

Frequentist Parameter Estimation of Two-Parameter Exponentiated Log-logistic Distribution BB

Arun Kumar Chaudhary, Ph.D.

Associate Professor

Nepal Commerce Campus, T.U.

Abstract

In this paper, we have considered frequentist parameter estimation of two-parameter exponentiated log-logistic distribution based on a complete sample. The parameters are estimated using a likelihood-based classical inferential procedure. This study has tried to compute MLEs along with their asymptotic confidence interval, contour plot and standard errors. The K-S test statistic, Quantile-Quantile (QQ) and Probability-Probability (PP) plots are used to check the validity of the model. All the computations are performed in R software. A real data set is considered for illustration of the purposed inferential procedures.

Keywords – *Exploratory data analysis, Contour plot, Exponentiated Log-logistic distribution, Maximum likelihood estimation, Model validation.*

Introduction

The log-logistic distribution is very useful in survival analysis since it has a nonmonotonic hazard function, (Bennett, 1983) and (Tadikamalla and Johnson, 1982). When dealing with censored data, which makes this distribution more convenient is that the shape of this distribution has a more manageable form than that of the log-normal, even though the shape is similar. (Srivastava and Shukla, 2008) studied the log-logistic distribution as a step-stress model. (Balakrishnan and Malik, 1987) gave the moments of order statistics from the truncated log-logistic distribution. This distribution has been also studied by Howlader and Weiss, (1992). Lawless, (2003), Lee and Wang, (2003) and Murthy et al., (2004) provide an excellent review for the log-logistic distribution.

Based on modifications such as adding parameters to the existing models, new classes of models have been proffered in recent years. Adding one or more parameters to distribution makes it richer and more flexible for modeling data.

There are different ways for adding a parameter(s) to a distribution. (Marshall and Olkin (1997, 2007) added one positive parameter to a given (general) survival function. As described by Marshall and Olkin, (2007) and Klugman et al., (2012), an exponentiated distribution can be easily constructed. It is based on the observation that by raising any baseline Cumulative Distribution Function (CDF) $F_{baseline}(x)$ to arbitrary power $\alpha > 0$, a new CDF

$$F(x) = (F_{baseline}(x))^{\alpha} ; \alpha > 0 \quad (1.1)$$

is obtained with the additional parameter α .

This idea led several authors to consider extensions from usual survival distributions. The exponentiated Weibull distribution was considered by (Mudholkar and Srivastava, 1993) as a generalization

of the Weibull distribution. The reliability test plan for exponentiated log-logistic distribution was studied by (Rosaiah, et al. (2006, 2007)). The Kumaraswamy-log-logistic distribution was introduced by (Santana et al., 2012) which includes exponentiated log-logistic distribution. Chaudhary and Kumar (2014) obtained the maximum likelihood and Bayes estimators for the parameters, the posterior predictive check procedure for evaluating the model fit and the reliability function of the three-parameter exponentiated log-logistic distribution using Markov Chain Monte Carlo simulation method.

The CDF of the log-logistic distribution is given by

$$F_{LL}(x; \beta, \lambda) = \frac{(x/\lambda)^\beta}{1+(x/\lambda)^\beta}; (\beta, \lambda) > 0, x > 0 \tag{1.2}$$

where $\beta > 0$ is the shape and $\lambda > 0$ is the scale parameter.

The CDF of the exponentiated log-logistic(ELL) distribution is defined by raising $F_{LL}(x)$ to the power of α , namely $F(x) = (F_{LL}(x))^\alpha$. The distribution function of ELL distribution with three parameters is given by

$$F(x; \alpha, \beta, \lambda) = \left\{ \frac{(x/\lambda)^\beta}{1+(x/\lambda)^\beta} \right\}^\alpha; (\alpha, \beta, \lambda) > 0, x > 0 \tag{1.3}$$

Where shape parameters are $\alpha > 0$ and $\beta > 0$ and the scale parameter is $\lambda > 0$. The model reduces to the log-logistic distribution when $\alpha = 1$. $\lambda = 1$ can be assumed without any loss of generality since λ it is the scale parameter. When $\lambda = 1$ three-parameter ELL distribution reduces to two-parameter ELL distribution and we shall denote it as $ELL(\alpha, \beta)$.

The rest of the article is arranged as given. The model and its features are introduced in Section 2. The real data set and its exploratory data analysis, maximum likelihood estimation (MLE) and model validation are described in Section 3. Conclusions are given in Section 4.

The Exponentiated Log-logistic Model

The Cumulative Distribution Function (CDF) of two-parameter exponentiated log-logistic (ELL) distribution is given by

$$F(x; \alpha, \beta) = \left\{ \frac{x^\beta}{1+x^\beta} \right\}^\alpha; (\alpha, \beta) > 0, x > 0 \tag{2.1}$$

where the shape parameters are $\alpha > 0$ and $\beta > 0$.

And the corresponding Probability Density Function (PDF) with shape parameters $\alpha > 0$ and $\beta > 0$ is given by

$$f(x; \alpha, \beta) = \frac{\alpha \beta x^{\alpha\beta}}{x [1+x^\beta]^{\alpha+1}}; (\alpha, \beta) > 0, x > 0. \tag{2.2}$$

The reliability/survival function is given by

$$R(x; \alpha, \beta) = 1 - \left\{ \frac{x^\beta}{1+x^\beta} \right\}^\alpha; (\alpha, \beta) > 0, x > 0. \tag{2.3}$$

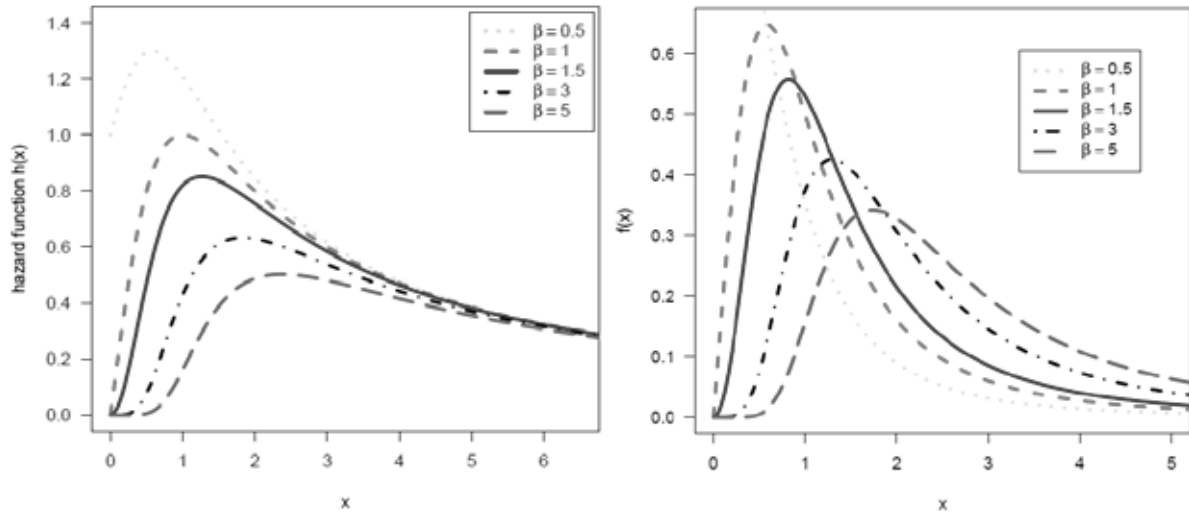


Figure 1 The probability density function (Left panel); The hazard function (Right panel) of $ELL(\alpha, \beta)$ distribution for $\alpha = 1$ and different values of β .

The hazard rate function with shape parameters $\alpha > 0$ and $\beta > 0$ is

$$h(x; \alpha, \beta) = \frac{\alpha \beta x^{\alpha\beta}}{x(1+x^\beta)^{\alpha+1}} \left(1 - \left\{ \frac{x^\beta}{1+x^\beta} \right\}^\alpha \right)^{-1}; (\alpha, \beta) > 0, x > 0. \tag{2.4}$$

The quantile function is given by

$$x_p = \left(p^{-1/\alpha} - 1 \right)^{-1/\beta}; 0 < p < 1. \tag{2.5}$$

The random deviate can be generated from $ELL(\alpha, \beta)$ by

$$x = \left(u^{-1/\alpha} - 1 \right)^{-1/\beta}; 0 < u < 1 \tag{2.6}$$

where u have the $U(0, 1)$ distribution.

The r^{th} moment and mode are given by

$$\mu'_r = \frac{\alpha}{\Gamma(\alpha+1)} \Gamma\left(\alpha + \frac{r}{\beta}\right) \Gamma\left(1 - \frac{r}{\beta}\right); -\beta < r < \beta \tag{2.7}$$

$$\text{and Mode} = \left(\frac{\alpha\beta - 1}{\beta + 1} \right)^{1/\beta}; \alpha\beta \geq 1.$$

Some of the typical $ELL(\alpha, \beta)$ density functions for different values of β and $\alpha = 1$ are depicted in Figure 1 (left panel). It is evident from Figure 1 that the density function of the ELL distribution can take different shapes. Figure 1(right panel) exhibits the different hazard rate functions of $ELL(\alpha, \beta)$ distribution.

Data, Maximum likelihood estimation and Model validation

The following real data set is considered for illustration of the proposed methodology. The data given below represent active repair times (in hours) for 46 repair times of an airborne communication

transceiver. (Chhikara and Folks, 1977) fitted a two-parameter inverse Gaussian distribution. The data are presented below:

0.2, 0.3, 0.5, 0.5, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 1.0, 1.0, 1.0, 1.0, 1.1, 1.3, 1.5, 1.5, 1.5, 2.0, 2.0, 2.2, 2.5, 2.7, 3.0, 3.0, 3.3, 3.3, 4.0, 4.0, 4.5, 4.7, 5.0, 5.4, 5.4, 7.0, 7.5, 8.8, 9.0, 10.3, 22.0, 24.5

Exploratory data analysis (EDA)

Exploratory data analysis (EDA) provides a variety of graphical tools for quickly summarizing and gaining insight into a set of data (Tukey, 1977). It allows data to speak for themselves, without making assumptions and conducting formal analyses. The descriptive statistical methods quantitatively describe the main features of data.

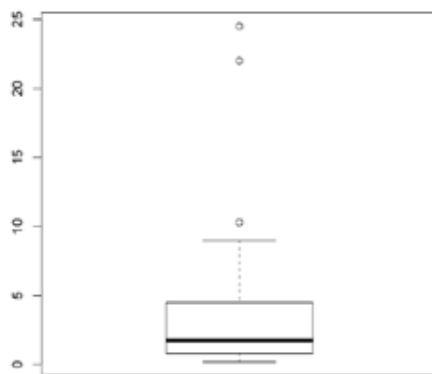


Table -1 Summary statistics

Minimum	0.200
First Quartile (Q ₁)	0.800
Median	1.750
Mean	3.607
Third Quartile (Q ₃)	4.375
Maximum	24.500
Kurtosis	8.295
Skewness	2.795

Figure 2 Boxplot

The main data features are (i) measures of central tendency(e.g. mean and median); (ii) measures of variability (e.g., standard deviation) and (iii) measures of relative standing (e.g.,quantiles). The descriptive statistics for the above data set are presented in Table 1. We have plotted the boxplot in Figure 2, which shows that the data set contains three "outliers".

The estimation of the parameter of the proposed model is obtained by the method of maximum likelihood(ML) estimation.

Maximum likelihood estimation (MLE) and asymptotic confidence intervals

In this section, we briefly discuss the maximum likelihood estimators (MLE's) of the two-parameter ELL distribution and discuss their asymptotic properties to obtain approximate confidence intervals based on MLE's.

Let a random sample $\underline{x}=(x_1, \dots, x_n)$ of size n can be taken from $ELL(\alpha, \beta)$, then the log-likelihood function $\ell(\alpha, \beta)$ can be written as;

$$\ell(\alpha, \beta) = n \log \alpha + n \log \beta + \alpha \sum_{i=1}^n \log(x_i^\beta) - \sum_{i=1}^n \log x_i - (\alpha + 1) \sum_{i=1}^n \log(1 + x_i^\beta) . \quad (4.1)$$

Therefore, in order to obtain the MLE's of α and β , we can maximize (4.1) directly with respect to α and β or the following two non-linear equations can be solved using iteration method e.g. Newton-Raphson method

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \log(x_i^\beta) - \sum_{i=1}^n \log(1 + x_i^\beta) = 0$$

$$\frac{\partial \ell}{\partial \beta} = \frac{n}{\beta} + \alpha \sum_{i=1}^n \log(x_i) - (\alpha + 1) \sum_{i=1}^n \left(\frac{x_i^\beta}{1 + x_i^\beta} \right) \log(x_i) = 0 .$$

Let us denote the parameter vector by $\underline{\theta} = (\alpha, \beta)$ and the corresponding MLE of $\underline{\theta}$ as $\hat{\underline{\theta}} = (\hat{\alpha}, \hat{\beta})$,

then the asymptotic normality results in

$$(\hat{\underline{\theta}} - \underline{\theta}) \rightarrow N_2\left(0, (I(\underline{\theta}))^{-1}\right) \tag{4.2}$$

where $I(\underline{\theta})$ is the Fisher's information matrix given by

$$I(\underline{\theta}) = - \begin{bmatrix} E\left(\frac{\partial^2 \ell}{\partial \alpha^2}\right) & E\left(\frac{\partial^2 \ell}{\partial \alpha \partial \beta}\right) \\ E\left(\frac{\partial^2 \ell}{\partial \beta \partial \alpha}\right) & E\left(\frac{\partial^2 \ell}{\partial \beta^2}\right) \end{bmatrix}. \tag{4.3}$$

In practice, as we do not know $\underline{\theta}$, it is useless that the MLE has asymptotic variance $(I(\underline{\theta}))^{-1}$. Hence, the asymptotic variance can be approximated by "plugging in" the estimated value of the parameters. The common procedure is to use the observed Fisher information matrix $O(\hat{\underline{\theta}})$ (as an estimate of the information matrix $I(\underline{\theta})$) given by

$$O(\hat{\underline{\theta}}) = - \begin{pmatrix} \frac{\partial^2 \ell}{\partial \alpha^2} & \frac{\partial^2 \ell}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \ell}{\partial \beta \partial \alpha} & \frac{\partial^2 \ell}{\partial \beta^2} \end{pmatrix} \Big|_{(\hat{\alpha}, \hat{\beta})} = -H(\theta) \Big|_{\underline{\theta}=\hat{\underline{\theta}}} \tag{4.4}$$

where H is the Hessian matrix, $\underline{\theta} = (\alpha, \beta)$ and $\hat{\underline{\theta}} = (\hat{\alpha}, \hat{\beta})$. To maximize the likelihood, the Newton-Raphson algorithm produces the observed information matrix. Therefore, the variance-covariance matrix is given by

$$\left(-H(\underline{\theta}) \Big|_{\underline{\theta}=\hat{\underline{\theta}}}\right)^{-1} = \begin{pmatrix} var(\hat{\alpha}) & cov(\hat{\alpha}, \hat{\beta}) \\ cov(\hat{\beta}, \hat{\alpha}) & var(\hat{\beta}) \end{pmatrix}. \tag{4.5}$$

Hence, approximate $100(1-\gamma)\%$ confidence intervals for α and β can be constructed from the asymptotic normality of MLEs as

$$\hat{\alpha} \pm z_{\gamma/2} \sqrt{var(\hat{\alpha})} \quad \text{and} \quad \hat{\beta} \pm z_{\gamma/2} \sqrt{var(\hat{\beta})} \tag{4.6}$$

where $z_{\gamma/2}$ is the upper percentile of standard normal variate.

Computation of MLE

We have started the iterative procedure by maximizing the log-likelihood function given in the equation (4.1) directly with an initial guess for $\alpha = 0.5$ $\beta = 0.5$ and, far away from the solution. We have used `optim()` function in R, (R Development Core Team, 2013) and (Rizzo, 2008), with the option Newton-Raphson method. The iterative process stopped only after 26 iterations. We obtain $\hat{\alpha} = 1.838$ and $\hat{\beta} = 1.3297$ and the corresponding log-likelihood value is $\ell(\hat{\alpha}, \hat{\beta}) = -100.474$. We have plotted the contour plot $\ell(\alpha, \beta)$ in Figure 3, the (+) indicates the MLE.

The 95% confidence interval is computed using (4.5) and (4.6). Table 2 shows the ML estimates, standard error (SE) and 95 % Confidence Intervals for the parameters alpha and beta.

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are defined as

$$AIC = -2 \ell(\hat{\theta}) + 2 p \text{ and}$$

$$BIC = -2 \ell(\hat{\theta}) + p \log(n)$$

where $\hat{\theta} = (\hat{\alpha}, \hat{\beta})$ is the ML estimate of $\theta = (\alpha, \beta)$ and p is the number of parameters estimated in the model. The smaller the value of AIC and BIC, the better will be the model. The values of the information measures are AIC= 204.9 and BIC = 208.6, respectively.

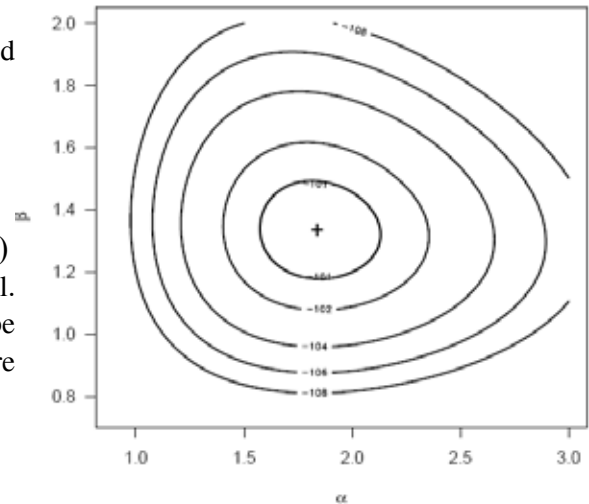


Table 2 MLE, standard error and 95% confidence interval

Parameter	MLE	Std. Error	95% Confidence Interval
alpha	1.8381	0.27185	(1.3053, 2.3709)
beta	1.3297	0.15382	(1.0282, 1.6312)

Model Validation

Once the parameters have been obtained by the method of maximum likelihood, the Kolmogorov-Smirnov (KS) distance between the fitted distribution function and the empirical distribution function is used to check the validity of the model.

The graphical methods Quantile-Quantile (QQ) and Probability-Probability (PP) plots are used for the suitability of the model under consideration. The value of the K-S test statistic is 0.0899 and the corresponding p-value is given by 0.8514. The higher p-value clearly indicates that ELL distribution can be used to analyze the given data set, and we have also plotted the empirical distribution function and the fitted distribution function in Figure 4. It is clear that the estimated ELL distribution provides a reasonable fit to the given data, (Kumar and Ligges, 2011).

Further support for this finding can be obtained by inspecting the probability-probability (P-P) and quantile-quantile (Q-Q) plots. The P-P plot shows the empirical and theoretical distribution functions. The estimated versus the observed quantiles can be shown by the Q-Q plot. As can be seen from the straight line pattern in Figure 5 the ELL fits the data well.

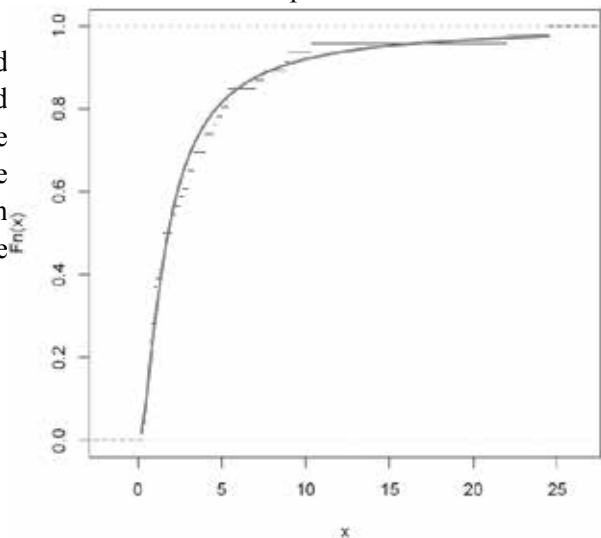


Figure 4 The empirical and fitted distribution function.

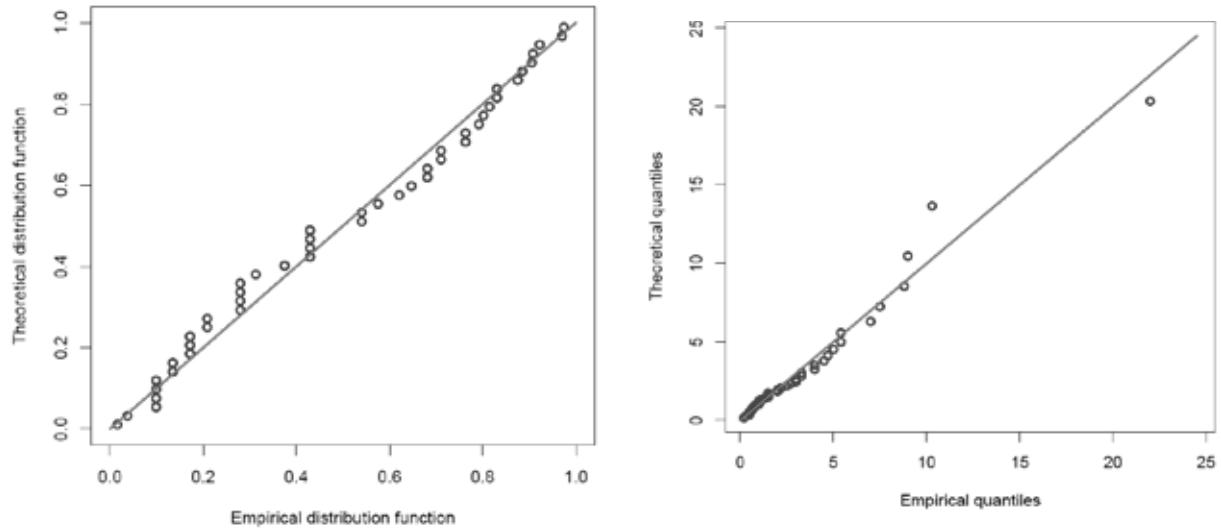


Figure 5 Probability-Probability (PP) plot (left panel); Quantile-Quantile (QQ) plot (right panel) using MLEs as estimate

Conclusions

We have discussed frequentist inferential procedures in order to compute parameters of exponentiated log-logistic distribution by the method of maximum likelihood (ML) estimation. We have obtained the approximate confidence intervals for parameters and contour plot of the log-likelihood function. We have used exploratory data analysis techniques and the descriptive statistical methods which describe the main features of data. The K-S test statistic, the graphical methods Q-Q and P-P plots are used for the suitability of the model under consideration. The value of K-S test statistic, Q-Q, and P-P plots show that the two-parameter exponentiated log-logistic distribution fits the data well.

References

- Balakrishnan, N. and Malik, H.J. (1987). Moments of order statistics from truncated log-logistic distribution. *Journal of Statistical Planning and Inference*. 17: 251-267.
- Bennete, S. (1983). Log-logistic regression models for survival data. *Applied Statistics*. 32: 165 – 171.
- Chaudhary, A.K. and Kumar, V.(2014), “Bayesian Estimation of Three- Parameter Exponentiated Log-logistic Distribution”*International journal of Statistika and Matematika*,9 (2),66-81.
- Chhikara, R.S. and Folks, J.L. (1977). The inverse Gaussian distribution as a lifetime model. *Technometrics*. 19: 461-468.
- Gupta, R. D., and Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*. 41(2): 173 - 188.
- Howlader, H.A. and Weiss, G.(1992). Log-logistic survival estimation based on failure-censored data. *Journal of Applied Statistics*. 19, No. 2: 231-240.
- Klugman, S.; Panjer, H. and Willmot, G. (2012) *Loss Models: From Data to Decisions*. (4th edition). John Wiley & Sons, New York.
- Kumar, V. and Ligges, U. (2011). *reliaR: A package for some probability distributions*. <http://cran.r-project.org/web/packages/reliaR/index.html>.
- Lawless, J. F., (2003). *Statistical Models and Methods for Lifetime Data*. 2nd ed., John Wiley and Sons, New York.
- Marshall, A. W., and Olkin, I. (1997). A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. *Biometrika*. 84(3): 641–652.

-
- Marshall, A. W., and Olkin, I. (2007). *Life Distributions: Structure of Nonparametric, Semiparametric and Parametric Families*. Springer, New York.
- Mudholkar, G.S. and Srivastava, D.K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*. 42(2): 299–302.
- Murthy, D.N.P., Xie, M. and Jiang, R. (2004). *Weibull Models*. Wiley, New York.
- Nadarajah, S. and Kotz, S. (2006). The exponentiated type distributions. *Acta Applicandae Mathematicae*. 92: 97-111.
- R Development Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rizzo, M. L.(2008). *Statistical computing with R*. Chapman & Hall/CRC.
- Rosaiah, K.; Kantam, R.R.L. and Kumar, Ch. S. (2006). Reliability test plans for exponentiated log-logistic distribution. *Economic Quality Control*. 21(2) : 165-175.
- Rosaiah, K.; Kantam, R.R.L. and Santosh Kumar, Ch. (2007). Exponentiated log-logistic distribution - an economic reliability test plan. *Pakistan Journal of Statistics*. 23 (2) : 147 -156.
- Santana, T.V.F; Ortega, E.M.M.; Cordeiro, G.M. and Silva, G.O. (2012). Kumaraswamy-Log-Logistic Distribution. *Journal of Statistical Theory and Applications*. 11(3) : 265-291.
- Srivastava, P.W. and Shukla, R. (2008). A Log-Logistic Step-Stress Model. *IEEE Transactions on Reliability*. 57(3) : 431-434
- Tukey, J.W. (1977). *Exploratory Data Analysis*. Reading, Mess. Addison-Wesley.