

The Zografos–Balakrishnan-log-logistic Distribution

Shahram Yaghoubzadeh Shahrestani;

Department of Statistics, Payame Noor University, Iran

Received: 12 Jun. 2014

Revised: 1 July 2017

Abstract

The Zografos–Balakrishnan-log-logistic (ZBLL) distribution is a new distribution of three parameters that has been introduced by Ramos et al. [1], and they presented some properties of the new distribution such as its probability density function, the cumulative distribution function, the moment generating function, its hazard (failure) rate function, quantiles and moments, rényi and shannon entropies, reliability, moments of order statistics and estimate the model parameters by maximum likelihood. In this paper, we obtain other several properties of the ZBLL distribution such as probability weighted moments, mean deviations and Bonferroni and Lorenz curves. We discuss estimation by method of minimum spacing square distance estimator. Also, we compare the results of fitting this distribution to some of other models, using to a real data set. We show that the ZBLL distribution fits better to this data set.

Keywords: Zografos–Balakrishnan-log-logistic distribution; mean deviations; Bonferroni and Lorenz curves, minimum spacing square distance estimator.

Introduction

Zografos and Balakrishnan [2] and Ristić and Balakrishnan [3] proposed two novel families of univariate distributions generated by gamma random variables. For any baseline cumulative distribution function (cdf) $G(x)$, and $x \in R$, Zografos and Balakrishnan [1] defined a distribution with probability density function (pdf) $f(x)$ and (cdf) $F(x)$ given by

$$f(x) = \frac{1}{\Gamma(a)} \{-\log[1 - G(x)]\}^{a-1} g(x) \quad (1)$$

and

$$F(x) = \frac{\gamma(a, -\log[1 - G(x)])}{\Gamma(a)} = \frac{1}{\Gamma(a)} \int_0^{-\log[1 - G(x)]} t^{a-1} e^{-t} dt \quad (2)$$

respectively, for $a > 0$, where, $g(x) = dG(x)/dx$, $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ denotes the gamma function, and $\gamma(a, z) = \int_0^z t^{a-1}e^{-t}dt$ denotes the incomplete gamma function. where $\Gamma(a, z) = \int_z^\infty t^{a-1}e^{-t}dt$ denotes the complementary incomplete gamma function. If X is a random variable with pdf, (1), we write $X \sim \text{Zografos-Balakrishnan-}G(a)$. For any baseline cdf $G(x)$, and $x \in R$, Ristić and Balakrishnan [3] defined a distribution with (pdf) $f(x)$ and cdf $F(x)$ given by

$$f(x) = \frac{1}{\Gamma(a)} \{-\log[G(x)]\}^{a-1} g(x) \quad (3)$$

and

$$F(x) = 1 - \frac{\gamma(a, -\log[G(x)])}{\Gamma(a)} = 1 - \frac{1}{\Gamma(a)} \int_0^{-\log[G(x)]} t^{a-1} e^{-t} dt \quad (4)$$

respectively, for $a > 0$. The Zografos-Balakrishnan- G and Ristić-Balakrishnan- G distributions have the same parameters of the G distribution plus an additional shape parameter $a > 0$. New Zografos-Balakrishnan- G and Ristić-Balakrishnan- G distributions can be obtained from a specified G distribution. For $a = 1$, the G distribution is a basic exemplar of the Zografos-Balakrishnan- G and Ristić-Balakrishnan- G distributions with a continuous crossover towards cases with different shapes (for example, a particular combination of skewness and kurtosis). In Section 2, we explain probability weighted moments. Mean deviations from the mean and median are obtained in Section 3. Section 4 is devoted to the Bonferroni and Lorenz curves of the ZBLL distribution. In Section 5, the model parameters are estimated by the minimum spacing distance method. In Section 6, the flexibility and potentiality of the ZBLL distribution is illustrated using to a real data sets and the new model will be compared with baseline distributions by various tools. Some concluding remarks are given in Section 7.

Probability weighted moments

The (pdf) and (cdf) of the log-logistic (LL) distribution are

$$g(x) = \frac{\beta}{\alpha^\beta} x^{\beta-1} \left[1 + \left(\frac{x}{\alpha} \right)^\beta \right]^{-2}$$

$$G(x) = 1 - \left[1 + \left(\frac{x}{\alpha} \right)^\beta \right]^{-1}$$

respectively, for $x > 0$, $\alpha > 0$, and $\beta > 0$. Inserting these expressions into (1) gives the Zografos-Balakrishnan-log-logistic (ZBLL) (Ramos et al., [1]) (pdf)

$$f_{ZBLL}(x) = \frac{\beta}{\alpha^\beta \Gamma(a)} x^{(\beta-1)} \left[1 + \left(\frac{x}{\alpha} \right)^\beta \right]^{-2} \left\{ \log \left[1 + \left(\frac{x}{\alpha} \right)^\beta \right] \right\}^{a-1} \quad (5)$$

for $x > 0$. The corresponding cumulative distribution function (7) is

$$F_{ZBLL}(x) = \frac{1}{\Gamma(a)} \int_0^x f_{ZBLL}(x) dx = \int_0^{\log \left[1 + \left(\frac{x}{\alpha} \right)^\beta \right]} t^{a-1} e^{-t} dt = \frac{\gamma \left(a, \log \left[1 + \left(\frac{x}{\alpha} \right)^\beta \right] \right)}{\Gamma(a)}. \quad (6)$$

Probability weighted moments (PWMs) are expectations of certain functions of a random variable defined when the ordinary moments of the random variable exist. The (PWMs) method can generally be used for estimating parameters of a distribution whose inverse form cannot be expressed explicitly. The (PWMs) method, which has been investigated by many researchers, was originally proposed by Greenwood et al.[4]. Since then it has been used widely in practice and for research purposes. Hosking et al. [5] investigated the properties of parameters estimated by the (PWMs) method for the generalized extreme value (GEV) distribution using fairly long observed series, and they gave a good summary of the (PWMs) method. In this paper we calculate the PWMs of the (ZBLL) distribution. For a random variable with the pdf $f(\cdot)$ and cdf $F(\cdot)$, the (PWMs) function are defined by

$$\tau_{s,r} = E \left[X^s (F(X))^r \right] = \int_0^\infty x^s (F(x))^r f(x) dx.$$

According to $\log \left[1 + \left(\frac{x}{\alpha} \right)^\beta \right] = u$ and the power series expansion

$$(1 - z)^\alpha = \sum_{j=0}^\infty (-1)^j \binom{\alpha}{j} z^j$$

gives

$$\begin{aligned} \tau_{s,r} &= \frac{1}{\Gamma(a)} \sum_{j=0}^\infty c_{r,j} \int_0^\infty \left(\alpha^s (e^u - 1)^{\frac{s}{\beta}} \right) u^{a+r+j-1} e^{-u} du \\ &= \sum_{j=0}^\infty \sum_{i=0}^\infty \binom{\frac{s}{\beta}}{i} \frac{\alpha^s (-1)^i \Gamma(a+r+j) c_{r,j}}{\Gamma(a) \left(i - \frac{s}{\beta} + 1 \right)^{a+r+j}} \end{aligned}$$

Mean deviations

The amount of scatter in a population can be measured by the totality of deviations from the mean and median. The mean deviation from the mean is a robust statistic, being more resilient to outliers in a data set than the standard deviation. In the standard deviation, the distances from the mean are squared, so on average, large deviations are weighted more heavily, and thus outliers can heavily influence it. In the mean deviation from the mean, the magnitude of the distances of a small number of outliers is

irrelevant. The mean deviation from the median is a measure of statistical dispersion. It is a more robust estimator of scale than the sample variance or standard deviation. Let X be a (ZBLL) random variable with mean $\mu = E(X)$ and median M . The median M is the solution of the nonlinear equation

$$\gamma\left(a, \log\left[1 + \left(\frac{x}{\alpha}\right)^\beta\right]\right) = 1/2.$$

The mean deviation from the mean can be defined as

$$\begin{aligned}\delta_1 &= E(|X - \mu|) = \int_0^\infty |X - \mu| f(x) dx = 2\mu F(\mu) - 2 \int_0^\mu x f(x) dx = 2 \int_0^\mu F(x) dx \\ &= 2 \int_0^\mu \sum_{k=0}^{\infty} \frac{(-1)^k}{k! (a+k)\Gamma(a)} \left\{ \log\left[1 + \left(\frac{x}{\alpha}\right)^\beta\right] \right\}^{a+k} dx\end{aligned}$$

Using the change of variable $\log\left[1 + \left(\frac{x}{\alpha}\right)^\beta\right] = u$ gives

$$\begin{aligned}\delta_1 &= \frac{2\alpha}{\beta} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \left(\frac{1-\beta}{\beta}\right)^j \frac{(-1)^{k+j}}{k! (a+k)\Gamma(a)} \int_0^\mu u^{a+k} e^{-\frac{u(\beta j-1)}{\beta}} du \\ &= \frac{2\alpha}{\beta} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \left(\frac{1-\beta}{\beta}\right)^j \frac{(-1)^{k+j} \beta^{a+k} \gamma\left(a+k+1, \frac{\mu(\beta j-1)}{\beta}\right)}{k! (a+k)\Gamma(a) (\beta j-1)^{a+k+1}}\end{aligned}$$

The mean deviation from the median is, also, defined by

$$\begin{aligned}\delta_2 &= E(|X - M|) = \int_0^\infty |x - M| f(x) dx = \mu - M + 2 \int_0^M F(x) dx = \mu - M \\ &\quad + \frac{2\alpha}{\beta} \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \left(\frac{1-\beta}{\beta}\right)^j \frac{(-1)^{k+j} \beta^{a+k} \gamma\left(a+k+1, \frac{M(\beta j-1)}{\beta}\right)}{k! (a+k)\Gamma(a) (\beta j-1)^{a+k+1}}\end{aligned}$$

The Bonferroni and Lorenz curves

Study of income inequality has gained a lot of importance over the last many years. Lorenz curve and the associated Gini index are undoubtedly the most popular indices of income inequality. However, there are certain measures which despite possessing interesting characteristics have not been used often for measuring inequality. Bonferroni curve and scaled total time on test transform are two such measures, which have the advantage of being represented graphically in the unit square and can also be related to the Lorenz curve and Gini ratio Giorgi [6]. These two measures have some applications in reliability and life testing as well Giorgi and Crescenzi [7]. The Bonferroni and Lorenz curves and Giniindex have many applications not only in economics to study

income and poverty, but also in other fields like reliability, medicine and insurance. For a random variable X with (cdf $(F(\cdot))$, the Bonferroni curve is given by

$B_F[F(x)] = \frac{1}{\mu F(x)} \int_0^x u f(u) du$. Using the change of variable $\log \left[1 + \left(\frac{x}{a} \right)^\beta \right] = u$ and $x = a \sum_{i=0}^{\infty} \binom{1/\beta}{i} (-1)^i e^{-\frac{u(\beta j-1)}{\beta}}$, The Bonferroni curve of the (GLL) distribution is given by

$$B_F[F(x)] = \frac{\alpha}{\Gamma(a)} \sum_{i=0}^{\infty} \binom{1/\beta}{i} (-1)^i \int_0^x u^{a-1} e^{-\frac{u(\beta j-1)}{\beta}} du$$

Using the change of variable $\frac{u(\beta j-1)}{\beta} = w$, We obtain

$$B_F[F(x)] = \frac{\alpha \beta^a}{\mu F(x) \Gamma(a)} \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{1}{\beta} + 1\right) (-1)^i \gamma\left(a, \frac{x(\beta j-1)}{\beta}\right)}{i! \Gamma\left(\frac{1}{\beta} - i + 1\right) (\beta j - 1)^a}$$

The Lorenz curve of the ZBLL distribution can be obtained via the expression

$$L_F[F(x)] = B_F[F(x)]F(x) = \frac{\alpha \beta^a}{\mu \Gamma(a)} \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{1}{\beta} + 1\right) (-1)^i \gamma\left(a, \frac{x(\beta j-1)}{\beta}\right)}{i! \Gamma\left(\frac{1}{\beta} - i + 1\right) (\beta j - 1)^a}$$

where μ is the mean of the (ZBLL) distribution.

Minimum spacing distance estimator

Torabi [8] introduced a general method for estimating parameters through spacing called maximum spacing distance estimator (MSDE). Torabi and Bagheri [9] and Torabi and Montazeri [10] used different MSDEs to compare with the MLEs. Here, we used a (MSDE), "minimum spacing absolute distance estimator" (MSSDE) of the (ZBLL) distribution.

Let X_1, X_2, \dots, X_{n-1} be a random sample from some continuous distribution function F_θ , $\theta \in \Theta$ with support on \mathcal{R} . Here, the unknown parameter θ maybe a vector. Just like in rank-theory, the continuity assumption eliminates the possibility of two observations being equal. Let the order statistics be denoted by Y_1, \dots, Y_{n-1} . Define

$$D_i(\theta) = F_\theta(Y_i) - F_\theta(Y_{i-1}), \quad i = 1, 2, \dots, n \quad (7)$$

where $F_\theta(Y_0) = 0$ and $F_\theta(Y_n) = 1$. We define minimum spacing distance estimator (MSDE) of θ by minimizing

$$T(\theta) = \sum_{i=1}^n h(D_i(\theta), 1/n), \quad (8)$$

in which $h(x, y)$ is a appropriate distance. Some standard choices of $h(x, y)$ are $h(x, y) = (x - y)^2$, $(\ln x - \ln y)^2$, $|x - y|$, $|\ln - \ln y|$ and $e^{x-y} - (x - y) - 1$, which is called square, square-log, absolute, absolute-log and line distance, respectively.

Minimum spacing square distance estimator

Let X_1, X_2, \dots, X_{n-1} be a random sample from (ZBLL) distribution, and let the order statistics be denoted by Y_1, \dots, Y_{n-1} . According to equations (7) and (8), for finding (MSSDE), note that

$$T(\theta) = \left[\frac{\gamma \left(a, \log \left(1 + \left(\frac{Y_1}{\alpha} \right)^\beta \right) \right)}{a} - \frac{1}{n} \right]^2 + \sum_{i=2}^{n-1} \left[\frac{\gamma \left(a, \log \left(1 + \left(\frac{Y_i}{\alpha} \right)^\beta \right) \right) - \gamma \left(a, \log \left(1 + \left(\frac{Y_{i-1}}{\alpha} \right)^\beta \right) \right)}{a} - \frac{1}{n} \right]^2 + \left[\frac{1 - \gamma \left(a, \log \left(1 + \left(\frac{Y_{n-1}}{\alpha} \right)^\beta \right) \right)}{a} - \frac{1}{n} \right]^2$$

The associated score function is given by $U_n(\theta) = \left(\frac{\partial T(\theta)}{\partial \alpha}, \frac{\partial T(\theta)}{\partial \beta}, \frac{\partial T(\theta)}{\partial a} \right)^T$,

where

$$\begin{aligned} \frac{\partial T(\theta)}{\partial \alpha} &= \frac{2}{a} \left\{ U_1 V_1 S_1 + \sum_{i=1}^{n-1} (U_i V_i - U_{i-1} V_{i-1}) \left(S_i - S_{i-1} - \frac{1}{n} \right) - U_{n-1} V_{n-1} \left(\frac{1}{a} - \frac{2}{n} - S_{n-1} \right) \right\} \\ \frac{\partial T(\theta)}{\partial \beta} &= \frac{2}{a} \left\{ \sum_{i=1}^{n-1} \left(U_{i-1} V_{i-1} \log \left(\frac{Y_{i-1}}{\alpha} \right)^\beta - U_i V_i \log \left(\frac{Y_i}{\alpha} \right)^\beta \right) \left(S_i - S_{i-1} - \frac{1}{n} \right) + U_{n-1} V_{n-1} \log \left(\frac{Y_{n-1}}{\alpha} \right)^\beta \left(\frac{1}{a} - \frac{2}{n} - S_{n-1} \right) - U_1 V_1 S_1 \log \left(\frac{Y_1}{\alpha} \right)^\beta \right\} \end{aligned}$$

The (MSSDE) of Θ , say $\hat{\Theta}$, is obtained by solving the nonlinear system $U_n(\Theta) = 0$.

In this section, we provide an application to real data to illustrate the (ZBLL) distribution. The (MLEs) of the parameters are computed and the goodness-of-fit statistics for this model is compared with other competing models. The data correspond to the age of death in years of retired women who died during 2004 with temporary disabilities. The data was provided by the Mexican Institute of Social Security (IMSS) to study the distributional behavior of the mortality of retired people on disability. The data set is:

22, 24, 25, 25, 27, 28, 29, 29, 29, 29, 30, 31, 31, 31, 31, 31, 31, 32, 32, 32, 32, 32, 32,
32, 33, 33, 33, 34, 34, 34, 34, 34, 34, 35, 35, 35, 35 , 36, 36, 36, 36, 36, 36, 36, 36, 36,
36, 36, 37, 37, 37, 37, 37, 38, 38, 38, 39, 39, 39, 39, 39, 39, 40, 40, 40, 40, 40, 40, 40,
40, 40, 40, 40, 41, 41, 41, 41, 41, 41, 41, 41, 41, 41, 41, 42, 42, 42, 42, 42, 42, 43,
43, 43, 43, 43, 44, 44, 44, 44, 44, 44, 44, 45, 45, 45, 45, 45, 45, 45, 45, 45, 45, 46, 46,
46, 46, 46, 46, 47, 47, 47, 47, 47, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 48, 49, 49, 49,
49, 49, 49, 49, 49, 50, 50, 50, 50, 50, 50, 50, 50, 51, 51, 51, 51, 51, 51, 51, 51, 52, 52,
52, 52, 52, 52, 52, 52, 52, 52, 52, 52, 52, 53, 53, 53, 53, 53, 53, 53, 53, 53, 53, 54,
54, 54, 54, 54, 54, 54, 54, 54, 54, 54, 54, 55, 55, 55, 55, 55, 55, 55, 55, 55, 55, 55,
56, 56, 56, 56, 56, 56, 56, 56, 56, 56, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57, 57,
57, 57, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 58, 59, 59, 59, 59, 59, 59, 59, 59, 59, 60,
60, 60, 60, 60, 60, 60, 70, 61, 61, 62, 63, 64, 64, 64, 64, 65, 65, 66, 66, 66, 71, 74, 75,
79, 86. The data set is given in Balakrishnan et al. [11].

We fit the (ZBLL), Exponentiated Log-Logistic (ELL), Log-Logistic Weibull (LLW) (Ouyede et al. (2016)) and LL distributions to data set. The densities of the (ELL) and (LLW) distributions are, respectively, given by:

$$f_{ELL}(x) = \frac{a\beta}{\alpha^{a\beta}} x^{a\beta-1} \left[1 + \left(\frac{x}{\alpha} \right)^\beta \right]^{-(a+1)}, x > 0, \alpha, \beta, a > 0,$$

$$f_{LLW}(x) = \frac{a\beta}{\alpha} \left(\frac{x}{\alpha} \right)^{\beta-1} \left[1 + \left(\frac{x}{\alpha} \right)^\beta \right]^{-(a+1)}, x > 0, \alpha, \beta, a > 0.$$

For all models, the MSSEs are computed. Further, Akaike information criterion (AIC), Consistent Akaike information criterion (CAIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC), Anderson-Darling (AD), Cramér_von Mises (CM) and Kolmogorov-Smirnov (K-S) (p-value) statistics are calculated to compare the fitted models. The statistics (AD) and (CM) are defined by Chen and Balakrishnan [12]. In general, the smaller the values of these statistics, the better the fit to the data. The required computations are carried out in R-language.

Table 1 list the MSSEs of the parameters. The values of the model selection statistics (AIC), (CAIC), (BIC), (HQIC), (AD), (CM) and (K-S) (p-value). We note from Table 1 that the

ZBLL distribution has the lowest values of the (AIC), (CAIC), (BIC), (HQIC), (AD), (CM) and (K-S) statistics among the fitted (ELL), (LLW) and (LL) distributions, thus suggesting that the

Table.1. MSSDE estimates of the model parameters for the data set and the (AIC), (CAIC), (BIC), (HQIC), (CM), (AD) and (K-S) (p-value) statistics.

	ZBLL	LLW	ELL	LL
$\hat{\alpha}$	0.03366	0.00345	65.2919	47.5155
$\hat{\beta}$	0.27960	0.02394	5.84693	7.56921
\hat{a}	3.59955	2.00791	4.44557	-----
AIC	1741.0	1808.8	2099.8	2122.6
BIC	1735.9	1836.6	2110.7	2139.1
CAIC	1741.1	1808.9	2099.9	2120.6
HQIC	1745.4	1813.2	2104.2	2123.5
CM	0.14554	0.25188	0.33129	0.51744
AD	0.15041	0.23529	0.23627	0.34457
K-S (p-value)	0.0336 (0.965)	0.0475(0.889)	0.1545(0.624)	0.1761(0.576)

(ZBLL) model provide the best fits, and therefore could be chosen as the most adequate model for the this data set. The histogram of this data and the estimated pdfs and cdfs of the (ZBLL) distribution and their competitive distributions are displayed in Figure 1. It is clear that the (ZBLL) distribution provides a better fit than the other distributios.

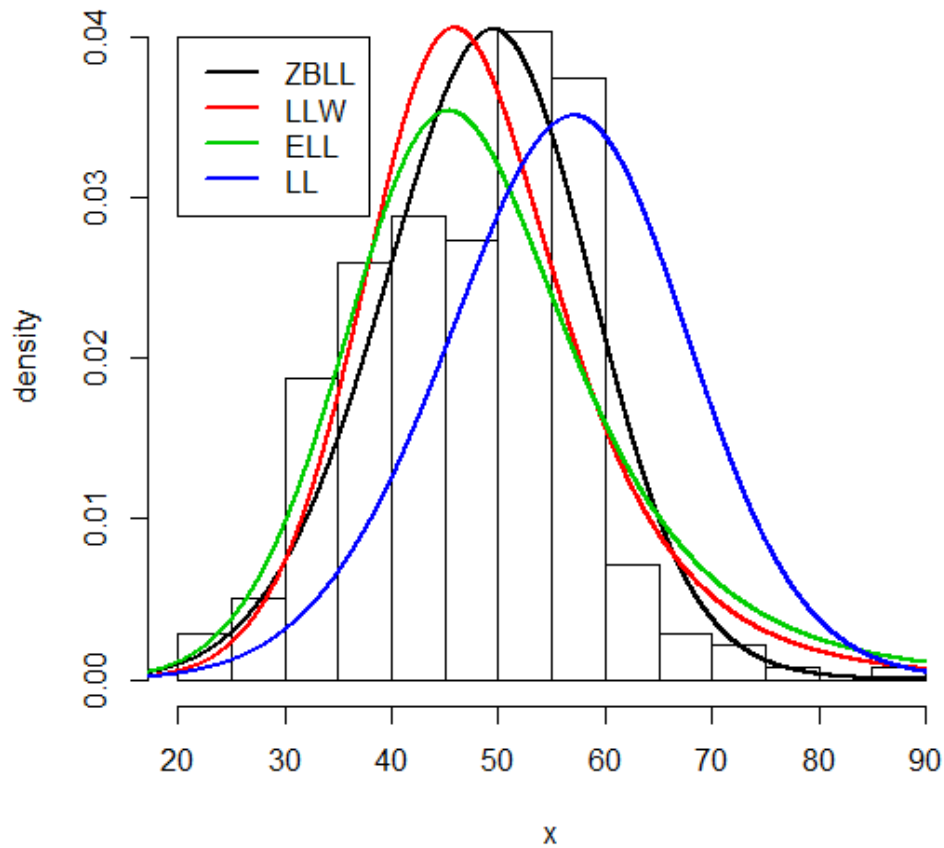


Figure1. Fitted pdf of the ZBLL, LLW, ELL and LL distributions for the data set corresponding to Tables 1

Conclusion

A new three-parameter distribution called the (ZBLL) distribution was introduced by Ramos et al. [1]. A characteristic of the (ZBLL) distribution is that its failure rate function can be decreasing, increasing, bathtub-shaped and unimodal depending on its parameter values Ramos et al. [1]. Several properties of this distribution such as probability weighted moments, mean deviations and Bonferroni and Lorenz curves is obtained. The minimum spacing square distance estimator procedure is presented. Fitting the (ZBLL) distribution to a real data set indicates the flexibility and capacity of the this distribution in reliability analysis, biological systems, data modeling, and related fields.

References

1. Ramos M. W. A., "The Zografos-Balakrishnan Log-Logistic Distribution", Properties and Application, Journal of Statistical Theory and Applications, 12(3) (2013) 255-244.
2. Zografos K., Balakrishnan N., "On families of beta- and generalized gamma-generated distributions and associated inference", Stat. Methodol., 6(2009) 344-362.
3. Ristić M. M., Balakrishnan N., "The gamma exponentiated exponential Distribution", J. Stat. Comput. Simul., 82(2012)1191-1206.
4. Greenwood J. A., Landwehr J. M., Matalas N. C., Wallis J. R., "Probability weighted moments: definition and relation to parameters of several distribution expressible in inverse form", Water Resources Research, 15(1979)1049-1054.
5. Hosking J. R. M., Wallis J. R., Wood E. F., "Estimation of the generalized extreme value distribution by the method of probability weighted moments", Technometrics, 27(1985) 251-261.
6. Giorgi G. M., "Concentration Index", Bonferroni, Encyclopedia of Statistical Sciences 2, Wiley, New York, (1998)141-146.
7. Giorgi G. M. M., Crescenzi M., "A look at the Bonferroni inequality measure in a reliability framework", Statistica LXL, 4(2001)571-583.
8. Torabi H., "A general method for estimating and hypotheses testing using spacings", Journal of Statistical Theory and Applications 8 (2008) 163-168.
9. Torabi H., Bagheri F. L., "Estimation of parameters for an extended generalized half-logistic distribution based on complete and censored data", Journal of the Iranian Statistical Society, 9 (2010) 171-195.
10. Torabi H., Montazari N. H., "The gamma-uniform distribution and its application", kybernetika, 48(2012) 16-30.
11. Balakrishnan N., Leiva V., Sanhueza A., Cabrera E., "Mixture inverse Gaussian distributions and its transformations", moments and applications, Statistics, 43(2009) 91-104.
12. Chen G., Balakrishnan N. A., "General purpose approximate goodness-of-fit test", Journal of Quality Technology, 27(1995) 154-161.