STATISTICA, anno LXXXI, n. 1, 2021

THE ZOGRAFOS-BALAKRISHNAN LINDLEY DISTRIBUTION: PROPERTIES AND APPLICATIONS

Muhammed Rasheed Irshad [1]
Department of Statistics, Cochin University of Science and Technology, Kochi-22, Kerala
Veena D'cruz
Department of Statistics, Cochin University of Science and Technology, Kochi-22, Kerala
Radhakumari Maya
Department of Statistics, Government College for Women, University of Kerala, Trivandrum-695014

1. Introduction

The Lindley distribution was originally proposed by Lindley (1958), in the context of Bayesian statistics, as a counter example of fiducial statistics. Lindley developed the distribution whose probability density function (pdf) is obtained by mixing densities, exponential (θ) and gamma $(2,\theta)$ with mixing probabilities $\frac{\theta}{1+\theta}$ and $\frac{1}{1+\theta}$ respectively. The pdf of Lindley distribution is given by

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

The corresponding cumulative distribution function (cdf) is given by

$$F(y)=1-e^{-\theta y}\left[1+\frac{\theta y}{1+\theta}\right], \qquad y>0,\theta>0. \tag{2}$$

Ghitany *et al.* (2008) discussed various properties of this distribution and showed that the Lindley distribution provides a better fit than the exponential distribution. A discrete version of the Lindley distribution was suggested by Gómez-Déniz and Calderín-Ojeda (2011) based on an application related to insurance. The Lindley mixture of Poisson

[1] Corresponding Author. E-mail: irshadmr@cusat.ac.in

distribution is obtained by Sankaran (1970). Ghitany *et al.* (2011) developed a two-parameter weighted Lindley distribution and discussed its applications to survival data. Nadarajah *et al.* (2011) have obtained the generalized Lindley distribution. Shibu and Irshad (2016) proposed extended new generalized Lindley distribution. Another extension of generalized Lindley distribution was proposed by Irshad and Maya (2017). Maya and Irshad (2017) developed the generalized Stacy-Lindley mixture distribution by mixing two generalized Stacy-gamma distributions.

Zografos and Balakrishnan (2009) proposed a class of distribution which is generated by gamma random variables with an extra positive shape parameter. For any baseline cdf $G(z), z \in R$, Zografos and Balakrishnan defined the Zografos-Balakrishnan G (ZBG) distribution with pdf f(z), cdf F(z) and for $a > 0$ is given by

$$f(z) = \frac{1}{\Gamma(a)}\{-\log[1-G(z)]\}^{a-1} g(z), \quad z > 0 \tag{3}$$

and

$$\begin{aligned} F(z) &= \frac{\gamma(a, -\log[1-G(z)])}{\Gamma(a)} \\ &= \frac{1}{\Gamma(a)} \int_0^{-\log[1-G(z)]} t^{a-1} e^{-t} dt, \quad a, z > 0, \end{aligned} \tag{4}$$

where $g(z)=\frac{d}{dz}(G(z))$, $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ denotes the gamma function, $\gamma(a,u) = \int_0^u t^{a-1} e^{-t} dt$ denotes the lower incomplete gamma function and $\Gamma(a,u) = \int_u^\infty t^{a-1} e^{-t} dt$ denotes the upper incomplete gamma function. The corresponding hazard rate function is given by

$$h(z) = \frac{\{-\log[1-G(z)]\}^{a-1} g(z)}{\Gamma(a, -\log[1-G(z)])}. \tag{5}$$

Nadarajah *et al.* (2015) developed various properties of ZBG family of distributions. Zografos-Balakrishnan loglogistic distribution was obtained by Hamedani (2013).

In this paper, the generator suggested by Zografos and Balakrishnan (2009) is used to define a new model, the Zografos-Balakrishnan Lindley (ZBL) distribution, which generalizes the classical one parameter Lindley distribution. The main motivation behind the construction of ZBG family of distributions is that, if a random variable Z have density (1), then a logarithmic transformation of G(z), that is., $u = -\log(1-G(z))$ have gamma $G(a,1)$. Also different structural properties are defined and the parameters are estimated using the method of maximum likelihood. It is obtained that the proposed model is more flexible than the classical one parameter Lindley distribution and can be effectively used to model certain real life data sets. The rest of the paper is organised as

follows.
In Section 2 ZBL distribution is defined and its moments and model identifiability is obtained. Some of the reliability properties of the model such as hazard rate, survival function followed by an expression for stress strength reliability are presented in Section 3. In Section 4 quantile function is derived. The parameters of ZBL distribution are estimated using the method of maximum likelihood and thus obtained observed Fisher information matrix and asymptotic confidence intervals which are given in Section 5. The results of montecarlo simulation is given in Section 6 and finally in Section 7 an application to the model using two real data sets is illustrated.

2. The Zografos-Balakrishnan Lindley distribution

In this Section we define the ZBL distribution and give useful expansions of the pdf.

2.1. *Definition*

A continuous random variable X is said to follow ZBL distribution with parameters a and θ if its pdf is of the form,

$$\begin{aligned} f(x) &= \frac{1}{\Gamma(a)}\left\{-\log\left[1-\left(1-e^{-\theta x}\frac{1+\theta+\theta x}{1+\theta}\right)\right]\right\}^{a-1}\frac{\theta^2}{\theta+1}(1+x)e^{-\theta x} \\ &= \frac{1}{\Gamma(a)}\left[\log\left(\frac{1+\theta}{1+\theta+\theta x}e^{\theta x}\right)\right]^{a-1}\frac{\theta^2}{\theta+1}(1+x)e^{-\theta x}, \quad x>0,\ a,\theta>0. \end{aligned} \tag{6}$$

The cdf is given by,

$$\begin{aligned} F(x) &= \frac{1}{\Gamma(a)}\gamma\left(a,\log\left[\frac{1+\theta}{1+\theta+\theta x}e^{\theta x}\right]\right) \\ &= \frac{1}{\Gamma(a)}\int_0^{\log(\frac{1+\theta}{1+\theta+\theta x}e^{\theta x})} t^{a-1}e^{-t}dt, \quad x>0,\ a,\theta>0. \end{aligned} \tag{7}$$

Remark 1. *When $a=1$, the ZBL distribution reduces to the one parameter Lindley distribution.*

The pdf of ZBL distribution, for different values of parameters, is plotted in Figure 1.

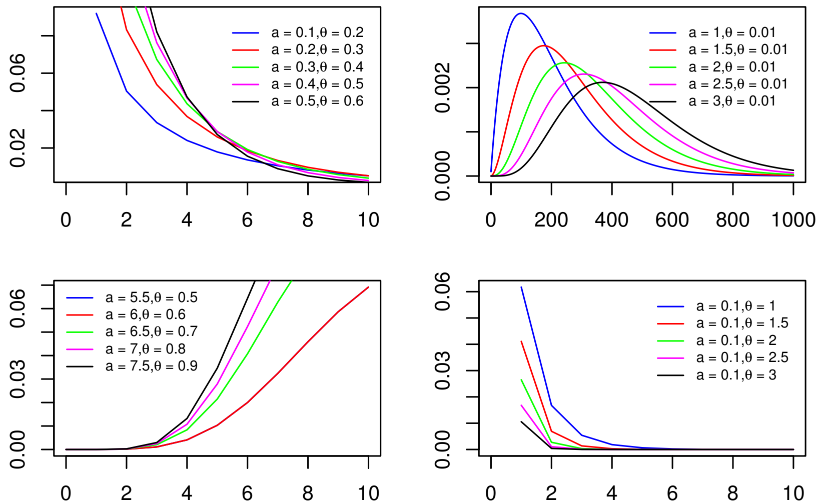

Figure 1 – The pdf of ZBL distribution is plotted for different values of a and θ.

2.2. *Expansion of pdf*

2.2.1. *Expansion in terms of power series*

The ZBL distribution with pdf given in (6) is expanded. We have

$$f(x) = \frac{1}{\Gamma(a)}\left\{-\log\left[1-\left(1-e^{-\theta x}\frac{1+\theta+\theta x}{1+\theta}\right)\right]\right\}^{a-1} \frac{\theta^2}{\theta+1}(1+x)e^{-\theta x}\ , x>0,\ a,\theta>0.$$

Now consider the following results for simplifications. Using the series representation,

$$-\log(1-y)=\sum_{i=0}^{\infty}\frac{y^{i+1}}{i+1},$$

we can write, (see Fagbamigbe *et al.*, 2018)

$$(-\log(1-y))^{a-1}=y^{a-1}\sum_{m=0}^{\infty}\binom{a-1}{m}y^m\left[\sum_{s=0}^{\infty}\frac{y^s}{s+2}\right]^m. \tag{8}$$

A power series raised to a positive integer m with $a_s=\frac{1}{s+2}$ can be written as (see Gradshteyn and Ryzhik, 2014),

$$\left(\sum_{s=0}^{\infty} a_s y^s\right)^m = \sum_{s=0}^{\infty} b_{s,m} y^s, \tag{9}$$

$$\text{where} \quad b_{s,m} = \frac{1}{s a_0} \sum_{i=1}^{s} (i(m+1)-s) a_i b_{s-i} \quad for \quad s \geq 1.$$

Also, we have the generalization of the binomial theorem,

$$(1-z)^{b-1} = \sum_{k=0}^{\infty} \frac{(-1)^k \Gamma(b)}{\Gamma(b-k)k!} z^k \tag{10}$$

and the basic results of power series and gamma function,

$$(1+x)^t = \sum_{r=0}^{t} \binom{t}{r} x^r \quad \text{and} \tag{11}$$

$$\int_0^{\infty} e^{-mx} x^{p-1} dx = \frac{\Gamma(p)}{x^p}. \tag{12}$$

Applying the results (8) to (12) in (6),we get the density function of ZBL as,

$$\begin{aligned} f(x) &= \frac{1}{\Gamma(a)} \sum_{m,s,p=0}^{\infty} \binom{a-1}{m} b_{s,m} (-1)^p \binom{a+m+s-1}{p} \\ &\quad \sum_{q=0}^{p} \frac{\binom{p}{q} \theta^{q+2}}{(1+\theta)^{p+1}} \sum_{t=0}^{q+1} \binom{q+1}{t} x^t e^{-\theta x(p+1)}. \end{aligned} \tag{13}$$

Hence (6) reduced to a constant multiplied by a gamma form which in turn is useful in determining various properties of the ZBL distribution.

2.2.2. *Expansion in terms of exponentiated Lindley distribution*

A random variable is said to have exponentiated-G distribution (exp-G) (see Cordeiro *et al.*, 2013), if its pdf and cdf are respectively given as,
for any $a > 0$,

$$h_a(x) = a G^{a-1}(x) g(x) \quad \text{and} \quad H_a(x) = G^a(x). \tag{14}$$

Nadarajah *et al.* (2015) expressed (3) as

$$f(z) = \sum_{k=0}^{\infty} b_k h_{a+k}(z) \tag{15}$$

$$\text{where} \quad b_k = \frac{\binom{k+1-a}{k}}{(a+k)\Gamma(a-1)} \sum_{j=0}^{k} \frac{(-1)^{j+k}\binom{k}{j} P_{j,k}}{(a-1-j)},$$

$P_{j,k}$s are obtained from the relation

$$p_{j,k} = \frac{1}{k} \sum_{m=1}^{k} \frac{(-1)^m [m(j+1)-k]}{(m+1)} c_m p_{j,k-m} \tag{16}$$

for k=1,2,... and $P_{j,0} = 1$. Here $h_{a+k}(x)$ denotes the pdf of exp-G$(a+k)$ distribution and its cdf can be represented as

$$F(x) = \sum_{k=0}^{\infty} b_k H_{a+k}(x), \tag{17}$$

where $H_{a+k}(x)$ denotes the cdf of exp-G $(a+k)$ distribution. Nadarajah *et al.* (2011) introduced exponentiated Lindley distribution. If $h(x)$ and $H(x)$ denote the pdf and cdf exponentiated Lindley distribution, then the pdf and cdf of ZBL distribution can be written as in (15) and (17).

2.3. *Identifiability*

The problem of identifiability lies in the uniqueness of distribution for the unknown parameters in the model. That is, the parameters should uniquely determine a distribution. The set of unknown parameters of a particular model is said to be identifiable if different sets of parameters gives different distribution for given x. Here, the identifiability property of ZBL is verified. Let $f(\Theta_1)$ and $f(\Theta_2)$ be different members of ZBL distribution indexed by $\Theta_1 = (a_1, \theta_1)$ and $\Theta_2 = (a_2, \theta_2)$ respectively. Then the likelihood ratio,

$$\begin{aligned} L &= \frac{f(\Theta_1)}{f(\Theta_2)} \\ &= \frac{\Gamma(a_2)(\theta_1 x + \log \frac{1+\theta_1}{1+\theta_1+\theta_1 x})^{a_1-1}(\theta_2+1)\theta_1^2}{\Gamma(a_1)(\theta_2 x + \log \frac{1+\theta_2}{1+\theta_2+\theta_2 x})^{a_2-1}(\theta_1+1)\theta_2^2} e^{-\theta_1 x + \theta_2 x}. \end{aligned} \tag{18}$$

Taking logarithm of likelihood,

$$\begin{aligned} \log L &= \log\left(\frac{\Gamma(a_2)}{\Gamma(a_1)}\right) + (a_1 - 1) \log\,(\theta_1 x + \log \frac{1+\theta_1}{1+\theta_1+\theta_1 x}) \\ &- (a_2 - 1) \log\,(\theta_2 x + \log \frac{1+\theta_2}{1+\theta_2+\theta_2 x}) + \log\left(\frac{\theta_2+1}{\theta_1+1}\right) \\ &+ \log\left(\frac{\theta_1^2}{\theta_2^2}\right) - \theta_1 x + \theta_2 x \end{aligned} \tag{19}$$

Partial derivative of log L with respect to x and equating it to 0,

$$\frac{\partial \log L}{\partial x} = 0 \implies$$

$$\frac{(a_1-1)\theta_1^2 - \theta_1(\theta_1 x + \log\left(\frac{1+\theta_1}{1+\theta_1+\theta_1 x}\right)(1+\theta_1+\theta_1 x)}{\left(\theta_1 x + \log\left(\frac{1+\theta_1}{1+\theta_1+\theta_1 x}\right)\right)(1+\theta_1+\theta_1 x)}$$
$$= \frac{(a_2-1)\theta_2^2 - \theta_2(\theta_2 x + \log\left(\frac{1+\theta_2}{1+\theta_2+\theta_2 x}\right)(1+\theta_2+\theta_2 x)}{\left(\theta_2 x + \log\left(\frac{1+\theta_2}{1+\theta_2+\theta_2 x}\right)\right)(1+\theta_2+\theta_2 x)}. \tag{20}$$

In (20), RHS=LHS iff $a_1 = a_2$ and $\theta_1 = \theta_2$. That is, the parameters uniquely determines the distribution. Therefore we conclude that the model is identifiable that is, $f(\Theta_1) = f(\Theta_2) \iff \Theta_1 = \Theta_2$.

2.4. *Moments*

THEOREM 2. *The r^{th} moment of the ZBL distribution is obtained as*

$$E(X^r) = \frac{\theta^2}{\theta+1}\frac{1}{\Gamma(a)}\sum_{m,s,p=0}^{\infty}\binom{a-1}{m} b_{s,m}(-1)^p \binom{a+m+s-1}{p}$$
$$\sum_{q=0}^{p}\frac{\binom{p}{q}\theta^q}{(1+\theta)^p}\sum_{t=0}^{q+1}\binom{q+1}{t}\frac{\Gamma(r+t+1)}{[\theta(p+1)]^{r+t+1}}. \tag{21}$$

where $b_{s,m}$ is defined in (9).

PROOF. The r^{th} moment of ZBL random variable is such that,

$$E(X^r) = \int_0^{\infty} x^r f(x) dx. \tag{22}$$

Using (8) and (9) we have

$$E(X^r) = \frac{\theta^2}{\theta+1}\frac{1}{\Gamma(a)}\sum_{m,s,p=0}^{\infty}\binom{a-1}{m} b_{s,m}(-1)^p \binom{a+m+s-1}{p}$$
$$\sum_{q=0}^{p}\frac{\binom{p}{q}\theta^q}{(1+\theta)^p}\int_{x=0}^{\infty} x^r (1+x)^{q+1} e^{-\theta(p+1)x} dx. \tag{23}$$

Using (11) and (12), and simplifying (23) we get (21). □

From this expression one can easily find out the various properties related to moments such as mean, variance, measures of skewness and kurtosis etc. Table 1 provides values of first five moments as well as standard deviation (SD), coefficient of variation (CV), coefficient of skewness (CS), coefficient of kurtosis (CK) for selected values of parameters of ZBL distribution.

TABLE 1
The moments related to ZBL distribution for selected values of parameters.

	$a=1,\theta=0.5$	$a=1,\theta=1$	$a=1,\theta=2$	$a=1,\theta=3$
μ_1'	3.333	1.5	0.667	0.417
μ_2'	18.667	4	0.833	0.333
μ_3'	144	15	1.5	0.389
μ_4'	1408	72	3.5	0.593
μ_5'	16639.999	420	10	1.111
SD	2.749	1.323	0.624	0.399
CV	0.825	0.882	0.935	0.96
CS	-6.306	-4.59	-2.215	-0.172
CK	6.343	6.796	7.469	7.889

	$a=2,\theta=0.5$	$a=2,\theta=1$	$a=2,\theta=2$	$a=2,\theta=3$
μ_1'	6.069	2.819	1.29	0.816
μ_2'	49.471	11	2.377	0.967
μ_3'	503.863	54.416	5.652	1.493
μ_4'	6146.4	325.11	16.399	2.831
μ_5'	87296.9	2273.32	56.068	6.350
SD	3.555	1.747	0.845	0.549
CV	0.586	0.62	0.655	0.673
CS	-13.971	-11.187	-7.751	-5.185
CK	4.869	4.995	5.232	5.406

From Table 1, we can conclude about skewness and kurtosis of ZBL distribution as,
(i) when a is fixed and θ is increasing, skewness and kurtosis are increasing,
(ii) when θ is fixed and a is increasing, skewness and kurtosis are decreasing.

3. Some measures of reliability

In this section, we discuss certain measures of reliability such as survival and hazard rate function for ZBL distribution. Also an expression for stress strength reliability is given.

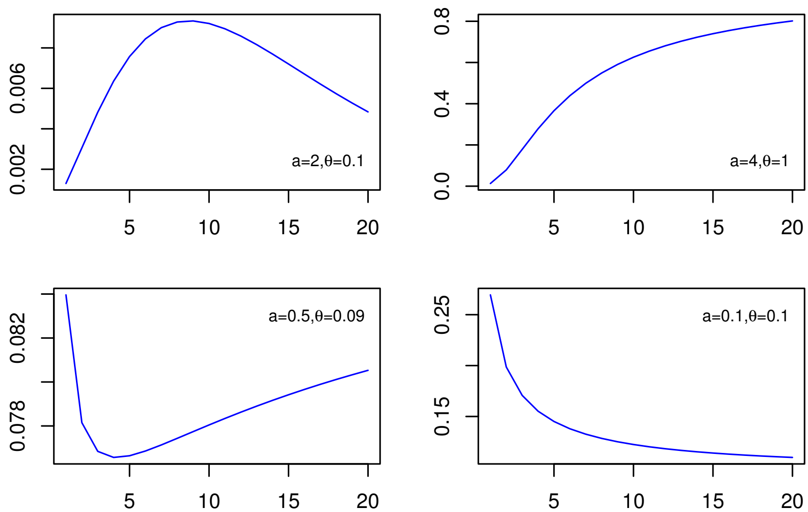

Figure 2 – The hrf is plotted for different values of a and θ.

3.1. Survival and hazard rate function

The survival function of the ZBL random variable is given by,

$$\overline{F}(x) = 1 - \frac{1}{\Gamma(a)}\gamma(a, \log[\frac{1+\theta}{1+\theta+\theta x}e^{\theta x}]). \tag{24}$$

The hazard rate function (hrf) $h(x)$ of a random variable X with cdf $F(x)$ and pdf $f(x)$ is defined as,

$$h(x) = \frac{f(x)}{\overline{F}(x)}, \tag{25}$$

where $\overline{F}(x) = 1 - F(x)$ is the survival function of X. The hrf of ZBL random variable is given by,

$$h(x) = \frac{\frac{1}{\Gamma(a)}[\log(e^{\theta x}\frac{1+\theta}{1+\theta+\theta x})]^{a-1}\frac{\theta^2}{\theta+1}(1+x)e^{-\theta x}}{1 - \frac{1}{\Gamma(a)}\gamma(a, \log[\frac{1+\theta}{1+\theta+\theta x}e^{\theta x}])}. \tag{26}$$

The hrf for ZBL distribution is plotted for different values of a and θ in Figure 2.

From the plot of hrf of ZBL distribution, it is clear that the distribution posses various shapes including increasing, decreasing, bathtub and upside-down bathtub shapes.

3.2. Stress strength reliability

Reliability have wide applications including engineering concepts. Let X_1 and X_2 be two independent random variables following ZBL(a_1, θ) and ZBL(a_2, θ) respectively. Then the stress strength probability or the reliability is defined by $P(X_2 < X_1)$. On the basis of the expressions (15) and (17) we can obtain the expression for reliability as, (see Nadarajah *et al.*, 2015)

$$\begin{aligned} R &= \sum_{j,k=0}^{\infty} C_{j,k} \int_{0}^{\infty} H_{a_2+j}(x) h_{a_1+k}(x) dx \\ &= \sum_{j,k=0}^{\infty} C_{jk} R_{jk}, \end{aligned} \tag{27}$$

where

$$c_{jk} = \frac{\binom{k+1-a_1}{k}}{(a_1+k)\Gamma(a_1-1)} \frac{\binom{j+1-a_2}{j}}{(a_2+j)\Gamma(a_2-1)} I_k(a_1) I_j(a_2),$$

$$I_k(a_1) = \sum_{i=0}^{k} \frac{(-1)^{i+k} p_{i,k}}{(a_1-1-i)} \binom{k}{i}$$

where $p_{i,k}$ is defined in (16) and $R_{j,k} = P(Y_j < Y_k)$ is the reliability between Y_j and Y_k which are distributed as exponentiated Lindley with parameters $a_2 + j$ and $a_1 + k$ respectively. Hence,

$$H_{a_2+j} = \left[1 - \frac{1+\theta+\theta x}{1+\theta} e^{-\theta x}\right]^{a_2+j} \tag{28}$$

$$h_{a_1+k} = \frac{(a_1+k)\theta^2}{\theta+1}(1+x)\left[1 - \frac{1+\theta+\theta x}{1+\theta} e^{-\theta x}\right]^{a_1+k-1} e^{-\theta x} \tag{29}$$

Substituting (28) and (29) in (27) and using (11) and (12) we obtain

$$\begin{aligned} R = & \sum_{p,j,k=0}^{\infty} C_{j,k} \frac{a_1+k}{\theta+1} \theta^2 (-1)^p \binom{a_1+a_2+j+k-1}{p} \\ & \sum_{q=0}^{p} \frac{\binom{p}{q} \theta^q e^{\theta(p+1)}}{(1+\theta)^q} \frac{\Gamma(q+2)}{(\theta(p+1))^{q+2}}. \end{aligned} \tag{30}$$

4. Quantile function

In this section we derive an expression for quantile function of ZBL distribution. The cdf of ZBL distribution is given in (7). Let

$$\begin{aligned} u &= \frac{\gamma(a, \log[\frac{1+\theta}{1+\theta+\theta x} e^{\theta x}])}{\Gamma(a)}, \\ Q^{-1}(a, 1-u) &= \log\left[\frac{1+\theta}{1+\theta+\theta x} e^{\theta x}\right], \end{aligned} \tag{31}$$

where $Q^{-1}(a,u)$ is the inverse gamma regularized function that is, inverse function of $Q(a,x) = 1 - \frac{\gamma(a,x)}{\Gamma(a)}$.
Now (31) can be reduced to,

$$-(1+\theta+\theta x)e^{-(1+\theta+\theta x)} = -(\theta+1)e^{-[1+\theta+Q^{-1}(a,1-u)]} \in (\frac{-1}{e}, 0).$$

That is,

$$-(1+\theta+\theta x) = W_{-1}[-(\theta+1)e^{-[1+\theta+Q^{-1}(a,1-u)]}],$$

where $W_{-1}(x)$ is the negative branch of Lambert-W function. Thus the quantile function of ZBL distribution is given by

$$x = -\left\{1 + \frac{1}{\theta} + \frac{1}{\theta} W_{-1}\left[-(1+\theta)e^{-[1+\theta+Q^{-1}(a,1-u)]}\right]\right\}. \tag{32}$$

Table 2 represents the quantiles of ZBL distribution for selected values of the parameters a and θ.

TABLE 2
The values of quantiles of ZBL distribution for selected values of the parameters.

u	$a=1,\theta=5$	$a=2,\theta=0.5$	$a=3,\theta=2$	$a=4,\theta=6$
0.1	0.025	2.141	0.755	0.333
0.2	0.053	3.053	1.029	0.436
0.3	0.085	3.844	1.262	0.522
0.4	0.121	4.616	1.487	0.605
0.5	0.164	5.421	1.719	0.689
0.6	0.216	6.31	1.972	0.781
0.7	0.283	7.356	2.268	0.888
0.8	0.377	8.713	2.648	1.024
0.9	0.534	10.837	3.236	1.234

5. Maximum likelihood estimation

Let $X_1, X_2, ..., X_n$ be a random sample taken from ZBL(a,θ). The likelihood function is given by

$$L = \frac{\theta^{2n}}{(\theta+1)^n}\frac{1}{(\Gamma(a))^n} e^{-\theta\sum_{i=0}^{n} x_i} \prod_{i=1}^{n}\left(\theta x_i + \log\frac{1+\theta}{1+\theta+\theta x_i}\right)^{a-1}(1+x_i) \tag{33}$$

and the log-likelihood function is given by,

$$\begin{aligned}\log L &= 2n\log\theta - n\log(\theta+1) - n\log(\Gamma(a)) - \theta\sum_{i=0}^{n} x_i \\ &+ (a-1)\sum_{i=1}^{n}\log\left(\theta x_i + \log\frac{1+\theta}{1+\theta+\theta x_i}\right) + \log(1+x_i).\end{aligned} \tag{34}$$

The normal equations are thus obtained by,

$$\begin{aligned}\frac{\partial \log L}{\partial\theta} = 0 \implies & \frac{2n}{\theta} - \frac{n}{\theta+1} - \sum_{i=0}^{n} x_i + (a-1) \\ & \frac{\partial}{\partial\theta}\left\{\sum_{i=0}^{n}\log\left[\theta x_i + \log\frac{1+\theta}{1+\theta+\theta x_i}\right]\right\} = 0 \\ \implies & \frac{2n}{\theta} - \frac{n}{\theta+1} - \sum_{i=0}^{n} x_i + (a-1)\sum_{i=0}^{n}\frac{1}{\theta x_i + \log\frac{1+\theta}{1+\theta+\theta x_i}} \\ & \left[\frac{\theta x_i(3\theta+3+\theta x_i^2 + x_i + 3\theta x_i + \theta^2 + \theta^2 x_i^2 + 2\theta^2 x_i)}{(1+\theta+\theta x_i)^2(1+\theta)}\right] = 0.\end{aligned} \tag{35}$$

$$\frac{\partial \log L}{\partial a} = 0 \implies -n\frac{d}{da}[\log\Gamma(a)] + \sum_{i=0}^{n}\log\left(\theta x_i + \log\frac{1+\theta}{1+\theta+\theta x_i}\right) = 0. \tag{36}$$

Solving (35) and (36) using mathematical softwares like `MATHEMATICA`, `MATHCAD` and `R` we can obtain the maximum likelihood estimates (MLEs) of the parameters a and θ.

5.1. *Observed Fisher information matrix*

The Fisher Information matrix is required for interval estimation of the parameters a and θ. The observed Fisher information matrix of $\hat{\lambda} = (\hat{a}, \hat{\theta})$ can be expressed as

$$I(\lambda) = \begin{bmatrix} I_{aa} & I_{a\theta} \\ I_{a\theta} & I_{\theta\theta}\end{bmatrix},$$

where

$$
\begin{aligned}
I_{aa} &= \frac{\partial^2 L}{\partial a^2} = n\psi'(\hat{a}),\ \psi'(\hat{a}) = \frac{\partial^2 \log(\Gamma(\hat{a}))}{\partial \hat{a}^2} \text{ is the trigamma function,}\\
I_{a\theta} &= -\sum_{i=1}^{n} \frac{\theta x_i(2+x_i+\theta+\theta x_i)}{(1+\theta)(1+\theta+\theta x_i)(\theta x_i+\log(\frac{1+\theta}{1+\theta+\theta x_i}))} \text{ and}\\
I_{\theta\theta} &= \frac{2n}{\theta^2} - \frac{n}{(\theta+1)^2} + (a-1)\sum_{i=1}^{n} x_i \big[\theta x_i(-2+2\theta+4\theta^2+\theta^3+\theta x_i^2(1+\theta)^2\\
&\quad + x_i(-1+2\theta+6\theta^2+2\theta^3)) - (2+x_i+2\theta+2\theta x_i)\log(\frac{1+\theta}{1+\theta+\theta x_i})\Big].
\end{aligned}
$$

5.2. *Asymptotic confidence intervals*

For large n, the asymptotic distribution of $\hat{\lambda} = (\hat{a}, \hat{\theta})$ is given by

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N_2(0, I^{-1}(\lambda)). \tag{37}$$

Thus the properties of $\hat{\lambda}$ can be derived based on this normal approximation. The 100(1-α)% confidence intervals of the parameters a and θ is thus obtained by

$$\hat{a} \pm Z_{\frac{\alpha}{2}} \sqrt{I_{aa}^{-1}(\hat{\lambda})} \tag{38}$$

and

$$\hat{\theta} \pm Z_{\frac{\alpha}{2}} \sqrt{I_{\theta\theta}^{-1}(\hat{\lambda})}, \tag{39}$$

where $I_{aa}^{-1}(\hat{\lambda})$ and $I_{\theta\theta}^{-1}(\hat{\lambda})$ are the diagonal elements of $I^{-1}(\hat{\lambda})$ and $Z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}^{\text{th}}$ percentile of standard normal distribution.

6. Monte Carlo simulation

The Monte Carlo simulation was done in order to prove the efficiency of the model. The estimates were calculated for true values of parameters ($a = 1, \theta = 5$) and ($a = 2, \theta = 10$) for N=1000 samples of sizes 25, 50, 100, 200, 400 and 800 and the following quantities are computed.

1. Mean of the MLE of parameters, $\hat{a} = \frac{1}{N}\sum_{i=1}^{N} \hat{a}_i$ and $\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}_i$.

2. Average bias of MLEs of parameters, $\text{bias}(a) = \frac{1}{N}\sum_{i=1}^{N}(\hat{a}_i - a)$ and $\text{bias}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)$.

3. Mean square error (MSE) of MLEs of parameters, $\text{MSE}(a) = \frac{1}{N}\sum_{i=1}^{N}(\hat{a}_i - a)^2$ and $\text{MSE}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2$.

4. Root mean square error (RMSE) of MLEs of parameters, $\text{RMSE}(a) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{a}_i - a)^2}$ and $\text{RMSE}(\theta) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i - \theta)^2}$.

Table 3 and Table 4 show the mean MLEs of the parameters with their bias, MSEs and RMSEs.

TABLE 3
The simulation results for $a = 1, \theta = 5$.

Parameter	n	Mean	Average bias	MSE	RMSE
a	25	1.107	0.107	0.110	0.332
	50	1.057	0.057	0.043	0.206
	100	1.023	0.023	0.019	0.137
	200	1.014	0.014	0.008	0.089
	400	1.008	0.008	0.004	0.067
	800	1.004	0.004	0.002	0.047
θ	25	5.673	0.673	3.738	1.934
	50	5.351	0.351	1.374	1.172
	100	5.115	0.115	0.562	0.749
	200	5.068	0.068	0.241	0.491
	400	5.046	0.046	0.136	0.369
	800	5.023	0.023	0.067	0.259

TABLE 4
The simulation results for $a = 2, \theta = 10$.

Parameter	n	Mean	Average Bias	MSE	RMSE
a	25	2.249	0.249	0.565	0.752
	50	2.092	0.092	0.170	0.412
	100	2.045	0.045	0.088	0.296
	200	2.026	0.026	0.345	0.186
	400	2.013	0.013	0.0181	0.134
	800	2.007	0.007	0.009	0.097
θ	25	11.408	0.408	15.691	3.961
	50	10.511	0.511	4.981	2.232
	100	10.297	0.297	2.468	1.157
	200	10.178	0.178	0.995	0.998
	400	10.091	0.091	0.504	0.710
	800	10.043	0.043	0.256	0.506

7. Application

In this section two real data sets are considered in order to prove the applicability of the proposed distribution by comparing it with other classical distributions. The first data set given in Table 5 represents the waiting times (in minutes) before service of 100 Bank customers and examined and analysed by Ghitany *et al.* (2008) for fitting the Lindley (1958) distribution.

And the second one given in Table 6 is the failure stresses of single carbon fibres of length 50mm, originally proposed by Bader and Priest (1982). For the two datasets, the estimates of the parameters, -log likelihood (-log L), Akaike Information criterion (AIC), Bayesian Information criterion (BIC), Corrected Akaike Information Criterion (AICc) and the Kolmogorov Smirnov (K-S) statistic values and p value are calculated. Table 7 and 8 gives the results obtained from ZBL, Generalized Lindley (GL) (see Zakerzadeh and Dolati, 2009), Extended Inverse Lindley (EIL) (see Alkarni, 2015) , Generalized Inverse Lindley (GIL) (see Sharma *et al.*, 2016), Power Lindley (PL) (see, Ghitany *et al.*, 2013), Marshall-Olkin Extended Lindley (MOEL) (see Ghitany *et al.*, 2012) , and Lindley (L) (see Lindley, 1958) distributions to the datasets. The plots of fitted densities and cumulative densities with respective to the given data sets are also plotted.

From Tables 7 and 8, we can see that ZBL gives a better fit to the data than other classical distributions considered here, since ZBL distribution have the smallest goodness of fit statistics and model adequacy measures.

Figures 3 and 4 depict the density and cumulative density plot which compares the fitted densities of the models with the empirical histogram and cumulative density of the real data.

TABLE 5
Data of waiting times of 100 bank customers.

0.08	2.09	3.48	4.87	6.94	8.66	13.11	23.63	0.20	2.23	3.52
4.98	6.97	9.02	13.29	0.40	2.26	3.57	5.06	7.09	9.22	13.80
25.74	0.50	2.46	3.64	5.09	7.26	9.47	14.24	25.82	0.51	2.54
3.70	5.17	7.28	9.74	14.76	6.31	0.81	2.62	3.82	5.32	7.32
10.06	14.77	32.15	2.64	3.88	5.32	7.39	10.34	14.83	34.26	0.90
2.69	4.18	5.34	7.59	10.66	15.96	36.66	1.05	2.69	4.23	5.41
7.62	10.75	16.62	43.01	1.19	2.75	4.26	5.41	07.63	17.12	46.12
1.26	2.83	4.33	5.49	7.66	11.25	17.14	79.05	1.35	2.87	5.62
7.87	11.64	17.36	1.40	3.02	4.34	5.71	7.93	11.79	18.10	1.46
4.4	5.85	8.26	11.98	19.13	1.76	3.25	4.50	6.25	8.37	12.02
2.02	3.31	4.51	6.54	8.53	12.03	20.28	2.02	3.36	6.76	12.07
21.73	2.07	3.36	6.93	8.65	12.63	22.69				

TABLE 6
Data of failure stresses of single carbon fibres.

1.339	1.434	1.549	1.574	1.589	1.613	1.746	1.753	1.764
1.807	1.812	1.84	1.852	1.852	1.862	1.864	1.931	1.952
1.974	2.019	2.051	2.055	2.058	2.088	2.125	2.162	2.171
2.172	2.18	2.194	2.211	2.27	2.272	2.28	2.299	2.308
2.335	2.349	2.356	2.386	2.39	2.41	2.43	2.431	2.458
2.471	2.497	2.514	2.558	2.577	2.593	2.601	2.604	2.620
2.633	2.670	2.682	2.699	2.705	2.735	2.785	3.02	3.042
3.116	3.174							

TABLE 7
Estimates and statistics for the data of waiting times of 100 customers.

Model	Estimates	-log L	AIC	BIC	AICc	K-S (p-value)
ZBL	$\hat{a}$=1.263 $\hat{\theta}$=0.223	317.876	639.751	644.961	639.875	0.050 (0.962)
GL	$\hat{\theta}$=0.23 $\hat{\alpha}$=1.491 $\hat{\beta}$=0.814	317.836	641.673	649.488	641.923	0.0511 (0.957)
EIL	$\hat{\theta}$=6.542 $\hat{\alpha}$=0.011 $\hat{\beta}$=1.163	334.381	674.762	682.578	675.012	0.117 (0.131)
GIL	$\hat{\theta}$=7.229 $\hat{\alpha}$=1.152	334.779	673.558	678.768	673.682	0.118 (0.124)
PL	$\hat{\theta}$=0.153 $\hat{\alpha}$=1.083	318.319	640.637	645.848	640.761	0.052 (0.95)
MOEL	$\hat{\theta}$=0.208 $\hat{\alpha}$=1.243	318.914	641.827	647.037	641.951	0.057 (0.906)
L	$\hat{\theta}$=0.187	319.037	640.075	642.68	640.116	0.068 (0.749)

TABLE 8
Estimates and statistics for the data of the failure stresses of single carbon fibres.

Model	Estimates	-log L	AIC	BIC	AICc	K-S (p-value)
ZBL	$\hat{a}$=25.715 $\hat{\theta}$=11.938	34.839	73.677	78.026	73.871	0.069 (0.915)
GL	$\hat{\theta}$=12.828 $\hat{\alpha}$=28.135 $\hat{\beta}$=24.003	35.065	76.13	82.653	76.523	0.072 (0.886)
EIL	$\hat{\theta}$=31.725 $\hat{\alpha}$=0.01 $\hat{\beta}$=4.995	43.860	93.720	100.243	94.114	0.125 (0.262)
GIL	$\hat{\theta}$=32.637 $\hat{\alpha}$=4.994	43.875	91.098	96.098	91.943	0.125 (0.260)
PL	$\hat{\theta}$=0.052 $\hat{\alpha}$=4.219	34.978	73.955	78.309	74.149	0.075 (0.905)
MOEL	$\hat{\theta}$=4.456 $\hat{\alpha}$=786	35.993	75.985	80.334	76.179	0.077 (0.838)
L	$\hat{\theta}$=0.707	106.999	215.999	218.173	216.062	0.42 (2.207×10^{-10})

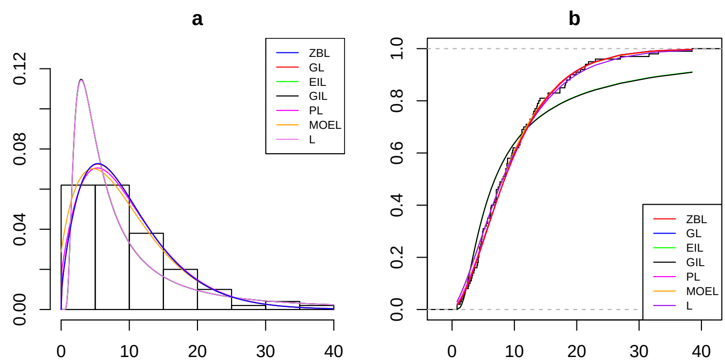

Figure 3 – Fitted densities (a) and cumulative densities (b) of data of waiting times of 100 customers.

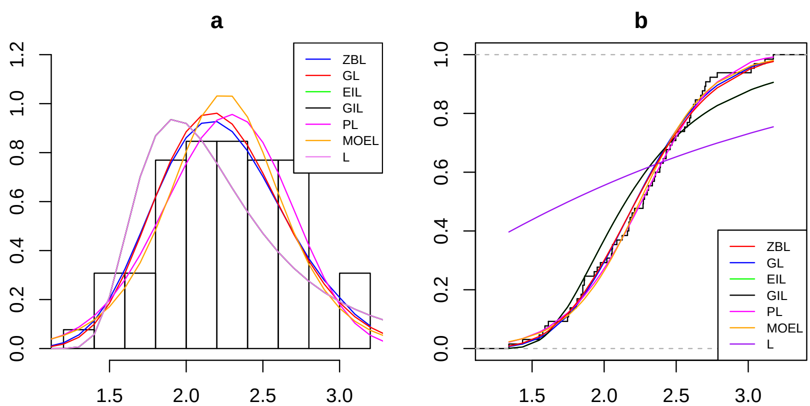

Figure 4 – Fitted densities (a) and cumulative densities (b) of data of the failure stresses of single carbon fibres.

We can see that the ZBL distribution is closer to the empirical histogram and the empirical cumulative densities than the fits of other classical distributions considered here. In summary, the new ZBL distribution may be an interesting alternative to the other models available in the literature for modelling positive real life data sets.

Acknowledgements

The authors express their gratefulness to the learned referee for many of the valuable comments and suggestions, which lead the way to improvements and thus acquire current version of the paper. First author acknowledges the Cochin University of Science and Technology for providing financial support in the form of seed money for new research initiatives. (No.PL.(UGC)I/SPG/SMNRI/2018- 2019).

REFERENCES

S. H. ALKARNI (2015). *Extended inverse Lindley distribution: Properties and application*. SpringerPlus, 4, no. 1, pp. 1–13.

M. G. BADER, A. M. PRIEST (1982). *Statistical aspects of fibre and bundle strength in hybrid composites*. Progress in Science and Engineering of Composites, pp. 1129–1136.

G. M. CORDEIRO, E. M. M. ORTEGA, D. C. C. DA CUNHA (2013). *The exponentiated generalized class of distributions*. Journal of Data Science, 11, no. 1, pp. 1–27.

A. F. FAGBAMIGBE, P. MELAMU, B. O. OLUYEDE, B. MAKUBATE (2018). *The Ristić and Balakrishnan Lindley-Poisson distribution: Model, theory and application*. Afrika Statistika, 13, no. 4, pp. 1837–1864.

M. E. GHITANY, D. K. AL-MUTAIRI, F. A. AL-AWADHI, M. M. AL-BURAIS (2012). *Marshall-Olkin extended Lindley distribution and its application*. International Journal of Applied Mathematics, 25, no. 5, pp. 709–721.

M. E. GHITANY, D. K. AL-MUTAIRI, N. BALAKRISHNAN, L. J. AL-ENEZI (2013). *Power Lindley distribution and associated inference*. Computational Statistics & Data Analysis, 64, pp. 20–33.

M. E. GHITANY, F. ALQALLAF, D. K. AL-MUTAIRI, H. A. HUSAIN (2011). *A two-parameter weighted Lindley distribution and its applications to survival data*. Mathematics and Computers in simulation, 81, no. 6, pp. 1190–1201.

M. E. GHITANY, B. ATIEH, S. NADARAJAH (2008). *Lindley distribution and its application*. Mathematics and Computers in Simulation, 78, no. 4, pp. 493–506.

E. GÓMEZ-DÉNIZ, E. CALDERÍN-OJEDA (2011). *The discrete Lindley distribution: Properties and applications*. Journal of Statistical Computation and Simulation, 81, no. 11, pp. 1405–1416.

I. S. GRADSHTEYN, I. M. RYZHIK (2014). *Table of Integrals, Series, and Products*. Academic Press, London.

G. HAMEDANI (2013). *The Zografos-Balakrishnan log-logistic distribution: Properties and applications*. Journal of Statistical Theory and Applications, 12, no. 3, pp. 225–244.

M. R. IRSHAD, R. MAYA (2017). *Extended version of generalised Lindley distribution*. South African Statistical Journal, 51, no. 1, pp. 19–44.

D. V. LINDLEY (1958). *Fiducial distributions and Bayes' theorem*. Journal of the Royal Statistical Society: Series B, 20, no. 1, pp. 102–107.

R. MAYA, M. R. IRSHAD (2017). *Generalized Stacy-Lindley mixture distribution*. Afrika Statistika, 12, no. 3, pp. 1447–1465.

S. NADARAJAH, H. S. BAKOUCH, R. TAHMASBI (2011). *A generalized Lindley distribution*. Sankhya B, 73, no. 2, pp. 331–359.

S. NADARAJAH, G. M. CORDEIRO, E. M. M. ORTEGA (2015). *The Zografos–Balakrishnan-G family of distributions: Mathematical properties and applications*. Communications in Statistics-Theory and Methods, 44, no. 1, pp. 186–215.

M. SANKARAN (1970). *The discrete Poisson-Lindley distribution*. Biometrics, 26, no. 1, pp. 145–149.

V. K. SHARMA, S. K. SINGH, U. SINGH, F. MEROVCI (2016). *The generalized inverse Lindley distribution: A new inverse statistical model for the study of upside-down bathtub data*. Communications in Statistics-Theory and Methods, 45, no. 19, pp. 5709–5729.

D. S. SHIBU, M. R. IRSHAD (2016). *Extended new generalized Lindley distribution*. Statistica, 76, no. 1, pp. 41–56.

H. ZAKERZADEH, A. DOLATI (2009). *Generalized Lindley distribution*. Journal of Mathematical Extension, 3, no. 2, pp. 13–25.

K. ZOGRAFOS, N. BALAKRISHNAN (2009). *On families of beta-and generalized gamma-generated distributions and associated inference*. Statistical methodology, 6, no. 4, pp. 344–362.

SUMMARY

The Lindley distribution was proposed in the context of Bayesian statistics as a counter example of fiducial statistics. In this paper, we propose the Zografos-Balakrishnan Lindley (ZBL) distribution in which Lindley distribution is a special case. Some properties of the new distribution is obtained such as moments, hazard rate function, stress strength reliability etc. The parameters are estimated using the method of maximum likelihood. Finally an application of the proposed distribution to two real data sets is illustrated and it is concluded that ZBL distribution provides better fit than other classical distributions.

Keywords: Lindley distribution; Hazard rate function; Stress strength reliability; Moments; Quantile function; Maximum likelihood estimation; Simulation.