

وزارة التعليم العالي والبحث العلمي

Université 20 Aout 1955 de Skikda

Faculté des Sciences

Département de Mathématiques



جامعة 20 أوت 1955 ، سكيكدة

كلية العلوم

قسم الرياضيات

N° : U.S/F.S/D.M/...../2022.

Faculté des Sciences
Département de Mathématiques

Mémoire

Présenté en vue de l'obtention du diplôme de
Master en Mathématiques

Nouvelle Distribution de Poisson composée: Propriétés et Applications

Option: COSD

Par:

1. Larkem Manel
2. Yakoubi Sarab

Encadré par: SEGHIER Fatma Zohra

MCB

ENSET. SKIKDA

Devant le jury:

Président: Bouzettouta Lamine

MCA

U. SKIKDA

Examineur: Lallouche Abdallah

MCB

ENSET. SKIKDA

Année: 2021/2022

Dédicace

Aux prunelles de mes yeux, à ma raison de vivre, à celle qui m'a toujours soutenu, encouragé et a cru en moi même si j'en doutais parfois.

Merci d'être la maman que tu es. Je sais que tu es fière de moi, je te rendrai plus fière encore inshallah.

A mon père, je te serai reconnaissante toute ma vie et je ne te remercierai jamais assez.

A mes deux femmes, à ceux que j'ai le plus cher au monde, en qui je peux compter : mes soeurs Amina et Maria.

A mon petit ange, Sarab, ma meilleure amie, ma confidente, ma binôme : la soeur que j'aurais aimé l'avoir, à qui je dois ma bonne humeur et mon sourire, avec laquelle j'ai pris beaucoup de plaisir à travailler, sans qui rien n'aurait été pareil.

A mon encadreur SEGHER Fatma Zohra, qui a eu la gentillesse de partager son savoir et de nous l'inculquer inlassablement

A mes meilleures amies, particulièrement Chaima Ikram Narimen Manel Roumaïssa Rahma merci d'être là.

Manel.

Dédicace

Aux prunelles de mes yeux, à ma raison de vivre, à celle qui m'a toujours soutenu, encouragé et a cru en moi même si j'en doutais parfois.

Merci d'être la maman que tu es. Je sais que tu es fière de moi, je te rendrai plus fière encore inshallah.

A mon père, je te serai reconnaissante toute ma vie et je ne te remercierai jamais assez.

A mon petit ange, Manel, ma meilleure amie, ma confidente, ma binôme : la soeur que j'aurais aimé l'avoir, à qui je dois ma bonne humeur et mon sourire, avec laquelle j'ai pris beaucoup de plaisir à travailler, sans qui rien n'aurait été pareil.

A mon encadeur SEGHIER Fatma Zohra, qui a eu la gentillesse de partager son savoir et de nous l'inculquer inlassablement

A mes meilleures amies, particulièrement Chaima Ikram merci d'être là.

Sarab.

Remerciements

De prime abord, nous tenons à remercier Allah pour la force et la volonté qu'il nous a donné pour mener à son terme ce travail.

Nous remercions notre encadreur, Madame F. Seghier, Maître de Conférence à l'école supérieure de Skikda, pour la confiance qu'elle nous a accordé, en acceptant de diriger ce mémoire et pour la patience dont elle a toujours fait preuve.

On remercie vivement Madame Merabet Farida, Maître de Conférence à l'école supérieure de Skikda, pour l'intérêt qu'elle a apporté à ce sujet en acceptant d'être le président du jury.

Nous sommes également très honorées que Monsieur Khouchemane Houssemeddine, Maître de Conférence à l'école supérieure de Skikda pour avoir bien voulu faire partie du jury.

Toute notre gratitude va aussi à nos familles, qui nous ont soutenu, encouragé, et poussé durant toutes nos années d'études.

Résumé

Dans ce mémoire, nous introduisons une nouvelle distribution nommée **Poisson Gamma Lindley (PGaL)**.

Diverses propriétés statistiques ont été données à savoir : la méthode des moments, l'estimation du maximum de vraisemblance. Enfin, une étude comparative entre les distributions à un et deux paramètres a été discutée.

Mots-clés : Distributions de Lindley, Distribution de Poisson Gamma Lindley, Distribution de Poisson Lindley, Méthode des moments, estimation du maximum de vraisemblance.

Abstract

In this memory, we introduce a new distribution called **Poisson Gamma Lindley (PGaL)**.

Various statistical properties : moments method, maximum likelihood estimation are established. An application of the model to a real data set is presented finally and compared with the fit attained by some others well-known one and two parameters distributions.

Keywords : Lindley distributions, Poisson Gamma Lindley distribution, Poisson Lindley distribution, moments method, maximum likelihood estimation.

TABLE DES MATIÈRES

Introduction générale	1
1 Généralités et quelques notions probabilistes	3
1.1 Variables aléatoires discrètes	3
1.2 Fonction de répartition	4
1.3 Moments d'une variable aléatoire discrète	4
1.3.1 Moment ordinaire	5
1.3.2 Moment factoriel	5
1.3.3 Moment centré	6
1.4 Fonction génératrice des moments	8
1.5 Distribution d'une durée de survie	9
1.5.1 Fonction de survie $S(t)$	10
1.5.2 Fonction de risque $h(t)$ (taux de hasard)	10
1.5.3 Le concept de censure	10
1.6 Estimation	11
1.6.1 Définition d'un estimateur	11
1.6.2 Propriétés d'un estimateur	12
1.7 Méthodes de construction d'un estimateur	13
1.7.1 Méthode du maximum de vraisemblance	13
1.7.2 Méthode des Moments	14
1.8 Lois de probabilités usuelles	14

1.8.1	Loi Exponentielle	15
1.8.2	Loi de Gamma	15
1.8.3	Loi de Poisson	16
2	Quelques nouvelles distributions discrètes et ses applications	18
2.1	Distribution Poisson Lindley	18
2.1.1	Moments et mesures connexes	19
2.1.2	Fonction génératrice	22
2.1.3	Estimations	22
2.2	Distribution de Poisson Pseudo Lindley	23
2.2.1	Moments et mesures connexes	24
2.2.2	Fonction génératrice	26
2.2.3	Estimations	27
2.3	Distribution Poisson Quasi Lindley	28
2.3.1	Moments et mesures connexes	29
2.3.2	Fonction génératrice	32
2.3.3	Fonction de survie et Fonction de hasard	32
2.3.4	Estimations	32
3	Distribution Poisson Gamma Lindley et ses applications	35
3.1	Distribution Poisson Gamma Lindley	35
3.1.1	Moments et mesures connexes	36
3.1.2	Estimations	37
3.2	Applications aux ensembles de données réelles	38

Les statistiques touchent tous les aspects de la vie moderne. Elles sous-tendent de nombreuses décisions des pouvoirs publics, des entreprises et des collectivités. Elles renseignent sur les tendances et les forces qui influent sur notre vie. La qualité des procédures utilisées dans une analyse statistique dépend fortement de modèle de la probabilité supposée ou la distribution. Pour 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 notre travail, 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) a 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.

Notre travail est structuré de la manière suivante :

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

Dans le deuxième chapitre, nous faisons une synthèse des résultats sur les nouvelles distributions : Poisson Lindley et Poisson Pseudo Lindley et Poisson Quasi Lindley dont on donne quelques propriétés.

Quant au troisième chapitre, il se focalise sur la nouvelle distribution à deux paramètres (Poisson Gamma Lindley) dont on donne quelques propriétés à savoir : moments, estimation, méthode des moments, estimation du maximum de vraisemblance. Par ailleurs, une comparaison entre les distributions est obtenue.

CHAPITRE 1

Généralités et quelques notions probabilistes

Dans ce chapitre, nous rappelons quelques notions et certaines propriétés de la théorie des probabilités que nous utiliserons par la suite. 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 notre travail 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 1.1.1 *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 pour cela elle est donnée dans les tables des lois de probabilité.

Définition 1.2.1 *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 moyennes, qui sont les caractéristiques de position, on utilise également certaines autres caractéristiques de répartition. Les moments sont les plus courants 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 :

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

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 :

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

où les coefficients $S(r, j)$ sont les nombres de Stirling de deuxième espèce [21].

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

$$\begin{aligned} \mu_1^{\text{!}} &= \mu'_{(1)}, \\ \mu_2^{\text{!}} &= \mu'_{(2)} + \mu'_{(1)}, \\ \mu_3^{\text{!}} &= \mu'_{(3)} + 3\mu'_{(2)} + \mu'_{(1)}, \\ \mu_4^{\text{!}} &= \mu'_{(4)} + 6\mu'_{(3)} + 7\mu'_{(2)} + \mu'_{(1)}. \end{aligned}$$

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

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

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 centré d'ordre 2 :

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

D'une manière analogue pour le moment centré d'ordre 3 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 utilisés 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 d'ordre 3 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 d'ordre 4 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 limité 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 tout ordres.

Définition 1.4.1 *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 1.4.1 *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 bien mener 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 1.5.1 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 1.5.2 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 censures existent et nous nous limiterons, dans le cadre de ce mémoire, 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'estimations 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 considérera 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 1.6.1 *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 1.6.2 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 $b_\theta(T_n) = 0$ quelque 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 1.6.3 Un estimateur T_n de θ est dit asymptotiquement sans biais si pour tout θ de :

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

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 1.6.4 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 :

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

1.7 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 1.7.1 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 1.7.2 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 :

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

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 Lois de probabilités usuelles

Dans cette section, on procède à la présentation de quelques lois de probabilités utiles pour la suite du travail.

1.8.1 Loi Exponentielle

Une loi exponentielle modélise la durée de vie d'un phénomène sans mémoire, ou sans vieillissement, ou sans usure. En d'autres termes, le fait que le phénomène ait duré pendant t heures ne change rien à son espérance de vie à partir du temps t .

Une variable aléatoire continue X suit une loi exponentielle de paramètre (d'intensité ou inverse de l'échelle) $\lambda > 0$ si elle admet pour densité de probabilité la fonction :

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \forall x \in \mathbb{R}^+ \\ 0 & \text{ailleurs} \end{cases}$$

On note $X \rightsquigarrow EXP(\lambda)$.

La fonction de répartition de X :

$$F_X(x) = \begin{cases} 1 - \exp(-\lambda x) & \forall x \in \mathbb{R}^+ \\ 0 & \text{ailleurs} \end{cases}$$

L'espérance mathématique (ou durée de vie moyenne) et la variance de X :

$$\mathbb{E}[X] = \frac{1}{\lambda}, \text{var}[X] = \frac{1}{\lambda^2}.$$

L'estimateur $\hat{\lambda}_{MoM}$ de paramètre λ obtenu par la méthode des moments est :

$$\hat{\lambda}_{MoM} = \frac{1}{\mathbb{E}[X]}.$$

Le logarithme de la vraisemblance d'un échantillon issu d'une loi exponentielle est donné par

$$\ln L(x_i, \lambda) = n \ln \lambda - \lambda \sum_{i=1}^n \ln x_i.$$

1.8.2 Loi de Gamma

On présente la famille de lois Gamma ou d'Euler très utiles pour les propriétés de décroissance rapide de leur fonction de survie. Une variable aléatoire continue suit une loi Gamma, de paramètres $\alpha, \beta \in \mathbb{R}_+^*$, le premier est appelé paramètre

d'échelle alors que β est le paramètre de forme, si elle admet pour densité de probabilité la fonction :

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

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

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

L'espérance mathématique et la variance de X :

$$\mathbb{E}[X] = \frac{\alpha}{\beta}, \text{var}[X] = \frac{\alpha}{\beta^2}.$$

Les estimateurs $(\hat{\alpha}_{MoM}, \hat{\beta}_{MoM})$ des paramètres (α, β) obtenu par la méthode des moments sont :

$$\hat{\alpha}_{MoM} = \frac{\mathbb{E}^2[X]}{\text{var}[X]}, \quad \hat{\beta}_{MoM} = \frac{\mathbb{E}[X]}{\text{var}[X]}.$$

Le logarithme de la vraisemblance d'un échantillon issu d'une loi Gamma est donné par

$$\ln L(x_i, \alpha, \beta) = n\alpha \ln \beta - n \ln \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^n \ln x_i - n\beta \bar{X}.$$

1.8.3 Loi de Poisson

La loi de Poisson est une distribution discrète très utile dans l'étude de la survenue dans le temps d'événements homogènes (le nombre d'absents par jour dans une entreprise, le nombre de clients dans une file d'attente durant des laps de temps de même durée).

Une variable aléatoire suit une loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$ (qui est à la fois la moyenne et la variance) si :

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}; \quad k \in \mathbb{N}.$$

L'espérance mathématique et la variance de X :

$$\mathbb{E}[X] = Var[X] = \lambda.$$

Le logarithme de la vraisemblance d'un échantillon issu d'une loi gamma est donné par :

$$\ln L(x_i, \lambda) = n\lambda + \ln \lambda + \sum_{i=1}^n x_i + \sum_{i=1}^n \ln(x_i!).$$

L'estimateur $\hat{\lambda}_{MoM}$ de paramètre λ obtenu par la méthode des moments est :

$$\hat{\lambda}_{MoM} = \mathbb{E}[X].$$

CHAPITRE 2

Quelques nouvelles distributions discrètes et ses applications

Dans ce chapitre on collecte des nouvelles distributions discrètes dont on donne quelques propriétés à savoir : moments et mesures connexes, estimation du méthode des moments, estimation du maximum de vraisemblance.

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) [15] 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 :

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

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 quatres premiers moments factoriels de PLD peuvent être obtenus comme suit :

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

Puis en utilisant la relation entre les moments factoriels et les moments d'origines, les quatres premiers moments d'origines 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 d'origines et les moments centrés de la distribution PL sont obtenus 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 obtenus comme suit :

$$\begin{aligned}C.V &= \frac{\sqrt{\theta^3 + 4\theta^2 + 6\theta + 2}}{\theta + 2}, \\ \sqrt{\beta_1} &= \frac{2(\theta + 1)^4(\theta + 2) - \theta^3(\theta + 2)(\theta + 3)}{[2(\theta + 1)^3 - \theta^2(\theta + 2)]^{\frac{3}{2}}}, \\ \beta_2 &= 3 + \frac{2(\theta + 1)^5[(\theta + 3)^2 - 3] - \theta^4(\theta + 2)[(\theta + 4)^2 - 3]}{[2(\theta + 1)^3 - \theta^2(\theta + 2)]^2}.\end{aligned}$$

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 Estimations

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 log-vraisemblance 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 [15] 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 [15] 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) peut avoir plusieurs solutions.

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 :

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

2.2 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.[24]

On a :

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

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 :

$$P_{PPsL}(x; \theta, \beta) = \frac{\theta}{\beta} \frac{\beta - 1 + \theta\beta + \theta x}{(\theta + 1)^{x+2}}, \quad x = 0, 1, \dots, \quad \theta > 0, \beta \geq 1. \quad (2.6)$$

La fonction de répartition correspondante est :

$$F_{PPL}(x) = 1 - \frac{\theta\beta + \theta x + \theta + \beta}{\beta(1 + \theta)^{x+2}}, x = 0, 1, \dots, \theta > 0, \beta \geq 1. \quad (2.7)$$

Remarque 2.2.1 Si $\beta = \theta + 1$, cette distribution est la distribution de Poisson Lindley.

2.2.1 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 :

$$\mu'_{(r)} = \frac{r!(\theta + r)}{\beta\theta^r}; \quad r = 1, 2, 3, \dots \quad (2.8)$$

En substituant $r = 1, 2, 3$ et 4 dans (2.8), les quatres 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'origines, les quatres premiers moments d'origines de la distribution PPsL sont donnés par :

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

En utilisant la relation entre les moments d'origines et les moments centrés de la distribution PPsL sont obtenus comme suit :

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

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

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

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

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

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

2.2.2 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 (2.6) est donnée par :

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

2.2.3 Estimations

Estimation par la méthode des moments (MoM)

Étant donné un échantillon aléatoire X_1, \dots, X_n , de la distribution de Poisson Pseudo Lindley (2.6), 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 tout $\mu'_2 > 0, \mu'_1 > 0$, pour trouver les estimateurs des moments $\hat{\theta}_{MoM}$ et $\hat{\beta}_{MoM}$ de θ et β respectivement comme suit :

$$\begin{cases} \hat{\theta}_{MoM} = -1 + \frac{\sqrt{2\mu'_1}}{\sqrt{\mu'_1 - n + 2\mu_1'^2}} \\ \hat{\beta}_{MoM} = \frac{1}{\mu'_1\theta - 1} \end{cases}.$$

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 log-vraisemblance est :

$$\ln L(x_i; \theta) = n \ln \frac{\theta}{\beta} - n(\bar{x} + 2) \ln(1 + \theta) + \sum_{i=1}^n \ln(\beta + \theta\beta - 1 + \theta x_i). \quad (2.9)$$

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

$$\frac{\partial \ln l(x_i; \theta, \beta)}{\partial \theta} = \frac{n}{\theta} - n\left(\frac{\bar{x} + 2}{1 + \theta}\right) + \sum_{i=1}^n \left(\frac{\beta + x_i}{\beta + \theta\beta - 1 + \theta x_i}\right), \quad (2.10)$$

$$\frac{\partial \ln l(x_i; \beta, \theta)}{\partial \beta} = \frac{-n}{\beta} + \sum_{i=1}^n \left(\frac{1 + \theta}{\beta + \theta\beta - 1 + \theta x_i}\right). \quad (2.11)$$

Les deux équations (2.10) et (2.11) ne peuvent pas être résolues directement, on doit utiliser la méthode de *Fisher Scoring*, on a :

$$\begin{bmatrix} \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta^2} & \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \theta \partial \beta} \\ \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta \partial \theta} & \frac{\partial^2 Lnl(x_i; \beta, \theta)}{\partial \beta^2} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\beta} - \beta_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial Lnl(x_i; \beta, \theta)}{\partial \theta} \\ \frac{\partial Lnl(x_i; \beta, \theta)}{\partial \beta} \end{bmatrix}_{\substack{\hat{\theta} = \theta_0 \\ \hat{\beta} = \beta_0}} \quad (2.12)$$

L'équation (2.12) 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 :

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

$$\begin{cases} \hat{\theta}_{MV} = \frac{1}{X} \\ \hat{\beta}_{MV} = \frac{X}{1-X^2} \end{cases} .$$

2.3 Distribution Poisson Quasi Lindley

Une distribution composée de Poisson peut être obtenue en composant la distribution de Poisson et une distribution Quasi Lindley. [5]

On a

$$dF(\lambda) = e^{\lambda\Phi} h(\lambda) B(\Phi) d\lambda, \text{ où } h(\lambda) = \beta + \theta\lambda \text{ et } B(\Phi) = \frac{-\Phi}{\beta + 1},$$

La distribution Poisson composée est :

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

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 par :

$$f_{PQL}(x; \theta, \beta) = \frac{\theta(\beta + \theta + \beta\theta + \theta x)}{(1 + \beta)(1 + \theta)^{x+2}}, \quad x = 0, 1, \dots, \quad \theta > 0, \quad \beta > -1. \quad (2.13)$$

Remarque 2.3.1 Si $\theta = \beta$, cette distribution est la distribution Poisson Lindley.

On peut déterminer facilement la fonction de répartition pour PQL :

$$F_{PQL}(x; \theta, \beta) = 1 - \frac{\beta + 2\theta + \beta\theta + \theta x + 1}{(1 + \beta)(1 + \theta)^{x+2}}, \quad x = 0, 1, \dots, \quad \theta > 0, \quad \beta > -1. \quad (2.14)$$

2.3.1 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 :

$$\mu'_{(r)} = \frac{r!(\beta + r + 1)}{(1+\beta)\theta^r}; \quad r = 1, 2, 3, \dots \quad (2.15)$$

En substituant $r = 1, 2, 3$ et 4 dans (2.15), les quatres premiers moments factoriels de PQL peuvent être obtenus comme suit :

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

Puis en utilisant la relation entre les moments factoriels et les moments d'origines, les quatres premiers moments d'origines de la distribution PQL sont donnés

par :

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

En utilisant la relation entre les moments d'origines et les moments centrés de la distribution PQL sont obtenus comme suit :

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

Le coefficient de variation ($C.V$), le coefficient de Skewness ($\sqrt{\beta_1}$), le coefficient de Kurtosis (β_2) de PQL (2.13) sont obtenus comme suit :

$$\begin{aligned}C.V &= \frac{\sqrt{2 + 4\beta + \beta^2 + \theta(\beta + 2)(1 + \beta)}}{\beta + 1}, \\ \sqrt{\beta_1} &= \frac{2(\theta + 1)^4(\theta + 2) - \theta^3(\theta + 2)(\theta + 3)}{[2(\theta + 1)^3 - \theta^2(\theta + 2)]^{\frac{3}{2}}}, \\ \beta_2 &= 3 + \frac{2(\theta + 1)^5[(\theta + 3)^2 - 3] - \theta^4(\theta + 2)[(\theta + 4)^2 - 3]}{[2(\theta + 1)^3 - \theta^2(\theta + 2)]^2}.\end{aligned}$$

2.3.2 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 (2.13) est donnée par :

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

2.3.3 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.

2.3.4 Estimations

Estimation par le maximum de vraisemblance

Dans cette partie on discute les points d'estimation pour les paramètres de PQL (θ, β) . Soit la fonction log-vraisemblance pour une seule observation (dite x_i)

pour le vecteur de paramètre (θ, β) peut être écrit comme suit :

$$\ln l(x; \theta, \beta) = \ln \theta + \ln(\beta + \theta + \beta\theta + \theta x) - \ln(1 + \beta) - 2 \ln(1 + \theta) - x \ln(1 + \theta), \quad (2.16)$$

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

$$\frac{\partial \ln l(x; \theta, \beta)}{\partial \theta} = \frac{1}{\theta} + \left(\frac{1 + \beta + x}{\beta + \theta + \beta\theta + \theta x} \right) - \frac{2 + x}{1 + \theta}, \quad (2.17)$$

$$\frac{\partial \ln l(x; \theta, \beta)}{\partial \beta} = \left(\frac{1 + \theta}{\beta + \theta + \beta\theta + \theta x} \right) - \frac{1}{1 + \beta}. \quad (2.18)$$

Les estimateurs du maximum de vraisemblance $\hat{\theta}$ de θ et $\hat{\beta}$ de β sont obtenus par la résolution de système non linéaire :

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

Estimation par la méthode des moments

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

$$\begin{cases} \mu'_1 = \frac{2+\beta}{(1+\beta)\theta} \\ \mu'_2 = \frac{\theta\beta+2\theta+2\beta+6}{\theta^2(\beta+1)} \end{cases} \quad (2.19)$$

où $\mu'_2 = \mu_2 + (\mu'_1)^2$ et μ_2 est la variance. On résout ce système non linéaire on trouve le couple $(\hat{\theta}, \hat{\beta})$, où $(\hat{\theta}, \hat{\beta}) > 0$ pour tout $\mu_2 > 0$, $\mu'_1 > 0$. La solution du système non linéaire (2.19) donne :

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

Il est facile de voir que la solution de :

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

Est :

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

Alors :

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

CHAPITRE 3

Distribution Poisson Gamma Lindley et ses applications

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

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.[11]

on a :

$$dF(\lambda) = e^{\lambda\Phi} h(\lambda) B(\Phi) d\lambda, \text{ où } h(\lambda) = 1 + \beta\lambda \text{ et } B(\Phi) = \frac{[-\Phi]^2}{\beta(1-\Phi)},$$

La distribution poisson composée est (voir Sankaran (1970)) :

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

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 par :

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

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

Remarque 3.1.1 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 Moments et mesures connexes

Proposition 3.1.1 Soit X_1, X_2, \dots, X_n les n variables aléatoires indépendantes 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 3.1.1 Soit $X \sim PGaL(\theta, \beta)$, l'espérance et la variance pour X est :

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

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

3.1.2 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 à 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 peuvent pas être résolues directement, on

doit utiliser la méthode de *Fisher Scoring*, on a :

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

L'équation (3.7) peut se résoudre de façon itérative où θ_0, β_0 sont les valeurs initiales de θ, β .

Estimation 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 résout ce système non linéaire on trouve le couple $(\hat{\theta}, \hat{\beta})$, où $\hat{\theta}, \hat{\beta} > 0$ pour tout $\mu_2 > 0, \mu'_1 > 0$. La solution du système non linéaire (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 Applications aux ensembles de données réelles

Dans cette section. Nous illustrons l'applicabilité de la distribution PGaL 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) et la distribution Poisson Gamma Lindley (PGaLD).

Selon le **tableau 3.3**, on peut observer que la distribution PGaL 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> $\hat{\theta} = 0.75$	<i>PLD</i> $\hat{\theta} = 1.8081$	<i>PPsLD</i> $\hat{\theta} = 6.9277$ $\hat{\beta} = 0.2383$	<i>PGaLD</i> $\hat{\theta} = 1.5257$ $\hat{\beta} = 5.93$
0	33	26.45	31.49	23.03	31.85
1	12	19.84	14.17	25.54	13.83
2	6	7.44	6.09	6.08	5.94
3	3	1.86	2.54	1.13	2.53
4	1	0.35	1.04	0.19	1.07
5	1	0.05	0.42	0.03	0.45
<i>Total</i>	56	56	56	56	56
χ^2		24.1	1.289	49.405	1.057

Tableau 3.3. Comparaisons des fréquences observées

Les fonctions de masse des probabilités ajustées correspondantes de PD, PPsLD et PGaLD ainsi que les points de données d'origines pour l'ensemble de données du tableau 3.3 sont données dans la figure 3.2.

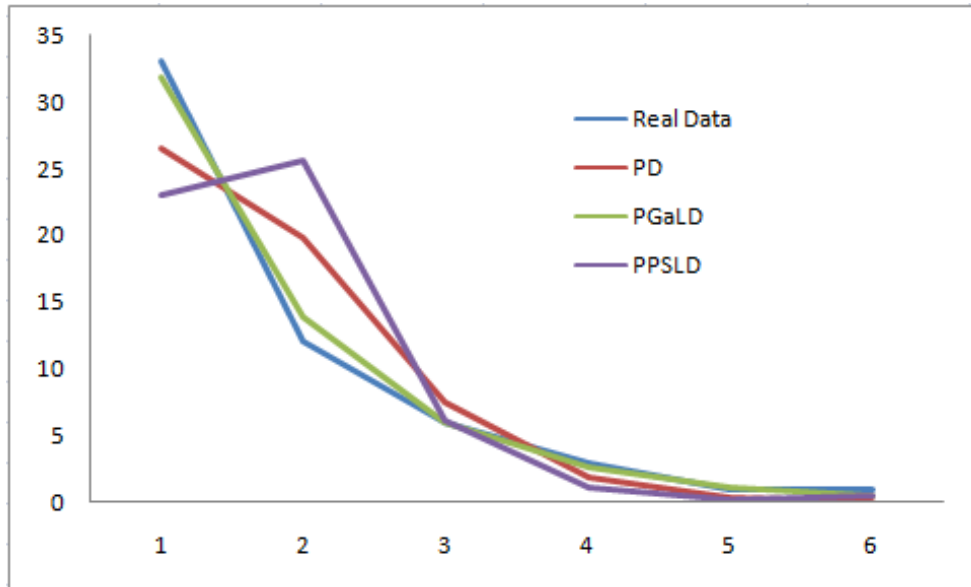


Figure 3.2. Comparaison des différents modèles *OBS*, *PD*, *PQLD*, *PP_sLD* et *SBPGaLD*

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 Bayésien (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

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 $PGaLD$ par rapport à $PPsLD$, 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 plausibles est le $PGaLD$. Donc la distribution $PGaL$ peut donc être considérée comme un modèle important pour les données de sciences biologiques.

Conclusion

Ainsi, nous avons réussi à introduire une nouvelle distribution nommée "**Poisson Gamma Lindley (PGaL)**".

L'estimation de ses paramètres a été discutée en utilisant la méthode du maximum de vraisemblance et la méthode des moments. 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.

BIBLIOGRAPHIE

- [1] **Akaike H. (1974)**, A new look at statistical model identification, IEEE Transactions on Automatic Control AU-19 : 716-722.
- [2] **Fisher R.A (1937)**, The effects of methods of ascertainment upon the estimation of frequency. Ann. Eugenics. 6(1) : 13-25.
- [3] **Ghitany M. E., Atieh B., Nadarajah S. (2008)**, Lindley distribution and its applications. *Math. Comput. Simulation* , 78, pp. 493-506.
- [4] **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.
- [5] **Grine R., Zeghdoudi H (2017)**, On Poisson Quasi-Lindley Distribution and its Applications. Journal of Modern Applied Statistical Methods, 16(2), 403-417.
- [6] **Johnson N L, Kotz S and Kemp A W.(2005)**, Univariate Discrete Distribution, third edition, John Wiley, New York.
- [7] **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.
- [8] **Lecoutre J.P. (2002)**, Statistique et probabilités. Dunod. Paris, 2^{ième} édition, ISBN 2-10-007277-3.
- [9] **Lejeune M. (2010)**, Statistique La théorie et ses applications. Springer-Verlag France, Paris, 2^{ième} édition. ISBN 978-2-8178-0156-8.

- [10] **Lindley D. V. (1958)**, Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1), 102-107.
- [11] **Nedjar S., Zeghdoudi H. (2020)**, New compound Poisson distribution : Properties, inflated distribution and applications. *International Journal of Agricultural and Statistics Sciences*. 16(2) :519-526
- [12] **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.
- [13] **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.
- [14] **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 ;
- [15] **Sankaran M. (1970)**, The discrete Poisson-Lindley distribution. *Biometrics*, 26, pp. 145-149.
- [16] **Saporta G. (2006)**, Probabilités analyse des données et statistiques. Editions Technip. Paris, 2^{ième} édition. ISBN 2-7108-0814-5.
- [17] **Schwarz G. (1978)**, Estimating the dimension of a model, *Annals of Statistics* 6 : 461-464.
- [18] **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.
- [19] **Shanker R., Shukla K.K (2019)**, A generalization of zero-truncated Poisson-Sujatha distribution. *Hungarian Statistical Review*, 2 (2). pp. 32-50.
- [20] **Sheldon M. Ross (1987)**, *Initiation aux probabilités*. Presses polytechniques romandes lière édition, ISBN 2-88074-118-8.
- [21] **Seghier F.Z.(2020)**, La nouvelle distribution de Poisson-Gamma Lindley : Simulation et Applications (Thèse de Doctorat), UBMA.

-
- [22] **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
- [23] **Zhang.T, M. Xie. (2011)**, On the upper truncated Weibull distribution and its reliability implications, Reliability Engineering and System Safety 96. 194–200.
- [24] **Zeghdoudi H. , Nedjar S. (2017)**. A Poisson pseudo Lindley distribution and its application . *Journal of Probability and Statistical Sciences* . Vol. 15, N° 1.