



UNIVERSIDAD
PABLO DE OLAVIDE
SEVILLA



REVISTA DE MÉTODOS CUANTITATIVOS PARA
LA ECONOMÍA Y LA EMPRESA (16). Páginas 121–142.
Diciembre de 2013. ISSN: 1886-516X. D.L: SE-2927-06.
URL: <http://www.upo.es/RevMetCuant/art.php?id=81>

The Compound DGL/Erlang Distribution in the Collective Risk Model

GÓMEZ DÉNIZ, EMILIO

Departamento de Métodos Cuantitativos en Economía y Gestión
Universidad de Las Palmas de Gran Canaria (Spain)
Correo electrónico: egomez@dmc.ulpgc.es

CALDERÍN OJEDA, ENRIQUE

Centre for Actuarial Studies, Department of Economics
The University of Melbourne (Australia)
Correo electrónico: enrique.calderin@unimelb.edu.au

ABSTRACT

In this paper the analysis of the collective risk model assuming Erlang loss, when the claim frequency follows the discrete generalized Lindley distribution, is considered. After providing some new results of this discrete model, analytical expressions for the aggregate claim size distribution in general insurance in the case that the discrete generalized Lindley distribution is assumed as the primary distribution while claim size, the secondary distribution, is modeled using an Erlang(r) distribution ($r = 1, 2$). Comparisons with the compound Poisson and compound negative binomial are developed to explain the viability of the new compound model in two examples in automobile insurance.

Keywords: automobile insurance; collective risk model; Lindley distribution.

JEL classification: C13; M20.

MSC2010: 97M10; 97M30.

La distribución compuesta DGL/Erlang en el modelo de riesgo colectivo

RESUMEN

En este artículo se analiza el modelo de riesgo colectivo asumiendo que la cantidad individual reclamada sigue una función de densidad Erlang y el número de reclamaciones es una variable aleatoria cuya función masa de probabilidad es la generalizada discreta Lindley. En la primera parte de este trabajo se presentan nuevas propiedades de esta distribución discreta; seguidamente, se calculan expresiones analíticas para la cantidad total reclamada en seguros generales cuando la distribución primaria es la generalizada discreta Lindley, asumiendo la densidad Erlang(r) ($r = 1, 2$) como distribución secundaria. En la ilustración numérica, el nuevo modelo expuesto en este artículo se compara con los modelos compuestos Poisson y Binomial Negativa en dos ejemplos, en el contexto de seguros de automóviles, para mostrar su efectividad.

Palabras clave: seguro de automóviles; modelo de riesgo colectivo; distribución Lindley.

Clasificación JEL: C13; M20.

MSC2010: 97M10; 97M30.



1 Introduction

One of the most significant goals of any insurance risk activity is to achieve a satisfactory model for the probability distribution of the total claim amount. The classical collective theory of risk is based on the assumption that the counting process representing the number of claims (known as primary distribution) is a Poisson process and the associated cumulative or compound process representing the total claim amount is thus compound Poisson. Nevertheless, it has been considered in many instances that the number of automobile claims process is not necessarily of Poisson type. It is well-known that collective claim frequencies are characterized by over-dispersion, i.e., the variance is greater than the mean (see Dionne and Vanasse (1989) and Meng *et al.*, (1999)). A key restriction of the Poisson distribution is that the variance equals the mean and, therefore, it seems not suitable for modeling automobile claim frequencies. Therefore, alternative assumptions need to be made concerning the primary distribution in the setting of the collective risk model.

In this paper the discrete generalized Lindley distribution proposed recently by Calderín-Ojeda and Gómez-Déniz (2013) is considered. This model can be viewed as a generalization of the geometric distribution and, thus, an alternative to the negative binomial distribution and the Poisson-inverse Gaussian distribution (see Willmot (1987)). The new distribution is unimodal with the possibility of a zero vertex or a mode greater than zero, depending on the values of the parameters of the distribution. These two features, zero vertex (high percentage of zero values in the empirical distribution (Boucher *et al.* (2007))) and over-dispersion are omnipresent in automobile insurance portfolios. Therefore, the new distribution can be considered as a useful alternative for modeling phenomena of this nature in the context of insurance.

When addressing the aggregate amount of claims for a compound class of policies, and when the new distribution acts as the primary one, a closed expression for the probability density function (pdf) of the total claim amount is obtained assuming that the secondary distribution is *Erlang*(r, γ). This family of loss distributions may arise in insurance settings when the individual claim amount is the sum of exponentially distributed claims. For example, in catastrophe insurance the aggregate claim size of a portfolio of r insured properties in a particular region where each property is prone to the same risk (e.g. storms, fires) and claim amounts follow an exponential

distribution with parameter γ , would follow an Erlang distribution with parameters r and γ . Apart from that, by choosing this family of distributions more flexibility is achieved when fitting empirical loss data. Additionally, from the numerical results obtained, it can be considered as alternative to compound Poisson and compound negative binomial models, traditionally used in actuarial literature. See Freifelder (1974), Rolski *et al.* (1999), Nadarajah and Kotz (2006a and 2006b) and Klugman *et al.* (2008) among others, for a comprehensive study of the compound models in the collective risk theory.

After reviewing some of its properties, we consider the question of parameter estimation when both, the moments and maximum likelihood method are considered. Then, the expected frequencies are calculated in two examples based on automobile claim frequencies, and the estimated values are used to plot the right-tail of the probability density function of the aggregate claim size of the compound discrete generalized Lindley distribution for different values of the parameter γ . Next, the results are compared with those obtained by using the compound Poisson and compound negative binomial models.

The rest of the paper is structured as follows. In Section 2, we give some properties of the discrete generalized Lindley distribution such as moments and inverse moments, among other characteristics. Then, parameters are estimated via moments and maximum likelihood methods. In Section 3, analytical expressions for the compound collective risk model are obtained by using the discrete generalized Lindley distribution and the *Erlang*(r, γ) (with $r = 1, 2$) distribution as primary and secondary distributions, respectively. Next in Section 4, the performance of the discrete generalized distribution is evaluated by using two examples in the context of automobile insurance claiming. Later, the obtained results are used to plot the density function of the aggregate claim size of the compound model. Finally, conclusions are given in the last Section.

2 DGL as primary distribution

The discrete generalized Lindley distribution (DGL), introduced not long ago by Calderín-Ojeda and Gómez-Déniz (2013), is obtained by discretizing the continuous generalized Lindley distribution proposed in Zakerzadeh and

Dolati (2009) whose survival function is given by

$$\bar{F}(x) = \frac{\alpha(1 + \theta x) + \theta}{\alpha + \theta} \exp(-\theta x), \quad x > 0, \quad (1)$$

for $\theta > 0$ and $\alpha \geq 0$, which generalizes the continuous Lindley distribution (Lindley (1958)) when $\alpha = 1$ and the classical exponential distribution is obviously obtained when $\alpha = 0$.

The DGL distribution has been built by discretizing the continuous survival function (1). This methodology has been used before in reliability and other fields of science and engineering. For instance, Nakagawa and Osaki (1975) derived the discrete Weibull distribution; Roy (2004) studied the discrete Rayleigh distribution; Gómez-Déniz (2010) analyzed a generalization of the geometric distribution using the Marshall and Olkin (1997) scheme; and Gómez-Déniz and Calderín-Ojeda (2011) deduced the discrete Lindley distribution. Properties of this distribution can be seen in Calderín-Ojeda and Gómez-Déniz (2013). The probability mass function (pmf) of the DGL distribution is given by

$$\Pr(X = x) = \lambda^x \frac{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - \log \lambda^{\alpha x + 1})}{\alpha - \log \lambda}, \quad x = 0, 1, 2, \dots, \quad (2)$$

for $\alpha \geq 0$, $0 < \lambda < 1$, $\bar{\lambda} = 1 - \lambda$ and where the reparameterization $\lambda = \exp(-\theta)$ has been considered. Observe that when $\alpha = 0$ equation (2) reduces to the geometric distribution and when $\alpha = 1$ we obtain the discrete Lindley distribution in Gómez-Déniz and Calderín-Ojeda (2011).

The cumulative distribution function is given by

$$\Pr(X \leq x) = 1 - \frac{\alpha - (1 + \alpha(1 + x)) \log \lambda}{\alpha - \log \lambda} \lambda^{1+x}.$$

In addition to the results provided by Calderín-Ojeda and Gómez-Déniz (2013), moments can be easily derived from the probability generating function, which is given by

$$G_X(z) = \frac{\alpha \bar{\lambda}(1 - z\lambda) - [1 - \lambda(1 + \alpha + z(1 - \alpha - \lambda))] \log \lambda}{(1 - \lambda z)^2(\alpha - \log \lambda)}, \quad |z| < 1. \quad (3)$$

The mean and second factorial moment can be obtained from (3) and are given by

$$\mathbb{E}(X) = \frac{\lambda[\alpha \bar{\lambda} - (\bar{\lambda} + \alpha) \log \lambda]}{\bar{\lambda}^2(\alpha - \log \lambda)}, \quad (4)$$

$$\mathbb{E}(X(X - 1)) = \frac{2\lambda^2[\alpha \bar{\lambda} - (\bar{\lambda} + 2\alpha) \log \lambda]}{\bar{\lambda}^3(\alpha - \log \lambda)}. \quad (5)$$

Successive computation of the factorial moments provides the general expression of the factorial moment of order $r > 0$

$$\mathbb{E}(X(X-1)\dots(X-r+1)) = \frac{r!\lambda^r[\alpha\bar{\lambda} - (\bar{\lambda} + r\alpha)\log\lambda]}{\bar{\lambda}^{r+1}(\alpha - \log\lambda)},$$

for $r = 1, 2, \dots$, which can be proven by induction on r .

Using Equation (4) together with Equation (5) we obtain, after some computations, the variance of the distribution,

$$\begin{aligned} \text{var}(X) &= \frac{\lambda}{\bar{\lambda}^4(\alpha - \log\lambda)^2} [\alpha^2\bar{\lambda}^2 - \alpha\bar{\lambda}(2 + \alpha - (2 - \alpha)\lambda)\log\lambda \\ &+ (\alpha + \bar{\lambda}^2 - \alpha^2\lambda - \alpha\lambda^2)\log^2\lambda]. \end{aligned} \quad (6)$$

Furthermore, it can be easily seen that the mean increases with α and λ since

$$\frac{\partial\mathbb{E}(X)}{\partial\alpha} = \frac{\lambda\log^2\lambda}{\bar{\lambda}^2(\alpha - \log\lambda)^2} > 0$$

and

$$\frac{\partial\mathbb{E}(X)}{\partial\lambda} = \frac{\log\lambda}{\bar{\lambda}^3(\alpha - \log\lambda)^2} [(\bar{\lambda} + \alpha(1 + \lambda))\log\lambda - \alpha(2\bar{\lambda} + \alpha(1 + \lambda))] > 0.$$

Besides, the ratio of successive probabilities is given by

$$\frac{\Pr(X = x)}{\Pr(X = x - 1)} = \lambda \left(1 - \frac{\alpha\bar{\lambda}\log\lambda}{\alpha\lambda\log\lambda + \bar{\lambda}(\alpha - (\alpha(x - 1) + 1)\log\lambda)} \right), \quad (7)$$

for $x = 1, 2, \dots$ where $\Pr(X = 0) = 1 + \frac{\lambda((1 + \alpha)\log\lambda - \alpha)}{\alpha - \log\lambda}$.

On the other hand, it is straightforward to see that as x increases, (7) decreases, and hence, the distribution is unimodal. In addition to this, from (7) it can be shown that the mode is at the origin if $M(\alpha, \lambda) < 0$, where

$$M(\alpha, \lambda) = \frac{1}{\log\lambda} - \frac{1}{\alpha} + \frac{1 + \lambda}{\bar{\lambda}}.$$

When $M(\alpha, \lambda) > 0$ the mode is at $[M(\alpha, \lambda)]$, where $[\cdot]$ denotes the integer part. If $[M(\alpha, \lambda)]$ is an integer, then there are joint modes at $M(\alpha, \lambda) - 1$ and $M(\alpha, \lambda)$. Finally, if $M(\alpha, \lambda) < 0$ then the mode is at zero.

The ratio between the variance and mean has been calculated for different values of the parameters and they are shown in Table 1. As it can be observed, the ratio is larger than one for the considered values of α and λ . Therefore, it seems that the coefficient of variation is always larger than 1.

Table 1: Ratio between variance and mean of the DGL distribution for different values of α and λ .

| $\alpha \backslash \lambda$ | 0.25 | 0.50 | 0.75 | 1.00 | 2.00 | 5.00 | 10.00 | 25.00 |
|-----------------------------|---------|---------|---------|---------|--------|--------|--------|--------|
| 0.10 | 1.1278 | 1.1300 | 1.1270 | 1.1220 | 1.0997 | 1.0579 | 1.0301 | 1.0066 |
| 0.20 | 1.2889 | 1.2922 | 1.2852 | 1.2749 | 1.2349 | 1.1713 | 1.1338 | 1.1045 |
| 0.30 | 1.4975 | 1.4998 | 1.4865 | 1.4698 | 1.4122 | 1.3324 | 1.2897 | 1.2579 |
| 0.40 | 1.7773 | 1.7743 | 1.7516 | 1.7267 | 1.6505 | 1.5575 | 1.5116 | 1.4888 |
| 0.50 | 2.1700 | 2.1540 | 2.1171 | 2.0816 | 1.9850 | 1.8806 | 1.8327 | 1.8000 |
| 0.60 | 2.7553 | 2.7139 | 2.6556 | 2.6063 | 2.4870 | 2.3722 | 2.3231 | 2.2903 |
| 0.70 | 3.7247 | 3.6260 | 3.5356 | 3.4681 | 3.3228 | 3.1986 | 3.1488 | 3.1165 |
| 0.80 | 5.6090 | 5.3982 | 5.2591 | 5.1676 | 4.9925 | 4.8594 | 4.8093 | 4.7777 |
| 0.90 | 10.9914 | 10.5434 | 10.3292 | 10.2056 | 9.9960 | 9.8544 | 9.8042 | 9.7733 |

Similarly, in Figure 1, the pmf (2) has been plotted for different values of λ and α . As it can be observed from the graphs, the distribution has longer right-tail when the parameter λ is close to 1 and α increases. Furthermore, the charts confirm that the distribution is always unimodal and the mode moves to the right when both parameters increase, showing a great versatility. On the other hand, for smaller values of α , there is a substantial effect on the probabilities and, of course, on the values of the mean, mode and variance.

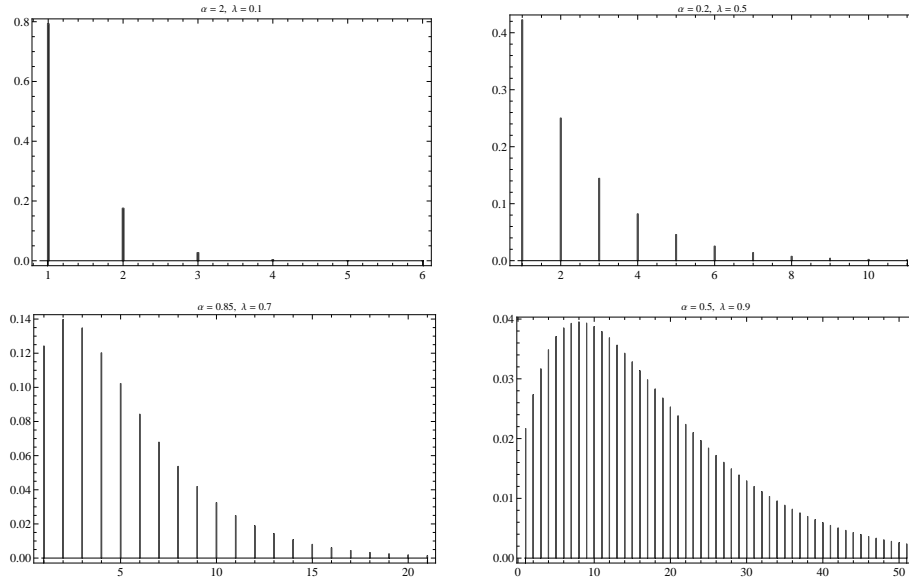


Figure 1: Probability mass function of DGL distribution for different values of α and λ

2.1 Parameter estimation

Moment estimators can be easily obtained by using equations (4) and (5). They are achieved by equating expression (4) to the sample mean \bar{x}_1 ; then, by isolating α in this equation and plugging its expression into equation (5), we obtain after some algebra that

$$\frac{2\lambda}{\bar{\lambda}^2}(2\bar{x}_1\bar{\lambda} - \lambda) = \bar{x}_2, \quad (8)$$

where \bar{x}_2 is the sample second factorial moment. Now, by solving the latter equation for λ we obtain the moment estimator of λ which is given by

$$\begin{aligned} \hat{\lambda}_1 &= \frac{2\bar{x}_1 + \bar{x}_2 + \sqrt{4\bar{x}_1^2 - 2\bar{x}_2}}{\bar{x}_2 + 4\bar{x}_1 + 2}, \\ \hat{\lambda}_2 &= \frac{2\bar{x}_1 + \bar{x}_2 - \sqrt{4\bar{x}_1^2 - 2\bar{x}_2}}{\bar{x}_2 + 4\bar{x}_1 + 2}. \end{aligned}$$

As it can be seen, it might not be unique. Finally, by replacing λ by $\hat{\lambda}$ in (8) the moment estimator for α is achieved.

Next, these moment estimators can be used as starting values to calculate maximum likelihood estimators. For that reason, let us assume that $\underline{x} =$

(x_1, x_2, \dots, x_n) is a random sample of size n from the pmf (2) with sample mean \bar{x} . The log-likelihood function is

$$\ell(\alpha, \lambda) = \sum_{i=1}^n x_i \log \lambda - n \log[\alpha - \log \lambda] + \sum_{i=1}^n \log[\alpha \lambda \log \lambda + (1 - \lambda)(\alpha - \log \lambda^{\alpha x_i + 1})]. \quad (9)$$

The normal equations are obtained by taking first derivative with respect to both parameters. They are given by

$$\begin{aligned} \frac{\partial \ell(\alpha, \lambda)}{\partial \alpha} &= -\frac{n}{\alpha - \log \lambda} \\ &+ \sum_{x=1}^n \frac{1 + \lambda(\log \lambda - 1) + (x_i(\lambda - 1)) \log \lambda}{\alpha \lambda \log \lambda + (1 - \lambda)(\alpha - (\alpha x_i + 1) \log \lambda)} = 0, \end{aligned} \quad (10)$$

$$\begin{aligned} \frac{\partial \ell(\alpha, \lambda)}{\partial \lambda} &= \frac{\bar{x}}{\lambda} + \frac{n}{\lambda(\alpha - \log \lambda)} \\ &+ \frac{1}{\lambda} \sum_{i=1}^n \frac{\lambda \log \lambda(\alpha(x_i + 1) + 1) - (1 - \lambda)(\alpha x_i + 1)}{\alpha \lambda \log \lambda + (1 - \lambda)(\alpha - (\alpha x_i + 1) \log \lambda)} = 0. \end{aligned} \quad (11)$$

The solutions of the normal equations (10) and (11) provide the maximum likelihood estimates of α and λ . These values can be easily obtained by a numerical method or direct numerical search for the global maximum of the log-likelihood surface given in (9). The second partial derivatives can be used to approximate standard errors of these estimates. They are provided in the Appendix.

3 Distribution of the total claim amount

As mentioned above, one of the primary targets of any insurance risk activity is to obtain a satisfactory model for the probability distribution of the total claim amount. The collective theory of risk is based on the assumption that the counting process representing the number of claims is a Poisson process and the associated cumulative or compound process representing the total claim amount is thus compound Poisson.

The classical collective risk model is described as follows. A distribution for the number of claims in the time interval $(0, t]$ which is denoted as $N(t)$, $t > 0$, and it is assumed that $N(0) = 0$. And the distribution of the total claim amount in the interval $(0, t]$, denoted by $S(t)$ if $t > 0$ and $S(0) = 0$.

In general, it is assumed that $S(t)$ has only step sample functions (i.e., the sample paths of $S(t)$ only change vertically at times of claims). Thus, $\{N(t), t \geq 0\}$ is a counting process, and $\{S(t), t \geq 0\}$ the associated cumulative process.

A relaxed simple form for the distribution of the total claim amount may be obtained if the individual claim size distribution is independent of time. Let us now denote by X_i the amount of the i^{th} claim for $i = 1, 2, \dots$. It is also assumed that N and X_i are independent (conditional on distribution parameters). There is an extensive body of literature on modeling the risk process. For instance, Freifelder (1974), Rolski *et al.* (1999), Nadarajah and Kotz (2006a and 2006b) and Klugman *et al.* (2008) among others. Estimation of the annual loss distribution by modeling the frequency and severity of losses is a well-known actuarial technique. It is also used for modeling solvency requirements in the insurance industry (Sandström (2006) and Wüthrich (2006)).

Then, the total claim amount S is the sum of the individual claim sizes, that is, $S = \sum_{i=1}^N X_i$ for $N > 0$ while $S = 0$ for $N = 0$. Traditionally, collective risk theory assumes Poisson and exponential distributions as primary and secondary distributions, respectively. In addition to this, it is well-known (see Klugman *et al.* (2008) and Rolski *et al.* (1999); among others) that the pdf of the sum S is given by

$$f_S(x) = \sum_{n=0}^{\infty} \Pr(N = n) f^{n*}(x), \quad (12)$$

where $\Pr(N = n)$ denotes the probability of n claims (primary distribution) and $f^{n*}(x)$ is the n^{th} fold convolution of $f(x)$, the pdf of the claim amount (secondary distribution). Recall that the convolution of two densities f and g on the positive real line is

$$(f * g)(z) = \int_0^z f(\tau)g(z - \tau) d\tau.$$

Erlang loss distributions may arise in insurance settings when the individual claim amount is the sum of exponentially distributed claims. For example, in catastrophe insurance the aggregate claim size of a portfolio of r insured properties in a particular region where each property is prone to the same risk (e.g. storms, fires) and claim amounts follows an exponential distribution with parameter γ , would follow an Erlang distribution with parameters r and γ .

Result 1 Let X_1, \dots, X_n be n independent and identically distributed random variables following an Erlang(r, γ) distribution. Then, the n^{th} fold convolution of exponential distribution has a closed-form expression and it is given by

$$f^{*n}(x) = \frac{\gamma^{rn}}{(rn-1)!} x^{rn-1} e^{-\gamma x}, \quad n = 1, 2, \dots$$

Proof. The result follows by summing n independent and identically distributed random variables with pdf $Erlang(r, \gamma)$. ■

In this regard, the compound Poisson model has been traditionally considered when the size of a single claim is modeled by an exponential distribution (e.g. $Erlang(1, \gamma)$), chiefly because of the complexity of the collective risk model under other probability distributions such as Pareto and log-normal distributions.

In the compound Poisson/exponential case (see Rolski *et al.* (1999) and Hernández-Bastida *et al.* (2009); among others) the density of the random variable aggregate claim size is given by

$$f_s(x) = \sqrt{\frac{\gamma\theta}{x}} e^{-(\theta+\gamma x)} I_1\left(2\sqrt{\theta\gamma x}\right), \quad x > 0, \quad (13)$$

while $f_s(0) = e^{-\theta}$. Here, $\theta > 0$ and $\gamma > 0$ are the parameters of the Poisson and exponential distributions, respectively and

$$I_\nu(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k+\nu}}{\Gamma(k+1)\Gamma(\nu+k+1)}, \quad z \in \mathbb{R}, \nu \in \mathbb{R},$$

represents the modified Bessel function of the first kind.

Additionally, the negative binomial distribution with parameters $r > 0$ and $0 < p < 1$ could also be assumed as primary distribution. In this case, the pdf of the random variable total claim amount (see Rolski *et al.* (1999)) is now given by the expression

$$f_s(x) = \gamma r p^r (1-p) e^{-\gamma x} {}_1F_1(1+r; 2; \gamma(1-p)x), \quad x > 0, \quad (14)$$

where ${}_1F_1(\cdot; \cdot; \cdot)$ is the confluent hypergeometric function and $f_s(0) = p^r$.

Our goal is to develop an alternative to the compound Poisson and negative binomial models by considering the DGL distribution proposed in this manuscript.

The next result shows that a closed-form expression is obtained when the DGL and exponential distributions are assumed as primary and secondary distributions, respectively.

Proposition 1 *If we assume a DGL distribution with parameters $0 < \lambda < 1$ and $\alpha \geq 0$ as primary distribution and an exponential distribution with parameter $\gamma > 0$ as secondary distribution, then the pdf of the random variable $S = \sum_{i=1}^N X_i$ is given by*

$$f_S(x) = \frac{\gamma\lambda}{\alpha - \log \lambda} [\alpha\bar{\lambda} - (\bar{\lambda} + \alpha(1 - (2 - x\gamma\bar{\lambda})\lambda)) \log \lambda] e^{-\gamma\bar{\lambda}x}, \quad x > 0, \quad (15)$$

while $f_S(0) = 1 + \frac{\lambda((1+\alpha)\log \lambda - \alpha)}{\alpha - \log \lambda}$.

Proof. By assuming that the claim amount follows an exponential distribution with parameter $\gamma > 0$, then the n^{th} fold convolution of exponential distribution is given by (see Klugman *et al.* (2008) and Rolski *et al.* (1999))

$$f^{*n}(x) = \frac{\gamma^n}{(n-1)!} x^{n-1} e^{-\gamma x}, \quad n = 1, 2, \dots$$

Then, the pdf of the random variable S is given by

$$\begin{aligned} f_S(x) &= \sum_{n=0}^{\infty} \lambda^n \frac{\alpha\lambda \log \lambda + \bar{\lambda}(\alpha - \log \lambda^{\alpha n+1})}{\alpha - \log \lambda} \frac{\gamma^n}{(n-1)!} x^{n-1} e^{-\gamma x} \\ &= \frac{e^{-\gamma x}}{\alpha - \log \lambda} \left[(\alpha\lambda \log \lambda) \sum_{n=0}^{\infty} \frac{(\lambda\gamma)^n}{(n-1)!} x^{n-1} \right. \\ &\quad \left. + \bar{\lambda} \sum_{n=0}^{\infty} (\alpha - \log \lambda^{\alpha n+1}) \frac{(\lambda\gamma)^n}{(n-1)!} x^{n-1} \right] \\ &= \frac{\gamma\lambda e^{-\gamma\bar{\lambda}x}}{\alpha - \log \lambda} [\alpha\lambda \log \lambda + \bar{\lambda}(\alpha - (1 + \alpha + x\alpha\gamma\lambda) \log \lambda)]. \end{aligned}$$

Then, after some algebra, the proposition holds. ■

Observe that the pdf of the aggregate claim amount has a jump of size $\Pr(S = 0)$ at the origin.

In the compound Poisson/*Erlang*(2, γ) the density of the random variable aggregate claim size is given by

$$f_s(x) = x\gamma^2 \lambda e^{-\lambda - \gamma x} {}_0F_2(; 3/2, 2; 1/4 x^2 \gamma^2 \lambda), \quad x > 0, \quad (16)$$

while $f_S(0) = e^{-\theta}$. Here, $\theta > 0$ and $\gamma > 0$ are the parameters of the Poisson and *Erlang*(2, γ) distributions respectively and ${}_0F_2(\cdot; \cdot; \cdot)$ the generalized hypergeometric function.

Moreover, in the compound negative binomial/*Erlang*(2, γ) the pdf of the random variable total claim amount is now given by the expression

$$f_s(x) = \gamma^2 r p^r (1-p) e^{-\gamma x} {}_1F_2(1+r; 3/2, 2; 1/4 \gamma^2 (1-p)x^2), \quad x > 0, \quad (17)$$

where ${}_1F_2(\cdot; \cdot; \cdot)$ is the generalized hypergeometric function and $f_S(0) = p^r$.

The next result shows that a closed-form expression is obtained when the DGL and $Erlang(2, \gamma)$ are assumed as primary and secondary distributions, respectively.

Proposition 2 *If we assume a DGL distribution with parameters $0 < \lambda < 1$ and $\alpha \geq 0$ as primary distribution and an $Erlang(2, \gamma)$ distribution with parameter $\gamma > 0$ as secondary distribution, then the pdf of the random variable $S = \sum_{i=1}^N X_i$ is given by*

$$f_S(x) = \frac{e^{-x\gamma}}{2(\alpha - \log \lambda)} \left[x \alpha \gamma^2 (-1 + \lambda) \lambda \cosh(x \gamma \sqrt{\lambda}) + \gamma \sqrt{\lambda} \sinh(x \gamma \sqrt{\lambda}) (-2\alpha(-1 + \lambda) + (-2 - \alpha + 2\lambda + 3\alpha\lambda) \log \lambda) \right],$$

$$x > 0, \quad (18)$$

$$\text{while } f_S(0) = 1 + \frac{\lambda((1+\alpha)\log \lambda - \alpha)}{\alpha - \log \lambda}.$$

Proof. Again, by assuming that the claim amount follows an $Erlang(2, \gamma)$ distribution with parameter $\gamma > 0$, then its n^{th} fold convolution is given by

$$f^{*n}(x) = \frac{\gamma^2 n}{(2n-1)!} x^{2n-1} e^{-\gamma x}, \quad n = 1, 2, \dots$$

Then, the pdf of the random variable S is given by

$$\begin{aligned} f_S(x) &= \sum_{n=0}^{\infty} \lambda^n \frac{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - \log \lambda^{\alpha n+1})}{\alpha - \log \lambda} \frac{\gamma^{2n}}{(2n-1)!} x^{2n-1} e^{-\gamma x} \\ &= \frac{e^{-\gamma x}}{\alpha - \log \lambda} \left[(\alpha \lambda \log \lambda) \sum_{n=0}^{\infty} \frac{(\lambda \gamma^2)^n}{(2n-1)!} x^{2n-1} \right. \\ &\quad \left. + \bar{\lambda} \sum_{n=0}^{\infty} (\alpha - \log \lambda^{\alpha n+1}) \frac{(\lambda \gamma^2)^n}{(2n-1)!} x^{2n-1} \right] \\ &= \frac{e^{-\gamma x}}{\alpha - \log \lambda} \left[(\alpha \lambda \log \lambda) \gamma \sqrt{\lambda} \sinh(x \gamma \sqrt{\lambda}) - \bar{\lambda} \right. \\ &\quad \left. \times \frac{\gamma \sqrt{\lambda}}{2} ((\alpha + 2) \log \lambda - 2\alpha) \sinh(\gamma \sqrt{\lambda} x) + \alpha \gamma \sqrt{\lambda} x \log \lambda \cosh(\gamma \sqrt{\lambda} x) \right] \end{aligned}$$

Then, after some algebra, the proposition holds. ■

Again, note that the pdf of the aggregate claim amount has a jump of size $\Pr(S = 0)$ at the origin.

4 Numerical applications

In this section, two set of real data based on a portfolio of automobile insurance claims are considered. These data were taken from a sample of 298 automobile liability policies (Klugman *et al.*, 1998, pp. 244) for the first example and, from a sample of 7842 automobile liability policies that appeared in Seal (1982) for the second case. The number of claims for each policy and the corresponding observed frequency are given in the first two columns (from left to right) of Table 2 and Table 3, respectively. These data are over-dispersed and right skewed; besides, the proportion of zeros in the sample is about one third and two thirds of the total number of claims for the first and second example respectively. Additionally, the first set shows a long and thick right-tail. Expected frequencies are also provided for the Poisson ($\mathcal{P}o$) distribution and the negative binomial (NB) distribution obtained after estimating parameters by maximum likelihood. As it can be inferred from the results, the DGL distribution provides a slightly better fit to data than NB distribution as judged by maximum of the log-likelihood (ℓ_{max}) function (-528.619 as opposed to -528.769) for the first data set and similar fit for the second one. As expected, the results obtained for the Poisson distribution are worse than the previous ones since this model is unable to capture the over-dispersion phenomenon.

Table 2: Fit to data for automobile insurance claims. Klugman *et al.* (2008)

| Claim number | Observed | Fitted | | |
|-----------------|----------|-----------------|-------|-------|
| | | \mathcal{P}_o | NB | DGL |
| 0 | 99 | 54.00 | 95.85 | 96.65 |
| 1 | 65 | 92.24 | 75.83 | 74.29 |
| 2 | 57 | 78.77 | 50.35 | 50.23 |
| 3 | 35 | 44.85 | 31.29 | 31.71 |
| 4 | 20 | 19.15 | 18.79 | 19.17 |
| 5 | 10 | 6.54 | 11.04 | 11.26 |
| 6 | 4 | 1.86 | 6.39 | 6.47 |
| 7 | 0 | 0.45 | 3.66 | 3.66 |
| 8 | 3 | 0.09 | 2.08 | 2.04 |
| 9 | 4 | 0.01 | 1.17 | 1.13 |
| 10 | 0 | 0.00 | 0.66 | 0.61 |
| 11 | 1 | 0.00 | 0.37 | 0.33 |
| 12 | 0 | 0.00 | 0.20 | 0.18 |
| Total | 298 | 298 | 298 | 298 |

Table 3: Fit to data for automobile insurance claims. Seal (1982)

| Claim number | Observed | Fitted | | |
|-----------------|----------|-----------------|---------|---------|
| | | \mathcal{P}_o | NB | DGL |
| 0 | 5147 | 4783.18 | 5147.94 | 5149.94 |
| 1 | 1859 | 2364.75 | 1859.84 | 1852.33 |
| 2 | 595 | 584.55 | 586.33 | 591.902 |
| 3 | 167 | 96.33 | 175.85 | 177.28 |
| 4 | 54 | 11.90 | 51.39 | 50.96 |
| 5 | 14 | 1.17 | 14.78 | 14.24 |
| 6 | 5 | 0.09 | 4.20 | 3.90 |
| 7 | 0 | 0.00 | 1.18 | 1.05 |
| 8 | 0 | 0.00 | 0.33 | 0.28 |
| 9 | 0 | 0.00 | 0.09 | 0.07 |
| 10 | 0 | 0.00 | 0.02 | 0.02 |
| 11 | 1 | 0.00 | 0.00 | 0.00 |
| Total | 7842 | 7842 | 7842 | 7842 |

A summary of results for each example is exhibited in Table 4. The standard errors of the parameters have been approximated from Fisher's information matrix.

Table 4: Summary of results

| | Table 1 | Table 2 |
|---|----------------------------------|----------------------------------|
| Estimators | | |
| $\hat{\theta}$ | 1.708 (0.075) | 0.494 (0.007) |
| (\hat{r}, \hat{p}) | (1.473, 0.463) (0.259, 0.046) | (1.341, 0.730) (0.098, 0.015) |
| $(\hat{\alpha}, \hat{\lambda})$ | (0.501, 0.695) (0.396, 0.031) | (0.239, 0.621) (0.148, 0.011) |
| $(\chi^2, \text{d.f.}, p - \text{value})$ | | |
| $\mathcal{P}o$ | (151.73, 5, 0.000) | (809.59, 5, 0.000) |
| NB | (3.14, 4, 0.533) | (0.75, 4, 0.944) |
| DGL | (2.64, 4, 0.618) | (0.90, 4, 0.923) |
| ℓ_{\max} | | |
| $\mathcal{P}o$ | -577.002 | -7608.77 |
| NB | -528.769 | -7429.60 |
| DGL | -528.619 | -7429.85 |

The estimates displayed in the latter Table have been used to plot the right tail of the probability density function of equations (16), (17) and (18) to calculate the right-tail cumulative probabilities for different values of γ . The graphs are displayed in Figure 2 and 3 for the first and second data set respectively. From the examples considered and the values of the parameter γ chosen, the compound DGL model provides higher probabilities in the right tail of the distribution than the compound NB and compound Poisson models

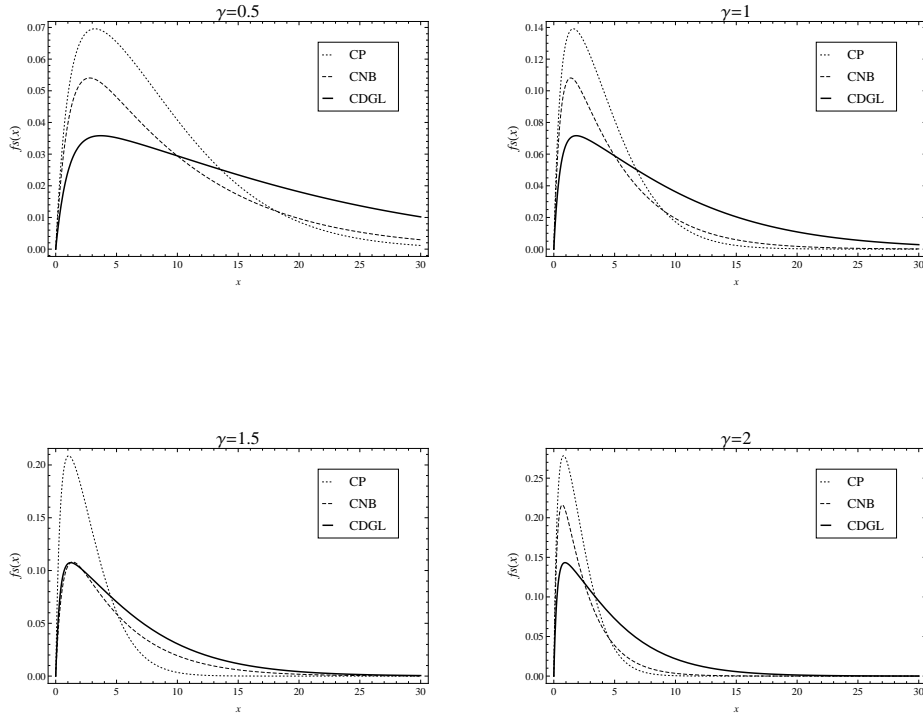


Figure 2: Right tail of the pdf of the aggregate claim size for different values of γ for Compound Poisson (CP), Compound Negative Binomial (CNB) and Compound DGL (CDGL) models. Dataset 1.

The compound model introduced in this paper seems suitable for modeling extreme data when the severity component follows an $Erlang(2, \gamma)$ distribution. This result will surely be improved if the same analysis is conducted with more heavy-tailed distributions to model the severity component. Likewise, distribution (15) seems a feasible model to be used in the framework of collective risk theory.

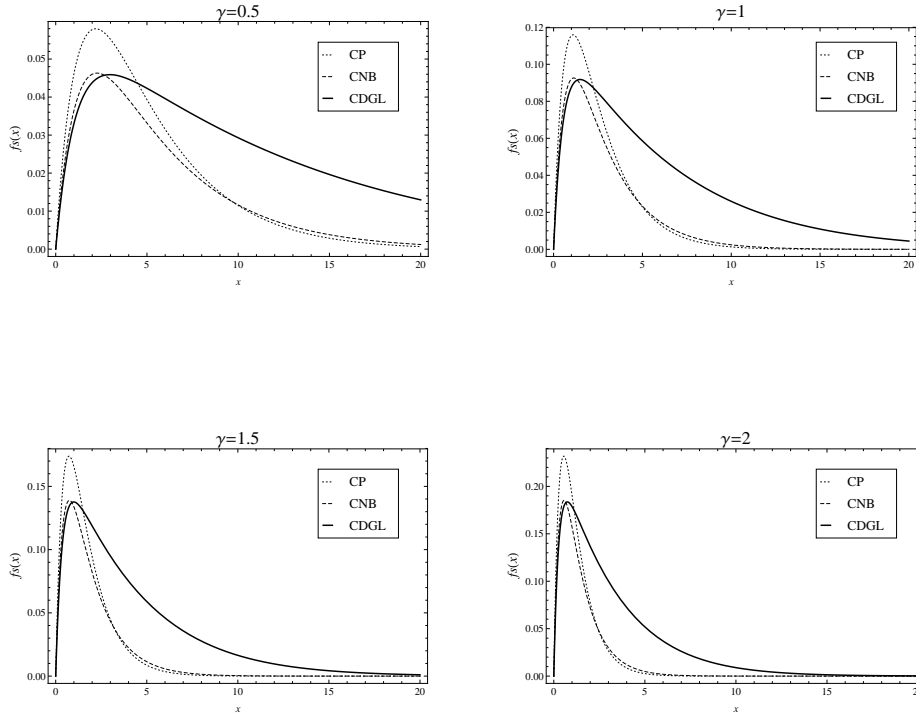


Figure 3: Right tail of the pdf of the aggregate claim size for different values of γ for Compound Poisson (CP), Compound Negative Binomial (CNB) and Compound DGL (CDGL) models. Dataset 2.

5 Conclusions

In this paper we have shown the applicability of the discrete generalized Lindley (DGL) distribution as a feasible model to describe the number of claims in automobile insurance framework. In addition to this, this discrete model seems appropriate to be used as a primary distribution in collective risk theory when the $Erlang(2, \gamma)$ distribution acts as secondary distribution. The compound model obtained holds no additional complications since its closed-form expression can be deduced analytically. From the numerical results obtained, it can be considered as alternative to the compound Poisson and compound negative binomial models.

Acknowledgements

EGD was partially funded by ECO2009-14152 (MICINN, Spain).

References

- Boucher, J.P., Denuit, M. and Guillén, M. (2007). Risk classification for claim counts: a comparative analysis of various zero-inflated mixed Poisson and hurdle models. *North American Actuarial Journal*, 11, 4, 110–131.
- Calderín-Ojeda, E. and Gómez-Déniz, E. (2013). An Extension of the Discrete Lindley Distribution with Applications. *Journal of the Korean Statistical Society*, 42, 3, 371–373.
- Dionne, G. and Vanasse, Ch. (1989). A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. *Astin Bulletin*, 19, 2, 199–212.
- Freifelder, F. (1974). Statistical decision theory and procedures, in *Credibility theory and applications*, P.M. Kahn, ed., Academic Press, New-York, 71–88.
- Gómez-Déniz, E. (2010). Another generalization of the geometric distribution. *Test*, 19, 399-415.
- Gómez-Déniz, E. and Calderín-Ojeda, E. (2011). The discrete Lindley distribution: properties and applications. *Journal of Statistical Computation and Simulation*, 81, 11, 1405–1416.
- Hernández-Bastida, A., Gómez-Déniz, E. and Pérez-Sánchez, J.M. (2009). Bayesian Robustness of the Compound Poisson Distribution under Bidimensional Prior: an Application to the Collective Risk Model. *Journal of Applied Statistics*, 36, 8, 853–869.
- Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2008). *Loss Models: From Data to Decisions*. Third Edition. Wiley.
- Lindley, D.V., (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society Series B*, 20, 3, 102–107.

- 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.
- Meng, S., Wei, Y. and Whitmore, G. A. (1999). Accounting for individual over-dispersion in a bonus-malus system. *Astin Bulletin*, 29, 2, 327–337.
- Nadarajah, S. and Kotz, S. (2006a). Compound mixed Poisson distributions I. *Scandinavian Actuarial Journal*, 3, 141–162.
- Nadarajah, S. and Kotz, S. (2006b). Compound mixed Poisson distributions II. *Scandinavian Actuarial Journal*, 3, 163–181.
- Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution. *IEEE Trans. Reliab.*, 24, 5, 300–301.
- Rolski, T., Schmidli, H. Schmidt, V. and Teugel, J. (1999). *Stochastic processes for insurance and finance*. John Wiley & Sons.
- Roy D. (2004). Discrete Rayleigh distribution. *IEEE Trans. Reliab.*, 53, 5, 255–260.
- Sandström, A. (2006). *Solvency: Models, Assessment and Regulation*, Chapman & Hall/CRC, Boca Raton.
- Seal, H.L. (1982). Mixed Poisson-an ideal distribution of claim number? *MVSV*, 293–295.
- Willmot, G.E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scandinavian Actuarial Journal*, 113-127.
- Wüthrich, M.V. (2006). Premium liability risks: Modelling small claims. *Bull. Swiss Association Actuaries*, 1, 27–38.
- Zakerzadeh, Y. and Dolati, A. (2009). Generalized Lindley Distribution. *Journal of Mathematical Extension*, 3, 2, 13–25.

Appendix

Here we show the second partial derivatives of the log-likelihood function. They are as follows.

$$\begin{aligned}
\frac{\partial^2 \ell(\alpha, \lambda)}{\partial \alpha^2} &= \frac{n}{(\alpha - \log \lambda)^2} - \sum_{x=1}^n \left[\frac{\bar{\lambda} + (\lambda - \bar{\lambda}x_i) \log \lambda}{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - (\alpha x_i + 1) \log \lambda)} \right]^2, \\
\frac{\partial^2 \ell(\alpha, \lambda)}{\partial \lambda^2} &= -\frac{\bar{x}}{\lambda^2} - \frac{n(\alpha - \log \lambda - 1)}{\lambda^2(\alpha - \log \lambda)^2} \\
&\quad - \frac{1}{\lambda^2} \sum_{i=1}^n \frac{\lambda \log \lambda (\alpha(x_i + 1) + 1) - \bar{\lambda}(\alpha x_i + 1)}{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - (\alpha x_i + 1) \log \lambda)} \\
&\quad + \frac{1}{\bar{\lambda}} \sum_{i=1}^n \frac{(\alpha(x_i + 1) + 1)(1 + \log \lambda) + \alpha x_i + 1}{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - (\alpha x_i + 1) \log \lambda)} \\
&\quad - \frac{1}{\lambda^2} \sum_{i=1}^n \frac{\lambda \log \lambda (\alpha(x_i + 1) + 1) - \bar{\lambda}(\alpha x_i + 1)}{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - (\alpha x_i + 1) \log \lambda)} \\
&\quad \times \frac{\alpha(\log \lambda + 1) - \alpha + (\alpha x_i + 1) \log \lambda - \bar{\lambda}(\alpha x_i + 1)}{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - (\alpha x_i + 1) \log \lambda)}, \\
\frac{\partial^2 \ell(\alpha, \lambda)}{\partial \lambda \partial \alpha} &= -\frac{n}{\lambda(\alpha - \log \lambda)^2} - \frac{1}{\lambda} \sum_{i=1}^n \frac{\lambda \log \lambda - \bar{\lambda}x_i + \lambda x_i \log \lambda}{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - (\alpha x_i + 1) \log \lambda)} \\
&\quad + \frac{1}{\bar{\lambda}} \sum_{i=1}^n \frac{\bar{\lambda} + (\lambda - \bar{\lambda}x_i) \log \lambda}{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - (\alpha x_i + 1) \log \lambda)} \\
&\quad \times \frac{(1 + \alpha(x_i + 1)) \log \lambda - \bar{\lambda}(1 + \alpha x_i)}{\alpha \lambda \log \lambda + \bar{\lambda}(\alpha - (\alpha x_i + 1) \log \lambda)},
\end{aligned}$$

from these expressions Fisher's information matrix can be approximated in the conventional way.