International Journal of Reliability and Applications
Vol. 8, No. 1, pp. 17-25, 2007

Parameters Estimators for the Generalized Exponential Distribution

A. Abuammoh
Department of Statistics and Operations Research
King Saud University, Riyadh 11451, Saudi Arabia

A. M. Sarhan * †
Department of Statistics and Operations Research
King Saud University, Riyadh 11451, Saudi Arabia

Abstract. Maximum likelihood method is utilized to estimate the two parameters of generalized exponential distribution based on grouped and censored data. This method does not give closed form for the estimates, thus numerical procedure is used. Reliability measures for the generalized exponential distribution are calculated. Testing the goodness of fit for the exponential distribution against the generalized exponential distribution is discussed. Relevant reliability measures of the generalized exponential distributions are also evaluated. A set of real data is employed to illustrate the results given in this paper.

Key Words : *Lifetime data, censored data, point estimate, interval estimate, maximum likelihood estimate.*

Acronyms

cdf	cumulative distribution function
pdf	probability density function
sf	survival function
MLE	maximum likelihood estimate
C.I.	confidence interval
GED(α, β)	generalized exponential distribution with parameters α, β
ED(α)	exponential distribution with parameter α
MTTF	mean time to failure

*Corresponding Author. *E-mail address:* asarhan0@yahoo.com
†*Home address:* Department of Mathematics, Faculty of Science, Mansoura University, Mansoura 35516, Egypt

1. INTRODUCTION

Tests, in reliability analysis, can be performed for units or systems either continuously or intermittently inspection for failure. If the life test can be based on continuous inspection throughout the experiment, then the sample life lengths, i.e. data, is said to be complete. In other hand, the data from intermittent inspections are known by grouped data. The number of failures expresses this data in each inspection interval is used more frequently than the first one, since it, generally, costs less and requires fewer efforts. Moreover, intermittent inspections is some times the only possible or available for units or their system, Ehrenfeld (1962).

It is known that the exponential distribution, GE(α), is the most frequently used distribution in reliability theory and applications. The mean and its confidence limit of this distribution have been estimated based on group and censored data by Seo and Yum(1993) and Chen and Mi (1996), respectively.

Recently the generalized exponential distribution GED(α, β) is introduced and studied quite extensively by Gupta and Kundu in a series of papers from (1999) to (2003) and Sarhan (2007). The cdf of GED(α, β) takes the following form

$$F(x; \alpha, \beta) = (1 - \exp\{-\alpha x\})^{\beta}, \quad x > 0, \alpha, \beta > 0. \tag{1.1}$$

The corresponding sf is

$$\bar{F}(x; \alpha, \beta) = 1 - (1 - \exp\{-\alpha x\})^{\beta}, \tag{1.2}$$

the pdf is

$$f(x; \alpha, \beta) = \alpha\beta \exp\{-\alpha x\} (1 - \exp\{-\alpha x\})^{\beta - 1}, \tag{1.3}$$

the hazard rate function is

$$h(x; \alpha, \beta) = \frac{\alpha\beta \exp\{-\alpha x\} (1 - \exp\{-\alpha x\})^{\beta - 1}}{1 - (1 - \exp\{-\alpha x\})^{\beta}}, \tag{1.4}$$

and the MTTF is

$$\text{MTTF} = \sum_{i=1}^{\infty} \binom{\beta}{i} \frac{(-1)^{i+1}}{i\alpha}. \tag{1.5}$$

Here α is the scale parameter and β is the shape parameter. As it seems GED(α, β) reduces to ED(α) when the shape parameter β equals 1. The GED(α, β) can have increasing and decreasing hazard rates depending on the shape parameter β. The hazard rate increase from 0 to α if $\beta > 1$ and if $\beta < 1$ it decreases from ∞ to α. This property leads to a good ability of using this distribution in reliability and life testing.

The main objective in this paper is to derive the MLE of the parameters of the GED(α, β), when the data are grouped and censored. The paper gives some reliability measures for the GED. At this situation, it is known that MLE does not give closed form, but iterative procedure can be used to give the required results. In fact numerical iterative procedure has been used for the GE by Ehrenfeld (1962) and Nelson (1977) among many others.

The rest of the paper is organized as follows. In section 2, we present the notations and assumptions used throughout this paper. Point and interval estimations of the unknown parameters are discussed in section 3. Testing for the goodness of fit of ED against the GED based on the estimated likelihood ratio test statistics is discussed in section 4. A set of real data is applied and a conclusion is drawn in section 5.

2. MODEL AND ASSUMPTIONS

Assumptions:

1. n independent and identical experimental units are put on a life test at time zero.
2. The lifetime of each unit follows a GED(α, β) with cdf given by (1.1).
3. The inspection times $0 < t_1 < t_2 < \cdots < t_k < \infty$ are predetermined.
4. The test is terminated at the predetermined time t_k. That is, the data is of Type-I censoring.
5. $t_0 = 0$ and $t_{k+1} = \infty$.
6. The number of failures in $(t_i, t_{i+1}]$ are recorded.

The data collected from the above test scheme consist of number of failures n_i in the interval $(t_{i-1}, t_i]$, $i = 1, 2, ...k$ and the number of units tested without failing up to t_k, n_{k+1} (censored units).

3. PARAMETER ESTIMATION

Based on the data collected in the previous section, the likelihood function takes the following form

$$L = C \prod_{i=1}^{k} [P\{t_{i-1} < T \leq t_i\}]^{n_i} \; [P\{T > t_k\}]^{n_{k+1}} \tag{3.1}$$

where $C = \frac{n!}{\prod_{\ell=1}^{k+1} n_\ell!}$ is a constant with respect to the parameters α and β.

But

$$P\{t_{i-1} < T \leq t_i\} = F(t_i) - F(t_{i-1}),$$

and

$$P\{T > t_k\} = 1 - F(t_k)\,.$$

Then based on assumption 2, we have

$$L = C\left[1-\left(1-e^{-\alpha t_k}\right)^{\beta}\right]^{n_{k+1}}\prod_{i=1}^{k}\left[\left(1-e^{-\alpha t_i}\right)^{\beta}-\left(1-e^{-\alpha t_{i-1}}\right)^{\beta}\right]^{n_i} \tag{3.2}$$

The log-likelihood function becomes

$$\mathcal{L} = \ln C + n_{k+1}\ln\left[1-\left(1-e^{-\alpha t_k}\right)^{\beta}\right]+\sum_{i=1}^{k} n_i \ln\left\{\left[\left(1-e^{-\alpha t_i}\right)^{\beta}-\left(1-e^{-\alpha t_{i-1}}\right)^{\beta}\right]\right\}, \tag{3.3}$$

The partial derivatives of $\mathcal{L}$ are

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial\alpha} &= -\frac{t_k n_{k+1}\beta e^{-\alpha t_k}\left(1-e^{-\alpha t_k}\right)^{\beta-1}}{1-\left(1-e^{-\alpha t_k}\right)^{\beta}} \\
&\quad +\beta\sum_{i=1}^{k}\frac{n_i\left[t_i e^{-\alpha t_i}\left(1-e^{-\alpha t_i}\right)^{\beta-1}-t_{i-1}e^{-\alpha t_{i-1}}\left(1-e^{-\alpha t_{i-1}}\right)^{\beta-1}\right]}{\left(1-e^{-\alpha t_i}\right)^{\beta}-\left(1-e^{-\alpha t_{i-1}}\right)^{\beta}}, \\
\frac{\partial\mathcal{L}}{\partial\beta} &= -\frac{n_{k+1}\left(1-e^{-\alpha t_k}\right)^{\beta}\ln\left(1-e^{-\alpha t_k}\right)}{1-\left(1-e^{-\alpha t_k}\right)^{\beta}}+\frac{n_1\left(1-e^{-\alpha t_1}\right)^{\beta}\ln\left(1-e^{-\alpha t_1}\right)}{\left(1-e^{-\alpha t_1}\right)^{\beta}} \\
&\quad +\sum_{i=2}^{k}\frac{n_i\left[\left(1-e^{-\alpha t_i}\right)^{\beta}\ln\left(1-e^{-\alpha t_i}\right)-\left(1-e^{-\alpha t_{i-1}}\right)^{\beta}\ln\left(1-e^{-\alpha t_{i-1}}\right)\right]}{\left(1-e^{-\alpha t_i}\right)^{\beta}-\left(1-e^{-\alpha t_{i-1}}\right)^{\beta}}.
\end{aligned}$$

Setting $\frac{\partial\mathcal{L}}{\partial\alpha}=0$ and $\frac{\partial\mathcal{L}}{\partial\beta}=0$, we get the likelihood equations, which should be solved to get the MLE of the parameters α and β. As it seems the likelihood equations have no closed form solutions in α and β. Therefore a numerical technique method should be used to get the solution.

Asymptotic confidence bounds: Since the MLE of the element of the vector of unknown parameters $\theta = (\alpha, \beta)$, are not obtained in closed forms, then it is not possible to derive the exact distributions of the MLE of these parameters. Thus we derive approximate confidence intervals of the parameters based on the asymptotic distributions of the MLE of the parameters. It is known that the asymptotic distribution of the MLE $\hat{\theta}$ is given by, see Miller (1981),

$$\left((\hat{\alpha}-\alpha),(\hat{\beta}-\beta)\right) \to N_2\left(0,\mathbf{I}^{-1}(\alpha,\beta)\right) \tag{3.4}$$

where $\mathbf{I}^{-1}(\alpha,\beta)$ is the variance covariance matrix of the unknown parameters $\theta = (\alpha,\beta)$. The elements of the 2×2 matrix $\mathbf{I}^{-1}$, $I_{ij}(\alpha,\beta)$, $i,j=1,2$, can be approximated by $I_{ij}(\hat{\alpha},\hat{\beta})$, where

$$I_{ij}(\hat{\theta}) = -\left.\frac{\partial^2 \mathcal{L}(\theta)}{\partial\theta_i\partial\theta_j}\right|_{\theta=\hat{\theta}} \tag{3.5}$$

From (3.3), we get the following

$$\frac{\partial^2 \mathcal{L}}{\partial\alpha^2} = \sum_{i=1}^{k+1} n_i \frac{P_i P_{i,\alpha^2} - [P_{i,\alpha}]^2}{P_i^2}, \tag{3.6}$$

$$\frac{\partial^2 \mathcal{L}}{\partial\beta^2} = \sum_{i=1}^{k+1} n_i \frac{P_i P_{i,\beta^2} - [P_{i,\beta}]^2}{P_i^2}, \tag{3.7}$$

$$\frac{\partial^2 \mathcal{L}}{\partial\alpha\partial\beta} = \sum_{i=1}^{k+1} n_i \frac{P_i P_{i,\alpha\beta} - P_{i,\alpha} P_{i,\beta}}{P_i^2}, \tag{3.8}$$

where

$$P_i = [g_i]^\beta - [g_{i-1}]^\beta, \qquad i = 1, 2, \dots, k+1,$$

$$P_{i,\alpha} = \beta\left\{[g_i]^{\beta-1} g_{i,\alpha} - [g_{i-1}]^{\beta-1} g_{i-1,\alpha}\right\},$$

$$P_{i,\beta} = [g_i]^\beta \ln g_i - [g_{i-1}]^\beta \ln g_{i-1},$$

$$P_{i,\beta^2} = [g_i]^\beta [\ln g_i]^2 - [g_{i-1}]^\beta [\ln g_{i-1}]^2,$$

$$P_{i,\alpha\beta} = [g_i]^{\beta-1} g_{i,\alpha} - [g_{i-1}]^{\beta-1} g_{i-1,\alpha} + \beta\Big\{[g_i]^{\beta-1} g_{i,\alpha} \ln g_i - [g_{i-1}]^{\beta-1} g_{i-1,\alpha} \ln g_{i-1}\Big\},$$

$$P_{i,\alpha^2} = \beta\Big\{(\beta-1)\Big[[g_i]^{\beta-2} g_{i,\alpha}^2 - [g_{i-1}]^{\beta-2} g_{i-1,\alpha}^2\Big] + [g_i]^{\beta-1} g_{i,\alpha^2} - [g_{i-1}]^{\beta-1} g_{i-1,\alpha^2}\Big\},$$

$$g_i = g(t_i;\alpha) = 1 - e^{-\alpha t_i}, i = 1, 2, \dots, k, \; g_{k+1} = 1, \; g_0 = 0,$$

$$g_{i,\alpha} = t_i\, e^{-\alpha t_i}, \quad g_{i,\alpha^2} = -t_i^2\, e^{-\alpha t_i}.$$

Therefore, the approximate $100(1-\gamma)\%$ two sided confidence intervals for α and β are, respectively, given by

$$\hat{\alpha} \pm Z_{\gamma/2}\sqrt{\mathbf{I}_{11}^{-1}(\hat{\alpha})}, \qquad \hat{\beta} \pm Z_{\gamma/2}\sqrt{\mathbf{I}_{22}^{-1}(\hat{\beta})}$$

Here, $Z_{\gamma/2}$ is the upper $(\gamma/2)$th percentile of a standard normal distribution.

4. GOODNESS OF FIT

The problem of testing goodness-of-fit of an exponential distribution model against the unrestricted class of alternative is complex. However, by restricting the alternative to a generalized exponential distribution, we can use the usual likelihood ratio test statistic, see (1973), to test the adequacy of an exponential distribution. The following are the null and the alternative hypotheses, respectively,

$$H_0 \;:\; \beta = 1, \text{ exponential distribution,}$$

$$H_1 \;:\; \beta \neq 1, \text{ generalized exponential distribution.}$$

In terms of the MLE, the likelihood ration test statistic for testing H_0 against H_1 is

$$\Lambda = \frac{L(\alpha, \beta = 1)}{L(\alpha, \beta)}. \tag{4.1}$$

Under the null hypothesis, $X_L = -2\ \ln(\Lambda) = 2(\mathcal{L}_{GE} - \mathcal{L}_E)$ follows a χ^2 distribution with 1 degree of freedom. Here $\mathcal{L}_E$ and $\mathcal{L}_{GE}$ are the log-likelihood functions under H_0 and H_1, respectively, after replacing the unknown parameters withe their MLE.

5. DATA ANALYSIS

Using the set of real data presented in Nelson (1982), which is a set of cracking data on 167 independent and identically parts in a machine. The test duration was 63.48 months and 8 unequally spaced inspections were conducted to obtain the number of cracking parts in each interval. The data were

$$(t_1, \cdots, t_8) = (6.12, 19.92, 29.64, 35.40, 39.72, 45.24, 52.32, 63.48)$$

and

$$(x_1, \cdots, x_8) = (5, 16, 12, 18, 18, 2, 6, 17, 73)$$

Assuming the ED(α) (or under H_0: $\beta = 1$), the MLE of α and MTTF are obtained as

$$\hat{\alpha} = 1.2097 \times 10^{-2}, \qquad \widehat{\text{MTTF}} = 82.6655.$$

The corresponding log-likelihood function is $\mathcal{L}_E = -316.6705$.

Assuming the GED(α, β), the MLE of the parameters α and β are obtained as

$$\hat{\alpha} = 2.0285 \times 10^{-2}, \text{ and } \hat{\beta} = 1.7839$$

The corresponding log-likelihood function is $\mathcal{L}_{GED} = -309.74$.
Therefore, the likelihood ratio test statistic is $X_L = 2(\mathcal{L}_{GED} - \mathcal{L}_{ED}) = 13.8592$ and the p-value is 1.9708×10^{-4}. Thus the GED($2.0285 \times 10^{-2}, 1.7839$) fits this data much better than ED(1.2097×10^{-2}).

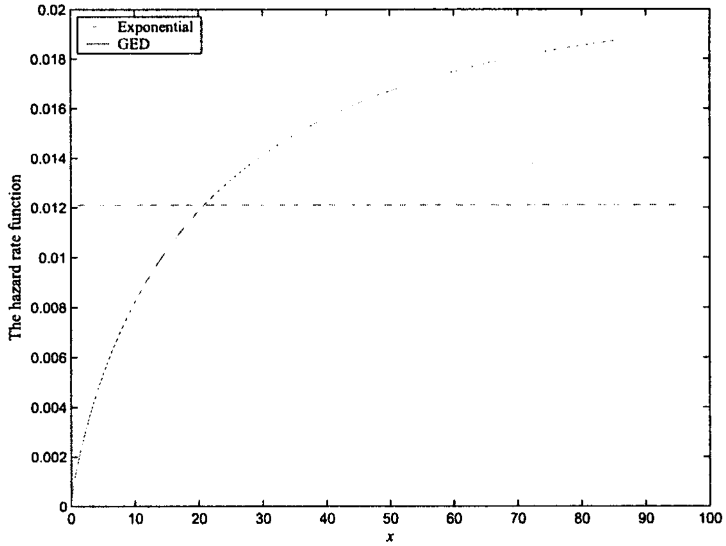

Figure 5.1. The estimated hazard rate functions.

The variance covariance matrix is computed as

$$I^{-1}=\begin{bmatrix} 8.1226\times 10^{-6} & 6.5626\times 10^{-4} \\ 6.5626\times 10^{-4} & 7.3470\times 10^{-2} \end{bmatrix}$$

Thus, the variances of the MLE of α and β become $\mathrm{Var}(\hat{\alpha}) = 8.1226\times 10^{-6}$ and $\mathrm{Var}\left(\hat{\beta}\right) = 7.3470\times 10^{-2}$. Therefore, the 95% C.I of α and β, respectively, are

$$\left[1.4699\times 10^{-2},\ 2.5871\times 10^{-2}\right],\quad [1.2526,\ 2.3151].$$

Figure 5.1 shows the hazard rate functions of the ED(α) and GED(α,β) computed when the MLE of the parameters replacing the unknown parameters.

Figure 5.2 shows the empirical estimate of survival function and estimation of survival function under H_0 and H_1. Also, we computed the Kolmogorov-Smirnov (K-S) distances of the empirical distribution function and the fitted distribution for the data set. The K-S distance between the empirical survival function and the fitted exponential survival function is 0.0547. The K-S distance between the empirical survival function and the fitted generalized exponential survival function is 0.0298.

Based on the 95% C.I of β and the values of the K-S distances, we get the same conclusion that the values of the likelihood ration test statistics and p-value which leads to that the GED($2.0285\times 10^{-2}, 1.7839$) fits this data rather than the ED(1.2097×10^{-2}).

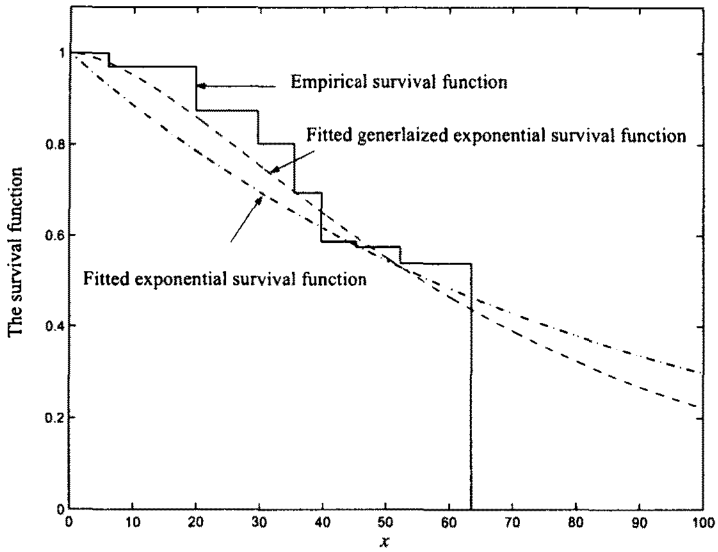

Figure 5.2. The empirical and fitted survival functions.

ACKNOWLEDGEMENTS

The authors would like to thank the Research Center at King Saud University for supporting this paper under the project number STAT/2006/37.

REFERENCES

Chen, Z. and Mi, J. (1996). Confidence interval for the mean of the exponential distribution, based on grouped data, *IEEE Trans. on Reliability,* **45**, 671-677.

Ehrenfeld, S. (1962). Some experimental design problems in attribute life testing, *J. Amer. Statistical Assoc.,* **57**, 668-679.

Gupta, R.D. and Kundu, D. (1999), Generalized exponential distributions, *Australian and New Zealand Journal of Statistics,* **41**, 173-188.

Gupta, R.D. and Kundu, D. (2001a), Exponentiated exponential distribution: an alternative to gamma and Weibull distributions, *Biometrical Journal,* **43**, 117-130.

Gupta, R.D. and Kundu, D. (2001b), Generalized exponential distributions: different methods of estimations, *Journal of Statistical Computation and Simulation,* **69**, 315-338.

Gupta, R.D. and Kundu, D. (2002), Generalized exponential distributions: Statistical inferences, *Journal of Statistical Theory and Applications,* **1**, 101-118.

Gupta, R.D. and Kundu, D. (2003), Closeness of gamma and generalized exponential distribution, *Communication in Statistics: Theory and Method,* **32**, 705-722.

Miller Jr., R. G., (1981). *Survival Analysis*, John Wiley, New York.

Nelson, W. (1982). *Applied Life Data Analysis*, John Weily & Sons.

Nelson, W. (1977). Optimum demonstration tests with grouped inspection data from an exponential distribution, *IEEE Trans. ov Reliability,* **26**, pp 226-231.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications, John Weily & Sons.*

Sarhan, A.M. (2007). Analysis of incomplete, censored data in competing risks models with generalized exponential distributions, *IEEE Trans. on Reliability,* **56**, 132-138.

Seo, S.K. and Yum, B.J. (1993). Estiamtion methods for the mean of the exponential distribution based on grouped and censored data, *IEEE Transactions on Reliability,* **42**, 97-96.