Volume 66, Issue 5, 2022

Journal of Scientific Research

of

The Banaras Hindu University

A Bayes Estimation for the Parameters of Misclassified Size-biased Poisson Lindley Distribution and its Application

B. S. Trivedi*[1], M. N. Patel[2], D. R. Barot[3]

*[1]Amrut Mody School of Management, Ahmedabad University, India, *E-mail:* bhaktida.trivedi@ahduni.edu.in
[2]Department of Statistics, School of Sciences, Gujarat University, Ahmedabad, *E-mail:* mnpatel.stat@gmail.com
[3]Amrut Mody School of Management, Ahmedabad University, India, *E-mail:* dinesh.barot@ahduni.edu.in

Abstract: **This paper introduces a misclassified size-biased Poisson-Lindley distribution by compounding the size-biased Poisson distribution with the size-biased Lindley distribution. The Bayes estimation for the misclassified size-biased Poisson-Lindley distribution is investigated and compared with the Maximum Likelihood estimation. A real dataset is discussed to demonstrate the suitability and applicability of the proposed distribution in the modeling count dataset. Finally, a Monte Carlo simulation study of 100000 simulated data is presented to investigate and compare the Bayes estimators and Maximum Likelihood estimators in terms of simulated risk for different sample sizes and varying parameters value.**

Keywords: **Poisson-Lindley distribution, Misclassification, Size biased, Maximum likelihood estimation, Bayes estimation, Simulation.**

MSC 2010: **62E15, 62F10, 60E05**

I. INTRODUCTION

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and /or space if these events occur with a known average rate and are independent of time since the last events. So the Poisson distribution can be applied to systems with a large number of possible events, each of which is rare in probability theory and statistics. Ladislaus Bortkiewicz (1898) used this distribution when he was tasked with investigating the number of soldiers in the Prussian army killed accidentally by horse kicks. This experiment introduced the Poisson distribution to the field of reliability engineering.

Shanker et al. (2013) introduced the Lindley distribution with two parameters by considering the survival and waiting time data. Ghitany et al. (2008) compared two models and showed that the Lindley distribution provides an effective model than the exponential distribution. Elbatal et al. (2013) proposed that Lindley distribution is a mixture of gamma and exponential distribution. Shanker et al. (2013) compared one parameter Lindley distribution with two-parameter Lindley distribution. Mervoci and Sharma (2014) extended the Lindley distribution called the beta Lindley distribution. At the same time, Singh et al. (2014) gave truncated Lindley distribution. Shanker *et al.* (2015) have done a critical and comparative study on applications of Lindley and exponential distributions for modeling lifetime data from biological sciences and engineering and observed that there are many lifetime data where exponential distribution gives a better fit than Lindley distribution.

Borah and DekaNath (2001) enlarged a Poisson Lindley distribution (PLD) with a further study called inflated Poisson Lindley distribution. Whereas Ghitany et al. (2008) examined the PLD to model count data, as well as Ghitany et al. (2008), aim their study for data does not include zero counts since Zakerzadeh and Dolati (2009) described a generalized form of Lindley distribution with three parameters. Therefore, Ghitany et al. (2011) worked on modeling survival data and introduced a Lindley distribution with two parameters called weighted Lindley distribution. However, Lord and Geedipally (2011) proposed a new distribution called negative binomial Lindley, which contains two parameters for crash count data. Ghitany*et al.* (2008) discussed statistical properties including moments based coefficients, hazard rate function, mean residual life function, mean deviations, stochastic ordering, Renyi entropy measure, order statistics, Bonferroni and Lorenz curves, stress-strength reliability, along with the estimation of parameter and application to model waiting for time data in a bank. A detailed and critical study on applications of PLD for count data relating to biological sciences has been done by Shanker and Hagos (2015), and found that PLD gives a much closer fit than Poisson distribution.

Fisher (1934) introduced Size-biased distribution to model ascertainment bias. It is a particular case of the more general form known as weighted distribution. Weighted distributions were later formalized in a unifying theory by Rao (1965). Such distributions arise naturally in practice when observations from a sample are recorded with unequal probability, such as from probability proportional to size (PPS) designs.

Shanker*et al.* (2015) have studied a size-biased Poisson Lindley distribution (SBPLD) and its applications in detail to model data relating to thunderstorms and found that the SBPLD is a suitable model for thunderstorms data. This distribution has wide applications in the theory of accident proneness. It arises from the Poisson distribution when its parameter follows a continuous Lindley distribution. Sankaran (1970) investigated this distribution with application to errors and accident data. Sankaran (1970) has obtained its moments and discussed some of the statistical properties, estimation of parameters, and applications to model count data.

The primary motivation of this paper is to study misclassification in Size Biased Poisson Lindley distribution. Bayes estimation method is applied to estimate the parameters of the distribution. A simulation study is carried out to study the effectiveness of the estimation method.

II. Model Description

A discrete random variable X is said to have a Poisson distribution with parameter $\lambda > 0$, if for $k = 0, 1, 2, \ldots.$, the probability mass function of X is given by

$$f(k;\lambda) = P_r(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (1)$$

The positive real number λ is equal to the expected value of X and also to its variance, i.e.$\lambda = E(X) = Var(X)$

A one-parameter Lindley distribution with parameter θ is defined by its probability density function given as

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

If the random variable X has distribution, $f\,(x;\,\theta)$, with unknown parameters θ, then the corresponding weighted distribution is of the form

$$f^w(x\,;\,\theta) = \frac{w(x)\,f(x\,;\,\theta)}{E(w(x))} \qquad (3)$$

Where $w(x)$ is a non–negative weight function such that $E(w(x))$exists.

By taking weight $w(x) = x$, the size–biased distribution can be determined with pmf.

$$f^*(x\,;\,\theta) = \frac{x\,f(x;\theta)}{E(x)} \qquad (4)$$

Ghitany and Al-Muttairi (2008) investigated and showed that the size-biased Poisson Lindley distribution also arises from the size-biased Poisson distribution with pmf:

$$g(x|\lambda) = e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!} \quad , x = 1, 2, 3 \ldots\ldots\ \lambda > 0 \qquad (5)$$

When its parameter λ follows a size–biased Lindley model with pdf.

$$h(\lambda\,;\,\theta) = \frac{\theta^3}{(\theta+2)}\lambda(1+\lambda)e^{-\theta\lambda}, \lambda > 0, \theta > 0 \qquad (6)$$

A size-biased Poisson Lindley distribution (SBPLD), introduced by Ghitany and Al- Mutairi (2008), having parameter θ is defined by its probability mass function (pmf)

$$P(x) = p(x;\theta) = \frac{x\theta^3(x+\theta+2)}{(\theta+2)(\theta+1)^{x+2}} \qquad (7)$$

$$x = 1, 2, 3, \ldots, \theta > 0$$

Where the mean of the Poisson Lindley distribution, $\mu = \frac{\theta+2}{\theta(1+\theta)}$ and θ is known as a parameter of the distribution. It would be noted that the SBPLD is a simple size-biased version of Poisson-Lindley distribution (PLD) having pmf

$$f_0(x;\theta) = \frac{\theta^2(x+\theta+2)}{(\theta+1)^{x+3}}, x = 0, 1, 2, \ldots and\ \ \theta > 0$$

and

$$\sum_{x=1}^{\infty} P_x(\theta) = \frac{\theta^3}{(\theta+2)(1+\theta)^2}\sum_{x=1}^{\infty}\frac{x(x+\theta+2)}{(1+\theta)^x} = 1$$

If the random variables follow the Poisson probability law, then the n problem of misclassification may arise where specific counts $(c+1)$ are sometimes reported as the count c. Cohen (1960) studied this situation in a general way called the misclassified Poisson distribution.

Suppose the number of defects in a unit during the production of the units follows a Poisson Lindley distribution with parameter θ and let α be the probability that the unit which contains $(c+1)$ defects is misclassified by reporting it as having only c defects. In all other cases, the observations and reporting of defects are found correct.

Let X denote the number of defects reported in a produced unit. Then

for $X = c$,

$$p(c;\theta,\alpha) = p(c) + \alpha\,p(c+1)$$
$$= \frac{\theta^3}{(\theta+2)(\theta+1)^{c+3}}[c(c+\theta+2)(\theta+1) + \alpha(c+1)(c+\theta+3)]; \qquad (8)$$

for $X = c+1$,

$$p(c+1;\theta,\alpha) = p(c+1) - \alpha\,p(c+1)$$

$$= (1-\alpha)(c+1)\left[\frac{\theta^3(c+\theta+3)}{(\theta+2)(\theta+1)^{c+3}}\right]; \quad (9)$$

and

for $X \in S$,

$$p(x;\theta,\alpha) = p(x) = \frac{x\theta^3(x+\theta+2)}{(\theta+2)(\theta+1)^{x+2}}, \quad \theta>0, \quad 0<\alpha<1 \quad (10)$$

Thus, from Eq. (8), Eq. (9), and Eq. (10), we get the pmf of Misclassified Size-Biased Poisson Lindley (MSBPL) distribution of the random variable x as

$$p(x;\theta,\alpha) = \begin{cases} \frac{\theta^3}{(\theta+2)(\theta+1)^{c+3}}[c(c+\theta+2)(\theta+1)+\alpha(c+1)(c+\theta+3)], & for\ x=c \\ (1-\alpha)(c+1)\left[\frac{\theta^3(c+\theta+3)}{(\theta+2)(\theta+1)^{c+3}}\right], & for\ x=c+1 \\ \frac{x\theta^3(x+\theta+2)}{(\theta+2)(\theta+1)^{x+2}}, & for\ x\in S \end{cases} \quad (11)$$

Where S is a subset of the set I of non–negative integers, not containing c and $c+1$, that is,$S = T - [c, c+1]$, where T is a set of non–negative integers.

The mean of the MSBPL distribution is

$$\mu_1' = E(x) = \frac{\theta^2+4\theta+6}{\theta(\theta+2)} - \frac{\alpha(c+1)\theta^3(c+\theta+3)}{(\theta+2)(\theta+1)^{c+3}} \quad (12)$$

and

$$\mu_2' = E(x^2) = \sum x^2 p(x;\theta,\alpha)$$

$$= c^2 p(1;\theta,\alpha) + c^2.p(2;\theta,\alpha) + \sum_{x=3}^{\infty} x^2.p(x;\theta,\alpha)$$

$$= \sum_{x=1}^{\infty} x^2 \frac{x\theta^3(x+\theta+2)}{(\theta+2)(\theta+1)^{x+2}} - \frac{\alpha(2c+1)(c+1)\theta^3(c+\theta+3)}{(\theta+2)(\theta+1)^{c+3}} \quad (13)$$

Now, the first four moments about the origin and the variance of the SBPLD are given by

$$\mu_1'^* = \frac{\theta^2+4\theta+6}{\theta(\theta+2)}$$

$$\mu_2'^* = \frac{\theta^3+8\theta^2+24\theta+24}{\theta^2(\theta+2)}$$

$$\mu_3'^* = \frac{\theta^4+16\theta^3+78\theta^2+168\theta+120}{\theta^3(\theta+2)}$$

$$\mu_4'^* = \frac{\theta^5+32\theta^4+240\theta^3+840\theta^2+1320\theta+720}{\theta^2(\theta+2)}$$

$$\mu_2^* = \mu_2'^* - (\mu_1'^*)^2 = \frac{2(\theta^3+6\theta^2+12\theta+6)}{\theta^2(\theta+2)^2}$$

So by using the results of SBPL distribution in Eq. (13), we get μ_2' of MSBPL distribution as

$$\mu_2' = \frac{2(\theta^3+6\theta^2+12\theta+6)}{\theta^2(\theta+2)^2} + \left(\frac{\theta^2+4\theta+6}{\theta(\theta+2)}\right)^2 - \frac{\alpha(2c+1)(c+1)\theta^3(c+\theta+3)}{(\theta+2)(\theta+1)^{c+3}} \quad (14)$$

Hence the variance of MSBPL distribution is

$$\mu_2 = \frac{2(\theta^3+6\theta^2+12\theta+6)}{\theta^2(\theta+2)^2} - \frac{\alpha(c+1)\theta^2(c+\theta+3)}{(\theta+1)^{c+3}(\theta+2)^2}\left[\theta(2c+1)(\theta+2) - 2(\theta^2+4\theta+6) + \frac{\alpha(c+1)\theta^4(c+\theta+3)}{(\theta+1)^{c+3}}\right] \quad (15)$$

III. MAXIMUM LIKELIHOOD ESTIMATION

Suppose that a random sample of N observations of the random variable X is taken from the distribution given in Eq. (4). Let $x_1, x_2, x_3, \ldots, x_i, \ldots, x_k$ ($where\ k \in \{1, 2, \ldots \infty\}$ be the possible values of random variable X in a random sample and n_k denotes the number of observations corresponding to the value x_k in the sample, then wehave$\sum_{i=1}^{k} n_i = N$.

The likelihood function L of such a sample is given by

$$L \propto \prod_{i=1}^{k} P_i^{n_i} \quad (16)$$

$$= \{P_{(C;\theta,\alpha)}\}^{n_c}\{P_{(C+1;\theta,\alpha)}\}^{n_{c+1}} \prod_{i\in S}\{P_{(i;\theta,\alpha)}\}^{n_i}$$

Where S is a subset of the set I of non–negative integers (not containing $c\ and\ c+1$).

That is, $S = T - \{c, c+1\}$, where T is a set of non–negative integers and c is constant, which is independent of θ and α and is given by

$$c = \frac{N!}{\prod_{i=1}^{k} n_i!}$$

Trivedi and Patel (2013) have derived the equations

$$\alpha = \left[\frac{\{n_c(c+1)(c+\theta+3)\} - \{cn_{c+1}(c+\theta+2)(\theta+1)\}}{(c+1)(c+\theta+3)(n_c+n_{c+1})}\right] \quad (17)$$

and

$$N\left[\frac{2(\theta+3)}{\theta(\theta+2)}\right] - \frac{(c+3)(n_c+n_{c+1})}{(\theta+1)} + \frac{n_{c+1}}{c+\theta+3} + \frac{n_c[c(c+2\theta+3)+\alpha(c+1)]}{[c(c+\theta+2)(1+\theta)+\alpha(c+1)(c+\theta+3)]} - \sum_{i\in S} n_i\left[\frac{(x+2)(x+\theta+2)-(\theta+1)}{(\theta+1)(x+\theta+2)}\right] = 0 \quad (18)$$

Substituting α from Eq. (17) in Eq. (18), we get an equation containing only parameter θ, say $g(\theta) = 0$, and by applying any iteration method (e.g., Newton – Raphson), we can solve it for θ; which we call MLE of θ. By substituting this value of θ in Eq. (17), we get MLE of α.

IV. BAYES ESTIMATION

In this section, the Bayesian approach is used to derive estimates of the parameters θ and α. Here assume θ and α are both unknown. It is customary to assume that a-priori θ and α are independent since whatever prior belief one may have about θ, is not likely to be significantly influenced by one's knowledge about α. (Box and Tiao, 1973). Thus, joint prior of θ and α is given by

$$g(\theta,\alpha) = g_1(\alpha)g_2(\theta)$$

The prior density should be chosen both for mathematical tractability and especially for the ability to represent the prior technical knowledge.

Here we assume a uniform prior for parameter α,

i.e. $g_1(\alpha) = 1,\ 0 < \alpha < 1$ (19)

and exponential prior for parameter θ,

i.e. $g_2(\theta) = \frac{1}{\beta}e^{-\frac{\theta}{\beta}}, \theta > 0, \beta > 0$ (20) Hence the joint prior of θ and α is given by

$$g(\theta,\alpha) = g_1(\alpha)g_2(\theta) \tag{21}$$

If L is the likelihood function indexed by a continuous parameter $\Theta = (\theta,\alpha)$, with prior density $g(\theta,\alpha)$ then the posterior density for Θ is given by

$$\pi(\Theta|x) = \frac{L(\Theta)g(\theta,\alpha)d\theta\, d\alpha}{\int_\Theta L(\Theta)g(\theta,\alpha)d\theta\, d\alpha} \tag{22}$$

The Bayes estimation of an arbitrary function of Θ, say u(Θ) under the squared error loss function, is given by

$$\hat{\theta}_{BS} = E_\pi u(\Theta) = \frac{\int_\Theta u(\Theta)L(\Theta)g(\theta,\alpha)d\theta\, d\alpha}{\int_\Theta L(\Theta)g(\theta,\alpha)d\theta\, d\alpha}$$

Due to the complex form of the likelihood function (16), neither the posterior distribution nor the Bayes estimate simplifies to a closed-form. Lindley's approximation provides a numerical approximation method that is useful when the number of parameters is small (≤ 5, Press (1989)). This method provides an approximation for

$$I(x_1, x_2, \dots\dots, x_n) = \frac{\int_\Theta u(\Theta)L(\Theta)g(\theta,\alpha)d\theta\, d\alpha}{\int_\Theta L(\Theta)g(\theta,\alpha)d\theta\, d\alpha} \tag{23}$$

For the particular case of two parameters $\Theta = (\theta_1, \theta_2)$ Lindley's approximation of (Eq. (23) reduces to

$$\begin{aligned} I(x_1, x_2, \dots\dots, x_n) \cong{}& u(\hat{\Theta}) + \frac{1}{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\left\{\left[\frac{\partial^2 u(\Theta)}{\partial\theta_i\partial\theta_j} + 2\left(\frac{\partial u(\Theta)}{\partial\theta_i}\right)\left(\frac{\partial \log g(\Theta)}{\partial\theta_j}\right)\right]\hat{\sigma}_{ij}\right\}\Bigg|_{\theta=\hat{\theta}} \\ &+ \frac{1}{2}\left\{\frac{\partial^3 \log L(\Theta)}{\partial\theta_1^{\,3}}\left[\frac{\partial u(\Theta)}{\partial\theta_1}\hat{\sigma}_{11}^{\,2} + \frac{\partial u(\Theta)}{\partial\theta_2}\hat{\sigma}_{11}\hat{\sigma}_{21}\right]\right. \\ &+ \frac{\partial^3 \log L(\Theta)}{\partial\theta_1^{\,2}\partial\theta_2}\left[3\frac{\partial u(\Theta)}{\partial\theta_1}\hat{\sigma}_{11}\hat{\sigma}_{12} + \frac{\partial u(\Theta)}{\partial\theta_2}\left(\hat{\sigma}_{11}\hat{\sigma}_{22} + 2\hat{\sigma}_{12}^{\,2}\right)\right] \\ &+ \frac{\partial^3 \log L(\Theta)}{\partial\theta_1\partial\theta_2^{\,2}}\left[3\frac{\partial u(\Theta)}{\partial\theta_2}\hat{\sigma}_{12}\hat{\sigma}_{22} + \frac{\partial u(\Theta)}{\partial\theta_1}\left(\hat{\sigma}_{11}\hat{\sigma}_{22} + 2\hat{\sigma}_{12}^{\,2}\right)\right] \\ &\left.+ \frac{\partial^3 \log L(\Theta)}{\partial\theta_2^{\,3}}\left[\frac{\partial u(\Theta)}{\partial\theta_1}\hat{\sigma}_{12}\hat{\sigma}_{22} + \frac{\partial u(\Theta)}{\partial\theta_2}\hat{\sigma}_{22}^{\,2}\right]\right\}\Bigg|_{\theta=\hat{\theta}} \end{aligned} \tag{24}$$

Where $\hat{\sigma}_{ij}$ denotes the (i,j) element of the inverse of the observed information matrix given as

$$\hat{\Lambda} = \left[\begin{matrix} \frac{\partial^2 \log L(\Theta)}{\partial\theta_1^{\,2}} & \frac{\partial^2 \log L(\Theta)}{\partial\theta_1\partial\theta_2} \\ \frac{\partial^2 \log L(\Theta)}{\partial\theta_1\partial\theta_2} & \frac{\partial^2 \log L(\Theta)}{\partial\theta_2^{\,2}} \end{matrix}\right]\Bigg|_{\theta=\hat{\theta}} \tag{25}$$

Under mild conditions, this observed information matrix is a consistent estimator of the Fisher information matrix.

In our case, we take $\theta_1 = \theta, \theta_2 = \alpha$ and $u(\Theta) = u(\theta,\alpha) =$ a function of θ and α.

All functions on the right-hand side of Eq. (24) and Eq. (25) are evaluated at $\hat{\theta}$ and $\hat{\alpha}$. Replacing the unknown parameters by their maximum likelihood estimates.

When $u(\Theta) = \theta$, Eq. (24) reduces to

$$\begin{aligned} I(x_1, x_2, \dots\dots, x_n) \cong{}& \hat{\theta} - \frac{\hat{\sigma}_{11}}{\beta^2} \\ &+ \frac{1}{2}\left\{\begin{matrix} \frac{\partial^3 \log L(\Theta)}{\partial\theta^3}\hat{\sigma}_{11}^{\,2} + \frac{\partial^3 \log L(\Theta)}{\partial\theta^2\partial\alpha}3\hat{\sigma}_{11}\hat{\sigma}_{12} \\ + \frac{\partial^3 \log L(\Theta)}{\partial\theta\partial\alpha^2}\left(\hat{\sigma}_{11}\hat{\sigma}_{22} + 2\hat{\sigma}_{12}^{\,2}\right) + \frac{\partial^3 \log L(\Theta)}{\partial\alpha^3}\hat{\sigma}_{12}\hat{\sigma}_{22} \end{matrix}\right\}\Bigg|_{\theta=\hat{\theta}} \end{aligned} \tag{26}$$

Eq. (26) provides the Bayes estimate for θ under squared error loss.

When $u(\Theta) = \alpha$, Eq. (24) reduces to

$$I(x_1, x_2, \dots\dots, x_n) \cong \hat{\alpha} - \frac{\hat{\sigma}_{21}}{\beta^2} + \frac{1}{2}\left\{\begin{array}{l} \frac{\partial^3 \log L(\Theta)}{\partial \theta^3}\hat{\sigma}_{11}\hat{\sigma}_{21} + \frac{\partial^3 \log L(\Theta)}{\partial \theta^2 \partial \alpha}\left(\hat{\sigma}_{11}\hat{\sigma}_{22} + 2\hat{\sigma}_{12}{}^2\right) \\ + \frac{\partial^3 \log L(\Theta)}{\partial \theta \partial \alpha^2} 3\hat{\sigma}_{12}\hat{\sigma}_{22} + \frac{\partial^3 \log L(\Theta)}{\partial \alpha^3}\hat{\sigma}_{22}{}^2 \end{array}\right\}\Bigg|_{\alpha=\hat{\alpha}} \quad (27)$$

Eq. (27) provides the Bayes estimate for α under squared error loss.

To evaluate Eq. (26) and Eq. (27), the following third derivatives are obtained from Eq. (16) to Eq. (18)

$$\frac{\partial^3 \ln L(\Theta)}{\partial \theta^3} = N\left[\frac{6}{\theta^3} - \frac{2}{(\theta+2)^3}\right] - \frac{2(c+3)(n_c + n_{c+1})}{(\theta+1)^3} + \frac{2n_{c+1}}{(c+\theta+3)^3} + n_c \frac{1}{\{[c(c+\theta+2)(\theta+1)] + \alpha(c+1)(c+\theta+3)\}^2}\Bigg[-2c[c(c+2\theta+3) + \alpha(c+1)] - \frac{[c(c+2\theta+3)+\alpha(c+1)]^3 2\big[[c(c+\theta+2)(\theta+1)] + \alpha(c+1)(c+\theta+3)\big]}{\{[c(c+\theta+2)(\theta+1)] + \alpha(c+1)(c+\theta+3)\}^2}\Bigg] - \sum_{i \in S} n_i \left[\frac{2(x+2)}{(\theta+1)^3} - \frac{2}{(x+\theta+2)^3}\right] \quad (28)$$

$$\frac{\partial^3 \log L(\Theta)}{\partial \alpha^3} = \frac{2n_c(c+1)^3(c+\theta+3)^3}{[c(c+\theta+2)(\theta+1) + \alpha(c+1)(c+\theta+3)]^3} - \frac{2n_{c+1}}{(1-\alpha)^3} \quad (29)$$

$$\frac{\partial^3 \log L(\Theta)}{\partial \theta \partial \alpha^2} = -n_c(c+1)^2\left(\frac{\big(2(c+\theta+3)\big)}{[c(c+\theta+2)(\theta+1) + \alpha(c+1)(c+\theta+3)]^2} - \frac{((c+\theta+3)^2[c(c+2\theta+3) + \alpha(c+1)\})}{[c(c+\theta+2)(\theta+1) + \alpha(c+1)(c+\theta+3)]^4}\right) \quad (30)$$

$$\frac{\partial^3 \log L(\Theta)}{\partial \theta^2 \partial \alpha} = n_c(c+1)\left\{\frac{-(c(c+2\theta+3) + \alpha(c+1))}{[c(c+\theta+2)(1+\theta) + \alpha(c+1)(c+\theta+3)]^2} - \frac{(c+\theta+3)2c + c(c+2\theta+3) + \alpha(c+1)}{[c(c+\theta+2)(1+\theta) + \alpha(c+1)(c+\theta+3)]^2} + \frac{2(c+\theta+3)[c(c+2\theta+3) + \alpha(c+1)]^2}{[c(c+\theta+2)(1+\theta) + \alpha(c+1)(c+\theta+3)]^3}\right\} \quad (31)$$

V. REAL DATA APPLICATION

This section analyzes a real data set to illustrate the proposed estimation method described in the preceding section. The size-biased distributions have been used in modeling data relating to situations when organisms occur in groups, and group size influences the probability of detection. Keith and Meslow (1968) studied animal abundance data in which snowshoe hares were captured over seven days. A hare was marked and released after it had been captured. Subsequently, the same hare may or may not have been recaptured. Those captured on a previous day were identified by marking done on their first day of capture. Among 261 hares caught over seven days, 184 were caught once, 55 were caught twice, 14 were caught three times, four were caught four times, and four were caught five times. One question arises about whether the animal abundance data fit the MSBPL distribution or not. The following table shows the fitting of the MSBPL distribution to the data.

Table 1: Summary of fitting the MSBPL distribution to animal abundance data.

X_i	O_i	E_i
		MSBPL
1	184	182.8471
2	55	53.2472
3	14	18.77413
4	4	4.717652
5	4	1.413953
Total	261	261
		$\tilde{\theta} = 4.84502463$
		$\tilde{\alpha} = 0.18986707$
χ^2		2.98961
$p-$ value		0.0875

The above-reported $p-value$ indicates that the MSBPL is adequate for the 5% significance level. The MLE for θ and α are respectively $\hat{\theta}$ = 4.900827 and $\hat{\alpha}$ =0.16719. To examine the behavior of this pair of estimators, we have generated 1000 random samples from the MSBPL distribution by taking $\theta = 4.8$ and $\alpha = 0.18$. The results are reported in the following table:

	$\hat{\theta}$	$\hat{\alpha}$	$SE(\hat{\theta})$	$SE(\hat{\alpha})$
MLE	4.5341	0.2130	0.4704	0.1051
Bayes	4.8737	0.1766	0.4996	0.1463

VI. SIMULATION STUDY

Since the performance of the different estimates cannot be compared theoretically, an extensive Monte Carlo simulation study is conducted to compare the performance of the proposed Bayes estimates with the ML estimates in terms of simulated risk for different values of the parameter α, parameterθ, and sample size (n).In this section, we have generated 100000 different random samples of varying sample sizes n, θ, α, and values of $c = 1$ from the MSBPLD. In the Bayes estimation, the hyperparameterβ is taken as 0.5. As one data set does not help to clarify the performance of an estimate, the simulated risks were computed based on 100000 simulated data sets according to the following formulae:

$$SR = \sqrt{\frac{\sum_{i=1}^{100000}(\hat{\theta}_i-\theta)^2}{100000}} \text{ and } SR = \sqrt{\frac{\sum_{i=1}^{100000}(\hat{\alpha}_i-\alpha)^2}{100000}}$$

We have also obtained a coverage probability of the true value of the parameters θ and α included in the 95% confidence intervals

$(\hat{\theta} - 1.96\, SE(\hat{\theta}),\ \hat{\theta} + 1.96\, SE(\hat{\theta}))$ and

$(\hat{\alpha} - 1.96\, SE(\hat{\alpha}), \hat{\alpha} + 1.96\, SE(\hat{\alpha}))$

In the case of the above combinations and all, they are found to be 1. The simulation results are reported respectively, in Tables 2-5.

Table 2:Simulated risk of ML and Bayes estimates of θ and α for $\alpha = 0.01$ with different choices of n and θ

$\alpha = 0.01$		ML		Bayes	
θ	n	$SR(\theta)$	$SR(\alpha)$	$SR(\theta)$	$SR(\alpha)$
1	10	0.3226	0.3904	0.9372	0.3737
1.5	10	0.5964	0.4312	1.3209	0.3785
2	10	0.6907	0.4611	1.8516	0.3888
1	20	0.2596	0.3552	0.5882	0.3315
1.5	20	0.3947	0.3729	0.8082	0.3427
2	20	0.4942	0.4706	1.6112	0.3612
1	100	0.1097	0.1627	0.2771	0.1500
1.5	100	0.1878	0.2359	0.3104	0.1963
2	100	0.2424	0.2593	0.8186	0.2345
1	300	0.0504	0.1506	0.1156	0.0936
1.5	300	0.1106	0.1839	0.1827	0.1147
2	300	0.1896	0.3010	0.4592	0.1357

Table 3:Simulated risk of ML and Bayes estimates of θ and α for $\alpha = 0.05$ with different choices of n and θ

$\alpha = 0.05$		ML		Bayes	
θ	n	$SR(\theta)$	$SR(\alpha)$	$SR(\theta)$	$SR(\alpha)$
1	10	0.2517	0.3541	0.8928	0.2744
1.5	10	0.4153	0.3627	1.2425	0.2948
2	10	0.6690	0.4107	1.7647	0.3345
1	20	0.2201	0.2859	0.5069	0.2496
1.5	20	0.3642	0.3055	0.9153	0.2824
2	20	0.4814	0.3895	1.4721	0.3058
1	100	0.1031	0.1150	0.1217	0.1202
1.5	100	0.1516	0.1256	0.3100	0.1224
2	100	0.1783	0.1921	0.7909	0.1854
1	300	0.0498	0.1194	0.1122	0.0522
1.5	300	0.1092	0.1807	0.1698	0.0894
2	300	0.1712	0.2840	0.4073	0.1310

Table 4:Simulated risk of ML and Bayes estimates of θ and α for $\alpha = 0.1$ with different choices of n and θ

$\alpha = 0.1$		ML		Bayes	
θ	n	$SR(\theta)$	$SR(\alpha)$	$SR(\theta)$	$SR(\alpha)$
1	10	0.2362	0.2967	0.6770	0.2048
1.5	10	0.3244	0.3223	1.2397	0.2624
2	10	0.5773	0.3627	1.6991	0.2958
1	20	0.1083	0.2201	0.4944	0.1976
1.5	20	0.3164	0.2688	0.8067	0.2377
2	20	0.4627	0.3392	1.4070	0.2853
1	100	0.0912	0.1213	0.1183	0.1118
1.5	100	0.1509	0.1254	0.3061	0.1148
2	100	0.1691	0.1962	0.7217	0.1917
1	300	0.0439	0.1166	0.1095	0.0473
1.5	300	0.1060	0.1797	0.1474	0.0891
2	300	0.1288	0.2562	0.3926	0.1205

Table 5:Simulated risk of ML and Bayes estimates of θ and α for $\alpha = 0.2$ with different choices of n and θ

$\alpha = 0.2$		ML		Bayes	
θ	n	$SR(\theta)$	$SR(\alpha)$	$SR(\theta)$	$SR(\alpha)$
1	10	0.2300	0.2864	0.5908	0.1416
1.5	10	0.3037	0.3144	1.2309	0.2104
2	10	0.4697	0.3478	1.6678	0.2514
1	20	0.0987	0.1813	0.4759	0.1354
1.5	20	0.2913	0.2044	0.7810	0.2023
2	20	0.4048	0.2557	1.3506	0.2492
1	100	0.0896	0.0814	0.1243	0.0725
1.5	100	0.1492	0.1646	0.3096	0.1607
2	100	0.1632	0.1691	0.5590	0.1862
1	300	0.0433	0.0733	0.0310	0.0629
1.5	300	0.0826	0.1785	0.1402	0.0783
2	300	0.1222	0.1969	0.3883	0.1153

Graphs of Simulated Risk of ML and Bayes estimators for a fixed value of θ with different choices of n and α

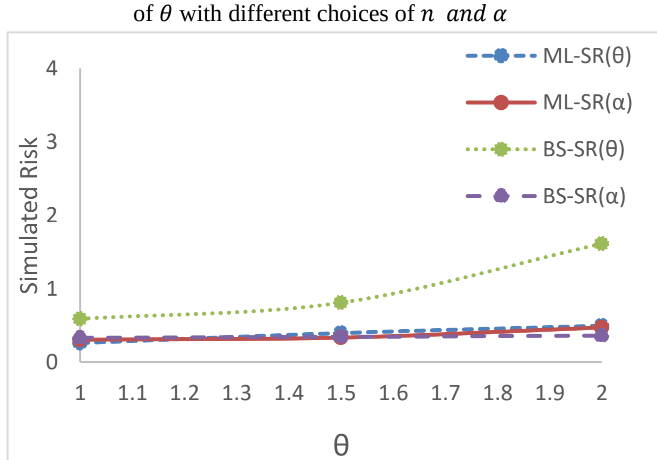


Fig. 1. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 10$ and $\alpha = 0.01$

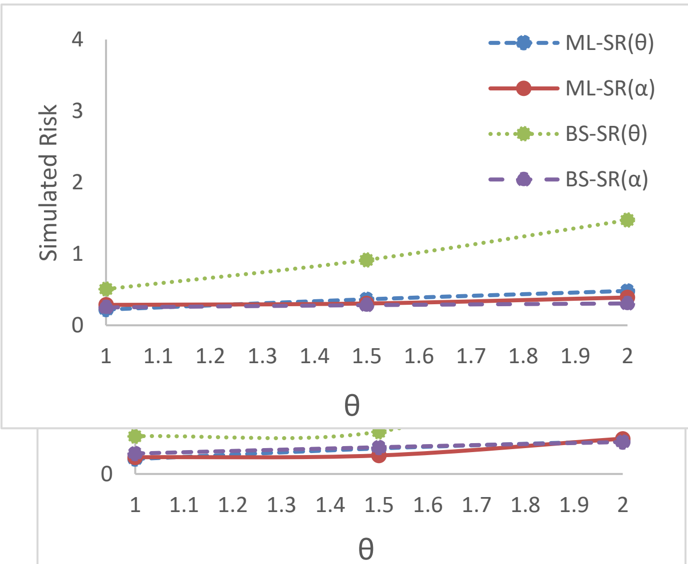
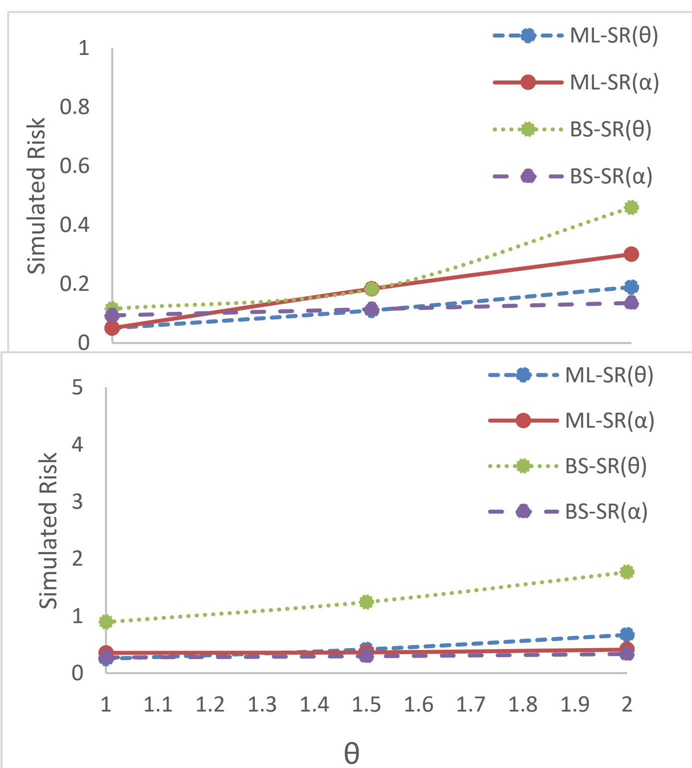


Fig. 3. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 100$ and $\alpha = 0.01$



Fig

Fig

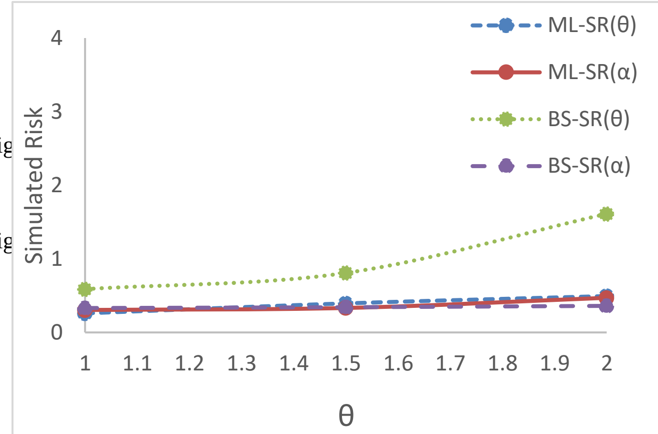


Fig. 6. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 20$ and $\alpha = 0.05$

Fig. 7. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 100$ and $\alpha = 0.05$

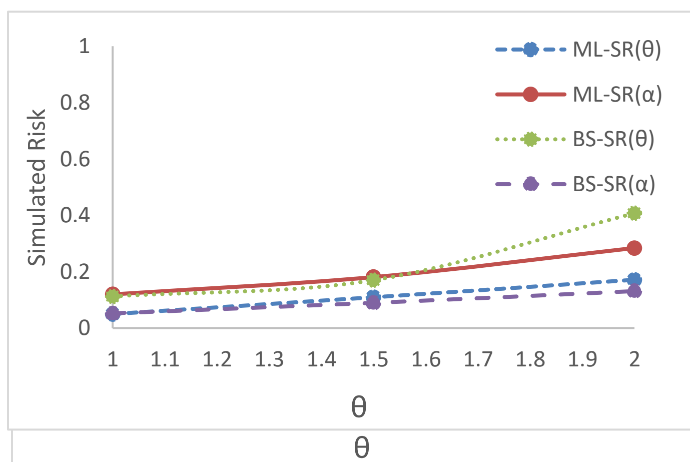


Fig. 8. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 300\ \ and\ \alpha = 0.05$

Fig. 9. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 10\ \ and\ \alpha = 0.1$

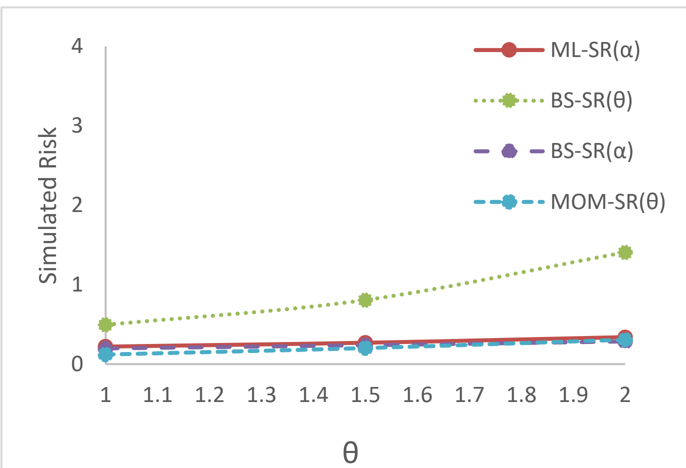


Fig. 10. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 20\ \ and\ \alpha = 0.1$

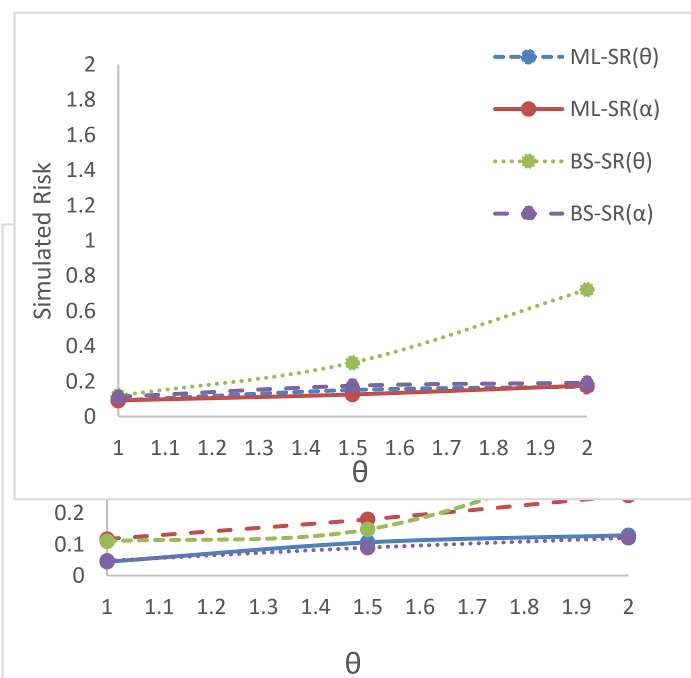


Fig. 11. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 100\ \ and\ \alpha = 0.1$

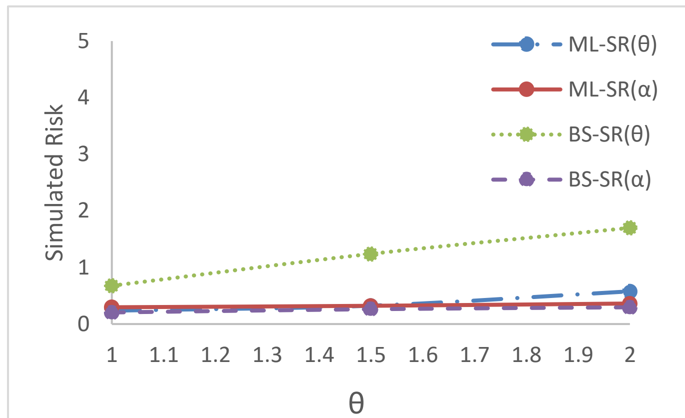


Fig. 12. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 300\ \ and\ \alpha = 0.1$

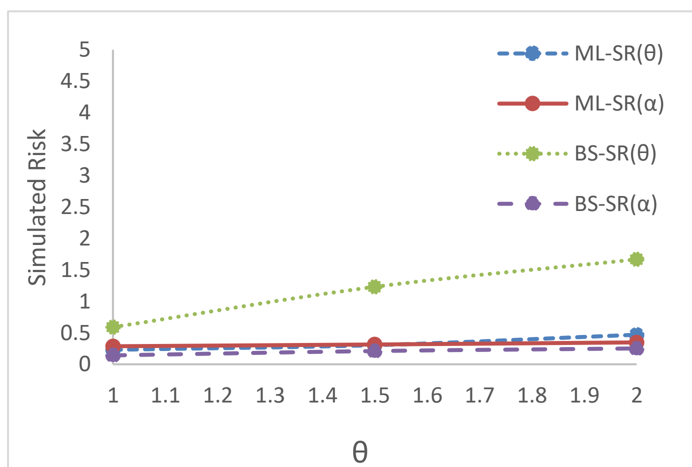


Fig. 13. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 10 \;\; and \;\alpha = 0.2$

Fig. 14. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 20 \;\; and \;\alpha = 0.2$

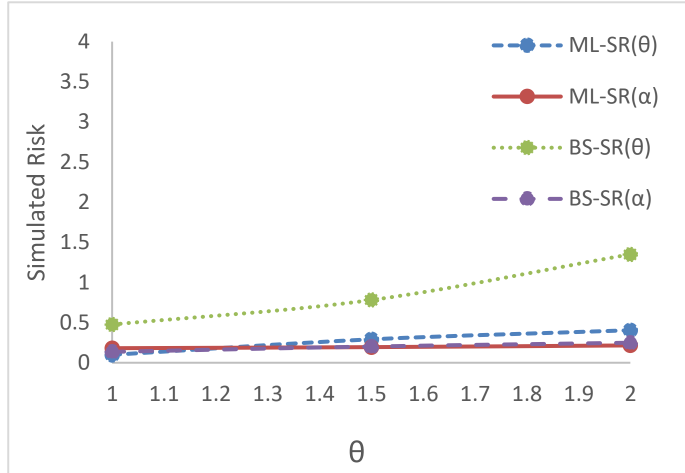
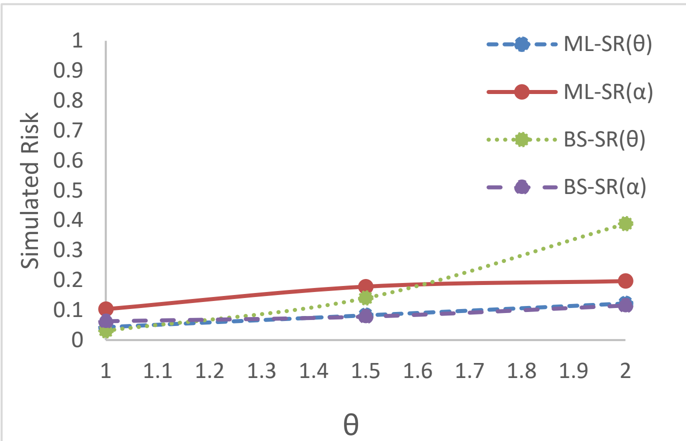


Fig. 15. Simulated Risk of ML and Bayes estimators for a fixed value of



θ with different $n = 100 \;\; and \;\alpha = 0.2$

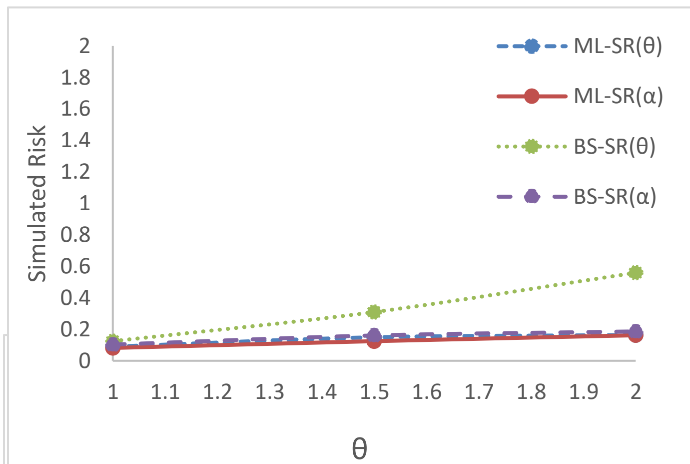
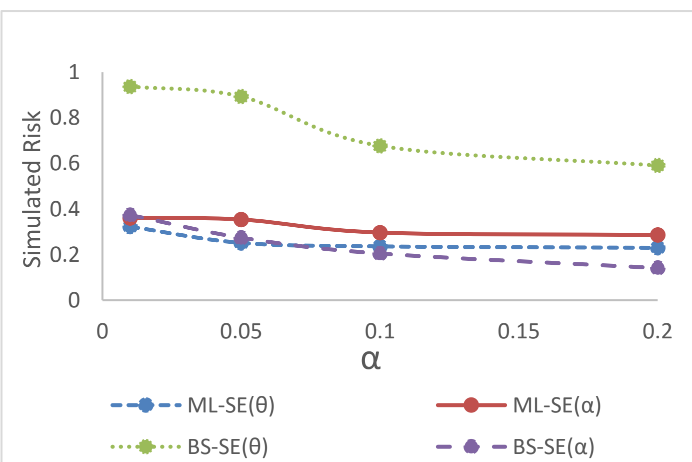


Fig. 16. Simulated Risk of ML and Bayes estimators for a fixed value of θ with different $n = 300 \;\; and \;\alpha = 0.2$



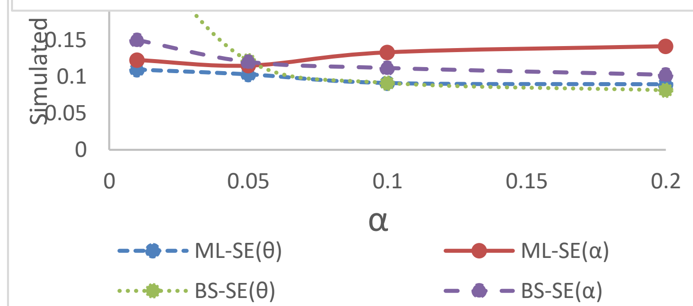
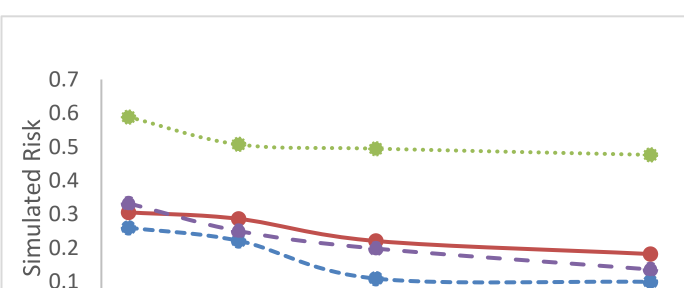


Fig. 17. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 10 \;\; and \;\theta = 1$

Fig. 18. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 20 \;\; and \;\theta = 1$



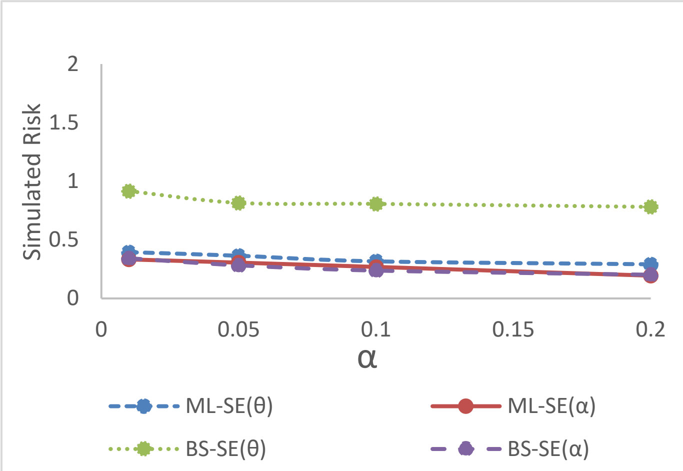
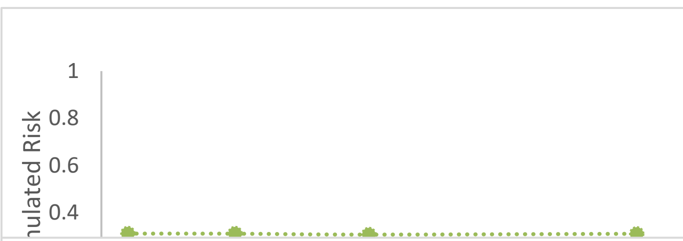


Fig. 19. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 100$ and $\theta = 1$

Fig. 20. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 300$ and $\theta = 1$

Fig. 21. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 10$ and $\theta = 1.5$

Fig. 22. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 20$ and $\theta = 1.5$



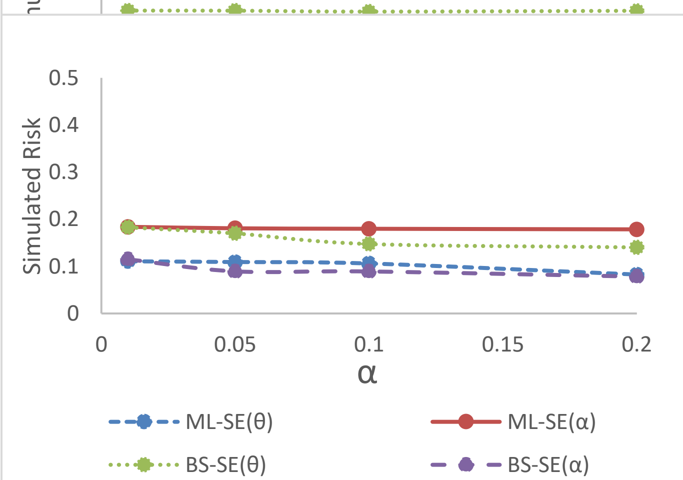
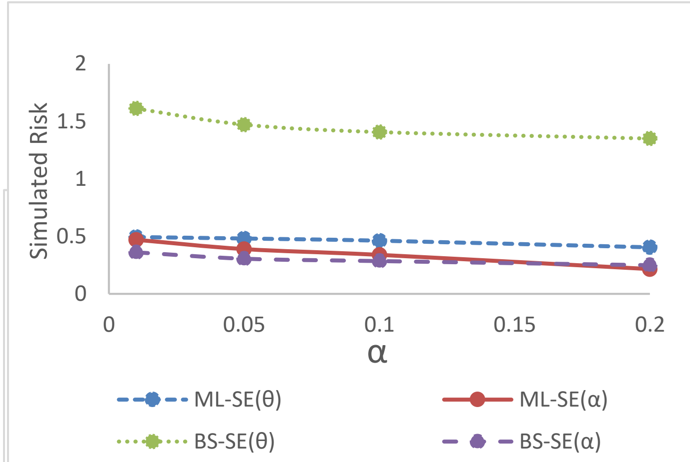


Fig. 23. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 100$ and $\theta = 1.5$

Fig. 24. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 300$ and $\theta = 1.5$

Fig. 25. Simulated Risk of ML and Bayes estimators for a fixed

value of α, with $n = 10$ and $\theta = 2$



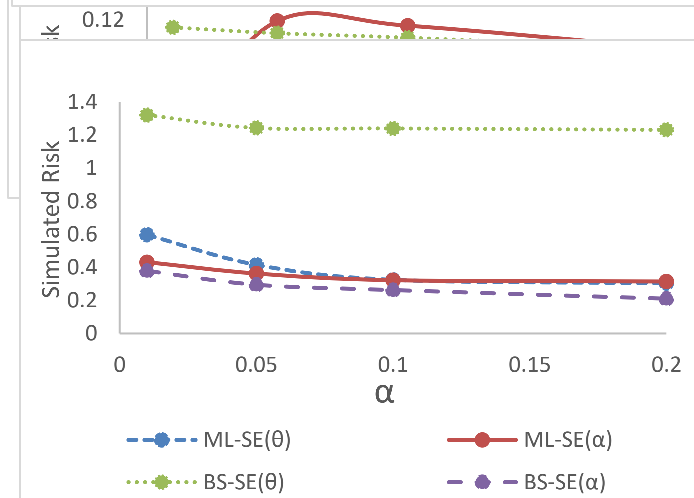


Fig. 26. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 20$ and $\theta = 2$

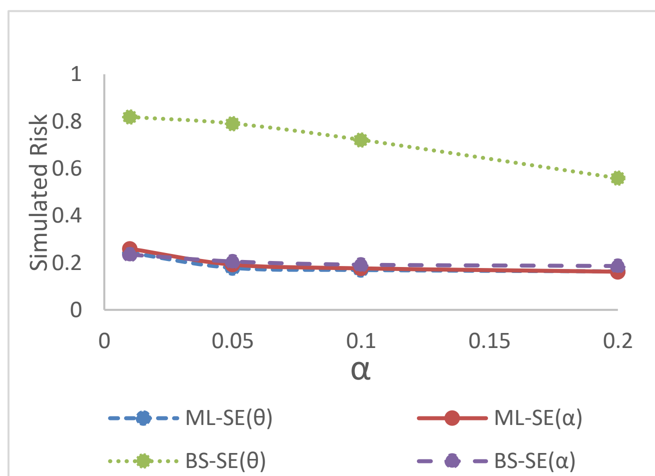
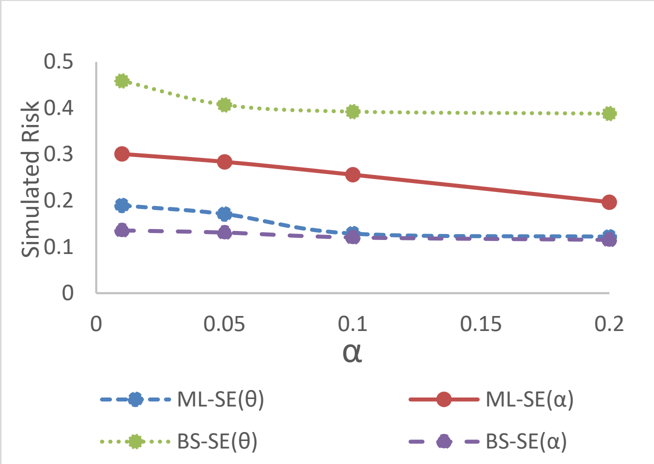


Fig. 27. Simulated Risk of ML and Bayes estimators for a fixed value of α, with $n = 100$ and $\theta = 2$

Fig. 28. Simulated Risk of ML and Bayes estimators for a fixed



value of α, with $n = 300$ and $\theta = 2$

VII. Simulation Results

From the results of the Monte Carlo simulation study presented in Tables 2.6 – 2.8 and graphs, the following points can be drawn:

1. For fixed misclassification error(α)as θ increases the simulated risk of ML, and Bayes estimates of θ increases.
2. For fixed misclassification error(α) as θ increases, the simulated risk of ML and Bayes estimates of α increase.
3. As sample size (n)increases, the simulated risk of ML and Bayes estimates of θ and α decreases.
4. In all cases, the simulated risk of ML estimates of θ is smaller than Bayes estimates of θ.
5. In all cases, the simulated risk of Bayes estimates of α is smaller than the simulated risk of ML estimates of θ.
6. For a large sample size (n), ML estimation gives better results compared to Bayes estimation in most cases.

CONCLUSION

A misclassified size-biased Poisson-Lindley distribution (MSBPLD) has been proposed, and its nature for varying values of parameters has been studied. The ML and Bayes estimates of θ and $\alpha$are obtained. A real data set has been analyzed for an illustrative purpose. The distribution's applications and goodness of fit have been explained through datasets relating to the size distribution of freely-forming small groups, and fit has been found satisfactory over SBPD and SBPLD. The simulation study is carried out to examine and compare the performance of ML and Bayes estimates in terms of simulated risk for different sample sizes and different values of the parameters α and θ. The simulation results show that the ML estimates perform better than their corresponding Bayes estimates for the large sample size. Moreover, the estimated risks of the estimates get smaller with the decreasing value of θ.

REFERENCES

Borah, M.,&Nathl, A.D. (2001). A Study on the inflated Poisson Lindley distribution. *Journal of the Indian Society of Agricultural Statistics*, 54, 317-323.

Bortkiewcz, L. Von. (1898). Das Gesetz der kleinenZahlen, Leipzig.*Teubner*.

Box, G.,&Tiao, G. (1973). Bayesian inference in statistical analysis, *John Wiley & Sons*

Cohen, A. C. (1960). Estimating the parameters of a modified Poisson distribution.*Journal of the American Statistical Association*, 55, 139 -143.

Elbatal, I., Merovci, F.,&Elgarhy, M. (2013). A new generalized Lindley distribution.*Mathematical Theory, and Modeling*, 13(3), 30-47.

Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies. *Annals of Eugenics*, 6, 13-25.

Ghitany, M. E., & Al-Mutairi, D. K. (2008). Size-Biased Poisson-Lindley distribution and its application.*Metron - International Journal of Statistics*,LXVI, 299 – 311.

Ghitany, M.E.,Atieh, B., &Nadarajah, S. (2008). Lindley distribution and its application. *Mathematics, and Computers in Simulation*, 78(4), 493-506.

Ghitany, M.E., Alqallaf, F., Al-Mutairi, D.K., & Husain, H.A. (2011). A two-parameter weighted Lindley distribution and its applications to survival data.*Mathematics, and Computers in Simulation*, 81, 1190-1201.

Keith, L. B., &Meslow, E. C. (1968).Trap response by snowshoe hares.*Journal of Wildlife Management*, 20, 795-801.

Lord, D.,&Geedipally, S.R. (2011). The Negative Binomial-Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros.*Accident Analysis & Prevention*, 43, 1738-1742.

Merovci, F., & Sharma, V.K. (2014). The Beta-Lindley distribution: properties and applications.*Journal of Applied Mathematics*, 2014.

Rao, C. R. (1965). On discrete distributionsarising out of methods of ascertainment,*Sankhyā: The Indian Journal of Statistics, Series A*, Vol. 27,311-324.

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

Shanker, R., Sharma, S., &Shanker, R. (2013). A two-parameter Lindley distribution for modeling waiting and survival times data. Applied *Mathematics*,4, 363-368.

Shanker, R., Hagos, F.,&Abrehe, Y. (2015). On Size Biased Poisson - Lindley distribution and its applications to model thunderstorms.*American Journal of Mathematics and Statistics*,5(6), 354-360.

Shanker, R., Hagos, F.,& Sujatha, S. (2015). On modeling of lifetime data using one parameter Akash, Lindley, and Exponential distribution.*Biometrics & Biostatistics International Journal*,2(5), 56-62.

Shanker, R., &Hagos, F.(2015). On Poisson-Lindley distribution and its applications to biological sciences. *Biometrics & Biostatistics International Journal*, 2(4),103-107.

Singh, S.K., Singh, U., &Sharma, V.K. (2014). The truncated Lindley distribution:Inference and application.*Journal of Statistics Applications & Probability*,3, 219-228.

Press, S. J. (1989), Bayesian statistics: Principles, Models, and applications: *John Wiley & Sons.*

Trivedi, B. S.,& Patel, M. N.(2013). Estimation in misclassified Size-Biased Generalized Negative Binomial distribution.*Mathematics and Statistics*,1 (2), 74-85.

Zakerzadeh, H. &Dolati, A. (2009). Generalized Lindley Distribution.*Journal of Mathematical Extension* 3, 13-25.

***