Compound Zero-Truncated Poisson Normal Distribution and its Applications

Mohammad Z. Raqab[a,b,*], Debasis Kundu[c], Fahimah A. Al-Awadhi[d]

[a] *Department of Mathematics, The University of Jordan, Amman 11942, JORDAN*

[b] *King Abdulaziz University, Jeddah, SAUDI ARABIA*

[c] *Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, Pin 208016, INDIA*

[d] *Department of Statistics and OR, Kuwait University, Safat 13060, KUWAIT*

Abstract

Here, we first propose three-parameter model and call it as the compound zero-truncated Poisson normal (ZTP-N) distribution. The model is based on the random sum of N independent Gaussian random variables, where N is a zero truncated Poisson random variable. The proposed ZTP-N distribution is a very flexible probability distribution function. The probability density function can take variety of shapes. It can be both positively and negatively skewed, moreover, normal distribution can be obtained as a special case. It can be unimodal, bimodal as well as multimodal also. It has three parameters. An efficient EM type algorithm has been proposed to compute the maximum likelihood estimators of the unknown parameters. We further propose a four-parameter bivariate distribution with continuous and discrete marginals, and discuss estimation of unknown parameters based on the proposed EM type algorithm. Some simulation experiments have been performed to see the effectiveness of the proposed EM type algorithm, and one real data set has been analyzed for illustrative purposes.

Keywords: Poisson distribution; skew normal distribution; maximum likelihood estimators; EM type algorithm; Fisher information matrix.

AMS Subject Classification (2010): 62E15; 62F10; 62H10.

* *Corresponding author. E-mail addresses: mraqab@ju.edu.jo (M. Raqab), kundu@iitk.ac.in (D. Kundu), mapfahaa@hotmail.com (F. Al-Awadhi).*

1 Introduction

The compound distributions have been used in several studies, including risk measurement, capital allocation and aggregate claims data. In addition, the compound distributions arise naturally in various queuing models as well as the theory of dams, see for example Willmot and Lin [17]. There are many cases where the classical normal distribution is an inadequate model for risks or losses and some risks also exhibit heavy tails and multimodality. It may be mentioned that the normal and some related distributions have been used quite extensively for analyzing claims data, see for example Kazemi and Noorizadeh [10] and see the references cited therein.

The skew-normal distribution was first introduced by Azzalini [4] as a natural extension of the normal density to accommodate asymmetry. Let us recall that a random variable X is said to have a skew-normal distribution with the skewness parameter $\lambda \in (-\infty, \infty)$, if it has the probability density function (PDF)

$$f_{SN}(x; \lambda) = 2\phi(x)\Phi(\lambda x); \qquad -\infty < x < \infty, \tag{1}$$

where $\phi(x)$ and $\Phi(x)$ denote the PDF and cumulative distribution function (CDF) of a standard normal distribution. The location and scale parameters also can be easily incorporated in (1). The PDF in (1) is positively and negatively skewed if $\lambda > 0$ and $\lambda < 0$, respectively. When $\lambda = 0$, it coincides with the standard normal PDF. The idea of obtaining the skewed version of a standard normal distribution can be applied to any symmetric distribution. In general if $g(x)$ is any symmetric PDF defined on $(-\infty, \infty)$ and $G(x)$ is its CDF, then for any $\lambda \in (-\infty, \infty)$, the PDF

$$f_{SG}(x; \lambda) = 2g(x)G(\lambda x); \qquad -\infty < x < \infty, \tag{2}$$

is a skewed version of the PDF $g(x)$.

For details on some developments on skew-normal distribution and other related distributions, interested readers may refer to Azzalini [5], Jamalizadeh, Behboodian, and Balakrishnan [8], Sharafi and Behboodian [16], Asgharzadeh, Esmaily, and Nadarajah [2] and Asgharzadeh, Esmaily, and Nadarajah [3]. The skew-normal distribution is a very flexible distribution and its PDF can take a variety of shapes, but it cannot have heavy tails or/and multimodality. Moreover, the maximum likelihood estimators (MLEs) may not always exist. It is well known that for any sample size, there is a positive probability that the MLEs will not exist, see for example Pewsey [15], Gupta and Gupta [7] and Kundu [12] in this respect. In fact the same problem exists in general for model (2) also.

The main aim of this paper is to propose a new model which is flexible, it can be both positively and negatively skewed, normal distribution can be obtained as a special case, can have heavy tails and can be multimodal also. We call it as the compound zero truncated Poisson-normal (ZTP-N) distribution with three parameters. It has an absolute continuous distribution function. Moreover, the normal distribution can be obtained as a special case of the compound ZTP-N distribution.

Estimation of the unknown parameters is always a challenging problem. The MLEs of the unknown parameters cannot be obtained in explicit forms, although it always exists. The MLEs can be obtained only by solving three non-linear equations simultaneously. Newton-Raphson method or some of its variants may be used to compute the MLEs, but it has the standard problem of non-convergence or even if it converges it may converge to a local minimum than the global minimum. Moreover, the choice of the initial guesses is also very important. To avoid that problem we have proposed to use an expectation maximization (EM) type algorithm to compute the MLEs in this case as it was originally proposed by Kundu and Nekoukhot [13] in case of discrete missing values. It is observed at each 'E'-step the corresponding 'M'-step can be obtained without solving any non-linear equations. Hence, the implementation of the proposed EM type algorithm is very simple and it is very efficient as well. Moreover, using the observed Fisher information matrix as suggested by Louis [14], the confidence intervals of the unknown parameters also can be obtained quite conveniently. We have performed some simulation experiments to see how the proposed EM type algorithm performs and also the analysis of a data set to see how the proposed model behaves, performances are quite satisfactory.

Furthermore, we will define a bivariate distribution which has ZTP-N and Poisson marginals. We call this new distribution as the bivariate zero-truncated Poisson-normal(BZTP-N) distribution with four parameters. The BZTP-N distribution can be used quite effectively to analyze bivariate insurance claim data when one component represents the amount of claims and the other component represents the number of claims. Applications of the proposed BZTP-N distribution range from finance to hydrology and climate. We derive different properties of the BZTP-N distribution including marginals and conditional distributions, joint integral transformations, product moment, infinite divisibility, stability with respect to Poisson summation etc. In this case also the MLEs of the unknown parameters cannot be obtained in closed form. We propose an efficient EM type algorithm similar to the previous case, to compute the MLEs of the parameters. One data set has been analyzed for illustrative purposes. Some simulation experiments have also been performed to see the behavior of the proposed EM type algorithm.

The rest of the paper is organized as follows. In Section 2 we define the compound ZTP-N distribution and derive its properties. We also provide the estimation of the unknown parameters of the ZTP-N model in this section. The BZTP-N distribution has been introduced and the estimation procedure of the model parameters is discussed in Section 3. Some simulation experiments have been performed and the results are presented in Section 4. One real data set has been analyzed in Section 5, and finally we conclude the paper in Section 6.

2 Poisson Skew Normal Distribution

2.1 Definition, CDF, PDF and MGF

Suppose $\{X_1, X_2, \ldots,\}$ is a sequence of independent identically distributed (i.i.d.) normal random variables with mean μ and variance σ^2 ($\mathrm{N}(\mu, \sigma^2)$), M is a Poisson random variable with mean θ ($\mathrm{Poi}(\theta)$) and M is independent of $X_1, X_2, \ldots$. Let us define

$$M^* \stackrel{d}{=} \{M | M \geq 1\}, \tag{3}$$

here $\stackrel{d}{=}$ means equal in distribution. Then the random variable

$$Y = \sum_{i=1}^{M^*} X_i \tag{4}$$

is said to have a ZTP-N distribution, and it will be denoted by ZTP-N(μ, σ, θ). Therefore, ZTP-N distribution can be seen as a random Poisson sum of normal random variables.

Now we would like to derive different properties of the ZTP-N distribution. Observe that

$$P(Y \leq y, M^* = m) = \frac{\theta^m}{m!(e^\theta - 1)} \Phi\left(\frac{y - m\mu}{\sigma\sqrt{m}}\right).$$

Note that

$$\lim_{\theta \to 0} P(Y \leq y, M^* = m) = \begin{cases} \Phi\left(\frac{y-\mu}{\sigma}\right) & \text{if} \quad m = 1 \\ 0 & \text{if} \quad m > 1. \end{cases}$$

The CDF of Y for $-\infty < y < \infty$, is given by

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= \sum_{m=1}^{\infty} P(Y \leq y, M^* = m) = \sum_{m=1}^{\infty} P(Y \leq y | M^* = m) P(M^* = m) \\ &= \frac{1}{e^\theta - 1} \sum_{m=1}^{\infty} \Phi\left(\frac{y - m\mu}{\sigma\sqrt{m}}\right) \frac{\theta^m}{m!}. \end{aligned} \tag{5}$$

The corresponding PDF can be obtained as

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{\sigma(e^\theta - 1)} \sum_{m=1}^{\infty} \phi\left(\frac{y - m\mu}{\sigma\sqrt{m}}\right) \frac{\theta^m}{\sqrt{m}m!}. \tag{6}$$

Hence, from (5), we immediately obtain

$$\lim_{\theta \to 0} F_Y(y) = \Phi\left(\frac{y - \mu}{\sigma}\right).$$

It is clear the PDF of a ZTP-N distribution is an infinite mixture of normal PDFs and the weights are truncated Poisson weights. We have provided the plots of the PDFs of the ZTP-N distribution for different parameter values in Figure 1. It can take variety of shapes. It is symmetric if $\mu = 0$, for μ positive it is positively skewed and for $\mu < 0$, it is negatively skewed. It can be multimodal and it can have heavy tails.

It is immediate that the hazard function of ZTP-N distribution is an increasing function for all values of μ, σ and θ. It simply follows as the hazard function of a normal distribution is an increasing function, and ZTP-N is an infinite mixture of normal distributions.

Now we derive the moment generating function (MGF) of the proposed ZTP-N distribution. The MGF can be used to compute different moments, and further we derive several properties of a ZTP-N distribution. If Y follows ZTP-N, then the moment generating function of Y can be obtained as

$$\varphi_Y(t) = P_{M^*}(\varphi_X(t)) = \frac{e^{\theta e^{\mu t + \sigma^2 t^2/2}} - 1}{e^\theta - 1}; \qquad -\infty < t < \infty.$$

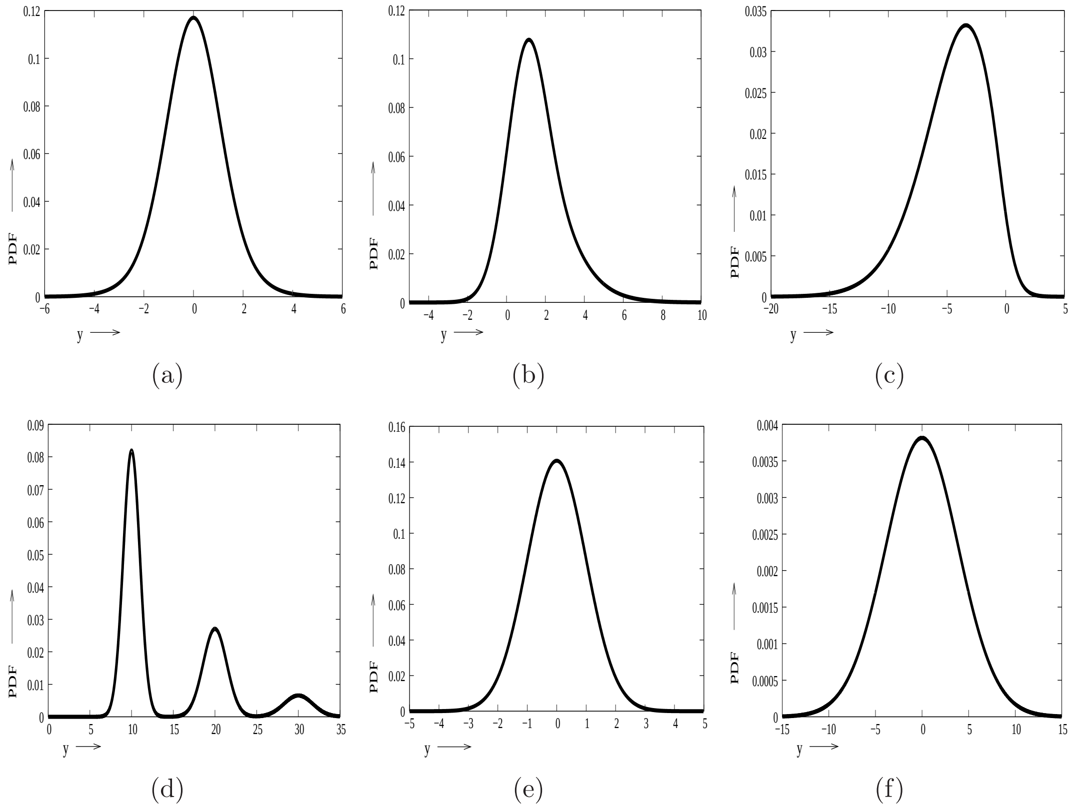

Figure 1: PDF plots of PSN(μ, σ, θ) distribution for different parameter values: (a) PSN(0.0,1.0,1.0) (b) (1.0,1.0,1.0) (c) (-5.0, 1.0, 1.0) (d) (10, 1.0, 1.0) (e) (0.0, 1.0, 0.001) (f) (0.0, 1.0, 20.0).

Here, P_{M^*} denotes the probability generating function of M^*, as defined in (3) and $\varphi_X(t)$ denotes the MGF of X, when X follows $N(\mu, \sigma^2)$.

Therefore, the mean and variance of Y can be obtained as

$$E(Y) = \frac{\theta \mu e^{\theta}}{e^{\theta} - 1}, \text{ and } Var(Y) = \theta^2 \left(1 - \frac{1}{(e^{\theta} - 1)^2}\right) \mu^2 + \frac{\theta \sigma^2}{(e^{\theta} - 1)},$$

respectively. In general, the k-th moment of Y can be expressed as

$$E(Y^k) = \sum_{n^*=1}^{\infty} \xi_k(n^* \mu, n^* \sigma^2) \frac{(\theta)^{n^*}}{n^*!(e^{\theta} - 1)},$$

where $\xi_k(n^* \mu, n^* \sigma^2)$ is the k-th moment of $N(n^* \mu, n^* \sigma^2)$, and they are available in Johnson, Kotz, and Balakrishnan (1995). If $\mu = 0$, and $\sigma = 1$, then

$$\xi_k(0, n^*) = \begin{cases} 0, & if \quad \text{k is odd}, \\ \frac{(2n^*)^{k/2}\, \Gamma(\frac{k+1}{2})}{\sqrt{\pi}}, & if \quad \text{k is even}. \end{cases}$$

The skewness measure of Y is also derived to be

$$\begin{aligned}\gamma_1 &= \left[\left(1+\theta-\frac{\theta}{1-e^{-\theta}}\right)\mu^2+\sigma^2\right]^{-3/2} \\ &\times \left\{\left(\frac{\theta}{1-e^{-\theta}}\right)^{-1/2}\left[(1+\theta)^2\mu^3+3(1+\theta)\mu\sigma^2+\theta\mu^3\right]\right. \\ &- \left. 3\left(\frac{\theta}{1-e^{-\theta}}\right)^{1/2}\mu\left[\left(1+\theta-\frac{\theta}{1-e^{-\theta}}\right)\mu^2+\sigma^2\right]-\left(\frac{\theta}{1-e^{-\theta}}\right)^{3/2}\mu^3\right\}.\end{aligned}$$

2.2 Distributional Properties

Now we present different conditional distributional properties and some conditional expectations of the proposed ZTP-N distribution, which may be useful in the implementation of the EM type algorithm, for some goodness of fit and some other purposes.

Consider the following bivariate mixed random variables (Y, M^*), where Y and M^* are same as defined before. First let us consider the following probability for Y for integers $m \geq n > 0$.

$$\begin{aligned}P(Y \leq y, M^* \leq n | M^* \leq m) &= \frac{P(Y \leq y, M^* \leq n)}{P(M^* \leq m)} \\ &= \frac{1}{\sum_{j=1}^{m}\theta^j/j!}\sum_{k=1}^{n}\Phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)\theta^k/k!.\end{aligned}$$

If $m < n$, then

$$\begin{aligned}P(Y \leq y, M^* \leq n | M^* \leq m) &= \frac{P(Y \leq y, M^* \leq m)}{P(M^* \leq m)} \\ &= \frac{1}{\sum_{j=1}^{m}\theta^j/j!}\sum_{k=1}^{m}\Phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)\theta^k/k! = P(Y \leq y | M^* \leq m).\end{aligned}$$

For any integer $n > 0$, and for $x \geq y$

$$\begin{aligned}P(Y \leq y, M^* \leq n | Y \leq x) &= \frac{P(Y \leq y, M^* \leq n)}{P(Y \leq x)} \\ &= \frac{1}{\sum_{j=1}^{\infty}\Phi\left(\frac{x-j\mu}{\sigma\sqrt{j}}\right)\theta^j/j!}\sum_{k=1}^{n}\Phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)\theta^k/k!.\end{aligned}$$

Similarly, for any integer $m > 0$ and for real $y \geq x$,

$$\begin{aligned} P(Y \leq y, M^* \leq n | Y \leq x) &= \frac{P(Y \leq x, M^* \leq n)}{P(Y \leq x)} \\ &= \frac{1}{\sum_{j=1}^{\infty} \Phi\left(\frac{x-j\mu}{\sigma\sqrt{j}}\right) \theta^j/j!} \sum_{k=1}^{n} \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) \theta^k/k! \\ &= P(M^* \leq n | Y \leq x). \end{aligned}$$

From the joint distribution of (Y, M^*), we can obtain

$$P(M^* \leq m | Y \leq y) = \frac{1}{\sum_{j=1}^{\infty} \Phi\left(\frac{x-j\mu}{\sigma\sqrt{j}}\right) \theta^j/j!} \sum_{k=1}^{m} \Phi\left(\frac{x-k\mu}{\sigma\sqrt{k}}\right) \theta^k/k!. \tag{7}$$

The conditional probability mass function (PMF) of M^* given $Y = y$ is given by

$$P(M^* = m | Y = y) = \frac{\frac{\theta^m}{m!} \phi\left(\frac{y-m\mu}{\sigma\sqrt{m}}\right)}{\sum_{k=1}^{\infty} \frac{\theta^k}{\sigma\sqrt{k}k!} \phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)} \tag{8}$$

Hence

$$E(M^* | Y = y) = \frac{\sum_{m=1}^{\infty} \frac{\theta^m}{(m-1)!} \phi\left(\frac{y-m\mu}{\sigma\sqrt{m}}\right)}{\sum_{k=1}^{\infty} \frac{\theta^k}{\sigma\sqrt{k}k!} \phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)} \tag{9}$$

and

$$E(1/M^* | Y = y) = \frac{\sum_{m=1}^{\infty} \frac{\theta^m}{m\ m!} \phi\left(\frac{y-m\mu}{\sigma\sqrt{m}}\right)}{\sum_{k=1}^{\infty} \frac{\theta^k}{\sigma\sqrt{k}k!} \phi\left(\frac{y-k\mu}{\sigma\sqrt{k}}\right)} \tag{10}$$

2.3 Maximum Likelihood Estimators

In this section we discuss about the maximum likelihood estimators (MLEs) of the unknown parameters based on a random sample $\{y_1, \ldots, y_n\}$ from a ZTP-N(μ, σ, θ). The MLEs of the unknown parameters can be obtained by maximizing the log-likelihood function

$$l(\mu, \sigma, \theta) = \sum_{i=1}^{n} \ln f_Y(y_i; \mu, \sigma, \theta), \tag{11}$$

here $f_Y(y; \mu, \sigma, \theta)$ is the right hand side of (6). The normal equations can be obtained and they will be non-linear in nature. Clearly, they cannot be obtained in explicit forms, they need to be solved by solving three non-linear equations simultaneously. The standard Newton-Raphson method or some of its variants may be used to solve these non-linear equations. But it is well known that the Newton-Raphson method requires a very good choice of the initial guesses otherwise it may not converge, moreover even when it converges it may converge to a local maximum rather than the global maximum. To avoid that problem we propose to use the EM type algorithm which can be incorporated quite easily in this case. We treat this problem as a missing value problem. It is assumed that the complete sample is of the form $\{(y_1, m_1), \ldots, (y_n, m_n)\}$ from (Y, M^*). But we observe only $\{y_1, \ldots, y_n\}$ and $m_1, \ldots, m_n$ are missing observations. Therefore, the following EM type algorithm can be used to compute the MLEs of the unknown parameters. Suppose at the k-th stage of the algorithm the estimates of μ, σ and θ are $\mu^{(k)}$, $\sigma^{(k)}$ and $\theta^{(k)}$, respectively. Let us write $\Theta^{(k)} = (\mu^{(k)}, \sigma^{(k)}, \theta^{(k)})$.

Now at the 'E'-step we need to obtain the 'pseudo' log-likelihood function, and it can be written as

$$l_{pseudo}(\Theta|\Theta^{(k)}) = -n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{1}{m_i^{(k)}} (y_i - m_i^{(k)} \mu)^2 + \left(\sum_{i=1}^{n} m_i^{(k)} \right) \ln \theta - n \ln \left(e^{\theta} - 1 \right). \quad (12)$$

Here $m_i^{(k)}$ is the missing m_i value at the k-th stage and in the traditional EM algorithm, the choice of $m_i^{(k)}$ is

$$m_i^{(k)} = \frac{\sum_{m=1}^{\infty} \frac{(\theta^{(k)})^m}{(m-1)!} \phi\left(\frac{y - m\mu^{(k)}}{\sigma^{(k)}\sqrt{m}} \right)}{\sum_{j=1}^{\infty} \frac{(\theta^{(k)})^j}{\sigma^{(k)}\sqrt{j}j!} \phi\left(\frac{y - j\mu^{(k)}}{\sigma^{(k)}\sqrt{j}} \right)}. \quad (13)$$

Note that (13) is obtained from (9) by replacing μ, σ and θ with $\mu^{(k)}$, $\sigma^{(k)}$ and $\theta^{(k)}$, respectively. There are two problems associated with it. First of all, $m_i^{(k)}$ as defined in (13) need not be an integer, and secondly, it involves computing two infinite series. To avoid that problem Kundu and Nekoukhou [13] proposed to use the following $m_i^{(k)}$ at the 'E'-step for a discrete missing value:

$$m_i^{(k)} = \arg\max_m P(M^* = m | Y = y_i, \Theta^{(k)}). \quad (14)$$

Note that $P(M^* = m|Y = y_i, \Theta^{(k)})$ is obtained from the expression (8) by replacing μ, σ and θ by $\mu^{(k)}$, $\sigma^{(k)}$ and $\theta^{(k)}$, respectively. Hence, $m_i^{(k+1)}$, will be always a positive integer. Here, $m_i^{(k)}$ is the maximum likelihood predictor rather than the expected predictor. It is observed by extensive simulations that it is very efficient and it converges to the MLEs quite quickly.

Now at the 'M'-step we can obtain $\Theta^{(k+1)}$ as

$$\Theta^{(k+1)} = \arg\max_{\Theta} l_{pseudo}(\Theta|\Theta^{(k)}).$$

Hence, it can be easily seen that

$$\mu^{(k+1)} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i^{(k)}} \quad \text{and} \quad \sigma^{(k+1)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i^{(k)}}(y_i - m_i^{(k)}\mu^{(k)})^2}, \tag{15}$$

and $\theta^{(k+1)}$ is obtained by maximizing the profile 'pseudo' log-likelihood function, namely,

$$g(\theta) = \left(\sum_{i=1}^{n} m_i^{(k)}\right)\ln\theta - n\ln\left(e^{\theta} - 1\right) \tag{16}$$

with respect to θ. It can be easily shown that for $\sum_{i=1}^{n} m_i^{(k)} > n$, $g(\theta)$ has a unique maximum and it can obtained as the solution of the non-linear equation

$$g'(\theta) = \left(\sum_{i=1}^{n} m_i^{(k)}\right)\frac{1}{\theta} - \frac{ne^{\theta}}{e^{\theta} - 1} = 0. \tag{17}$$

Now at the last step of the EM type algorithm using the method of Louis [14] it is possible to obtain the approximate confidence intervals of the unknown parameters. For completeness purposes, we provide the observed Fisher information matrix which can be used to obtain the approximate confidence intervals of the unknown parameters. Now, the log-likelihood of the complete data is

$$l_c(\mu, \sigma, \theta) = -n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{1}{m_i}(y_i - m_i\mu)^2 + \left(\sum_{i=1}^{n} m_i\right)\ln\theta - n\ln\left(e^{\theta} - 1\right). \tag{18}$$

The observed Fisher information matrix can be written in the following form:

$$\mathbf{I}_{obs} = \mathbf{B} - \mathbf{S}\mathbf{S}^T.$$

Here $\mathbf{B} = ((b_{ij}))$ is the Hessian matrix and $\mathbf{S} = ((s_i))$ is the gradient vector, i.e.

$$\mathbf{B} = \begin{bmatrix} \frac{\partial^2 l_c}{\partial\mu^2} & \frac{\partial^2 l_c}{\partial\mu\partial\sigma} & \frac{\partial^2 l_c}{\partial\mu\partial\theta} \\ \frac{\partial^2 l_c}{\partial\sigma\partial\mu} & \frac{\partial^2 l_c}{\partial\sigma^2} & \frac{\partial^2 l_c}{\partial\sigma\partial\theta} \\ \frac{\partial^2 l_c}{\partial\theta\partial\mu} & \frac{\partial^2 l_c}{\partial\theta\partial\sigma} & \frac{\partial^2 l_c}{\partial\theta^2} \end{bmatrix}, \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} \frac{\partial l_c}{\partial\mu} & \frac{\partial l_c}{\partial\sigma} & \frac{\partial l_c}{\partial\theta} \end{bmatrix}^T,$$

with

$$
\begin{aligned}
\frac{\partial^2 l_c}{\partial \mu^2} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n} m_i, \quad \frac{\partial^2 l_c}{\partial \mu \partial \sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{n}(y_i - m_i\mu), \quad \frac{\partial^2 l_c}{\partial \mu \partial \theta} = 0,\\
\frac{\partial^2 l_c}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}\frac{1}{m_i}(y_i - m_i\mu)^2, \quad \frac{\partial^2 l_c}{\partial \sigma \partial \theta} = 0,\\
\frac{\partial^2 l_c}{\partial \theta^2} &= -\frac{\sum_{i=1}^{n} m_i}{\theta^2} + \frac{ne^{-\theta}}{(1-e^{-\theta})^2},\\
\frac{\partial l_c}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - m_i\mu), \quad \frac{\partial l_c}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}\frac{1}{m_i}(y_i - m_i\mu)^2, \quad \frac{\partial l_c}{\partial \theta} = \frac{\sum_{i=1}^{n} m_i}{\theta} - \frac{n}{1-e^{-\theta}}.
\end{aligned}
$$

Note that once the observed Fisher information matrix is obtained, then based on the asymptotic normality results the approximate confidence intervals of the unknown parameters can be readily obtained.

3 Bivariate ZTP-N Distribution

In this section we propose a new bivariate model with zero truncated Poisson and ZTP-N as marginals. The motivation of the proposed bivariate distribution came from Barreto-Souza [6] and Kozubowski, Anorska, Podgórki [11]. The proposed bivariate distribution can be defined as follows. Suppose $\{X_1, X_2, \ldots.\}$ is a sequence of i.i.d. normal random variables with mean μ and variance σ^2. N is Poisson random variable with mean λ and it is independent of X_i's. Suppose M is a random variable defined as follows: $\{M|N = n\}$ is a binomial random variable with parameter n and p, for $n \geq 0$ and $0 < p < 1$. Consider a new bivariate random model as follows:

$$
(Z, K) \stackrel{d}{=} \{(S_M, N)|M \geq 1\}. \tag{19}
$$

Here $\stackrel{d}{=}$ means equal in distribution and

$$
S_M = \sum_{i=1}^{M} X_i.
$$

Here, N is observed, but M is not observed. We call this new distribution as the BZTP-N distribution.

Now first we provide the joint and marginal distributions of the above BZTP-N distribution. The joint cumulative distribution function of (Z, K) for $-\infty < z < \infty$ and for $k = 1, 2, \ldots$, is

$$\begin{aligned}
P(Z \leq z, K \leq k) &= P(S_M \leq z, N \leq k | M \geq 1) = \sum_{n=1}^{k} P(S_M \leq z, N = n | M \geq 1) \\
&= \frac{1}{1-e^{-\theta}} \sum_{n=1}^{k} P(S_M \leq z, N = n, M \geq 1) \\
&= \frac{1}{1-e^{-\theta}} \sum_{n=1}^{k} \sum_{m=1}^{n} P(S_M \leq z, N = n, M = m) \\
&= \frac{1}{1-e^{-\theta}} \sum_{n=1}^{k} \sum_{m=1}^{n} P(S_M \leq z | N = n, M = m) P(M = m | N = n) P(N = n) \\
&= \frac{1}{1-e^{-\theta}} \sum_{n=1}^{k} \sum_{m=1}^{n} \Phi\left(\frac{(z - m\mu)}{\sigma\sqrt{m}}\right) \binom{n}{m} p^m (1-p)^{n-m} \frac{e^{-\lambda}\lambda^n}{n!}. \qquad (20)
\end{aligned}$$

Hence, we also have for $k = 1, 2, \cdots$,

$$P(Z \leq z, K = k) = \frac{1}{1-e^{-\theta}} \sum_{m=1}^{k} \Phi\left(\frac{(z - m\mu)}{\sigma\sqrt{m}}\right) \binom{k}{m} p^m (1-p)^{k-m} \frac{e^{-\lambda}\lambda^k}{k!}. \qquad (21)$$

The marginal CDF of Z for $-\infty < z < \infty$ when $\theta = \lambda p$, can be obtained as

$$\begin{aligned}
P(Z \leq z) &= \sum_{k=1}^{\infty} P(Z \leq z, K = k) \\
&= \frac{1}{1-e^{-\theta}} \sum_{k=1}^{\infty} \sum_{m=1}^{k} \Phi\left(\frac{(z - m\mu)}{\sigma\sqrt{m}}\right) \binom{k}{m} p^m (1-p)^{k-m} \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \frac{1}{1-e^{-\theta}} \sum_{m=1}^{\infty} \sum_{k=m}^{\infty} \Phi\left(\frac{(z - m\mu)}{\sigma\sqrt{m}}\right) \binom{k}{m} p^m (1-p)^{k-m} \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \frac{1}{1-e^{-\theta}} \sum_{m=1}^{\infty} \frac{e^{-\lambda}\theta^m}{m!} \Phi\left(\frac{(z - m\mu)}{\sigma\sqrt{m}}\right) \sum_{j=0}^{\infty} \frac{\lambda^j (1-p)^j}{j!} \\
&= \frac{1}{1-e^{-\theta}} \sum_{m=1}^{\infty} \Phi\left(\frac{(z - m\mu)}{\sigma\sqrt{m}}\right) \frac{e^{-\theta}\theta^m}{m!} \\
&= \frac{1}{e^{\theta}-1} \sum_{m=1}^{\infty} \Phi\left(\frac{(z - m\mu)}{\sigma\sqrt{m}}\right) \frac{\theta^m}{m!} \qquad (22)
\end{aligned}$$

Therefore, as expected $Z \sim \text{ZTPN}(\mu, \sigma, \theta)$. Moreover, using (21) the PMF of K for $k = 1, 2, \cdots$, can be obtained as

$$\begin{aligned} P(K = k) &= \frac{1}{1 - e^{-\theta}} \sum_{m=1}^{k} \binom{k}{m} p^m (1-p)^{k-m} \frac{e^{-\lambda}\lambda^k}{k!} \\ &= \frac{e^{-\lambda}\lambda^k}{k!(1 - e^{-\theta})} \sum_{m=1}^{k} \binom{k}{m} p^m (1-p)^{k-m} \\ &= \frac{e^{-\lambda}}{k!(1 - e^{-\theta})} \left(\lambda^k - (\lambda - \theta)^k\right). \end{aligned} \tag{23}$$

It can be easily verified that $\sum_{k=1}^{\infty} P(K = k) = 1$. Note that if $\lambda = \theta$, i.e. $p = 1$, then

$$P(K = k) = \frac{e^{-\lambda}\lambda^k}{k!(1 - e^{-\lambda})},$$

a truncated Poisson with parameter λ. It may be mentioned that $p = 1$ indicates that $\{M = N | N = n\}$.

The conditional CDF of Z given K can be obtained for $-\infty < z < \infty$, as

$$\begin{aligned} P(Z \le z | K = k) &= \frac{1}{\lambda^k - (\lambda - \theta)^k} \sum_{m=1}^{k} \Phi\left(\frac{z - m\mu}{\sigma\sqrt{m}}\right) \binom{k}{m} \theta^m (\lambda - \theta)^{k-m} \\ &= \frac{1}{1 - (1-p)^k} \sum_{m=1}^{k} \Phi\left(\frac{z - m\mu}{\sigma\sqrt{m}}\right) \binom{k}{m} p^m (1-p)^{k-m} \end{aligned} \tag{24}$$

and the conditional PDF of Z given K is

$$\begin{aligned} f_{Z|K=k}(z) = \frac{d}{dz} P(Z \le z | K = k) &= \frac{1}{\lambda^k - (\lambda - \theta)^k} \sum_{m=1}^{k} \frac{1}{\sigma\sqrt{m}} \phi\left(\frac{z - m\mu}{\sigma\sqrt{m}}\right) \binom{k}{m} \theta^m (\lambda - \theta)^{k-m} \\ &= \frac{1}{1 - (1-p)^k} \sum_{m=1}^{k} \frac{1}{\sigma\sqrt{m}} \phi\left(\frac{z - m\mu}{\sigma\sqrt{m}}\right) \binom{k}{m} p^m (1-p)^{k-m}. \end{aligned} \tag{25}$$

From (24) it easily follows that the conditional distribution of $Z | K = k$ has the following representation

$$\{Z | K = k\} \stackrel{d}{=} \sum_{i=1}^{M} V_i, \tag{26}$$

here $V_1, \ldots, V_k$ are i.i.d. $\mathrm{N}(\mu, \sigma^2)$ and M is a binomial random variable with parameters k and p and it is truncated at 0. Moreover, Z_i's and M are independently distributed.

Using the above (26) representation, the joint MGF of Z and K for $-\infty < t, s < \infty$ can be obtained for

$$c = \left(1 - p + pe^{\mu s + \frac{\sigma^2 s^2}{2}}\right),$$

as follows:

$$\begin{aligned}
\varphi_{Z,K}(s,t) &= E_{Z,K} e^{sZ+tK} = E_K E_{Z|K} e^{sZ+tK} \\
&= E_K \left\{ \frac{e^{tK}}{1-(1-p)^K} \left(c^K - (1-p)^K\right) \right\} \\
&= \frac{e^{-\lambda}}{1-e^{\theta}} \sum_{k=1}^{\infty} \left\{ \frac{\lambda^k}{1-(1-p)^k} \left(c^k - (1-p)^k\right) \right\} \frac{1}{k!}.
\end{aligned}$$

Although, the joint MGF of Z and K cannot be obtained in closed form, the MGF of K can be obtained quite conveniently for $-\infty < t < \infty$, and it is

$$M_K(t) = \frac{e^{-\lambda}}{1-e^{-\theta}} \left[e^{\lambda e^t} - e^{(\lambda-\theta)e^t}\right]. \tag{27}$$

Hence, the first two moments of K can be obtained as

$$\begin{aligned}
E(K) &= \lambda + \frac{\theta}{e^{\theta}-1} \\
E(K^2) &= \lambda + \lambda^2 + \frac{\theta(1+2\lambda-\theta)}{e^{\theta}-1}.
\end{aligned}$$

Using the ratio

$$\frac{P(K=k+1)}{P(K=k)} = \frac{\lambda}{k+1} \left[\frac{1-(1-p)^{k+1}}{1-(1-p)^k}\right],$$

it can be easily obtained that the PMF of K will be always unimodal. If $\lambda p < 2$, the mode will be at 1, otherwise the mode will be at $k_0 > 1$.

3.1 Maximum Likelihood Estimators

In this section we discuss about the estimation procedure of the unknown parameters based on a random sample $\{(z_1, k_1), \ldots, (z_n, k_n)\}$ from a BZTP-N$(\mu, \sigma, \theta, \lambda)$. The joint log-likelihood function can be written as

$$l(\mu, \sigma, \lambda, p) = \sum_{i=1}^{n} \ln \left[f_{Z|K=k_i}(z_i) P(K=k_i)\right], \tag{28}$$

where $f_{Z|K=k_i}(z_i)$ and $P(K = k_i)$ can be obtained (23) and (25), respectively. The MLEs of the unknown parameters can be obtained by maximizing the right hand side of (28) with respect to the unknown parameters. They have to be obtained by solving a four dimensional optimization problem.

In this case also we propose to use EM type algorithm to avoid solving a four dimensional optimization problem. The basic idea of the proposed EM type algorithm comes from the following observations. Suppose $\{(z_1, k_1, m_1), \ldots, (z_n, k_n, m_n)\}$ is a random sample of size n from (Z, K, M). Here Z, K and M are same as defined before. Note that in this case

$$\{Z|K = k, M = m\} \sim \mathrm{N}(m\mu, m\sigma^2), \tag{29}$$

$$P(M = m|K = k) = \frac{1}{1-(1-p)^k}\binom{k}{m}p^m(1-p)^{k-m}; \quad m = 1, \ldots, k, \tag{30}$$

$$P(K = k) = \frac{1}{1-e^{-\lambda}}\frac{e^{-\lambda}\lambda^k}{k!}; \quad k = 1, 2, \ldots. \tag{31}$$

Therefore, in practice we only observe $\{(z_1, k_1), \ldots, (z_n, k_n)\}$, and $m_1, \ldots, m_n$ are missing. We propose to use the EM type algorithm to compute the MLEs of the unknown parameters. Suppose $\Theta^{(k)} = (\mu^{(k)}, \sigma^{(k)}, \lambda^{(k)}, p^{(k)})$ denotes the estimate of the unknown parameter vector and $m_i^{(k)}$ denotes the missing value of m_i for $i = 1, \ldots, n$, at the k-th stage of the EM type algorithm. Then at the 'E'-step of the EM type algorithm, the 'pseudo' log-likelihood function at the k-th iterate can be written without the additive constant as

$$\begin{aligned} l_{pseudo}(\Theta|\Theta^{(k)}) = & -n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^n \frac{1}{m_i^{(k)}}(z_i - m_i^{(k)}\mu)^2 + \left(\sum_{i=1}^n m_i^{(k)}\right)\ln p - n\ln(1-e^{-\lambda}) + \\ & \left(\sum_{i=1}^n (k_i - m_i^{(k)})\right)\ln(1-p) - \sum_{i=1}^n \ln(1-(1-p)^{k_i}) - n\lambda + \left(\sum_{i=1}^n k_i\right)\ln\lambda. \end{aligned} \tag{32}$$

Here

$$m_i^{(k)} = \arg\max_m P(M = m|K = k_i, p^{(k)}),$$

same as before. Now at the 'M'-step of the EM type algorithm, we need to maximize the 'pseudo' log-likelihood function to obtain $\Theta^{(k+1)}$. Note that $\mu^{(k+1)}$ and $\sigma^{(k+1)}$ can be obtained same as in (15) with y_i's are replaced by z_i's. Moreover, $p^{(k+1)}$ can be obtained as the solution of the non-linear equation

$$\left(\sum_{i=1}^n m_i^{(k)}\right)\frac{1}{p} - \left(\sum_{i=1}^n (k_i - m_i^{(k)})\right)\frac{1}{1-p} - \sum_{i=1}^n \frac{k_i(1-p)^{k_i-1}}{1-(1-p)^{k_i}} = 0, \tag{33}$$

and $\lambda^{(k+1)}$ can be obtained as the solution of

$$-\frac{ne^{-\lambda}}{1-e^{-\lambda}} + \frac{1}{\lambda}\left(\sum_{i=1}^{n} k_i\right) - n = 0. \tag{34}$$

In this case also at the last step of the EM type algorithm using the method of Louis [14] we can obtain approximate confidence intervals of the unknown parameters. Below we provide the observed Fisher information matrix which can be used to obtain the approximate confidence intervals of the unknown parameters. The observed Fisher information matrix can be written in the following form:

$$\mathbf{I}_{obs} = \mathbf{B} - \mathbf{S}\mathbf{S}^T.$$

Here $\mathbf{B} = ((b_{ij}))$ is the Hessian matrix and $\mathbf{S} = ((s_i))$ is the gradient vector of the 'pseudo' log-likelihood function, i.e.

$$\mathbf{B} = \begin{bmatrix} \frac{\partial^2 l_c}{\partial \mu^2} & \frac{\partial^2 l_c}{\partial \mu \partial \sigma} & \frac{\partial^2 l_c}{\partial \mu \partial \lambda} & \frac{\partial^2 l_c}{\partial \mu \partial p} \\ \frac{\partial^2 l_c}{\partial \sigma \partial \mu} & \frac{\partial^2 l_c}{\partial \sigma^2} & \frac{\partial^2 l_c}{\partial \sigma \partial \lambda} & \frac{\partial^2 l_c}{\partial \sigma \partial p} \\ \frac{\partial^2 l_c}{\partial \lambda \partial \mu} & \frac{\partial^2 l_c}{\partial \lambda \partial \sigma} & \frac{\partial^2 l_c}{\partial \lambda^2} & \frac{\partial^2 l_c}{\partial \lambda \partial p} \\ \frac{\partial^2 l_c}{\partial p \partial \mu} & \frac{\partial^2 l_c}{\partial p \partial \sigma} & \frac{\partial^2 l_c}{\partial p \partial \lambda} & \frac{\partial^2 l_c}{\partial p^2} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \frac{\partial l_c}{\partial \mu} & \frac{\partial l_c}{\partial \sigma} & \frac{\partial l_c}{\partial \lambda} & \frac{\partial l_c}{\partial p} \end{bmatrix}^T,$$

$$\begin{aligned} l_c(\mu,\sigma,\lambda,p) \;=\; & -n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{1}{m_i}(z_i - m_i\mu)^2 + \left(\sum_{i=1}^{n} m_i\right)\ln p - n\ln(1-e^{-\lambda}) + \\ & \left(\sum_{i=1}^{n}(k_i - m_i)\right)\ln(1-p) - \sum_{i=1}^{n}\ln(1-(1-p)^{k_i}) - n\lambda + \left(\sum_{i=1}^{n} k_i\right)\ln\lambda. \end{aligned} \tag{35}$$

and

$$\begin{aligned} \frac{\partial^2 l_c}{\partial \mu^2} &= -\frac{1}{\sigma^2}\sum_{i=1}^{n} m_i, \quad \frac{\partial^2 l_c}{\partial \mu \partial \sigma} = -\frac{2}{\sigma^3}\sum_{i=1}^{n}(y_i - m_i\mu), \quad \frac{\partial^2 l_c}{\partial \mu \partial \lambda} = 0, \; \frac{\partial^2 l_c}{\partial \mu \partial p} = 0, \\ \frac{\partial^2 l_c}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}\frac{1}{m_i}(y_i - m_i\mu)^2, \quad \frac{\partial^2 l_c}{\partial \sigma \partial \lambda} = 0, \; \frac{\partial^2 l_c}{\partial \sigma \partial p} = 0, \\ \frac{\partial^2 l_c}{\partial \lambda^2} &= -\frac{\sum_{i=1}^{n} k_i}{\lambda^2} + \frac{ne^{-\lambda}}{(1-e^{-\lambda})^2}, \end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2 l_c}{\partial p^2} &= -\frac{\sum_{i=1}^n m_i}{p^2} - \frac{\sum_{i=1}^n (k_i - m_i)}{(1-p)^2} + \sum_{i=1}^n \frac{k_i\left[(k_i-1)(1-p)^{k_i-2} + (1-p)^{2k_i-2}\right]}{[1-(1-p)^{k_i}]^2}, \\
\frac{\partial l_c}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^n (y_i - m_i\mu), \quad \frac{\partial l_c}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n \frac{1}{m_i}(y_i - m_i\mu)^2, \\
\frac{\partial l_c}{\partial \lambda} &= \frac{-ne^{-\lambda}}{1-e^{-\lambda}} - n + \frac{\sum_{i=1}^n k_i}{\lambda}, \\
\frac{\partial l_c}{\partial p} &= \frac{\sum_{i=1}^n m_i}{p} - \frac{\sum_{i=1}^n (k_i - m_i)}{1-p} - \sum_{i=1}^n \frac{k_i(1-p)^{k_i-1}}{1-(1-p)^{k_i}}.
\end{aligned}
$$

From the above observed Fisher information matrix, the asymptotic confidence intervals of the unknown parameters can be obtained.

4 Simulation Results

In this section we present some simulation results for different sample sizes and for different parameter values for the univariate ZTP-N and BZTP-N models to see how the proposed methods work in practice.

4.1 ZTP-N Model

In this section we present some simulation results for the ZTP-N model. In this case for a given set of parameter values and for a given sample size, we first generate a random sample of size n from a ZTP-N model. Now for the given generated sample we obtain the maximum likelihood estimates of the unknown parameters based on the proposed EM type algorithm. In each case we stop the EM type algorithm, if the difference between the two successive estimates is less than 10^{-4}, for all the three estimates. In all the cases the EM type algorithm converges within twenty iterations. We obtain the average estimates and the mean squared errors based on 1000 replications. The results are recorded in Tables 1 and 2. In both the cases we have kept σ and λ fixed, and we have changed μ.

From the tables it is clear in all the cases as the sample sizes increase the biases and MSEs decrease. It seems the proposed EM type algorithm works quite well. It is observed that as μ changes from zero to some non-zero number, the average estimates and the MSEs

Table 1: The average estimates and the mean squared errors (MSEs) of the MLEs for the ZTP-N model. In this case $\mu = 0.0$, $\sigma^2 = 1.0$ and $\lambda = 1.0$

n	μ	σ^2	λ
25	0.4123	1.6451	1.5481
	(0.1115)	(0.2654)	(0.1565)
50	0.1132	1.2173	1.1324
	(0.0574)	(0.1143)	(0.0784)
75	0.0254	1.0149	1.0715
	(0.0413)	(0.0845)	(0.0539)
100	0.0117	1.0041	1.0156
	(0.0312)	(0.0665)	(0.0401)

Table 2: The average estimates and the mean squared errors (MSEs) of the MLEs for the ZTP-N model. In this case $\mu = 1.0$, $\sigma^2 = 1.0$ and $\lambda = 1.0$

n	μ	σ^2	λ
25	1.5453	1.7216	1.7143
	(0.2541)	(0.3289)	(0.2754)
50	1.2143	1.3549	1.1468
	(0.1352)	(0.1641)	(0.1359)
75	1.1150	1.1248	1.0879
	(0.0812)	(0.1109)	(0.0978)
100	1.0275	1.0143	1.0287
	(0.0558)	(0.0875)	(0.0669)

increase for all the estimates.

4.2 BZTP-N Model

In this section we present some simulation results for the BZTP-N model. In this case also for a given set of parameter values we have generated a random bivariate sample from the BZTP-N model and for the given generated data set we compute the MLEs based on the proposed EM type algorithm. We have used the same stopping criterion as the previous model. For the BZTP-N model, the EM type algorithm in all the cases converges within twenty five iterations. In each case we compute the average estimates and the MSEs based on 1000 replications. The results are reported in Tables 3 - 4

Table 3: The average estimates and the mean squared errors (MSEs) of the MLEs for the BZTP-N model. In this case $\mu = 0.0$, $\sigma^2 = 1.0$, $\lambda = 1.0$, $p = 0.75$

n	μ	σ^2	λ	p
25	0.4378	1.5712	1.6117	0.8212
	(0.1356)	(0.2736)	(0.1693)	(0.0511)
50	0.1872	1.1423	1.1311	0.7914
	(0.0664)	(0.1318)	(0.0887)	(0.0267)
75	0.0187	1.0278	1.0659	0.7638
	(0.0487)	(0.0921)	(0.0569)	(0.0179)
100	0.0104	1.0031	1.0122	0.7535
	(0.0323)	(0.0614)	(0.0399)	(0.0154)

Table 4: The average estimates and the mean squared errors (MSEs) of the MLEs for the BZTP-N model. In this case $\mu = 1.0$, $\sigma^2 = 1.0$, $\lambda = 1.0$ and $p = 0.75$

n	μ	σ^2	λ	p
25	1.6133	1.7732	1.7578	0.8278
	(0.2466)	(0.3378)	(0.2987)	(0.0589)
50	1.2217	1.3443	1.1398	0.7878
	(0.1278)	(0.1715)	(0.1469)	(0.0295)
75	1.1156	1.1143	1.0781	0.7661
	(0.0823)	(0.1145)	(0.0999)	(0.0194)
100	1.0145	1.0121	1.0217	0.7512
	(0.0558)	(0.0875)	(0.0669)	(0.0142)

In this case also it is observed that the proposed EM type algorithm works quite well. The biases and the MSEs decrease as the sample size increases. It verifies the consistency property of the MLEs.

5 Data Analysis

In this section we present the analysis of a data set to see the effectiveness of the proposed models. The data set represents the third party motor insurance for 1977 for one of seven geographical zones. The data represent the total amount paid by the insurance company and the total number claims received by the insurance company. The exact number of claims against which the payment has been made is not available. The data set is available in Andrews and Herzberg [1]. The amount here is in Swedish kroner (Skr). All the data points are divided by 10000 for computational purposes. It is not going to affect the inference procedure. The data set is presented in Table 5.

Table 5: Aggregate payments by third party motor insurance for Sweden in 1977

Z	20.871	48.767	6.560	50.908	4.399	14.788	48.713	52.076	13.161
K	5	7	2	6	3	6	9	9	3
Z	27.919	103.910	38.065	14.620	40.258	20.011	4.496	12.268	6.279
K	7	7	4	6	5	3	3	4	3
Z	39.939	47.551	47.589	40.634	50.067	55.603	2.802	15.193	11.839
K	3	5	7	3	6	8	2	4	4
Z	77.588	14.084	2.059						
K	7	6	5						

Before progressing further first we provide some basic statistics of the observed value of Z, it is presented in Table 6

Table 6: Basic statistics of the payments

Mean	Stand. Dev.	Median	Q_1	Q_3
31.100	24.062	23.965	12.053	48.739

It is clear from the basic statistics of the payment data that the payment data is skewed. We try to fit BZTP-N distribution to the bivariate data set. We have used the EM type algorithm to compute the MLEs of the unknown parameters. We obtain the following estimates of μ, λ, σ and p and the associated 95% confidence intervals. The corresponding log-likelihood (ll) value also has been reported below:

$$\begin{aligned} \widehat{\mu} &= 10.971(\mp 2.413), \quad \widehat{\lambda} = 4.861(\mp 0.761), \quad \widehat{\sigma} = 8.039(\mp 1.786), \quad \widehat{p} = 0.541(\mp 0.051) \\ \text{ll} &= -127.671 \end{aligned}$$

We fit ZTP-N distribution to the univariate payment data set. We have used the EM type algorithm to compute the MLEs of the unknown parameters. We obtain the following estimates of μ, θ and σ, and the associated 95% confidence intervals. The corresponding log-likelihood (ll) value also has been reported below:

$$\widehat{\mu} = 11.491(\mp 2.679), \quad \widehat{\theta} = 2.468(\mp 0.479), \quad \widehat{\sigma} = 8.481(\mp 1.878), \quad \text{ll} = -134.409.$$

To verify whether the proposed EM type algorithm actually converges to the MLEs, we have performed extensive grid search method with a grid size 0.001, and it is observed that they match. Which ensures that the proposed EM type algorithm actually converges to the MLEs.

We fit the geometric skew normal (GSN) distribution to the univariate payment data set (see, Kundu (2014)). We have used the EM type algorithm to compute the MLEs of the unknown parameters. We obtain the following estimates of μ, p and σ, and the associated 95% confidence intervals. The corresponding log-likelihood (ll) value also has been reported below:

$$\widehat{\mu} = 8.116(\mp 1.213), \quad \widehat{p} = 0.738(\mp 0.024), \quad \widehat{\sigma} = 4.873(\mp 0.656), \quad \text{ll} = -133.514$$

We fit normal distribution to the univariate payment data set. We obtain the following estimates of μ, and σ, and the associated 95% confidence intervals. The corresponding log-likelihood (ll) value also has been reported below:

$$\widehat{\mu} = 31.100(\mp 8.610), \quad \widehat{\sigma} = 24.062(\mp 5.478), \quad \text{ll} = -137.909. \tag{36}$$

We have also reported the Kolmogorov-Smirnov (K-S) distances of the empirical distribution function and the fitted distribution functions, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for each case in Table 7. Based on the K-S distances and the associated p values it is clear that all the four models provide good fit the data set. Now comparing the AIC and BIC values we can conclude that BZTP-N distribution provides the best fit among these four models. In Figure 2 we have plotted the empirical survival function (ESF) and fitted survival functions for different models.

Now for the BZTP-N model we want to test the hypothesis

$$H_0 : p = 1, \quad vs. \quad H_1 : p < 1.$$

Table 7: The goodness of fit tests for the motor insurance data set

Model	K-S	p-value	AIC	BIC
BZTP-N	0.1236	0.7435	263.342	268.947
ZTP-N	0.1283	0.7067	274.818	279.022
GSN	0.1289	0.7006	273.028	277.232
Normal	0.1790	0.2912	279.818	282.620

Under the null hypothesis, the MLEs of the unknown parameters, and the corresponding log-likelihood value are provided below.

$$\widetilde{\mu} = 16.0887, \quad \widetilde{\sigma} = 6.1383, \quad \widetilde{\lambda} = 5.0335, \quad l_{H_0} = -132.121.$$

Therefore, the test statistic in this case becomes $-2(l_{H_0 \cup H_1} - l_{H_0}) = 8.9$. Since p-value $< 10^{-6}$, we reject the null hypothesis.

6 Conclusions

In this paper first we have proposed a new three-parameter model based on a random Poisson sum of normal random variables. It is observed that the proposed model has a very flexible PDFs. We have derived different properties of the model and proposed the inference procedures of the unknown parameters. We have further generalized it two dimensions with four parameters. It is observed that the proposed model has an interesting physical interpretations. We have proposed inference procedure of the bivariate model also. Based on the simulation experiments, it is observed that the EM type algorithm work quite well in practice. Further, we have analyzed one real life insurance data set and shown that the BZTP-N model is the best fitting model when compared to ZTP-N, GSN and normal models.

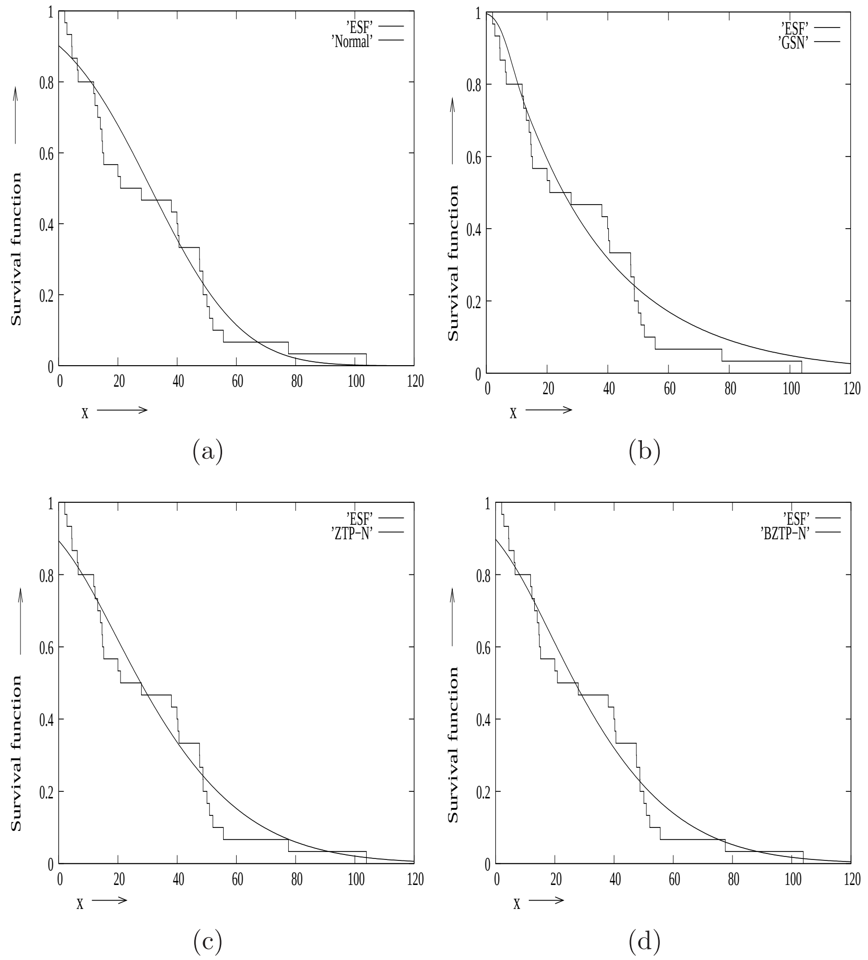

Figure 2: Empirical survival functions (ESF) and the fitted survival functions for different distributions: (a) Normal (b) GSN (c) ZTN-P (d) BZTN-P

Acknowledgements:

The authors would like to thank unknown reviewers for their constructive suggestions which have helped us to improve the manuscript. Part of the work of the second author has been funded by a research grant from the Science and Engineering Research Board, Government of India.

Appendix: Pseudo Codes

In this Appendix we have provided the pseudo codes of computing the EM type algorithm for ZTP-N and BZTP-N models.

ZTP-N Model: Pseudo Code of the EM Algorithm

- Set J, an integer, denoting the maximum number of iterations allowed. It may be chosen 50. Set M, an integer, denoting the upperbound where the mode of a ZTP-N model can occur. It may be chosen 20. Set ϵ, the tolerance limit, i.e. if the absolute difference between the two successive log-likelihood values is less than ϵ, then we stop the iteration. It may be chosen $\epsilon = 10^{-5}$. Set sample size n.

- Read data vector $\{y_1, \ldots, y_n\}$.

- Set initial values of μ, σ and θ, say $\mu^{(1)}$, $\sigma^{(1)}$ and $\theta^{(1)}$.

- DO k = 1:J

  DO i = 1:n

  Find $m_i^{(k)} = \min\{m\} \in \{1, 2, \ldots\}$, such that $\frac{P(M^*=m+1|y_i,\mu^{(k)},\sigma^{(k)},\theta^{(k)})}{P(M^*=m|y_i,\mu^{(k)},\sigma^{(k)},\theta^{(k)})} < 1$

  ENDDO

  Compute $\mu^{(k+1)} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i^{(k)}}, \quad \sigma^{(k+1)} = \sqrt{\frac{1}{n}\sum_{i=1}^n \frac{1}{m_i^{(k)}}(y_i - m_i^{(k)}\mu^{(k)})^2}$ and $\theta^{(k+1)}$ as a solution of (17), by using bi-section method.

  Compute the log-likelihood function $l(\mu^{(k+1)}, \sigma^{(k+1)}, \theta^{(k+1)})$.

  If $|l(\mu^{(k+1)}, \sigma^{(k+1)}, \theta^{(k+1)}) - l(\mu^{(k)}, \sigma^{(k)}, \theta^{(k)})| < \epsilon$, stop the iteration, otherwise CONTINUE.

- ENDDO

Pseudo code of finding the maximum of $g(\theta)$, using the bi-section Method

- Set ϵ, the tolerance limit, to indicate when to stop the iteration. We may choose $\epsilon = 10^{-5}$. Set N a large integer indicating the maximum number of iterations allowed. We may choose $N = 1000$.
- Choose θ_1 and θ_2, such that $g'(\theta_1) > 0$ and $g'(\theta_2) < 0$.
- DO ITER: = 1, N

  If $\theta_1 - \theta_2| < \epsilon$, stop the iteration.

  If $g'((\theta_1 + \theta_2)/2) > 0$, then set $\theta_1 := (\theta_1 + \theta_2)/2$, otherwise set $\theta_2 := (\theta_1 + \theta_2)/2$.
- ENDDO

BZTP-N Model: Pseudo Code of the EM Algorithm

- Set J, an integer, denoting the maximum number of iterations allowed. It may be chosen 50. Set L, an integer, denoting the upperbound where the mode of a BZTP-N model can occur. It may be chosen 20. Set ϵ, the tolerance limit, i.e. if the absolute difference between the two successive log-likelihood values is less than ϵ, then we stop the iteration. It may be chosen $\epsilon = 10^{-5}$. Set sample size n.
- Read data vector $\{(z_1, k_1), \ldots, (z_n, k_n)\}$.
- Set initial values of μ, σ, λ and p, say $\mu^{(1)}$, $\sigma^{(1)}$, $\lambda^{(1)}$ and $p^{(1)}$.
- DO k = 1:J

  DO i = 1:n

  Find $m_i^{(k)} = \min\{m\} \in \{1, 2, \ldots\}$, such that $\frac{P(M=m+1|K=k_i,p^{(k)})}{P(M=m|K=k_i,p^{(k)})} < 1$

  ENDDO

  Compute $\mu^{(k+1)} = \frac{\sum_{i=1}^n z_i}{\sum_{i=1}^n m_i^{(k)}}$, $\quad \sigma^{(k+1)} = \sqrt{\frac{1}{n}\sum_{i=1}^n \frac{1}{m_i^{(k)}}(z_i - m_i^{(k)}\mu^{(k)})^2}$ and $p^{(k+1)}$ as a solution of (33), by using bi-section method.

$\lambda^{(k+1)}$ as a solution of (34), by using bi-section method.

Compute the log-likelihood function $l(\mu^{(k+1)}, \sigma^{(k+1)}, p^{(k+1)}, \lambda^{(k+1)})$.

If $|l(\mu^{(k+1)}, \sigma^{(k+1)}, p^{(k+1)}, \lambda^{(k+1)}) - l(\mu^{(k)}, \sigma^{(k)}, p^{(k)}, \lambda^{(k)})| < \epsilon$, stop the iteration, otherwise CONTINUE.

- ENDDO

References

[1] Andrews, D. F. and Herzberg, A. M. 1985. *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer-Verlag, New York.

[2] Asgharzadeh, A., Esmaily, L., and Nadarajah, S. 2013. Approximate MLEs for the location and scale parameters of the skew logistic distribution. *Statistical Papers* 54(2): 391–411.

[3] Asgharzadeh, A., Esmaily, L., and Nadarajah, S. 2016. Balakrishnan skew logistic distribution. *Communications in Statistics- Theory and Methods* 45(2): 444–464.

[4] Azzalini, A. A. 1985. A Class of distributions which include the normal. *Scandinavian Journal of Statistics* 12: 171–178.

[5] Azzalini, A. 2005. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* 32: 159-200.

[6] Barreto-Souza, W. 2012. The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics* 109, 130-145.

[7] Gupta, R.C. and Gupta, R.D. 2008. Analyzing skewed data by power normal model. *Test* 17: 197 - 210.

[8] Jamalizadeh, A., Behboodian, J., Balakrishnan, N. 2008. A two-parameter generalized skew-normal distribution. *Statistics & Probabability Letters* 78: 1722-1726.

[9] Johnson, N. L., Kotz, S. and Balakrishnan, N. 1995. *Continuous Univariate Distribution*, Vol. 1, John Wiely & Sons, New York, USA.

[10] Kazemi, R. and Noorizadeh, M. 2015. A comparison between skew-logistic and skew-normal distributions. *Mathematika* 31: 15 - 24.

[11] Kozubowski, T.J., Anorska, A.K. and Podgórki, K. 2008, "A bivariate Lévy process with negative binomial and gamma marginals", *Journal of Multivariate Analysis*, vol. 99, 1418 – 1437.

[12] Kundu, D. 2014. Geometric skew normal distribution. *Sankhya B* 76(2): 167 - 189.

[13] Kundu, D. and Nekoukhou, V. (2018), "Univariate and bivariate geometric discrete generalized exponential distribution", *Journal of Statistical Theory and Practice*, vol. 12, no. 3, 595 – 614.

[14] Louis, T. A. 1982. Finding the observed information matrix when uisng the EM algorithm. *Journal of the Royal Statistical Society*, Ser. B, Vol. 44, 226–233.

[15] Pewsey, A. 2000. Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics* 27: 859-870.

[16] Sharafi, M., and Behboodian, J. 2008. The Balakrishnan skew-normal density. *Statistical Papers* 49: 769-778.

[17] Willmot, G. E. and Lin, X. S. 2001. *Lundberg Approximations for Compound Distributions with Insurance Applications*, Lecture Notes in Statistics, V. 156, Springer, New York.