4(\theta + 1)}.\end{aligned}$$

Puis en utilisant la relation entre les moments factoriels et les moments d'origine, les quatre premiers moments d'origine de la distribution PL sont donnés par :

$$\begin{aligned}\mu_1 &= \frac{\theta + 2}{\theta(\theta + 1)}, \\ \mu_2 &= \frac{\theta + 2}{\theta(\theta + 1)} + \frac{2(\theta + 3)}{\theta^2(\theta + 1)}, \\ \mu_3 &= \frac{\theta + 2}{\theta(\theta + 1)} + \frac{6(\theta + 3)}{\theta^2(\theta + 1)} + \frac{6(\theta + 4)}{\theta^3(\theta + 1)}, \\ \mu_4 &= \frac{\theta + 2}{\theta(\theta + 1)} + \frac{14(\theta + 3)}{\theta^2(\theta + 1)} + \frac{36(\theta + 4)}{\theta^3(\theta + 1)} + \frac{24(\theta + 5)}{\theta^4(\theta + 1)}.\end{aligned}$$

En utilisant la relation entre les moments de l'origine et les moments centrés de la distribution PL sont obtenues comme suit :

$$\begin{aligned}\mu_2 &= \frac{\theta^3 + 4\theta^2 + 6\theta + 2}{\theta^2(\theta + 1)^2}, \\ \mu_3 &= \frac{2(\theta + 1)^4(\theta + 2) - \theta^3(\theta + 2)(\theta + 3)}{\theta^3(\theta + 2)^3}, \\ \mu_4 &= \frac{3(\theta^3 + 4\theta^2 + 6\theta + 2) + 2(\theta + 1)^5[(\theta + 3)^2 - 3] - \theta^4(\theta + 2)[(\theta + 4)^2 - 3]}{\theta^4(\theta + 2)^4}.\end{aligned}$$

Le coefficient de variation ($C.V$), le coefficient de Skewness ($\sqrt{\beta_1}$), le coefficient

de Kurtosis (β_2) de PLD (2.1) sont obtenues comme suit :

$$C.V = \frac{\sqrt{\theta^3 + 4\theta^2 + 6\theta + 2}}{\theta + 2},$$

$$\sqrt{\beta_1} = \frac{2(\theta + 1)^4(\theta + 2) - \theta^3(\theta + 2)(\theta + 3)}{[2(\theta + 1)^3 - \theta^2(\theta + 2)]^{\frac{3}{2}}},$$

$$\beta_2 = 3 + \frac{2(\theta + 1)^5 [(\theta + 3)^2 - 3] - \theta^4(\theta + 2) [(\theta + 4)^2 - 3]}{[2(\theta + 1)^3 - \theta^2(\theta + 2)]^2}.$$

2.1.2 Fonction génératrice

La fonction génératrice de probabilité de la distribution PL est obtenue comme suit :

$$\begin{aligned} G_X(t) &= \mathbb{E}(t^X) = \sum_{x=1}^{\infty} t^X P(x; \theta) \\ &= \frac{\theta^2}{(\theta + 1)^3} \left[\sum_{x=1}^{\infty} x \left(\frac{t}{1 + \theta} \right)^x + (2 + \theta) \sum_{x=1}^{\infty} \left(\frac{t}{1 + \theta} \right)^x \right] \\ &= \frac{\theta^2}{\theta + 1} \frac{2 + \theta - t}{(\theta + 1 - t)^2}. \end{aligned}$$

La fonction génératrice de moment de la distribution PL (2.1) est donnée par :

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

2.1.3 Estimation

Maximum de vraisemblance

Étant donné un échantillon aléatoire X_1, \dots, X_n , de la distribution de Poisson Lindley discrète (2.1), la fonction de logvraisemblance est :

$$\ln L(x_i; \beta, \theta) = 2n \ln \theta - n(\bar{x} + 3) \ln(1 + \theta) + \sum_{i=1}^n \ln[x_i + \theta + 2].$$

Sankaran [32] a montré que l'estimateur de la méthode du maximum de vraisemblance $\hat{\theta}$ de θ est une solution de l'équation :

$$\frac{\partial \ln L(x_i, \theta)}{\partial \theta} = \frac{2n}{\theta} - n\left(\frac{\bar{x} + 3}{1 + \theta}\right) + \sum_{i=1}^n \frac{1}{x_i + \theta + 2} = 0. \quad (2.4)$$

Sankaran [32] a déclaré que la résolution de l'équation (2.4) est équivalente à la résolution d'un polynôme de degré $(n + 1)$, et l'équation (2.4) peuvent avoir plusieurs solutions. Dans ce qui suit, on montre que l'équation (2.4) a une solution unique pour tout n .

Théorème 2.1 [13] *L'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ est unique pour tout n .*

Démonstration. l'équation (2.4) est équivalent à

$$\xi(\theta) = 2n - n\bar{x}\theta - \sum_{i=1}^n \frac{(x_i + 1)\theta}{x_i + \theta + 2} = 0.$$

La fonction $\xi(\theta)$ est strictement décroissante en θ , puisque

$$\xi'(\theta) = -n\bar{x} - \sum_{i=1}^n \frac{(x_i + 1)(x_i + 2)}{(x_i + \theta + 2)^2} < 0.$$

Comme $\xi(0) = 2n$ et $\xi(\infty) = -\infty$, il en résulte que $\xi(\theta)$ traversera l'axe θ qu'une seule fois, c'est-à-dire il existe un unique $\hat{\theta}$ tel que $\xi(\hat{\theta}) = 0$. ■

Remarque 2.2 *Bien qu'efficace de l'estimateur du maximum de vraisemblance $\hat{\theta}$ mais ce dernier n'est pas simple obtenu et il peut être suffisant pour des raisons pratiques d'utiliser l'estimation des moments.*

Méthode des moments

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

$$\hat{\theta}_{MoM} = \frac{-(\bar{X} - 1) + \sqrt{(\bar{X} - 1)^2 + 8\bar{X}}}{2\bar{X}}, \quad \bar{X} > 0. \quad (2.5)$$

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

Théorème 2.3 [13] *L'estimateur $\hat{\theta}_{MoM}$ de θ est positivement biaisée, c'est-à-dire*

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

Démonstration. Soit $\hat{\theta} = g(\bar{X})$, et $g(t) = \frac{-(t-1) + \sqrt{(t-1)^2 + 8t}}{2t} \quad \forall t > 0$.

Comme

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

alors $g(t)$ est strictement convexe.

Ainsi, par l'inégalité de Jensen, on a

$$\mathbb{E}(g(\bar{X})) > g[\mathbb{E}(\bar{X})].$$

Enfin, étant donné que

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

on obtient

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

■

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

Théorème 2.4 [13] *L'estimateur $\hat{\theta}_{MoM}$ de θ est convergent et asymptotiquement normal :*

$$\sqrt{n}(\hat{\theta}_{MoM} - \theta) \xrightarrow{D} \mathcal{N}(0, \vartheta^2(\theta)).$$

Où

$$\vartheta^2(\theta) = \frac{\theta^2(\theta + 1)^2(\theta^3 + 4\theta^2 + 6\theta + 2)}{(\theta^2 + 4\theta + 2)^2}.$$

Démonstration. Convergence :

Étant donné $\mu < \infty$, $\bar{X} \xrightarrow{\mathbb{P}} \mu$, aussi, puisque $g(t)$ est une fonction continue à $t = \mu$, alors $g(\bar{X}) \xrightarrow{\mathbb{P}} g(\mu)$, c'est-à-dire

$$\hat{\theta}_{MoM} \xrightarrow{\mathbb{P}} \theta.$$

Asymptotique normal :

Comme $\sigma^2 < \infty$, par le théorème central limite, on a :

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

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

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

Enfin, étant donné que

$$g(\bar{X}) = \hat{\theta}_{MoM}, g(\mu) = \theta,$$

et

$$g'(\mu) = \frac{-1}{2\mu^2} \left[1 + \frac{1 + 3\mu}{\sqrt{(\mu - 1)^2 + 8\mu}} \right] = -\frac{\theta^2 (\theta + 1)^2}{(\theta^2 + 4\theta + 2)}.$$

■

2.1.4 Simulation

Dans ce paragraphe, on étudie le comportement des estimateurs MoM pour une taille d'échantillon n fini : La simulation est réalisée pour chaque couple $(\theta; n)$, où

$\theta = \frac{1}{3}, 1, 3$ et $n = 20, 40, 60, 80, 100$.

Alors on a l'algorithme suivant :

- Choisissez les valeurs initiales de θ_0 pour spécifier la distribution de PLD ;
- Choisir la taille de l'échantillon n ;
- Générer N échantillons indépendants de taille n de $PLD(\theta)$;
- Calculer les estimateurs $\hat{\theta}_{MV}$ de θ pour chacun des N échantillons;
- Calculer :

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

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

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

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

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

n	$\theta = \frac{1}{3}$	$\theta = 1$	$\theta = 3$
20	0.011908	0.064915	0.588291
40	0.005658	0.033095	0.230338
60	0.003531	0.021611	0.149065
80	0.002471	0.014849	0.107402
100	0.001887	0.011892	0.083455

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

n	$\theta = \frac{1}{3}$	$\theta = 1$	$\theta = 3$
20	0.004637	0.080496	4.424323
40	0.002104	0.032186	0.963864
60	0.001349	0.020220	0.538328
80	0.000999	0.014581	0.366871
100	0.00077800	0.011507	0.278360

Remarque 2.5 (i) *Le tableau 2.1 présente un biais positif, comme indiqué dans le théorème 2.1.2. Le tableau montre également que le biais diminue (augmente) quand $n(\theta)$ augmente (respectivement).*

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

2.2 Distribution de Size Baised poisson Lindley

Dans cette section, on introduit la distribution de taille biaisée Poisson Lindley, cette distribution a été introduite par Ghitany M.E. (2008) [12], et étudier ses propriétés de base. Plusieurs propriétés de cette distribution telles que les moments, la mesure de l'asymétrie, le kurtosis, la fonction génératrice de moments, la fonction génératrice de caractéristiques, le coefficient de la variation, de la fonction de survie et

de la fonction de risque qui sont dérivés pour comprendre plus brièvement la structure de la distribution proposée.

En utilisant (1.1), (1.2) et (2.1), la fonction de masse de la distribution biaisée (size biased) de Poisson Lindley (SBPL) avec le paramètres θ peut être obtenue comme suit :

$$f_{SBPL}(x, \theta) = P_1(x; \theta) = \frac{\theta^3}{\theta + 2} \frac{x(x + \theta + 2)}{(1 + \theta)^{x+2}}, \quad x = 1, 2, \dots, \theta > 0. \quad (2.6)$$

La fonction des probabilité de SBPL a été obtenu en mélangeant une distribution biaisée (size biased) de Poisson [12] avec la distribution biaisée de Lindley.

Supposons que le paramètre λ de la distribution biaisés de Poisson d'une fonction de probabilité définit par :

$$g(x | \lambda) = e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!}; \quad x = 1, 2, 3, \dots, \theta > 0. \quad (2.7)$$

Lorsque son paramètre λ suit la distribution biaisée (size biased) de Lindley (SBLD) ayant la fonction de masse suivante :

$$h(\lambda; \theta) = \frac{\theta^3}{\theta + 2} \lambda (1 + \lambda) e^{-\theta\lambda}; \quad \lambda > 0, \theta > 0. \quad (2.8)$$

Ainsi, la fonction de densité de probabilité de la distribution SBPL peut être

obtenu comme suit :

$$\begin{aligned}
P_1(x; \theta) &= \int_0^{\infty} g(x | \lambda) h(\lambda; \theta) d\lambda \\
&= \frac{\theta^3}{(\theta + 2)(x - 1)!} \int_0^{\infty} \lambda^{x-1} e^{-\lambda} \lambda(\lambda + 1) e^{-\theta\lambda} d\lambda \\
&= \frac{\theta^3}{(\theta + 2)(x - 1)!} \int_0^{\infty} (\lambda^x + \lambda^{x+1}) e^{-(1+\theta)\lambda} d\lambda \\
&= \frac{\theta^3}{\theta + 2} \left[\frac{x}{(1 + \theta)^{x+1}} + \frac{x(x + 1)}{(1 + \theta)^{x+2}} \right] \\
&= \frac{\theta^3}{\theta + 2} \frac{x(x + \theta + 2)}{(1 + \theta)^{x+2}}, \quad x = 1, 2, \dots, \theta > 0.
\end{aligned}$$

Qui est la distribution biaisée par la taille de Poisson-Lindley (SBPLD).

2.2.1 Moments et mesures connexes

Le moment factoriel d'ordre r de la distribution SBPL peut être obtenu en tant que :

$$\begin{aligned}
\mu'_{(r)} &= \mathbb{E}(\mathbb{E}(X^{(r)} | \lambda)) = \int_0^{\infty} \left[\sum_{x=1}^{\infty} x^r e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \right] \cdot h(\lambda; \theta) d\lambda \\
&= \int_0^{\infty} \left[\sum_{x=1}^{\infty} x^r e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \right] \cdot \frac{\theta^3}{\theta + 2} \lambda(1 + \lambda) e^{-\theta\lambda} d\lambda \\
&= \frac{\theta^3}{\theta + 2} \int_0^{\infty} \lambda^{r-1} \left[\sum_{x=r}^{\infty} x e^{-\lambda} \frac{\lambda^{x-r}}{(x-r)!} \right] \cdot \lambda(1 + \lambda) e^{-\theta\lambda} d\lambda.
\end{aligned}$$

En prenant $y = x - r$, on a :

$$\begin{aligned}\mu'_{(r)} &= \frac{\theta^3}{\theta + 2} \int_0^{\infty} \lambda^{r-1} \left[\sum_{y=0}^{\infty} (y+r) e^{-\lambda} \frac{\lambda^y}{y!} \right] \cdot \lambda(\lambda+1) e^{-\theta\lambda} d\lambda \\ &= \frac{\theta^3}{\theta + 2} \int_0^{\infty} \lambda^r (\lambda+r) (\lambda+1) e^{-\theta\lambda} d\lambda.\end{aligned}$$

Nous obtenons finalement une expression générale du moment factoriel du distribution SBPL comme :

$$\mu'_{(r)} = \frac{r! [(\theta + r + 2)(r + 1) + r\theta(\theta + r + 1)]}{(\theta + 2)\theta^r}. \quad (2.9)$$

En substituant $r = 1, 2, 3$ et 4 dans (2.9), les quatre premiers moments factoriels de SBPLD peuvent être obtenus comme :

$$\begin{aligned}\mu'_{(1)} &= 1 + \frac{2(\theta + 3)}{\theta(\theta + 2)}, \\ \mu'_{(2)} &= \frac{4\theta^2(3\beta - 2) + 18\theta(2\beta - 1) + 24\beta}{\theta^2(2\beta(1 + \theta) - \theta)}, \\ \mu'_{(3)} &= \frac{18\theta^2(4\beta - 3) + 96\theta(2\beta - 1) + 120\beta}{\theta^3(2\beta(1 + \theta) - \theta)}, \\ \mu'_{(4)} &= \frac{96\theta^2(5\beta - 4) + 600\theta(2\beta - 1) + 720\beta}{\theta^4(2\beta(1 + \theta) - \theta)}.\end{aligned}$$

Puis en utilisant la relation entre les moments factoriels et les moments d'origine, les quatre premiers moments d'origine de la distribution SBPL sont donnés par :

$$\begin{aligned}\mu'_1 &= 1 + \frac{2(\theta + 3)}{\theta(\theta + 2)}, \\ \mu'_2 &= 1 + \frac{6(\theta + 3)}{\theta(\theta + 2)} + \frac{6(\theta + 4)}{\theta^2(\theta + 2)}, \\ \mu'_3 &= 1 + \frac{14(\theta + 3)}{\theta(\theta + 2)} + \frac{36(\theta + 4)}{\theta^2(\theta + 2)} + \frac{24(\theta + 5)}{\theta^3(\theta + 2)}, \\ \mu'_4 &= 1 + \frac{30(\theta + 3)}{\theta(\theta + 2)} + \frac{126(\theta + 4)}{\theta^2(\theta + 2)} + \frac{240(\theta + 5)}{\theta^3(\theta + 2)} + \frac{120(\theta + 6)}{\theta^4(\theta + 2)}.\end{aligned}$$

En utilisant la relation entre les moments de l'origine et les moments centrés de la distribution SBPGaL sont obtenues comme suit :

$$\begin{aligned}\mu_2 &= \frac{2(\theta^3 + 6\theta^2 + 12\theta + 6)}{\theta^2(\theta + 2)^2}, \\ \mu_3 &= \frac{\theta^5 + 10\theta^4 + 42\theta^3 + 84\theta^2 + 72\theta + 24}{\theta^3(\theta + 2)^3}, \\ \mu_4 &= \frac{\theta^7 + 22\theta^6 + 184\theta^5 + 780\theta^4 + 1800\theta^3 + 2256\theta^2 + 1440\theta + 360}{\theta^4(\theta + 2)^4}.\end{aligned}$$

Le coefficient de variation ($C.V$), le coefficient de Skewness ($\sqrt{\beta_1}$), le coefficient de Kurtosis (β_2) de SBPLD (2.6) sont obtenues comme suit :

$$\begin{aligned}C.V &= \frac{\sqrt{2(\theta^3 + 6\theta^2 + 12\theta + 6)}}{\theta^2 + 4\theta + 6}, \\ \sqrt{\beta_1} &= \frac{\theta^5 + 10\theta^4 + 42\theta^3 + 84\theta^2 + 72\theta + 24}{\sqrt{2}(\theta^3 + 6\theta^2 + 12\theta + 6)^{\frac{3}{2}}}, \\ \beta_2 &= \frac{\theta^7 + 22\theta^6 + 184\theta^5 + 780\theta^4 + 1800\theta^3 + 2256\theta^2 + 1440\theta + 360}{2(\theta^3 + 6\theta^2 + 12\theta + 6)^2}.\end{aligned}$$

2.2.2 Fonction génératrice

La fonction génératrice de probabilité de la distribution SBPL est obtenue comme suit :

$$\begin{aligned}
 G_X(t) &= \mathbb{E}(t^X) = \sum_{x=1}^{\infty} t_1^x P(x; \theta) \\
 &= \frac{\theta^3}{(\theta + 2)} \left[\sum_{x=1}^{\infty} t^x \frac{x(x + \theta + 2)}{(1 + \theta)^{x+2}} \right] \\
 &= \frac{\theta^3}{(\theta + 2)(1 + \theta)^2} \left[\sum_{x=1}^{\infty} x^2 \left(\frac{t}{1 + \theta} \right)^x + (\theta + 2) \sum_{x=1}^{\infty} x \left(\frac{t}{1 + \theta} \right)^x \right] \\
 &= \frac{\theta^3}{(\theta + 2)(1 + \theta)^2} \frac{t(\theta + 2 - t)}{(\theta + 1 - t)^2}.
 \end{aligned}$$

La fonction génératrice de moment de la distribution SBPL (2.6) est donnée par :

$$G_X(t) = \mathbb{E}(e^{tX}) = \frac{\theta^3}{(\theta + 2)(1 + \theta)^2} \frac{e^t(\theta + 2 - e^t)}{(\theta + 1 - e^t)^2}.$$

2.2.3 Estimation du paramètre

Estimation par la méthode de maximum de vraisemblance

Soient $X_i \sim SBPLD(\theta)$, $i = \overline{1, n}$, n variables aléatoires. Soit f_x la fréquence observée dans l'échantillon correspondant à $X = x$ ($x = 1, 2, 3, \dots, k$) tel que $\sum_{x=1}^k f_x = n$, où k est la plus grande valeur observée ayant une fréquence non nulle. On a :

$$P(x; \theta) = \frac{\theta^3}{\theta + 2} \frac{x(x + \theta + 2)}{(1 + \theta)^{x+2}}, \quad x = 1, 2, \dots, \theta > 0.$$

La fonction de vraisemblance sera :

$$L(x_i; \theta) = \left(\frac{\theta^3}{\theta + 2} \right)^n \frac{1}{\sum_{x=1}^k f_x(x+2)} \prod_{x=1}^k [x^2 + x(\theta + 2)]^{f_x}.$$

La fonction de log-vraisemblance est :

$$\ln L(x; \theta) = n \ln \left(\frac{\theta^3}{\theta + 2} \right) - \sum_{x=1}^k f_x(x+2) \ln(1 + \theta) + \sum_{x=1}^k f_x \ln [x^2 + x(\theta + 2)].$$

La dérivation de $\ln L(x; \theta)$ par rapport a θ est :

$$\frac{\partial \ln L}{\partial \theta} = \frac{3n}{\theta} - \frac{n}{\theta + 2} - \frac{n(2 + \bar{X})}{(1 + \theta)} + \sum_{x=1}^k \frac{f_x}{(x + \theta + 2)}.$$

L'estimation de maximum vraisemblance (MV), $\hat{\theta}$ de θ est la solution de l'équation $\frac{\partial \ln L}{\partial \theta} = 0$ et est donnée par la solution de l'équation non linéaire.

Donc

$$\frac{\partial \ln L}{\partial \theta} = \frac{3n}{\theta} - \frac{n}{\theta + 2} - \frac{n(2 + \bar{X})}{(1 + \theta)} + \sum_{x=1}^k \frac{f_x}{(x + \theta + 2)} = 0, \quad (2.11)$$

où \bar{X} est la moyenne de l'échantillon. Ghitany et Mutairi (2008) ont montré que l'estimateur $\hat{\theta}$ de MV de θ est cohérent et asymptotiquement normal.

Estimation par la méthode des moments

Étant donné un échantillon aléatoire X_1, \dots, X_n , de la distribution SBPL, en utilisant le premier moment μ'_1 . On a :

$$\mu'_1 = \frac{\theta^2 + 4\theta + 6}{\theta(\theta + 2)}.$$

Assimiler le premier moment originale à la moyenne de l'échantillon, l'estimateur $\hat{\theta}$ de θ et donné par :

$$\hat{\theta} = \frac{2 - \bar{X} + \sqrt{\bar{X}^2 + 2\bar{X} - 2}}{\bar{X} - 1}, \quad \bar{X} > 1.$$

Notons que $\mu'_1 = \bar{X} = 1$ si et seulement si $x_i = 1$ pour tous $i = 1, 2, \dots, n$.

Théorème 2.6 [12] *L'estimateur $\hat{\theta}_{MoM}$ de θ est positivement biaisée.*

Démonstration. Soit $\hat{\theta} = \Psi(\bar{X})$, et $\Psi(t) = \frac{2-t+\sqrt{t^2+2t-2}}{t-1}$, $t > 1$.

Comme

$$\Psi''(t) = \frac{1 + 4t^3 + 3t^2 - 6t + 2(t^2 + 2t - 2)\sqrt{t^2 + 2t - 2}}{(t-1)^3(t^2 + 2t - 2)^{\frac{3}{2}}} > 0,$$

alors $\Psi(t)$ est strictement convexe.

Ainsi, par l'inégalité de Jensen, on a

$$\mathbb{E}(\Psi(\bar{X})) > \Psi[\mathbb{E}(\bar{X})].$$

Enfin, étant donné que

$$\mathbb{E}(\Psi(\bar{X})) = \Psi(\mu) = \Psi\left(\frac{\theta^2 + 4\theta + 2}{\theta(\theta + 2)}\right) = \theta,$$

on obtient

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

■

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

Théorème 2.7 [12] *L'estimateur $\hat{\theta}_{MoM}$ de θ est convergent et asymptotiquement normal :*

$$\sqrt{n} \left(\hat{\theta}_{MoM} - \theta \right) \xrightarrow{D} \mathcal{N} \left(0, \vartheta^2 (\theta) \right).$$

Où

$$\vartheta^2 (\theta) = \frac{\theta^2 (\theta + 2)^2 (\theta^3 + 6\theta^2 + 12\theta + 6)}{2 (\theta^2 + 6\theta + 6)^2}.$$

Démonstration. Convergence :

Etant donné $\mu < \infty$, $\bar{X} \xrightarrow{\mathbb{P}} \mu$. aussi, puisque $\Psi (t)$ est une fonction continue à $t = \mu$, alors $\Psi (\bar{X}) \xrightarrow{\mathbb{P}} \Psi (\mu)$, c'est-à-dire

$$\hat{\theta}_{MoM} \xrightarrow{\mathbb{P}} \theta.$$

Asymptotique normal :

Comme $\sigma^2 < \infty$, par le théorème central limite, on a :

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

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

$$\sqrt{n} (\Psi (\bar{X}) - \Psi (\mu)) \xrightarrow{D} \mathcal{N} \left(0, [\Psi' (\mu)]^2 \sigma^2 \right).$$

Enfin, étant donné que

$$\Psi (\bar{X}) = \hat{\theta}_{MoM}, \quad \Psi (\mu) = \theta,$$

et

$$\Psi'(\mu) = \frac{1 - 2\mu - \sqrt{\mu^2 + 2\mu - 2}}{(\mu - 1)^2 \sqrt{\mu^2 + 2\mu - 2}} = -\frac{\theta^2 (\theta + 2)^2}{2(\theta^2 + 6\theta + 6)}.$$

■

2.2.4 Simulation

Dans ce paragraphe, on étudie le comportement des estimateurs *MoM* pour une taille d'échantillon n fini : La simulation est réalisée pour chaque couple $(\theta; n)$, où $\theta = 0.5, 2, 8$ et $n = 20, 40, 60, 80, 100$.

Alors on a l'algorithme suivant :

- Choisissez les valeurs initiales de θ_0 pour spécifier la distribution de *SBPL* ;
- Choisir la taille de l'échantillon n ;
- Générer N échantillons indépendants de taille n de *SBPLD*(θ) ;
- Calculer les estimateurs $\hat{\theta}_{MV}$ de θ pour chacun des N échantillons ;
- Calculer :

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

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

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

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

Tableau 2.3 : les biais moyens de l'estimateur $\hat{\theta}$

n	$\theta = 0,5$	$\theta = 2$	$\theta = 8$
20	0.012466	0.119956	2.182868
40	0.006323	0.051995	0.963021
60	0.004458	0.037846	0.591916
80	0.003854	0.028590	0.424201
100	0.003365	0.023547	0.343903

Tableau 2.4 : la moyenne EQM de l'estimateur $\hat{\theta}$

n	$\theta = 0,5$	$\theta = 2$	$\theta = 8$
20	0.007049	0.311509	47.847213
40	0.003294	0.119653	14.948719
60	0.002130	0.076943	6.867637
80	0.001625	0.056354	4.081337
100	0.001276	0.044042	3.122239

Remarque 2.8 (i) **Le tableau 2.3.** présente un biais positif, comme indiqué dans le théorème 2.2.1. Le tableau montre également que le biais diminue (augmente) quand $n(\theta)$ augmente (respectivement).

(ii) **Le tableau 2.4.** montre que l'erreur quadratique moyenne diminue (augmente) lorsque $n(\theta)$ augmente respectivement.

2.3 Distribution de Poisson Pseudo Lindley

Une distribution composée de Poisson peut être obtenue en composant la distribution de Poisson et une distribution due à Pseudo Lindley.[54]

On a $dF(\lambda) = e^{\lambda\Phi}h(\lambda)B(\Phi)d\lambda$, où $h(\lambda) = (\beta - 1 - \Phi\lambda)$ et $B(\Phi) = \frac{-\Phi}{\beta}$.

Alors la distribution de Poisson composée est :

$$\begin{aligned} P_x(\Phi) &= \int_0^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} dF(\lambda) \\ &= \frac{B(\Phi)}{x!} \left[(\beta - 1) \int_0^{\infty} e^{(\Phi-1)\lambda} \lambda^x d\lambda + (-\Phi) \int_0^{\infty} e^{(\Phi-1)\lambda} \lambda^{x+1} d\lambda \right] \\ &= \frac{\Phi}{\beta} \left(\frac{1 - \beta - \Phi\beta - \Phi x}{(1 - \Phi)^{x+2}} \right). \end{aligned}$$

Alors on remplace Φ par $-\theta$ on trouve :

$$P_x(\theta) = \frac{\theta}{\beta} \left(\frac{\beta - 1 + \theta\beta + \theta x}{(\theta + 1)^{x+2}} \right).$$

La fonction de masse de la distribution de Poisson-Pseudo Lindley ($PPsLD$) est :

$$P_{PPsL}(x; \theta, \beta) = \frac{\theta}{\beta} \frac{\beta - 1 + \theta\beta + \theta x}{(\theta + 1)^{x+2}}, x = 0, 1, \dots, \theta > 0, \beta \geq 1. \quad (2.12)$$

La fonction de répartition correspondante est :

$$F_{PPL}(x) = 1 - \frac{\theta\beta + \theta x + \theta + \beta}{\beta(1 + \theta)^{x+2}}, x = 0, 1, \dots, \theta > 0, \beta \geq 1. \quad (2.13)$$

Remarque 2.9 Si $\beta = \theta + 1$, cette distribution est la distribution de Poisson Lindley.

2.3.1 Moments et Mesures connexes

Le moment factoriel d'ordre r de la distribution Poisson Pseudo Lindley peut être obtenu en tant que :

$$\begin{aligned}
 \mu'_{(r)} &= \mathbb{E}(\mathbb{E}(X^{(r)} | \lambda)) = \int_0^{\infty} \left[\sum_{x=0}^{\infty} x^r e^{-\lambda} \frac{\lambda^x}{x!} \right] \cdot h(\lambda; \theta) d\lambda \\
 &= \int_0^{\infty} \left[\sum_{x=0}^{\infty} x^r e^{-\lambda} \frac{\lambda^x}{x!} \right] \cdot \frac{\theta}{\beta} (\beta - 1 + \theta\lambda) e^{-\theta\lambda} d\lambda \\
 &= \frac{\theta}{\beta} \int_0^{\infty} \lambda^r \left[\sum_{x=r}^{\infty} e^{-\lambda} \frac{\lambda^{x-r}}{(x-r)!} \right] \cdot (\beta - 1 + \theta\lambda) e^{-\theta\lambda} d\lambda.
 \end{aligned}$$

En prenant $y = x - r$, on a :

$$\begin{aligned}
 \mu'_{(r)} &= \frac{\theta}{\beta} \int_0^{\infty} \lambda^r \left[\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} \right] \cdot (\beta - 1 + \theta\lambda) e^{-\theta\lambda} d\lambda \\
 &= \frac{\theta}{\beta} \int_0^{\infty} \lambda^r (\beta - 1 + \theta\lambda) e^{-\theta\lambda} d\lambda \\
 &= \frac{\theta}{\beta} \left[(\beta - 1) \int_0^{\infty} \lambda^r e^{-\theta\lambda} d\lambda + \theta \int_0^{\infty} \lambda^{r+1} e^{-\theta\lambda} d\lambda \right] \\
 &= \frac{\theta}{\beta} \left[\frac{r! (\beta - 1)}{\theta^{r+1}} + \frac{\theta r! (r + 1)}{\theta^{r+2}} \right] \\
 &= \frac{r!}{\beta} \left[\frac{\beta - 1}{\theta^r} + \frac{r + 1}{\theta^r} \right],
 \end{aligned}$$

nous obtenons finalement une expression générale du moment factoriel du distribution PPsL comme :

$$\mu'_{(r)} = \frac{r!(\theta + r)}{\beta\theta^r}; \quad r = 1, 2, 3, \dots \quad (2.14)$$

En substituant $r = 1, 2, 3$ et 4 dans (2.14), les quatre premiers moments factoriels

de PPsL peuvent être obtenus comme :

$$\begin{aligned}\mu'_{(1)} &= \frac{\beta + 1}{\theta\beta}, \\ \mu'_{(2)} &= \frac{2(\beta + 2)}{\theta^2\beta}, \\ \mu'_{(3)} &= \frac{6(\theta + 3)}{\theta^3\beta}, \\ \mu'_{(4)} &= \frac{24(\theta + 4)}{\theta^4\beta}.\end{aligned}$$

Puis en utilisant la relation entre les moments factoriels et les moments d'origine, les quatre premiers moments d'origine de la distribution PPsL sont donnés par :

$$\begin{aligned}\mu_1 &= \frac{\beta + 1}{\theta\beta}, \\ \mu_2 &= \frac{\theta\beta + \theta + 2\beta + 4}{\theta\beta}, \\ \mu_3 &= \frac{\beta + 1}{\theta\beta} + \frac{6(\beta + 2)}{\theta^2\beta} + \frac{6(\theta + 3)}{\theta^3\beta}, \\ \mu_4 &= \frac{\beta + 1}{\theta\beta} + \frac{14(\beta + 2)}{\theta^2\beta} + \frac{36(\theta + 3)}{\theta^3\beta} + \frac{24(\theta + 4)}{\theta^4\beta}.\end{aligned}$$

En utilisant la relation entre les moments de l'origine et les moments centrés de la distribution PPsL sont obtenues comme suit :

$$\begin{aligned}\mu_2 &= \left(\frac{\beta^2\theta + \beta^2 + \beta\theta + 2\beta - 1}{\theta^2\beta^2} + \theta + \beta + 1 \right), \\ \mu_3 &= \frac{2(\theta + 1)^4(\theta + 2) - \theta^3(\theta + 2)(\theta + 3)}{\theta^3(\theta + 2)^3}, \\ \mu_4 &= \frac{3(\theta^3 + 4\theta^2 + 6\theta + 2) + 2(\theta + 1)^5[(\theta + 3)^2 - 3] - \theta^4(\theta + 2)[(\theta + 4)^2 - 3]}{\theta^4(\theta + 2)^4}.\end{aligned}$$

Le coefficient de variation ($C.V$), le coefficient de Skewness ($\sqrt{\beta_1}$), le coefficient de Kurtosis (β_2) de PPsLD (2.12) sont obtenues comme suit :

$$C.V = \frac{\sqrt{\theta^3\beta^2 + \theta^2\beta^3 + \theta^2\beta^2 + \beta^2\theta + \beta^2 + \beta\theta + 2\beta - 1}}{\beta + 1},$$

$$\sqrt{\beta_1} = \frac{2(\theta+1)^4(\theta+2) - \theta^3(\theta+2)(\theta+3)}{[2(\theta+1)^3 - \theta^2(\theta+2)]^{\frac{3}{2}}},$$

$$\beta_2 = 3 + \frac{2(\theta+1)^5[(\theta+3)^2 - 3] - \theta^4(\theta+2)[(\theta+4)^2 - 3]}{[2(\theta+1)^3 - \theta^2(\theta+2)]^2}.$$

2.3.2 Fonction génératrice

La fonction génératrice de probabilité de la distribution PPsL est obtenue comme suit :

$$\begin{aligned} G_X(t) &= \mathbb{E}(t^X) = \sum_{x=1}^{\infty} t^X P(x; \theta, \beta) \\ &= \frac{\theta}{\beta(\theta+1)^2} \left[(\beta-1 + \theta\beta) \sum_{x=1}^{\infty} \left(\frac{t}{1+\theta}\right)^x + \theta \sum_{x=1}^{\infty} x \left(\frac{t}{1+\theta}\right)^x \right] \\ &= \frac{\theta(1-\beta)t + \theta\beta + \beta - 1}{\beta(\theta+1-t)^2}. \end{aligned}$$

La fonction génératrice de moment de la distribution PPsL (2.12) est donnée par :

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

2.3.3 Estimation

Estimation par la méthode des moments (MoM)

Étant donné un échantillon aléatoire X_1, \dots, X_n , de la distribution de Poisson Pseudo Lindley (2.12), en utilisant les moments d'ordre 1 et 2 respectivement on

a

$$\mu_1' = \frac{\beta+1}{\theta\beta}, \quad \mu_2' = \frac{\theta+2\beta+\theta\beta+4}{\beta\theta^2}.$$

On résout ce système non linéaire pour tous $\mu'_2 > 0, \mu'_1 > 0$, pour trouver les estimateurs des moments $\hat{\theta}_{MoM}$ et $\hat{\beta}_{MoM}$ de θ et β respectivement comme suit :

$$\begin{cases} \hat{\theta}_{MoM} = -1 + \frac{\sqrt{2}\mu'_1}{\sqrt{\mu'_1 - n + 2\mu'_1}} \\ \hat{\beta}_{MoM} = \frac{1}{\mu'_1\theta - 1} \end{cases} .$$

L'estimation par la méthode du maximum de vraisemblance

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

$$\ln L(x_i; \theta) = n \ln \frac{\theta}{\beta} - n(\bar{x} + 2) \ln(1 + \theta) + \sum_{i=1}^n \ln(\beta + \theta\beta - 1 + \theta x_i). \quad (2.15)$$

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

$$\frac{\partial \ln l(x_i; \theta, \beta)}{\partial \theta} = \frac{n}{\theta} - n \left(\frac{\bar{x} + 2}{1 + \theta} \right) + \sum_{i=1}^n \left(\frac{\beta + x_i}{\beta + \theta\beta - 1 + \theta x_i} \right), \quad (2.16)$$

$$\frac{\partial \ln l(x_i; \beta, \theta)}{\partial \beta} = \frac{-n}{\beta} + \sum_{i=1}^n \left(\frac{1 + \theta}{\beta + \theta\beta - 1 + \theta x_i} \right). \quad (2.17)$$

Les deux équations (2.16) et (2.17) ne peut pas résoudre directement, on doit utiliser la méthode de *Fisher Scoring*, on a :

$$\begin{bmatrix} \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta^2} & \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta \partial \beta} \\ \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta \partial \theta} & \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta^2} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial Lnl(x_i; \beta, \theta)}{\partial \theta} \\ \frac{\partial Lnl(x_i; \beta, \theta)}{\partial \beta} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \quad (2.18)$$

L'équation (2.18) peut se résoudre de façon itérative où θ_0, β_0 sont les valeurs initiales de θ, β .

Les estimateurs dans le cas de l'observation unique (disons x_i) pour le vecteur de paramètre (θ, β) peut être écrit comme suit :

$$\ln l(x; \theta, \beta) = \ln \frac{\theta}{\beta} - (x + 2) \ln(1 + \theta) + \ln(\beta + \theta\beta - 1 + \theta x).$$

$$\left\{ \begin{array}{l} \hat{\theta}_{MV} = \frac{1}{X} \\ \hat{\beta}_{MV} = \frac{X}{1-X^2} \end{array} \right. .$$

2.3.4 Simulations

Dans cette section, nous étudions le comportement des estimateurs *MoM* pour un échantillon de taille (n) finis. L'étude de la simulation est effectuée pour chaque triplet $(\theta, \beta; n)$ de la distribution *PPsLD*, où $\theta = 0.5, 1.5, 2, 3, 6$ $\beta = 1.5, 2, 3, 6$ et $n = 10, 30, 50$, avec :

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

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

Tableau 2.5 : Biais moyenne des estimations simulées

	$\theta = 0.5, \beta = 1.5$		$\theta = 1.5, \beta = 1.5$		$\theta = 3, \beta = 1.5$	
	$biais(\theta)$	$biais(\beta) \leq$	$biais(\theta)$	$biais(\beta) \leq$	$biais(\theta)$	$biais(\beta) \leq$
$n = 10$	-2.4074×10^{-2}	-0.13148	-0.082	-0.106	-0.2	-0.0875
$n = 30$	-8.0247×10^{-3}	-4.3827×10^{-2}	-2.7333×10^{-2}	-3.5333×10^{-2}	-6.6667×10^{-2}	-2.9167×10^{-2}
$n = 50$	-4.8148×10^{-3}	-2.6296×10^{-2}	-0.0164	-0.0212	-0.04	-0.0175
	$\theta = 3, \beta = 2$		$\theta = 3, \beta = 3$		$\theta = 3, \beta = 6$	
	$biais(\theta)$	$biais(\beta) \leq$	$biais(\theta)$	$biais(\beta) \leq$	$biais(\theta)$	$biais(\beta) \leq$
$n = 10$	-0.20625	-0.13438	-0.2125	-0.23125	-0.21875	-0.52813
$n = 30$	-0.06875	-4.4792×10^{-2}	-7.0833×10^{-2}	-7.7083×10^{-2}	-7.2917×10^{-2}	-0.17604
$n = 50$	-0.04125	-2.6875×10^{-2}	-0.0425	-0.04625	-0.04375	-0.10563
	$\theta = 0.5, \beta = 2$		$\theta = 2, \beta = 2$		$\theta = 6, \beta = 2$	
	$biais(\theta)$	$biais(\beta) \leq$	$biais(\theta)$	$biais(\beta) \leq$	$biais(\theta)$	$biais(\beta) \leq$
$n = 10$	-2.2222×10^{-2}	-0.17778	-0.12222	-0.14444	-0.48367	-0.12041
$n = 30$	-7.4074×10^{-3}	-5.9259×10^{-2}	-4.0741×10^{-2}	-4.8148×10^{-2}	-0.16122	-4.0136×10^{-2}
$n = 50$	-4.4444×10^{-3}	-3.5556×10^{-2}	-2.4444×10^{-2}	-2.8889×10^{-2}	-9.6735×10^{-2}	-2.4082×10^{-2}

Tableau 2.6 : Erreur quadratique moyenne des estimations simulées

	$\theta = 0.5, \beta = 1.5$		$\theta = 1.5, \beta = 1.5$		$\theta = 3, \beta = 1.5$	
	$EQM(\theta)$	$EQM(\beta) \leq$	$EQM(\theta)$	$EQM(\beta) \leq$	$EQM(\theta)$	$EQM(\beta) \leq$
$n = 10$	5.7956×10^{-3}	0.17287	0.06724	0.11236	0.4	7.6563×10^{-2}
$n = 30$	1.9319×10^{-3}	5.7625×10^{-2}	2.2413×10^{-2}	3.7453×10^{-2}	0.13333	2.5521×10^{-2}
$n = 50$	1.1591×10^{-3}	3.4575×10^{-2}	1.3448×10^{-2}	2.2472×10^{-2}	0.08	1.5313×10^{-2}
	$\theta = 3, \beta = 2$		$\theta = 3, \beta = 3$		$\theta = 3, \beta = 6$	
	$EQM(\theta)$	$EQM(\beta) \leq$	$EQM(\theta)$	$EQM(\beta) \leq$	$EQM(\theta)$	$EQM(\beta) \leq$
$n = 10$	0.42539	0.18057	0.45156	0.53477	0.47852	2.7892
$n = 30$	0.14180	6.0189×10^{-2}	0.15052	0.17826	0.15951	0.92972
$n = 50$	8.5078×10^{-2}	3.6113×10^{-2}	9.0313×10^{-2}	0.10695	9.5703×10^{-2}	0.55783
	$\theta = 0.5, \beta = 2$		$\theta = 2, \beta = 2$		$\theta = 6, \beta = 2$	
	$EQM(\theta)$	$EQM(\beta) \leq$	$EQM(\theta)$	$EQM(\beta) \leq$	$EQM(\theta)$	$EQM(\beta) \leq$
$n = 10$	4.9383×10^{-3}	0.31605	0.14938	0.20864	2.3394	0.14498
$n = 30$	1.6461×10^{-3}	0.10535	4.9794×10^{-2}	6.9547×10^{-2}	0.7798	4.8327×10^{-2}
$n = 50$	9.8765×10^{-4}	6.3210×10^{-2}	2.9877×10^{-2}	4.1728×10^{-2}	0.46788	2.8996×10^{-2}

Remarque 2.10 selon les **tableaux 2.5.** et **2.6.** on remarque que :

- le biais de $(\hat{\theta})$ et le biais de $(\hat{\beta})$ diminué si (θ, β) augmente.

2.4 Distribution Poisson Quasi Lindley

Une distribution composée de Poisson peut être obtenue en composant la distribution de Poisson et une distribution quasi Lindley. [15]

On a $dF(\lambda) = e^{\lambda\Phi}h(\lambda)B(\Phi)d\lambda$, ou $h(\lambda) = \beta + \theta\lambda$ et $B(\Phi) = \frac{-\Phi}{\beta + 1}$, alors la distribution poisson composée est :

$$\begin{aligned} P_x(\Phi) &= \int_0^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} dF(\lambda) \\ &= \frac{B(\Phi)}{x!} \left[\beta \int_0^{\infty} e^{(\Phi-1)\lambda} \lambda^x d\lambda + (-\Phi) \int_0^{\infty} e^{(\Phi-1)\lambda} \lambda^{x+1} d\lambda \right] \\ &= \frac{-\Phi}{\beta + 1} \left(\frac{\beta(1 - \Phi) - \Phi(x + 1)}{(1 - \Phi)^{x+2}} \right). \end{aligned}$$

Alors on remplace Φ par $-\theta$ on trouve :

$$P_x(\theta) = \frac{\theta}{\beta + 1} \left(\frac{\beta(1 + \theta) + \theta(x + 1)}{(1 + \theta)^{x+2}} \right).$$

Alors la fonction de masse de la distribution Poisson-Quasi Lindley est donnée :

$$f_{PQL}(x; \theta, \beta) = \frac{\theta(\beta + \theta + \beta\theta + \theta x)}{(1 + \beta)(1 + \theta)^{x+2}}, \quad x = 0, 1, \dots, \quad \theta > 0, \quad \beta > -1. \quad (2.19)$$

Remarque 2.11 Si $\theta = \beta$, cette distribution est la distribution Poisson Lindley.

On peut déterminer facilement la fonction de répartition pour PQL

$$F_{PQL}(x; \theta, \beta) = 1 - \frac{\beta + 2\theta + \beta\theta + \theta x + 1}{(1 + \beta)(1 + \theta)^{x+2}}, x = 0, 1, \dots, \theta > 0, \beta > -1. \quad (2.20)$$

2.4.1 Moments et Mesures connexes

Le moment factoriel d'ordre r de la distribution Poisson Quasi Lindley peut être obtenu en tant que :

$$\begin{aligned} \mu'_{(r)} &= \mathbb{E}(\mathbb{E}(X^{(r)} | \lambda)) = \int_0^{\infty} \left[\sum_{x=0}^{\infty} x^r e^{-\lambda} \frac{\lambda^x}{x!} \right] \cdot h(\lambda; \theta) d\lambda \\ &= \int_0^{\infty} \left[\sum_{x=0}^{\infty} x^r e^{-\lambda} \frac{\lambda^x}{x!} \right] \cdot \frac{\theta}{1 + \beta} (\beta + \theta\lambda) e^{-\theta\lambda} d\lambda \\ &= \frac{\theta}{(1 + \beta)} \int_0^{\infty} \lambda^r \left[\sum_{x=r}^{\infty} e^{-\lambda} \frac{\lambda^{x-r}}{(x-r)!} \right] \cdot (\beta + \theta\lambda) e^{-\theta\lambda} d\lambda. \end{aligned}$$

En prenant $y = x - r$, on a :

$$\begin{aligned} \mu'_{(r)} &= \frac{\theta}{(1 + \beta)} \int_0^{\infty} \lambda^r \left[\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!} \right] \cdot (\beta + \theta\lambda) e^{-\theta\lambda} d\lambda \\ &= \frac{\theta}{(1 + \beta)} \int_0^{\infty} \lambda^r (\beta + \theta\lambda) e^{-\theta\lambda} d\lambda \\ &= \frac{\theta}{(1 + \beta)} \left[\beta \int_0^{\infty} \lambda^r e^{-\theta\lambda} d\lambda + \theta \int_0^{\infty} \lambda^{r+1} e^{-\theta\lambda} d\lambda \right] \\ &= \frac{\theta}{(1 + \beta)} \left[\frac{r! \beta}{\theta^{r+1}} + \frac{\theta r! (r+1)}{\theta^{r+2}} \right] \\ &= \frac{r!}{(1 + \beta)} \left[\frac{\beta}{\theta^r} + \frac{r+1}{\theta^r} \right] \end{aligned}$$

Nous obtenons finalement une expression générale du moment factoriel du distribution PPsL comme :

$$\mu'_{(r)} = \frac{r!(\beta + r + 1)}{(1 + \beta)\theta^r}; \quad r = 1, 2, 3, \dots \quad (2.21)$$

En substituant $r = 1, 2, 3$ et 4 dans (2.21), les quatre premiers moments factoriels de PQL peuvent être obtenus comme :

$$\begin{aligned} \mu'_{(1)} &= \frac{2 + \beta}{(1 + \beta)\theta}, \\ \mu'_{(2)} &= \frac{2(\beta + 3)}{(1 + \beta)\theta^2}, \\ \mu'_{(3)} &= \frac{6(\beta + 4)}{(1 + \beta)\theta^3}, \\ \mu'_{(4)} &= \frac{24(\beta + 5)}{(1 + \beta)\theta^4}. \end{aligned}$$

Puis en utilisant la relation entre les moments factoriels et les moments d'origine, les quatre premiers moments d'origine de la distribution PQL sont donnés par :

$$\begin{aligned} \mu'_1 &= \frac{\beta + 2}{\theta(\beta + 1)}, \\ \mu'_2 &= \frac{\theta\beta + 2\theta + 2\beta + 6}{\theta^2(\beta + 1)}, \\ \mu'_3 &= \frac{\theta + 2}{\theta(\theta + 1)} + \frac{6(\beta + 3)}{(1 + \beta)\theta^2} + \frac{6(\beta + 4)}{(1 + \beta)\theta^3}, \\ \mu'_4 &= \frac{\beta + 2}{\theta(\beta + 1)} + \frac{14(\beta + 3)}{(1 + \beta)\theta^2} + \frac{36(\beta + 4)}{(1 + \beta)\theta^3} + \frac{24(\beta + 5)}{(1 + \beta)\theta^4}. \end{aligned}$$

En utilisant la relation entre les moments de l'origine et les moments centrés de la distribution PQL sont obtenues comme suit :

$$\mu_2 = \frac{2 + 4\beta + \beta^2 + \theta(\beta + 2)(1 + \beta)}{(1 + \beta)^2\theta^2},$$

$$\mu_3 = \frac{2(\theta+1)^4(\theta+2) - \theta^3(\theta+2)(\theta+3)}{\theta^3(\theta+2)^3},$$

$$\mu_4 = \frac{3(\theta^3 + 4\theta^2 + 6\theta + 2) + 2(\theta+1)^5[(\theta+3)^2 - 3] - \theta^4(\theta+2)[(\theta+4)^2 - 3]}{\theta^4(\theta+2)^4}.$$

Le coefficient de variation ($C.V$), le coefficient de Skewness ($\sqrt{\beta_1}$), le coefficient de Kurtosis (β_2) de PQL (2.19) sont obtenues comme suit :

$$C.V = \frac{\sqrt{2 + 4\beta + \beta^2 + \theta(\beta + 2)(1 + \beta)}}{\beta + 1},$$

$$\sqrt{\beta_1} = \frac{2(\theta+1)^4(\theta+2) - \theta^3(\theta+2)(\theta+3)}{[2(\theta+1)^3 - \theta^2(\theta+2)]^{\frac{3}{2}}},$$

$$\beta_2 = 3 + \frac{2(\theta+1)^5[(\theta+3)^2 - 3] - \theta^4(\theta+2)[(\theta+4)^2 - 3]}{[2(\theta+1)^3 - \theta^2(\theta+2)]^2}.$$

2.4.2 Fonction génératrice

La fonction génératrice de probabilité de la distribution PQL est obtenue comme suit :

$$\begin{aligned} G_X(t) &= \mathbb{E}(t^X) = \sum_{x=1}^{\infty} t^X P(x; \theta, \beta) \\ &= \frac{\theta}{(1+\beta)(1+\theta)^2} \left[(\beta + \theta + \theta\beta) \sum_{x=1}^{\infty} \left(\frac{t}{1+\theta}\right)^x + \theta \sum_{x=1}^{\infty} x \left(\frac{t}{1+\theta}\right)^x \right] \\ &= \frac{\theta\beta t + \theta^2}{(1+\beta)(t + \theta t - t^2)}. \end{aligned}$$

La fonction génératrice de moment de la distribution PQL (2.19) est donnée par :

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

2.4.3 Fonction de survie et Fonction de hasard

Soit

$$S_{PQL}(x) = 1 - F_{PQL}(x) = \frac{\beta + 2\theta + \beta\theta + \theta x + 1}{(1 + \beta)(1 + \theta)^{x+2}}$$

et

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

la fonction de survie et hasard respective.

2.4.4 Estimations

Estimation par le maximum de vraisemblance

Dans cette partie on discute les points d'estimation pour les paramètres de $PQL(\theta, \beta)$. Soit la fonction log-vraisemblance pour une seule observation (dite x_i) pour le vecteur de paramètre (θ, β) peut être écrit comme suit :

$$\ln l(x; \theta, \beta) = \ln \theta + \ln(\beta + \theta + \beta\theta + \theta x) - \ln(1 + \beta) - 2 \ln(1 + \theta) - x \ln(1 + \theta), \quad (2.22)$$

la dérivation de $\ln l(x; \theta, \beta)$ par rapport a θ et β est :

$$\frac{\partial \ln l(x; \theta, \beta)}{\partial \theta} = \frac{1}{\theta} + \left(\frac{1 + \beta + x}{\beta + \theta + \beta\theta + \theta x} \right) - \frac{2 + x}{1 + \theta}, \quad (2.23)$$

$$\frac{\partial \ln l(x; \theta, \beta)}{\partial \beta} = \left(\frac{1 + \theta}{\beta + \theta + \beta\theta + \theta x} \right) - \frac{1}{1 + \beta}. \quad (2.24)$$

L'estimateur de maximum de vraisemblance $\hat{\theta}$ de θ et $\hat{\beta}$ de β est obtenu par la résolution de système non linière

$$\begin{cases} \hat{\theta} = \frac{1}{x} \\ \hat{\beta} = \frac{-x}{1+x} \end{cases}$$

Estimations par la méthode des moments

Utilisant μ'_1 et μ'_2 le premier et le deuxième moment de la distribution PQL, on a

$$\begin{cases} \mu'_1 = \frac{2+\beta}{(1+\beta)\theta} \\ \mu'_2 = \frac{\theta\beta+2\theta+2\beta+6}{\theta^2(\beta+1)} \end{cases} \quad (2.25)$$

où $\mu'_2 = \mu_2 + (\mu'_1)^2$ et μ_2 est la variance. on résoudre ce système non linéaire on trouve le couple $(\hat{\theta}, \hat{\beta})$, où $(\hat{\theta}, \hat{\beta}) > 0$ pour toute $\mu_2 > 0$, $\mu'_1 > 0$. La solution du système non linéaire (2.25) donne :

$$(\mu'_2 - \mu'_1) \theta^2 - 4\mu'_1 \theta + 2 = 0 \quad \text{et} \quad \beta = \frac{2 - \theta\mu'_1}{\theta\mu'_1 - 1}.$$

il est facile d'avoir que la solution de :

$$(\mu'_2 - \mu'_1) \theta^2 - 4\mu'_1 \theta + 2 = 0$$

est

$$\theta = -\frac{1}{\mu'_1 - \mu'_2} \left(2\mu'_1 + \sqrt{2} \sqrt{\mu'_1 - \mu'_2 + 2\mu'^2_1} \right), \quad \text{si } \mu'_1 - \mu'_2 \neq 0,$$

alors

$$\begin{cases} \hat{\theta} = -\frac{1}{\mu'_1 - \mu'_2} \left(2\mu'_1 + \sqrt{2} \sqrt{\mu'_1 - \mu'_2 + 2(\mu'_1)^2} \right) \\ \hat{\alpha} = \frac{2 - \hat{\theta}\mu'_1}{\hat{\theta}\mu'_1 - 1}. \end{cases}$$

2.4.5 Simulation

Dans cette section, nous étudions le comportement des estimateurs *MoM* pour un échantillon de taille (n) finis. Une étude de simulation comporte les étapes suivantes est effectuée pour chaque triple $(\theta, \alpha; n)$, ou $\theta = 0.01, \alpha = 0.1, 0.01, 1$ et pour $\alpha = 0.5, \theta = 0.05, 1, 5$ et $n = 10, 30, 50$.

- Choisir les valeurs initiale de θ_0, α_0 pour les éléments correspondants du vecteur de $\Theta = (\theta, \alpha)$ pour spécifier la distribution *PQL*.
- choisir la taille de l'échantillon n ;
- générer N échantillon indépendantes de taille n à partir de *PQL* (θ, α) ;
- Calculer l'estimation MM $\hat{\Theta}_n$ de Θ_0 pour chacun de N échantillons ;
- Calculer la moyenne des estimateurs obtenus sur tous les N échantillon,

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

et l'erreur quadratique moyenne

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

Tableau 2.7 : Biais moyenne des estimations simulées.

	$\theta = 0.01, \beta = 0.1$		$\theta = 0.01, \beta = 0.01$		$\theta = 0.01, \beta = 1$	
	$biais(\theta)$	$biais(\beta)$	$biais(\theta)$	$biais(\beta)$	$biais(\theta)$	$biais(\beta)$
$n = 10$	0.00002556	-0.0056129	0.0000025	-0.00050509	0.000264	-0.088393
$n = 30$	0.00000852	-0.00187096	0.000000833	-0.000168363	0.000088	-0.02946433
$n = 50$	0.000005112	-0.00112258	0.0000005	-0.000101018	0.0000528	-0.0176786
	$\theta = 0.05, \beta = 0.5$		$\theta = 1, \beta = 0.5$		$\theta = 5, \beta = 0.5$	
	$biais(\theta)$	$biais(\beta)$	$biais(\theta)$	$biais(\beta)$	$biais(\theta)$	$biais(\beta)$
$n = 10$	0.0006842	-0.038232	0.02743	-0.061022	0.5756	-0.11132
$n = 30$	0.00029806	-0.012744	0.009143333	-0.02034066	0.1918666	-0.0371066
$n = 50$	0.00013684	-0.0076464	0.005486	-0.0122044	0.11512	-0.022264

Tableau 2.8 : Biais moyenne des estimations simulées

	$\theta = 0.01, \beta = 0.1$		$\theta = 0.01, \beta = 0.01$		$\theta = 0.01, \beta = 1$	
	$EQM(\theta)$	$EQM(\beta)$	$EQM(\theta)$	$EQM(\beta)$	$EQM(\theta)$	$EQM(\beta)$
$n = 10$	0.0000000066	0.000315046	$6.3 \cdot 10^{-11}$	0.0000025511	0.000000697	0.078133224
$n = 30$	0.0000000022	0.000105015	$2 \cdot 10^{-11}$	0.0000008504	0.0000002323	0.026044408
$n = 50$	0.0000000013	0.000063009	10^{-11}	0.0000005102	0.0000001394	0.015626644
	$\theta = 0.05, \beta = 0.5$		$\theta = 1, \beta = 0.5$		$\theta = 5, \beta = 0.5$	
	$EQM(\theta)$	$EQM(\beta)$	$EQM(\theta)$	$EQM(\beta)$	$EQM(\theta)$	$EQM(\beta)$
$n = 10$	0.0000046813	0.01461685	0.00752404	0.03723684	3.3131536	0.123921424
$n = 30$	0.0000015604	0.00487228	0.002508016	0.0124228	1.10438453	0.041307141
$n = 50$	0.0000009363	0.00292337	0.001504809	0.00744736	0.66263072	0.024784284

Remarque 2.12 - selon les *tableaux 2.7*, l'estimateur $\hat{\theta}$ est positivement biaisée avec $bias(\theta) \rightarrow 0$ pour $\theta \rightarrow 0$, et $\hat{\beta}$ est négativement biaisée avec $bias(\beta) \rightarrow 0$ pour $\beta \rightarrow 0$.

- selon les *tableaux 2.8*, $EQM(\theta)$ et $EQM(\beta) \rightarrow 0$ où $\theta \rightarrow 0$ et $n \rightarrow \infty$.

Chapitre 3

Distribution size biased

Poisson-Gamma Lindley et ses

applications

La qualité des procédures utilisées dans une analyse statistique dépend fortement sur le modèle de probabilité supposé ou la distribution. À cause de cela, des efforts considérables ont été déployés dans le développement de grandes classes de distributions de probabilité standard ainsi que des méthodologies statistiques pertinentes. Cependant, il reste beaucoup de problèmes importants où les données réelles ne suivent aucune des modèles de probabilité standard.

Dans ce chapitre on introduit une nouvelle distribution à deux paramètres où on va étudier quelques propriétés comme : moments, estimation du moments, estimation

du maximum de vraisemblance. Des simulations sont établies pour examiner le biais et l'erreur quadratique moyenne des estimateurs des paramètres obtenus par la méthode du maximum de vraisemblance et leur application dans l'analyse de survie.

Une application du modèle à un ensemble de données réelles est finalement présentée et comparée à l'ajustement obtenu par certaines autres distributions.

3.1 Distribution Poisson Gamma Lindley

En (2020), Nedjar et Zeghdoudi ont introduit une autre distribution discrète appelée la distribution de Poisson Gamma Lindley. On présente dans ce qui suit quelques aspects statistiques concernant cette distribution.

Une distribution composée de Poisson peut être obtenue en composant la distribution de Poisson et une distribution Gamma-Lindley.[27]

on a $dF(\lambda) = e^{\lambda\Phi}h(\lambda)B(\Phi)d\lambda$, où $h(\lambda) = 1 + \beta\lambda$ et $B(\Phi) = \frac{[-\Phi]^2}{\beta(1-\Phi)}$, alors la distribution poisson composée est (voir Sankaran (1970)) :

$$\begin{aligned} P_x(\Phi) &= \int_0^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} dF(\lambda) \\ &= \frac{B(\Phi)}{x!} \left[\int_0^{\infty} e^{(\Phi-1)\lambda}\lambda^x d\lambda + \beta \int_0^{\infty} e^{(\Phi-1)\lambda}\lambda^{x+1} d\lambda \right] \\ &= \frac{\Phi^2}{\beta(1-\Phi)} \left(\frac{1-\Phi+\beta x+\beta}{(1-\Phi)^{x+2}} \right) \end{aligned}$$

alors on remplace Φ par $-\theta$ on trouve :

$$P_x(\theta) = \frac{\theta^2}{\beta} \left(\frac{(\beta + \beta\theta - \theta)x + \beta(1 + \theta) + 1}{(1 + \theta)^{x+3}} \right)$$

alors la fonction de densité Poisson-Gamma Lindley est donnée :

$$f_{PGaL}(x; \theta, \beta) = \frac{\theta^2}{\beta} \left(\frac{(\beta + \beta\theta - \theta)x + \beta(1 + \theta) + 1}{(1 + \theta)^{x+3}} \right), \quad (3.1)$$

où $x = 0, 1, \dots$, $\theta > 0$, $\beta > \frac{\theta}{1+\theta}$.

Remarque 3.1 si $\beta = 1$, cette distribution est la distribution Poisson-Lindley.

On peut déterminer facilement la fonction de répartition pour $PGaL$

$$F_{PGaL}(x; \theta, \beta) = 1 - \left(\frac{\theta(\beta + \beta\theta - \theta)x + 2\theta^2\beta + 3\theta\beta + \beta - \theta^2}{\beta(1 + \theta)^{x+3}} \right), \quad (3.2)$$

où $x = 0, 1, \dots$, $\theta > 0$, $\beta > \frac{\theta}{1+\theta}$.

3.1.1 Fonction de survie et Fonction de hasard

Soit

$$S_{PGaL}(x) = 1 - F_{PGaL}(x) = \frac{\theta(\beta + \beta\theta - \theta)x + 2\theta^2\beta + 3\theta\beta + \beta - \theta^2}{\beta(1 + \theta)^{x+3}},$$

et

$$h_{PGaL}(x) = \frac{f_{PGaL}(x)}{1 - F_{PGaL}(x)} = \frac{\theta^2((\beta + \beta\theta - \theta)x + \beta(1 + \theta) + 1)}{\theta(\beta + \beta\theta - \theta)x + 2\theta^2\beta + 3\theta\beta + \beta - \theta^2}.$$

la fonction de survie et hasard respective.

3.1.2 Moments et mesures connexes

Proposition 3.2 Soit X_1, X_2, \dots, X_n les n variable aléatoires indépendante de la distribution $PGaL(\theta, \beta)$. Alors la fonction génératrice des moments de $S = \sum_{i=1}^n X_i$, est donnée par

$$G_S(t) = \left(\frac{\theta^2}{\beta(1+\theta)} \right)^n \left(\frac{1 + \theta\beta + \beta - e^t}{(1 + \theta - e^t)^2} \right)^n,$$

d'où

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

Corollaire 3.3 Soit $X \sim PGaL(\theta, \alpha)$, l'espérance et la variance pour X est :

$$\mathbb{E}(X) = \mu'_1 = \frac{-\theta + 2\beta(1+\theta)}{\theta\beta(1+\theta)}, \quad \mathbb{E}(X^2) = \mu'_2 = \frac{-4\theta + 8\theta\beta + 2\beta\theta^2 + 6\beta - \theta^2}{\theta^2\beta(1+\theta)}, \quad (3.3)$$

$$Var(X) = \mu_2 = \frac{2\theta^3\beta^2 - \theta^3\beta + 6\theta^2\beta^2 - \theta^2\beta - \theta^2 + 6\theta\beta^2 + 2\beta^2}{\theta^2\beta^2(1+\theta)^2}. \quad (3.4)$$

3.1.3 Estimations

Estimation par le maximum de vraisemblance

Dans cette partie on discute les points d'estimation pour les paramètres de $PGaL(\theta, \beta)$.

Soit la fonction log-vraisemblance pour une seule observation (dite x_i) pour le vecteur de paramètre (θ, β) peut être écrit comme suit :

$$\ln l(x_i; \theta, \beta) = n \ln \frac{\theta^2}{\beta} - (n\bar{X} + 3) \ln(1 + \theta) + \sum_{i=1}^n \ln [(\beta + \beta\theta - \theta)x_i + \beta(1 + \theta) + 1],$$

la dérivation de $\ln l(x_i; \theta, \beta)$ par rapport a θ et β est :

$$\frac{\partial \ln l(x_i; \theta, \beta)}{\partial \theta} = \frac{2n}{\theta} - n \left(\frac{\bar{X} + 3}{1 + \theta} \right) + \sum_{i=1}^n \frac{(\beta - 1) x_i + \beta}{(\beta + \beta\theta - \theta) x_i + \beta(1 + \theta) + 1}, \quad (3.5)$$

$$\frac{\partial \ln l(x_i; \theta, \beta)}{\partial \beta} = \frac{2n}{\beta} + \sum_{i=1}^n \frac{(1 + \theta)(1 + x_i)}{(\beta + \beta\theta - \theta) x_i + \beta(1 + \theta) + 1}. \quad (3.6)$$

Les deux équations (3.5) et (3.6) ne peut pas résoudre directement, on doit utiliser la méthode de *Fisher Scoring*, on a :

$$\begin{bmatrix} \frac{\partial^2 \ln l(x_i; \theta, \beta)}{\partial \theta^2} & \frac{\partial^2 \ln l(x_i; \theta, \beta)}{\partial \theta \partial \beta} \\ \frac{\partial^2 \ln l(x_i; \theta, \beta)}{\partial \beta \partial \theta} & \frac{\partial^2 \ln l(x_i; \theta, \beta)}{\partial \beta^2} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial \ln l(x_i; \theta, \beta)}{\partial \theta} \\ \frac{\partial \ln l(x_i; \theta, \beta)}{\partial \beta} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \quad (3.7)$$

L'équation (3.7) peut se résoudre de façon itérative où θ_0, β_0 sont les valeurs initiales de θ, β . Pour des raisons pratiques, on va utiliser l'observation singulière ou unique.

Estimations par la méthode des moments

Utilisant μ'_1 et μ'_2 le premier et le deuxième moment de la distribution PGaL, on a

$$\begin{cases} \mu'_1 = \frac{-\theta + 2\beta(1 + \theta)}{\theta\beta(1 + \theta)} \\ \mu'_2 = \frac{-4\theta + 8\theta\beta + 2\beta\theta^2 + 6\beta - \theta^2}{\theta^2\beta(1 + \theta)} \end{cases} \quad (3.8)$$

où $\mu'_2 = \mu_2 + (\mu'_1)^2$ et μ_2 est la variance. on résoudre ce système non linère on trouve le couple $(\hat{\theta}, \hat{\beta})$, où $\hat{\theta}, \hat{\beta} > 0$ pour toute $\mu_2 > 0, \mu'_1 > 0$. La solution du système non linère (3.8) donne :

$$\begin{cases} \hat{\theta} = \frac{1}{\mu'_2 - \mu'_1} \left(2\mu'_1 + \sqrt{2} \sqrt{\mu'_1 - \mu'_2 + 2(\mu'_1)^2} \right) \\ \hat{\beta} = \frac{\theta}{2\theta - \mu'_1\theta(1 + \theta) + 2} \end{cases} .$$

3.2 Distribution de Size-Biased de Poisson-Gamma

Lindley

Soit $P_0(x; \theta, \beta)$ la distribution originale définie dans (3.1) avec un support des entiers non négatifs, alors sa version biaisée à la taille (size biased) est définie par :

$$P_1(x; \theta; \beta) = \frac{xP_0(x; \theta; \beta)}{\mu'_1}, \quad (3.9)$$

où $\mu'_1 = \mathbb{E}(X) = \sum_{x=0}^{\infty} xP_0(x; \theta; \beta)$ est la moyenne de la distribution originale (*PGaLD*).

En utilisant (3.1) et (3.9), la fonction de masse de la distribution biaisée (size biased) de Poisson-Gamma Lindley (SBPGaL) avec des paramètres θ, β peut être obtenue comme suit :

$$\begin{aligned} P_1(x; \theta; \beta) &= \frac{\theta^2}{\beta - \theta + 2\beta(1 + \theta)} \frac{x((\beta + \beta\theta - \theta)x + \beta(1 + \theta) + 1)}{(\theta + 1)^{x+3}} \\ &= \frac{\theta^3}{-\theta + 2\beta(1 + \theta)} \frac{x((\beta + \beta\theta - \theta)x + \beta(1 + \theta) + 1)}{(\theta + 1)^{x+2}}. \end{aligned}$$

Alors

$$P_1(x; \theta; \beta) = \frac{\theta^3}{-\theta + 2\beta(1 + \theta)} \frac{x((\beta + \beta\theta - \theta)x + \beta(1 + \theta) + 1)}{(\theta + 1)^{x+2}}, \quad (3.10)$$

où $\mu'_1 = \frac{-\theta + 2\beta(1 + \theta)}{\beta\theta(1 + \theta)}$ et $x = 1, 2, 3, \dots, \theta > 0, \beta > \frac{\theta}{1 + \theta}$.

La figure suivante présente la forme de la fonction de probabilité pour différentes valeurs des paramètres θ et β .

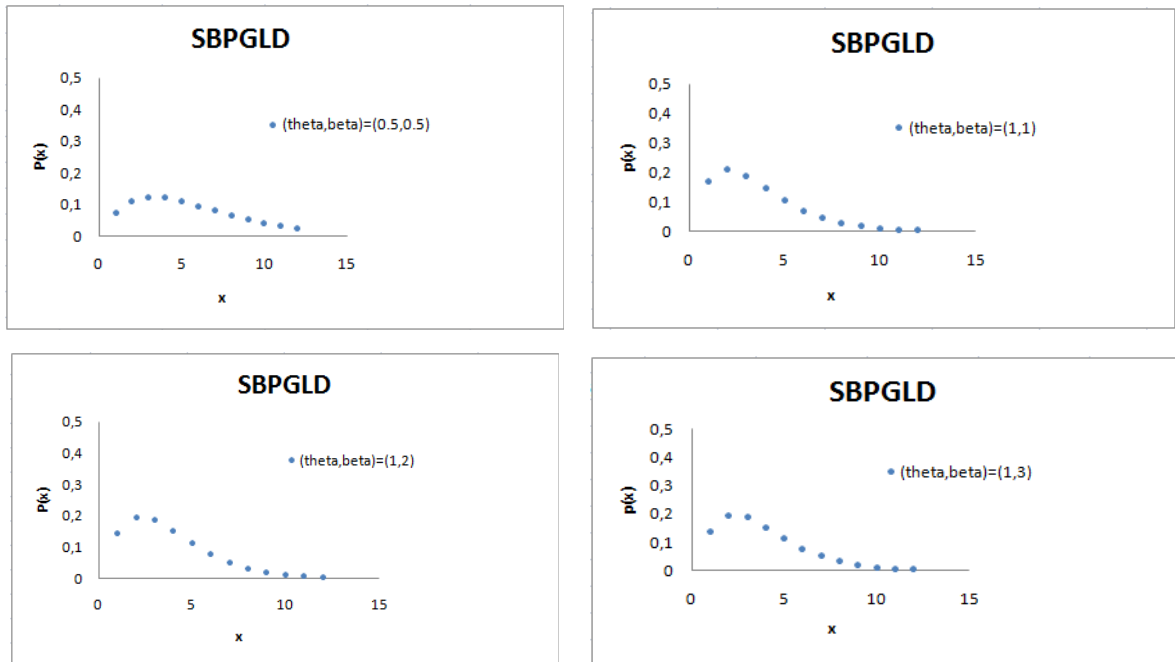


Figure.3.1. La représentation graphique de la distribution de Size biased Poisson Gamma Lindley pour différentes valeurs de θ et β .

La nature et le comportement de la distribution SBPGaL pour la variation des valeurs des paramètres et sont expliqués graphiquement à la figure (3.1). Il ressort des graphiques de la fonction des probabilités de la distribution SBPGaL que celle-ci décroît de façon monotone lorsque les valeurs des paramètres et sont croissantes.

Remarque 3.4 Lorsque $\beta = 1$, on peut déduire facilement que la distribution donnée par (3.10) revient à la distribution Poisson Lindley.

La fonction de probabilité de SBPGaL a été obtenue en mélangeant une distribution de size biased Poisson [29] (G.P. Patil, C.R. Rao, 1978) avec une distribution proposée.

Supposons que le paramètre λ de la distribution de size biased Poisson (SBP) d'une fonction de probabilité définit par :

$$g(x | \lambda) = e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!}; \quad x = 1, 2, 3, \dots, \quad \theta > 0 \quad (3.11)$$

Lorsque son paramètre λ suit la distribution de size biased Gamma Lindley (SB-GaLD) ayant la fonction de masse suivante :

$$h(\lambda; \theta, \beta) = \frac{\theta^3 \lambda ((\beta + \beta\theta - \theta)\lambda + 1) e^{-\theta\lambda}}{2\beta(1+\theta) - \theta}; \quad \lambda > 0, \quad \theta > 0, \quad \beta > \frac{\theta}{1+\theta}. \quad (3.12)$$

Ainsi, la fonction de masse de SBPGaLD est obtenue en tant que :

$$\begin{aligned} P_1(x; \theta; \beta) &= \int_0^{\infty} g(x | \lambda) h(\lambda; \theta, \beta) d\lambda \\ &= \frac{\theta^3}{(2\beta(1+\theta) - \theta)(x-1)!} \int_0^{\infty} \lambda^{x-1} e^{-\lambda} \lambda ((\beta + \beta\theta - \theta)\lambda + 1) e^{-\theta\lambda} d\lambda \\ &= \frac{\theta^3}{(2\beta(1+\theta) - \theta)(x-1)!} \int_0^{\infty} \lambda^x e^{-(1+\theta)\lambda} d\lambda + (\beta + \beta\theta - \theta) \int_0^{\infty} \lambda^{x+1} e^{-(1+\theta)\lambda} d\lambda \\ &= \frac{\theta^3}{(2\beta(1+\theta) - \theta)(x-1)!} \left[\frac{x!}{(1+\theta)^{x+1}} + \frac{(\beta + \beta\theta - \theta)x!(x+1)}{(1+\theta)^{x+2}} \right] \\ &= \frac{\theta^3}{(2\beta(1+\theta) - \theta)} \left[\frac{x(1+\theta)}{(1+\theta)^{x+2}} + \frac{(\beta + \beta\theta - \theta)x(x+1)}{(1+\theta)^{x+2}} \right] \\ &= \frac{\theta^3}{(2\beta(1+\theta) - \theta)} \left[\frac{x((\beta + \beta\theta - \theta)x + \beta(1+\theta) + 1)}{(\theta + 1)^{x+2}} \right], \end{aligned}$$

où $x = 1, 2, 3, \dots, \theta > 0, \beta > \frac{\theta}{1+\theta}$.

3.2.1 Moments et mesures connexes

Le moment factoriel d'ordre r de la distribution SBPGaL peut être obtenu en tant que :

$$\begin{aligned}
\mu'_{(r)} &= \mathbb{E}(X^{(r)} | \lambda) = \int_0^{\infty} \left[\sum_{x=1}^{\infty} x^r e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \right] \cdot h(\lambda; \theta, \beta) d\lambda \\
&= \int_0^{\infty} \left[\sum_{x=1}^{\infty} x^r e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} \right] \cdot \frac{\theta^3 \lambda ((\beta + \beta\theta - \theta)\lambda + 1) e^{-\theta\lambda}}{2\beta(1+\theta) - \theta} d\lambda \\
&= \frac{\theta^3}{2\beta(1+\theta) - \theta} \int_0^{\infty} \lambda^{r-1} \left[\sum_{x=r}^{\infty} x e^{-\lambda} \frac{\lambda^{x-r}}{(x-r)!} \right] \cdot \lambda ((\beta + \beta\theta - \theta)\lambda + 1) e^{-\theta\lambda} d\lambda \\
&= \frac{\theta^3}{2\beta(1+\theta) - \theta} \int_0^{\infty} \lambda^r \left[\sum_{x=r}^{\infty} x e^{-\lambda} \frac{\lambda^{x-r}}{(x-r)!} \right] \cdot ((\beta + \beta\theta - \theta)\lambda + 1) e^{-\theta\lambda} d\lambda.
\end{aligned}$$

En prenant $y = x - r$, on a :

$$\begin{aligned}
\mu'_{(r)} &= \frac{\theta^3}{2\beta(1+\theta) - \theta} \int_0^{\infty} \lambda^r \left[\sum_{y=0}^{\infty} (y+r) e^{-\lambda} \frac{\lambda^y}{y!} \right] \cdot ((\beta + \beta\theta - \theta)\lambda + 1) e^{-\theta\lambda} d\lambda \\
&= \frac{\theta^3}{2\beta(1+\theta) - \theta} \int_0^{\infty} \lambda^r (\lambda + r) ((\beta + \beta\theta - \theta)\lambda + 1) e^{-\theta\lambda} d\lambda \\
&= \frac{\theta^3}{2\beta(1+\theta) - \theta} \int_0^{\infty} \lambda^r ((\beta + \beta\theta - \theta)\lambda^2 + (r(\beta + \beta\theta - \theta) + 1)\lambda + r) e^{-\theta\lambda} d\lambda \\
&= \frac{r! [(\beta + \beta\theta - \theta)(r+1)(r+2) + (r(\beta + \beta\theta - \theta) + 1)\theta(r+1) + r\theta^2]}{(2\beta(1+\theta) - \theta)\theta^r}.
\end{aligned}$$

Nous obtenons finalement une expression générale du moment factoriel du distribution SBPGL comme :

$$\mu'_{(r)} = \frac{r! [(\beta + \beta\theta - \theta)(r+1)(r+2) + (r(\beta + \beta\theta - \theta) + 1)\theta(r+1) + r\theta^2]}{(2\beta(1+\theta) - \theta)\theta^r}. \quad (3.13)$$

En substituant $r = 1, 2, 3$ et 4 dans (3.13), les quatre premiers moments factoriels de SBPGaLD peuvent être obtenus comme :

$$\begin{aligned}\mu'_{(1)} &= \frac{\theta^2 (2\beta - 1) + 4\theta (2\beta - 1) + 6\beta}{\theta (2\beta(1 + \theta) - \theta)}, \\ \mu'_{(2)} &= \frac{4\theta^2 (3\beta - 2) + 18\theta (2\beta - 1) + 24\beta}{\theta^2 (2\beta(1 + \theta) - \theta)}, \\ \mu'_{(3)} &= \frac{18\theta^2 (4\beta - 3) + 96\theta (2\beta - 1) + 120\beta}{\theta^3 (2\beta(1 + \theta) - \theta)}, \\ \mu'_{(4)} &= \frac{96\theta^2 (5\beta - 4) + 600\theta (2\beta - 1) + 720\beta}{\theta^4 (2\beta(1 + \theta) - \theta)}.\end{aligned}$$

Puis en utilisant la relation entre les moments factoriels et les moments d'origine, les quatre premiers moments d'origine de la distribution SBPGaL sont donnés par :

$$\begin{aligned}\mu_1 &= \frac{\theta^2 (2\beta - 1) + 4\theta (2\beta - 1) + 6\beta}{\theta (2\beta(1 + \theta) - \theta)}, \\ \mu_2 &= \frac{\theta^3 (2\beta - 1) + 4\theta^2 (5\beta - 3) + 6\theta (7\beta - 3) + 24\beta}{\theta^2 (2\beta(1 + \theta) - \theta)}, \\ \mu_3 &= \frac{\theta^4 (2\beta - 1) + 4\theta^3 (11\beta - 7) + 6\theta^2 (31\beta - 18) + 24\theta (11\beta - 4) + 120\beta}{\theta^3 (2\beta(1 + \theta) - \theta)}, \\ \mu_4 &= \frac{\left[\begin{aligned} &\theta^5 (2\beta - 1) + 4\theta^4 (23\beta - 15) + 30\theta^3 (23\beta - 15) + 120\theta^2 (15\beta - 8) \\ &+ 120\theta (16\beta - 5) + 720\beta \end{aligned} \right]}{\theta^4 (2\beta(1 + \theta) - \theta)}.\end{aligned}$$

En utilisant la relation entre les moments de l'origine et les moments centrés de la distribution SBPGaL sont obtenues comme suit :

$$\mu_2 = \frac{2\theta^3 (6\beta^2 - 7\beta + 2) + 2\theta^2 (18\beta^2 - 13\beta + 1) + 12\theta (3\beta^2 - \beta) + 12\beta^2}{\theta^2 (2\beta(1 + \theta) - \theta)^2},$$

$$\mu_3 = \frac{\left[\begin{aligned} &2\theta^5 (12\beta^3 - 20\beta^2 + 11\beta - 2) + 2\theta^4 (72\beta^3 - 94\beta^2 + 35\beta - 3) \\ &+ 4\theta^3 (84\beta^3 - 82\beta^2 + 21\beta - 2) + 4\theta^2 (96\beta^3 - 63\beta^2 + 9\beta) \\ &+ 72\theta (3\beta^3 - \beta^2) + 48\beta^3 \end{aligned} \right]}{\theta^3 (2\beta(1 + \theta) - \theta)^3},$$

$$\mu_4 = \frac{\left[\begin{aligned} &2\theta^7 (24\beta^4 - 52\beta^3 + 41\beta^2 + 15\beta + 2) \\ &+ 2\theta^6 (9898\beta^4 - 996\beta^3 + 760\beta^2 - 253\beta + 31) \\ &+ 4\theta^5 (1200\beta^4 - 2094\beta^3 + 1273\beta^2 - 311\beta + 24) \\ &+ 8\theta^4 (1410\beta^4 - 1957\beta^3 + 871\beta^2 - 132\beta + 3) \\ &+ 16\theta^3 (905\beta^4 - 932\beta^3 + 270\beta^2 - 18\beta) \\ &+ 16\theta^2 (678\beta^4 - 449\beta^3 + 63\beta^2) \\ &+ 32\theta (130\beta^4 - 45\beta^3) + 720\beta^4 \end{aligned} \right]}{\theta^4 (2\beta(1 + \theta) - \theta)^4}.$$

Le coefficient de variation ($C.V$), le coefficient de Skewness ($\sqrt{\beta_1}$), le coefficient de Kurtosis (β_2) de SBPGaLD (3.10) sont obtenues comme suit :

$$C.V = \frac{\sqrt{2\theta^3 (6\beta^2 - 7\beta + 2) + 2\theta^2 (18\beta^2 - 13\beta + 1) + 12\theta (3\beta^2 - \beta) + 12\beta^2}}{\theta^2 (2\beta - 1) + 4\theta (2\beta - 1) + 6\beta},$$

$$\sqrt{\beta_1} = \frac{\left[\begin{aligned} &2\theta^5 (12\beta^3 - 20\beta^2 + 11\beta - 2) + 2\theta^4 (72\beta^3 - 94\beta^2 + 35\beta - 3) \\ &+ 4\theta^3 (84\beta^3 - 82\beta^2 + 21\beta - 2) + 4\theta^2 (96\beta^3 - 63\beta^2 + 9\beta) \\ &+ 72\theta (3\beta^3 - \beta^2) + 48\beta^3 \end{aligned} \right]}{(2\theta^3 (6\beta^2 - 7\beta + 2) + 2\theta^2 (18\beta^2 - 13\beta + 1) + 12\theta (3\beta^2 - \beta) + 12\beta^2)^{\frac{3}{2}}},$$

$$\beta_2 = \frac{\begin{bmatrix} 2\theta^7 (24\beta^4 - 52\beta^3 + 41\beta^2 + 15\beta + 2) \\ +2\theta^6 (9898\beta^4 - 996\beta^3 + 760\beta^2 - 253\beta + 31) \\ +4\theta^5 (1200\beta^4 - 2094\beta^3 + 1273\beta^2 - 311\beta + 24) \\ +8\theta^4 (1410\beta^4 - 1957\beta^3 + 871\beta^2 - 132\beta + 3) \\ +16\theta^3 (905\beta^4 - 932\beta^3 + 270\beta^2 - 18\beta) \\ +16\theta^2 (678\beta^4 - 449\beta^3 + 63\beta^2) \\ +32\theta (130\beta^4 - 45\beta^3) + 720\beta^4 \end{bmatrix}}{(2\theta^3 (6\beta^2 - 7\beta + 2) + 2\theta^2 (18\beta^2 - 13\beta + 1) + 12\theta (3\beta^2 - \beta) + 12\beta^2)^2}.$$

3.2.2 Propriétés statistiques de SBPGaLD

Nous donnons maintenant quelques propriétés de base de la distribution SBPGaL.

I) Pour tout x , on peut déduire que :

$$\frac{P_1(x+1; \theta; \beta)}{P_1(x; \theta; \beta)} = \left(\frac{1}{1+\theta} \right) \left(1 + \frac{1}{x} \right) \left[1 + \frac{(\beta + \beta\theta - \theta)}{(\beta + \beta\theta - \theta)x + \beta(1+\theta) + 1} \right],$$

est une fonction décroissante dans x , alors la distribution SBPGaL est unimodale.

(voir Johnson, Kotz and Kemp (2005)).

II) La fonction génératrice de probabilité de la distribution SBPGaL est obtenue comme suit :

$$\begin{aligned} G_X(t) &= \mathbb{E}(t^X) = \sum_{x=1}^{\infty} t_1^x P(x; \theta; \beta) \\ &= \frac{\theta^3}{(2\beta(1+\theta) - \theta)(\theta+1)^2} \left[(\beta + \beta\theta - \theta) \sum_{x=1}^{\infty} x^2 \left(\frac{t}{1+\theta} \right)^x + (\beta(1+\theta) + 1) \sum_{x=1}^{\infty} x \left(\frac{t}{1+\theta} \right)^x \right] \\ &= \frac{\theta^3}{(2\beta(1+\theta) - \theta)(\theta+1)^2} \left[\frac{(\beta + \beta\theta - \theta)t(\theta+1+t)(1+\theta)}{(\theta+1-t)^3} + \frac{(\beta(1+\theta)+1)t(1+\theta)}{(\theta+1-t)^2} \right] \end{aligned}$$

$$= \frac{\theta^3 t}{(2\beta(1+\theta)-\theta)(\theta+1)} \left[\frac{(\beta+\beta\theta-\theta)(\theta+1+t)}{(\theta+1-t)^3} + \frac{(\beta+\beta\theta+1)}{(\theta+1-t)^2} \right].$$

La fonction génératrice de moment de la distribution SBPGaL (3.10) est donnée par :

$$G_X(t) = \mathbb{E}(e^{tX}) = \frac{\theta^3 e^t}{(2\beta(1+\theta)-\theta)(\theta+1)} \left[\frac{(\beta+\beta\theta-\theta)(\theta+1+e^t)}{(\theta+1-e^t)^3} + \frac{(\beta+\beta\theta+1)}{(\theta+1-e^t)^2} \right].$$

3.2.3 Estimation du paramètre

Estimation par la méthode de maximum de vraisemblance

Soient $X_i \sim SBPGaLD(\theta, \beta)$, $i = \overline{1, n}$, n variables aléatoires. On a :

$$P(x; \theta, \beta) = \frac{\theta^3}{-\theta + 2\beta(1+\theta)} \frac{x((\beta + \beta\theta - \theta)x + \beta(1+\theta) + 1)}{(\theta + 1)^{x+2}},$$

où $x = 1, 2, 3, \dots$, $\theta > 0$, $\beta > \frac{\theta}{1+\theta}$.

La fonction de vraisemblance sera :

$$L(x_i; \theta, \beta) = \left(\frac{\theta^3}{2\beta(1+\theta) - \theta} \cdot \frac{1}{(\theta + 1)^2} \right)^n \prod_{i=1}^n \frac{1}{(\theta + 1)^{x_i}} \cdot x_i ((\beta + \beta\theta - \theta)x_i + \beta + \beta\theta + 1).$$

La fonction de logvraisemblance est :

$$\begin{aligned} \ln L(x; \theta, \beta) &= 3n \ln(\theta) - n \ln(2\beta(1+\theta) - \theta) - 2n \ln((1+\theta)) \\ &\quad + \sum_{i=1}^n \ln x_i + \sum_{i=1}^n \ln((\beta + \beta\theta - \theta)x_i + \beta + \beta\theta + 1) - \sum_{i=1}^n x_i \ln(1+\theta), \end{aligned}$$

la dérivation de $\ln L(x; \theta, \beta)$ par rapport a θ et β est :

$$\frac{\partial \ln L}{\partial \beta} = -\frac{2n(1+\theta)}{2\beta(1+\theta) - \theta} + \sum_{i=1}^n \frac{(1+\theta)x_i + (1+\theta)}{(\beta + \beta\theta - \theta)x_i + \beta + \beta\theta + 1} = 0, \quad (3.14)$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{3n}{\theta} + \frac{n(2\beta - 1)}{2\beta(1 + \theta) - \theta} - \frac{n(2 + \bar{X})}{(1 + \theta)} + \sum_{i=1}^n \frac{(\beta - 1)x_i + \beta}{(\beta + \beta\theta - \theta)x_i + \beta + \beta\theta + 1},$$

donc

$$\frac{\partial \ln L}{\partial \theta} = \frac{3n}{\theta} + \frac{n(2\beta - 1)}{2\beta(1 + \theta) - \theta} - \frac{n(2 + \bar{X})}{(1 + \theta)} + \sum_{i=1}^n \frac{(\beta - 1)x_i + \beta}{(\beta + \beta\theta - \theta)x_i + \beta + \beta\theta + 1} = 0, \quad (3.15)$$

où \bar{X} est la moyenne de l'échantillon.

Les estimateurs du maximum vraisemblance des deux paramètres θ et β sont les solutions de l'équation $\frac{\partial \ln L(x; \theta, \beta)}{\partial \theta} = 0$, $\frac{\partial \ln L(x; \theta, \beta)}{\partial \beta} = 0$.

Les deux équations (3.14) et (3.15) ne peut pas résoudre directement, on doit utiliser la méthode Fisher Scoring, on a :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{3n}{\theta^2} - \frac{n(2\beta - 1)^2}{(2\beta(1 + \theta) - \theta)^2} + \frac{n(2 + \bar{X})}{(1 + \theta)^2} - \sum_{i=1}^n \frac{((\beta - 1)x_i + \beta)^2}{((\beta + \beta\theta - \theta)x_i + \beta + \beta\theta + 1)^2}, \quad (3.16)$$

$$\frac{\partial^2 \ln L}{\partial \beta^2} = \frac{4n(1 + \theta)^2}{(2\beta(1 + \theta) - \theta)^2} - \sum_{i=1}^n \frac{((1 + \theta)x_i + (1 + \theta))^2}{((\beta + \beta\theta - \theta)x_i + \beta + \beta\theta + 1)^2}, \quad (3.17)$$

$$\frac{\partial^2 \ln L}{\partial \theta \partial \beta} = \frac{-2n}{(2\beta(1 + \theta) - \theta)^2} + \sum_{i=1}^n \frac{(x_i + 1)^2}{((\beta + \beta\theta - \theta)x_i + \beta + \beta\theta + 1)^2}. \quad (3.18)$$

Les équations suivantes pour $\hat{\theta}$ et $\hat{\beta}$ peuvent être résolues

$$\begin{bmatrix} \frac{\partial^2 \ln L}{\partial \theta^2} & \frac{\partial^2 \ln L}{\partial \theta \partial \beta} \\ \frac{\partial^2 \ln L}{\partial \beta \partial \theta} & \frac{\partial^2 \ln L}{\partial \beta^2} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\beta}=\beta_0}} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial \ln L}{\partial \theta} \\ \frac{\partial \ln L}{\partial \beta} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\beta}=\beta_0}}$$

Où θ_0 et β_0 sont les valeurs initiales de θ et β respectivement.

Après les simplifications, nous trouvons qu'il n'y a pas des solutions théoriques de ce système. Donc, nous utilisons les méthodes numériques pour obtenir les valeurs approximées aux estimateurs de maximum de vraisemblance $\hat{\theta}$ et $\hat{\beta}$ des paramètres θ et β respectivement.

Estimation par la méthode des moments

Étant donné un échantillon aléatoire X_1, \dots, X_n , de la distribution SBPGaL, en utilisant le premier moment μ'_1 et le deuxième moment μ'_2 . On a :

$$\left\{ \begin{array}{l} \mu'_1 = \frac{\theta^2(2\beta-1)+4\theta(2\beta-1)+6\beta}{\theta(2\beta(1+\theta)-\theta)} \\ \mu'_2 = \frac{\theta^3(2\beta-1)+4\theta^2(5\beta-3)+6\theta(7\beta-3)+24\beta}{\theta^2(2\beta(1+\theta)-\theta)} \end{array} \right., \quad (3.19)$$

où $\mu'_2 = \mu_2 + (\mu'_1)^2$ et μ_2 et la variance. Nous résolvons ce système non linéaire. On retrouve le couple $(\hat{\theta}, \hat{\beta})$, où $(\hat{\theta}, \hat{\beta}) > 0$ pour tous $\mu'_1 > 0$, $\mu_2 > 0$.

$$\mu'_1 = \frac{\theta^2(2\beta-1) + 4\theta(2\beta-1) + 6\beta}{\theta(2\beta(1+\theta) - \theta)} \rightarrow \beta = \frac{\theta^2\mu'_1 - \theta^2 + 4\theta}{2\mu'_1(\theta^2 + \theta) - 2\theta^2 - 8\theta - 6}$$

Alors

$$\hat{\beta} = \frac{\theta^2\mu'_1 - \theta^2 + 4\theta}{2\mu'_1(\theta^2 + \theta) - 2\theta^2 - 8\theta - 6}.$$

3.2.4 Simulation

Dans cette section, On procède à des simulations pour obtenir des estimateurs des paramètres de la distribution SBPGaL développés dans les sections précédentes. On va générer la distribution de SBPGaL de paramètres θ, β pour plusieurs valeurs.

Nous étudions le comportement des estimateurs MV pour un échantillon de taille (n) finis. Une étude de simulation comporte les étapes suivantes est effectuée pour chaque triple ($n; \theta, \beta$), ou $\theta = 0.1, 0.5, 5$ et $\beta = 0.2, 0.5, 2$ et $n = 20, 50, 100$.

- Choisir les valeurs initiale de θ_0, β_0 pour les éléments correspondants du vecteur de $\Theta = (\theta, \beta)$ pour spécifier la distribution SBPGaL.

- Choisir la taille de l'échantillon n ;

- Générer N échantillon indépendantes de taille n à partir de $SBPGaL(\theta, \beta)$;

- Calculer l'estimation MV $\hat{\Theta}_n$ de Θ_0 pour chacun de N échantillons;

- Calculer la moyenne des estimateurs obtenus sur tous les N échantillon,

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

et l'erreur quadratique moyenne

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

Tableau 3.1 : Biais moyenne des estimations simulées.

	$\theta = 0.1, \beta = 0.2$		$\theta = 0.1, \beta = 0.5$		$\theta = 0.1, \beta = 2$	
	<i>biais</i> (θ)	<i>biais</i> (β)	<i>biais</i> (θ)	<i>biais</i> (β)	<i>biais</i> (θ)	<i>biais</i> (β)
$n = 20$	0.002	0.085	0.086	0.0086	0.025	-0.014
$n = 50$	0.0006	0.042	2.5×10^{-2}	0.001	1.6×10^{-2}	-0.001
$n = 100$	0.0001	0.0071	4×10^{-3}	0.0006	0.01	-0.0008
	$\theta = 0.1, \beta = 2$		$\theta = 0.5, \beta = 2$		$\theta = 5, \beta = 2$	
	<i>biais</i> (θ)	<i>biais</i> (β)	<i>biais</i> (θ)	<i>biais</i> (β)	<i>biais</i> (θ)	<i>biais</i> (β)
$n = 20$	2×10^{-4}	-0.005	10^{-2}	-0.03	0.092	-0.056
$n = 50$	0.6×10^{-4}	-0.01	2×10^{-3}	-0.013	0.022	-0.01
$n = 100$	5×10^{-5}	-0.006	10^{-4}	-0.0064	0.007	-0.008
	$\theta = 5, \beta = 0.2$		$\theta = 0.5, \beta = 0.5$		$\theta = 0.1, \beta = 0.5$	
	<i>biais</i> (θ)	<i>biais</i> (β)	<i>biais</i> (θ)	<i>biais</i> (β)	<i>biais</i> (θ)	<i>biais</i> (β)
$n = 20$	0.45	-0.426	0.015	-0.0049	0.0005	-0.024
$n = 50$	0.12	-0.122	3.3×10^{-3}	-0.0029	1.2×10^{-4}	-0.073
$n = 100$	0.05	-0.067	0.0001	-0.0004	0.00002	-0.0005

Tableau 3.2 : Erreur quadratique moyenne des estimations simulées.

	$\theta = 0.1, \beta = 0.2$		$\theta = 0.1, \beta = 0.5$		$\theta = 0.1, \beta = 2$	
	<i>EQM</i> (θ)	<i>EQM</i> (β)	<i>EQM</i> (θ)	<i>EQM</i> (β)	<i>EQM</i> (θ)	<i>EQM</i> (β)
$n = 20$	0.8×10^{-4}	1.76	0.14	0.0011	0.017	0.00252
$n = 50$	2.1×10^{-5}	0.23	0.03	2.7×10^{-4}	2.1×10^{-3}	0.0013
$n = 100$	4×10^{-6}	0.03	0.002	0.00001	0.0003	0.0006
	$\theta = 0.1, \beta = 2$		$\theta = 0.5, \beta = 2$		$\theta = 5, \beta = 2$	
	<i>EQM</i> (θ)	<i>EQM</i> (β)	<i>EQM</i> (θ)	<i>EQM</i> (β)	<i>EQM</i> (θ)	<i>EQM</i> (β)
$n = 20$	1.2×10^{-6}	0.046	0.0005	0.024	0.142	0.035
$n = 50$	2.8×10^{-7}	0.012	1.1×10^{-4}	0.0075	0.051	0.015
$n = 100$	10^{-8}	0.006	0.00002	0.0003	0.008	0.005
	$\theta = 5, \beta = 0.2$		$\theta = 0.5, \beta = 0.5$		$\theta = 0.1, \beta = 0.5$	
	<i>EQM</i> (θ)	<i>EQM</i> (β)	<i>EQM</i> (θ)	<i>EQM</i> (β)	<i>EQM</i> (θ)	<i>EQM</i> (β)
$n = 20$	3.0	3.6	1.1×10^{-3}	0.001	1×10^{-5}	0.08
$n = 50$	0.5	0.7	0.4×10^{-3}	0.0003	3×10^{-6}	0.002
$n = 100$	0.06	0.08	0.00006	0.00004	2×10^{-6}	0.0004

Remarque 3.5 - Le *tableau 3.1* montre que θ est positivement biaisé et que le $\text{biais}(\theta) \rightarrow 0$, $\theta \rightarrow 0$, et β est négativement biaisé si $\beta > 0,5$ et le $\text{biais}(\beta) \rightarrow 0$, $\beta \rightarrow 0$ et $\theta \rightarrow 0$.

- Le *tableau 3.2* montre que $EQM(\theta)$ et $EQM(\beta) \rightarrow 0$, pour $\theta \rightarrow 0$ et $n \rightarrow \infty$.

3.3 Application aux ensembles de données réelles

Dans cette section. Nous illustrons l'applicabilité de la distribution SBPGaL sur l'ensemble des données réelles des individus infectés par le **virus Nipah** (*world health organization source*).

Le **tableau 3.3.** représente les données de durée de survie (par mois) de 56 individus de Kerala (état indien) infectés par le virus Nipah en 2017, où on a comparé entre la distribution de Poisson (PD), la distribution Poisson Lindley (PLD), la distribution Poisson Pseudo Lindley (PPsLD), la distribution Poisson Quasi Lindley (PQLD) et la distribution SBPGaLD.

Selon le **tableau 3.3**, on peut observer que la distribution SBPGaL donne la plus petite valeur de χ^2 lorsqu'on compare avec les autres distributions et par conséquent un meilleur ajustement des données.

<i>Nombre de particules attachées</i>	<i>Fréquence observées</i>	<i>PD</i>	<i>PLD</i>	<i>PPsLD</i>	<i>PQLD</i>	<i>SBPGaLD</i>
		$\hat{\theta} = 0.75$	$\hat{\theta} = 1.8081$	$\hat{\theta} = 6.9277$ $\hat{\beta} = 0.2383$	$\hat{\theta} = 1.5257$ $\hat{\beta} = 5.93$	$\hat{\theta} = 1.53$ $\hat{\beta} = 0.75$
0	33	26.45	31.49	23.03	31.85	32.56
1	12	19.84	14.17	25.54	13.83	13.522
2	6	7.44	6.09	6.08	5.94	5.84
3	3	1.86	2.54	1.13	2.53	2.51
4	1	0.35	1.04	0.19	1.07	1.08
5	1	0.05	0.42	0.03	0.45	0.47
<i>Total</i>	56	56	56	56	56	56
χ^2		24.1	1.289	49.405	1.057	0.756

Tableau 3.3. *Comparaisons des fréquences observées*

Les fonctions de masse des probabilités ajustés correspondants de PD, PPsLD, PQLD et SBPGaLD ainsi que les points de données d'origine pour l'ensemble de données du tableau 3.3 sont donnés à la figure 3.2.

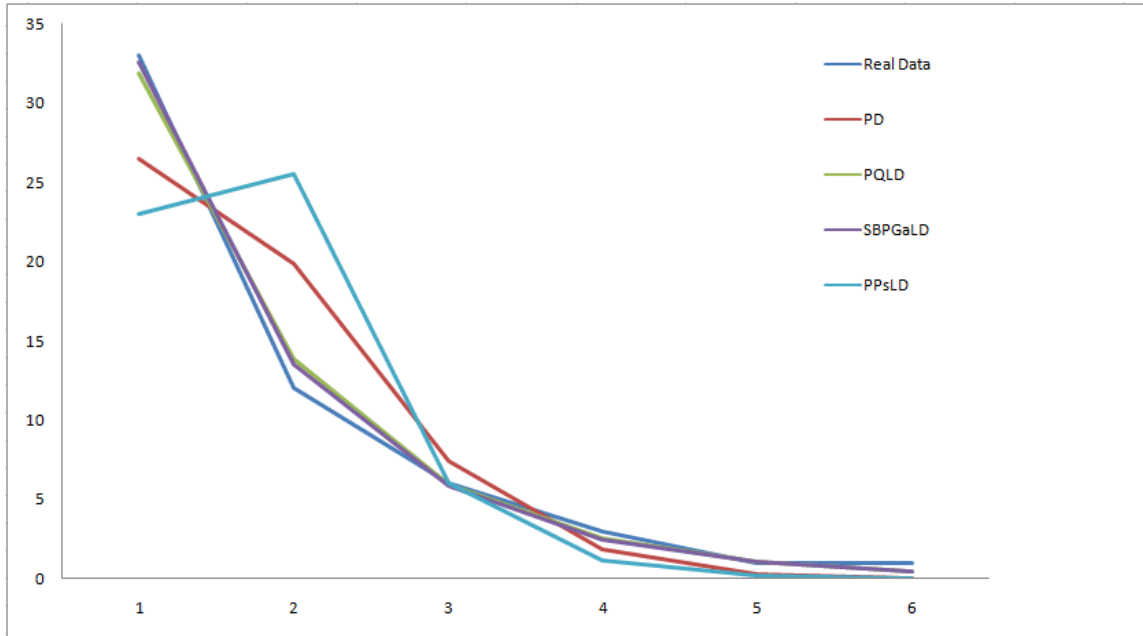


Figure 3.2. Comparaison des différents modèles *OBS*, *PD*, *PQLD*, *PPsLD* et *SBPGaLD*

Pour donner une signification à cette comparaison, on utilise les valeurs de log-vraisemblance négatives ($-LL$), le critère d'information Akaike (AIC) Le critère d'information d'Akaike corrigé ($AICc$) et le critère d'information Bayésienne (BIC) qui sont définis par :

$$AIC = -2LL + 2q \quad , \quad AICc = AIC + \frac{2q(q+1)}{n-q-1} \quad \text{et} \quad BIC = -2LL + q \log(n)$$

où q est le nombre de paramètres estimés et n est la taille de l'échantillon.[1]

	$\hat{\theta}$	$\hat{\beta}$	$-LL$	AIC	$AICc$	BIC
PD	0.75		71.58235	145.1647	145.2388	147.1901
PLD	1.8081		66.9821	135.9642	136.0383	137.9896
$PPsLD$	6.9277	0.2383	77.03363	158.0673	158.2937	162.118
$PQLD$	1.53	5.93	66.92797	137.8559	138.0824	141.9066
$SBPGaLD$	1.53	0.75	12.784	29.568	29.9442	33.6187

Tableau 3.4. *Estimations du maximum de vraisemblance, valeurs statistiques AIC , $AICc$, BIC .*

Le log-vraisemblance négatif ($-\log L$), le critère d'information Akaike (AIC), le critère d'information Akaike corrigé ($AICc$) et le critère d'information Bayésien (BIC) sont donnés aussi dans le **tableau 3.4**.

Dans le **tableau 3.4**, les valeurs du log-vraisemblance négatif sont minimales pour $SBPGaLD$ par rapport à $PQLD$, $PPsLD$, PLD et PD . Selon les valeurs de l' AIC , de l' $AICc$ et du BIC , le meilleur modèle de prévision et de génération de données plausible est le $SBPGaLD$. Notre distribution peut donc être considérée comme un modèle important pour les données de science biologique.

Chapitre 4

Nouvelle distribution tronquée à zéro et ses applications

4.1 Introduction

Les distributions tronquées sont utilisées assez efficacement lorsqu'une variable aléatoire est restreinte pour être observée sur une plage donnée. et ces situations sont courantes dans divers domaines. Par exemple, dans l'analyse de survie, les défaillances de la période de garantie peut ne pas être compté. Les articles peuvent également être remplacés après un certain temps suivant la politique de remplacement, de sorte que les défaillances de la article soient ignorés.

De nombreux chercheurs, attirés par le problème de l'analyse de telles données tronquées rencontrées dans divers disciplines, ont proposé les versions tronquées des

distributions statistiques habituelles. En (2010), S. E. Ahmed, C. Castro-Kuriss, E. Flores, V. Leiva et A. Sanhueza ont discuté de l'application de la version tronquée de la distribution de Birnbaum-Saunders (BS) pour améliorer un modèle actuariel de prévision, en particulier pour la modélisation des données provenant de paiements d'assurance établissant une franchise. Aban, Meerschaert, Panorska en (2006) et Zaninetti, M. Ferraro en (2008) ont discuté de l'application de la distribution de Pareto tronquée à l'analyse statistique de masses d'étoiles et de diamètres d'astéroïdes. La distribution de Weibull tronquée s'est avérée être appliquée dans divers domaines, par exemple pour analyser les données de diamètre des arbres, tronquer le seuil spécifique, pour prédire la distribution en hauteur de petits arbres sur la base de données de balayage laser incomplètes, pour modéliser la distribution en diamètre de la forêt, pour caractériser la distribution de la taille du feu observé au Portugal, pour des données sismologiques, sur l'évolution des profondeurs de la fosse sur une conduite d'eau, etc. En (2011) T. Zhang, M. Xie basé sur la distribution tronquée de Weibull. En (2014), S.K.Singh, U.Singh, et V.K.Sharma ont étudié la version tronquée de la distribution de Lindley, ainsi que ces caractéristiques statistiques telles que les moments, la fonction quantile et les estimateurs du maximum de vraisemblance pour différents cas de troncature.

Dans la théorie des probabilités, les distributions tronquées à zéro sont certaines distributions discrètes prenant en charge l'ensemble des entiers positifs. Les distributions zéro tronquées sont des modèles appropriés pour la modélisation de données

lorsque les données à modéliser proviennent d'un mécanisme qui génère des données à compte nul. Prenons comme des exemples, les nombres enregistrés d'infraction au code de la route pour les conducteurs au cours d'une certaine période où nous avons constaté qu'il n'y aura aucune trace de ceux qui n'ont pas reçu de billets (Rider, (1953)), le nombre de cas de choléra dans un ménage que l'appareil d'observation n'est activé que lorsqu'il y a un cas (Dahiya et Gross, (1973)).

En (2015), Shanker R, Hagos F et Sujatha S. ont mené une étude approfondie sur la comparaison entre la distribution zéro tronquée Poisson et la distribution Poisson Lindley en ce qui concerne leurs applications dans des ensembles de données excluant le zéro et ont montré qu'en démographie et en sciences biologiques, le zero tronquée Poisson Lindley donne un meilleur ajustement que le zero tonquée Poisson. tandis que dans les sciences sociales, la distribution zero tronquée de Poisson donne un meilleur ajustement que la distribution zéro tronquée Poisson Lindley.

Dans ce chapitre, nous nous proposons une nouvelle distribution de durée de survie basée sur les modèles tronqués cette distribution est nommée la distribution de Poison Quasi-Lindley tronquée à zéro. Cette distribution dépend de deux paramètres, l'un est un paramètre de forme et l'autre c'est un paramètre d'échelle.[44]

4.2 Modèles zero tronquées discrètes

Dans cette section, nous nous intéressons à quelques distribution discrètes tronqués à zéro à un seul paramètre, tel que la distribution de poisson tronqué à zéro et la

distribution de poisson lindley tronqué à zéro.

Soit $P_0(x; \theta)$ la distribution originale avec un support des entiers non négative, alors ça versson tronquée par un zéro (Zéro truncated) est définit par :

$$P(x; \theta) = \frac{P_0(x; \theta)}{1 - P_0(0; \theta)}; \quad x = 1, 2, 3, \dots \quad (4.1)$$

4.2.1 Distribution zero tronquée Poisson

En utilisant (4.1) et la fonction de masse de probabilité de la distribution de Poisson (1.3) on peut déduire la distribution zéro tronquée de Poisson (*ZTPD*) comme suit :

$$P(x; \theta) = \frac{\theta^x}{(e^\theta - 1)x!}; \quad x = 1, 2, 3, \dots, \quad \theta > 0. \quad (4.2)$$

Le moment factoriel d'ordre r de la distribution ZTP peut être obtenu en tant que :

$$\mu'_{(r)} = \frac{\theta^r e^\theta}{e^\theta - 1}. \quad (4.3)$$

En substituant $r = 1, 2, 3$ et 4 dans (4.3), les quatre premiers moments factoriels de ZTPD peuvent être obtenus et, puis en utilisant la relation entre les moments factoriels et les moments d'origine, les quatre premiers moments d'origine de la distribution ZTP

sont donnés par :

$$\begin{aligned}\mu_1' &= \frac{\theta e^\theta}{e^\theta - 1}, \\ \mu_2' &= \frac{\theta e^\theta}{e^\theta - 1} (\theta + 1), \\ \mu_3' &= \frac{\theta e^\theta}{e^\theta - 1} (\theta^2 + 3\theta + 1), \\ \mu_4' &= \frac{\theta e^\theta}{e^\theta - 1} (\theta^3 + 6\theta^2 + 7\theta + 1).\end{aligned}$$

En utilisant la relation entre les moments de l'origine et les moments centrés de la distribution ZTP sont obtenues comme suit :

$$\mu_2 = \sigma^2 = \frac{\theta e^\theta}{(e^\theta - 1)^2} (e^\theta - \theta - 1).$$

Estimation par la méthode de maximum de vraisemblance

Soient $X_i \sim ZTPD(\theta)$, $i = \overline{1, n}$, n variables aléatoires. L'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ est donné par la solution de l'équation non linéaire suivante :

$$e^\theta (\bar{X} - \theta) - \bar{X} = 0$$

où \bar{X} est la moyenne de l'échantillon.

Estimation par la méthode des moments

Étant donné un échantillon aléatoire X_1, \dots, X_n , de la distribution ZTP, en utilisant le premier moment μ_1' . L'estimateur du moment $\hat{\theta}$ de θ est donné par la solution de l'équation non linéaire suivante :

$$e^\theta (\bar{X} - \theta) - \bar{X} = 0$$

où \bar{X} est la moyenne de l'échantillon.

Ainsi, MV et MOM donnent tous deux la même estimation du paramètre θ de ZTPD (4.2).

4.2.2 Distribution zéro tronquée Poisson Lindley

À parti de (4.1) et la fonction de masse de probabilité de la distribution de Poisson Lindley (2.1) on peut déduire la distribution zéro tronquée de Poisson Lindley (*ZTPLD*) comme suit :

$$P(x; \theta) = \frac{\theta^2}{\theta^2 + 3\theta + 1} \frac{x + \theta + 2}{(\theta + 1)^x}; \quad x = 1, 2, 3, \dots, \quad \theta > 0 \quad (4.4)$$

Le moment factoriel d'ordre r de la distribution ZTPL peut être obtenu en tant que :

$$\mu'_{(r)} = \frac{r! (\theta + 1)^2 (r + \theta + 1)}{\theta^r (\theta^2 + 3\theta + 1)}; \quad r = 1, 2, 3, \dots \quad (4.5)$$

En substituant $r = 1, 2, 3$ et 4 dans (4.5), les quatre premiers moments factoriels de ZTPLD peuvent être obtenus et, puis en utilisant la relation entre les moments factoriels et les moments d'origine, les quatre premiers moments d'origine de la dis-

tribution ZTPL sont donnés par :

$$\begin{aligned}\mu_1' &= \frac{(\theta + 1)^2 (r + 2)}{\theta (\theta^2 + 3\theta + 1)} \\ \mu_2' &= \frac{(\theta + 1)^2 (\theta^2 + 4\theta + 6)}{\theta^2 (\theta^2 + 3\theta + 1)} \\ \mu_3' &= \frac{(\theta + 1)^2 (\theta^3 + 8\theta^2 + 24\theta + 24)}{\theta^3 (\theta^2 + 3\theta + 1)} \\ \mu_4' &= \frac{(\theta + 1)^2 (\theta^4 + 16\theta^3 + 78\theta^2 + 168\theta + 120)}{\theta^4 (\theta^2 + 3\theta + 1)}\end{aligned}$$

En utilisant la relation entre les moments de l'origine et les moments centrés de la distribution ZTPL sont obtenues comme suit :

$$\mu_2 = \sigma^2 = \frac{(\theta + 1)^2 (\theta^2 + 4\theta + 6)}{\theta^2 (\theta^2 + 3\theta + 1)} - \left(\frac{(\theta + 1)^2 (r + 2)}{\theta (\theta^2 + 3\theta + 1)} \right)^2$$

Estimation par la méthode de maximum de vraisemblance

Soient $X_i \sim ZTPLD(\theta)$, $i = \overline{1, n}$, n variables aléatoires. L'estimateur du maximum vraisemblance $\hat{\theta}$ de θ est donné par la solution de l'équation non linéaire suivante :

$$\frac{2n}{\theta} - \frac{n(2\theta + 3)}{\theta^2 + 3\theta + 1} - \frac{n\bar{X}}{\theta + 1} + \sum_{i=1}^n \frac{1}{x_i + \theta + 2} = 0$$

où \bar{X} est la moyenne de l'échantillon.

Estimation par la méthode des moments

Étant donné un échantillon aléatoire X_1, \dots, X_n , de la distribution ZTPL, en utilisant le premier moment μ_1' . L'estimateur du moment $\hat{\theta}$ de θ est donné par la solution de l'équation non linéaire suivante :

$$(\bar{X} - 1) \theta^3 + (3\bar{X} - 4) \theta^2 + (\bar{X} - 5) \theta - 2 = 0; \bar{X} > 1$$

où \bar{X} est la moyenne de l'échantillon.

4.3 Distribution Poisson Quasi-Lindley zéro tronquée

Dans cette section, la distribution Poisson Quasi-Lindley zéro tronquée (*ZTPQL*) a été introduite avec ses propriétés statistiques, y compris l'expression générale des probabilités, les moments, les paramètres de la distribution du Quasi-Lindley de Poisson tronquée par un zéro ont été estimées à l'aide de méthode des moments. Une application du modèle à un ensemble de données réelles est finalement présentée et comparée à l'ajustement obtenu par certaines autres distributions.

Soit $P_0(x; \theta, \beta)$ la distribution original définit dans (2.19) avec un support des entiers non négative, alors ça version tronquée à zéro est définit par :

$$P(x; \theta, \beta) = \frac{P_0(x; \theta, \beta)}{1 - P_0(0; \theta, \beta)}; x = 1, 2, .. \quad (4.6)$$

A partir de (2.19) et (4.6) , on peut déduire la distribution zero tronquée de Poisson Quasi-Lindley avec :

$$P(x; \theta, \beta) = \frac{\theta(\theta\beta + \beta + \theta + \theta x)}{(1 + 2\theta + \beta + \theta\beta)(\theta + 1)^x}; x = 1, 2, 3, \dots, \theta > 0, 1 + 2\theta + \beta + \theta\beta > 0 \quad (4.7)$$

Remarque 4.1 Lorsque $\beta = \theta$, on peut déduire facilement que la distribution donnée par (4.7) revient à la distribution zéro tronquée Poisson Lindley.

La fonction des probabilités de ZTPQLD a été obtenue en mélangeant une distribution de size-biased de Poisson [29] avec une distribution proposée. Comme le paramètre de la distribution de Poisson suppose des valeurs non négatives, il a été supposé que le paramètre de la distribution de Poisson suit une distribution continue ayant des valeurs de support non négatives.

Supposons que le paramètre λ de la distribution size-biased de Poisson d'une fonction de probabilité définit par :

$$g(x | \lambda) = \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!}; x = 1, 2, 3, \dots, \lambda > 0 \quad (4.8)$$

suit une distribution continue ayant la fonction de masse suivante :

$$h(\lambda; \theta, \beta) = \frac{\theta}{1 + 2\theta + \beta + \theta\beta} [(\theta\beta + \beta + \theta) + \theta(\theta + 1)\lambda] e^{-\lambda\theta} \quad (4.9)$$

où $\lambda > 0$, $\theta > 0$, $1 + 2\theta + \beta + \theta\beta > 0$.

Ainsi, la fonction de masse de ZTPQLD est obtenue en tant que :

$$\begin{aligned} P(X = x) &= \int_0^{\infty} g(x | \lambda) \cdot h(\lambda, \theta, \beta) d\lambda \\ &= \frac{\theta}{(1 + 2\theta + \beta + \theta\beta)(x-1)!} \int_0^{\infty} [\theta(\theta + 1)\lambda^x + (\theta\beta + \beta + \theta)\lambda^{x-1}] e^{-(\theta+1)\lambda} d\lambda \\ &= \frac{\theta}{(1 + 2\theta + \beta + \theta\beta)(x-1)!} \left[\frac{\theta(\theta + 1)\Gamma(x+1)}{(\theta + 1)^{x+1}} + \frac{(\theta\beta + \beta + \theta)\Gamma(x)}{(\theta + 1)^{x+1}} \right] \\ &= \frac{\theta}{(1 + 2\theta + \beta + \theta\beta)} \left[\frac{\theta x}{(\theta + 1)^x} + \frac{(\theta\beta + \beta + \theta)}{(\theta + 1)^x} \right] \end{aligned}$$

où $x = 1, 2, 3, \dots$, $\theta > 0$, $1 + 2\theta + \beta + \theta\beta > 0$.

La figure suivante présente la forme de la fonction de probabilité pour différents valeurs des paramètres θ et β .

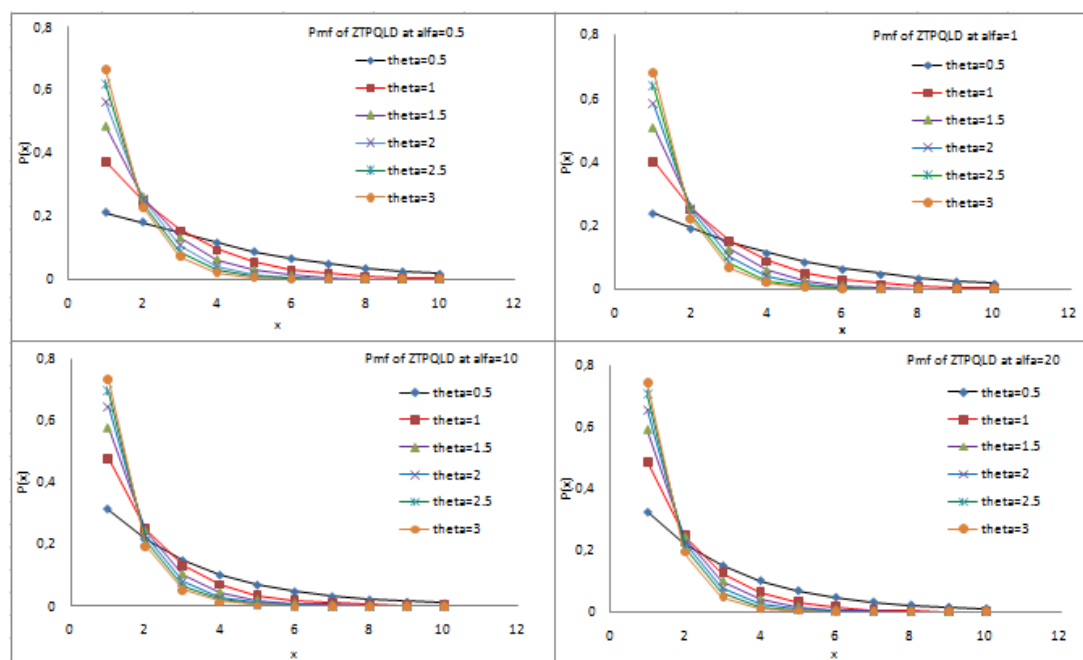


Figure.4.1. La représentation graphique de la distribution zéro tronquée Poisson Quasi Lindley pour différentes valeurs de θ et β .

4.3.1 Moments et mesures connexes

Le moment factoriel d'ordre r de la distribution ZTPQL peut être obtenu en tant que :

$$\begin{aligned}
 \mu'_{(r)} &= \int_0^{\infty} \left[\sum_{x=1}^{\infty} x^r \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!} \right] \cdot h(\lambda; \theta, \beta) d\lambda \\
 &= \int_0^{\infty} \left[\sum_{x=1}^{\infty} x^r \frac{e^{-\lambda} \lambda^{x-1}}{(x-1)!} \right] \cdot \frac{\theta}{1 + 2\theta + \beta + \theta\beta} [(\theta\beta + \beta + \theta) + \theta(\theta + 1)\lambda] e^{-\lambda\theta} d\lambda \\
 &= \frac{\theta}{1 + 2\theta + \beta + \theta\beta} \int_0^{\infty} \lambda^{r-1} \left[\sum_{x=r}^{\infty} x e^{-\lambda} \frac{\lambda^{x-r}}{(x-r)!} \right] \cdot [(\theta\beta + \beta + \theta) + \theta(\theta + 1)\lambda] e^{-\lambda\theta} d\lambda
 \end{aligned}$$

En prenant $y = x - r$, on a :

$$\begin{aligned}
 \mu'_{(r)} &= \frac{\theta}{1 + 2\theta + \beta + \theta\beta} \int_0^{\infty} \left[\lambda^{r-1} \sum_{y=0}^{\infty} (y+r) e^{-\lambda} \frac{\lambda^y}{y!} \right] \cdot [(\theta\beta + \beta + \theta) + \theta(\theta + 1)\lambda] e^{-\lambda\theta} d\lambda \\
 &= \frac{\theta}{1 + 2\theta + \beta + \theta\beta} \int_0^{\infty} \lambda^{r-1} (\lambda + r) ((\theta\beta + \beta + \theta) + \theta(\theta + 1)\lambda) e^{-\lambda\theta} d\lambda \\
 &= \frac{\theta}{1 + 2\theta + \beta + \theta\beta} \int_0^{\infty} (\lambda^r + r\lambda^{r-1}) ((\theta\beta + \beta + \theta) + \theta(\theta + 1)\lambda) e^{-\lambda\theta} d\lambda \\
 &= \frac{r!}{1 + 2\theta + \beta + \theta\beta} \left[\frac{(\theta + 1)^2 (r + \beta + 1)}{\theta^r} \right]
 \end{aligned}$$

nous obtenons finalement une expression générale du moment factoriel du distribution ZTPQL comme :

$$\mu'_{(r)} = \frac{r!}{1 + 2\theta + \beta + \theta\beta} \left[\frac{(\theta + 1)^2 (r + \beta + 1)}{\theta^r} \right] \quad (4.10)$$

En substituant $r = 1, 2, 3$ et 4 dans (4.10), les quatre premiers moments factoriels

de ZTPQL peuvent être obtenus comme :

$$\begin{aligned}\mu'_{(1)} &= \frac{(\theta + 1)^2 (\beta + 2)}{\theta (1 + 2\theta + \beta + \theta\beta)} \\ \mu'_{(2)} &= \frac{(\theta + 1)^2 (2(\beta + 3))}{\theta^2 (1 + 2\theta + \beta + \theta\beta)} \\ \mu'_{(3)} &= \frac{(\theta + 1)^2 (6(\beta + 4))}{\theta^3 (1 + 2\theta + \beta + \theta\beta)} \\ \mu'_{(4)} &= \frac{(\theta + 1)^2 (24(\beta + 5))}{\theta^4 (1 + 2\theta + \beta + \theta\beta)}\end{aligned}$$

puis en utilisant la relation entre les moments factoriels et les moments d'origine, les quatre premiers moments d'origine de la distribution ZTPQL sont donnés par :

$$\begin{aligned}\mu'_1 &= \frac{(\theta + 1)^2 (\beta + 2)}{\theta (1 + 2\theta + \beta + \theta\beta)} \\ \mu'_2 &= \frac{(\theta + 1)^2 (\theta (\beta + 2) + 2(\beta + 3))}{\theta^2 (1 + 2\theta + \beta + \theta\beta)} \\ \mu'_3 &= \frac{(\theta + 1)^2 (\theta^2 (\beta + 2) + 6\theta (\beta + 3) + 6(\beta + 4))}{\theta^3 (1 + 2\theta + \beta + \theta\beta)} \\ \mu'_4 &= \frac{(\theta + 1)^2 (\theta^3 (\beta + 2) + 14\theta^2 (\beta + 3) + 36\theta (\beta + 4) + 24(\beta + 5))}{\theta^4 (1 + 2\theta + \beta + \theta\beta)}\end{aligned}$$

En utilisant la relation entre les moments de l'origine et les moments centrés de la distribution ZTPQL sont obtenues comme suit :

$$\begin{aligned}\mu_2 &= \frac{(\theta + 1)^2 (\theta (\beta^2 + 5\beta + 6) + 2(\beta^2 + 4\beta + 2))}{\theta^2 (1 + 2\theta + \beta + \theta\beta)^2} \\ \mu_3 &= \frac{(\theta + 1)^2 \left(\begin{array}{l} \theta^3 (\beta^3 + 7\beta^2 + 16\beta + 12) + \theta^2 (4\beta^3 + 28\beta^2 + 59\beta + 38) \\ \theta (5\beta^3 + 63\beta^2 + 54\beta + 22) + (2\beta^3 + 12\beta^2 + 12\beta + 4) \end{array} \right)}{\theta^3 (1 + 2\theta + \beta + \theta\beta)^3}\end{aligned}$$

$$\mu_4 = \frac{(\theta + 1)^2 \left(\begin{array}{l} \theta^5 (\beta^4 + 9\beta^3 + 30\beta^2 + 44\beta + 24) \\ + \theta^4 (12\beta^4 + 114\beta^3 + 389\beta^2 + 572\beta + 308) \\ \theta^3 (39\beta^4 + 243\beta^3 + 907\beta^2 + 1497\beta + 686) \\ + \theta^2 (55\beta^4 + 252\beta^3 + 896\beta^2 + 1508\beta + 554) \\ \theta (36\beta^4 + 186\beta^3 + 480\beta^2 + 636\beta + 192) \\ + (9\beta^4 + 72\beta^3 + 132\beta^2 + 96\beta + 24) \end{array} \right)}{\theta^4 (1 + 2\theta + \beta + \theta\beta)^4}$$

le coefficient de variation ($C.V$), le coefficient de Skewness ($\sqrt{\beta_1}$), le coefficient de Kurtosis (β_2) de ZTPQL (4.7) sont obtenues comme suit :

$$C.V = \frac{\sqrt{\theta (\beta^2 + 5\beta + 6) + 2 (\beta^2 + 4\beta + 2)}}{(\theta + 1) (\beta + 2)}$$

$$\sqrt{\beta_1} = \frac{\left[\begin{array}{l} \theta^3 (\beta^3 + 7\beta^2 + 16\beta + 12) + \theta^2 (4\beta^3 + 28\beta^2 + 59\beta + 38) \\ \theta (5\beta^3 + 63\beta^2 + 54\beta + 22) + (2\beta^3 + 12\beta^2 + 12\beta + 4) \end{array} \right]}{(\theta + 1) (\theta (\beta^2 + 5\beta + 6) + 2 (\beta^2 + 4\beta + 2))^{\frac{3}{2}}}$$

$$\beta_2 = \frac{\left[\begin{array}{l} \theta^5 (\beta^4 + 9\beta^3 + 30\beta^2 + 44\beta + 24) \\ + \theta^4 (12\beta^4 + 114\beta^3 + 389\beta^2 + 572\beta + 308) \\ \theta^3 (39\beta^4 + 243\beta^3 + 907\beta^2 + 1497\beta + 686) \\ + \theta^2 (55\beta^4 + 252\beta^3 + 896\beta^2 + 1508\beta + 554) \\ \theta (36\beta^4 + 186\beta^3 + 480\beta^2 + 636\beta + 192) \\ + (9\beta^4 + 72\beta^3 + 132\beta^2 + 96\beta + 24) \end{array} \right]}{(\theta + 1)^2 (\theta (\beta^2 + 5\beta + 6) + 2 (\beta^2 + 4\beta + 2))^2}$$

4.3.2 Propriétés statistiques de ZTPQLD

Nous donnons maintenant quelques propriétés de base de la distribution ZTPQL.

I) Pour tout x et avec (4.7), on peut déduire que :

$$\frac{P(x+1, \theta, \beta)}{P(x, \theta, \beta)} = \frac{1}{1+\theta} \left[1 + \frac{\theta}{(\theta\beta + \beta + \theta)} \right]$$

est une fonction décroissante dans x , alors la distribution ZTPQL est unimodale. (voir Johnson, Kotz and Kemp (2005)).

II) La fonction génératrice de probabilité de la distribution ZTPQL est obtenue comme suit :

$$G_X(t) = \mathbb{E}(t^X) = \frac{\theta}{1+2\theta+\beta+\theta\beta} \left[(\theta\beta + \beta + \theta) \sum_{x=1}^{\infty} \left(\frac{t}{\theta+1} \right)^x + \theta \sum_{x=1}^{\infty} x \left(\frac{t}{\theta+1} \right)^x \right]$$

Donc, la fonction génératrice de probabilité de la distribution ZTPQL est :

$$G_X(t) = \frac{\theta t}{1+2\theta+\beta+\theta\beta} \left[\frac{(\theta\beta + \beta + \theta)}{(\theta+1-t)} + \frac{\theta(\theta+1)}{(\theta+1-t)^2} \right]$$

La fonction génératrice de moment de la ZTPQL est donc donnée par :

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

4.3.3 Estimation du paramètre

Estimation par la méthode de maximum de vraisemblance

On considère un n -échantillon X_1, X_2, \dots, X_n de la distribution tronquée définie dans (4.7). La fonction de maximum de vraisemblance est donc :

$$L(x_i; \theta, \beta) = \left(\frac{\theta}{1 + 2\theta + \beta + \theta\beta} \right)^n \prod_{i=1}^n \left(\frac{\theta\beta + \beta + \theta + \theta x_i}{(\theta + 1)^{\sum_{i=1}^n x_i}} \right)$$

Et le logarithme de la vraisemblance est :

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

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

$$\frac{\partial \ln L(x_i; \theta, \beta)}{\partial \theta} = \frac{n}{\theta} - \frac{n(\beta + 2)}{1 + 2\theta + \beta + \theta\beta} - \frac{n\bar{X}}{(\theta + 1)} + \sum_{i=1}^n \left(\frac{1 + \beta + x_i}{\theta\beta + \beta + \theta + \theta x_i} \right) \quad (4.11)$$

$$\frac{\partial \ln L(x_i; \theta, \beta)}{\partial \beta} = -\frac{n(\theta + 1)}{1 + 2\theta + \beta + \theta\beta} + \sum_{i=1}^n \left(\frac{\theta + 1}{\theta\beta + \beta + \theta + \theta x_i} \right) \quad (4.12)$$

Les estimateurs du maximum vraisemblance des deux paramètres θ et β sont les solutions de l'équation $\frac{\partial \ln L(x; \theta, \beta)}{\partial \theta} = 0$, $\frac{\partial \ln L(x; \theta, \beta)}{\partial \beta} = 0$.

Les deux équations (4.11) et (4.12) ne peut pas résoudre directement, on doit utiliser la méthode Fisher Scoring, on a :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{n}{\theta^2} + \frac{n(\beta + 2)^2}{(1 + 2\theta + \beta + \theta\beta)^2} + \frac{n\bar{X}}{(1 + \theta)^2} - \sum_{i=1}^n \frac{(x_i + \beta + 1)^2}{(\theta x_i + \beta + \beta\theta + \theta)^2} \quad (4.13)$$

$$\frac{\partial^2 \ln L}{\partial \beta^2} = \frac{n(1 + \theta)^2}{(1 + 2\theta + \beta + \theta\beta)^2} - \sum_{i=1}^n \frac{(1 + \theta)^2}{(\theta x_i + \beta + \beta\theta + \theta)^2} \quad (4.14)$$

$$\frac{\partial^2 \ln L}{\partial \theta \partial \beta} = \frac{\partial^2 \ln L}{\partial \beta \partial \theta} = \frac{-n}{(1 + 2\theta + \beta + \theta\beta)^2} - \sum_{i=1}^n \frac{x_i + 1}{(\theta x_i + \beta + \beta\theta + \theta)^2} \quad (4.15)$$

Les équations suivantes pour $\hat{\theta}$ et $\hat{\beta}$ peuvent être résolues

$$\begin{bmatrix} \frac{\partial^2 \ln L}{\partial \theta^2} & \frac{\partial^2 \ln L}{\partial \theta \partial \beta} \\ \frac{\partial^2 \ln L}{\partial \beta \partial \theta} & \frac{\partial^2 \ln L}{\partial \beta^2} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\beta}=\beta_0}} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial \ln L}{\partial \theta} \\ \frac{\partial \ln L}{\partial \beta} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\beta}=\beta_0}}$$

où θ_0 et β_0 sont les valeurs initiales de θ et β respectivement.

Après les simplifications, on trouve qu'il n'y a pas des solutions théoriques de ce système. Donc, on utilise les méthodes numérique pour obtenir les valeurs approximées aux estimateurs de maximum de vraisemblance $\hat{\theta}$ et $\hat{\beta}$ des paramètres θ et β respectivement.

Estimations par la méthode des Moments

Comme la distribution de ZTPQL définit dans (4.7) de deux paramètres θ et β à estimé, la méthode du moment est basée sur l'équation du rapport entre les deux premières probabilités et la moyenne. De (4.7) nous avons :

$$p_1 = \frac{\theta(\theta\beta + \beta + 2\theta)}{(1 + 2\theta + \beta + \theta\beta)(\theta + 1)}$$

Et

$$p_2 = \frac{\theta(\theta\beta + \beta + 3\theta)}{(1 + 2\theta + \beta + \theta\beta)(\theta + 1)^2}$$

En prenant le rapport de ces probabilités, on obtient :

$$\frac{p_2}{p_1} = \frac{(\theta\beta + \beta + 3\theta)}{(\theta + 1)(\theta\beta + \beta + 2\theta)}$$

Après une petite simplification algébrique, l'équation ci-dessus se réduit à :

$$\beta(\theta + 1) = \frac{3\theta p_1 - 2\theta(\theta + 1)p_2}{p_2(\theta + 1) - p_1} \quad (4.16)$$

Encore une fois, en simplifiant l'expression de la moyenne de ZTPQL, on obtient :

$$\beta(\theta + 1) = \frac{2\theta^2(1 - \mu) + \theta(4 - \mu) + 2}{(\mu\theta - \theta - 1)} \quad (4.17)$$

Maintenant, en assimilant les équations (4.16) et (4.17), nous obtenons une équation quadratique de θ comme :

$$A\theta^2 + B\theta + C = 0 \quad (4.18)$$

Tellque

$$\begin{cases} A = (1 - \mu)p_1 + (2 - \mu)p_2 \\ B = (\mu - 1)p_1 + (4 - \mu)p_2 \\ C = 2(p_1 - p_2) \end{cases} \quad (4.19)$$

D'après l'équation (4.18), l'estimateur de moment du paramètre θ peut être obtenu de :

$$\hat{\theta} = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \quad (4.20)$$

En remplaçant la moyenne de la population μ par la moyenne correspondante \bar{X} et p_1 et p_2 de l'échantillon des probabilités respectives de l'ensemble de données, les valeurs de A , B et C dans (4.19) peuvent être obtenues et le remplacement de ces valeurs par (4.20) donnera l'estimateur des moments $\hat{\theta}$ du paramètre θ . Après avoir obtenu l'estimateur des moments de θ , l'estimateur des moments $\hat{\beta}$ de β peut être

obtenu par :

$$\hat{\beta} = \frac{1}{\theta + 1} \left[\frac{3\theta p_1 - 2\theta(\theta + 1)p_2}{p_2(\theta + 1) - p_1} \right] \quad (4.21)$$

ou bien

$$\hat{\beta} = \frac{1}{\theta + 1} \left[\frac{2\theta^2(1 - \mu) + \theta(4 - \mu) + 2}{(\mu\theta - \theta - 1)} \right] \quad (4.22)$$

4.4 Application aux ensembles de données réelles

Une comparaison de la qualité de l'ajustement des distributions ZTP, ZTPL, ZTPPsL et ZTPQL a été établie pour un ensemble de données dû à Finney et Varley (1955), qui ont considéré un certain nombre de fleurs et le nombre d'oeufs des mouches dans chaque fleur.

Selon le **tableau 4.1**, on peut observer que la distribution ZTPQL donne la plus petite valeur de χ^2 lorsqu'on compare avec les autres distributions et par conséquent nous donne le meilleur ajustement des données.

Pour donner une signification à cette comparaison, on utilise les valeurs de log-vraisemblance négatives ($-LL$), le critère d'information Akaike (AIC) Le critère d'information d'Akaike corrigé ($AICC$) et le critère d'information Bayésienne (BIC).

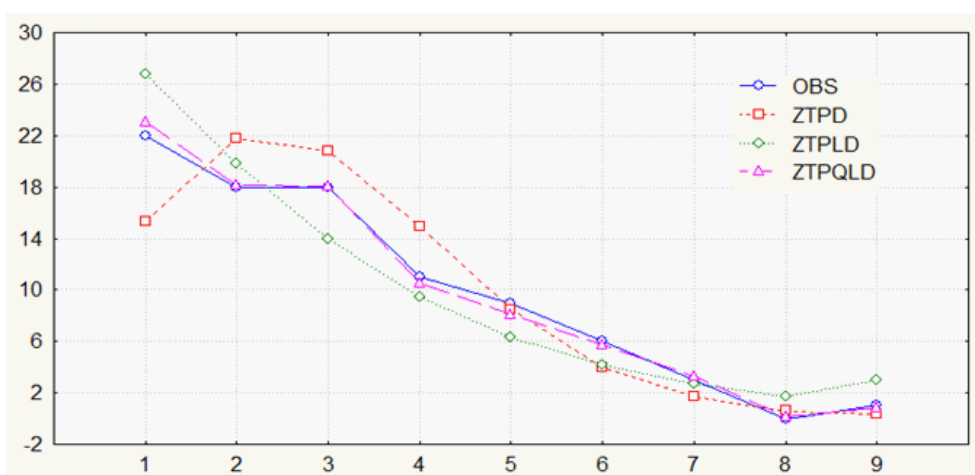
Dans le **tableau 4.1**, les valeurs du log-vraisemblance négatif sont minimales pour ZTPQLD par rapport à ZTPPsLD, ZTPLD et ZTPD. Selon les valeurs de l' AIC , de l' $AICC$ et du BIC , le meilleur modèle de prévision et de génération de données

plausible est le ZTPQLD. Notre distribution peut donc être considérée comme un modèle important pour les données de science biologique.

<i>Nombre de</i>	<i>Fréquence</i>	<i>ZTPD</i>	<i>ZTPLD</i>	<i>ZTPPsLD</i>	<i>ZTPQLD</i>
<i>particules</i>	<i>observées</i>	$\hat{\theta} = 2,8604$	$\hat{\theta} = 0,7186$	$\hat{\theta} = 1.0208$	$\hat{\theta} = 1,24$
<i>attachées</i>				$\hat{\beta} = 0.4237$	$\hat{\beta} = 0,87$
1	22	15.3	26.8	20.8	23.1
2	18	21.8	19.8	22.3	18.2
3	18	20.8	14.0	17	18.1
4	11	14.9	9.5	11.3	10.5
5	9	8.5	6.3	7.1	8.1
6	6	4.0	4.2	4.20	5.7
7	3	1.7	2.7	2.5	3.3
8	0	0.6	1.7	1.4	0.2
9	1	0.4	3.0	1	0.8
<i>Total</i>	88	88	88	88	88
χ^2		6.648	3.780	3.744	0.443
$-2 \log L$		333.09	334.76	165.24	123.34
<i>AIC</i>		336.76	335.09	334.48	142.43
<i>AICC</i>		335.14	336.80	334.62	142.57
<i>BIC</i>		335.04	336.70	334.37	132.29

Tableau 4.1 : Comparaisons des fréquences observées.

Les fonctions de masse des probabilités ajustés correspondants de ZTPD, ZTPLD et ZTPQLD ainsi que les points de données d'origine pour l'ensemble de données du tableau 4.1 sont donnés à la figure 4.2.

**figure 4.2.** Comparaison des différents modèles *OBS*, *ZTPD*, *ZTPLD* et *ZTPQLD*

Conclusion et Perspectives

Ainsi, nous avons réussi à introduire deux nouvelles distributions nommées **Size Baised Poisson Gamma Lindley (SBPGaL)** et **Poisson Quasi-Lindley zéro tronquée (ZTPQLD)**. L'estimation de ses paramètres a été discutée en utilisant la méthode du maximum de vraisemblance et la méthode des moments. Nous avons réalisé des simulations pour examiner le biais et l'erreur quadratique moyenne des estimateurs du maximum de vraisemblance des paramètres ce qui nous a permis de montrer l'efficacité de nos deux distributions. Plusieurs applications du modèle à un ensemble de données réelles sont enfin présentées et comparées à l'ajustement obtenu par quelques autres paramètres bien connus.

Nous pourrions dans nos recherches futures utiliser d'autres distributions à savoir :

- Poisson XLindley processus distribution with applications.
- Size baised Poisson Pseudo Lindley distribution and its applications.
- Zero-truncated Poisson Gamma-Lindley distribution and its applications.

Bibliographie

- [1] **Akaike H. (1974)**, A new look at statistical model identification, IEEE Transactions on Automatic Control AU-19 : 716-722.
- [2] **Alavi SMR, Chinipardaz R (2009)**, Form-invariance under weighted sampling Statistics 43 (1), 81-90
- [3] **Ayesha A. (2017)**, Size biased Lindley distribution and its properties a special case of weighted distribution. Appl. Math. 8(6) : 808-819.
- [4] **Beghriche A., Zeghdoudi H. (2019)**, A Size Biased Gamma Lindley Distribution. Thailand Statistician; 17(2) : 179-189.
- [5] **Dahiya R. C., A. J. Gross. (1973)**, Estimating the Zero Class from a Truncated Poisson Sample. Journal of the American Statistical Association, 68, 731–733.
- [6] **David F. N., Johnson L.N. (1952)**, The truncated Poisson. Biometrics 8 : 272-285.
- [7] **Ducey MJ, Gove JH (2015)**, Size-biased distributions in the generalized beta distribution family with applications to forestry. Int. J. Forest Res. 88 : 143-151.

- [8] **Finney D.J., Varley G.C. (1955)**, An example of the truncated Poisson distribution. *Biometric*, 11, 387-394.
- [9] **Fisher R.A (1937)**, The effects of methods of ascertainment upon the estimation of frequency. *Ann. Eugenics*. 6(1) : 13-25.
- [10] **Ghitany M. E., Atieh B., Nadarajah S. (2008a)**, Lindley distribution and its applications. *Math. Comput. Simulation* , 78, pp. 493-506.
- [11] **Ghitany M. E., Al-Mutairi D. K., Nadarajah S. (2008b)**, Zero-truncated Poisson-Lindley distribution and its applications. *Mathematics and Computers in Simulation*, 79(3), 279-287.
- [12] **Ghitany M.E., Al-Mutairi, D.K. (2008)**, Size-biased Poisson-Lindley distribution and Its Applications, *Metron - International Journal of Statistics*, LXVI (3), 299-311.
- [13] **Ghitany M.E., Almutairi D.K. (2009)**, Estimation methods for the discrete Poisson-Lindley distribution, *Journal of Statistical Computation and Simulation*, 79 :1, 1-9.
- [14] **Ghouar A., Zeghdoudi H., Bouras M,C (2019)**, New Zero -Truncated Distribution : Properties and Applications. *Journal of modern applied statistical methods JMASM*. Vol 18 issue 2.
- [15] **Grine R., Zeghdoudi H (2017)**, On Poisson Quasi-Lindley Distribution and its Applications. *Journal of Modern Applied Statistical Methods*, 16(2), 403-417.

- [16] **Jodrá P. (2010)**, Computer generation of random variables with Lindley or Poisson-Lindley distribution via the Lambert W function. *Mathematics and Computers in Simulation*, 81(4), 851-859.
- [17] **Johnson N L, Kotz S and Kemp A W.(2005)**, *Univariate Discrete Distribution*, third edition, John Wiley, New York.
- [18] **Khurshid A. Mir and Munir Ahmad (2009)**, Sise-Biazed distributions and their applications, *Pakistan Journal of Statistics*, Vol. 25(3), 283-294.
- [19] **Lappi J. , Bailey R. L. (1987)**, Estimation of the Diameter Increment Function or Other Tree Relations Using Angle-Count Samples. *Forest Science* 33(3) :725-739.
- [20] **Lecoutre J.P. (2002)**, *Statistique et probabilités*. Dunod. Paris, 2^{ième} édition, ISBN 2-10-007277-3.
- [21] **Lejeune M. (2010)**, *Statistique La théorie et ses applications*. Springer-Verlag France, Paris, 2^{ième} édition. ISBN 978-2-8178-0156-8.
- [22] **Lémeray E. M. (1897)**, Racines de quelques équations transcendentes. Intégration d'une équation aux différences mêlées. Racines imaginaires, *Nouvelles Ann. Math.* 16 540-546.
- [23] **Lindley D. V. (1958)**, Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1), 102-107.

- [24] **Mathews JNS., Appleton DR. (1993)**, An application of the truncated Poisson distribution to Immunogold assay. *Biometrics* 49(2), 617-621.
- [25] **Munir A. and A.(1968)**, The maximum likelihood estimation of right hand truncated Poisson distribution. Unpublished paper. Ames, Iowa, Department of Statistics, Iowa State University.
- [26] **Munir A. (1968)**, Truncated multivariate Poisson distributions, Iowa state university of science and technology, Ames, Iowa.
- [27] **Nedjar S., Zeghdoudi H. (2020)**, New compound Poisson distribution and Study its Inflated Distribution. *Thailand Statistician*, (submitted).
- [28] **Patil GP, Rao CR (1977)**, The weighted distributions : a survey and their applications. In : Krishnaiah PR, editor. *Applications of statistics*. Amsterdam : North Holland ; p.383-405.
- [29] **Patil GP, Rao CR (1978)**, Weighted distributions and size-biased sampling with applications to wild life populations and human families. *Biometrics* ; 34(2) : 179-189.
- [30] **Rao CR (1965)**, On discrete distributions arising out of methods of ascertainment. In : Patil GP, editor. *Classical and contagious discrete distributions*. Calcutta : Statistical Publishing Society ; p.320-332 ;
- [31] **Rider P. R. (1953)**, Truncated Poisson Distributions. *Journal of the American Statistical Association*, 48, 826–830.

- [32] **Sankaran M. (1970)**, The discrete Poisson-Lindley distribution. *Biometrics*, 26, pp. 145-149.
- [33] **Saporta G. (2006)**, Probabilités analyse des données et statistiques. Editions Technip. Paris, 2^{ième} édition. ISBN 2-7108-0814-5.
- [34] **Schwarz G. (1978)**, Estimating the dimension of a model, *Annals of Statistics* 6 : 461-464.
- [35] **Shanker R., Mishra A (2015)**, On Size- Biased Poisson-Lindley Distribution and Its Applications to Model Thunderstorms; *American Journal of Mathematics and Statistics* , 5(6) : 354-360.
- [36] **Shanker R., Hagos F., Sujatha, S. (2015)**, On Zero-Truncated of Poisson and Poisson-Lindley distribution and their applications, 11(6),1-14.
- [37] **Shanker R. (2016)**, A Size-Biased Poisson-Amarendra Distribution and Its Applications; *International Journal of Statistics and Applications*, 6(6) : 376-385
- [38] **Shanker R. (2017)**, A Size-Biased Poisson-Shanker Distribution and Its Applications, *International Journal of Probability and Statistics* , 6(3) : 33-44
- [39] **Shanker R., Shukla KK (2017)**, Size-Biased Poisson-Garima Distribution with Applications; *BiomBiostatInt J* 2017, 6(3) : 00167
- [40] **Shanker R., Mishra A. (2017)**, On Size- Biased Two Parameter Poisson-

- Lindley Distribution and Its Applications; American Journal of Mathematics and Statistics , 7(3) : 99-107.
- [41] **Shanker R., Shukla K.K (2018)**, A Generalized Size-Biased Poisson-Lindley Distribution and Its Applications to Model Size Distribution of Freely-Forming Small Group; J. Sci. Res. 10(2), 145-157.
- [42] **Shanker R., Shukla K.K (2019)**, A generalization of zero-truncated Poisson-Sujatha distribution. Hungarian Statistical Review, 2 (2). pp. 32-50.
- [43] **Seghier F.Z., Zeghdoudi H, Benchaabane A. (2020)**, A new Size- Biased approach in generalizing Poisson Lindley's distribution with application. Journal of Modern Applied Statistical Methods, 19(2).
- [44] **Seghier F.Z., Zeghdoudi H, Berkane A. (2020)**, A Zero-Truncated Poisson Quasi-Lindley distribution with Application. International Journal of Agricultural and Statistical Sciences. Vol.16, No.2, December 2020.
- [45] **Sheldon M. Ross (1987)**, Initiation aux probabilités. Presses polytechniques romandes 1ière édition, ISBN 2-88074-118-8.
- [46] **Singh.S.K, U.Singh, V.K.Sharma (2014)**, The Truncated Lindley Distribution : Inference and Application. Journal of Statistics Applications and Probability. 2 : 219-228.
- [47] **Tanka Raj Adhikari, R.S. Srivastava (2014)**, Size-biased discrete two parameter Poisson-Lindley Distribution for modeling and waiting survival times data; IOSR Journal of Mathematics (IOSR-JM); PP 39-45.

- [48] **Thionet P. (1967)**, Application des nombres de Stirling de 2e espèce à un problème de sondage *Revue de statistique appliquée*, tome 15, no 4 , p. 35-46
- [49] **Van Deusen J. (1986)**, Occupational Therapy Theory : Values Held by Undergraduate Students. *The Occupational Therapy Journal of Research*, Volume : 6 issue : 2, page(s) : 115-122.
- [50] **Zaninetti L., Ferraro M. (2008)**, On the truncated Pareto distribution with applications, *Central European Journal of Physics* 6. 1–6.
- [51] **Zhang.T, M. Xie. (2011)**, On the upper truncated Weibull distribution and its reliability implications, *Reliability Engineering and System Safety* 96. 194–200.
- [52] **Zeghdoudi H., Nedjar S. (2016a)**, Gamma Lindley distribution and its application. *Journal of Applied Probability and Statistics* Vol. 11, N° 1, 129-138
- [53] **Zeghdoudi H., Nedjar S. (2016c)**, On Gamma Lindley distribution :properties and simulation. *Journal of Computational and Applied Mathematics* 298, pp167-174.
- [54] **Zeghdoudi H. , Nedjar S. (2017)**. A Poisson pseudo Lindley distribution and its application . *Journal of Probability and Statistical Sciences* . Vol. 15, N° 1.