Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online): 2229 5666 Vol. 11, Issue 1 (2018): 117-132

DISCRETE XGAMMA DISTRIBUTIONS: PROPERTIES, ESTIMATION AND AN APPLICATION TO THE COLLECTIVE RISK MODEL

Sudhansu S. Maiti[1*], Mithu Dey[2] and Seema Sarkar (Mondal)[3]
[1]Department of Statistics, Visva-Bharati University, Santiniketan, India
[2]Department of Basic Sciences (Mathematics), Asansol Engineering College, Vivekananda Sarani, Kanyapur, India
[3]Department of Mathematics, National Institute of Technology Durgapur, Burdwan, India
E Mail: [1]dssm1@rediffmail.com[*]; [2]mithu.dey07@gmail.com; [3]seemasarkarmondal@gmail.com
*Corresponding Author

Received December 25, 2017
Modified March 20, 2018
Accepted April 30, 2018

Abstract

In this paper, discrete versions of xgamma distribution [c.f. Sen et al., 2016] have been studied. Two discrete versions, namely discrete xgamma-I and discrete xgamma-II and their structural and reliability properties have been studied. Estimation procedures of the parameter of these discrete distributions have been mentioned. Compound discrete xgamma distributions in the context of collective risk model have been obtained in closed form. The new compound distributions have been compared with the classical compound Poisson, compound Negative binomial and compound discrete Lindley distributions regarding suitability of modelling extreme data with the help of some automobile claim.

Key Words: Discrete Analogue Approach, Discrete Concentration Approach, Collective Risk Model, Heavy-Tailed Distribution, Reinsurance Premium.

1. Introduction

There is need for finding out discrete distributions to cater the need of fitting real life data as most of the times we come across situations where measurements are discrete in nature. Aiming at this a number of approaches has been adopted to find out discrete version of some standard continuous probability distributions. A recent review of approaches has been discussed in Chakraborty (2015). Out of these approaches we concentrate here on two, viz., discrete concentration approach and discrete analogue approach for finding the discrete versions of the xgamma distribution proposed by Sen et al. (2016).

a) Discrete Concentration Approach: A continuous life time variable is used to generate a discrete model by introducing a grouping on the time axis. Let us consider the underlying continuous failure time X has the survival function S(x) = P[X>x] and times are grouped into unit intervals such that observed discrete variable is dX=[X], the largest integer contained in X, the probability mass function (pmf) of dX can be written as,

$$p(x)=P[dX=x]=P[x\le dX<x+1]=S(x)-S(x+1);\quad x=0,1,2,3 \qquad (1)$$

In fact, the probability mass function (pmf) of random variable dX is the discrete concentration of the probability density of X. Exponential distribution, when discretized in this approach, takes the form of geometric distribution. Following this approach, Weibull, half-normal, Rayleigh, Burr and Lindley distributions have been discretized by Nakagawa and Osaki (1975), Kemp (2008), Roy (2004), Krishna and Pundir(2009), and Gomez-Deniz and Calderin-Ojeda (2011), respectively.

b) Discrete Analogue Approach: In this approach , if the probability density function (pdf) of continuous random variable X is f(x), then the pmf of discrete random variable Y may be written as

$$p(k)=P[Y=k]=\frac{f(k)}{\sum_{j=-\infty}^{\infty} f(j)};k=0,\pm1,\pm2,\pm3,..... \qquad (2)$$

Following this approach, Good, normal, log-normal, exponential and gamma, Laplace(double exponential), and Skew Laplace distributions have been discretized by Good (1953), Kemp (1997), Bi et al. (2001), Sato et al. (1999), Inusah and Kozubowski (2006), and Kozubowski and Inusah (2006), respectively.

The xgamma distribution was introduced by Sen et al. (2016) and is given by the pdf

$$f(x)=\frac{\theta^2}{1+\theta}(1+\frac{\theta}{2}x^2)e^{-\theta x};x>0,\theta>0. \qquad (3)$$

The corresponding cumulative distribution function (cdf) is:

$$F(x)=1-\frac{1+\theta+\theta x+\frac{\theta^2x^2}{2}}{1+\theta}e^{-\theta x};x>0,\theta>0. \qquad (4)$$

It has been shown that (3) is more flexible than Lindley distribution proposed by Lindley (1958) and is given by the pdf

$$f(x)=\frac{\theta^2}{1+\theta}(1+x)e^{-\theta x};x>0,\theta>0. \qquad (5)$$

The exponential distribution is widely discussed and is given by the pdf

$$f(x)=\theta e^{-\theta x};x>0,\theta>0 \qquad (6)$$

and is closed in form to (5). Ghitaney et al. (2008) have shown that from flexibility and application point of view, the Lindley distribution is better than the exponential distribution based model.

The present article is organized as follows. In section 2, the discrete xgamma-I distribution has been derived using discrete concentration approach and structural and reliability properties have been studied. Section 3 is concentrated on deriving the discrete xgamma-II distribution using discrete analogue approach and on studying its properties. The two discrete distributions have been compared with Poisson, Negative binomial and discrete Lindley distributions in connection to the collective risk in section 4. Estimation of model parameter by method of moment and method of maximum likelihood has been mentioned in section 5. Some data sets have been analyzed to compare the distributions in term of negative log-likelihood in section 6. Section 7 concludes.

2. Discrete xgamma-I Distribution (dxgamma-I)

Using discrete concentration approach discussed in (a) of section 1, we have the pmf as

$$p_x^I = P(X=x) = \frac{1}{1+\theta}[1+\theta-e^{-\theta}(1+2\theta+\frac{\theta^2}{2})+\theta(1-e^{-\theta}(1+\theta))x+\frac{\theta^2}{2}(1-e^{-\theta})x^2]e^{-\theta x}; \theta>0, x=0,1,2,...$$

$$=\frac{1}{1-\ln p}[1-\ln p-p(1-2\ln p+\frac{(\ln p)^2}{2})-\ln p(1-p(1-\ln p))x+\frac{(\ln p)^2}{2}(1-p)x^2]p^x; 0<p<1, x=0,1,2...$$

Here we find that $p_0^I = \dfrac{1-\ln p - p(1-2\ln p+\frac{(\ln p)^2}{2})}{1-\ln p}$, and the other probabilities are to be calculated using the recursive relationship

$$p_{x+1}^I = [\frac{1-\ln p-p(1-2\ln p+\frac{(\ln p)^2}{2})-(x+1)\ln p\{1-p(1-\ln p)\}+(x+1)^2\frac{(1-p)(\ln p)^2}{2}}{1-\ln p-p(1-2\ln p+\frac{(\ln p)^2}{2}-x\ln p\{1-p(1-\ln p)\}+x^2\frac{(1-p)(\ln p)^2}{2})}]p_x^I$$

This pmf can be re-written with the linear combination of negative binomial distributions as

$$p_x^I = a_1 G(x;1-p)+b_1 NB(x;2,1-p)+c_1 NB(x;3,1-p); 0<p<1, x=0,1,2, ...$$

with $a_1 = \dfrac{1-p+p\ln p+\frac{(\ln p)^2}{2}}{(1-p)(1-\ln p)}$, $b_1 = \dfrac{(1-p)\ln p+(3-p)\frac{(\ln p)^2}{2}}{(1-p)^2(1-\ln p)}$,

and $c_1 = \dfrac{(1-p)(\ln p)^2}{(1-p)^2(1-\ln p)}$.

Here G(x;1-p), NB(x;2,1-p) and NB(x;3,1-p) are the pmfs of geometric, negative binomial with success 2 and that with success 3 having success probability 1-p. Constants a_1, b_1 and c_1 are such that $a_1+b_1+c_1=1$, but are not necessarily proportions. We call this distribution as the Discrete xgamma-I Distribution (xgamma-I).
The cdf of this distribution is given by

$$F(x)=P(X\le x)=1-S(x+1)=1-\frac{1-\ln p-\ln p(x+1)+\frac{(\ln p)^2}{2}(x+1)^2}{1-\ln p}p^{x+1};\quad 0<p<1,\ x=0,1,2...$$

Probability Generating function (pgf),

$$P_I(s)=a_1\frac{1-p}{1-ps}+b_1\frac{(1-p)^2}{(1-ps)^2}+c_1\frac{(1-p)^3}{(1-ps)^3}, |s|<\frac{1}{p}$$

Moment generating function (mgf),

$$M_1(t) = a_1 \frac{1-p}{1-pe^t} + b_1 \frac{(1-p)^2}{(1-pe^t)^2} + c_1 \frac{(1-p)^3}{(1-pe^t)^3}, t < -\ln p$$

Characteristic function (cf),

$$\phi_1(t) = a_1 \frac{1-p}{1-pe^{it}} + b_1 \frac{(1-p)^2}{(1-pe^{it})^2} + c_1 \frac{(1-p)^3}{(1-pe^{it})^3}, |it| < -\ln p$$

First four moments are

$$\mu_1' = [a_1 + 2b_1 + 3c_1]\frac{p}{1-p}$$

$$\mu_2' = [(1+p)a_1 + 2(1+2p)b_1 + 3(1+3p)c_1]\frac{p}{(1-p)^2}$$

$$\mu_3' = [(1+4p+p^2)a_1 + 2(1+7p+4p^2)b_1 + 3(1+10p+9p^2)c_1]\frac{p}{(1-p)^3}$$

$$\mu_4' = [(1+11p+11p^2+p^3)a_1 + 2(1+18p+33p^2+8p^3)b_1 + 3(1+25p+67p^2+27p^3)c_1]\frac{p}{(1-p)^4}$$

Here, survival function,

$$S_I(x) = P(X \geq x) = \frac{1-\ln p - \ln px + \frac{(\ln p)^2}{2}x^2}{1-\ln p} p^x; 0<p<1, x=0,1,2,...$$

Hazard function,

$$h_I(j) = \frac{1-\ln p - p(1-2\ln p + \frac{(\ln p)^2}{2}) - \ln p(1-p(1-\ln p))j + \frac{(\ln p)^2}{2}(1-p)j^2}{1-\ln p - \ln pj + \frac{(\ln p)^2}{2}j^2}$$

Note that $1 - \frac{[1-2\ln p + \frac{(\ln p)^2}{2}]p}{1-\ln p} < h_I(j) < (1-p) \quad for \ \forall j.$

Mean remaining life,

$$m_I(t) = \frac{\sum_{x=t+1}^{\infty} S_I(x)}{S_I(t)}$$

$$= p.\frac{\frac{1-\ln p}{1-p} - \ln p\{\frac{1+t}{1-p} + \frac{p}{(1-p)^2}\} + \frac{(\ln p)^2}{2}\{\frac{(1+t)^2}{1-p} + \frac{2(1+t)p}{(1-p)^2} + \frac{p}{(1-p)^2} + \frac{2p^2}{(1-p)^3}\}}{1-\ln p - t\ln p + t^2\frac{(\ln p)^2}{2}}$$

It is to be noted that $\frac{1}{1-p} < m_I(t) < \frac{p}{1-p}[1 - \frac{\ln p(1-\frac{1+p}{2}\ln p)}{(1-p)(1-\ln p)}] \ for \ \forall t$

In risk management theory, the risk measures, for example, the Value-at-risk (VaR), the Tail-at-risk (TVaR), play important roles. The VaR at u , 0<u<1 is defined as x_u=VaR[X;u]=inf{x €R:F(x)≥u} which is the uth quantile of the random variable X and the TVaR is defined as $TVaR[X;u] = E(X \mid X \geq x_u) = x_u + \frac{\int_{z_u}^{\infty} xp(x)dx}{1-u}$.

For this distribution,

$$TVaR[X;u] = x_u + p\frac{\frac{1-\ln p}{1-p} - \ln p\{\frac{1+z_u}{1-p} + \frac{p}{(1-p)^2}\} + \frac{(\ln p)^2}{2}\{\frac{(1+z_u)^2}{1-p} + \frac{2(1+z_u)p}{(1-p)^2} + \frac{p}{(1-p)^2} + \frac{2p^2}{(1-p)^2}\}}{1-\ln p - x_u \ln p + x_u^2 \frac{(\ln p)^2}{2}}.$$

3. Discrete xgamma-II Distribution (dxgamma-II)

Using discrete analogue approach in (b) of section 1, we have pmf as

$$p_x^{II} = P(X = x) = \frac{2(1-e^{-\theta})^3(1+\frac{\theta x^2}{2})}{2(1-e^{-\theta})^2 + \theta e^{-\theta}(1+e^{-\theta})} e^{-\theta x}; \theta > 0, x = 0,1,2,...$$

$$= \frac{2(1-p)^3(1-\frac{x^2 \ln p}{2})}{2(1-p)^2 - p(1+p)\ln p} p^x; 0 < p < 1, x = 0,1,2,...$$

Here we find that $p_0^{II} = \frac{2(1-p)^2}{2(1-p)^2 - p(1+p)\ln p}$, and the other probabilities are to be calculated using the recursive relationship $p_{x+1}^{II} = p\frac{2-(x+1)^2 \ln p}{2-x^2 \ln p} p_x^{II}$.

This pmf can also be re-written as a linear function of negative binomial distributions as

$$p_x^{II} = a_2 G(x;1-p) + b_2 NB(x;2,1-p) + c_2 NB(x;3,1-p); 0 < p < 1, \quad x = 0,1,2,...$$

with $a_2 = \frac{2(1-p)^2(1-\frac{\ln p}{2})}{2(1-p)^2 - p(1+p)\ln p}, b_2 = \frac{3(1-p)\ln p}{2(1-p)^2 - p(1+p)\ln p}$

and $c_2 = -\frac{2\ln p}{2(1-p)^2 - p(1+p)\ln p}$

Here the constants are a2, b2and c2 with $a_2+b_2+c_2=1$ and we call this distribution as Discrete xgamma-II (dxgamma-II) Distribution. The cdf of this distribution is given by

$$F_{II}(x)=P(X\le x)=\frac{2(1-p)^3}{2(1-p)^2-p(1+p)\ln p}[\frac{1-p^{x+1}}{1-p}-\frac{\ln p}{2}\{\frac{2p^2+2(x^2-1)p^{x+2}-x(x+1)p^{x+1}-x(x-1)p^{x+3}}{(1-p)^3}$$
$$+\frac{p-(x+1)p^{x+1}+xp^{x+2}}{(1-p)^2}\}]; \qquad x=0,1,2....$$

Probability Generating function (pgf),

$$P_{II}(s)=a_2\frac{1-p}{1-ps}+b_2\frac{(1-p)^2}{(1-ps)^2}+c_2\frac{(1-p)^3}{(1-ps)^3},|s|<\frac{1}{p}$$

Moment Generating function (mgf),

$$M_{II}(t)=a_2\frac{1-p}{1-pe^t}+b_2\frac{(1-p)^2}{(1-pe^t)^2}+c_2\frac{(1-p)^3}{(1-pe^t)^3},t<-\ln p$$

Characteristic function (cf),

$$\phi_{II}(t)=a_2\frac{1-p}{1-pe^{it}}+b_2\frac{(1-p)^2}{(1-pe^{it})^2}+c_2\frac{(1-p)^3}{(1-pe^{it})^3},|it|<-\ln p$$

First four raw moments are

$$\mu_1^{'}=[a_2+2b_2+3c_2]\frac{p}{1-p}$$

$$\mu_2^{'}=[(1+p)a_2+2(1+2p)b_2+3(1+3p)c_2]\frac{p}{(1-p)^2}$$

$$\mu_3^{'}=[(1+4p+p^2)a_2+2(1+7p+4p^2)b_2+3(1+10p+9p^2)c_2]\frac{p}{(1-p)^3}$$

$$\mu_4^{'}=[(1+11p+11p^2+p^3)a_2+2(1+18p+33p^2+8p^3)b_2+3(1+25p+67p^2+27p^3)c_2]\frac{p}{(1-p)^4}$$

Survival function,

$$S_{II}(k)=\frac{2(1-p)^3}{2(1-p)^2-p(1+p)\ln p}[\frac{1}{1-p}-\frac{\ln p}{2}\{\frac{k^2}{1-p}+\frac{(2k+1)p}{(1-p)^2}+\frac{2p^2}{(1-p)^3}\}]p^k.$$

Hazard function, $h_{II}(j)=\dfrac{1-\dfrac{j^2\ln p}{2}}{\dfrac{1}{1-p}-\dfrac{\ln p}{2}\{\dfrac{j^2}{1-p}+\dfrac{(2j+1)p}{(1-p)^2}+\dfrac{2p^2}{(1-p)^3}\}}.$

Note that $\dfrac{2(1-p)^3}{2(1-p)^2-p(1+p)\ln p}<h_{II}(j)<(1-p)\, for \forall j$

Mean remaining life,

$$m_{II}(t) = E[X - t \mid X \geq t] = \frac{p}{1-p} \frac{2(1-p)^2 - \ln p\{5p^2 + p + 4p(1-p)(t+1) + (1-p)^2(t+1)^2\}}{2(1-p)^2 - \ln p\{2p^2 + p(1-p)(2t+1) + (1-p)^2 t^2\}}.$$

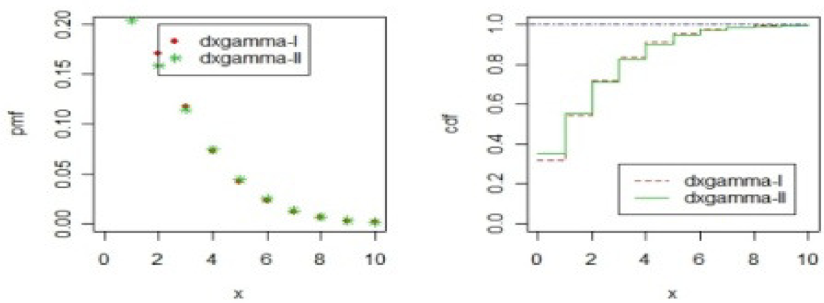
Also note that $\frac{p}{1-p} < m_{II}(t) <$

$$\frac{p}{1-p} \frac{2(1-p)^2 - \ln p\{5p^2 + p + 4p(1-p)(t+1) + (1-p)^2(t+1)^2\}}{2(1-p)^2 - \ln p\{2p^2 + p(1-p)(2t+1) + (1-p)^2 t^2\}} \text{ for } \forall \text{ t.}$$

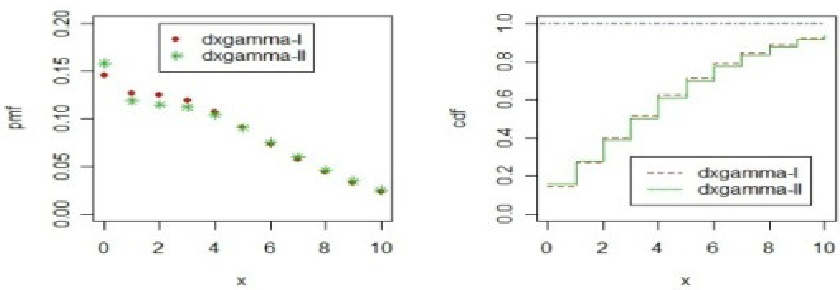
For this distribution,

$$TVaR[X; u] = x_u + \frac{p}{1-p} \frac{2(1-p)^2 - \ln p\{5p^2 + p + 4p(1-p)(x_u+1) + (1-p)^2(x_u+1)^2\}}{2(1-p)^2 - \ln p\{2p^2 + p(1-p)(2x_u+1) + (1-p)^2 x_u^2\}}$$

The pmf and cdf of dxgamma-I and dxgamma-II distributions for different values of p have been shown in figure1.



(a) pmf and cdf for p=0.4



(b) pmf and pdf for p=0.6

Fig.1. Probability mass function (pmf) and Cumulative density function (cdf) of dxgamma-I and dxgamma-II.

4. Comparing with the Poisson, Negative binomial and Discrete Lindley distributions in the collective risk model

The collective risk theory mainly deals with the aggregate claim $S = \sum_{i=1}^{N} X_{i,}$ where N denotes the claim numbers and X_i (i=1,2,…..) denotes the ith claim amount or size. When the random variables X_1,X_2,…….are independent and identically distributed and are independent of the number of claims N, then the pdf of S is $f_S(x) = \sum_{n=0}^{\infty} p_n f^{n^*}(x)$ where the probability of n claims (primary distribution) is p_n and $f^{n^*}(x)$ is the nth fold convolution of f(x) which is the pdf of the claim size (secondary distribution) . The mgf of S is $M_S(t) = M_N(\ln M_X(t))$. Hence, the mean and variance of the random aggregate claim amount are E(S)=E(N).E(X) and $Var(S) = Var(X)E(N) + E^2(X)Var(N)$. For details , see, Bowers et al. (1997).

In the context of reinsurance, large claim amounts play an important role. In case of reinsurance premiums, insurer is always interested to use a long and heavy-tailed distribution. From this consideration, in reinsurance premium calculation the pareto and log normal distributions are generally used.

The compound Poisson (sometimes negative binomial) model is generally used when a single claim size is assumed to follow an exponential distribution. Gomez-Deniz and Calderin -Ojeda (2011) developed models by considering the discrete Lindley distributions. Some new models have been derived using the discrete xgamma distributions discussed in sections 2 and 3.

The next results show that one can have a closed form expression for aggregate claim assuming a discrete xgamma as primary and exponential as secondary distribution.

Theorem 4.1: If the primary distribution is a discrete xgamma-I distribution with parameter 0 < p < 1 and the secondary distribution is an exponential distribution with parameter γ > 0 as secondary distribution, then the pdf of $S = \sum_{i=1}^{n} X_i$ is given by

$$
\begin{aligned}
f_s(x) &= \frac{\gamma p}{1-\ln p}[1-(2-3p)\ln p + \frac{1-4p}{2}(\ln p)^2 \\
&+ \ln p(p-1+\frac{3-5p}{2}\ln p)\gamma px + (1-p)\frac{(\ln p)^2}{2}(\gamma px)^2]e^{-\gamma(1-p)x} \quad for\ \ x>0 \\
&= \frac{1}{1-\ln p}[1-\ln p - p\{1-2\ln p + \frac{(\ln p)^2}{2}\}] \ for\ \ x=0 \qquad (7)
\end{aligned}
$$

Proof: If the distribution of claim amount is exponential with parameter γ > 0, then nth fold convolution of exponential distribution is a gamma distribution and is given by

$$f^{n*}(x) = \frac{\gamma^n}{(n-1)!} e^{-\gamma x}, \qquad x > 0, \qquad n = 1, 2, ...$$

Then, with algebraic simplification, we have the theorem.
The mgf of ***S*** is,

$$M_s(t) = a_1 \frac{1-p}{1-p(1-\frac{t}{\gamma})^{-1}} + b_1 \frac{(1-p)^2}{\{1-p(1-\frac{t}{\gamma})^{-1}\}^2} + c_1 \frac{(1-p)^3}{\{1-p(1-\frac{t}{\gamma})^{-1}\}^3}$$

$t < \gamma(1-p)$

Hence Mean, $E(s) = [a_1 + 2b_1 + 3c_1] \frac{p}{\gamma(1-p)}$ and

the Variance,

$$\text{var}(s) = \frac{p}{\gamma^2(1-p)^2}[(1-p)(a_1 + 2b_1 + 3c_1) + (1+p)a_1 + 2(1+2p)b_1 + 3(1+3p)c_1 - p(a_1 + 2b_1 + 3c_1)^2]$$

Theorem 4.2: If the primary distribution is a discrete xgamma-II distribution with parameter 0 < p < 1 and the secondary distribution is an exponential distribution with parameter γ> 0, then the pdf of $S = \sum_{i=1}^{N} X_i$ is

$$f_s(x) = \frac{2(1-p)^3 \gamma p}{2(1-p)^2 - p(1-p)\ln p}[1 - \frac{\ln p}{2} - \frac{3\ln p}{2}(\gamma px) - \frac{\ln p}{2}(\gamma px)^2] e^{-\gamma(1-p)x} \quad for \;\; x > 0$$

$$= \frac{2(1-p)^3}{2(1-p)^2 - p(1-p)\ln p} \quad for \;\; x = 0 \qquad (8)$$

Proof: It is to be done by the same way as in case of Theorem 4.1.
The mgf of S is,

$$M_s(t) = a_2 \frac{1-p}{1-p(1-\frac{t}{\gamma})^{-1}} + b_2 \frac{(1-p)^2}{\{1-p(1-\frac{t}{\gamma})^{-1}\}^2} + c_2 \frac{(1-p)^3}{\{1-p(1-\frac{t}{\gamma})^{-1}\}^3},$$

$t < \gamma(1-p)$
and the Variance,

$$\text{var}(s) = \frac{p}{\gamma^2(1-p)^2}[(1-p)(a_2 + 2b_2 + 3c_2) + (1+p)a_2 + 2(1+2p)b_2 + 3(1+3p)c_2$$
$$-p(a_2 + 2b_2 + 3c_2)^2]$$

Theorem 4.3: [Gòmez-D´eniz and Calderin-Ojeda (2011)] If the primary distribution is a discrete Lindley distribution with parameter 0 < λ< 1 and the secondary distribution is an exponential distribution with parameter γ > 0, then the pdf of $S = \sum_{i=1}^{N} X_i$ is

$$f_s(x) = \frac{\gamma\lambda(1-\gamma+(\lambda^2\gamma x+(3-\lambda x)-2)\ln\lambda)}{1-\ln\lambda} e^{-\gamma(1-\lambda)x} \quad for \quad x>0$$

$$= \frac{1-\lambda+(2\lambda-1)\ln\lambda}{1-\ln\lambda} \quad for \quad x=0 \tag{9}$$

The mfg of S is,

$$M_s(t) = \frac{1}{1-\ln\lambda}\left[\frac{2(1-\lambda+\lambda\ln\lambda)-\ln\lambda}{1-\lambda(1-\frac{t}{\gamma})^{-1}} - \frac{e^t\lambda\ln\lambda}{\{1-p(1-\frac{t}{\gamma})^{-1}\}^2}\right], \quad t<\gamma(1-\lambda)$$

Hence, Mean $E(S) = \dfrac{\lambda\{1-\lambda+(\lambda-2)\ln\lambda\}}{\gamma(1-\lambda)^2(1-\ln\lambda)}$ and Variance,

$$Var(S) = \frac{\lambda}{\gamma^2(1-\lambda)^4(1-\ln\lambda)^2}[(1-\lambda^2)(1-\ln\lambda\{1-\lambda+(\lambda-2)\}+$$
$$(1-\lambda)^2-(3-4\lambda+\lambda^2)\ln\lambda+(2-3\lambda)(\ln\lambda)^2].$$

Aggregate claim model has been obtained using the Poisson as primary and exponential as secondary distribution [see, Rolski et al. (1999)]. Hence, the aggregate claim size distribution is

$$f_s(x) = \sqrt{\frac{\gamma\alpha}{x}} I_1(2\sqrt{\gamma\alpha x})e^{-(\alpha+\gamma x)} \quad for \quad x>0$$
$$= e^{-\alpha} \quad for \quad x=0 \tag{10}$$

Here, $\alpha > 0$ and $\gamma > 0$ are the parameters of the Poisson and exponential distributions, respectively, and

$$I_v(z) = \sum_{k=0}^{\infty} \frac{(\frac{z}{2})^{2k+v}}{\Gamma(k+1)\Gamma(v+k+1)}, z\in\mathbb{R}, v\in\mathbb{R}$$

represents the modified Bessel function of the first kind.

The mfg of S is,

$$M_s(t) = e^{-\alpha\{1-(1-\frac{t}{\gamma})^{-1}\}}, \quad t<\gamma(1-\alpha)..$$

Hence Mean, $E(s) = \dfrac{\alpha}{\gamma}$ and Variance, $Var(s) = \dfrac{2\alpha}{\gamma^2}$.

Another aggregate claim model is obtained when the negative binomial with parameters r and $0 < 1-p < 1$ is the primary distribution, and the exponential distribution is the secondary distribution. Hence, the aggregate claim size distribution is

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

Here, ${}_1F_1(.;.;.)$ is the confluent hypergeometric function.

The mfg of S is,

$$M_s(t) = (1-p)^r \{1 - p(1-\frac{t}{\gamma})^{-1}\}^{-r}, \quad t < \gamma(1-p)$$

Hence, Mean $E(s) = \dfrac{rp}{\gamma(1-p)}$ and Variance $Var(s) = \dfrac{rp(2-p)}{\gamma^2(1-p)^2}$.

5. Estimation of Parameter

If $X_1, X_2, \ldots, X_n$, be a random sample from the discrete xgamma-I distribution , the method of moments (MoM) and the maximum likelihood (ML) estimators of the parameter p are to be obtained by solving numerically

$$(a_1 + 2b_1 + 3c_1)\frac{p}{1-p} = \overline{X}$$

and

$$\frac{n}{p(1-\ln p)} + \frac{\sum X_i}{p} +$$

$$\sum_{i=1}^{n} \frac{-\frac{1+p}{p} + 2(1+\ln p) - \frac{1}{2}\ln p(2+\ln p) + \{-\frac{1}{p} + 1 - \ln p - (\ln p)^2\}X_i}{1 - \ln p - p(1 - 2\ln p + \frac{(\ln p)^2}{2}) - \ln p(1 - p(1-\ln p))X_i + \frac{(\ln p)^2}{2}(1-p)} +$$

$$\frac{\frac{1}{2}\{\frac{2(1-p)\ln p}{p} - (\ln p)^2\}X_i^2}{1 - \ln p - p(1 - 2\ln p + \frac{(\ln p)^2}{2}) - \ln p(1 - p(1-\ln p))X_i + \frac{(\ln p)^2}{2}(1-p)X_i^2} = 0$$

respectively. If the sample is from the discrete xgamma-II distribution, the method of moments (MoM) and the maximum likelihood (ML) estimators of the parameter p are to be obtained by solving numerically

$$(a_2 + 2b_2 + 3c_2)\frac{p}{1-p} = \overline{X}$$

and

$$-\frac{3n}{1-p} + n.\frac{4(1-p) + \{(1+p)(1+\ln p) + p\ln p\}}{2(1-p)^2 - p(1+p)\ln p} + \frac{\sum X_i}{p} - \sum_{i=1}^{n}\frac{X_i^2}{p(2 - X_i^2 \ln p)} = 0,$$

respectively.

6. Data Analysis

In this section, we will fit six data sets [four from Wilmot (1987), one from Bermu’dez (2009), remaining from Boucher et al. (2007)] and compared these with Poisson, Negative binomial, discrete Lindley, dxgamma-I and dxgamma-II distributions. We have shown the fitted probabilities for each data set and comparisons have been made in term of negative log-likelihood. Summarized results have been shown in Tables 1-6. In all the data sets; it is noticed that there is large proportion of zero values. We have estimated the parameter of each model by the method of maximum likelihood. From the Tables1-6, it is found that the discrete xgamma-II distribution performs better than the other distributions in negative log-likelihood sense. Following is the approximate ordering:

Worst: Poisson → Negative binomial →Discrete xgamma-I → Discrete Lindley →Discrete xgamma-II: **Best**.

So, the model described in Theorem 4.2 seems to be appropriate for collective risk modelling in the actuarial literature than the others discussed in section 4.

7. Conclusion

Two discrete versions of the xgamma distribution, viz., discrete xgamma-I and discrete xgamma-II using discrete concentration and discrete analogue approaches have beenderived. The structural and reliability properties of these distributions have been studied. Estimation procedures of the parameter have been mentioned. Compound discrete xgamma distributions in connection to collective risk model have been obtained. The new compound discrete xgamma-I and xgamma-II distributions have been compared with the classical compound Poisson, compound Negative binomial and compound discrete Lindley distributions with the help of some automobile claim data. It has been observed that the discrete xgamma-II distribution is comparatively more appropriate for study of collective risk data empirically.

Number of Claims	Observed	Fitted Poisson	Fitted Negative Binomial	Fitted dLindley	Fitted dxgamma-I	Fitted dxgamma-II
0	103704	102627.9	103217.2	103350.1	103321.3	103840.8
1	14075	15923.36	14861.67	14626.28	14607.95	13715.99
2	1766	1235.304	1604.886	1681.739	1748.554	2046.422
3	255	63.8884	154.0523	175.7091	161.5036	226.9247
4	45	2.478171	13.86321	17.36953	12.75092	20.96838
5	6	0.076900074	1.197651	1.656130	0.9115403	1.731256
6	2	0.001988605	0.1005918	0.1539398	0.06091039	0.1326198
Negative log-likelihood	-	55108.46	54697.39	54659.61	54678.22	54652.51

Table1: Data Set-I

Number of Claims	Observed	Fitted Poisson	Fitted Negative Binomial	Fitted dLindley	Fitted dxgamma-I	Fitted dxgamma-II
0	7840	7635.46	7718.056	7735.846	7730.347	7797.46
1	1317	1636.852	1494.167	1463.028	1455.685	1343.856
2	239	175.4500	216.9461	225.6468	240.053	271.7930
3	42	12.53737	27.99961	31.65992	31.0723	41.80549
4	14	0.6719245	3.387843	4.205013	3.457364	5.389682
5	4	0.02880876	0.3935191	0.5388292	0.3491404	0.6221004
6	4	0.001029313	0.04443999	0.06732125	0.03299364	0.06667545
7	1	3.152271e-05	0.004916174	0.008254485	0.00297182	0.006779088
Negative log-likelihood	-	5490.781	5388.843	5377.51	5384.057	5367.253

Table 2: Data Set-II

Number of Claims	Observed	Fitted Poisson	Fitted Negative Binomial	Fitted dLindley	Fitted dxgamma-I	Fitted dxgamma-II
0	3719	3668.600	3675.159	3676.208	3671.561	3676.639
1	232	317.2765	304.7798	302.4305	306.8567	296.5420
2	38	13.71973	18.95647	20.08010	20.50728	25.19934
3	7	0.3955141	1.048036	1.201059	1.029558	1.538505
4	3	0.00855144	0.0543208	0.0688194	0.0438056	0.0776410
5	1	0.00014791	0.002702885	0.003776502	0.001681779	0.003491711
Negative log-likelihood	-	1246.077	1221.197	1217.698	1221.520	1211.224

Table 3: Data Set-III

Number of Claims	Observed	Fitted Poisson	Fitted Negative Binomial	Fitted dLindley	Fitted dxgamma-I	Fitted dxgamma-II
0	20592	20417.77	20522.27	20544.93	20540.09	20632.49
1	2651	2947.815	2760.887	2720.243	2718.655	2560.743
2	297	212.7954	278.5692	292.3791	302.3779	355.8121
3	41	10.24078	24.98418	28.54859	25.85673	36.56728
4	7	0.3696281	2.100720	2.637138	1.887721	3.127832
5	0	0.01067301	0.1695675	0.2349461	0.1247285	0.2389654
6	1	0.0002568194	0.01330708	0.02040511	0.007701431	0.01693577
Negative log-likelihood	-	10297.85	10233.72	10228.45	10232.25	10221.59

Table 4: Data Set-IV

Number of Claims	Observed	Fitted Poisson	Fitted Negative Binomial	Fitted dLindley	Fitted dxgamma-I	Fitted dxgamma-II
0	71087	67424.99	67960.82	68052.5	67775.91	68179.48
1	6744	12363.00	11415.29	11225.99	11366.38	10612.39
2	2067	1133.436	1438.059	1507.513	1646.199	1910.340
3	690	69.27539	161.0327	184.0637	185.8067	259.1091
4	248	3.175573	16.90529	21.26899	17.98504	29.38657
5	95	0.1164542	1.703736	2.370814	1.578403	2.981471
6	34	0.003558829	0.1669351	0.2576519	0.1295657	0.2807841
7	17	9.322066e-05	0.01602279	0.02747802	0.01013466	0.02508084
8	4	2.136610e-06	0.001513872	0.002888127	0.0007644298	0.00215381
9	3	4.352973e-08	0.0001412684	0.0003000664	5.604773e-05	0.0001794176
10	3	7.981583e-10	1.305077e-05	3.088308e-05	4.017203e-06	1.458903e-05
11	2	1.330453e-11	1.195703e-06	3.15703e-06	2.826347e-07	1.163190e-06
Negative log-likelihood	-	44481.26	42392.02	42097.6	42256.75	41522.34

Table 5: Data Set-V

Number of Claims	Observed	Fitted Poisson	Fitted Negative Binomial	Fitted dLindley	Fitted dxgamma-I	Fitted dxgamma-II
0	530642	528917.3	529526.7	529646.2	529458.2	530005.8
1	33495	36734.96	35556.89	35318.90	35557.22	34536.13
2	2585	1275.679	1790.692	1896.219	1867.180	2302.655
3	211	29.53329	80.16146	92.25659	72.9354	108.9847
4	25	0.512749	3.364198	4.241825	2.40716	4.252032
Negative log-likelihood	-	146704.8	146051.2	145973.0	146039.8	145826.3

Table 6: Data Set-VI

Acknowledgement

The authors are thankful to the anonymous referees and the editor for their valuable comments that helped to improve the earlier version.

References

1. Bermu'dez, L. (2009). A priori rate making using bivariate Poisson regression models, Insurance Mathematics and Economics, 44, p. 135–141.
2. Bi, Z., Faloutsos, C. and Korn, F. (2001). The DGX distribution for mining massive skewed data, 7th Conference on Knowledge discovery and data Mining, San Francisco.
3. Boucher, J. P., Denuit, M. and Guill´en, 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), p. 110–131.
4. Bowers, JR. N. L., Gerber, H. U., Hickman, J. C., Jones, D. A. and Nesbitt, C. J.(1997). Actuarial Mathematics, Second Edition, The Society of Actuaries. Illinois: Sahaumburg.
5. Chakraborty, S. (2015). Generating discrete analogues of continuous probability distributions−A survey of methods and constructions, Journal of Statistical Distributions and Applications, 2(6), 30 pages.
6. Ghitany, M. E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its application, Mathematics and Computers in Simulation, 78, p. 493–506.
7. G´omez-D´eniz, E. and Calder´in-Ojeda, E. (2011). The discrete Lindley distribution: properties and applications, Journal of Statistical Computation and Simulation, 81, p. 1405–1416.
8. Good, I. J. (1953). The population frequencies of species and the estimation of population parameters, Biometrika, 40, p. 237–264.
9. Inusah, S. and Kozubowski, T. J. (2006). A discrete analogue of the Laplace distribution, Journal of Statistical Planning and Inference, 136, p. 1090–1102.
10. Kemp, A. W. (1997). Characterization of a discrete normal distribution, Journal of Statistical Planning and Inference, 63, p. 223–229.

11. Kemp, A. W. (2008). The Discrete Half-normal Distribution, Advances in Mathematical and Statistical Modeling, Birkh¨auser, Boston, MA, p. 353–365.
12. Kozubowski, T. J. and Inusah, S. (2006). A skew Laplace distribution on Integers, Annals of the Institute of Statistical Mathematics, 58, p. 555–571.
13. Krishna, H. and Pundir, P. S. (2009). Discrete Burr and discrete Pareto distributions, Statistical Methodology, 6, p. 177–188.
14. Lindley, D. V. (1958). Fiducial distributions and Bayes theorem, Journal of Royal Statistical Society Series B, 20(1), p. 102–107.
15. Nakagawa, T. and Osaki, S. (1975). The discrete Weibull distribution, IEEE Transactions on Reliability, 24(5), p. 300–301.
16. Rolski, T., Schmidli, H., Schmidt, V. and Teugel, J. (1999). Stochastic Processes for Insurance and Finance, John Wiley & Sons.
17. Roy, D. (2004). Discrete Rayleigh distribution, IEEE Transactions on Reliability, 53(2), p. 255–260.
18. Sato, H., Ikota, M., Aritoshi, S. and Masuda, H. (1999). A new defect distribution metrology with a consistent discrete exponential formula and its applications, IEEE Transactions on Semiconductor Manufacturing, 12(4), p. 409–418.
19. Sen, S., Maiti, S. S. and Chandra, N. (2016). The xgamma distribution: Statistical properties and application, Journal of Modern Applied Statistical Methods, 15(1), p. 774–788.
20. Willmot, G. E. (1987). The Poisson-inverse Gaussian distribution as an alternative to the negative binomial, Scandinavian Actuarial Journal, 2, p. 113–127.