REVSTAT – Statistical Journal
Volume 0, Number 0, Month 0000, 000-000

A Study on Zografos-Balakrishnan Log-Normal Distribution: Properties and Application to Cancer Dataset

Authors: D. S. Shibu
– Department of Statistics, University College,
Thiruvananthapuram, India - 695034 (statshibu@gmail.com)

S. L. Nitin ✉*
– Department of Statistics, University College,
Thiruvananthapuram, India - 695034 (nitinstat24@gmail.com)

M. R. Irshad
– Department of Statistics, CUSAT,
Cochin, India - 682022 (irshadm24@gmail.com)

Received: Month 0000 Revised: Month 0000 Accepted: Month 0000

Abstract:

- In this article, we studied a generalization of the log-normal distribution called Zografos-Balakrishnan log-normal distribution, and investigate its various important properties and functions including moments, quantile function, various reliability measures, Rényi entropy, and some inequality measures. The estimation of unknown parameters is discussed by the methods of maximum likelihood, and the Bayesian technique and their simulation studies are also carried out. The applicability of the distribution is illustrated utilizing a real dataset. A likelihood ratio test is utilized for testing the efficiency of the third parameter. The effectiveness of this model for the dataset is also established using the parametric bootstrap approach.

Key-Words:

- *Zografos-Balakrishnan-G family; reliability measures; maximum likelihood estimation; Bayesian estimation; bootstrap confidence interval; likelihood ratio test.*

AMS Subject Classification:

- 60E05, 62F10, 62F15.

*Corresponding author

1. INTRODUCTION

Recently, an increasing interest can be observed for the art of adding parameters to some well-known existing distributions for getting different shapes of hazard rate or failure rate functions for applying it in various real-life situations and also for analyzing data with a high degree of skewness and kurtosis. In principle, the log-normal distribution is defined as the continuous probability distribution of a random variable whose logarithm is normally distributed. It is one of the most widely used distributions for asymmetric datasets. Thus, it has been widely applied in many different aspects of life sciences, including biology, geology, ecology, and meteorology as well as in economics, finance, and risk analysis (see [15]), and also attracts attention quite often in environmental sciences, physics, astrophysics, and cosmology (see [3], [4], [22]).

On many occasions, the significance of the LN distribution in biological science has been acknowledged. Bentley (1954) ([9]) provides numerous generic resources for statistical data generated from biological and agricultural sources. A study on the complexities of the biochemical mechanisms associated with gene expression has created an emergent LN distribution of expression levels, according to [2]. Carvalho (2018) ([5]) found that a form of the LN distribution fits the postpartum blood loss data from numerous geographical areas quite well, suggesting that the LN distribution may fit postpartum blood loss generally. Hence, in this article, we utilize a cancer dataset as an application that is related to biological science.

The probability density function (pdf) for a log-normal random variable W is given by

$$q(w) = \frac{1}{\sqrt{2\pi}\sigma w}\, exp\left[-\frac{(\log w - \mu)^2}{2\sigma^2}\right], \quad w > 0,\ \mu \in \mathbb{R},\ \sigma > 0.$$

Zografos and Balakrishnan (2009) (see [26]) proposed a novel family of univariate distributions generated by gamma random variables. Further Nadarajah et. al.(2015) (see [20]) provides a comprehensive treatment of the general mathematical properties of this family and denote it with the prefix "Zografos-Balakrishnan-G" or "ZB-G" distributions. They discuss the estimation of parameters by maximum likelihood and provide an application to a real dataset and also propose a bivariate generalization. For any baseline cumulative distribution function (cdf) $G(x)$, and $x \in \mathbb{R}$, Zografos and Balakrishnan (2009)([26]) defined a distribution with pdf $f(x)$ and cumulative distribution function (cdf) $F(x)$ given by

$$f(x) = \frac{1}{\Gamma(\alpha)}\{-\log[1 - G(x)]\}^{\alpha-1} g(x), \tag{1.1}$$

and

$$F(x) = \frac{\gamma(\alpha, -\log[1 - G(x)])}{\Gamma(\alpha)} = \frac{1}{\Gamma(\alpha)} \int_0^{-\log[1-G(x)]} t^{\alpha-1} \exp(-t)dt, \tag{1.2}$$

respectively for $\alpha > 0$, where $g(x) = dG(x)/dx$, $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}\exp(-t)dt$ denotes the gamma function, and $\gamma(\alpha, z) = \int_0^\infty t^{\alpha-1}\exp(-t)dt$ denotes the incomplete gamma function. The corresponding hazard rate function (hrf) is

$$h(x) = \frac{\{-\log[1-G(x)]\}^{\alpha-1} g(x)}{\Gamma(\alpha, -\log[1-G(x)])},$$

where $\Gamma(\alpha, z) = \int_0^z t^{\alpha-1}\exp(-t)dt$ denotes the complementary incomplete gamma function.

Moreover, using the generalization in (1.1) and considering the immense applicability of the log-normal distribution, Nadarajah et. al.(2015)([20]) also suggests the generalization of log-normal distribution called Zografos-Balakrishnan log-normal (ZBLN) distribution. However, little is known in terms of general mathematical properties and in terms of application for this generalization.

The aim of this article is to derive some mathematical properties of Zografos-Balakrishnan log-normal distribution in the most simple, explicit and general forms and apply it to biological sciences and other reliability analyses. The main motivation for considering this lifetime model is to study the flexibility of the distribution that can be used to model lifetime data in a wider class of biological data and reliability problems.

The rest of the paper is organized as follows. In section-2, we present the definition of the ZBLN distribution and obtain the weighted form of the same. The moments of the distribution are obtained in Section-3. The quantile function and some of its associated measures are obtained in Section-4. The various functions and the moments related to the reliability measures are discussed in Section-5. Section-6 deals with the derivation of the Rényi entropy, and Section-7 deals with the discussion of some inequality measures associated with the ZBLN distribution. The distributions of order statistics are derived in Section-8. In order to estimate the unknown parameters of the ZBLN model, the method of maximum likelihood estimation, and the Bayesian estimation procedure are employed, and also a parametric bootstrap method of simulation is presented in Section-9. To analyze the longstanding performances of maximum likelihood estimators, and the Bayesian estimators of the parameters, a simulation study has been conducted in Section-10. To illustrate the potentiality of the ZBLN distribution over competing distributions, one real dataset is analyzed in Section-11. The final concluding remarks are given in Section-12.

2. DEFINITION OF THE DISTRIBUTION

In this section, we present the definition and some important features of the ZBLN distribution.

Definition 2.1. Let X be a random variable which follows ZBLN distribution (see [20]) with parameters α, μ and σ, then its pdf is given by

$$f(x) = \frac{1}{\sigma\ x\ \Gamma(\alpha)} \left\{ -\log \left[1 - \Phi\left(\frac{\log x - \mu}{\sigma} \right) \right] \right\}^{\alpha - 1} \phi\left(\frac{\log x - \mu}{\sigma} \right), \tag{2.1}$$

and the cdf is given by

$$\begin{aligned} F(x) &= \frac{\gamma\left(\alpha, -\log\left[1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]\right)}{\Gamma(\alpha)} \\ &= \frac{1}{\Gamma(\alpha)} \int_0^{-\log\left[1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]} t^{\alpha - 1} \exp(-t) dt, \end{aligned} \tag{2.2}$$

where $x > 0$, $\mu \in \mathbb{R}$ and $\alpha, \sigma > 0$. Also, $\Phi(.)$ and $\phi(.)$ are respectively the cdf and pdf of the standard normal distribution.

Note that, ZBLN distribution reduces to the two-parameter log-normal if $\alpha = 1$. The plot in Figure 1 portrays the pdf of ZBLN distribution, and we observe that the pdf may be decreasing and unimodal with a certain flexibility in the mode and tails. It is, however, mainly right-skewed or almost symmetrical.

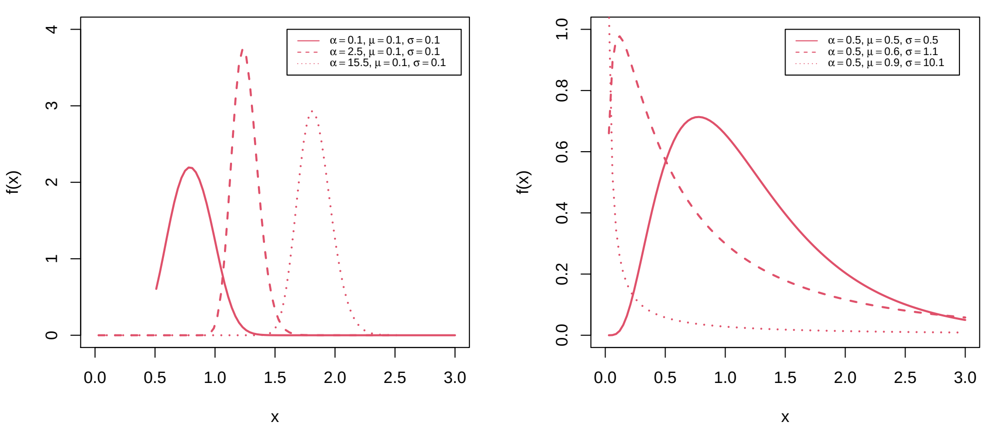

Figure 1: Plots of pdf of the ZBLN distribution

2.1. Expansions for pdf and cdf

Nadarajah et. al. (2015)([20]) derived some useful expansions for (1.1) and (1.2) using the concept of exponentiated distributions. For an arbitrary baseline cdf $G(x)$, a random variable is said to have the exponentiated-G distribution with parameter $\alpha > 0$, say $X \sim \exp\text{-}G(\alpha)$, if its pdf and cdf are respectively given by

$$f^*_\alpha(x) = \alpha G^{\alpha-1}(x) g(x), \quad \text{and} \quad F^*_\alpha(x) = G^\alpha(x).$$

The important properties of exponentiated distributions have been studied by several authors; for examples, see [18] for exponentiated Weibull, [10] for exponentiated Pareto, [12] for exponentiated exponential, [19] for exponentiated Gumbel and [21] for exponentiated gamma distributions.
Note that, for any real parameter $\alpha > 0$, the following formula holds.

$$\{-\log[1 - G(x)]\}^{\alpha-1} = (\alpha - 1) \sum_{k=0}^{\infty} \binom{k+1-\alpha}{k} \sum_{j=0}^{k} \binom{k}{j} \frac{(-1)^{j+k}\ p_{j,k}}{(\alpha - 1 - j)} \{G(x)\}^{\alpha+k-1},$$

where the constants $p_{j,k}$ can be calculated recursively through the relation,

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

for $k = 1, 2, ...$ with $p_{j,0} = 1$ and $c_k = (-1)^{k+1}(k+1)^{-1}$. Thus, Nadarajah et. al. (2015)([20]) demonstrated that (1.1), and the corresponding (1.2) can be expressed as

$$f(x) = \sum_{k=0}^{\infty} b_k\ f^*_{\alpha+k}(x), \quad \text{and} \quad F(x) = \sum_{k=0}^{\infty} b_k\ F^*_{\alpha+k}(x),$$

where $f^*_{\alpha+k}(x)$ and $F^*_{\alpha+k}(x)$ respectively denotes the corresponding pdf and cdf of the exp-$G(\alpha + k)$ distribution and for any real parameter $\alpha > 0$, and

$$b_k = \frac{\binom{k+1-\alpha}{k}}{(\alpha + k)\Gamma(\alpha - 1)} \sum_{j=0}^{k} \binom{k}{j} \frac{(-1)^{j+k}\ p_{j,k}}{(\alpha - 1 - j)}. \tag{2.3}$$

Thus, the cdf and pdf of the ZBLN distribution respectively obtained as

$$F(x) = \sum_{k=0}^{\infty} b_k \left[\Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]^{\alpha+k}, \tag{2.4}$$

and

$$f(x) = \sum_{k=0}^{\infty} b_k \frac{(\alpha + k)}{\sigma x} \phi\left(\frac{\log x - \mu}{\sigma}\right) \left[\Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]^{\alpha+k-1}. \tag{2.5}$$

Thus, ZBLN distribution can be expressed as the infinite weighted sum of Exponentiated log-normal distributions indexed by power parameter $\alpha + k$.

3. MOMENTS

In this section, we derive the expression for the r^{th} raw moment of ZBLN distribution. From Equation (2.5), the moments of the ZBLN distribution can be written as the weighted sum of probability-weighted moments of the log-normal distribution. Thus, the r^{th} raw moment of the distribution is given by

$$\mu_r' = E(X^r) = \sum_{k=0}^{\infty}(\alpha+k)b_k\ \mu_{r,\alpha+k}',$$

where

$$\mu_{r,\alpha+k}' = E\left\{X^r\left[\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]^{\alpha+k-1}\right\}$$

$$\Rightarrow \qquad \mu_{r,\alpha+k}' = \int_0^{\infty}\frac{x^r}{\sigma x}\,\phi\left(\frac{\log x-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]^{\alpha+k-1}dx,$$

is the probability weighted moments of the log-normal distribution.

4. QUANTILE FUNCTION AND ASSOCIATED MEASURES

Generally, a probability distribution can be specified either in terms of the distribution function or by the quantile function. Quantile functions have several interesting properties that are not shared by distributions, which makes them more convenient and flexible for analysis. Moreover, the random numbers from any distribution can be generated using appropriate quantile functions. So, in this section, we derive an explicit expression for the quantile function of ZBLN distribution and some of its associated measures.

Theorem 4.1. *If X follows ZBLN distribution as given in (2.2), then the p^{th} quantile, $Q_p = F^{-1}(p)$ of the distribution is given by*

$$Q_p = \exp\left\{\mu+\sigma\ \Phi^{-1}\left[1-\exp\left(-Q^{-1}(\alpha,1-p)\right)\right]\right\},$$

where $\Phi^{-1}(.)$ is the quantile function of standard normal variate.

Proof: For the ZBLN distribution, Q_p is the solution of the equation

$$Q\left(\alpha,-\log\left[1-\Phi\left(\frac{\log(Q_p)-\mu}{\sigma}\right)\right]\right) = 1-u, \quad p\in(0,1)$$

$$\Rightarrow -\log\left[1-\Phi\left(\frac{\log(Q_p)-\mu}{\sigma}\right)\right] = Q^{-1}(\alpha,1-p) \tag{4.1}$$

On simplifications, (4.1) reduces to

$$\Phi\left(\frac{\log(Q_p)-\mu}{\sigma}\right)=1-\exp\left(-Q^{-1}(\alpha,1-p)\right)$$

$$\Rightarrow \frac{\log(Q_p)-\mu}{\sigma}=\Phi^{-1}\left[1-\exp\left(-Q^{-1}(\alpha,1-p)\right)\right]$$

$$\Rightarrow Q_p=\exp\left\{\mu+\sigma\ \Phi^{-1}\left[1-\exp\left(-Q^{-1}(\alpha,1-p)\right)\right]\right\}. \tag{4.2}$$

□

Remark 4.1. Since $\Phi^{-1}(.)$ is the quantile function of standard normal variate, Q_p in Equation (4.2) also written in the form

$$Q_p=\exp\left\{\mu+\sigma\sqrt{2}\ \operatorname{erf}^{-1}\left[1-2\exp\left(-Q^{-1}(\alpha,1-p)\right)\right]\right\}, \tag{4.3}$$

where $\operatorname{erf}^{-1}(.)$ is the inverse error function.

Now, by putting $p=0.5$, in Equation (4.3), we get the median (M) of ZBLN distribution and is given by

$$\mathrm{M}=Q_{0.5}=\exp\left\{\mu+\sigma\sqrt{2}\ \operatorname{erf}^{-1}\left[1-2\exp\left(-Q^{-1}(\alpha,1/2)\right)\right]\right\}.$$

For $p=1/4$ and $p=3/4$, Equation (4.3) respectively gives first and third quartiles of the ZBLN distribution.

5. RELIABILITY MEASURES

Many domains of practical studies, such as physics, engineering, psychology, and others, rely heavily on reliability measures. As a reason, providing expressions for various reliability measures is critical. Due to these facts, in this section, we derive expressions for various measures of reliability.

5.1. Hazard rate function

The hazard rate provides the instantaneous risk that the event of interest happens, within a very narrow time frame. As a function of age x, the hazard rate function is also referred to as the failure rate function, instantaneous death rate, force of mortality, and intensity function in other areas of study like survival analysis, actuarial science, biosciences, demography, and extreme value theory.

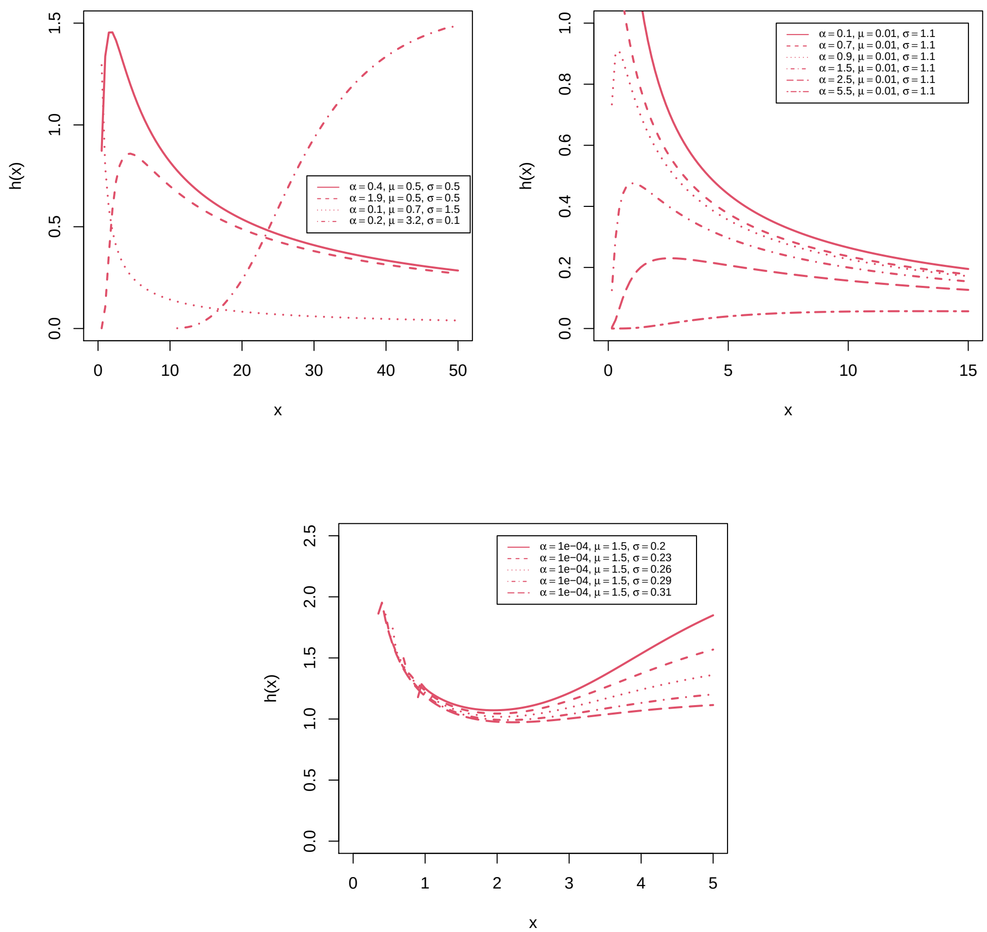

Figure 2: Plots of the hazard rate function of the ZBLN distribution

Thus, it also plays a substantial role in lifetime data analysis, mainly in survival and reliability studies. Indeed, the mathematical characterization of a lifetime distribution for a certain life phenomenon can be made on the basis of its failure rate pattern. Most commonly, the hazard function can be increasing, decreasing, upside-down bathtub or bathtub shaped.

By definition, the hazard function $h(x)$ can be defined as $h(x) = f(x)/S(x)$, where $S(x) = 1 - F(x)$ is the survival function. Obviously, the survival function of ZBLN distribution is given as

$$S(x) = 1 - \frac{\gamma\left(\alpha, -\log\left[1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]\right)}{\Gamma(\alpha)}.$$

Thus, the hazard function of ZBLN distribution is obtained as

$$h(x) = \frac{\phi\left(\frac{\log x-\mu}{\sigma}\right)\left\{-\log\left[1-\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}}{\sigma x\;\Gamma\left(\alpha, -\log\left[1-\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]\right)},$$

where $\Gamma(\alpha, z) = \int_z^{\infty} t^{\alpha-1}\exp(-t)dt$ denotes the complementary incomplete gamma function. Also, plots in Figure 3 refers the hazard rate function and observed that ZBLN distribution possess increasing, decreasing, bathtub, and upside-down bathtub shapes. In this scenario, the capability of our model to construct a bathtub-shaped failure rate function with a significantly longer flat region is one of its unique advantages. Nonetheless, this region is crucial in real-world applications, underscoring the importance of effective flat region modeling (see [14]). Again, from Figure 2, it can be seen in further detail that the hazard rate function graph for the shape bathtub happens when $\alpha = 0.0001$, $\mu = 1.5$, $0.2 \le \sigma \le 0.31$. When $\alpha \ge 0.1$, $\mu = 0.01$, and $\sigma = 1.1$, the shapes also change from decreasing to increasing via an upside-down bathtub.

5.2. Cumulative hazard rate function

The cumulative hazard rate function, also known as the integrated hazard function, is the overall number of failures or deaths over a period of time. Like the hazard function, the cumulative hazard function $H(x)$ is not a probability, but still a measure of risk. The greater the value of $H(x)$, the greater the risk of failure by time x.

By definition, $H(x) = -\log\{S(x)\}$. Thus, the cumulative hazard rate function of ZBLN distribution is given by

$$H(x) = -\log\left\{1 - \frac{\gamma\left(\alpha, -\log\left[1-\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]\right)}{\Gamma(\alpha)}\right\}. \tag{5.1}$$

Note that, $\log(1-x) = -\sum_{n=1}^{\infty}\frac{x^n}{n}$, and also from Equation (2.4), $H(x)$ in Equation (5.1) can be simplified as

$$H(x) = \sum_{n=1}^{\infty}\frac{1}{n}\left\{\sum_{k=0}^{\infty} b_k\left[\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]^{\alpha+k}\right\}^{n}.$$

5.3. Reversed hazard rate function

Reversed hazard rate (RHR) function is an important measure as a tool in the analysis of the reliability of both natural and man-made systems. Recently, the properties of the RHR have attracted considerable interest from researchers (see for examples [6] and [11]). The RHR function is defined as $r(x) = f(x)/F(x)$. Thus, the RHR function of ZBLN distribution is given by

$$r(x) = \frac{\phi\left(\frac{\log x - \mu}{\sigma}\right)\left\{-\log\left[1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]\right\}^{\alpha - 1}}{\sigma x\ \gamma\left(\alpha, -\log\left[1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]\right)}.$$

5.4. Conditional moments

For lifetime distributions, it is of greater interest to know the conditional moments which are important in prediction. The conditional moments of any distribution is defined as

$$E(X^r|X > t) = \frac{1}{S(t)} \int_t^\infty x^r f(x) dx.$$

Thus, the conditional moments of ZBLN distribution is given by

$$E(X^r|X > t) = \frac{1}{S(t)} \sum_{k=0}^{\infty} \left(\frac{\alpha + k}{\sigma}\right)\ b_k\ I_1(r, k), \tag{5.2}$$

where $S(.)$ is the survival function, b_k is given in Equation (2.3) and $I_1(r, k)$ is given as

$$I_1(r, k) = \int_t^\infty x^{r-1}\ \phi\left(\frac{\log x - \mu}{\sigma}\right)\left[\Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]^{\alpha + k - 1} dx. \tag{5.3}$$

5.5. Vitality function

In modeling lifetime data, the vitality function is a very valuable tool. This function plays important role in reliability engineering, biomedical science, and survival analysis. It is worth mentioning that the rapid aging of a component needs to low vitality relatively, whereas high vitality implies relatively slow aging during the given time period. For more details on the vitality function see [16].

For $r = 1$, in Equation (5.2), gives the vitality function of ZBLN distribution, and is given by

$$V(t) = E(X|X > t) = \frac{1}{S(t)} \int_t^\infty x f(x) dx = \frac{1}{S(t)} \sum_{k=0}^{\infty} \left(\frac{\alpha + k}{\sigma}\right) b_k\ I_1(1, k), \tag{5.4}$$

where $I_1(1,k)$ is obtained by putting $r=1$ in Equation (5.3), and is given by

$$I_1(1,k)=\int_t^\infty \phi\left(\frac{\log x-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]^{\alpha+k-1} dx.$$

5.6. Geometric vitality function

The concept of geometric vitality function is based on the geometric mean of the residual lifetime. If X be a random variable that represents the lifetime of a component, then $\log G(t)=E(\log X|X>t)$ represents the geometric mean of lifetimes of components that have survived up to time t. For a non-negative random variable X follows an absolutely continuous distribution function, with $E(\log X)<1$, the geometric vitality function is defined as

$$\log G(t)=E(\log X|X>t)=\frac{1}{S(t)}\int_t^\infty \log x\ f(x)dx.$$

Now, the geometric vitality function of the ZBLN distribution is given by

$$\log G(t)=\frac{1}{S(t)}\sum_{k=0}^{\infty}(\alpha+k)\ b_k\ I_2(k),$$

where $I_2(k)$ can be expressed as

$$I_2(k)=\int_t^\infty \left(\frac{\log x}{\sigma x}\right)\phi\left(\frac{\log x-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]^{\alpha+k-1} dx.$$

5.7. Moments of residual life

In reliability theory, the concept of residual life is very noteworthy. It represents the life remaining in a unit after it has attained age t.

The r^{th} order moment of the residual life of the ZBLN distribution is given as

$$\begin{aligned}\mu_r(t)=E\left[(X-t)^r|X>t\right]&=\frac{1}{S(t)}\int_t^\infty (x-t)^r\ f(x)dx\\&=\frac{1}{S(t)}\sum_{i=0}^{r}\binom{r}{i}(-1)^{r-i}\ t^{r-i}\int_t^\infty x^i\ f(x)dx,\end{aligned}$$

which can be simplified as

$$\mu_r(t)=\frac{1}{S(t)}\sum_{i=0}^{r}\sum_{k=0}^{\infty}\binom{r}{i}(-1)^{r-i}\ t^{r-i}\left(\frac{\alpha+k}{\sigma}\right)\ b_k\ I_1(i,k),$$

where $I_1(r,k)$ is given in Equation (5.3). Now, for $r=1$ and using Equation (2.5), we get the expression for mean residual life (MRL) function, and is given by

$$\mu_1(t) = E\left(X-t|X>t\right) = \frac{1}{S(t)}\int_t^{\infty}(x-t)\ f(x)dx$$
$$= \frac{1}{S(t)}\sum_{k=0}^{\infty}\left(\frac{\alpha+k}{\sigma}\right)\ b_k\ I_3(k),$$

where

$$I_3(k) = \int_t^{\infty}\frac{(x-t)}{x}\phi\left(\frac{\log x-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]^{\alpha+k-1}dx.$$

Hence, $\mu_1(t)$ also gets the form, $\mu_1(t)=V(t)-t$, where $V(t)$ is given in Equation (5.4). Similarly, the second moment of the residual lifetime of the ZBLN distribution is given by

$$\mu_2(t) = \frac{1}{S(t)}\sum_{k=0}^{\infty}\left(\frac{\alpha+k}{\sigma}\right) b_k\ I_1(2,k) - \frac{2t\ V(t)}{S(t)} + t^2,$$

where $I_1(2,k)$ is given as

$$I_1(2,k) = \int_t^{\infty} x\ \phi\left(\frac{\log x-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]^{\alpha+k-1}dx.$$

Thus, the variance of the residual life function of the ZBLN distribution can be obtained using $\mu_1(t)$ and $\mu_2(t)$.

5.8. Moments of reversed residual life

The r^{th} order moment of the reversed residual life of the ZBLN distribution is given by

$$m_r(t) = E\left[(t-X)^r|X\leq t\right] = \frac{1}{F(t)}\int_0^t (t-x)^r\ f(x)dx$$
$$= \frac{1}{F(t)}\sum_{i=0}^{r}\binom{r}{i}(-1)^i\ t^{r-i}\int_0^t x^i\ f(x)dx.$$

On simplification, $m_r(t)$ gets the form

$$m_r(t) = \frac{1}{F(t)}\sum_{i=0}^{r}\sum_{k=0}^{\infty}\binom{r}{i}(-1)^{r-i}\ t^{r-i}\left(\frac{\alpha+k}{\sigma}\right)\ b_k\ I_4(,k), \tag{5.5}$$

where $I_4(i,k)$ is given as

$$I_4(i,k) = \int_0^t x^{i-1}\ \phi\left(\frac{\log x-\mu}{\sigma}\right)\left[\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]^{\alpha+k-1}dx.$$

Now, the mean $(m_1(t))$ and second moment $(m_2(t))$ of the reversed residual life of the ZBLN distribution can be obtained by setting $r=1,2$; respectively in Equation (5.5). Again, using $m_1(t)$ and $m_2(t)$, one can obtain the variance of the reversed residual life function of the distribution.

6. RÉNYI ENTROPY

Entropy is considered to be the measure of uncertainty of a system and it is typically used in physical sciences. The study of entropy has gained momentum in the theoretical perspective as well as in terms of its applications in the field of applied research. Among the number of entropies available in the literature, one of the most popular entropy measures is Rényi entropy (see [23]).

By definition, for any random variable Y with pdf $g(y)$, the Rényi entropy is defined as

$$H_\gamma(y) = \frac{1}{1-\gamma} \log \int_{\mathbb{R}} g^\gamma(y)dy; \quad \text{for} \quad \gamma > 0 \quad \text{and} \quad \gamma \neq 1.$$

Let $f(x)$ be the density function of the ZBLN distribution, then standard calculations show that the Rényi entropy of the distribution can be written as

$$H_\gamma(x) = \frac{1}{1-\gamma} \log \int_0^\infty f^\gamma(x)dx$$

in which, by using (2.1),

$$\int_0^\infty f^\gamma(x)dx = \left(\frac{1}{\sigma\ \Gamma(\alpha)}\right)^\gamma \int_0^\infty \tau^\gamma(x)dx,$$

where

$$\tau^\gamma(x) = \left\{\frac{1}{x}\ \phi\left(\frac{\log x - \mu}{\sigma}\right)\left\{-\log\left[1 - \Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]\right\}^{\alpha-1}\right\}^\gamma.$$

On simplification, the Rényi entropy of ZBLN distribution gets the expression

$$H_\gamma(x) = (1-\gamma)^{-1} \log \int_0^\infty \tau^\gamma(x)dx - \gamma(1-\gamma)^{-1}\log(\sigma) - \gamma(1-\gamma)^{-1}\log\left(\Gamma(\alpha)\right).$$

7. INEQUALITY MEASURES

Lorenz and Bonferroni curves are income inequality measures that are widely useful and applicable to some other areas including reliability, demography, medicine, and insurance. Also, the Zenga curve introduced by Zenga (2007) (see [25]) is another widely used inequality measure. The Lorenz, Bonferroni, and Zenga curves for the ZBLN distribution will be derived in this section. The Lorenz curve is defined by

$$L_F(x) = \frac{1}{E(X)} \int_0^x t\ f(t)dt.$$

Simple algebra provides the Lorenz curve for ZBLN distribution, and is given by

$$L_F(x) = \frac{\sum\limits_{k_1=0}^{\infty} (\alpha + k_1)\ b_{k_1}\ I_4(k_1)}{\sum\limits_{k_2=0}^{\infty} (\alpha + k_2)\ b_{k_2}\ I_5(k_2)},$$

where b_k is given in equation (2.3),

$$I_4(k_1) = \int_0^x \phi\left(\frac{\log t - \mu}{\sigma}\right)\left[\Phi\left(\frac{\log t - \mu}{\sigma}\right)\right]^{\alpha+k_1-1}, \quad \text{and}$$

$$I_5(k_2) = \int_0^\infty \phi\left(\frac{\log x - \mu}{\sigma}\right)\left[\Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]^{\alpha+k_2-1}.$$

Also, the Bonferroni curve is defined by

$$B_F(x) = \frac{1}{E(X)F(x)}\int_0^x t\ f(t)dt.$$

Thus, the Bonferroni curve of ZBLN distribution gets expression given by

$$B_F(x) = \frac{\sum\limits_{k_1=0}^{\infty} (\alpha + k_1)\ b_{k_1}\ I_4(k_1)}{\left\{\sum\limits_{k_2=0}^{\infty} (\alpha + k_2)\ b_{k_2}\ I_5(k_2)\right\}\left\{\sum_{k=0}^{\infty} b_k\left[\Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]^{\alpha+k}\right\}}.$$

Now, the Zenga curve is defined as

$$A_F(x) = 1 - \frac{\mu^-(x)}{\mu^+(x)}, \tag{7.1}$$

where

$$\mu^-(x) = \frac{1}{F(x)}\int_0^x t\ f(t)dt, \quad \text{and} \quad \mu^+(x) = \frac{1}{S(x)}\int_x^\infty t\ f(t)dt.$$

Therefore, $\mu^-(x)$ and $\mu^+(x)$ of ZBLN distribution are respectively given by

$$\mu^-(x) = \frac{\sum\limits_{k_1=0}^{\infty}\left(\frac{\alpha+k_1}{\sigma}\right)\ b_{k_1}\ I_4(k_1)}{\sum\limits_{k=0}^{\infty} b_k\left[\Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]^{\alpha+k}}, \quad \text{and} \quad \mu^+(x) = V(x),$$

where $V(x)$ is the vitality function of ZBLN distribution in x, such that the expression for vitality function of the distribution is given in (5.4). Substituting the values of $\mu^-(x)$ and $\mu^+(x)$ in (7.1), gets the expression of $A_F(x)$ for the ZBLN distribution.

8. ORDER STATISTICS

Let $X_1, X_2, ..., X_n$ be a random sample from the ZBLN distribution and its order statistics is $X_{1:n}, X_{2:n}, ..., X_{n:n}$. Let $F_{i:n}(x)$ and $f_{i:n}(x)$ denote the cdf and pdf of the i^{th} order statistic $X_{i:n}$, respectively. Hence, using the standard expressions of order statistics, $F_{i:n}(x)$ and $f_{i:n}(x)$ of ZBLN distribution is respectively given by

$$F_{i:n}(x) = \sum_{j=i}^{n} \binom{n}{j} F^j(x) \; [1 - F(x)]^{n-j},$$

and

$$\begin{aligned} f_{i:n}(x) &= \frac{n!}{(i-1)! \; (n-i)!} \; [F(x)]^{i-1} \; [1 - F(x)]^{n-i} \; f(x) \\ &= \frac{1}{\mathcal{B}(i, n-i+1)} \sum_{k_3=0}^{n-i} (-1)^{k_3} \binom{n-i}{k_3} [F(x)]^{k_3+i-1} f(x) \\ &= \frac{1}{\mathcal{B}(i, n-i+1)} \sum_{k_3=0}^{n-i} (-1)^{k_3} \binom{n-i}{k_3} \times \\ & \left\{ \sum_{k=0}^{\infty} b_k \left[\Phi\left(\frac{\log x - \mu}{\sigma} \right) \right]^{\alpha+k} \right\}^{k_3+i-1} \times \\ & \sum_{k=0}^{\infty} b_k \frac{(\alpha+k)}{\sigma x} \phi\left(\frac{\log x - \mu}{\sigma} \right) \left[\Phi\left(\frac{\log x - \mu}{\sigma} \right) \right]^{\alpha+k-1}, \end{aligned}$$

where $\mathcal{B}(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the Beta function. Now, for $i = 1$ and n, one can get the pdf of $X_{(1)} = min\{X_1, X_2, ..., X_n\}$ and $X_{(n)} = max\{X_1, X_2, ..., X_n\}$ for ZBLN distribution, respectively.

9. ESTIMATION OF PARAMETERS

In this section, we'll look at how to estimate the parameters of the ZBLN distribution using two widely used methods: maximum likelihood (ML) and Bayesian methods.

9.1. Maximum likelihood estimation

This subsection considers the maximum likelihood estimation for the ZBLN model parameters α, μ, and σ. Let $X_1, X_2, ..., X_n$ be a random sample taken from the ZBLN distribution, and $x_1, x_2, ..., x_n$ are the corresponding observed values. Then the log-likelihood function can be expressed as

$$\mathcal{L}_n = -n\log(\sigma) - n\log\left(\Gamma(\alpha)\right) - \sum_{i=1}^{n}\log(x_i) + \sum_{i=1}^{n}\log\left[\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right]$$
$$+ (\alpha-1)\sum_{i=1}^{n}\log\left\{-\log\left[1-\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]\right\}.$$

The score function associated with the log-likelihood function is

$$\mathbf{U} = \left(\frac{\partial\mathcal{L}_n}{\partial\alpha}, \frac{\partial\mathcal{L}_n}{\partial\mu}, \frac{\partial\mathcal{L}_n}{\partial\sigma}\right)^T.$$

Now, by solving $\frac{\partial\mathcal{L}_n}{\partial\alpha} = 0$, $\frac{\partial\mathcal{L}_n}{\partial\mu} = 0$ and $\frac{\partial\mathcal{L}_n}{\partial\sigma} = 0$, we get the associated nonlinear log-likelihood equations and are respectively given by

$$\sum_{i=1}^{n}\log\left\{-\log\left[1-\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]\right\} - n\ \psi(\alpha) = 0, \tag{9.1}$$

(9.2)
$$\sum_{i=1}^{n}\frac{\log(x_i)-\mu}{\sigma^2} + \left(\frac{\alpha-1}{\sigma}\right)\sum_{i=1}^{n}\frac{\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{\left[1-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right]\log\left[1-\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]} = 0,$$

(9.3)
$$\frac{-n}{\sigma} + \sum_{i=1}^{n}\frac{(\log(x_i)-\mu)^2}{\sigma^3} + \sum_{i=1}^{n}\frac{(\alpha-1)\left(\frac{\log(x_i)-\mu}{\sigma}\right)\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{\left[1-\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right]\log\left[1-\Phi\left(\frac{\log x-\mu}{\sigma}\right)\right]} = 0,$$

where $\psi(\alpha) = d\left\{\log\Gamma(\alpha)\right\}/d\alpha$ is the digamma function. Now, by solving the equations (9.1), (9.2) and (9.3) simultaneously, we obtain the maximum likelihood estimators (MLEs) $(\hat{\alpha}, \hat{\mu}, \hat{\sigma})$ of the model parameters (α, μ, σ).

Now, we construct the asymptotic confidence intervals for parameters α, μ and σ. On taking the second partial derivatives of equations (9.1), (9.2) and (9.3), the Hessian matrix of ZBLN distribution can be obtained, and denoted as $H(\Theta)$, where $\Theta = \{\alpha, \mu, \sigma\}$. Now, the observed Fisher's information matrix $J(\Theta)$ can be obtained by taking the negative of Hessian matrix. That is, $J(\Theta) = -H(\Theta)$. Hence, the inverse of observed Fisher's information matrix will provide the variance-covariance matrix of the MLEs, which is given by

$$\Sigma = J^{-1}(\Theta) = \left\{\Sigma_{ij},\ i,j = 1,2,3\right\},$$

and $\Sigma_{ij} = \Sigma_{ji}$ for $i \neq j = 1,2,3$. Again, it is well established that the MLEs are asymptotically normally distributed. That is, $\sqrt{n}(\Theta - \hat{\Theta}) \sim N_3(0, \Sigma)$, where n is the sample size and $\hat{\Theta}$ is the MLEs of Θ.

Thus, we obtain $100 \times (1-\delta)\%$ asymptotic confidence intervals of the parameters using the following formulae:

$$\alpha \in \left\{\hat{\alpha} \mp Z_{\delta/2}\sqrt{\Sigma_{11}}\right\},\ \mu \in \left\{\hat{\mu} \mp Z_{\delta/2}\sqrt{\Sigma_{22}}\right\},\ \text{and } \sigma \in \left\{\hat{\sigma} \mp Z_{\delta/2}\sqrt{\Sigma_{33}}\right\},$$

where Z_δ is the upper δ^{th} percentile of the standard normal distribution.

9.2. Bayesian estimation

The Bayesian analysis for the ZBLN model parameters is performed in this subsection. Each parameter should have a prior density in order to do so. For this, we utilize two types of priors: half-Cauchy (HC) and normal (N) priors. The pdf of the HC distribution with scale parameter a is defined as

$$f_{HC}(x_*) = \frac{2a}{\pi(x_*^2 + a^2)}, \quad x_* > 0,\ a > 0.$$

The HC distribution has no mean or variance. Meanwhile, its mode is equal to 0. Since the pdf of the HC is virtually flat but not totally flat at scale value equals 25, which verges on acquiring adequate information for the numerical approximation algorithm to continue looking at the target posterior pdf, the HC distribution with $a = 25$ is recommended as a noninformative prior. Gelman and Hill (2006)([8]) suggested that the uniform distribution, or whether more information is required, is a superior alternative to the HC distribution. As a result, for the parameters α and σ, the HC distribution with $a = 25$ is chosen as a noninformative prior distribution in this article. Thus, we set the prior distributions of the parameters to be $\mu \sim N(0, 1000)$, and $\alpha, \sigma \sim HC(25)$. Thus, we obtain the joint posterior pdf as given by

$$\pi(\mu, \alpha, \sigma|x) \propto L_n \times \pi(\mu) \times \pi(\alpha) \times \pi(\sigma), \tag{9.4}$$

where L_n is the likelihood funtion for ZBLN distribution. From Equation (9.4), it is obvious that there is no analytical solution to find out the Bayesian estimates. Thus, we use a remarkable method of simulation, namely the Metropolis-Hastings algorithm of the Markov Chain Monte Carlo (MCMC) method.

9.3. Bootstrap confidence intervals

In this subsection, we use the parametric bootstrap method to approximate the distribution of the maximum likelihood estimators of the ZBLN parameters. Then, we can use the bootstrap distribution to estimate the confidence intervals on each parameter of the fitted ZBLN distribution. Let $\hat{\Theta} = \Theta(X)$ be a ML estimator of the set of parameters of interest $\Theta = \{\alpha, \mu, \sigma\}$ using a given dataset $X = \{x_1, x_2, ..., x_n\}$. The bootstrap is a method to estimate the distribution of the statistic $\hat{\Theta}$ by getting a random sample $\Theta_1^*, \Theta_2^*, ..., \Theta_B^*$ for Θ based on B random samples that are drawn with replacement from the original data $X = \{x_1, x_2, ..., x_n\}$ (see [24]). The bootstrap sample $\Theta_1^*, \Theta_2^*, ..., \Theta_B^*$ can be used to construct bootstrap confidence intervals for the parametric set $\Theta = \{\alpha, \mu, \sigma\}$ of ZBLN distribution.

Thus, we obtain $100 \times (1-\delta)\%$ bootstrap confidence intervals of the parameters using the following formulae:

$$\alpha \in \left\{\hat{\alpha} \mp z_{\delta/2}\ \hat{se}_{\alpha,boot}\right\},\ \mu \in \left\{\hat{\mu} \mp z_{\delta/2}\ \hat{se}_{\mu,boot}\right\},\ \text{and } \sigma \in \left\{\hat{\sigma} \mp z_{\delta/2}\ \hat{se}_{\sigma,boot}\right\},$$

where z_δ denotes the δ^{th} percentile of the bootstrap sample and for $\Theta = \{\alpha, \mu, \sigma\}$

$$\hat{se}_{\Theta,boot} = \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\Theta_b^* - \frac{1}{B}\sum_{b=1}^{B}\Theta_b^*\right)^2}.$$

10. PERFORMANCE OF THE ESTIMATORS USING SIMULATION STUDY

In this section, we conduct simulation experiments to assess the long-run performances of MLEs and Bayesian estimates of the ZBLN parameters for some finite sample sizes. We have simulated datasets of sizes $n = 50, 100$, and 250 from the ZBLN distribution for the parameter values $\alpha = 0.2,\ \mu = 3.5,\ \sigma = 0.5$ and iterated each sample for 500 times. Then, we compute the average biases and MSEs for the MLEs to all replications in the relevant sample sizes. That is, the

Parameters	**Sample Size**	**Estimates**	**Bias**	**M.S.E**
α	50	0.6892	0.4892	1.9199
	100	0.4514	0.2514	0.7520
	250	0.2597	0.0597	0.0384
μ	50	3.0965	-0.4035	1.1623
	100	3.2030	-0.2970	0.7376
	250	3.3702	-0.1298	0.2029
σ	50	0.5418	0.0418	0.0385
	100	0.5345	0.0345	0.0297
	250	0.5215	0.0215	0.0112

Table 1: Estimates, Average bias and MSE values of MLEs from simulation of the ZBLN distribution.

analysis computes the values by the given formulae. The equation for average bias of the simulated estimates equals $\frac{1}{500}\sum_{i=1}^{500}(\hat{\Theta}_i - \Theta)$, and the equation for average MSE of the simulated estimates equals $\frac{1}{500}\sum_{i=1}^{500}(\hat{\Theta}_i - \Theta)^2$, where $\hat{\Theta} = (\hat{\alpha}, \hat{\mu}, \hat{\sigma})$ are estimates of the parameter vector $\Theta = (\alpha, \mu, \sigma)$. The results to the simulation for MLEs are reported in Table 1. It can be concluded that the M.S.E of all the estimators decreases with increasing sample size. This shows the consistency of the estimators.

Now, in the case of Bayesian simulation, we consider the prior distributions for the ZBLN parameters as given in Subsection 9.2. For the respective sample sizes, the posterior summary results such as mean, standard deviation (SD), Monte Carlo error (MCE), 95% confidence interval (CI), and median are presented in Table 2. It is observed that the SD and MCE decrease as the sample size increases, which predicts the consistency of Bayesian estimates of the ZBLN distribution parameters.

Parameters	n	**Mean**	**SD**	**MCE**	95% **CI**	**Median**
	50	1.6267	1.5925	1.0016	(0.4615, 5.0629)	1.4301
α	100	0.5894	0.2646	0.1755	(0.1748, 0.8089)	0.7889
	250	0.2115	0.0919	0.0458	(0.1748, 0.4371)	0.1931
	50	2.5282	0.7701	0.5653	(1.1535, 3.4592)	2.3215
μ	100	2.9731	0.3274	0.2064	(2.7282, 3.5331)	2.8252
	250	3.4679	0.1632	0.0813	(3.0676, 3.5331)	3.4131
	50	0.6999	0.1494	0.0667	(0.4728, 0.8246)	0.7105
σ	100	0.6444	0.1065	0.0591	(0.4704, 0.7237)	0.6218
	250	0.4959	0.0640	0.0319	(0.4704, 0.6530)	0.5715

Table 2: Posterior summary results for Bayesian simulation

11. APPLICATION AND EMPIRICAL STUDY

To demonstrate the applicability of the ZBLN distribution, we consider a real dataset based on a cancer survival study, and the parameters are estimated by using maximum likelihood, and the Bayesian estimation methods to compare the data modeling ability of the ZBLN distribution over some competitive distributions. The dataset is taken from Lee & Wang (2003) (see [17]), which corresponds to the remission times (in months) of a random sample of 128 bladder cancer patients. The summary statistics of the dataset is given in Table 3.

n	M	Md	SD	Sk	Ku	min	max
128	9.2094	6.28	10.4026	3.3987	16.3942	0.08	79.05

Table 3: Summary statistics of real dataset.

Now, we study the empirical hazard function of the datasets using the concept of total time on test (TTT) plot. The TTT plot is a graph that mainly serves to discriminate between different types of aging represented in hazard rate shapes. For details, the readers are referred to [1]. The TTT plot is drawn by plotting

$$T\left(\frac{i}{n}\right) = \frac{\sum_{r=1}^{i} x_{r:n} + (n-i)x_{i:n}}{\sum_{r=1}^{n} x_{r:n}}$$

against i/n, where $i = 1, 2, ..., n$ and $x_{r:n}$, $r = 1, 2, ..., n$ are the order statistics of the sample. Figure (3) indicates that the above-given dataset has an upside-down bathtub shape for the empirical hazard function. Therefore, the ZBLN distribution can be a credible pick for the given dataset, since its hazard function satisfies the upside-down bathtub shape.

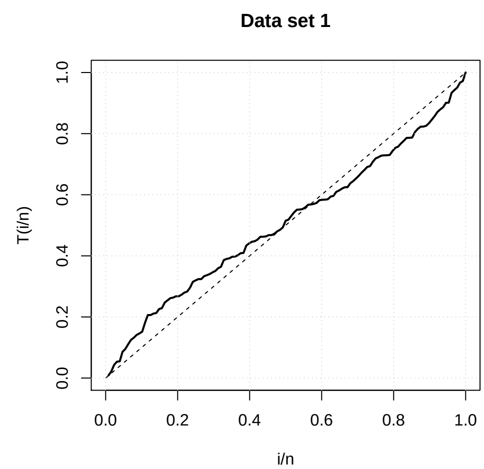

Figure 3: The TTT plot of real dataset.

11.1. Maximum likelihood estimation

To illustrate the potentiality of the ZBLN distribution, the following distributions are considered for comparison.

- The two-parameter Log-normal (LN) distribution with pdf

$$f(x) = \frac{1}{\sqrt{2\pi}\ \sigma x}\ exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right], \quad x > 0,\ \mu \in \mathbb{R},\ \sigma > 0.$$

- The Exponentiated Log-normal (ELN) distribution with pdf

$$f(x) = \frac{\alpha}{x\sigma}\phi\left(\frac{\log x - \mu}{\sigma}\right)\left[\Phi\left(\frac{\log x - \mu}{\sigma}\right)\right]^{\alpha-1},\ x > 0,\ \mu \in \mathbb{R},\ \alpha, \sigma > 0.$$

- The Weibull distribution with pdf

$$f(x) = \frac{\alpha}{\sigma}\left(\frac{x}{\sigma}\right)^{\alpha-1} e^{-(x/\sigma)^{\alpha}}, \quad x > 0,\ \alpha, \sigma > 0.$$

- The New Generalized Lindley distribution (NGLD) (see [7]) with pdf

$$f(x) = \frac{e^{-\mu x}}{1+\mu}\left(\frac{\mu^{\alpha+1}x^{\alpha-1}}{\Gamma(\alpha)} + \frac{\mu^{\sigma}x^{\sigma-1}}{\Gamma(\sigma)}\right), \quad x > 0,\ \alpha, \mu, \sigma > 0.$$

- The Zografos-Balakrishnan Lindley distribution (ZBLD) (see [13]) with pdf

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

We apply the following statistical tools in order to find out the goodness-of-fit of distributions to the real dataset; log-likelihood (LL), Kolmogorov-Smirnov (KS), Cramér-von Misses (W^*), Anderson-Darling (A^*) statistics, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values, and are presented in Table (4). We use the RStudio software for numerical evaluations.

Estimates	**LN**	**ELN**	**Weibull**	**NGLD**	**ZBLD**	**ZBLN**
$\hat{\alpha}$	-	0.1516	1.0514	1.1852	0.7353	0.2425
$\hat{\mu}$	1.7422	3.0494	-	0.1287	-	2.9666
$\hat{\sigma}$	1.0646	0.5404	9.4172	1.1850	0.1569	0.6730
LL	-412.6565	-410.0441	-411.8925	-411.0846	-413.5513	-409.3414
KS	0.0644	0.0562	0.0721	0.0760	0.0901	0.0542
W^*	0.1313	0.0846	0.1666	0.1416	0.2230	0.0736
A^*	0.8708	0.5589	1.0488	0.8235	1.2465	0.4828
AIC	829.3131	826.0883	827.7849	828.1691	831.1025	824.6828
BIC	835.0171	834.6444	833.4890	836.7252	836.8066	833.2389

Table 4: Maximum-likelihood estimates, goodness-of-fit statistics, AIC and BIC values based on the bladder cancer dataset.

Moreover, Table 4 shows the MLEs and goodness-of-fit statistics of the distributions for the corresponding dataset. It can be seen that the KS, W^*, A^*, AIC, and BIC values of the ZBLN distribution are smaller than that of other distributions. We also present other important graphs which consist of empirical density plot, empirical cdf plot, Q-Q, and P-P plots for the real dataset in Figure (4). It again gives some superimposed curves of those fitted and empirical functions. Thus, we conclude that the ZBLN is the most suitable distribution for the given dataset while comparing other distributions.

We also utilized the likelihood ratio (LR) test for comparing ZBLN distribution having additional parameter α with LN distribution. That is, we test H_0 : LN against H_A : ZBLN and obtain critical values for the LR test statistics for the cancer dataset. Thus we get the LR test statistic value as 6.663 and the corresponding p-value as 0.0098 for the given dataset. Given the value of the test

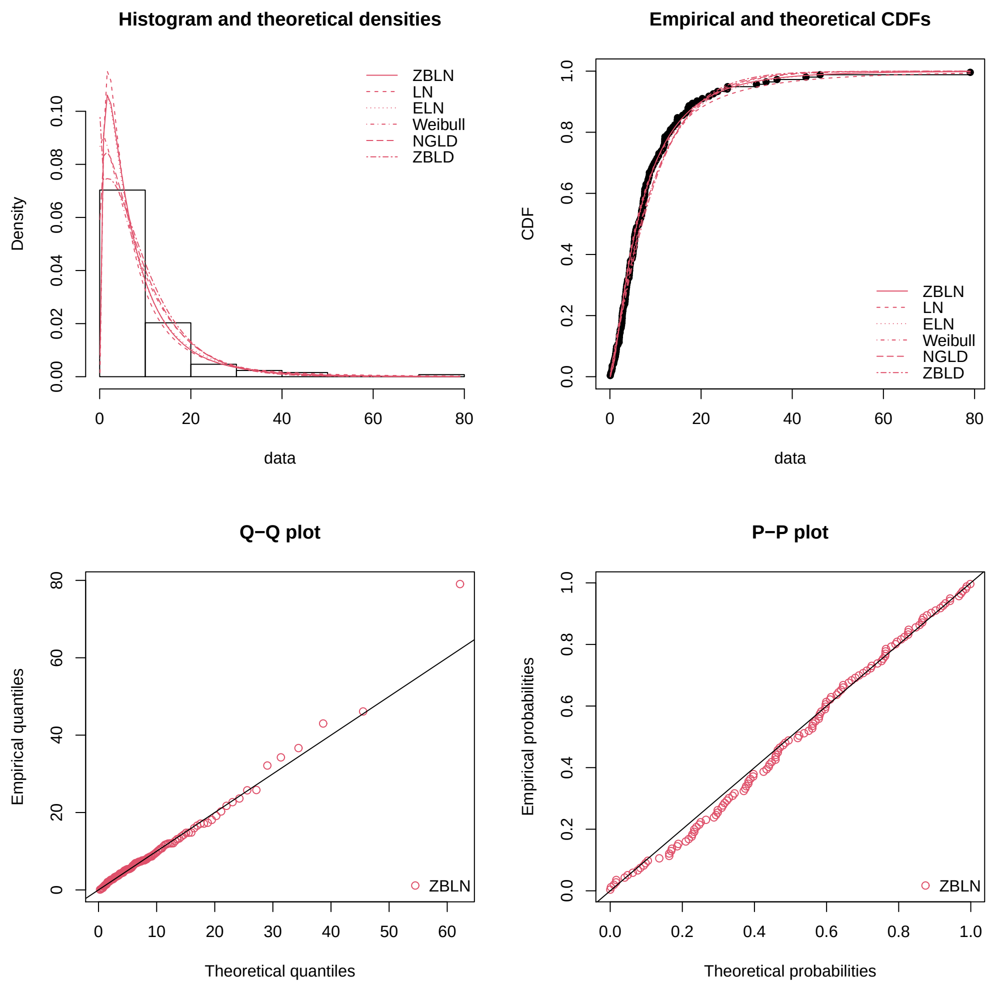

Figure 4: Various empirical plots of bladder cancer dataset.

statistics and the associated p-value, we reject the null hypotheses for the dataset and conclude that the ZBLN model provides a significantly better representation for the dataset than the LN distribution.

Now, the Hessian matrix corresponding to real dataset is obtained as

$$H(\Theta) = \begin{pmatrix} 2330.9119 & 794.1654 & -1451.9253 \\ 794.1654 & 131391.5635 & -191.6978 \\ -1451.9253 & -191.6978 & 132403.6280 \end{pmatrix}.$$

Hence, the asymptotic variance-covariance matrix for real dataset is obtained as

$$\Sigma = \begin{pmatrix} 0.0793 & -0.0065 & 0.0118 \\ -0.0065 & 0.0189 & -0.0005 \\ 0.0118 & -0.0005 & 0.0189 \end{pmatrix}.$$

Again, the 95% asymptotic confidence intervals of the ZBLN parameters are given in Table (5).

Parameter	Lower	Upper
α	0.2017	0.2833
μ	2.9612	2.9720
σ	0.6676	0.6784

Table 5: The 95% asymptotic confidence intervals of the ZBLN parameters based on bladder cancer dataset.

Now, we use the obtained MLEs to derive the 95% bootstrap confidence intervals for the parameters α, μ, and σ. We simulate 1001 samples of size as in the real dataset we studied, from ZBLN distribution with true values of the parameters taken as MLEs of the parameters. For each obtained sample, we have estimated the MLEs $\hat{\alpha}_b^*$, $\hat{\mu}_b^*$, and $\hat{\sigma}_b^*$, for $b \in \{1, 2, ..., 1001\}$. The median and 95% bootstrap confidence interval for parameters α, μ, and σ of the given dataset is presented in Table 6. It is also interesting to look at the joint distribution of the bootstrapped values in a matrix of scatter plots in order to understand the potential structural correlation between parameters. The plots in Figure 5 consist of matrix scatterplots of the bootstrapped values of ZBLN parameters providing a representation of the joint uncertainty distribution of the fitted parameters.

Parameter	Median	Bootstrap CI
α	0.2294	(0.1203, 3.9249)
μ	2.9653	(-0.4722, 3.3680)
σ	0.6549	(0.4867, 1.2079)

Table 6: The median and 95% bootstrap confidence interval for ZBLN parameters of the bladder cancer dataset.

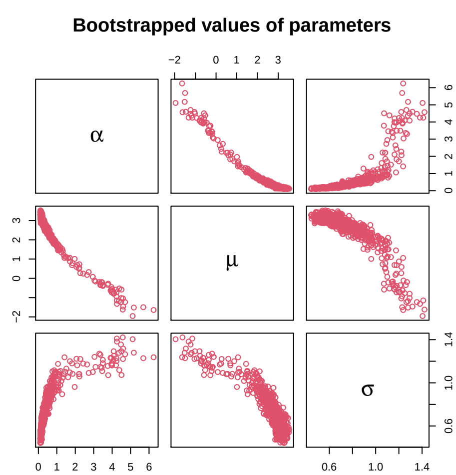

Figure 5: Matrix scatter plots of bootstrappped values of ZBLN parameters due to the bladder cancer dataset.

11.2. Bayesian estimation

Here, we focus on estimating the parameters of the ZBLN distribution using the Bayesian procedure based on the same univariate bladder cancer survival dataset which we discussed in the above subsection. In the context of Bayesian estimation, the analysis was performed using the Metropolis-Hastings algorithm of the MCMC method with 1001 iterations. For comparing Bayes estimates with the MLEs, both the estimates of the ZBLN parameters with corresponding standard error (SE) and Monte Carlo standard error (MCSE) for the real dataset are given in Table 7. The numerical computations on Bayesian estimation are also done using RStudio software.

Parameter	**MLE (SE)**	**Bayes (MCSE)**
α	0.2425 (0.0208)	0.2471 (0.0402)
μ	2.9666 (0.0028)	3.0206 (0.10003)
σ	0.6730 (0.0028)	0.7127 (0.0343)

Table 7: MLEs and Bayesian estimates of the ZBLN parameters on bladder cancer dataset.

12. CONCLUDING REMARKS

In this paper, we studied a distribution that generalizes the log-normal distribution. We refer to the model as the Zografos-Balakrishnan log-normal (ZBLN) distribution and study its mathematical and statistical properties. We provide explicit expressions for the moments, quantile function, various reliability measures, Rényi entropy, and some inequality measures associated with the ZBLN distribution. It is worth noting that the hazard rate function supports all of the standard shapes, including increasing, decreasing, bathtub, and upside-down bathtub. The model parameters are estimated by using the Bayesian technique, and the method of maximum likelihood, and also, the observed information matrix is presented. Further, we adopt the parametric bootstrap technique to obtain confidence intervals for the model parameters. Moreover, the simulation studies based on the defined estimation methods are also done to confirm the parameter consistencies. The usefulness of the new model is illustrated by an application to the real dataset based on a cancer survival study using goodness-of-fit tests. The model provides a consistently better fit than other models available in the literature. We hope the model may attract wider applications for modeling positive real datasets in many areas such as physics, engineering, medicine, survival analysis, hydrology, economics, and so on.

ACKNOWLEDGMENTS

The authors would like to thank the Editor and the unknown reviewers for the constructive comments, which greatly improved the present version of this article.

REFERENCES

[1] Aarset, M.V. (1987). How to Identify a Bathtub Hazard Rate, *IEEE Transactions on Reliability*, **R-36**(1), 106–108.

[2] Beal, J. (2017). Biochemical complexity drives log-normal variation in genetic expression, *Engineering Biology*, **1**(1), 55–60.

[3] Bernardeau, F. and Kofman, L. (1995). Properties of the Cosmological Density Distribution Function, *The Astrophysical Journal*, **443**, 479.

[4] Blasi, P.; Burles, S.M. and Olinto, A.V. (1999). Cosmological Magnetic Field Limits in an Inhomogeneous Universe, *The Astrophysical Journal*, **514**(2), L79–L82.

[5] Carvalho, J.; Piaggio, G.; Wojdyla, D.; Widmer, M. and Gülmezoglu, A. (2018). Distribution of postpartum blood loss: modeling, estimation and application to clinical trials, *Reproductive Health*, **15**(1), 199.

[6] Chandra, N. and Roy, D. (2001). Some results on reversed hazard rate, *Probability in the Engineering and Informational Sciences*, **15**(1), 95–102.

[7] Elbatal, I.; Merovci, F. and Elgarhy, M. (2013). A new generalized Lindley distribution, *Mathematical Theory and Modeling*, **3**, 30–47.

[8] Gelman A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models (Analytical Methods for Social Research)*, Cambridge University Press.

[9] Glass, B. (1954). Statistics for Medical and Other Biological Students. L. Bernstein , M. Weatherall, *The Quarterly Review of Biology*, **29**(3), p. 303.

[10] Gupta, R.; Gupta, P. and Gupta, R. (1998). Modeling failure time data by lehman alternatives, *Communications in Statistics: Theory and Methods*, **27**(4), 887–904.

[11] Gupta R. and Nanda, A. (2001). Some results on reversed hazard rate ordering, *Communications in Statistics: Theory and Methods*, **30**(11), 2447–2457.

[12] Gupta R.D. and Kundu, D. (1999). Theory & Methods: Generalized exponential distributions, *Australian & New Zealand Journal of Statistics*, **41**, 173–188.

[13] Irshad, M. R.; D'cruz, V., and Maya, R. (2021). The Zografos-Balakrishnan Lindley Distribution: Properties and Applications. *Statistica*, **81**(1), 45–64.

[14] Irshad, M.R.; Maya, R. and Krishna, A. (2021). Exponentiated Power Muth Distribution and Associated Inference, *Journal of the Indian Society for Probability and Statistics*, **22**, 265–302.

[15] Jobe, J.; Crow, E.L. and Shimizu, K. (1989). Lognormal Distributions: Theory and Applications, *Technometrics*, **31**, 392.

[16] Kupka J. and Loo, S. (1989). The Hazard and Vitality Measures of Ageing, *Journal of Applied Probability*, **26**(3), 532–542.

[17] Lee E.T. and Wang, J.W. (2013). *Statistical Methods for Survival Data Analysis*, 4th ed., Wiley Publishing.

[18] Mudholkar G. and Srivastava, D. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data, *Reliability, IEEE Transactions*, **42**(2), 299–302.

[19] Nadarajah, S. (2006). The exponentiated Gumbel distribution with climate application, *Environmetrics*, **17**, 13–23.

[20] Nadarajah, S.; Cordeiro, G. and Ortega, E. (2015). The Zografos-Balakrishnan-G Family of Distributions: Mathematical Properties and Applications, *Communications in Statistics: Theory and Methods*, **44**(1), 186–215.

[21] Nadarajah S. and Gupta, A. (2007). The Exponentiated Gamma Distribution with Application to Drought Data, *Calcutta Statistical Association Bulletin*, **59**, 29–54.

[22] Parravano, A.; Sanchez, N. and Alfaro, E. (2012). The dependence of prestellar core mass distributions on the structure of the parental cloud, *Astrophysical Journal*, **754**, 150.

[23] Rényi, A. (1961). *On Measures of Entropy and Information*, *In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Berkeley, Calif. University of California Press, 547–561.

[24] Wasserman, L. (2006). *All of Nonparametric Statistics*, Springer Texts in Statistics, Springer New York.

[25] Zenga, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means, *Statistica & Applicazioni* **4**, 3–27.

[26] Zografos K. and Balakrishnan, N. (2009). On families of beta- and generalized gamma-generated distributions and associated inference, *Statistical Methodology*, **6**(4), 344–362.