By submitting this paper to W-OSDCE I confirm that (i) I and any other co-author(s) are responsible for its content and its originality; (ii) any possible co-authors agreed to its submission to W-OSDCE.

INFERENCE ON THE PARAMETERS OF THE GENERALIZED LOGISTIC DISTRIBUTION BASED ON LEFT CENSORED DATA

BABAYI, S.[1]*, GHOLAMI, GH.[1]

[1] *Department of Mathematics, Faculty of Science, Urmia University, Iran; babayi@urmia.ac.ir; gh.gholami@urmia.ac.ir*

Abstract. The generalized Logistic distribution is an important lifetime distribution in survival analysis. This paper investigates the estimation of the parameters of the generalized Logistic distribution with a left-censored data. The maximum likelihood estimator (MLE) procedure of the parameters is considered and the Fisher information matrix of the unknown parameters is used to construct asymptotic confidence intervals. Bayes estimator of the parameters and the corresponding credible intervals are obtained by using the Gibbs sampling technique. A real data set is provided to illustrate the two proposed methods.

1. Introduction

The Logistic distribution is one of the most important statistical distributions because of its simplicity and also its historical importance as a growth curve (Erkelens, [6]). Some applications of Logistic distribution have been mentioned by Johnson and Kotz [7]. Babayi et al. [1] used generalized Logistic (GL) distribution for analysing stress-strength problem.

The random variable X has the GL distribution if it has the following cumulative distribution function (cdf)

$$F(x;\mu,\sigma,\alpha) = \frac{1}{(1+e^{-\frac{x-\mu}{\sigma}})^{\alpha}}, \quad -\infty < x < +\infty \tag{1.1}$$

2010 *Mathematics Subject Classification.* Primary 62N01; Secondary 62N02, 62N05.

Key words and phrases. Generalized Logistic distribution; Maximum likelihood estimator; Bayesian estimation; Left censoring.

* Speaker.

where $\mu \in \mathbb{R}$ and $\sigma, \alpha \in (0, +\infty)$. The probability density function (pdf) corresponding to the cdf (1.1) is

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha e^{-\frac{x-\mu}{\sigma}}}{\sigma(1 + e^{-\frac{x-\mu}{\sigma}})^{\alpha+1}}, \quad -\infty < x < +\infty.$$

Here μ, σ and α are the location, scale and shape parameters, respectively. The GL distribution with the shape parameter α and the scale parameter σ will be denoted by $GL(\alpha, \sigma)$. In the particular case of $\alpha = 1$, F corresponds to the usual Logistic distribution. Zelterman [11] showed that the maximum likelihood estimates do not exist for (μ, σ, α). Therefore for convenience, without loss of generality it is supposed $\mu = 0$.

There is a large number of applications and use of left censoring or left censored data in survival analysis and reliability theory. For example, in a medical study investigating Patterns of Health Insurance Coverage among Rural and Urban Children[**?**] faces this problem due to the existence of a higher proportion of children living in small villages whose spells were left censored in the sample (i.e., those children who entered the sample uninsured), and who remained so throughout the sample. Mitra and Kundu [10] discussed the maximum likelihood estimator for parameters of generalized exponential distribution in the presence of left censoring.

The rest of this paper is organized as follows. In Section 2, we derive the MLEs of the unknown parameters of the GL distribution for left censored data. In this case, the MLEs cannot be obtained in explicit form and the MLE of the scale parameter can be achieved by solving a non-linear equation using an iterative procedure. Once the MLE of the scale parameter is extracted, the MLE of the shape parameter can be obtained in explicit form. We have also obtained the explicit expression of the Fisher information matrix and it has been used to construct the asymptotic confidence intervals of the unknown parameters. In Section 3, we propose the Bayes estimator of parameters and the corresponding credible interval applying the Gibbs sampling technique. In Section 4, analysis of a real data set is given for illustrative purposes.

2. Maximum likelihood estimation

In this section, the MLEs of the parameters are extracted in the presence of left censored observations.

Let $X_{(r+1)}, \ldots, X_{(n)}$ be the last $n - r$ order statistics from a random sample of size n from $GL(\alpha, \sigma)$ distribution. Therefore, the joint probability density function of $X_{(r+1)}, \ldots, X_{(n)}$ becomes

$$\begin{aligned} f(x_{(r+1)}, \ldots, x_{(n)}; \alpha, \sigma) =& \frac{n!}{r!}(F(x_{(r+1)}))^r f(x_{(r+1)}) \ldots f(x_{(n)}) \\ =& \frac{n!}{r!}(1 + e^{-\frac{x_{(r+1)}}{\sigma}})^{-r\alpha} \prod_{i=r+1}^{n} \frac{\alpha e^{-\frac{x_{(i)}}{\sigma}}}{\sigma(1 + e^{-\frac{x_{(i)}}{\sigma}})^{\alpha+1}} \end{aligned}$$

Then, the log likelihood function is

$$\begin{aligned} L(\alpha, \sigma) =& \ln n! - \ln r! - r\alpha \ln(1 + e^{-\frac{x_{(r+1)}}{\sigma}}) + (n - r) \ln \alpha - (n - r) \ln \sigma \\ & - \frac{1}{\sigma} \sum_{i=r+1}^{n} x_{(i)} - (\alpha + 1) \sum_{i=r+1}^{n} \ln(1 + e^{-\frac{x_{(i)}}{\sigma}}) \end{aligned}$$

Hence, the likelihood equations are

$$\frac{\partial L}{\partial \alpha} = -r\ln(1+e^{-\frac{x_{(r+1)}}{\sigma}}) + \frac{n-r}{\alpha} - \sum_{i=r+1}^{n} \ln(1+e^{-\frac{x_{(i)}}{\sigma}}) = 0 \tag{2.1}$$

$$\frac{\partial L}{\partial \sigma} = \frac{-\alpha r x_{(r+1)} e^{-\frac{x_{(r+1)}}{\sigma}}}{\sigma^2(1+e^{-\frac{x_{(r+1)}}{\sigma}})} - \frac{n-r}{\sigma} + \frac{1}{\sigma^2}\sum_{i=r+1}^{n} x_{(i)} - \frac{(\alpha+1)}{\sigma^2}\sum_{i=r+1}^{n} \frac{x_{(i)}e^{-\frac{x_{(i)}}{\sigma}}}{1+e^{-\frac{x_{(i)}}{\sigma}}} = 0 \tag{2.2}$$

From (2.1) and (2.2), we get

$$\hat{\alpha}_{(\hat{\sigma})} = \frac{n-r}{\sum_{i=r+1}^{n} \ln(1+e^{-\frac{X_{(i)}}{\hat{\sigma}}}) + r\ln(1+e^{-\frac{X_{(r+1)}}{\hat{\sigma}}})} \tag{2.3}$$

and $\hat{\sigma}$ can be given as the solution of the following non-linear equation

$$h(\sigma) = \sigma, \tag{2.4}$$

where

$$h(\sigma) = \frac{-\frac{\alpha_{(\sigma)} r x_{(r+1)} e^{-\frac{x_{(r+1)}}{\sigma}}}{1+e^{-\frac{x_{(r+1)}}{\sigma}}} + \sum_{i=r+1}^{n} x_{(i)} - (\alpha_{(\sigma)}+1)\sum_{i=r+1}^{n} \frac{x_{(i)}e^{-\frac{x_{(i)}}{\sigma}}}{1+e^{-\frac{x_{(i)}}{\sigma}}}}{n-r} \tag{2.5}$$

Since $\hat{\sigma}$ is a fixed point solution of the non-linear equation (2.5), therefore it can be achieved by using an iterative scheme as follows:

$$h(\sigma_{(j)}) = \sigma_{(j+1)}, \tag{2.6}$$

where $\sigma_{(j)}$ is the jth iterate of $\hat{\sigma}$. The iteration procedure should be stopped when $|\sigma_{(j)} - \sigma_{(j+1)}|$ is sufficiently small. First, $\hat{\sigma}$ is obtained, then $\hat{\alpha}$ can be resulted from (2.3).

2.1. **Fisher information matrix.** In this section, we first obtain the Fisher information matrix of the unknown parameters of GL distribution when the data are left censored, which can be used to construct asymptotic confidence intervals. We denote the Fisher information matrix of $\theta = (\alpha, \sigma)$ as $I(\theta) = [I_{ij}(\theta)], i, j = 1, 2$. Therefore,

$$I(\theta) = -\begin{bmatrix} E(\frac{\partial^2 L}{\partial \alpha^2}) & E(\frac{\partial^2 L}{\partial \alpha \partial \sigma}) \\ E(\frac{\partial^2 L}{\partial \sigma \partial \alpha}) & E(\frac{\sigma^2 L}{\partial \sigma^2}) \end{bmatrix} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$$

where

$$E(\frac{\partial^2 L}{\partial \alpha^2}) = -\frac{n-r}{\alpha^2}$$

$$E(\frac{\partial^2 L}{\partial \alpha \partial \sigma}) = E(\frac{\partial^2 L}{\partial \sigma \partial \alpha}) = -\frac{r}{\sigma}E[\frac{Z_{(r+1)}e^{-Z_{(r+1)}}}{1+e^{-Z_{(r+1)}}}] - \frac{1}{\sigma}\sum_{i=r+1}^{n} E[\frac{Z_{(i)}e^{-Z_{(i)}}}{1+e^{-Z_{(i)}}}]$$

$$E(\frac{\partial^2 L}{\partial \sigma^2}) = \frac{2r\alpha}{\sigma^2}E[\frac{Z_{(r+1)}e^{-Z_{(r+1)}}}{1+e^{-Z_{(r+1)}}}] - \frac{r\alpha}{\sigma^2}E[\frac{Z^2_{(r+1)}e^{-Z_{(r+1)}}}{(1+e^{-Z_{(r+1)}})^2}] + \frac{n-r}{\sigma^2} - \frac{2}{\sigma^2}\sum_{i=r+1}^{n} E[Z_{(i)}] + \frac{2(\alpha+1)}{\sigma^2}\sum_{i=r+1}^{n} E[\frac{Z_{(i)}e^{-Z_{(i)}}}{1+e^{-Z_{(i)}}}] - \frac{\alpha+1}{\sigma^2}\sum_{i=r+1}^{n} E[\frac{Z^2_{(i)}e^{-Z_{(i)}}}{1+e^{-Z_{(i)}}}]$$

where $Z_{(i)} = \frac{X_{(i)}}{\sigma}$.

Using the results of Balakrishnan and Leung [3] the pdf of $Z_{(i)}$ $(r+1 \leq i \leq n)$ is obtained as

$$f_{Z_{(i)}}(z) = \frac{n!}{(i-1)!\,(n-i)!} \sum_{j=0}^{n-i} (-1)^j \left(\begin{array}{c} n-i \\ j \end{array} \right) \frac{\alpha e^{-z}}{(1+e^{-z})^{\alpha(i+j)+1}} \quad -\infty < z < +\infty \tag{2.7}$$

From (2.7) and some algebraic operations we get

$$E[Z_{(i)}] = \frac{n!}{(i-1)!\,(n-i)!} \sum_{j=0}^{n-i} (-1)^j \left(\begin{array}{c} n-i \\ j \end{array} \right) \frac{\psi(\alpha b_{ij}) - \psi(1)}{b_{ij}}$$

$$E[\frac{Z_{(i)} e^{-Z_{(i)}}}{1+e^{-Z_{(i)}}}] = \frac{n!}{(i-1)!\,(n-i)!} \sum_{j=0}^{n-i} (-1)^j \left(\begin{array}{c} n-i \\ j \end{array} \right) \frac{\psi(\alpha b_{ij}) - \psi(2)}{b_{ij}(\alpha b_{ij}+1)}$$

$$E[\frac{Z_{(i)}^2 e^{-Z_{(i)}}}{(1+e^{-Z_{(i)}})^2}] = \frac{n!}{(i-1)!\,(n-i)!} \sum_{j=0}^{n-i} (-1)^j \left(\begin{array}{c} n-i \\ j \end{array} \right) C(\alpha, b_{ij})$$

where $b_{ij} = i+j$, $C(\alpha, b_{ij}) = \frac{\alpha\{\psi'(\alpha b_{ij}+1)+\psi'(2)+[\psi(\alpha b_{ij}+1)-\psi(2)]^2\}}{(\alpha b_{ij}+1)(\alpha b_{ij}+2)}$, $\psi(x) = \frac{d \ln \Gamma(x)}{dx}$, with $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$.

3. Bayes estimation

In this section, we attempt to find the Bayes estimator of the parameters under the assumption that the shape parameters α and the scale parameter σ are random variables. It is assumed that α and σ have independent gamma priors with the parameters $\alpha \sim Gamma(a_1, b_1)$ and $\sigma \sim Gamma(a_2, b_2)$. Therefore,

$$\pi(\alpha) = \frac{b_1^{a_1}}{\Gamma(a_1)} \alpha^{a_1-1} e^{-b_1\alpha}; \qquad \alpha > 0 \tag{3.1}$$

and

$$\pi(\sigma) = \frac{b_2^{a_2}}{\Gamma(a_2)} \sigma^{a_2-1} e^{-b_2\sigma}; \qquad \sigma > 0. \tag{3.2}$$

Here $a_1, b_1, a_2, b_2 > 0$.
Based on the above assumptions, the likelihood function of the observed data is

$$L(data\,|\alpha, \sigma\,) = \frac{n!}{r!} (1+e^{-\frac{x_{(r+1)}}{\sigma}})^{-r\alpha} \prod_{i=r+1}^{n} \frac{\alpha e^{-\frac{x_{(i)}}{\sigma}}}{\sigma(1+e^{-\frac{x_{(i)}}{\sigma}})^{\alpha+1}}$$

The joint density of the data, α and σ can be achieved as

$$L(data, \alpha, \sigma) = L(data\,|\alpha, \sigma\,) \times \pi(\alpha) \times \pi(\sigma)$$

Therefore, the joint posterior density of α and σ given the data is

$$L(\alpha, \sigma\,|data\,) = \frac{L(data, \alpha, \sigma)}{\int_0^\infty \int_0^\infty L(data, \alpha, \sigma) d\alpha d\sigma} \tag{3.3}$$

Since (3.3) cannot be obtained analytically, we adopt the Gibbs sampling technique to compute the Bayes estimate of α and σ the corresponding credible interval of them. The posterior pdfs of α and σ are as follows:

$$\alpha\,|\sigma, data \sim Gamma(n-r+a_1, b_1 + r\ln(1+e^{-\frac{x_{(r+1)}}{\sigma}}) + \sum_{i=r+1}^{n} \ln(1+e^{-\frac{x_{(i)}}{\sigma}}))$$

and

$$\pi(\sigma|\alpha, data) \propto \quad \sigma^{a_2-(n-r)-1} \exp\{-b_2\sigma - \frac{1}{\sigma}\sum_{i=r+1}^{n} x_{(i)}\}$$
$$\times \exp\{-\alpha r \ln(1+e^{-\frac{x_{(r+1)}}{\sigma}}) - (\alpha+1)\sum_{i=r+1}^{n} \ln(1+e^{-\frac{x_{(i)}}{\sigma}})\}$$

The posterior pdf of σ is not known but its plot, as shown in Figure 1, is similar to the normal distribution, so to generate random numbers from this distribution, we apply the Metropolis method with normal proposal distribution. Therefore, the algorithm of Gibbs sampling is as follows.

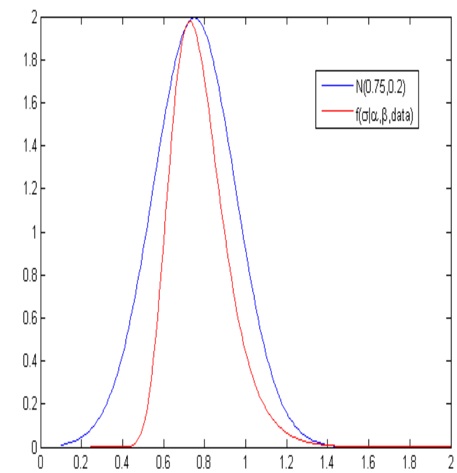

FIGURE 1. Proposal and posterior density functions of scale parameter

Step 1: Start with an initial guess $(\alpha^{(0)}, \sigma^{(0)})$.
Step 2: Set $t = 1$.
Step 3: Generate $\alpha^{(t)}$ from $Gamma(n-r+a_1, b_1 + r\ln(1+e^{-\frac{x_{(r+1)}}{\sigma^{(t-1)}}}) + \sum_{i=r+1}^{n} \ln(1+e^{-\frac{x_{(i)}}{\sigma^{(t-1)}}}))$.
Step 4: Using Metropolis-Hastings, generate $\sigma^{(t)}$ from $\pi(\sigma\,\big|\alpha^{(t-1)}, \beta^{(t-1)}, data)$ with the $N(\sigma^{(t-1)}, 1)$ as a proposal distribution.
Step 5: Set $t = t+1$.
Step 6: Repeat steps 3-5, T times.

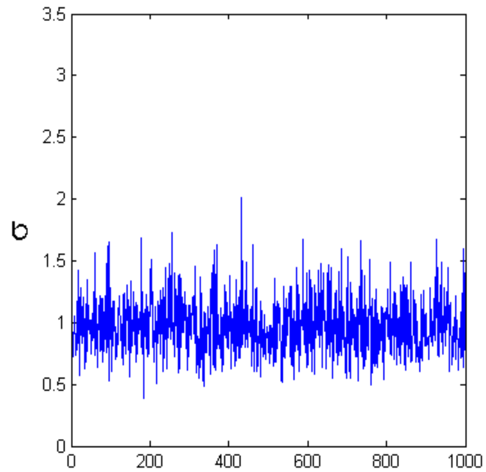

FIGURE 2. sequence of 1000 generations from posterior density functions of scale parameter

The corresponding graph for a sequence of 1000 generations from posterior density functions of scale parameter is displayed in Figure 2. Now the approximate posterior mean, and posterior variance of α become

$$\hat{E}(\alpha\,|data) = \frac{1}{T-K}\sum_{t=K+1}^{T}\alpha^{(t)}$$

and

$$\hat{V}(\alpha\,|data) = \frac{1}{T-K}\sum_{t=K+1}^{T}(\alpha^{(t)} - \hat{E}(\alpha\,|data))^2,$$

where K is the burn-in period. The burn-in is considered, here, to avoid the effect of the starting values on the generated scale parameter.

Based on T, and R values, using the method proposed by Chen and Shao [4], the approximate highest posterior density (HPD) credible interval of R can be easily constructed. Based on T, and α values, using the method proposed by Chen and Shao [4], the approximate highest posterior density (HPD) credible interval of α can be easily constructed. Let $\alpha_{(K+1)} < \alpha_{(K+2)} < \ldots < \alpha_{(T-K)}$ be the ordered $\alpha^{(t)}$, and suppose we would like to construct a $100(1-\gamma)\%$ approximate HPD credible interval of α, then consider the following:

$$\{(\alpha_{(T-K)}, \alpha_{((1-\gamma)(T-K))}), \ldots, (\alpha_{(\gamma(T-K))}, \alpha_{(T-K)})\}$$

Choose that interval which has the shortest length. The Bayes estimation and credible interval for σ are exactly the same as α.

4. Data Analysis

Here, analysis of the strength data, which was originally reported by Badar & Priest [2] is presented. Kundu & Gupta [9] observed that Weibull distribution works quite well for these strength data which are presented in Table 1. The GL distribution model for the data set, is fitted. The estimated scale and shape parameters are proposed assuming the location parameter to be known as the sample median for the data set. We also obtained Kolmogrov-Smirnov (K-S) distance between the empirical distribution function and the fitted distribution, and corresponding p value. All the results have reported in Table 2. For comparison purposes, we also compute the observed and the expected frequencies, the corresponding chi-square value based on the fitted model in Table 3.

Table 1. The real data set.

Data Set													
1.312	1.314	1.479	1.552	1.700	1.803	1.861	1.865	1.944	1.958	1.966	1.997	2.006	2.021
2.027	2.055	2.063	2.098	2.140	2.179	2.224	2.240	2.253	2.270	2.272	2.274	2.301	2.301
2.359	2.382	2.382	2.426	2.434	2.435	2.478	2.490	2.511	2.514	2.535	2.554	2.566	2.570
2.586	2.629	2.633	2.642	2.648	2.684	2.697	2.726	2.770	2.773	2.800	2.809	2.818	2.821
2.848	2.880	2.954	3.012	3.067	3.084	3.090	3.096	3.128	3.233	3.433	3.585	3.585	

It is clear that the GL distribution fits quite well to the data set. For the data set, the fitted empirical cdf plot of the GL distribution model is shown in Figure 3. As it is seen, a satisfactory fit for the GL distribution model is provided.

Table 2. Sample Median, Scale Parameter, Shape Parameter, K-S and p value of the fitted GL distribution to data set.

Sample Median	Scale Parameter	Shape Parameter	K-S	p value
2.478	0.2745	0.9489	0.0492	0.9933

Table 3. Observed Frequencies, and Expected Frequencies for modified data set when fitting the GL distribution.

Intervals	Observed Frequencies	Expected Frequencies	Chi-Square
< 1.76	5	2.3900	0.6452
1.76-2.22	15	15.2904	
2.22-2.68	27	26.9100	
2.68-3.14	18	16.0011	
> 3.14	4	5.4027	

For illustrative purposes, for left censoring, we have left out about 20% of the data set ($r = 14$). From (2.3) and (2.4), the MLEs of α and σ become 0.9162 and 0.2826, respectively. Also from Section 4, the Bayes estimations of α and σ become 0.9148 and 0.2845, respectively. To compute the Bayes estimate, as mentioned above, we have adopted the suggestion of Congdon ([5], p. 20) and Kundu and Gupta [8], that is, $a_1 = a_2 = b_1 = b_2 = 0.0001$. The 95% confidence intervals corresponding MLEs of α and σ become (0.6702,1.1623) and (0.1646,0.4006), respectively. Also, the 95% credible intervals of α and σ become (0.6840,1.1537) and (0.2273,0.3631), respectively. Comparing the average confidence lengths to the average credible lengths, we observe the average credible lengths is less than the average confidence lengths. We observe that the results are not significantly different from the corresponding results obtained from completed data.

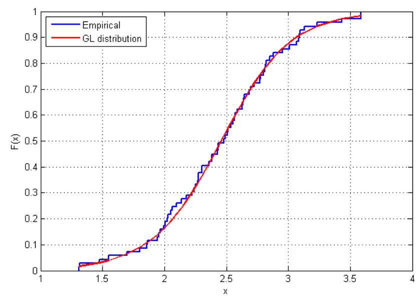

FIGURE 3. Empirical cdf plot of GL distribution for data set.

REFERENCES

1. Babayi, S., Khorram, E. and Tondro, F. (2014), Inference of $R = P(X < Y)$ for generalized Logistic distribution. *Statistics*, **48**(4), 862–871.
2. Badar, M.G and Priest, A.M. (1982), Statistical aspects of fiber and bundle strength in hybrid composites. In: Hayashi T, Kawata K, Umekava S, editors. *Progress in science and engineering of composites.* Tokyo: Japanese Society for Composite Materials, 1129–1136.
3. Balakrishnan, N. and Leung, M.Y. (1988), Order statistics from the type I generalized Logistic distribution.*Communication in Statistics: Simulation and Computation*, **17**(1), 25–50.
4. Chen, M.H. and Shao, Q.M. (1999), Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8, 69–92.
5. Congdon, p. (2001), *Bayesian Statistical Modeling*, Wiley, New York.
6. Erkelens, J. (1968), A method of calculation for the Logistic curve. *Statistics Neerlandica*, 22, 213–217.
7. Johnson, N.L. and Kotz, S. (1970), *Distribution in Statistics: Continuous Univariate Distributions*, Vol. 2, John Wiley, New York.
8. Kundu, D. and Gupta, R.D. (2005), Estimation of $P[Y < X]$ for generalized exponential distribution. *Metrika*, 61, 291–308.
9. Kundu, D. and Gupta, R.D. (2006), Estimation of $P[Y < X]$ for Weibull distributions. *IEEE Transactions on Reliability*, **55**(2), 270-280.
10. Mitra, S. and Kundu, D. (2008), Analysis of left censored data from the generalized exponential distribution. *Journal of Statistical Computation and Simulation*, **78**(7), 669–679.
11. Zelterman, D. (1987), Parameter estimation in the generalized Logistic distribution. *Computational Statistics & Data Analysis*, **5**(3), 177–184.