The Poisson-Conjugate Lindley Mixture Distribution

E. Gómez-Déniz[a] and E. Calderín-Ojeda[b]

[a] *Department of Quantitative Methods in Economics and TiDES Institute, University of Las Palmas de Gran Canaria*
[b] *Centre for Actuarial Studies, Department of Economics, The University of Melbourne, Australia*

Abstract

A new discrete distribution that depends on two parameters is introduced in this paper. From this new distribution the geometric distribution is obtained as a special case. After analyzing some of its properties such as moments and unimodality, recurrences for the probability mass function and differential equations for its probability generating function are derived. In addition to this, parameters are estimated by maximum likelihood estimation numerically maximizing the log-likelihood function. Expected frequencies are calculated for different sets of data to prove the versatility of this discrete model.

Keywords: Geometric Distribution, Mixture, Natural Exponential Family, EM Algorithm.

Acknowledgements

EGD thanks MICINN (Spain) (project ECO2009-14152) for partial support of this work. The authors are indebted to the Associate Editor and the anonymous referees for helpful comments which, without doubt, helped to improve an earlier version of the paper.

Address for correspondence: Emilio Gómez Déniz, Department of Quantitative Methods, University of Las Palmas de Gran Canaria, 35017-Las Palmas de Gran Canaria, Spain. E-mail: `egomez@dmc.ulpgc.es`

The Poisson-Conjugate Lindley Mixture Distribution

Abstract

A new discrete distribution that depends on two parameters is introduced in this paper. From this new distribution the geometric distribution is obtained as a special case. After analyzing some of its properties such as moments and unimodality, recurrences for the probability mass function and differential equations for its probability generating function are derived. In addition to this, parameters are estimated by maximum likelihood estimation parameters are estimated by maximum likelihood estimation numerically maximizing the log-likelihood function. Expected frequencies were calculated for different sets of data to prove the versatility of this discrete model.

Keywords: Geometric Distribution, Mixture, Natural Exponential Family, EM Algorithm.

1 Introduction

Mixtures of distributions define one of the most important approaches to obtain new probability distributions within the fields of applied probability and operational research. Among all of them, mixture of Poisson distribution with parameter $\theta > 0$ and mixing distribution $g(\theta)$ defined by

$$\Pr(X = x) = \frac{1}{x!} \int_0^\infty \exp(-\theta)\theta^x \, g(\theta) \, d\theta, \quad x = 0, 1, \ldots,$$

has played an important role in the statistical literature (see Antzoulakos and Chadjiconstantinidis (2004), Gupta and Ong (2005), Karlis and Xekalaki (2005) and Willmot (1993), among others). In this case, the density function $g(\theta)$ is intended to capture dependencies on hidden variables or incorporating individual heterogeneity.

The applicability of Poisson distribution to different scientific scenarios has been accepted over more than a century. However, modification of the usual scheme when the parameter involved in the distribution fluctuates randomly as a gamma variate, was first considered by Greenwood and Yule (1920)

in the study of accident proneness with reference to happenings of multiple attacks of a disease or repeated accidents; the distribution obtained in this chapter was a negative binomial (mixing a Poisson distribution with the gamma distribution). Alternatively, different models similarly built by considering other mixing distributions have been proposed in the statistical literature. For example, the Poisson-inverse Gaussian distribution (giving the Sichel's distribution), the Poisson-beta distribution, the Poisson-Pareto and the Poisson-generalized Pareto among others. Certainly, the main purpose of these models is to search for a more flexible alternative to the Poisson distribution, especially under over-dispersion (variance larger than the mean) phenomena.

In order to provide another competitive alternative to the models described above, a new mixed Poisson model is considered in this paper. In this regard, a new mixture of Poisson distribution by mixing its parameter with the generalized exponential distribution proposed not long ago by Gómez-Déniz (2012) is considered. In the last decades, a lot of attempts have been made to define new families of probability distributions (discrete or continuous) that extend the well-known families of distributions. In this sense, the new distribution proposed here generalizes the geometric distribution and it is competitive with other two-parameters discrete distributions such as the negative binomial, generalized Poisson, Poisson-inverse Gaussian, generalized Poisson-Lindley, etc. Different generalizations of the geometric distribution have been proposed over the last decades; see, for example Gómez-Déniz (2010), Jain and Consul (1971), Makčutek (2008), Philippou et al. (1983), Tripathi et al. (1987).

The new distribution has a closed-form expression for the probability mass function (pmf) associated with special functions (Tricomi confluent hypergeometric function) of mathematical physics. These functions, that show excellent properties, have been recently proposed by many authors. Two important features of this discrete distribution are unimodality and the presence of thick tails. Furthermore, since $d \log g(\theta)/d\theta$ can be expressed as a ratio of polynomials, i.e. it is a generalization of the Pearson system, recursive expressions for compound mixed Poisson distribution and the evaluation of the i-th moments of stop-loss transforms can be achieved. Finally, ordinary differential equations involving the probability generating function of the new distribution are given.

In addition to this, the question of parameter estimation is analyzed via maximum likelihood estimation (MLE) numerically maximizing the log-

likelihood surface; this can be done either by directly maximizing the log-likelihood function or by using an expectation-maximization (EM) algorithm. Besides, from a numerical perspective, the discrete model introduced in this paper provides a satisfactory performance, being therefore considered competitive with other two-parameters mixed Poisson models such as the negative binomial, Poisson-inverse Gaussian, discrete Weibull, generalized Poisson and hyper-Poisson distributions.

It is already known (see for instance, Sankaran (1970)) that when Poisson distribution with parameter $\theta > 0$ is mixed with a member of the natural exponential family of distributions, i.e. with probability density function (pdf) distribution

$$g(\theta) = q(\theta)\, \exp[-\phi\, \theta - \log \psi(\phi)], \tag{1}$$

being ϕ the natural parameter, $q(\theta)$ a function which depends on θ but not on the natural parameter and $\psi(\phi) = \int_0^\infty q(\theta)\, \exp(-\phi\, \theta)\, d\theta$ the normalization constant, then the resulting distribution is given by

$$p_x = \frac{\psi(\phi+1)}{\psi(\phi)} \frac{\mu_x(\phi+1)}{x!}, \quad x = 0, 1, \ldots \tag{2}$$

where $\mu_x(\phi+1)$ represents the raw moment of order x of (1) with $\phi+1$ as parameter. Additionally, the probability generating function is given by

$$G_X(s) = \frac{\psi(\phi - s + 1)}{\psi(\phi)}. \tag{3}$$

The remainder of the paper proceeds as follows. Section 2 briefly summarizes the mixing distribution proposed not long ago by Gómez-Déniz (2012). Section 3 describes the theoretical development of the new discrete distribution. Next, parameter estimation together with the derivation of a simulation algorithm of the discrete model introduced in this paper are provided in Section 4. Then, numerical applications of the proposed model is examined in section 5 and finally, conclusions are discussed in the last section.

2 The background

Recently, Gómez-Déniz (2012) has proposed a new generalization of the exponential distribution whose pdf is given by

$$g(\theta; \lambda, \sigma) = \kappa(\sigma, \lambda) \left(\frac{\theta^2}{1+\theta}\right)^\sigma \exp(-\lambda\, \theta), \quad \theta \in \Theta > 0, \tag{4}$$

where

$$\kappa(\sigma,\lambda)=\frac{1}{\Gamma(2\sigma+1)\,\mathcal{U}(2\sigma+1,2+\sigma,\lambda)}$$

is the normalization constant. Here $\sigma>-1/2$, $\lambda>0$ and

$$\begin{aligned}\Gamma(z) &= \int_0^\infty t^{z-1}\exp(-t)\,dt,\\ \mathcal{U}(a,b,z) &= \frac{1}{\Gamma(a)}\int_0^\infty t^{a-1}(1+t)^{-a+b-1}\exp(-z\,t)\,dt,\quad z>0,\ a>0 \quad (5)\end{aligned}$$

are the Euler gamma function and the Tricomi confluent hypergeometric function, respectively.

It is not difficult to prove that (4) is a member of the NEF of distributions given by (1) being $\phi=\lambda$ the natural parameter, while $q(\phi)=\left(\frac{\phi^2}{1+\phi}\right)^\sigma$ and $\psi(\phi)=[\kappa(\sigma,\phi)]^{-1}$.

It can be observe that for $\sigma=0$ the normalization constant reduces to λ and, therefore, (4) is the exponential distribution with parameter $\lambda>0$. This pdf can therefore be considered as an alternative to both, the generalized exponential distribution in Marshall and Olkin (1997) and the generalized exponential distribution in Gupta and Kundu (1999). Besides, as proven in Gómez-Déniz (2012), it is also competitive with the classical gamma distribution and the two-parameter Weibull distribution.

As it is shown in the same paper, the generalized exponential distribution with pdf given in (4) can be obtained as the natural conjugate prior distribution of the continuous Lindley distribution (Lindley (1958)). Moreover, the r-th moment around the origin of a random variable following the pdf (4) is

$$E(X^r)=\frac{\Gamma(2\sigma+r+1)\,\mathcal{U}(2\sigma+r+1,\sigma+r+2,\lambda)}{\Gamma(2\sigma+1)\mathcal{U}(2\sigma+1,2+\sigma,\lambda)}. \quad (6)$$

Furthermore, since Lindley distribution and the generalized exponential distribution with pdf given in (4) distributions are members of the NEF of distributions, we have that for the samples $(Y_1,Y_2,\ldots,Y_n)$ and $(X_1,X_2,\ldots,X_n)$, the statistics $T(Y)=\sum_{i=1}^n Y_i$ and $T(X)=\sum_{i=1}^n X_i$ are sufficient. Therefore, both statistics are also part of the NEF of distributions.

3 The main results

Basic properties of a Poisson-generalized exponential distribution are introduced in this section. It is easy to see, after simple manipulations, that when (6) is applied to (2), the resulting pmf is given by

$$p_x = \Pr(X = x) = \frac{\Gamma(x + 2\sigma + 1)}{x!\,\Gamma(2\sigma + 1)} \frac{\mathcal{U}(1 + 2\sigma + x, 2 + \sigma + x, 1 + \lambda)}{\mathcal{U}(1 + 2\sigma, 2 + \sigma, \lambda)}, \qquad (7)$$

with $x = 0, 1, 2, \ldots$; $\sigma > -1/2$ and $\lambda > 0$. Furthermore, from (3) the probability generating function of a random variable following (7) is

$$G_X(s) = \frac{\mathcal{U}(1 + 2\sigma, 2 + \sigma, 1 - s + \lambda)}{\mathcal{U}(1 + 2\sigma, 2 + \sigma, \lambda)}. \qquad (8)$$

Besides, the factorial moment of order k is given by

$$\begin{aligned} \mu_{[k]}(X) &= E[X(X-1)\cdots(X-k+1)] \\ &= \frac{\lambda^{1+\sigma}\Gamma(k + 2\sigma + 1)}{\Gamma(1 + 2\sigma)} \frac{\mathcal{U}(1 + k + 2\sigma, 2 + k + \sigma, \lambda)}{\mathcal{U}(\sigma, -\sigma, \lambda)} \end{aligned} \qquad (9)$$

with $k = 1, 2, \ldots$.

From (8) or alternatively (9), the first two moments of X about the origin are found to be,

$$\mu_1' = \frac{(2\sigma + 1)\,\mathcal{U}(2\sigma + 2, \sigma + 3, \lambda)}{\mathcal{U}(2\sigma + 1, \sigma + 2, \lambda)}, \qquad (10)$$

$$\mu_2' = \frac{(2\sigma + 1)\,\mathcal{U}(2 + 2\sigma, \sigma + 4, \lambda)}{\mathcal{U}(2\sigma + 1, \sigma + 2, \lambda)}. \qquad (11)$$

By using the identity $\mu_r = E[(X - \mu)^r] = \sum_{k=0}^{r} \binom{r}{k} \mu_k' (-\mu_1')^{r-k}$, the central moments, which are not reproduced here can be obtained after some algebra.

Some important indices of the shape of the distribution, apart of the mean and variance, are the skewness ($\sqrt{\beta_1} = \mu_3/(\mu_2)^{3/2}$), the kurtosis ($\beta_2 = \mu_4/(\mu_2)^2$) and the coefficient of variation ($C.V. = \sigma/\mu$). Different values of these coefficients are shown in Table 1 for selected values of the parameters.

In addition to this, by using the nice property provided by Chao and Strawderman (1972) which relates the inverse moments of a discrete random

Table 1: Mean, variance, skewness, kurtosis and coefficient of variation for selected values of parameters

	σ	μ_1'	μ_2	$\sqrt{\beta_1}$	β_2	C.V.
	–0.25	6.66	77.20	5.15	23.04	1.32
	0.0	10.00	110.00	5.73	22.92	1.05
	0.5	15.54	166.96	6.96	25.47	0.83
$\lambda = 0.1$	1.0	20.73	222.10	8.26	29.10	0.72
	2.0	30.87	331.82	11.09	38.24	0.59
	3.0	40.91	441.60	14.22	49.37	0.51
	4.0	50.94	551.46	17.60	62.31	0.46
	5.0	60.95	661.37	21.24	76.99	0.42
	–0.25	0.56	1.17	4.58	22.59	1.91
	0.0	1.00	2.00	4.59	18.75	1.41
	0.5	1.72	3.34	5.25	18.80	1.06
$\lambda = 1$	1.0	2.35	4.52	6.03	20.55	0.90
	2.0	3.51	6.70	7.67	25.26	0.73
	3.0	4.60	8.80	9.38	30.72	0.64
	4.0	5.67	10.85	11.14	36.73	0.58
	5.0	6.71	12.89	12.97	43.26	0.53
	–0.25	0.27	0.41	4.52	23.42	2.37
	0.0	0.50	0.75	4.23	17.77	1.73
	0.5	0.89	1.31	4.51	16.30	1.28
$\lambda = 2$	1.0	1.24	1.81	4.99	16.96	1.08
	2.0	1.88	2.72	6.05	19.62	0.87
	3.0	2.47	3.57	7.17	22.88	0.76
	4.0	3.03	4.38	8.30	26.46	0.68
	5.0	3.58	5.18	9.47	30.30	0.63

variable to the integration of the probability generating function (see also Cressie et al. (1981) and Kabe (1976)), it is possible to obtain that

$$E\left(\frac{1}{X+1}\right) = \frac{\lambda}{2\sigma}\frac{\mathcal{U}(\sigma,1-\sigma,\lambda)-\lambda^{\sigma}\mathcal{U}(2\sigma,1+\sigma,1+\lambda)}{\mathcal{U}(\sigma,-\sigma,\lambda)}.$$

Now, by using the relation

$$\mathcal{U}(a,b,z)=z^{1-b}\mathcal{U}(1+a-b,2-b,z),$$

it is possible to rewrite (7) as

$$p_x=\frac{\Gamma(x+2\sigma+1)}{x!\,\Gamma(2\sigma+1)}\left(\frac{\lambda}{\lambda+1}\right)^{1+\sigma}\left(\frac{1}{\lambda+1}\right)^{x}\frac{\mathcal{U}(\sigma,-\sigma-x,1+\lambda)}{\mathcal{U}(\sigma,-\sigma,\lambda)}. \qquad (12)$$

Note that for $\sigma=0$ the geometric distribution with pmf $p_x=\frac{\lambda}{1+\lambda}\left(\frac{1}{1+\lambda}\right)^x$ is obtained. Furthermore, since $\mathcal{U}(a,b,z)\sim z^{-a}$ for large values of z, the negative binomial distribution with parameters $2\sigma+1$ and $\lambda/(1+\lambda)$ is derived for large values of the parameter λ. In Figure 1 first (left hand side graph) and second (right hand side graph) order moments about the origin, written as a function of λ, for different values of σ are shown. As it can be observed both moments decrease with λ. Besides, the larger is the value of σ, the greater are the values of μ_1' and μ_2'. It is also noted that the latter expressions, for each value of λ and positive values of σ ($\sigma=1$ dashed line and $\sigma=2$ dotdashed line) are always higher than μ_1' and μ_2' for $\sigma=0$ (geometric distribution with thick solid line). The opposite occurs when σ takes negative values ($\sigma=-0.25$ dotted line).

Some examples of the graphs of pmf (7) for different values of the parameters, σ and λ are displayed in Figure 2.

It is important to point out that $p_0<\frac{\lambda}{1+\lambda}$ when $\sigma>0$ and, therefore, (7) has a lower value at $x=0$ than the geometric distribution for $\lambda>0$.

Proposition 1 *The distribution with pmf given in (7) is unimodal.*

Proof: It is not difficult to observe that

$$(\log(g(\theta;\lambda,\sigma))''=-\sigma\frac{2+x(x+4)}{x^2(1+x)^2}<0.$$

Therefore the mixing distribution is log–concave and also strongly unimodal and unimodal. Now the result follows by applying a result in Holgate (1970). ■

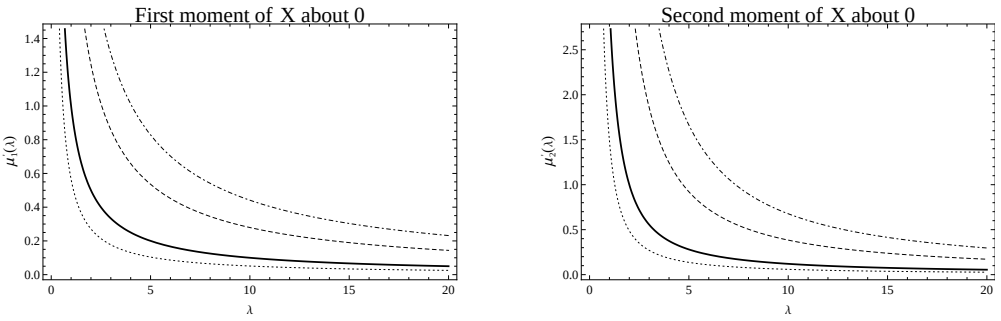

Figure 1: First and second order moments (as a function of λ) about the origin of (7) for different values of σ (top to bottom $\sigma = 2$ (dotdashed) , $\sigma = 1$ (dashed), $\sigma = 0$ (solid) and $\sigma = -0.25$ (dotted)) are shown.

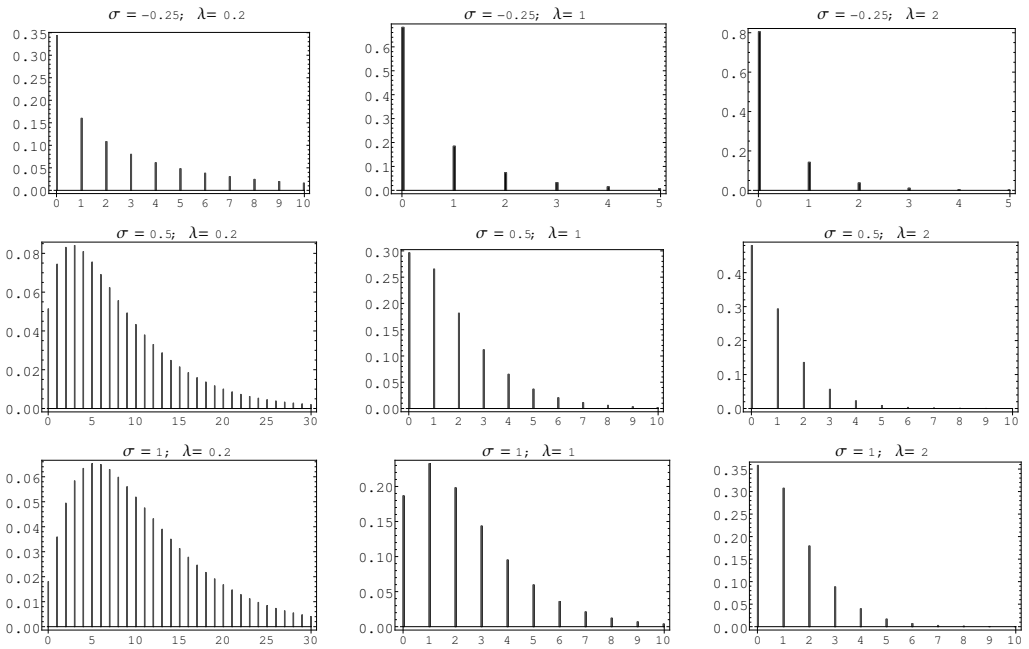

Figure 2: Graph of the pmf (7) for selected values of parameters σ and λ

Now, let p_k be the probability function (7) and $\Pr(k|m)$ be the probability mass function of a simple Poisson distribution with the same mean, say m. Then, as shown by Feller (1943), $p_0 \geq \Pr(0|m)$ and $p_1/p_0 \leq \Pr(1|m)/\Pr(0|m) = m$. The asymptotic tail behavior of Poisson distributions has been studied by Willmot (1990) by assuming that a mixed Poisson distribution has a longer right tail that the Poisson distribution (see, for instance, Gupta and Ong (2005) and Karlis and Xekalaki (2005)).

As it is described in the next proposition, the pmf (7) can be recursively computed in terms of σ and λ.

Proposition 2 *By denoting p_x a random variable following the pmf (7), it is satisfied that*

$$(2\sigma + x)p_{x-1} + (\sigma - \lambda + x)xp_x - (\lambda + 1)(x + 1)xp_{x+1} = 0, \quad x = 1, 2, \ldots \quad (13)$$

Proof: It is simple to see that

$$\frac{d}{d\theta} \log \pi(\theta) = \frac{\phi(\theta)}{\psi(\theta)} = \frac{\sum_{i=1}^{r} \phi_i \, \theta^i}{\sum_{i=1}^{r} \psi_i \, \theta^i}, \quad (14)$$

being

$$\begin{aligned} \phi(\theta) &= -\lambda \, \theta^2 + (\sigma - \lambda) \, \theta + 2\sigma \text{ and} \\ \psi(\theta) &= \theta^2 + \theta. \end{aligned}$$

Now, the result follows by using the recursive relation (see Willmot (1993) and Antzoulakos and Chadjiconstantinidis (2004)

$$\sum_{i=-1}^{r} [\phi_i - \psi_i + (x + i + 1)\psi_{i+1}](x + i)^{(i)} p_{x+i} = 0,$$

where $a^{(b)} = \prod_{i=1}^{b}(a + 1 - i)$ and $p_{-1} = 0$. ■

In the following, an ordinary differential equation satisfied by the probability generating function (8) is obtained. For that reason, let us denote as $G_X(s|\sigma, \lambda)$ the probability generating function in (8).

Theorem 1 *The probability generating function satisfies the following differential equation*

$$(s - \lambda - 1) \, G''_X(s|\sigma, \lambda) + (s + \sigma - \lambda + 1) \, G'_X(s|\sigma, \lambda) + (2\sigma + 1) \, G_X(s|\sigma, \lambda) = 0. \quad (15)$$

Proof: From (13), it is straightforward that

$$(\lambda+1)x(x-1)p_x = (2\sigma+x-1)p_{x-2} + (\sigma-\lambda+x-1)p_{x-1}.$$

Consequently, we have that

$$\begin{aligned} s^2(\lambda+1)\sum_{x=2}^{\infty} x(x-1)s^{x-2}p_x &= (2\sigma-1)s^2 G_X(s) + \sum_{x=2}^{\infty} x s^x p_{x-2} \\ &+ (\sigma-\lambda-1)s^2\sum_{x=2}^{\infty}(x-1)s^{x-2}p_{x-1} \\ &+ s^2\sum_{x=2}^{\infty} x(x-1)s^{x-2}p_{x-1}. \end{aligned}$$

Now, by having into account that

$$\begin{aligned} \sum_{x=2}^{\infty} x(x-1)s^{x-2}p_{x-1,\sigma,\lambda} &= \frac{d^2}{ds^2}\sum_{x=2}^{\infty} s^x p_{x-1,\sigma,\lambda} = \frac{d^2}{ds^2}\left\{ s\sum_{x=2}^{\infty} s^{x-1}p_{x-1,\sigma,\lambda}\right\} \\ &= \frac{d^2}{ds^2}\left\{ s[G_X(s|\sigma,\lambda) - p_0]\right\} = 2G'_X(s|\sigma,\lambda) \\ &+ sG''_X(s|\sigma,\lambda), \qquad (16) \\ \sum_{x=2}^{\infty}(x-1)s^{x-2}p_{x-1,\sigma,\lambda} &= G'_X(s|\sigma,\lambda), \qquad (17) \end{aligned}$$

and considering that

$$\sum_{x=2}^{\infty} x(x-1)s^{x-2}p_x = G''_X(s),$$

the result is obtained after simple algebra. ■

Similarly, differential equations for moment and cumulant generating functions may be calculated after appropriate change of variable in equation (15).

Although the cumulative distribution function of (7) cannot be obtained in a closed-form, it can be calculated by using the following expression

$$F(x;\lambda,\sigma) = \frac{1}{x!}\int_0^{\infty} \Gamma(x+1,\theta)g(\theta;\lambda,\sigma)\,d\theta, \qquad (18)$$

where we have used the fact that the cumulative distribution function of the Poisson distribution with parameter $\theta > 0$ is $F(x;\theta) = \Gamma(x+1,\theta)/x!$, where $\Gamma(a,z)$ is the incomplete gamma function given by $\Gamma(a,z) = \int_z^\infty t^{a-1}e^{-t}dt$.

The survival function is obtained from (18) and it is given by $\bar{F}(x;\lambda,\sigma) = 1 - F(x-1;\lambda,\sigma)$. By using (7) together with $\bar{F}(x;\lambda,\sigma)$ we obtain the failure rate given by $h(x) = p_x/\bar{F}(x)$. In Figure 3 different graphs of the hazard function of the new distribution are shown for selected values of the parameters. It seems that it is decreasing for $\sigma < 0$, increasing for $\sigma > 0$ and constant (the geometric case) for $\sigma = 0$. Besides, although difficult to prove, it might be conjectured that the hazard rate is reverse J–shaped for $\sigma \leq 0$.

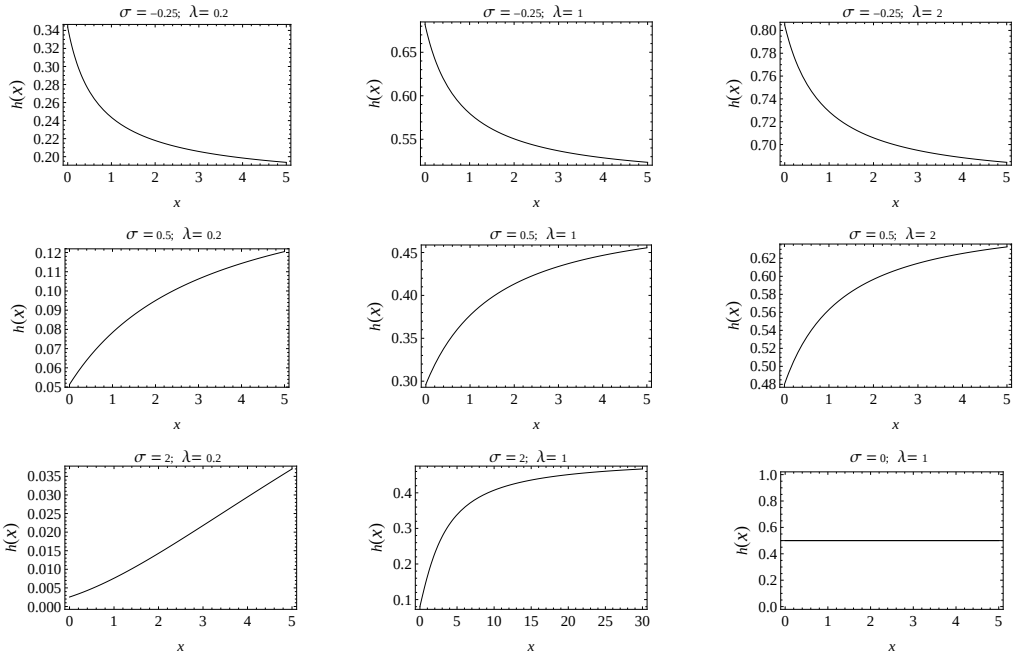

Figure 3: Different graphs of the hazard function for selected values of parameters

Finally posterior moments of parameter θ can be easily derived. On this subject, let the random variable X follow a Poisson distribution with parameter $\theta > 0$ and prior distribution $g(\theta)$; then, by using Proposition 10 in Karlis and Xekalaki (2005), the posterior expectation of θ^r given $X = x$

can be expressed as

$$E(\theta^r|X=x)=\frac{(x+r+2\sigma)!x!}{(1+\lambda)^r(x+r)!(x+2\sigma)!}\frac{\mathcal{U}(\sigma,-\sigma-x-r,1+\lambda)}{\mathcal{U}(\sigma,-\sigma-x,1+\lambda)}.$$

In particular, the posterior mean is provided by

$$E(\theta|X=x)=\frac{(x+2\sigma+1)}{(1+\lambda)(x+1)}\frac{\mathcal{U}(\sigma,-\sigma-x-1,1+\lambda)}{\mathcal{U}(\sigma,-\sigma-x,1+\lambda)}.$$

3.1 Left–truncated version of the distribution

As Mahmoudi and Zakerzadeh (2010) pointed out, in some occasions the practitioner requires a distribution with some kind of truncation. These distributions are interesting when modelling different situations, for example the number of offspring per family, the number of occupants per car, etc. Perhaps, one of the most well-known method of truncation is based on deletion of zeros from the distribution, i.e. left–truncation. The probability mass function of the zero–truncated distribution is given by

$$\Pr(X=x)=\frac{1}{1-p_0}\frac{\Gamma(x+2\sigma+1)}{x!\,\Gamma(2\sigma+1)}\left(\frac{\lambda}{\lambda+1}\right)^{1+\sigma}\left(\frac{1}{\lambda+1}\right)^{x}\frac{\mathcal{U}(\sigma,-\sigma-x,1+\lambda)}{\mathcal{U}(\sigma,-\sigma,\lambda)},$$

for $x=1,2,\ldots$, and where

$$p_0=\left(\frac{\lambda}{1+\lambda}\right)^{1+\sigma}\frac{\mathcal{U}(\sigma,-\sigma,1+\lambda)}{\mathcal{U}(\sigma,-\sigma,\lambda)}.$$

Although the left–truncated distribution with support $x=r,r+1,\ldots$ can be straightforward derived, right–truncation requires more effort since a closed-form expression for the cumulative distribution function is needed.

Additionally, zero-inflated models can be built starting with the discrete distribution proposed here in the conventional way.

3.2 Compound model

In actuarial statistics the distribution of the aggregate claims $S=\sum_{i=1}^{X}Y_i$, known as the compound distribution, is usually of interest to practitioners. Let us assume that X is the number of claims in a portfolio of policies at a time period; let us also consider that Y_i, $i=1,2,\ldots$ is a sequence of

independent and identically distributed non-negative random variables with common pdf $f(y)$, $y = 0, 1, 2, \ldots$, denoting the amount of the i-th claim. Moreover, X and $\{Y_i\}_{i\geq 1}$ are assumed to be stochastically independent.

There exists an extensive literature dealing with compound mixture Poisson distributions (Willmot (1986, 1993) and Antzoulakos and Chadjiconstantinidis (2004)). An extensive review of the topic can be found in Sundt and Vernic (2009).

Furthermore, by having into account (14) together with a result provided in Willmot (1993) (see also expression (1.7) in Antzoulakos and Chadjiconstantinidis (2004)), the following Proposition is not difficult to prove.

Proposition 3 *If the claim size is a discrete random variable with pmf $f(x)$ for $x > 0$, then the pmf, $g_s(x)$, of the compound Poisson-generalized exponential distribution satisfies the following recursion:*

$$g_S(x) = \frac{1}{1 - f(0) - \sigma + \lambda} \sum_{y=1}^{x} \left(1 + \frac{2\sigma y}{x}\right) f(y)\, g_S(x-y), \quad x \geq 1.$$

By following the works of Willmot (1986, 1993) and Antzoulakos and Chadjiconstantinidis (2004) recursive evaluation of the j-th order cumulative distribution function and the j-th order tail probabilities can also be calculated.

4 Estimation

The estimation of the two parameters by the method of moments (MM) can be achieved from (10) and (11) by setting equal their theoretical expressions to the sample first and second order sample moments respectively. Certainly, the resulting system of equations must be solved numerically and non-explicit values for the estimators are obtained. Although they are relatively easy to obtain, they tend to behave weakly since only limited information is used. In order to overcome this issue, parameters are estimated via maximum likelihood estimation (MLE) numerically maximizing the log-likelihood surface; firstly by directly maximizing the log-likelihood function, and secondly by using an expectation-maximization (EM) algorithm. They are described in the next subsection.

4.1 Maximum likelihood estimation

Let us assume that $\underline{x} = (x_1, \ldots, x_n)$ is a random sample of size n from the discrete distribution (7). The maximum likelihood estimates of the model are obtained by maximizing the following log-likelihood function

$$\begin{aligned} \ell &= \sum_{i=1}^{n} \log \frac{\Gamma(x_i + 2\sigma + 1)}{x_i!\,\Gamma(2\sigma+1)} + \sum_{i=1}^{n} \log \mathcal{U}(1 + 2\sigma + x_i, 2 + \sigma + x_i, 1 + \lambda) \\ &- n \log \mathcal{U}(1 + 2\sigma, 2 + \sigma, \lambda). \end{aligned}$$

It is already known that $\frac{\partial}{\partial c}\mathcal{U}(a, b, c) = -a\,\mathcal{U}(1 + a, 1 + b, c)$, from which the score equations are obtained after differentiating with respect to each parameter and setting the results equal to zero. These equations cannot be explicitly solved. They must be solved either by numerical method or by directly maximizing the log-likelihood function. Since the global maximum of the log-likelihood surface is not guaranteed, different initial values of the parametric space can be considered as a seed point. In this sense, by using the `FindMaximum` function of Mathematica software package v.8.0 (the derivative of the Trinomi confluent hypergeometric function is available in this package). Besides, by using other different methods such as Newton, PrincipalAxis and QuasiNewton the same result is obtained.

The second partial derivatives can be used to obtain an approximation of the Fisher's information matrix in the conventional way. This is calculated by using of an approximation of the Hessian matrix by means of the Cholesky factors. This package is available on the web upon request.

In addition to this, an expectation-maximization (EM) algorithm can be used to find maximum likelihood estimates of parameters in situations where data contains missing values.It is based on an expectation (E) step, which produces an expression for the expectation of the log-likelihood evaluated using the current estimates, and a maximization (M) step, which updates parameter estimates by maximizing the expected log-likelihood computed on the E–step. This methodology is suitable for distributions arising as mixtures since the mixing operation produces missing data. The algorithm is based on the structure behind the mixed Poisson distributions (see Karlis(2005)). In this case the problem of estimation is reduced to one of estimation of the mixing distribution. One of the main advantages of the EM algorithm is its numerical stability, increasing the likelihood of the observed data in each iteration. However, it presents slow convergence rate in a neighborhood

of the optimal point. Although, in general, the convergence to the global maximum is not guaranteed, for this particular model, the log-concavity implies that the log-likelihood has a unique maximum. It can be usually reached by starting the parameters at method of moments estimates.

In mixed Poisson distributions the unobserved quantities are the realizations of θ_i of the unobserved mixing parameter for each data point x_i, $i = 1 \dots n$. We assume the distribution of $X_i|\theta_i$ is the Poisson distribution where θ_i follows (4). Then, given the observations $\mathbf{X} = (x_1, ..., x_n)$ and the missing observations $\mathbf{\Theta} = (\theta_1, ..., \theta_n)$, to implement the algorithm we define the hypothetical complete-data distribution,

$$f(x, \theta) = g(\theta; \lambda, \sigma)\, f(x|\theta),\ \theta > 0 \text{ and } x = 0, 1, \dots \tag{19}$$

where $f(x|\theta)$ denotes the probability distribution of the Poisson distribution.

Following Karlis (2005), at the E–step of the $(j+1) - th$ iteration, the conditional expectations of some function of θ_i's are calculated. Later, the log–likelihood of the complete data model is maximized, this problem is equivalent to maximizing $g(\theta; \lambda, \sigma)$. As we have mixtures from the exponential family, these conditional expectations coincide with the sufficient statistics needed for maximum likelihood estimation of the mixing distribution. In this case θ_i and $\log \frac{\theta_i^2}{1+\theta_i}$.

The EM type algorithm for this model can be described as follows. From the current estimates $\hat{\sigma^{(j)}}$ and $\hat{\lambda^{(j)}}$.

- **E-step:** Calculate the pseudo-values t_i and s_i. As it can be seen in Karlis (2005), for linear functions of θ the conditional posterior expectations can be easily obtained, then

$$\begin{aligned} t_i &= E(\theta_i|x_i, \hat{\lambda}^{(j)}, \hat{\sigma}^{(j)}) \\ &= (x_i + 2\hat{\sigma}^{(j)} + 1)\, \frac{\mathcal{U}(2 + 2\hat{\sigma}^{(j)} + x_i, 3 + \hat{\sigma}^{(j)}, 1 + \hat{\lambda}^{(j)})}{\mathcal{U}(1 + 2\hat{\sigma}^{(j)} + x_i, 2 + \hat{\sigma}^{(j)}, 1 + \hat{\lambda}^{(j)})}. \end{aligned}$$

  To determine the pseudo-value s_i, as the exact solution is not available,

numerical integration is required,

$$\begin{aligned} s_i &= E(\log \frac{\theta_i^2}{1+\theta_i}|x_i, \hat{\lambda}^{(j)}, \hat{\sigma}^{(j)}) \\ &= \frac{\displaystyle\int_0^\infty \log\left(\frac{\theta_i^2}{1+\theta_i}\right) \frac{\theta^{2\hat{\sigma}^{(j)}+x_i}}{(1+\theta)^{\hat{\sigma}^{(j)}}} \exp(-\theta(\hat{\lambda}^{(j)}+1))\, d\theta}{\displaystyle\int_0^\infty \frac{\theta^{2\hat{\sigma}^{(j)}+x_i}}{(1+\theta)^{\hat{\sigma}^{(j)}}} \exp(-\theta(\hat{\lambda}^{(j)}+1))\, d\theta}. \end{aligned}$$

- **M-step:** Find the new estimates $\sigma^{(j+1)}$ and $\lambda^{(j+1)}$ by maximizing the complete data log-likelihood.

  Then, we have

$$\sigma^{(j+1)} = \frac{1}{2}\left(\bar{t}\,\frac{\mathcal{U}(1+2\sigma^{(j)}, 2+\sigma^{(j)}, \lambda^{(j)})}{\mathcal{U}(2+2\sigma^{(j)}, 3+\sigma^{(j)}, \lambda^{(j)})} - 1\right),$$

  being $\bar{t} = \dfrac{\Sigma_{i=1}^n t_i}{n}$;

  and

$$\lambda^{(j+1)} = \left\{\frac{\mathcal{U}(\sigma^{(j)}, -\sigma^{(j)}, \lambda^{(j)})}{\kappa_1(\sigma^{(j)}, \lambda^{(j)})}\left(\bar{s} - 2\,\Psi(1+2\sigma^{(j)})\right)\right\}^{\dfrac{1}{1+\sigma^{(j)}}},$$

  where $\bar{s} = \dfrac{\Sigma_{i=1}^n s_i}{n}$, $\kappa_1(\sigma^{(j)}, \lambda^{(j)}) = \dfrac{\partial}{\partial\sigma}\mathcal{U}(1+2\sigma, 2+\sigma, \lambda)_{|(\sigma,\lambda)=(\sigma^{(j)},\lambda^{(j)})}$ and $\Psi(\cdot)$ is the digamma function given by $\Psi(\alpha) = \Gamma'(\alpha)/\Gamma(\alpha)$.

- The convergence is assumed when the absolute differences between successive estimates are less than a given error tolerance. Otherwise move back to the E-step for another iteration.

4.2 Deriving a simulation algorithm

The Acceptance-Rejection method of simulation can be used to generate random variates from (4). We begin by simulating a value from a gamma

distribution with shape parameter $\sigma + 1$ and rate parameter λ. Note that the pdf of this gamma distribution can be written as

$$h(\theta; \sigma + 1, \lambda) = \kappa(\sigma, \lambda) \frac{\mathcal{U}(\sigma, -\sigma, \lambda)\, \Gamma(2\sigma + 1)}{\Gamma(\sigma + 1)} \theta^{\sigma} \exp(-\lambda\, \theta)$$

where σ and λ are defined as in (4) and $\theta > 0$.

Then, having chosen an alternative random variable that has a gamma probability distribution to simulate from, we define a constant c in the following way

$$c = \max_{x} \frac{g(x)}{h(x)} = \frac{\Gamma(\sigma + 1)}{\Gamma(2\,\sigma + 1)\, \mathcal{U}(\sigma, -\sigma, \lambda)}.$$

The algorithm for simulating a value from the distribution with probability mass function (7) is as follows:

1. Generate a random variate from the gamma distribution

$$\theta_i \sim Gamma(\sigma + 1, \lambda).$$

2. Generate a random variate from the standard uniform distribution. Call this value u_1.

3. If $u_1 \leq \left(\dfrac{\theta_i}{1 + \theta_i} \right)^{\sigma}$ then set the simulated value from (4) equal to θ_i. Otherwise return to step 1.

4. Generate X_i where $X_i \sim Poisson(\theta_i)$.

5 Applications

This section contains several numerical applications of the discrete distribution introduced in this paper. The first set of data has been considered to compare the performance of the different method of estimation given in this paper. This set of data appears in Rodríguez et al. (2008). It describes the spread of European corn borer larvae *Pyrausta naubilalis* in field corn. This data set is over-dispersed and positively skewed with a long and thick right tail. In Table 2 the number of corn borer and the corresponding observed frequency appears in the first two columns. Expected frequencies have been

computed by directly maximizing the log-likelihood function, they are shown in the right hand side column. Same values are achieved by using the EM algorithm described in the previous section. By using this procedure, the solution was found after 16 iterations when the relative change of the estimates between two successive iterations was smaller than 10^{-7} and taking initial starting values in the neighborhood of the moment estimates.

Table 2: Fit of number of European corn borer *Pyrausta Naubilalis*. Rodríguez et al. (2008)

Count	Observed	Fitted
0	10	7.72
1	18	21.57
2	39	34.94
3	33	43.02
4	42	44.68
5	56	41.30
6	36	35.06
7	26	27.89
8	19	21.08
9	19	15.29
10	7	10.72
11	4	7.31
12	4	4.86
13	4	3.17
14	2	2.03
15	1	1.27
16	2	0.79
17	1	0.48
18	1	0.29
Total	324	324

Summary of results for this first example is given in Table 3. Parameter estimates together with standard errors (in brackets) obtained by the procedure explained above are provided. Furthermore, in order to compare the fit to data obtained when new discrete distribution with other competitor models two measures of model selection criteria have been included in the

bottom part of Table 3: maximum of the log–likelihood and Akaike's Information Criterion (AIC). It is important to mention that the latter measure adjusts for the number of parameters in the distribution and it allows models with different numbers of parameters to be more fairly compared. Two two-parameter discrete distributions have been considered for this purpose, the negative binomial distribution (NB$(r, p),\ r > 0,\ 0 < p < 1$) and the generalized Poisson-Lindley distribution (GPL$(\alpha, \delta),\ \alpha > 0,\ \delta > 0$) in Mahmoudi and Zakerzadeh (2010). The discrete model given by (7) has been denoted by NGG. As it can be observed in Table 3, the discrete distribution presented in this manuscript outperforms both NB and GPL distributions since a marginally larger value of the maximum of the log-likelihood is achieved. Besides, as the three models considered have equal number of parameters, similar conclusions can be derived from AIC.

Table 3: Summary of the results

Estimates		NGG	
$\hat{\sigma}$		3.574	
		(0.29)	
$\hat{\lambda}$		1.002	
		(0.07)	
Selection Criteria	GPL	NB	NGG
Maximum log-likelihood	–807.430	—807.420	–807.407
AIC	1618.860	1618.840	1618.815

In the following, two additional over-dispersed sets of data have been considered to compare again the fit to data by using the discrete NGG model with NB and GPL distributions. The estimates based on maximum likelihood estimation have been calculated for these three discrete distributions and the results are shown in Tables 4 and 5. Then, by taking the maximum of the log–likelihood as criterion of comparison NGG model provides a marginal better fit to data than the other two models. In addition to this, it is also important to mention that NGG model outperforms the group of distributions provided by Karlis (2005) for the set of data in Table 5 with the exemption of the Poisson-inverse Gaussian distribution.

Furthermore, three different QQ-Plots for the latter set of data are given in Figure 4. The three discrete models described above have been considered.

As it can be seen, the three distributions provide similar results as it has been confirmed by the small differences between sample and theoretical quantiles.

Table 4: Fit of automobile claim frequency data in Germany 1960. Kokonendji and Khoudar (2004)

Count	Observed frequencies	Expected frequencies		
		GPL	NB	NGG
0	20592	20596.50	20596.80	20596.70
1	2651	2631.33	2631.03	2631.15
2	297	318.45	318.36	318.32
3	41	37.70	37.81	37.80
4	7	4.40	4.45	4.45
5	0	0.51	0.52	0.52
6	1	0.06	0.06	0.06
Parameters estimation		$\hat{\alpha} = 1.10$	$\hat{r} = 1.11$	$\hat{\lambda} = 7.71$
		$\hat{\delta} = 8.38$	$\hat{p} = 0.88$	$\hat{\sigma} = 0.06$
Maximum log-likelihood		–10223.50	–10223.40	–10223.40
Pearson's Chi-squared		3.72	3.61	3.60
Degrees of freedom		2	2	2
p-value		0.155	0.164	0.165

6 Conclusions and comments

In this paper a new generalization of the geometric distribution has been introduced. This distribution has been obtained by mixing the Poisson distribution with a member of the natural exponential family of distributions. This distribution satisfies interesting properties and its pmf can be recursively calculated. Besides, ordinary differential equations satisfied by the its probability generating function have been obtained. The issue of parameter estimation has been studied by using maximum likelihood estimation two perspectives; this has been done either by directly maximizing the log-likelihood function or by using an expectation maximization (EM) algorithm. Both methods require the use of special functions. From the numerical results analyzed in the previous section, it can be inferred that the discrete

Table 5: Number of monthly crimes the period 1982-1993 in Greece. Karlis (2005)

Count	Observed frequencies	Expected frequencies		
		GPL	NB	NGG
0	21	23.53	23.52	23.45
1	41	35.16	35.19	35.30
2	32	32.16	32.17	32.23
3	16	23.18	23.17	23.14
4	19	14.46	14.44	14.38
5	8	8.17	8.16	8.12
6	4	4.30	4.30	4.28
7	1	2.14	2.14	2.14
8	2	1.02	1.02	1.03
9	1	0.47	0.47	0.48
Parameters estimation		$\hat{\alpha} = 4.39$	$\hat{r} = 4.49$	$\hat{\lambda} = 1.78$
		$\hat{\delta} = 2.10$	$\hat{p} = 0.66$	$\hat{\sigma} = 2.23$
Maximum log-likelihood		–274.509	–274.506	–274.487
Pearson's Chi-squared		4.89	4.89	4.86
Degrees of freedom		4	4	4
p-value		0.298	0.298	0.301

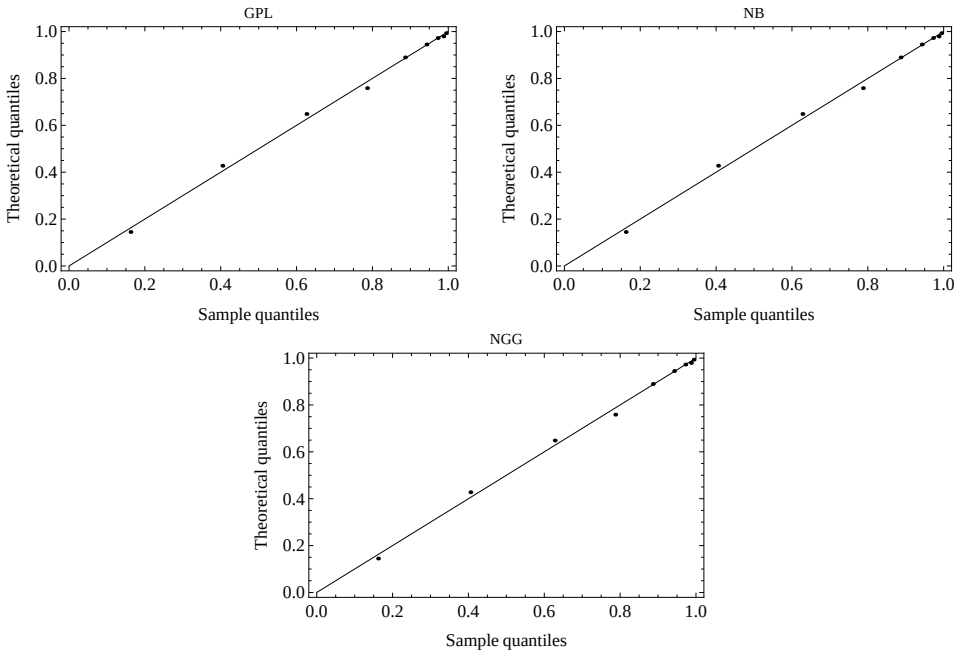

Figure 4: Quantitle–quantile plot (QQ-Plot) under the three models considered for the set of data in Karlis (2005)

distribution introduced in this manuscript shows a similar behavior than other discrete distributions traditionally used for that purpose; but having the advantage that only two parameters are required.

References

Abramowitz, M. and Stegun, I. (1972). *Handbook of Mathematical Functions*. Dover Publications, Inc., New York.

Antzoulakos, D. and Chadjiconstantinidis, S. (2004). On mixed and compound mixed Poisson distributions. *Scandinavian Actuarial Journal*, 3, 161–188.

Chao, M.T. and Strawderman, W.E. (1972). Negative moments of positive random variables. *Journal of the American Statistical Association*, 67, 338, 429–431.

Cressie, N., Davis, A.S., Leroy, J. and Policello, G.E. (1981). The moment–generating function and negative integer moments. *The American Statistician*, 35, 3, 148–150.

Feller, W. (1943). On a general class of "contagious" distributions. *Annals of Mathematical Statistics*, 16, 319–329.

Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. *Test*, 19, 399-415.

Gómez-Déniz, E. (2012). Adding a parameter to the exponential and Weibull distributions with applications. *Mathematics and Computers in Simulation* (submitted).

Greenwood, M., and Yule,G.U. (1920). An inquiry into the nature of frequency distributions of multiple happenings, with particular reference to the ocurrence of multiple attacks of disease or repeated accidents. *Journal of the Royal Statistical Society A*, 83, 255–279.

Gupta, R.C. and Ong, S.H. (2005). Analysis of long-tailed count data by Poisson mixtures. *Communications in Statistics–Theory and Methods*, 34, 557–573.

Gupta, R.D and Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, 41, 2, 173–188.

Holgate, P., 1970. The modality of some compound Poisson distribution. Biometrika, 57, 666-667.

Jain, G.C. and Consul, P.C. (1971). A Generalized Negative Binomial Distribution. *Siam Journal of Applied Mathematics*, 21, 501–513.

Kabe, D.G. (1976). Inverse moments of discrete distributions. *The Canadian Journal of Statistics*, 4, 1, 133–141.

Karlis, D. (2005). EM algorithm for mixed Poisson and other discrete distributions. *Astin Bulletin*, 35, 1, 3–24.

Karlis, D. and Xekalaki, E. (2005). Mixed Poisson distributions. *International Statistical Review*, 73, 35–58.

Kokonendji, C. and Khoudar, M. (2004). On strict arcsine distribution. *Communications in Statistics–Theory and Methods*, 33, 5, 993–1006.

Mahmoudi, E. and Zakerzadeh, H. (2010). Generalized Poisson–Lindley distribution. *Communications in Statistics–Theory and Methods*, 39, 1785–1798.

Makčutek, J. (2008). A generalization of the geometric distribution and its application in quantitative linguistics. *Romanian Reports in Physics*, 60, 3, 501–509.

Marshall, A.W. and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*, 84, 3, 641–652.

Philippou, A.N., Georghiou, C and Philippou, G.N. (1983). A generalized geometric distribution and some of its properties. *Statistics & Probability Letters*, 1, 171–175.

Rodríguez, J., Olmo, J.M., Conde, A. and Sáez, A.J. (2008). The ${}_3F_2$ with complex parameters as generating function of discrete distribution. *Communications in Statistics–Theory and Methods*, 37, 3009–3022.

Sankaran, M., (1970). The discrete Poisson-Lindley distribution. *Biometrics*, 26, 145–149.

Sundt, B. and Vernic, R. (2009). Recursions for Convolutions and Compound Distributions with Insurance Applications. Springer–Verlag, New York.

Tripathi, R.C., Gupta, R.C. and White, T.J. (1987). Some Generalizations of the Geometric Distribution *Sankhyā*, Series B, 49, 3, 218–223.

Willmot, G. (1986). Mixed compound Poisson distributions. *Astin Bulletin Supplement*, 16, 59–79.

Willmot, G. (1990). Asymptotic tail behaviuor of Poisson mixtures with applications. *Advances in Applied Probability*, 22, 1, 147–159.

Willmot, G. (1993). On recursive evaluation of mixed Poisson probabilities and related quantities. *Scandinavian Actuarial Journal*, 18, 114-133.