International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Discrimination Between Logistic and Gumbel Distribution

S. A. Al-Subh[a*], M. T. Alodat[b]

[a]*Dept. of Math. & Stat., Mutah University, Karak, Jordan*

[b]*Dept. of Stat., Yarmouk University, Irbid, Jordan*

[a]*Email: salsubh@yhoo.com*

[b]*Email: alodatmts@yahoo.com*

Abstract

When two distributions have ,approximately ,the same characteristics, it is often difficult to discriminate between them. In this study, we use the ratio of likelihoods for selecting between the logistic and Gumbel distributions for describing a set of data. The parameters for the logistic and Gumbel distributions are estimated by using maximum likelihood (ML), moments (MOM) and order statistic (OS) methods. In addition, by using Monte Carlo simulations, discriminating between the two distributions is investigated in terms of the probability of correct selection (PCS) as found based on the different methods of estimation. In general, it is found that the method of ML outperforms all the other methods when the estimators considered are compared in term of efficiency.

Keywords: Logistic distribution; Gumbel distribution; Discriminating; Maximum likelihood; Moment; Order statistic.

1. Introduction

According to Reference [19], the logistic distribution has interesting application in many fields, such as public health, graduation of mortality statistics, survival data, income distribution; human population and biology. When compared to other distributions, including logistic, the normal distribution is the most widely used family of distributions in statistics and many statistical tests are based on the assumption of normality.

----

* Corresponding author.

There are many areas of application of Gumbel distribution such as environmental sciences, system reliability and hydrology. In hydrology, for example, the Gumbel distribution may be used to represent the distribution of the minimum level of a river in a particular year based on minimum values for the past few years. It is useful for predicting the occurrence of that an extreme earthquake, flood or other natural disaster. The potential applicability of the Gumbel distribution to represent the distribution of minima relates to extreme value theory which indicates that it is likely to be useful if the distribution of the underlying sample data is of the normal or exponential type. Logistic and Gumbel distributions are quite similar since the shapes of these two density functions are similar for certain ranges of the parameters. The problem of estimation of the unknown parameters of the distribution function is considered by many authors. In this section, several articles particularly regarding estimation of the logistic and the Gumbel parameters are reported. Reference [11] studied the maximum likelihood (ML) method for estimation of the parameters of a logistic distribution from censored samples. Reference [15] proposed linear unbiased estimators which maintain high efficiency relative to the best linear unbiased estimators for the parameters of the logistic distribution. Reference [20] discussed the problem of estimating the mean and standard deviation of a logistic population based on multiply Type-II censored samples. Also, they studied the best linear unbiased estimation and the ML estimation methods. They showed that these estimators are also quite efficient, and derived the asymptotic variances and covariance of the estimators. Reference [30] studied the logistic parameters estimation and suggested many estimators for the location and scale parameters. The problem of estimation of the unknown parameters of the Gumbel distribution is considered by many authors under simple random sampling. Reference [14] considered the estimates of the parameters of the Gumbel distribution by the methods of probability weighted moments, moments, and maximum likelihood. They used both independent and serially correlated Gumbel numbers to derive the results from Monte Carlo experiments. They found the method of probability weighted moments estimator is more efficient than the estimators. Reference [17] derived the MLE (maximum likelihood estimator)of Gumbel distribution parameters in case of censored samples and he gave expressions for their large-sample standard errors. Reference [16] given some modifications of the MLE the Gumbel distribution parameters to reduce the bias of the estimators. Reference [12] estimated the parameters of the Gumbel distribution by moments, MLE, maximum entropy and probability weighted moments. He derived the asymptotic variance-covariance matrix of the MLEs and used simulation to compare between the various estimators. He found that the MLE is best in terms of the root MSE (mean square error). Reference [9] discussed the MLE and Cramer-Rao (CR) bounds for the location and scale parameters of the Gumbel distribution. Reference[18] found the Bayesian estimation for the two parameters of the Gumbel distribution based on record values.

Many authors studied the discrimination between two skewed distributions. References [24,22,23,24] introduced and studied quite extensively the generalized exponential distribution in a series of papers. The readers may refer to References [27, 7 , 29, 5, 13] for more studies in discrimination between distributions. Reference [1] discussed the use of the coefficient of skewness as a goodness of fit test to distinguish between the gamma and lognormal distributions. Reference [8] considered the discrimination between the logistic and the normal distributions based on likelihood ratio when the MLE's. Reference [2] introduced Log-normal and log-logistic distributions and found, for certain ranges of the parameters, the shape of the probability density functions or the hazard functions can be very similar in nature.

In the present paper, the problem of discriminating between the logistic and the Gumbel distributions is considered. The ratio of ML is used for discriminating between the two distributions. The methods of estimation that are considered are ML, MOM and OS. Monte Carlo simulation experiments are conducted for various combinations of sample sizes, and the performance of the ratio of maximized likelihood procedures is investigated in terms of the probability of correct selection (PCS) which is the ratio of the number of simulation experiments in which the procedure selects the true distribution relative to the total number of simulation runs.

2. Material and methods

a) Ratio of the maximized likelihoods procedures

The ratio of likelihoods is determined for the logistic and the Gumbel distributions. The two tests of hypotheses are given by

$$H_o: \text{Logistic vs } H_1: \text{Gumbel}, \tag{1}$$

and

$$H_o: \text{Gumbel vs } H_1: \text{Logistic}. \tag{2}$$

Let $X_1, X_2, ..., X_n$ be a random sample of size *n* from any one of the two distribution functions. The probability density function (pdf) of a logistic random variable, denoted by $L(\alpha, \beta)$, is given by

$$f_L(x;\alpha,\beta) = \frac{1}{\beta}\exp\left(-\frac{x-\alpha}{\beta}\right)\left(1+\exp\left(-\frac{x-\alpha}{\beta}\right)\right)^{-2}, \quad x \in \Re, \tag{3}$$

where $\alpha \in \Re$ and $\beta > 0$ are the location and scale parameters respectively. The pdf of a Gumbel random variable, denoted by $G(\eta, \theta)$ is given by

$$f_G(x;\alpha,\beta) = \frac{1}{\theta}\exp\left(-\frac{x-\eta}{\theta} - \exp\left(-\frac{x-\eta}{\theta}\right)\right), \tag{4}$$

where $\eta \in \Re$ and $\theta > 0$ are the location and scale parameters respectively.

Both logistic and Gumbel distributions are assumed to be effective in analyzing the same set of data since the shapes of these two density functions are quite close. It is clear that the shapes of the pdf and cumulative distribution function (cdf) of these two distributions are similar for certain range of the parameters. For example, as shown in Figures 1 and 2 the pdf and cdf for $G(0,1)$ and $L(0,0.68)$, it is found that these distributions are very close.

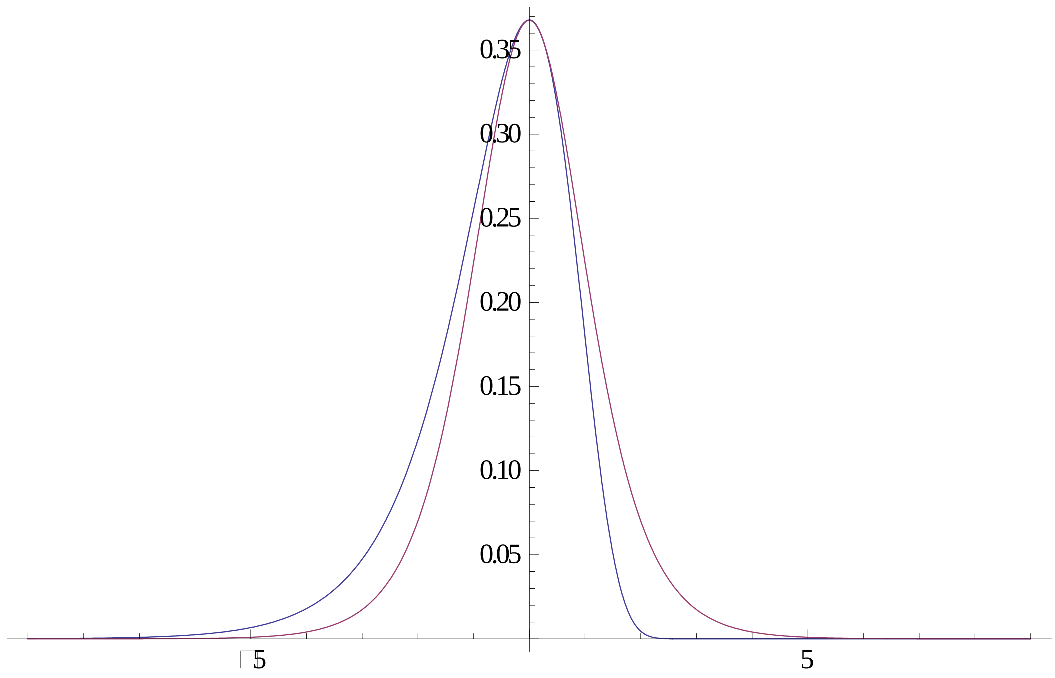

Figure 1: pdf for a random variable X which follows either G(0,1) and L(0,0.68)

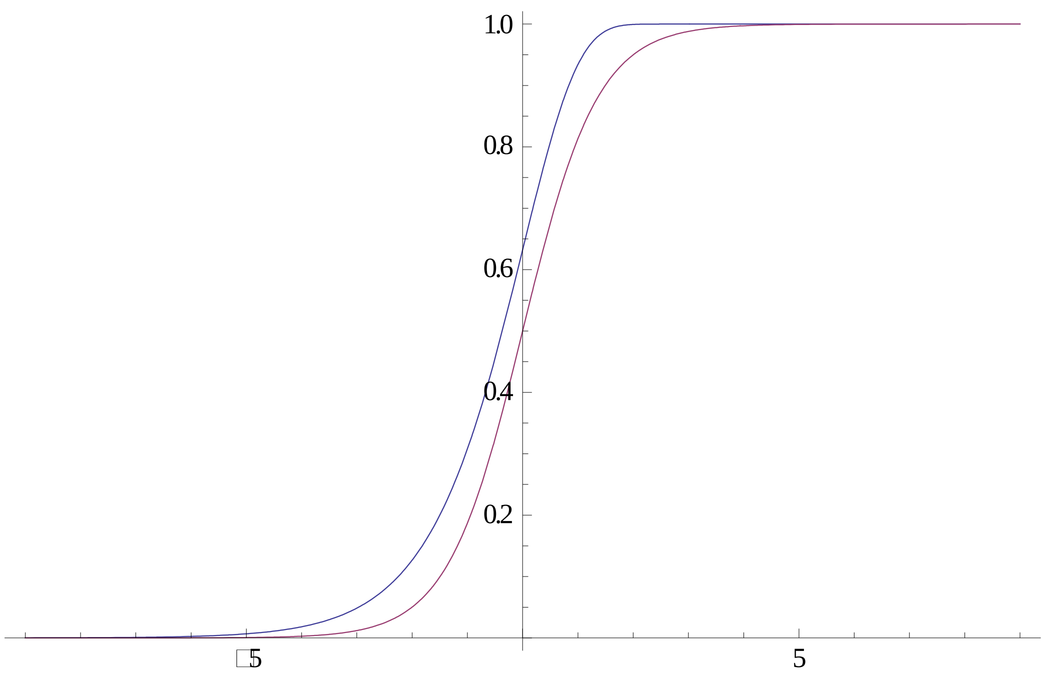

Figure 2: cdf for a random variable X which follows either G(0,1) and L(0,0.68)

Assuming that the data come from $L(\alpha, \beta)$ or $G(\eta, \theta)$, the likelihood functions are given by

$$l_L(\alpha,\beta) = \prod_{i=1}^{n} f_L(X_i;\alpha,\beta), \tag{5}$$

and

$$l_G(\eta,\theta) = \prod_{i=1}^{n} f_G(X_i;\eta,\theta), \tag{6}$$

respectively. The logarithm of the likelihood functions in (5) and (6) for $L(\alpha,\beta)$ or $G(\eta,\theta)$ are given by

$$L_L(\alpha,\beta) = -n\log(\beta) - \sum_{i=1}^{n}\left(\frac{X_i-\alpha}{\beta}\right) - 2\sum_{i=1}^{n}\log\left(1+\exp\left(-\frac{X_i-\alpha}{\beta}\right)\right), \tag{7}$$

and

$$L_G(\eta,\theta) = -n\log(\theta) - \sum_{i=1}^{n}\left(\frac{x_i-\eta}{\theta}\right) - \sum_{i=1}^{n}\exp\left(-\frac{x_i-\eta}{\theta}\right). \tag{8}$$

respectively.

The test statistic that is applied for discriminating the two distributions is called the ratio of maximized likelihoods, as based on the work of Reference [27], for example. An introduction to this approach with several examples was given in References [3,4], where the statistic considered was the logarithm of the ratio of the maximized likelihoods of two separate families of distributions.

If the method of parameter estimation used is MLE, then the test statistic is given by

$$T_1 = \log\left(\frac{l_L(\hat{\alpha}_{mle},\hat{\beta}_{mle})}{l_G(\hat{\eta}_{mle},\hat{\theta}_{mle})}\right), \tag{9}$$

where $(\hat{\alpha}_{mle},\hat{\beta}_{mle})$ and $(\hat{\eta}_{mle},\hat{\theta}_{mle})$ are the estimators found based on MLE for the logistic and Gumbel distributions respectively. Accordingly, the other test statistics, denoted as T_2 and T_3 can be determined by substitution of the estimators found by MOME and OSE in equation (9). For the logistic distribution, the estimators are denoted by $(\hat{\alpha}_{mle},\hat{\beta}_{mle})$, $(\hat{\alpha}_{moe},\hat{\beta}_{moe})$ and $(\hat{\alpha}_{ose},\hat{\beta}_{ose})$ respectively. For the Gumbel parameters, the respective estimators are denoted as $(\hat{\eta}_{mle},\hat{\theta}_{mle})$, $(\hat{\eta}_{moe},\hat{\theta}_{moe})$ and $(\hat{\eta}_{ose},\hat{\theta}_{ose})$.

b)Methods of estimation

1. Maximum likelihood method

By taking the partial derivatives for the log-likelihood function of the logistic distribution with respect to α and β and equating the resulting quantities to zero and with some algebraic manipulation, we obtain the following equations:

$$\frac{\partial L(\alpha,\beta)}{\partial \alpha} = \frac{n}{2} - \sum_{i=1}^{n} \frac{\exp\left(-\frac{X_i-\alpha}{\beta}\right)}{\left(1+\exp\left(-\frac{X_i-\alpha}{\beta}\right)\right)} = 0,$$

and

$$\frac{\partial L(\alpha,\beta)}{\partial \beta} = \frac{n}{2} - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_i-\alpha}{\beta}\right) + \sum_{i=1}^{n} \frac{\left(\frac{X_i-\alpha}{\beta}\right)\exp\left(-\frac{X_i-\alpha}{\beta}\right)}{\left(1+\exp\left(-\frac{X_i-\alpha}{\beta}\right)\right)} = 0. \tag{10}$$

There is no explicit solution for these simultaneous equations. So, for logistic distribution no closed form can be found for the MLE's. For the Gumbel distribution, MLE is given by

$$\hat{\eta}_{mle} = \bar{x} - \sum_{i=1}^{n} x_i w_i \quad \text{and} \quad \hat{\theta}_{mle} = -\hat{\eta}_{mle} \log\left(\bar{z}\right). \tag{11}$$

where $z_i = \exp\left(-\frac{x_i}{\hat{\beta}_{mle,S}}\right)$, $\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$ and $w_i = \frac{z_i}{n\bar{z}}$.

2. Method of moment

The moment estimators for logistic and normal distributions are given by

$$\hat{\alpha}_{moe} = \bar{X} \quad \text{and} \quad \hat{\beta}_{moe} = \frac{\sqrt{3}}{\pi} S, \tag{12}$$

and

$$\hat{\theta}_{\text{moe}} = \frac{\sqrt{6}}{\pi} s \quad \text{and} \quad \hat{\eta}_{moe} = \bar{x} - \gamma\ \hat{\theta}_{\text{moe}} \tag{13}$$

where

s, $\bar{x}$ are the sample standard deviation and mean, respectively, and $\gamma = 0.57721566$ is Euler's constant.

3. Order statistic method likelihood m

The p^{th} quantile for the logistic distribution is given by

$$Q(p;\alpha,\beta)=F^{-1}(p)=\alpha+\beta\log\left(\frac{p}{1-p}\right). \qquad (14)$$

Based on (3.14), if we substitute $p=0.25$ and $p=0.75$, corresponding to the first and third quartiles, we will have

$F^{-1}(0.25)=\alpha-\beta\log(3)$ and $F^{-1}(0.75)=\alpha+\beta\log(3)$

respectively. Accordingly, the two estimators of α and β, denoted as $\hat{\alpha}_{ose}$ and $\hat{\beta}_{ose}$, are

$$\hat{\alpha}_{ose}=\frac{1}{2}\left(\hat{F}^{-1}(0.75)+\hat{F}^{-1}(0.25)\right), \qquad (15)$$

and

$$\hat{\beta}_{ose}=\frac{1}{2\log(3)}\left(\hat{F}^{-1}(0.75)-\hat{F}^{-1}(0.25)\right). \qquad (16)$$

The p^{th} quantile for the Gumbel distribution is

$$Q(p;\eta,\theta)=F^{-1}(p)=\eta-\theta\log\left(-\log p\right), \qquad 0<p<1 \qquad (17)$$

It is known that the lower quartile, median and upper quartile, denoted by $F^{-1}(.25)$, $F^{-1}(.5)$ and $F^{-1}(.75)$ respectively, and the distributional limits as given by $F^{-1}(0)$ and $F^{-1}(1)$ provide the information on the feel of spread of the distribution over the axis. Since the interquartile range (IQR) given by

$IQR=F^{-1}(0.75)-F^{-1}(0.25),$

is independent of the location parameter , the scale parameter can be estimated using IQR. Using (17), if we substitute $p=0.25$ and $p=0.75$, corresponding to the first and third quartiles, or namely lower and upper quartiles, we will have

$F^{-1}(0.25)=\eta-\theta\log\left(-\log 0.25\right)=\eta-0.3266\theta,$

and

$F^{-1}(0.75) = \eta - \theta \log\left(-\log 0.75\right) = \eta + 1.2459\ \theta,$ respectively.

By taking $p = 0.25$ and $0.75,$ the estimators of η, θ are

$$\hat{\eta}_{ose} = \frac{1}{1.5725}\left(0.3266F^{-1}(0.75) + 1.2459F^{-1}(0.25)\right) \approx \frac{1}{5}F^{-1}(0.75) - \frac{1}{5}F^{-1}(0.25) \quad (18)$$

and

$$\hat{\theta}_{ose} = \frac{1}{1.5725}\left(F^{-1}(0.75) - F^{-1}(0.25)\right). \quad (19)$$

3. Results

In this simulation, without any loss of generality, in assessing the relative performance of the selection procedure, random samples are generated from $L(1,1)$ and $G(1,1).$ To calculate the PCS when the true sampled distribution is logistic, the following algorithm is introduced:

a) Let $\{X_i,\ i = 1,...,n\}$ be a sample of size *n* from $L(1,\ 1).$

b) The parameters (α, β) and (η, θ) are estimated by MLE, MOME or OSE.

c) The test statistics T_1 are calculated.

d) Check whether $T_1 > 0.$

e) The steps (a)-(d) are repeated 10, 000 times to get $T_{1t},\ t = 1,...,10,000.$

f) PCS is calculated by

$$T_1^* \approx \frac{1}{10,000}\sum_{t=1}^{10,000} I\,(T_{1\,t} > 0),$$

where $I\,(.)$ stands for indicator function. The same procedure can be repeated for the other test statistics.

In order to assess the performances of these procedures, estimates of the PCS can be based on the simulation results for several different cases using a Monte Carlo simulation of 10,000 runs according to the algorithm of Section 4 and also based on the asymptotic approximations. For different values of the sample size, i.e.

$n = 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500$. The results of PCS under both null hypotheses are calculated. First, the case when the true sampled distribution is logistic is considered and the results are summarized in Table 1. Second, the case when the true sampled distribution is Gumbel is studied and the results found are presented in Table 2.

Based on the simulation, with 10,000 iterations, as reported in Tables 1 and 2, the following remarks can be made:

- It is observed that the two densities and distributions are very close.
- When testing for logistic against Gumbel for the data generated, it is found that PCS is high and it increases as n increases when either MLE, MOME or OSE is used. PCS is equal one for a case OSE.
- When the true sampled distribution is Gumbel for the data generated, it is found that PCS is high and it increases as n increases when either MLE or MOME is used. PCS is equal to zero for a case OSE.

Table 1: The values of PCS when the true sampled distribution is logistic based on estimators found using MLE, MOME and OSE

n	H_o: Logitic(1,1)		
	MLE	$MOME$	OSE
10	0.552	0.610	1
20	0.694	0.679	1
30	0.779	0.709	1
40	0.837	0.738	1
50	0.878	0.759	1
60	0.912	0.779	1
70	0.926	0.790	1
80	0.942	0.808	1
90	0.954	0.825	1
100	0.969	0.836	1
150	0.989	0.873	1
200	0.997	0.903	1
250	1	0.931	1
300	1	0.961	1
400	1	0.986	1
500	1	1	1

Table 2: The values of PCS when the true sampled distribution is Gumbel based on estimators found using MLE, MOME and OSE

n	H_o: Gumbl(1,1)		
	MLE	$MOME$	OSE
10	0.772	0.336	0
20	0.843	0.457	0
30	0.885	0.551	0
40	0.910	0.622	0
50	0.932	0.675	0
60	0.942	0.721	0
70	0.958	0.762	0
80	0.968	0.797	0
90	0.978	0.825	0
100	0.981	0.849	0
150	0.994	0.921	0
200	0.998	0.959	0
250	1	0.979	0
300	1	0.987	0
400	1	0.998	0
500	1	1	0

4. Discussion

Determining a correct distribution for a given set of data is an important issue. When describing a data set, it is often difficult to choose between two distributions which have the same characteristics. In this study, the problem of discriminating between the two distributions, namely, the logistic and Gumbel distributions is considered. The statistics based on the logarithm of the ratio of the maximized likelihoods is considered. The performance of the ratio of maximized likelihood procedures is investigated in terms of the PCS which is the ratio of number of simulation experiments in which the procedure selects the true distribution to the total number of simulation runs.

The PCS's that are obtained are compared based on different estimators, namely, MLE, MOME and OSE, using Monte Carlo simulations. Based on the Monte Carlo simulations when the null hypothesis is logistic, with either MLE, MOME or OSE are being used as the method of estimation, the PCS is found to be high and gets larger as the sample size increases. If the null hypothesis is Gumbel and when these estimators are compared in terms of their contribution to the PCS, it appears that MLE slightly outperforms all the other estimators.

References

[1] A. Alzaid and K.S. Sultan." Discriminating between gamma and lognormal distributions with applications". Journal of King Saud University (Science), vol. 21, pp. 99-108, 2009.

[2] A. K. Dey and D. Kundu. "Discriminating between the log-normal and log-logistic distributions ". Communications in Statistics- Theory and Methods, vol. 39, pp. 280-292, 2010.

[3] D.R. Cox. "Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability". University of California Press, Berkeley, pp. 105–123, 1961.

[4] D. R. Cox." Further results on tests of separate families of hypotheses". J. of Royal Statistical Society, Ser. B, vol. 24, pp. 406–424,1962.

[5] D. Kundu and M. Z. Raqab. "Discriminating between the generalized Rayleigh and Log-normal distribution". Statistics, Vol. 41(6), pp. 505-515, 2007.

[6] D. Kundu and A. Manglick. Discriminating between the Weibull and Log-normal distributions. Navel Research Logistics, Vol. 51 (6), pp.893-905, 2004.

[7] D. Kundu, R. D. Gupta and A.Manglick." Discriminating between Log-normal and generalized exponential distributions. Journal of Statistical planning and inference 127: 213-227, 2005.

[8] F. Aucoin and F. Ashkar." Discriminating between the logistic and the normal distributions based on likelihood ratio". http://interstat.statjournals.net/YEAR/2010/articles/1008003.pdf, 2010.

[9] G. Corsini, F. Gini, M. V. Greco and L. Verrazzani. "Cramer-Rao bounds and estimation of the parameters of the Gumbel distribution". Aerospace and Electronic Syst., Vol. 31 (3), 1202-1204,1995.

[10] G.S. Mudholkar & A.D. Hutson. "LQ-moments: Analogs of L-moments". Journal of Statistical Planning and Inference, vol. 71, pp. 191-208,1998.

[11] H.L. Harter & A.H. Moore. "Maximum-likelihood estimation, from censored samples, of the parameters of a logistic distribution". J. Amer. Statist. Assoc., vol. 62, pp. 675–684, 1967.

[12] H. N Phien." A review of methods of parameter estimation for the extreme value type-1 distribution". Journal of Hydrology, vol. 90 (3–4), pp. 251–268,1987.

[13] J. S. Kim, and B. J. Yum. "Selection between Weibull and lognormal distributions: A comparative simulation study". Computational Statistics and Data Analysis, vol. 53, pp. 477-485, 2008.

[14] J. Maciunas, N. C. Matalas and J. R. Wallis. Probability weighted moments compared with some traditional techniques in estimating Gumbel Parameters and quantiles. Water Resources Research, Vol. 15(5), pp. 1055–1064, 1979.

[15] M. Engelhardt." Simple Linear Estimation of the Parameters of the Logistic Distribution from a Complete or Censored Sample". Journal of the American Statistical Association, Vol. 70(352), pp. 899- 902, 1975.

[16] M. Fiorentino and S. Gabriele." A correction for the bias of maximum-likelihood estimators of Gumbel parameters. Journal of Hydrology. Vol. 73 (1-2),pp. 39–49, 1984.

[17] M. N. Leese. "Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series". Water Resources Research. Vol. 9 (6), pp. 1534–1542,1973.

[18] M. A. M. Mousa, Z. F. Jaheen and A.A Ahmad." Bayesian Estimation, Prediction and Characterization for the Gumbel Model Based on Records". A Journal of Theoretical and Applied Statistics, vol. 36(1),

pp. 65-74, 2002.

[19] N. Balakrishnan. Handbook of the logistic distribution. New York, Marcel Dekker, Inc, 1992.

[20] N. Balakrishnan, S.S. Gupta & S. Panchapakesan. 1995. "Estimation of the Mean and Standard Deviation of the Logistic Distribution Based on Multiply Type-II Censored Samples". Statistics, Vol. 27(1), pp. 127 – 142, 1995.

[21] R. D. Gupta and D. Kundu." Generalized Exponential Distributions". Australian and New Zealand Journal of Statistics, vol. 41, pp. 173-188, 1999.

[22] R. D. Gupta and D. Kundu. "Exponentiated Exponential Distribution; An alternative to gamma or Weibull distribution". Biometrical Journal,vol. 43, pp. 117-130,2001b.

[23] R. D. Gupta and D. Kundu. "Generalized Exponential Distributions; Different Methods of Estimation". Journal of Statistical Computation and Simulation, vol. 69, pp.315-338, 2001a.

[24] R. D. Gupta and D. Kundu. "Generalized exponential distributions". Statistical Inferences. Journal of Statistical Theory and Applications, vol. 1, pp. 101-118,2002.

[25] R. D. Gupta and D. Kundu." Closeness of gamma and generalized exponential distributions". Commun. Stat.-Theory Methods, vol. 32, pp. 705–721,2003a.

[26] R. D. Gupta and D. Kundu." Discriminating between Weibull and generalized exponential distributions". Comput. Stat. Data Anal., vol. 43, pp. 179–196, 2003b.

[27] R. D. Gupta and D. Kundu. "Discriminating between gamma and generalized exponential distributions". Journal of Statistical Computations and Simulations, vol. 74, pp. 107–121,2004.

[28] R.L. Prentice." Discrimination among Some Parametric Models". Biometrika, vol. 62 (3), pp. 607-614,1975.

[29] W. G. Strupczewski, H. T. Mitosek, K. Kochanek, V. P. Singh and S. Weglarczyk. "Probability of correct selection from lognormal and convective diffusion models based on the likelihood ratio". Stochastic Environmental Research and Risk Assessment,Vol. 20(3), pp. 152-163,2006.

[30] W. Abu-Dayyeh, S. Al-Subh and H. Muttlak. "Logistic parameters estimation using simple random sampling and ranked set sampling data". Applied Mathematics and Computation, vol. 150(2), pp. 543-554, 2004.