



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة الشاذلي بن جديد - الطارف

Université Chadli Bendjedid – El Tarf

كلية العلوم والتكنولوجيا

Faculté des Sciences et de la Technologie

قسم الرياضيات

Département de Mathématiques

Mémoire de fin d'études

En vue de l'obtention du diplôme de Master

Domaine : Mathématiques et Informatique

Filière : Mathématiques

Spécialité : Analyse fonctionnelle et calcul stochastique

Thème

NOUVELLE DISTRIBUTION ZERO TRONQUEE

Présenté par :

Sahel Malak

Devant le Jury :

Dr. Grabsia Imen	MCB	Univ Chadli Bendjedid El Tarf	Présidente
Dr. Grine Razika	MCB	Univ Chadli Bendjedid El Tarf	Rapporteur
Dr. Zidani Nesrine	MCA	Univ Chadli Bendjedid El Tarf	Examinatrice

Année Universitaire 2024-2025

Nouvelle distribution zero tronquée

Sahel malek

Département de Mathématiques, Université de Chadli Bendjedid El-Tarf

Juin 2025

Encadré par : Dr Grine Razika

TABLE DES MATIÈRES

1	Introduction	9
1	Généralités et quelques notions de Probabilités	11
1	Variables aléatoires discrètes	12
2	Fonction de répartition	12
3	Moments d'une variable aléatoire discrète	13
3.1	Moment ordinaire	13
3.2	Moment factoriel	13
3.3	Moment centré	15
4	Fonction génératrice des moments	17
5	Distribution d'une durée de survie	17
5.1	Fonction de survie $S(t)$	18
5.2	Fonction de risque $h(t)$ (taux de hasard)	19
6	Estimation	19
6.1	Définition d'un estimateur	19
6.2	Propriétés d'un estimateur	20
6.3	Méthodes de construction d'un estimateur	22
7	Quelques lois de probabilités usuelles	24
7.1	Loi de Poisson	24

7.2	Distribution de Lindley	24
7.3	Loi de Gamma	25
8	Sélection de model avec différents critères	26
8.1	Le critère d'information d'Akaike (AIC) et (AICc)	26
8.2	Le critère BIC	27
2	Quelques distribution tronquée à zéro et ses applications	29
1	Modèles zéro tronquées discrètes	29
2	Distribution zéro tronquée Poisson	29
3	Distribution zéro tronquée Poisson Lindley	31
4	Distribution zéro tronquée Poisson Quasi-Lindley	33
4.1	Moments et mesures connexes	33
4.2	Estimation du paramètre	36
3	Nouvelle distribution à deux paramètres tronquée à zéro et ses applications	39
1	Distribution à deux paramètres	39
2	Distribution à deux parametres zéro tronquée	41
2.1	Moments et mesures connexes	42
2.2	Propriétés statistiques de ZTPD	44
2.3	Estimation des paramètres	45
3	Application aux ensembles de données réelles	48

REMERCIEMENT

Avant tout, je tiens à remercier dieu le tout puissant, de m'avoir donné le courage et la patience pour pouvoir mener à terme ce travail.

Je tiens à remercier Dr.Grine Razika, mon encadreur, pour tout le soutien, l'aide ; l'orientation ; la guidance ; ainsi que pour ses précieux conseils et ses encouragements lors de la réalisation de ce travail.

*Je remercie les membres du jury Dr Grabsia Imen et Dr. Zidani Nesrine .
Ensuite un grand merci à mes parents pour le soutien, l'encouragement tout au long de mes études, et à tous ma famille.*

Je remercie également toutes les personnes de près ou de loin, ont participé à l'élaboration de ce mémoire.

Je remercie tous les amis et les collègues du département de mathématiques

Dédicace

Je dédie ce mémoire

À ma famille, source de vie et de lumière, Vous êtes mon espoir, mon refuge, ma prière.

À ma mère si douce, au cœur plein de tendresse, Et à mon père, pilier de force et de sagesse.

À mes frères et sœurs, trésors de mon chemin, compagnons de l'âme, soutien de mes lendemains.

Grâce à vous, mes jours ont fleuri de beauté , recevez ce travail avec amour et fierté.

Sahel Malek

ملخص

في هذا العمل، نقترح توزيعاً جديداً ثنائي المعلمات يسمى التوزيع المقطوع عند الصفر. يتم دراسة خصائصه من خلال حساب اللحظات وتقدير المعلمات باستخدام طريقة أقصى احتمال وطريقة اللحظات. يتم إجراء محاكاة لتقييم أداء النموذج

الكلمات الرئيسية: نموذج مقطوع عند الصفر، توزيع ثنائي المعلمات، أقصى احتمال، طريقة اللحظات

Résumé

Dans ce travail nous proposons une nouvelle distribution à deux paramètres en nommée zéro-tronquée. ses propriétés sont étudiées à travers le calcul des moments et l'estimation des paramètres par la méthode de maximum de vraisemblance et la méthode de moments.une simulation est réalisée pour évaluer les performances du modèle

Key Words : Modèle zéro tronquée, Distribution à deux paramètres, maximum de vraisemblance, méthode de moment..

Abstract

In this paper, we propose a new two-parameter distribution called zero-truncated. Its properties are studied through the calculation of moments and the estimation of parameters using the maximum likelihood method and the method of moments. A simulation is performed to evaluate the performance of the model.

Key Words : Zero-truncated model, two-parameter distribution, maximum likelihood estimator, method of moments.

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 verssion tronquée de la distribution de Lindley, ainssi que ces caractéristique statistique tellque les moment, la fonction quantile et les estimateurs du maximum de vraisemblance pour différentes 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 zéro tronquée Poisson Lindley donne un meilleur ajustement que le zéro tronquée Poisson. tandis que dans les sciences sociales, la distribution zéro tronquée de Poisson donne un meilleur ajustement que la distribution zéro tronquée Poisson Lindley. Dans ce travail, nous nous proposons une nouvelle distribution de durée de survie basée sur les modèles tronqués cette distribution est nommée la distribution à deux paramètres 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 (Grine et Zeghdoudi 2024).

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 distributions zéro tronquée : Poisson Lindley zéro tronquée, Poisson Pseudo Lindley zéro tronquée et Poisson Quasi Lindley zéro tronquée dont on donne quelques propriétés. Dans le chapitre 3, nous nous intéressons à la nouvelle distribution tronquée nommé novel distribution à deux paramètres zéro tronquée 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 de Probabilités

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.

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 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).*

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 événement. C'est pourquoi c'est elle qui est donnée dans les tables des lois de probabilité.

Définition 2.1 *Soit X une variable aléatoire, on appelle fonction de répartition de*

3. Moments d'une variable aléatoire discrète

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.

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.

3.1 Moment ordinaire

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

$$\mu_r^i = \mathbb{E}(X^r) = \sum_{x \in I} x^r \mathbb{P}[X = x]. \quad (1.1)$$

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.

3. Moments d'une variable aléatoire discrète

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

$$\mu_r^{\cdot} = \sum_{j=1}^r S(r, j) \mu_{(j)}^{\cdot}, \quad \text{pour } r = 1, 2, \dots$$

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 :

$$\begin{aligned}\mu_1^{\cdot} &= \mu_{(1)}^{\cdot}, \\ \mu_2^{\cdot} &= \mu_{(2)}^{\cdot} + \mu_{(1)}^{\cdot}, \\ \mu_3^{\cdot} &= \mu_{(3)}^{\cdot} + 3\mu_{(2)}^{\cdot} + \mu_{(1)}^{\cdot}, \\ \mu_4^{\cdot} &= \mu_{(4)}^{\cdot} + 6\mu_{(3)}^{\cdot} + 7\mu_{(2)}^{\cdot} + \mu_{(1)}^{\cdot}.\end{aligned}$$

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

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

3.3 Moment centré

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

$$\mu_r = \mathbb{E}([X - \mathbb{E}(X)]^r) = \sum_{i=1}^n [x_i - \mathbb{E}(X)]^r p_i. \quad (1.2)$$

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 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 moment 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.

3. Moments d'une variable aléatoire discrète

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 :

$$C.V. = \frac{\sqrt{\mu_2}}{\mu'_1}.$$

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

$$G_X(t) = \mathbb{E}(e^{tX}). \quad (1.3)$$

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

Proposition 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 $t = 0$.

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.

5.1 Fonction de survie $S(t)$

Définition 5.1 *La fonction de survie est, pour t fixé, la probabilité de survivre jusqu'à l'instant t , c'est à dire :*

$$S(t) = \mathbb{P}(X > t); \quad t \geq 0. \quad (1.4)$$

Définition 5.2 *La fonction de répartition F fait correspondre à un temps t la probabilité de décéder avant le temps t :*

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

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) :

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

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.

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 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.

6.2 Propriétés d'un estimateur

Biais d'un estimateur

Définition 6.2 *Soit une v.a. X de loi de fonction de probabilité $\mathbb{P}(x; \theta)$ 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 6.3 *Un estimateur T_n de θ est dit asymptotiquement sans biais si pour tout θ de :*

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

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

6. Estimation

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 6.4 *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 6.1 [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 6.5 *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 :

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

$$\text{car } \mathbb{E}_{\theta}[T_n - \mathbb{E}_{\theta}(T_n)] = 0.$$

6.3 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.

Méthode du maximum de vraisemblance

Définition 6.6 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 :

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n \mathbb{P}(X_i = x_i), \quad (1.5)$$

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 6.7 On appelle *estimateur du maximum de vraisemblance (EMV)* toute

6. Estimation

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 : $\forall j \in \{1, \dots, d\}$,

$$\frac{\partial}{\partial \theta_j} \ln L(\theta; x_1, \dots, x_n) = 0,$$

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

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

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

$$\hat{\theta}_n = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i. \quad (1.6)$$

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 $S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

7 Quelques lois de probabilités 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.

7.1 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 :

$$\mathbb{P}_X(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k \in \mathbb{N}^*, \quad (1.7)$$

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

On déduit aisément :

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k \mathbb{P}_X(X = k) = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda, \quad (1.8)$$

et

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

7.2 Distribution de Lindley

Soient Y_1 et Y_2 deux variables aléatoires indépendantes. Pour $\theta > 0$, on considère la variable aléatoire $X = Y_1$ et $X = Y_2$ avec les probabilités respectivement $P_1 = \frac{\theta}{1+\theta}$ et $P_2 = \frac{1}{1+\theta}$. ou $Y_1 \sim exp(\theta)$ et $Y_2 \sim Gamma(2, \theta)$

La distribution lindley est spécifiée par la fonction densité suivante

$$f(x; \theta) = \frac{\theta^2(1+x)e^{-\theta x}}{1+\theta}; \quad x, \theta > 0 \quad (1.9)$$

7. Quelques lois de probabilités usuelles

ou la fonction de répartition est donné par :

$$F(x) = 1 - \frac{\theta + 1 + \theta x}{1 + \theta} e^{-\theta x}; x > 0, \theta > 0 \quad (1.10)$$

Moments et mesures connexes

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

$$\mu_r^l = E(X^r) = \frac{r! (\theta + r + 1)}{\theta^r (\theta + 1)}, r = 1, 2, \dots \quad (1.11)$$

En particulier on a

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

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

$$\mu_k = E\{(X - \mu)^k\} = \sum_{r=0}^k \binom{k}{r} \mu_r^l (-\mu)^{k-r}. \quad (1.12)$$

En particulier on a

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

7.3 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 $\alpha, \beta \in R_+^*$, le premier est appelé paramètre d'échelle alors que β est le paramètre de forme, si elle admet pour densité de probabilité la

fonction :

$$f_X(x) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \forall x > 0 \\ 0 & \text{ailleurs} \end{cases} \quad (1-13)$$

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

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

L'espérance mathématique et la variance de X :

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

8 Sélection de model avec différents critères

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

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

8.1 Le critère d'information d'Akaike (AIC) et (AICc)

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

$$AIC = -2 \log(L) + 2k, \quad (1.14)$$

8. Sélection de model avec différents critères

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

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

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

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

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

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

$$AICc = AIC + \frac{2k(k+1)}{n-k-1}. \quad (1.15)$$

8.2 Le critère BIC

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

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

Le Bayesian information criterion BIC est défini par :

$$BIC = -2\log(L) + k\log(n). \quad (1.15)$$

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

8. *Sélection de model avec différents critères*

de décision vis-à-vis de la signification statistique des paramètres retenus dans le modèle.

CHAPITRE 2

Quelques distribution tronquée à zéro et ses applications

1 Modèles zéro 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 $\mathbb{P}_0(x; \theta)$ la distribution originale avec un support des entiers non négative, alors ça version tronquée par un zéro (Zéro truncated) est définit par :

$$\mathbb{P}(x; \theta) = \frac{\mathbb{P}_0(x; \theta)}{1 - \mathbb{P}_0(0; \theta)}; \quad x = 1, 2, 3, \dots \quad (2.1)$$

2 Distribution zéro tronquée Poisson

En utilisant (2.1) 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 :

$$\mathbb{P}(x; \theta) = \frac{\theta^x}{(e^\theta - 1) x!}; \quad x = 1, 2, 3, \dots, \quad \theta > 0. \quad (2.2)$$

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

$$\mu'_{(r)} = \frac{\theta^r e^\theta}{e^\theta - 1}. \quad (2.3)$$

En substituant $r = 1, 2, 3$ et 4 dans (2.3), 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 vraisemblance $\hat{\theta}$ de θ est donné par la solution de l'équation non linéaire suivante :

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

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

3. Distribution zéro tronquée Poisson Lindley

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.

3 Distribution zéro tronquée Poisson Lindley

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

$$\mathbb{P}(x; \theta) = \frac{\theta^2}{\theta^2 + 3\theta + 1} \frac{x + \theta + 2}{(\theta + 1)^x}; \quad x = 1, 2, 3, \dots, \quad \theta > 0 \quad (2.4)$$

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

$$\mu'_{(r)} = \frac{r! (\theta + 1)^2 (r + \theta + 1)}{\theta^r (\theta^2 + 3\theta + 1)}; \quad r = 1, 2, 3, \dots \quad (2.5)$$

En substituant $r = 1, 2, 3$ et 4 dans (2.5), 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 Distribution zéro tronquée Poisson Quasi-Lindley

Soit $\mathbb{P}_0(x; \theta, \beta)$ la distribution original définit dans (2.1) avec un support des entiers non négative, alors ça version tronquée à zéro est définit par :

$$\mathbb{P}(x; \theta, \beta) = \frac{\mathbb{P}_0(x; \theta, \beta)}{1 - \mathbb{P}_0(0; \theta, \beta)}; x = 1, 2, .. \quad (2.6)$$

la distribution zero tronquée de Poisson Quasi-Lindley avec :

$$\mathbb{P}(x; \theta, \beta) = \frac{\theta (\theta\beta + \beta + \theta + \theta x)}{(1 + 2\theta + \beta + \theta\beta) (\theta + 1)^x}; x = 1, 2, 3, \dots, \theta > 0, 1 + 2\theta + \beta + \theta\beta > 0 \quad (2.7)$$

Remarque 4.1 Lorsque $\beta = \theta$, on peut déduit facilement que la distribution donner par (2.7) revient à la distribution zero tronquated Poisson Lindley.

4.1 Moments et mesures connexes

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

$$\mu'_{(r)} = \frac{r!}{1 + 2\theta + \beta + \theta\beta} \left[\frac{(\theta + 1)^2 (r + \beta + 1)}{\theta^r} \right] \quad (2.8)$$

En substituant $r = 1, 2, 3$ et 4 dans (2.8), les quatre premiers moments factoriels

4. Distribution zéro tronquée Poisson Quasi-Lindley

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 :

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

4. Distribution zéro tronquée Poisson Quasi-Lindley

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

le coefficient de variation ($C.V$), le coefficient de Skewness ($\sqrt{\beta_1}$), le coefficient de Kurtosis (β_2) de ZTPQL (2.7) sont obtenues comme suit :

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

Fonctions génératrice

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

$$G_X(t) = \mathbb{E}(t^X) = \frac{\theta}{1 + 2\theta + \beta + \theta\beta} \left[(\theta\beta + \beta + \theta) \sum_{x=1}^{\infty} \left(\frac{t}{\theta + 1}\right)^x + \theta \sum_{x=1}^{\infty} x \left(\frac{t}{\theta + 1}\right)^x \right] \quad (2.9)$$

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 :

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

4.2 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 (2.7). La fonction de maximum de vraisemblance est donc :

$$L(x_i; \theta, \beta) = \left(\frac{\theta}{1 + 2\theta + \beta + \theta\beta} \right)^n \prod_{i=1}^n \left(\frac{\theta\beta + \beta + \theta + \theta x_i}{(\theta + 1)^{\sum_{i=1}^n x_i}} \right) \quad (2.10)$$

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 :

$$\frac{\partial \ln L(x_i; \theta, \beta)}{\partial \theta} = \frac{n}{\theta} - \frac{n(\beta + 2)}{1 + 2\theta + \beta + \theta\beta} - \frac{n\bar{X}}{(\theta + 1)} + \sum_{i=1}^n \left(\frac{1 + \beta + x_i}{\theta\beta + \beta + \theta + \theta x_i} \right) \quad (2.11)$$

$$\frac{\partial \ln L(x_i; \theta, \beta)}{\partial \beta} = -\frac{n(\theta + 1)}{1 + 2\theta + \beta + \theta\beta} + \sum_{i=1}^n \left(\frac{\theta + 1}{\theta\beta + \beta + \theta + \theta x_i} \right) \quad (2.12)$$

Les estimateurs du maximum de 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 (2.9) et (2.10) ne peuvent pas être résolues directement, on doit utiliser la méthode Fisher Scoring, on a :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{n}{\theta^2} + \frac{n(\beta + 2)^2}{(1 + 2\theta + \beta + \theta\beta)^2} + \frac{n\bar{X}}{(1 + \theta)^2} - \sum_{i=1}^n \frac{(x_i + \beta + 1)^2}{(\theta x_i + \beta + \beta\theta + \theta)^2} \quad (2.13)$$

$$\frac{\partial^2 \ln L}{\partial \beta^2} = \frac{n(1 + \theta)^2}{(1 + 2\theta + \beta + \theta\beta)^2} - \sum_{i=1}^n \frac{(1 + \theta)^2}{(\theta x_i + \beta + \beta\theta + \theta)^2} \quad (2.14)$$

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

4. Distribution zéro tronquée Poisson Quasi-Lindley

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 (2.7) 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 (2.7) 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 :

$$\beta (\theta + 1) = \frac{3\theta p_1 - 2\theta (\theta + 1) p_2}{p_2 (\theta + 1) - p_1} \quad (2.16)$$

4. Distribution zéro tronquée Poisson Quasi-Lindley

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

$$\beta(\theta + 1) = \frac{2\theta^2(1 - \mu) + \theta(4 - \mu) + 2}{(\mu\theta - \theta - 1)} \quad (2.17)$$

Maintenant, en assimilant les équations (2.16) et (2.17), nous obtenons une équation quadratique de θ comme :

$$A\theta^2 + B\theta + C = 0 \quad (2.18)$$

Tellque

$$\begin{cases} A = (1 - \mu)p_1 + (2 - \mu)p_2 \\ B = (\mu - 1)p_1 + (4 - \mu)p_2 \\ C = 2(p_1 - p_2) \end{cases} \quad (2.19)$$

D'après l'équation (2.18), l'estimateur de moment du paramètre θ peut être obtenu de :

$$\hat{\theta} = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \quad (2.20)$$

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 (2.17) peuvent être obtenues et le remplacement de ces valeurs par (2.18) 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 :

$$\hat{\beta} = \frac{1}{\theta + 1} \left[\frac{3\theta p_1 - 2\theta(\theta + 1)p_2}{p_2(\theta + 1) - p_1} \right] \quad (2.21)$$

ou bien

$$\hat{\beta} = \frac{1}{\theta + 1} \left[\frac{2\theta^2(1 - \mu) + \theta(4 - \mu) + 2}{(\mu\theta - \theta - 1)} \right] \quad (2.22)$$

CHAPITRE 3

Nouvelle distribution à deux paramètres tronquée à zéro et ses applications

1 Distribution à deux paramètres

En (2024), Grine et Zeghdoudi ont introduit une autre nouvelle distribution discrète appelée distribution à deux paramètres. Nous présentons ci-dessous quelques aspects statistiques concernant cette distribution.

alors la fonction de densité de la distribution à deux paramètres est donnée par :

$$f(x; \theta, \gamma) = \theta (\gamma + (1 - \gamma) \theta x) \exp -\theta x, \quad (3.1)$$

where $x = 0, 1, \dots$,

La fonction de distribution cumulative (CDF), la fonction de survie (SF) et la fonction de taux de risque (HRF) correspondantes sont données par :

$$F(x; \theta, \gamma) = 1 - e^{-\theta x} (1 + (1 - \gamma) \theta^2 x), \quad x, \theta > 0.0 < \gamma < 1 \quad (3.2)$$

$$S(t; \theta, \gamma) = e^{-\theta x} (1 + (1 - \gamma) \theta^2 x), \quad x, \theta > 0.0 < \gamma < 1$$

$$h(t; \theta, \gamma) = \frac{\theta^2 (\gamma + (1 - \gamma) \theta x)}{(1 - \gamma) \theta^2 x + 1}$$

Let $X \sim TPD$, Then the i th moment of X is determined as follows

Ainsi, les quatre premiers moments de la variable aléatoire TPD peuvent être obtenus en substituant respectivement $i = 1, 2, 3, 4$ dans l'équation (3.3). Ils sont utilisés pour déterminer respectivement la variance, l'asymétrie, l'aplatissement et le coefficient de variation de TPD, comme suit

$$\mathbb{E}(X^i) = \frac{\Gamma(1+i)}{\theta^{(2+i)}} [\theta\gamma + (1-\gamma)\theta(1+i)] \quad (3.3)$$

$$\mu_1' = \frac{-\theta + 2\beta(1+\theta)}{\theta\beta(1+\theta)}, \quad \mu_2' = \frac{-4\theta + 8\theta\beta + 2\beta\theta^2 + 6\beta - \theta^2}{\theta^2\beta(1+\theta)},$$

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

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

$$Skewness = \sqrt{\beta_1} = \frac{\mathbb{E}(X^3)}{(\mathbb{E}(X))^{\frac{3}{2}}} = \frac{6\theta^4(4(1-\gamma)\theta + \gamma\theta)}{(2\gamma\theta^3 + 6(1-\gamma)\theta^3 - \gamma^2\theta^2 - 4\gamma(1-\gamma)\theta^2 - 4(1-\gamma)^2\theta^2)^{\frac{3}{2}}},$$

$$Kurtosis = \beta_2 = \frac{\mathbb{E}(X^4)}{(\mathbb{E}(X))^2} = \frac{24\theta^6(5(1-\gamma)\theta + \theta\gamma)}{(2\gamma\theta^3 + 6(1-\gamma)\theta^3 - \gamma^2\theta^2 - 4\gamma(1-\gamma)\theta^2 - 4(1-\gamma)^2\theta^2)^2},$$

2. Distribution à deux paramètres zéro tronquée

$$\begin{aligned}
 C.V &= K = \frac{\sqrt{Var(X)}}{\mathbb{E}(X)} = \frac{\sqrt{2\gamma\theta^3 + 6(1-\gamma)\theta^2 - \gamma^2\theta^2 - 4\gamma(1-\gamma)\theta^2 - 4(1-\gamma)^2\theta^2}}{2(1-\gamma)\theta + \theta\gamma} \\
 &= \frac{\sqrt{2\gamma\theta + 6(1-\gamma) - \gamma^2 - 4\gamma(1-\gamma) - 4(1-\gamma)^2}}{2-\gamma}
 \end{aligned}$$

2 Distribution à deux paramètres zéro tronquée

Dans cette section, la distribution deux paramètres zéro tronquée (*ZTTPD*) a été introduite avec ses propriétés statistiques, y compris l'expression générale des probabilités, les moments, les paramètres de la distribution deux paramètres 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 $\mathbb{P}_0(x; \theta, \beta)$ la distribution original définit dans (2.1) avec un support des entiers non négative, alors ça version tronquée à zéro est définit par :

$$\mathbb{P}(x; \theta, \beta) = \frac{\mathbb{P}_0(x; \theta, \beta)}{1 - \mathbb{P}_0(0; \theta, \beta)}; x = 1, 2, \dots \quad (3.4)$$

A partir de (2.1) et (3.4), on peut déduire la distribution à deux paramètres zero tronquée avec :

$$\mathbb{P}(x; \theta, \gamma) = f(x; \theta, \gamma) = \frac{\theta[\gamma + (1-\gamma)\theta x]e^{-\theta x}}{1 - \theta\gamma}; x = 1, 2, 3, \dots, \theta > 0, 1 - \theta\gamma > 0 \quad (3.5)$$

2.1 Moments et mesures connexes

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

$$\begin{aligned} \mu'_{(r)} &= \mu'_{(r)} = E [x^{(r)}] = \int_0^{\infty} x^{(r)} f(x) dx \\ x^{(r)} &= x(x-1)(x-2)\dots\dots(x-r+1) \\ \mu'_{(r)} &= \frac{\theta}{1-\theta\gamma} \left[\gamma \frac{\Gamma(r+1)}{\theta^{r+1}} + (1-\gamma)\theta \frac{\Gamma(r+2)}{\theta^{r+2}} \right] \\ &= \frac{\gamma\Gamma(r+1) + (1-\gamma)\Gamma(r+2)}{\theta^r(1-\theta\gamma)} \end{aligned}$$

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

$$\mu'_{(r)} = \frac{\gamma\Gamma(r+1) + (1-\gamma)\Gamma(r+2)}{\theta^r(1-\theta\gamma)} \tag{3.6}$$

En substituant $r = 1, 2, 3$ et 4 dans (3.6), les quatre premiers moments factoriels de ZTTPD peuvent être obtenus comme :

$$\begin{aligned} \mu'_{(1)} &= \frac{2-\gamma}{\theta(1-\theta\gamma)} \\ \mu'_{(2)} &= \frac{6-4\gamma}{\theta^2(1-\theta\gamma)} \\ \mu'_{(3)} &= \frac{24-18\gamma}{\theta^3(1-\theta\gamma)} \\ \mu'_{(4)} &= \frac{120-96\gamma}{\theta^4(1-\theta\gamma)} \end{aligned}$$

puis en utilisant la relation entre les moments factoriels et les moments d'origine, les

2. Distribution à deux paramètres zéro tronquée

quatre premiers moments d'origine de la distribution ZTTPD sont donnés par :

$$\begin{aligned}\mu'_1 &= \frac{2 - \gamma}{\theta(1 - \theta\gamma)} \\ \mu'_2 &= \frac{(2 - \gamma)\theta + (6 - 4\gamma)}{\theta^2(1 - \theta\gamma)} \\ \mu'_3 &= \frac{(24 - 18\gamma) + 3(6 - 4\gamma)\theta + (2 - \gamma)\theta^2}{\theta^3(1 - \theta\gamma)} \\ \mu'_4 &= \frac{(120 - 96\gamma) + 6(24 - 18\gamma)\theta + 7(6 - 4\gamma)\theta^2 + (2 - \gamma)\theta^3}{\theta^4(1 - \theta\gamma)}\end{aligned}$$

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

$$\begin{aligned}\mu_2 &= \frac{(\theta^2\gamma^2 - 2\theta^2\gamma + 4\theta\gamma^2 - 7\theta\gamma + 2\theta - \gamma^2 + 2)}{\theta^2(\theta\gamma - 1)^2} \\ \mu_3 &= \frac{\left(\begin{array}{l} \theta^4\gamma^3 - 2\theta^4\gamma^2 + 12\theta^3\gamma^3 - 20\theta^3\gamma^2 + 4\theta^3\gamma + 15\theta^2\gamma^3 - 36\theta^2\gamma^2 + 25\theta^2\gamma \\ -2\theta^2 - 12\theta\gamma^3 + 9\theta\gamma^2 + 12\theta\gamma - 6\theta + 2\gamma^3 - 4 \end{array} \right)}{\theta^3(\theta\gamma - 1)^3} \\ \mu_4 &= \frac{\left(\begin{array}{l} \theta^6\gamma^4 - 2\theta^6\gamma^3 + 28\theta^5\gamma^4 - 45\theta^5\gamma^3 + 6\theta^5\gamma^2 + 104\theta^4\gamma^4 \\ -212\theta^4\gamma^3 + 113\theta^4\gamma^2 - 6\theta^4\gamma + 48\theta^3\gamma^4 - 268\theta^3\gamma^3 + 340\theta^3\gamma^2 - 95\theta^3\gamma + 2\theta^3 \\ -66\theta^2\gamma^4 + 12\theta^2\gamma^3 + 224\theta^2\gamma^2 - 204\theta^2\gamma + 26\theta^2 + 24\theta\gamma^4 \\ +6\theta\gamma^3 + 36\theta\gamma^2 - 132\theta\gamma + 48\theta - 3\gamma^4 - 12\gamma^2 + 24 \end{array} \right)}{\theta^4(\theta\gamma - 1)^4}\end{aligned}$$

le coefficient de variation ($C.V$), le coefficient de Skewness ($\sqrt{\beta_1}$), le coefficient de Kurtosis (β_2) de ZTTPD (3.5) sont obtenues comme suit :

$$C.V = \frac{\sqrt{\theta^2\gamma^2 - 2\theta^2\gamma + 4\theta\gamma^2 - 7\theta\gamma + 2\theta - \gamma^2 + 2}}{2 - \gamma} \quad (3.7)$$

$$\sqrt{\beta_1} = \frac{\left(\begin{array}{c} \theta^4\gamma^3 - 2\theta^4\gamma^2 + 12\theta^3\gamma^3 - 20\theta^3\gamma^2 + 4\theta^3\gamma + 15\theta^2\gamma^3 - 36\theta^2\gamma^2 \\ + 25\theta^2\gamma - 2\theta^2 - 12\theta\gamma^3 + 9\theta\gamma^2 + 12\theta\gamma - 6\theta + 2\gamma^3 - 4 \end{array} \right)}{\theta^3(\theta\gamma - 1)^3 \left(\frac{1}{\theta^2(\theta\gamma - 1)^2} (\theta^2\gamma^2 - 2\theta^2\gamma + 4\theta\gamma^2 - 7\theta\gamma + 2\theta - \gamma^2 + 2) \right)^{\frac{3}{2}}} \quad (3.8)$$

$$\beta_2 = \frac{\left(\begin{array}{c} \theta^6\gamma^4 - 2\theta^6\gamma^3 + 28\theta^5\gamma^4 - 45\theta^5\gamma^3 + 6\theta^5\gamma^2 + 104\theta^4\gamma^4 \\ - 212\theta^4\gamma^3 + 113\theta^4\gamma^2 - 6\theta^4\gamma + 48\theta^3\gamma^4 - 268\theta^3\gamma^3 + 340\theta^3\gamma^2 - 95\theta^3\gamma + 2\theta^3 \\ - 66\theta^2\gamma^4 + 12\theta^2\gamma^3 + 224\theta^2\gamma^2 - 204\theta^2\gamma + 26\theta^2 + 24\theta\gamma^4 \\ + 6\theta\gamma^3 + 36\theta\gamma^2 - 132\theta\gamma + 48\theta - 3\gamma^4 - 12\gamma^2 + 24 \end{array} \right)}{(\theta^2\gamma^2 - 2\theta^2\gamma + 4\theta\gamma^2 - 7\theta\gamma + 2\theta - \gamma^2 + 2)^2} \quad (3.9)$$

2.2 Propriétés statistiques de ZTPPD

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

Propriétés de fiabilité

Pour tout x et avec (3.5), on peut déduire que :

$$\frac{\mathbb{P}(x+1, \theta, \gamma)}{\mathbb{P}(x, \theta, \gamma)} = \left(1 + \frac{(1-\gamma)\theta}{\gamma + (1-\gamma)\theta x} e^{-\theta} \right) \quad (3.10)$$

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

Fonctions génératrice

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

$$G_X(t) = \mathbb{E}(t^X) = \frac{\theta}{1-\theta\gamma} \left[\gamma \frac{1}{\theta - \ln(t)} + (1-\gamma)\theta \frac{1}{(\theta - \ln(t))^2} \right] \quad (3.11)$$

2. Distribution à deux paramètres zéro tronquée

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

$$G_x(t) = E[e^{tx}] = \frac{\theta}{1 - \theta\gamma} \left[\gamma \frac{1}{\theta - t} + (1 - \gamma) \theta \frac{1}{(\theta - t)^2} \right]$$

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

$$G_X(t) = \mathbb{E}(e^{tX}) = \frac{\theta e^t}{1 + 2\theta + \gamma + \theta\gamma} \left[\frac{(\theta\gamma + \gamma + \theta)}{(\theta + 1 - e^t)} + \frac{\theta(\theta + 1)}{(\theta + 1 - e^t)^2} \right] \quad (3.12)$$

2.3 Estimation des paramètres

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 (3.5). La fonction de maximum de vraisemblance est donc :

$$L(x_i; \theta, \gamma) = \prod_{i=1}^n \frac{\theta[\gamma + (1 - \gamma)\theta x_i]e^{-\theta x_i}}{1 - \theta\gamma}$$

Et le logarithme de la vraisemblance est :

$$n \ln(\theta) + \sum_{i=1}^n \ln(\gamma + (1 - \gamma)\theta x_i) - \theta \sum_{i=1}^n x_i - n \ln(1 - \theta\gamma)$$

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

$$\frac{\partial \ln L(x_i; \theta, \gamma)}{\partial \theta} = \frac{\partial \ln l(x_i; \theta, \gamma)}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^n \frac{(1 - \gamma)x_i}{\gamma + (1 - \gamma)\theta x_i} - \sum_{i=1}^n x_i - \frac{n\gamma}{1 - \theta\gamma} \quad (3.13)$$

$$\frac{\partial \ln l(x_i; \theta, \gamma)}{\partial \gamma} = \sum_{i=1}^n \frac{1 - \theta x_i}{\gamma + (1 - \gamma)\theta x_i} + \frac{n\theta}{1 - \theta\gamma} \quad (3.14)$$

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

Les deux équations (3.13) et (3.14) ne peuvent pas être résolues directement, on doit

2. Distribution à deux paramètres zéro tronquée

utiliser la méthode Fisher Scoring, on a :

$$\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{-n}{\theta^2} - \sum_{i=1}^n \frac{[(1-\gamma)x_i]^2}{(\gamma + (1-\gamma)\theta x_i)^2} - \frac{n\gamma^2}{(1-\theta\gamma)^2} \quad (3.15)$$

$$\frac{\partial^2 \ln l}{\partial \gamma^2} = - \sum_{i=1}^n \frac{(1-\theta x_i)^2 \theta x_i}{[\gamma + (1-\gamma)\theta x_i]^2} + \frac{n\theta^2}{(1-\theta\gamma)^2} \quad (3.16)$$

$$\frac{\partial^2 \ln l}{\partial \theta \partial \gamma} = \frac{\partial^2 \ln l}{\partial \gamma \partial \theta} = - \sum_{i=1}^n \frac{x_i[\gamma + (1-\gamma)\theta x_i] - (1-\gamma)\theta x_i^2}{[\gamma + (1-\gamma)\theta x_i]^2} - \frac{n}{(1-\theta\gamma)^2} \quad (3.17)$$

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

$$\begin{bmatrix} \frac{\partial^2 \ln L}{\partial \theta^2} & \frac{\partial^2 \ln L}{\partial \theta \partial \gamma} \\ \frac{\partial^2 \ln L}{\partial \gamma \partial \theta} & \frac{\partial^2 \ln L}{\partial \gamma^2} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\gamma}=\gamma_0}} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\gamma} - \gamma_0 \end{bmatrix} = \begin{bmatrix} \frac{\partial \ln L}{\partial \theta} \\ \frac{\partial \ln L}{\partial \gamma} \end{bmatrix}_{\substack{\hat{\theta}=\theta_0 \\ \hat{\gamma}=\gamma_0}}$$

où θ_0 et γ_0 sont les valeurs initiales de θ et γ 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 approxi-mées aux estimateurs de maximum de vraisemblance $\hat{\theta}$ et $\hat{\gamma}$ des paramètres θ et γ respectivement.

Estimations par la méthode des Moments

Utilisant m et m_2 le premier et le deuxième moment de la distribution ZTTPD, on a

$$\begin{cases} m = \frac{2-\gamma}{\theta(1-\theta\gamma)} \\ m_2 = \frac{(2-\gamma)\theta + (6-4\gamma)}{\theta^2(1-\theta\gamma)} \end{cases} \quad (3-18)$$

où $m_2 = S^2 + m^2$ et S^2 est la variance. on résolut ce système non linière on trouve le couple $(\hat{\theta}, \hat{\gamma})$, où $(\hat{\theta}, \hat{\gamma}) \succ 0$ pour toute $S \succ 0, m \succ 0$. La solution du système non linière (3-18) donne :

2. Distribution à deux paramètres zéro tronquée

$$(-5m - m_2)\theta^2 + (4m + 4)\theta - 2 = 0 \quad \text{and} \quad \gamma = \frac{\theta m - 2}{\theta^2 m - 1}.$$

il est facile d'avoir que la solution de :

$$(-5m - m_2)\theta^2 + (4m + 4)\theta - 2 = 0$$

est $\theta = \frac{1}{5m + m_2} (2m + \sqrt{2}\sqrt{-m - m_2 + 2m^2 + 2} + 2)$ si $-5m - m_2 \neq 0$
alors,

$$\hat{\theta} = \frac{1}{5m + m_2} \left(2m + \sqrt{2}\sqrt{-m - m_2 + 2m^2 + 2} + 2 \right) \quad \text{et} \quad \hat{\gamma} = \frac{2 - \hat{\theta}m}{\hat{\theta}m - 1} \quad (3-19)$$

Simulation

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

- Choisir les valeurs initiale de θ_0, α_0 pour les éléments correspondants du vecteur de $\Theta = (\theta, \gamma)$ pour spécifier la distribution ZTPD.

- choisir la taille de l'échantillon n ;
- générer N échantillon indépendantes de taille n à partir de $ZTPD(\theta, \gamma)$;
- 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

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

3. Application aux ensembles de données réelles

	$\theta=0.01, \gamma=0.1$		$\theta=0.01, \gamma=0.01$		$\theta=0.01, \gamma=1$	
	bais(θ)	bais(γ)	bais(θ)	bais(γ)	bais(θ)	bais(γ)
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

	$\theta=0.05, \gamma=0.5$		$\theta=1, \gamma=0.5$		$\theta=5, \gamma=0.5$	
	bais(θ)	bais(γ)	bais(θ)	bais(γ)	bais(θ)	bais(γ)
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

Tableau1. Biais moyenne des estimations simulées de la distribution *ZTPD*

	$\theta=0.01, \gamma=0.1$		$\theta=0.01, \gamma=0.01$		$\theta=0.01, \gamma=1$	
	EQM(θ)	EQM(γ)	EQM(θ)	EQM(γ)	EQM(θ)	EQM(γ)
n = 10	0.000000066	0.000315046	$6.3 \cdot 10^{-11}$	0.0000025511	0.000000697	0.078133224
n = 30	0.000000022	0.000105015	$2 \cdot 10^{-11}$	0.0000008504	0.0000002323	0.026044408
n = 50	0.000000013	0.000063009	10^{-11}	0.0000005102	0.0000001394	0.015626644

	$\theta=0.05, \gamma=0.5$		$\theta=1, \gamma=0.5$		$\theta=5, \gamma=0.5$	
	EQM(θ)	EQM(γ)	EQM(θ)	EQM(γ)	EQM(θ)	EQM(γ)
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

Tableau2. l'erreur quadratique moyenne

3 Application aux ensembles de données réelles

La qualité de l'ajustement de la distribution *ZTPD* à été discutée avec des données de comptage et l'ajustement a été comparé aux distributions *ZTPD* et *ZTPLD*. Dans cette section, la qualité de l'ajustement des distributions *ZTPD*, *ZTPLD* et *ZTTPD* a été comparée pour un ensemble de données dû à Finney et Varley (1955), qui ont donné des chiffres fleur ayant le nombre d'oeufs de mouche. Les valeurs de

3. Application aux ensembles de données réelles

l'estimation du maximum de vraisemblance des paramètres, le χ^2 , les p valeurs sont indiquées dans le **tableau 3**, où la valeur du χ^2 de la distribution ZTTPD est inférieure à celle de la distribution ZTPD et ZTPLD mais la p valeur est autrement. Cela est dû à la différence dans le nombre de paramètres. Les distributions ZTP et ZTPL sont des distributions à un paramètre, tandis que ZTTPD est une distribution à deux paramètres. 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**.

Dans le tableau **3**, les valeurs du log-vraisemblance négatif sont minimales pour ZTTPD par rapport à 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 ZTTPD. Notre distribution peut donc être considérée comme un modèle important

3. Application aux ensembles de données réelles

pour les données de science biologique.

<i>Nombre de</i>	<i>Fréquence</i>	<i>ZTPD</i>	<i>ZTPLD</i>	<i>ZTTPD</i>
<i>particules</i>	<i>observées</i>	$\hat{\theta} = 2,8604$	$\hat{\theta} = 0,7186$	$\hat{\theta} = 1,24$
<i>attachées</i>				$\hat{\gamma} = 0,87$
1	22	15.3	26.8	23.1
2	18	21.8	19.8	18.2
3	18	20.8	14.0	18.1
4	11	14.9	9.5	10.5
5	9	8.5	6.3	8.1
6	6	4.0	4.2	5.7
7	3	1.7	2.7	3.3
8	0	0.6	1.7	0.2
9	1	0.4	3.0	0.8
<i>Total</i>	88	88	88	88
χ^2		6.648	3.780	0.443
<i>d.l</i>		4	4	4
<i>P – value</i>		0.1557	0.4366	0.05
$-2 \log L$		333.09	334.76	123.34
<i>AIC</i>		336.76	335.09	142.43
<i>AICC</i>		335.14	336.80	142.57
<i>BIC</i>		335.04	336.70	132.29

Tableau 3 : Nombre de têtes de fleurs correspondant au nombre d'œufs de mouche rapporté par Finney et Varley (1955).

Conclusion et Perspectives

Nous avons réussi à introduire une nouvelle distribution nommée "Distribution à deux paramètres zéro tronquée" . L'estimation de ses paramètres à été discutée en utilisant la méthode du maximum de vraisemblance et la méthode des moments. Une étude de simulation est réalisée pour examiner le biais et l'erreur quadratique moyenne des estimateurs du maximum de vraisemblance des paramètres. 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 pouvons montrer que la Distribution à deux paramètres zéro tronquée par la taille peut être utilisée assez efficacement pour analyser des données réelles et la science actuarielle.

BIBLIOGRAPHIE

- [1] **Akaike H. (1974)**, A new look at statistical model identification, IEEE Transactions on Automatic Control AU-19 : 716-722.
- [2] **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.
- [3] **David F. N., Johnson L.N. (1952)**, The truncated Poisson. Biometrics 8 : 272-285.
- [4] **Finney D.J., Varley G.C. (1955)**, An example of the truncated Poisson distribution. Biometric, 11, 387-394.
- [5] **Fisher R.A (1937)**, The effects of methods of ascertainment upon the estimation of frequency. Ann. Eugenics. 6(1) : 13-25.
- [6] **Ghitany M. E., Atieh B., Nadarajah S. (2008a)**, Lindley distribution and its applications. *Math. Comput. Simulation* , 78, pp. 493-506.
- [7] **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.

-
- [8] **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.
- [9] **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.
- [10] **Grine R., Zeghdoudi H (2017)**, On Poisson Quasi-Lindley Distribution and its Applications. *Journal of Modern Applied Statistical Methods*, 16(2), 403-417.
- [11] **Grine R., Zeghdoudi H (2024)**, Novel Two Parameter Model : Statistical Properties, Fuzzy Reliability and Applications. *MAS Journal of Applied Sciences*. Vol 9, 965–973.
- [12] **Johnson N L, Kotz S and Kemp A W.(2005)**, *Univariate Discrete Distribution*, third edition, John Wiley, New York.
- [13] **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.
- [14] **Lecoutre J.P. (2002)**, *Statistique et probabilités*. Dunod. Paris, 2^{ième} édition, ISBN 2-10-007277-3.
- [15] **Lejeune M. (2010)**, *Statistique La théorie et ses applications*. Springer-Verlag France, Paris, 2^{ième} édition. ISBN 978-2-8178-0156-8.
- [16] **Mathews JNS., Appleton DR. (1993)**, An application of the truncated Poisson distribution to Immunogold assay. *Biometrics* 49(2), 617-621.
- [17] **Munir A. and A.(1968)**, The maxiaum likelihood estimation of right hand truncated Poisson distribution. Unpublished paper. Ames, Iowa, Department of Statistics, Iowa State University.
- [18] **Munir A. (1968)**, Truncated multivariate Poisson distributons, Iowa state university of science and technology,Ames,Iowa.
- [19] **Nedjar S., Zeghdoudi H. (2020)**, New compound Poisson distribution and Study its Inflated Distribution. *Thailand Statistician*, (submitted).

- [20] **Rider P. R. (1953)**, Truncated Poisson Distributions. *Journal of the American Statistical Association*, 48, 826–830.
- [21] **Sankaran M. (1970)**, The discrete Poisson-Lindley distribution. *Biometrics*, 26, pp. 145-149.
- [22] **Saporta G. (2006)**, Probabilités analyse des données et statistiques. Editions Technip. Paris, 2^{ième} édition. ISBN 2-7108-0814-5.
- [23] **Schwarz G. (1978)**, Estimating the dimension of a model, *Annals of Statistics* 6 : 461-464.
- [24] **Shanker R., Hagos F., Sujatha, S. (2015)**, On Zero-Truncated of Poisson and Poisson-Lindley distribution and their applications, 11(6),1-14.
- [25] **Shanker R., Shukla K.K (2019)**, A generalization of zero-truncated Poisson-Sujatha distribution. *Hungarian Statistical Review*, 2 (2). pp. 32-50.
- [26] **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.
- [27] **Sheldon M. Ross (1987)**, Initiation aux probabilités. Presses polytechniques romandes 1ière édition, ISBN 2-88074-118-8.
- [28] **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.
- [29] **Zaninetti L., Ferraro M. (2008)**, On the truncated Pareto distribution with applications, *Central European Journal of Physics* 6. 1–6.