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 $S(x) = \mathbb{P}[X > x]$ and times are grouped into unit intervals such that observed discrete variable is $dX = \lfloor X \rfloor$, the largest integer contained in X, the probability mass function (pmf) of $dX$ can be written as,
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

$$p(k) = P[Y = k] = \frac{f(k)}{\sum_{j=-\infty}^{\infty} f(j)}; k = 0, \pm 1, \pm 2, \pm 3, \ldots$$

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

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

The corresponding cumulative distribution function (cdf) is:

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

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

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

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

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

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
Discrete xgamma distributions: properties, estimation and an application ...

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

\[ p'_X(x) = \frac{1}{1+\theta} \left[ 1+\theta e^{-(1+\theta)x} + \frac{\theta^2}{2} (1-e^{-(1+\theta)x})^2 \right] e^{\theta x} ; \theta > 0, x = 0, 1, 2, ... \]

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

Here we find that \( p'_0 = \frac{1 - \ln p - p(1 - 2 \ln p + \frac{(\ln p)^2}{2})}{1 - \ln p} \), and the other probabilities are to be calculated using the recursive relationship

\[ p'_{x+1} = \frac{1 - \ln p - p(1 - 2 \ln p + \frac{(\ln p)^2}{2}) - (x + 1) \ln p(1 - p(1 - \ln p)) + (x + 1)^2 \frac{(1 - p)(\ln p)^2}{2}}{1 - \ln p - p(1 - 2 \ln p + \frac{(\ln p)^2}{2}) - x \ln p(1 - p(1 - \ln p)) + x^2 \frac{(1 - p)(\ln p)^2}{2}} p'_x \]

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

\[ p'_X(x) = a_G(x; 1-p) + b_NB(x; 2, 1-p) + c_NB(x; 3, 1-p) ; 0 < p < 1, x = 0, 1, 2, ... \]

with

\[ a_1 = \frac{1 - p + p \ln p + \frac{(\ln p)^2}{2}}{(1 - p)(1 - \ln p)} \], \quad b_1 = \frac{(1 - p) \ln p + (3 - p) \frac{(\ln p)^2}{2}}{(1 - p)^2 (1 - \ln p)} \], \quad c_1 = \frac{(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

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

Probability Generating function (pgf),

\[ P'_s(s) = a_1 \frac{1 - p}{1 - ps} + b_1 \frac{(1 - p)^2}{(1 - ps)^2} + c_1 \frac{(1 - p)^3}{(1 - ps)^3}, |s| < \frac{1}{p} \]

Moment generating function (mgf),
\[ M_1(t) = a_1 \frac{1 - p}{1 - pe^t} + b_1 \left( \frac{(1 - p)^2}{(1 - pe^t)^2} \right) + c_1 \left( \frac{(1 - p)^3}{(1 - pe^t)^3} \right), t < -\ln p \]

Characteristic function (cf),
\[ \phi_1(t) = a_1 \frac{1 - p}{1 - pe^{ita}} + b_1 \left( \frac{(1 - p)^2}{(1 - pe^{ita})^2} \right) + c_1 \left( \frac{(1 - p)^3}{(1 - pe^{ita})^3} \right), |t| < -\ln p \]

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,
\[ S_j(x) = P(X \geq x) = \frac{1 - \ln p - \ln px + \frac{\ln(p)^2}{2} x^2}{1 - \ln p}; 0 < p < 1, x = 0, 1, 2, \ldots \]

Hazard function,
\[ h_j(j) = \frac{1 - \ln p - p(1 - 2 \ln p + \frac{\ln(p)^2}{2} - \ln p(1 - (1 - ln p))j + \frac{\ln(p)^2}{2} (1 - p)j^2}{1 - \ln p - \ln pj + \frac{\ln(p)^2}{2} j^2} \]

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

Mean remaining life,
\[ m_j(t) = \frac{\sum_{k=t+1}^{\infty} S_j(x)}{S_j(t)} \]
\[ = p, \frac{1 - \ln p - \ln p(1 + t + \frac{p}{1 - p}) + \frac{\ln(p)^2}{2} (1 + t)^2 + 2(1 + t)p + \frac{p}{(1 - p)^2} + \frac{2p^2}{(1 - p)^3}}{1 - \ln p - t \ln p + t^2 \frac{\ln(p)^2}{2}} \]

It is to be noted that \[ \frac{1}{1 - p} < m_j(t) < \frac{p}{1 - p} \left[ 1 - \frac{\ln(p(1 + \frac{1 + p}{2} \ln p))}{(1 - p)(1 - \ln p)} \right] \] 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

\[
\text{VaR}_u[X] = \inf\{x \in \mathbb{R} : F(x) \geq u\}
\]

which is the \( u \)th quantile of the random variable \( X \)

and the TVaR is defined as

\[
TVaR_u[X] = E(X \mid X \geq x_u) = x_u + \frac{z}{1 - u}.
\]

For this distribution, \( TVaR_u[X] \) can be calculated as

\[
TVaR_u[X] = \frac{1 - \ln p - \ln [p(1 - p) + p(1 - p)^2]}{2} + \frac{(\ln p)^2}{2} - \frac{(1 + x_u)^2}{2 - \ln p - (1 + p) \ln p}.
\]

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

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

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

Here we find that

\[
p^{\mu}_0 = \frac{2(1 - p)^2}{2(1 - p)^2 - p(1 + p) \ln p},
\]

and the other probabilities are to be calculated using the recursive relationship

\[
p^{\mu}_{x+1} = p \frac{2 - (x + 1)^2 \ln p}{2 - x^2 \ln p} p^{\mu}_x.
\]

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

\[
p^{\mu}_x = a_x G(x; 1 - p) + b_x NB(x; 2, 1 - p) + c_x NB(x; 3, 1 - p); 0 < p < 1, \quad x = 0, 1, 2, ...
\]

with

\[
a_x = \frac{2(1 - p)^2(1 - \ln p)}{2(1 - p)^2 - p(1 + p) \ln p},
\]

\[
b_x = \frac{3(1 - p) \ln p}{2(1 - p)^2 - p(1 + p) \ln p},
\]

and

\[
c_x = \frac{2 \ln p}{2(1 - p)^2 - p(1 + p) \ln p}.
\]

Here the constants are \( a_x, b_x \) and \( c_x \) with \( a_x + b_x + c_x = 1 \) and we call this distribution as Discrete xgamma-II (dxgamma-II) Distribution. The cdf of this distribution is given by
Probability Generating function (pgf),
\[ P_s(s) = a_2 \frac{1-p}{1-ps} + b_2 \frac{(1-p)^2}{(1-ps)^2} + c_2 \frac{(1-p)^3}{(1-ps)^3}, |s| < \frac{1}{p} \]

Moment Generating function (mgf),
\[ M_s(t) = a_2 \frac{1-p}{1-pe^t} + b_2 \frac{(1-p)^2}{(1-pe^t)^2} + c_2 \frac{(1-p)^3}{(1-pe^t)^3}, t < -ln p \]

Characteristic function (cf),
\[ \phi_s(t) = a_2 \frac{1-p}{1-pe^{it}} + b_2 \frac{(1-p)^2}{(1-pe^{it})^2} + c_2 \frac{(1-p)^3}{(1-pe^{it})^3}, |t| < -ln p \]

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,
\[ S_s(k) = \frac{2(1-p)^3}{2(1-p)^2 - p(1+p) \ln p} \left[ \frac{1}{1-p} \frac{-\ln p}{2} + \frac{k^2}{1-p} + \frac{(2k+1)p}{(1-p)^2} + \frac{2p^2}{(1-p)^3} \right] \]

Hazard function, \( h_s(j) = \frac{1 - j^2 \ln p}{1-p} \left[ \frac{1}{1-p} - \frac{\ln p}{2} \frac{j^2}{1-p} + \frac{(2j+1)p}{(1-p)^2} + \frac{2p^2}{(1-p)^3} \right] \)

Note that \( \frac{2(1-p)^3}{2(1-p)^2 - p(1+p) \ln p} < h_s(j) < (1-p) \) for all \( j \)
Mean remaining life,
\[ m_{ij}(t) = E[X - t \mid X \geq t] = \frac{p}{1 - p} \frac{2(1 - p)^2 - \ln p(5p^2 + p + 4p(1 - p)(t + 1) + (1 - p)^2(t + 1)^2)}{2(1 - p)^2 - \ln p(2p^2 + p(1 - p)(2t + 1) + (1 - p)^2t^2)}. \]

Also note that
\[
\frac{p}{1 - p} < m_{ij}(t) < \frac{p}{1 - p} \frac{2(1 - p)^2 - \ln p(5p^2 + p + 4p(1 - p)(t + 1) + (1 - p)^2(t + 1)^2)}{2(1 - p)^2 - \ln p(2p^2 + p(1 - p)(2t + 1) + (1 - p)^2t^2)} \text{ for } \forall t.
\]

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

The pmf and cdf of dxgamma-I and dxgamma-II distributions for different values of \( p \) have been shown in figure 1.

![Figure 1](image.png)

(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 $i$th 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 $n$th 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 $Var(S) = Var(X)E(N) + \gamma^2XE^2(X)Var(N)$. 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 $\gamma > 0$ as secondary distribution, then the pdf of $S = \sum_{i=1}^{N} X_i$ is given by

$$f_S(x) = \frac{\gamma p}{1-\ln p} [1-(2-3p)\ln p + \frac{1-4p}{2}(\ln p)^2 + \ln p(p-1)+\frac{3-5p}{2}\ln p\gamma px + (1-p)\frac{\ln p^2}{2}(\gamma px^2)e^{-\gamma(1-p)x} \text{ for } x > 0$$

$$= \frac{1}{1-\ln p} [1-\ln p - p(1-2\ln p + \frac{\ln p^2}{2})] \text{ for } x = 0 \quad (7)$$
**Proof:** If the distribution of claim amount is exponential with parameter $\gamma > 0$, then the $n$th fold convolution of exponential distribution is a gamma distribution and is given by

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

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, $E(S) = \frac{[a_1 + 2b_1 + 3c_1]}{\gamma(1-p)}$ and the Variance,

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

**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 $\gamma > 0$, then the pdf of $S = \sum_{i=1}^{N} X_i$ is

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

$$= \frac{2(1-p)^3}{2(1-p)^2 - p(1-p) \ln p} \text{ for } x = 0$$

(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,

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

**Theorem 4.3:** [Gómez-D´eniz and Calderin-Ojeda (2011)] If the primary distribution is a discrete Lindley distribution with parameter $0 < \lambda < 1$ and the secondary distribution is an exponential distribution with parameter $\gamma > 0$, then the pdf of $S = \sum_{i=1}^{N} X_i$ is
\( f_i(x) = \frac{\gamma^x}{x!} I_1(2\sqrt{\gamma\alpha x}) e^{-(\alpha + \gamma x)} \quad \text{for} \quad x > 0 \\
= e^{-\alpha} \quad \text{for} \quad x = 0 \) 

(10)

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

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

represents the modified Bessel function of the first kind.

The mgf of \( S \) is,

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

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

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

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

\[ f_i(x) = \frac{\lambda^{x-1} e^{-\lambda}}{x!} I_1(2\sqrt{\lambda x}) e^{-(\alpha + \gamma x)} \quad \text{for} \quad x > 0 \\
= e^{-\alpha} \quad \text{for} \quad x = 0 \]
\( M_s(t) = (1 - p)^r \{1 - p(1 - t)^{-1}\}^{-r}, \quad t < \gamma(1 - p) \)

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

### 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

\[
(a_i + 2b_i + 3c_i) - \frac{p}{1 - p} = \overline{X}
\]

and

\[
\frac{n}{p(1 - \ln p)} + \sum X_i + \frac{-1 + p}{p} + 2(1 + \ln p) - \frac{1}{2} \ln p(2 + \ln p) + \{ - \frac{1}{p} + 1 - \ln p - (\ln p)^2 \} X_i + \frac{(\ln p)^2}{2} (1 - p) + \frac{1}{2} \frac{2(1 - p) \ln p}{p} - (\ln p)^2 X_i^2 - \frac{1}{2} \frac{2(1 - p) \ln p}{p} - (\ln p)^2 X_i^2 = 0
\]

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

\[
(a_i + 2b_i + 3c_i) - \frac{p}{1 - p} = \overline{X}
\]

and

\[
\frac{3n}{1 - p} + n \frac{4(1 - p) + \{1 + p(1 + \ln p) + p \ln p\}}{2(1 - p)^2 - p(1 + p) \ln p} + \sum X_i - \frac{\sum X_i^2}{p(2 - X_i \ln p)} = 0
\]

respectively.

### 6. Data Analysis

In this section, we will fit six data sets [four from Wilmot (1987), one from Bermu’dex (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 Tables 1-6, 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 been derived. 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.

<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>Observed</th>
<th>Fitted Poisson</th>
<th>Fitted Negative Binomial</th>
<th>Fitted dLindley</th>
<th>Fitted dxgamma-I</th>
<th>Fitted dxgamma-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>103704</td>
<td>102627.9</td>
<td>103217.2</td>
<td>103350.1</td>
<td>103321.3</td>
<td>103840.8</td>
</tr>
<tr>
<td>1</td>
<td>14075</td>
<td>15923.36</td>
<td>14861.67</td>
<td>14626.28</td>
<td>14607.95</td>
<td>13715.99</td>
</tr>
<tr>
<td>2</td>
<td>1766</td>
<td>1235.304</td>
<td>1604.886</td>
<td>1681.739</td>
<td>1748.554</td>
<td>2046.422</td>
</tr>
<tr>
<td>3</td>
<td>255</td>
<td>63.8884</td>
<td>154.0523</td>
<td>175.7091</td>
<td>161.5036</td>
<td>226.9247</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>2.478171</td>
<td>13.86321</td>
<td>17.36953</td>
<td>12.75092</td>
<td>20.96838</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0.07690007</td>
<td>1.197651</td>
<td>1.656130</td>
<td>0.9115403</td>
<td>1.731256</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0.00198860</td>
<td>0.100591</td>
<td>0.153939</td>
<td>0.0609103</td>
<td>0.132619</td>
</tr>
<tr>
<td>Negative log-likelihood</td>
<td>-</td>
<td>55108.46</td>
<td>54697.39</td>
<td>54659.61</td>
<td>54678.22</td>
<td>54652.51</td>
</tr>
</tbody>
</table>

Table 1: Data Set-I
<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>Observed</th>
<th>Fitted Poisson</th>
<th>Fitted Negative Binomial</th>
<th>Fitted dLindley</th>
<th>Fitted dxgamma-I</th>
<th>Fitted dxgamma-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7840</td>
<td>7635.46</td>
<td>7718.056</td>
<td>7735.846</td>
<td>7730.347</td>
<td>7797.46</td>
</tr>
<tr>
<td>1</td>
<td>1317</td>
<td>1636.852</td>
<td>1494.167</td>
<td>1463.028</td>
<td>1455.685</td>
<td>1343.856</td>
</tr>
<tr>
<td>2</td>
<td>239</td>
<td>175.4500</td>
<td>216.9461</td>
<td>225.6468</td>
<td>240.053</td>
<td>271.7930</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>12.53737</td>
<td>27.99961</td>
<td>31.65992</td>
<td>31.0723</td>
<td>41.80549</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>0.6719245</td>
<td>3.387843</td>
<td>4.205013</td>
<td>3.457364</td>
<td>5.389682</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.0288087</td>
<td>0.393519</td>
<td>0.5388292</td>
<td>0.349140</td>
<td>0.6221004</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>0.0010293</td>
<td>0.0444399</td>
<td>0.0673212</td>
<td>0.032993</td>
<td>0.0666754</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3.152271e-05</td>
<td>0.0049161</td>
<td>0.0082544</td>
<td>0.002971</td>
<td>0.006790</td>
</tr>
<tr>
<td>Negative log-likelihood</td>
<td>-</td>
<td>5490.781</td>
<td>5388.843</td>
<td>5377.51</td>
<td>5384.057</td>
<td>5367.253</td>
</tr>
</tbody>
</table>

Table 2: Data Set-II

<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>Observed</th>
<th>Fitted Poisson</th>
<th>Fitted Negative Binomial</th>
<th>Fitted dLindley</th>
<th>Fitted dxgamma-I</th>
<th>Fitted dxgamma-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3719</td>
<td>3668.600</td>
<td>3675.159</td>
<td>3676.208</td>
<td>3671.561</td>
<td>3676.639</td>
</tr>
<tr>
<td>1</td>
<td>232</td>
<td>317.2765</td>
<td>304.7798</td>
<td>302.4305</td>
<td>306.8567</td>
<td>296.5420</td>
</tr>
<tr>
<td>2</td>
<td>38</td>
<td>13.71973</td>
<td>18.95647</td>
<td>20.08010</td>
<td>20.50728</td>
<td>25.19934</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>0.3955141-01</td>
<td>1.048036</td>
<td>1.201059</td>
<td>1.029558</td>
<td>1.538505</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.00855144</td>
<td>0.0543208</td>
<td>0.0688194</td>
<td>0.0438056</td>
<td>0.0776410</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.00014791</td>
<td>0.00270285</td>
<td>0.003776502</td>
<td>0.00168179</td>
<td>0.003491711</td>
</tr>
<tr>
<td>Negativ e log-likelihood</td>
<td>-</td>
<td>1246.077</td>
<td>1221.197</td>
<td>1217.698</td>
<td>1221.520</td>
<td>1211.224</td>
</tr>
</tbody>
</table>

Table 3: Data Set-III
### Table 4: Data Set-IV

<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>Observed</th>
<th>Fitted Poisson</th>
<th>Fitted Negative Binomial</th>
<th>Fitted dLindley</th>
<th>Fitted dxgamma-I</th>
<th>Fitted dxgamma-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20592</td>
<td>20417.77</td>
<td>20522.27</td>
<td>20544.93</td>
<td>20540.09</td>
<td>20632.49</td>
</tr>
<tr>
<td>1</td>
<td>2651</td>
<td>2947.815</td>
<td>2760.887</td>
<td>2720.243</td>
<td>2718.655</td>
<td>2560.743</td>
</tr>
<tr>
<td>2</td>
<td>297</td>
<td>217.2954</td>
<td>278.5692</td>
<td>292.3791</td>
<td>302.3779</td>
<td>355.8121</td>
</tr>
<tr>
<td>3</td>
<td>41</td>
<td>10.24078</td>
<td>24.98418</td>
<td>28.54859</td>
<td>25.85673</td>
<td>36.56728</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>0.3696281</td>
<td>2.100720</td>
<td>2.637138</td>
<td>1.887721</td>
<td>3.127832</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.01067301</td>
<td>0.169567</td>
<td>0.234946</td>
<td>0.1247285</td>
<td>0.238965</td>
</tr>
</tbody>
</table>

### Table 5: Data Set-V

<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>Observed</th>
<th>Fitted Poisson</th>
<th>Fitted Negative Binomial</th>
<th>Fitted dLindley</th>
<th>Fitted dxgamma-I</th>
<th>Fitted dxgamma-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>71087</td>
<td>67424.99</td>
<td>67960.82</td>
<td>68052.5</td>
<td>67775.91</td>
<td>68179.48</td>
</tr>
<tr>
<td>1</td>
<td>6744</td>
<td>12363.00</td>
<td>11415.29</td>
<td>11225.99</td>
<td>11366.38</td>
<td>10612.39</td>
</tr>
<tr>
<td>2</td>
<td>2067</td>
<td>1133.436</td>
<td>1438.059</td>
<td>1507.513</td>
<td>1646.199</td>
<td>1910.340</td>
</tr>
<tr>
<td>3</td>
<td>690</td>
<td>69.27539</td>
<td>161.0327</td>
<td>184.0637</td>
<td>185.8067</td>
<td>259.1091</td>
</tr>
<tr>
<td>4</td>
<td>248</td>
<td>3.175573</td>
<td>16.90529</td>
<td>21.26899</td>
<td>17.98504</td>
<td>29.38657</td>
</tr>
<tr>
<td>5</td>
<td>95</td>
<td>0.1164542</td>
<td>1.703736</td>
<td>2.370814</td>
<td>1.578403</td>
<td>2.981471</td>
</tr>
<tr>
<td>6</td>
<td>34</td>
<td>0.003558829</td>
<td>0.1669351</td>
<td>0.2576519</td>
<td>0.129567</td>
<td>0.2807841</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>9.322066e-05</td>
<td>0.0160227</td>
<td>0.0274780</td>
<td>0.0101346</td>
<td>0.0250808</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2.136610e-06</td>
<td>0.0015138</td>
<td>0.0028881</td>
<td>0.0007644</td>
<td>0.0021538</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>4.352973e-08</td>
<td>0.0001412</td>
<td>0.0003000</td>
<td>5.604773e-05</td>
<td>0.0001794</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>7.981583e-10</td>
<td>1.305077e-05</td>
<td>3.088308e-05</td>
<td>4.017203e-06</td>
<td>1.458903e-05</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>1.330453e-11</td>
<td>1.195703e-06</td>
<td>3.15703e-06</td>
<td>2.826347e-07</td>
<td>1.163190e-06</td>
</tr>
</tbody>
</table>

### Table 6: Data Set-VI

<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>Observed</th>
<th>Fitted Poisson</th>
<th>Fitted Negative Binomial</th>
<th>Fitted dLindley</th>
<th>Fitted dxgamma-I</th>
<th>Fitted dxgamma-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20592</td>
<td>20417.77</td>
<td>20522.27</td>
<td>20544.93</td>
<td>20540.09</td>
<td>20632.49</td>
</tr>
<tr>
<td>1</td>
<td>2651</td>
<td>2947.815</td>
<td>2760.887</td>
<td>2720.243</td>
<td>2718.655</td>
<td>2560.743</td>
</tr>
<tr>
<td>2</td>
<td>297</td>
<td>217.2954</td>
<td>278.5692</td>
<td>292.3791</td>
<td>302.3779</td>
<td>355.8121</td>
</tr>
<tr>
<td>3</td>
<td>41</td>
<td>10.24078</td>
<td>24.98418</td>
<td>28.54859</td>
<td>25.85673</td>
<td>36.56728</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>0.3696281</td>
<td>2.100720</td>
<td>2.637138</td>
<td>1.887721</td>
<td>3.127832</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.01067301</td>
<td>0.169567</td>
<td>0.234946</td>
<td>0.1247285</td>
<td>0.238965</td>
</tr>
</tbody>
</table>

### Table 7: Data Set-VII

<table>
<thead>
<tr>
<th>Number of Claims</th>
<th>Observed</th>
<th>Fitted Poisson</th>
<th>Fitted Negative Binomial</th>
<th>Fitted dLindley</th>
<th>Fitted dxgamma-I</th>
<th>Fitted dxgamma-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20592</td>
<td>20417.77</td>
<td>20522.27</td>
<td>20544.93</td>
<td>20540.09</td>
<td>20632.49</td>
</tr>
<tr>
<td>1</td>
<td>2651</td>
<td>2947.815</td>
<td>2760.887</td>
<td>2720.243</td>
<td>2718.655</td>
<td>2560.743</td>
</tr>
<tr>
<td>2</td>
<td>297</td>
<td>217.2954</td>
<td>278.5692</td>
<td>292.3791</td>
<td>302.3779</td>
<td>355.8121</td>
</tr>
<tr>
<td>3</td>
<td>41</td>
<td>10.24078</td>
<td>24.98418</td>
<td>28.54859</td>
<td>25.85673</td>
<td>36.56728</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>0.3696281</td>
<td>2.100720</td>
<td>2.637138</td>
<td>1.887721</td>
<td>3.127832</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.01067301</td>
<td>0.169567</td>
<td>0.234946</td>
<td>0.1247285</td>
<td>0.238965</td>
</tr>
</tbody>
</table>
Discrete xgamma distributions: properties, estimation and an application ...

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

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