Int. J. Nonlinear Anal. Appl. 14 (2023) 1, 739–751
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.23746.2607

Parameter estimation for the generalized exponential distribution in the presence of interval censored data and covariate

Hussein Ali AL-Hakeem[a,*], Jayanthi Arasan[b], Mohd Shafie Bin Mustafa[b], Lim Fong Peng[b]

[a]*Institute for Mathematical Research,43400 UPM Serdang, Selangor, Malaysia*

[b]*Department of Mathematics & Statistics, Faculty of Science, 43400 UPM Serdang, Selangor, Malaysia*

(Communicated by Madjid Eshaghi Gordji)

Abstract

This paper concentrates on the parameter estimation for the Generalized Exponential Distribution (GED) in the presence of interval-censored data with covariate. Interval-censored data usually arises in clinical and epidemiological studies. This research attempts to investigate a conservative imputation technique to deal with interval censored data. This is achieved by comparing the bias, standard error (SE) and root mean square error (RMSE) of the maximum likelihood estimation (MLE) obtained without using imputation with the proposed imputation method at various censoring proportion and sample sizes. The results indicate that the proposed imputation method performs better than the traditional method at all sample sizes and censoring proportions.

Keywords: Generalized Exponential distribution, Interval censoring, Covariate, MLE
2020 MSC: 62E15, 62H10

1 Introduction

Survival modelling studies the relation between lifetime with one or more covariates. However, a decisive analytical problem occurs in analysing survival models when data is censored. One of the most common types of censoring in survival analysis is interval censoring. Interval censored data arises when a survival time T cannot be observed but can be specified to lie within the interval $[L_i, R_i]$ obtained from a sequence of examination times, where $L_i < T < R_i$. L_i and R_i are defined as the left and right endpoints respectively. Clinical trial assessments demonstrate good examples of these occurrences when patients are only evaluated at pre-scheduled visits and appointments. Interval-censored data can be analysed by utilizing non-parametric, semiparametric and parametric models.

In this paper, we are mainly interested in exploring the Generalized Exponential distribution in the presence of interval-censored data with covariates. The hazard function for this model is given by,

$$h(t;\alpha,\lambda,\mu) = \frac{\alpha}{\lambda}\frac{\left(1-e^{\frac{-(t-\mu)}{\lambda}}\right)^{\alpha-1} \quad e^{\frac{-(t-\mu)}{\lambda}}}{1-\left(1-e^{\frac{-(t-\mu)}{\lambda}}\right)^{\alpha}} \qquad (t>\mu). \tag{1.1}$$

*Corresponding author
Email addresses: ha828147@gmail.com (Hussein Ali AL-Hakeem), jayanthi@upm.edu.my (Jayanthi Arasan), mshafie@upm.edu.my (Mohd Shafie Bin Mustafa), fongpeng@upm.edu.my (Lim Fong Peng)

Received: May 2021 *Accepted:* July 2021

where λ is the scale parameter, α is the shape parameter, and μ is the location parameter. If the shape parameter $\alpha > 1$ the hazard (risk) function increases from 0 to λ, if $\alpha < 1$ the hazard (Risk) function is decreases from ∞ to λ and if $\alpha = 1$ the hazard (Risk) function remains unchanged [3]. The survival function for this model is given by,

$$S(t;\alpha,\lambda,\mu) = 1 - \left(1 - e^{\frac{-(t-\mu)}{\lambda}}\right)^{\alpha} \qquad (t > \mu, \alpha > 0, \lambda > 0). \tag{1.2}$$

and the probability density function for this model is given by,

$$f(t;\alpha,\lambda,\mu) = \frac{\alpha}{\lambda}\left(1 - e^{\frac{-(t-\mu)}{\lambda}}\right)^{\alpha-1} \quad e^{\frac{-(t-\mu)}{\lambda}} \qquad (t > \mu, \alpha > 0, \lambda > 0). \tag{1.3}$$

Originally, Generalized Exponential distribution was first discussed by [11] as an alternative to the popularly utilized Gamma and Weibull distributions. The three parameter Generalized Exponential distribution was developed by the [3],where they compared the theoretical characteristics of this family to the well-studied characteristics of the Weibull and the Gamma distributions. [5] studied the performance of the different estimators of the two unknown parameters of a Generalized Exponential distribution for various parameter values and various sample sizes.[2] focused on the estimation of parameters of the Generalized Exponential distribution via maximum likelihood estimates and moment methods in the presence of first type of interval censoring. [6, 4] mentioned that the two-parameters Generalized Exponential distribution could be used effectively in analysing certain survival time data, especially, in place of a two-parameters Weibull distribution or two-parameters Gamma distribution and in several cases the two-parameters Generalized Exponential distribution. [7] examined the statistical inference of the log-exponentiated Weibull regression model (LEWM) with interval censored data. [10] described how to assess survey process from a true interval-censored data combination. [12] implemented the goodness-of-fit tests with two parameters Gompertz model by using MLE based on interval and right-censored data. [8] reviewed estimation in interval censoring models, including nonparametric and semiparametric estimations of a distribution function and regression models.

The objective of this paper is to evaluate the performance of the maximum likelihood estimation for the Generalized exponential distribution in the presence of interval-censored data via simulation study at various censoring proportions and sample sizes using bias, standard error and root mean square error. In this research we will be comparing two methods of dealing with interval censored data. The first method (method (1)) will incorporate the probability of a subject i failing within the interval (L_i, R_i) directly into the likelihood function. The second method (method (2)) will approximate the failure time for the i^{th} subject, by randomly generating $\hat{t_i}$ from the Uniform (L_i, R_i) distribution.

2 Methodology

2.1 Generalized Exponential Distribution with Covariate and Interval Censored Data

Let T be a positive continuous random variable representing the survival time of individuals in a homogeneous population. Let $x_i = (x_{i1}, x_{i2}, ..., x_{iq})$ be a vector of q covariates for the i^{th} individuals, where; $i = 1, 2, ..., n$. The effect of covariates can be incorporated into the model by letting the scale parameter be a function of the covariates as shown below,

$$\lambda = e^{-(\beta_0 + \boldsymbol{\beta} x_i)}, \tag{2.1}$$

where $\boldsymbol{\beta} = \beta_1, \beta_2, ..., \beta_q$. The hazard function with covariates is given as:

$$h(t_i; x_i, \alpha, \boldsymbol{\beta}, \mu) = \frac{\alpha}{e^{-(\beta_0+\beta_1 x_i)}} \frac{\left(1 - e^{\frac{-(t_i-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha-1} \quad e^{\frac{-(t_i-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}}{1 - \left(1 - e^{\frac{-(t_i-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha}}. \tag{2.2}$$

the survival function with covariates can be expressed as follows:

$$S(t_i; x_i, \alpha, \boldsymbol{\beta}, \mu) = 1 - \left(1 - e^{\frac{-(t_i-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha}. \tag{2.3}$$

the probability density function with covariates can be expressed as follows:

$$f(t_i; x_i, \alpha, \boldsymbol{\beta}, \mu) = \frac{\alpha}{e^{-(\beta_0+\beta_1 x_i)}}\left(1 - e^{\frac{-(t_i-\mu)}{\exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha-1} \quad e^{\frac{-(t_i-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}. \tag{2.4}$$

2.2 Likelihood Equation and Estimation

The estimated parameters for the GED with covariates and interval censored data can be obtained by the method of maximum likelihood. Let $i = 1, 2, ..., n$ be censored and uncensored observation with covariates. The following indicator variables will be used to identify if observation was censored or uncensored,

$$\delta_{I_i} = \begin{cases} 1, & \text{if observation is interval censored;} \\ 0, & \text{Otherwise.} \end{cases}$$

$$\delta_{E_i} = \begin{cases} 1, & \text{if observation is uncensored;} \\ 0, & \text{Otherwise.} \end{cases}$$

The likelihood function of the full sample for the GED with covariates and interval censored data for method (1) is given by:

$$\begin{aligned} L(t_i, x_i, \boldsymbol{\psi}) &= \prod_{i=1}^{n} \{f(t_i, x_i, \psi)\}^{\delta_{E_i}} \quad \{S(t_{L_i}, x_i, \psi) - S(t_{R_i}, x_i, \psi)\}^{\delta_{I_i}} \\ &= \prod_{i=1}^{n} \left\{ \frac{\alpha}{e^{-(\beta_0+\beta_1 x_i)}} \left(1 - e^{\frac{-(t_i-\mu)}{\exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha-1} \quad e^{\frac{-(t_i-\mu)}{exp(-(\beta_0+\beta_1 x_i))}} \right\}^{\delta_{E_i}} \\ &\quad \times \left\{ \left[1 - \left(1 - e^{\frac{-(t_{L_i}-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha}\right] - \left[1 - \left(1 - e^{\frac{-(t_{R_i}-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha}\right] \right\}^{\delta_{I_i}}. \end{aligned}$$

where $\boldsymbol{\psi} = (\alpha, \beta_0, \beta_1, \mu)$ and log-likelihood function is,

$$\begin{aligned} \ell(t_i, x_i, \boldsymbol{\psi}) &= \sum_{i=1}^{n} \delta_{E_i} \log \left\{ \frac{\alpha}{e^{-(\beta_0+\beta_1 x_i)}} \left(1 - e^{\frac{-(t_i-\mu)}{\exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha-1} \quad e^{\frac{-(t_i-\mu)}{exp(-(\beta_0+\beta_1 x_i))}} \right\} \\ &\quad + \sum_{i=1}^{n} \delta_{I_i} \log \left\{ \left[1 - \left(1 - e^{\frac{-(t_{L_i}-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha}\right] - \left[1 - \left(1 - e^{\frac{-(t_{R_i}-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha}\right] \right\}. \end{aligned}$$

The likelihood function of the full sample for the GED with covariates and interval censored data for method (2) is given by:

$$\begin{aligned} L(t_i, x_i, \boldsymbol{\psi}) &= \prod_{i=1}^{n} \{f(t_i, x_i, \psi)\}^{\delta_{E_i}} \quad \{S(\hat{t_i}, x_i, \psi)\}^{\delta_{I_i}} \\ &= \prod_{i=1}^{n} \left\{ \frac{\alpha}{e^{-(\beta_0+\beta_1 x_i)}} \left(1 - e^{\frac{-(t_i-\mu)}{\exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha-1} \quad e^{\frac{-(t_i-\mu)}{exp(-(\beta_0+\beta_1 x_i))}} \right\}^{\delta_{E_i}} \\ &\quad \times \left\{ \left[1 - \left(1 - e^{\frac{-(\hat{t_i}-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha}\right] \right\}^{\delta_{I_i}}. \end{aligned}$$

and log-likelihood function is,

$$\begin{aligned} \ell(t_i, x_i, \boldsymbol{\psi}) &= \sum_{i=1}^{n} \delta_{E_i} \log \left\{ \frac{\alpha}{e^{-(\beta_0+\beta_1 x_i)}} \left(1 - e^{\frac{-(t_i-\mu)}{\exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha-1} \quad e^{\frac{-(t_i-\mu)}{exp(-(\beta_0+\beta_1 x_i))}} \right\} \\ &\quad + \sum_{i=1}^{n} \delta_{I_i} \log \left\{ 1 - \left(1 - e^{\frac{-(\hat{t_i}-\mu)}{exp(-(\beta_0+\beta_1 x_i))}}\right)^{\alpha} \right\} \end{aligned}$$

where $\hat{t_i} \sim$Uniform (L_i, R_i). The Newton-Raphson iterative procedure will be used for solving the non linear likelihood equations. The variance-covariance matrix can be estimated by taking the inverse of the observed information matrix $i(\hat{\alpha}, \hat{\beta_0}, \hat{\beta_1}, \hat{\mu})$ which can be obtained from the second partial derivative of the log-likelihood function evaluated at $\hat{\alpha}, \hat{\beta_0}, \hat{\beta_1}, \hat{\mu}$. Following that $\hat{var}(\hat{\alpha}, \hat{\beta_0}, \hat{\beta_1}, \hat{\mu})$, can be obtained from $[i(\hat{\alpha}, \hat{\beta_0}, \hat{\beta_1}, \hat{\mu})]^{-1}$

3 Simulation Study

A simulation study with 1000 replication each with n=40, 100, 150, 180 and 200 were performed to assess the performance of the GED with covariates and interval censored data using method (1) and method (2). The values of the covariates x_i were simulated independently from the Standard Normal distribution $x_i \sim N(0,1)$. To simulate real-life survival data, the values 1, -0.05, 0.03 and 0.05 were selected as parameters of α, β_0, β_1 and μ respectively. A series of random numbers, u_i, was generated from Uniform Distribution $U(0,1)$ to produce lifetime t_i for i = 1, 2,.....,n subjects. The lifetime t_i was generated via on inverse transform method as following,

$$t_i = \mu - log(1-(u_i)^{\frac{1}{\alpha}})e^{-(\beta_0+\beta_1 x_i)}.$$

The censoring time, C_i, was generated from exponential distribution $exp(\omega)$ where the value ω would be adjusted to obtain the desired approximate censoring proportion (CP) for the data with $10\%, 20\%, 30\%, 40\%$ and 50%, where $C_i \leq t_i$. Thus the data will consist of uncensored and interval censored data. In order to evaluate the performance of the estimator at different combination of sample sizes and censoring proportions, the bias, standard error (SE) and root mean square error (RMSE) of the parameter were calculated.

4 Results and Discussions

Table 1: Bias values of the parameters with covariate for method (1)

Estimates	n	Bias				
		CP				
		10%	20%	30%	40%	50%
α	40	-0.016405	-0.016445	-0.017546	-0.017593	-0.018505
	100	-0.007808	-0.009354	-0.009523	-0.009890	-0.010789
	150	-0.003373	-0.003602	-0.003345	-0.003283	-0.003609
	180	-0.001768	-0.001801	-0.002067	-0.002271	-0.002351
	200	-0.001250	-0.001682	-0.001325	-0.001765	-0.001844
β_0	40	0.002064	0.002097	0.002000	0.002054	0.004859
	100	0.000783	0.000977	0.000751	0.000777	0.001125
	150	0.000330	0.000336	0.000295	0.000303	0.000311
	180	0.000193	0.000200	0.000230	0.000261	0.000273
	200	0.000132	0.000127	0.000144	0.000204	0.000211
β_1	40	0.000962	0.001054	-0.002239	-0.006221	-0.049284
	100	0.000538	0.000471	0.000546	0.000592	0.000477
	150	0.000207	0.000211	0.000227	0.000224	0.000287
	180	0.000128	0.000128	0.000137	0.000150	0.000159
	200	0.000088	0.000112	0.000122	0.000150	0.000155
μ	40	0.020893	0.021078	0.021169	0.021139	0.022477
	100	0.009309	0.011347	0.011430	0.011591	0.012834
	150	0.003724	0.003769	0.003919	0.003840	0.004231
	180	0.002153	0.002205	0.002562	0.002867	0.002981
	200	0.001541	0.001406	0.001597	0.002216	0.002607

The results of the simulation study are given in tables (1)-(6). We calculated the value of bias, standard error (SE) and root mean square error (RMSE) at all combination of censoring proportion and sample sizes. From the previous tables (1)-(6), we can easily see that the values of bias, standard error (SE), and root mean square error

Table 2: SE values of the parameters with covariate for method (1)

Estimates	n	SE				
		CP				
		10%	20%	30%	40%	50%
α	40	0.012736	0.011948	0.013135	0.013199	0.013544
	100	0.008431	0.007947	0.008028	0.008351	0.010276
	150	0.006083	0.006098	0.006673	0.006448	0.006778
	180	0.004346	0.004374	0.004618	0.004599	0.004619
	200	0.003833	0.003913	0.004068	0.004237	0.004269
β_0	40	0.002649	0.002653	0.006761	0.008598	0.021410
	100	0.001151	0.001188	0.001253	0.000990	0.001463
	150	0.000782	0.000793	0.000713	0.000720	0.000724
	180	0.000522	0.000542	0.000580	0.000623	0.000633
	200	0.000417	0.000490	0.000514	0.000631	0.000633
β_1	40	0.007088	0.007441	0.006612	0.013679	0.035109
	100	0.001380	0.001367	0.001421	0.001533	0.001626
	150	0.000755	0.000769	0.000817	0.000820	0.000936
	180	0.000481	0.000501	0.000532	0.000504	0.000516
	200	0.000324	0.000550	0.000589	0.000549	0.000552
μ	40	0.015630	0.015850	0.015975	0.015947	0.016693
	100	0.010498	0.009976	0.009993	0.010049	0.012836
	150	0.007283	0.007323	0.007658	0.007622	0.008071
	180	0.005366	0.005436	0.005817	0.005967	0.006012
	200	0.004069	0.004729	0.004995	0.005546	0.005562

(RMSE) increase with the increase in censoring proportion (CP) and decrease with the increase in sample size. This means better performance of the estimates at lower censoring proportions and higher sample sizes. In addition, most values of the bias, SE and RMSE of the parameter estimates of the method (2) are slightly lower compared to the parameters of the method (1).This indicates that the method (2) would have higher accuracy and efficiency compared to the method (1) of the parameter estimates.

5 Application with Real Data

In this dataset 197 patients were chosen at random as 50% from a population patient with "high-risk" diabetic retinopathy as defined by the diabetic retinopathy study (DRS). One eye of each patient was randomly treated with a laser while the other eye got no treatment. For each eye, the time from the start of therapy to the time when visual acuity went below 5/200 for two visits in a row was the event of interest. In the event of death, dropout, or end of the study, a subject was censored. The Clayton Oakes model with exponential and Weibull marginals was previously used to analyse the DRS data by [9] whereas [13] examined a few models based on the bivariate exponential and Weibull distributions using Bayesian approaches, As for [1] they studied a parallel system survival model based on the bivariate exponential in the Presence of a time-varying covariate. In this study, the objective of analysing the DRS data is to see whether age affects the Fit of the Generalized Exponential Model. Where Diabetic Retinopathy data consists of 394 observations on time to event or last follow-up which measured in months, Age is one of the covariates available with this data set, Where the diagnosed in years, As for the status, where the 0= censored or 1 = visual loss. Firstly we will fit the model with complete data set with respect to age covariate.

Firstly, the non-parametric Kaplan-Meier (KM) estimates for the survival function was obtained and the Generalized Exponential Model was fitted to the data with full age covariate. Then, the two survival curves were plotted on

Table 3: RMSE values of the parameters with covariate for method (1)

Estimates	n	RMSE				
		CP				
		10%	20%	30%	40%	50%
α	40	0.020171	0.020327	0.021917	0.021994	0.022932
	100	0.011491	0.012273	0.012567	0.012944	0.014901
	150	0.006861	0.006888	0.007286	0.007236	0.007679
	180	0.004693	0.004730	0.005059	0.005129	0.005183
	200	0.003973	0.004090	0.004279	0.004590	0.004650
β_0	40	0.003358	0.003382	0.007051	0.008840	0.021954
	100	0.001392	0.001539	0.001637	0.001259	0.001845
	150	0.000849	0.000861	0.000771	0.000779	0.000788
	180	0.000557	0.000578	0.000625	0.000676	0.000691
	200	0.000437	0.000506	0.000534	0.000663	0.000668
β_1	40	0.007153	0.007515	0.006615	0.013693	0.035452
	100	0.001481	0.001445	0.001520	0.001643	0.001694
	150	0.000783	0.000801	0.000848	0.000848	0.000979
	180	0.000498	0.000514	0.000550	0.000526	0.000552
	200	0.000336	0.000571	0.000602	0.000569	0.000574
μ	40	0.026093	0.026454	0.026520	0.026480	0.027998
	100	0.014031	0.015108	0.015301	0.015341	0.018151
	150	0.008180	0.008236	0.008602	0.008535	0.009112
	180	0.005782	0.005866	0.006357	0.006620	0.006911
	200	0.004559	0.004934	0.005244	0.005972	0.006021

the same graph, Which is shown in Figure 3. Secondly, the non-parametric Kaplan-Meier estimates for the survival function was obtained and the Generalized Exponential distribution was fitted to the data when we classified DRS into two general groups by the age, adult DRS (age>20 years) and juvenile DRS (age$\leq$ 20 years). Then, the two survival curves were plotted on the same graph, Which is shown in Figure 4 and Figure 5

Table (7) show the values of the parameter estimates, standard errors and the length of 95% and 90% confidence intervals for the parameters α, β_0, β_1 and μ using the Z-test.

Figures 3-5 depict the plot of the survival functions obtained using the KM and the GED estimator for the modified Diabetic Retinopathy data. From Figure 3 It can be seen that the reduced Generalized Exponential distribution provides a satisfactory fit for the data as the estimated survival functions are approximately close to the values obtained using the nonparametric KM estimator. Hence, the GED is appropriate to be fitted into Diabetic Retinopathy data. Figure 4 and Figure 5, show that the plot of survival function obtained from GED to adult (age>20 years) is close to the non-parametric Kaplan-Meier plot than for juvenile (age$\leq$20 years). This indicates that the GED to adult (age>20 years) is appropriate to be fitted into Diabetic Retinopathy data more than juvenile (age$\leq$20 years).

Table 4: Bias values of the parameters with covariate for method (2)

Estimates	n	Bias				
		CP				
		10%	20%	30%	40%	50%
α	40	-0.011422	-0.012781	-0.013541	-0.017219	-0.016236
	100	-0.006623	-0.007584	-0.007811	-0.008347	-0.008883
	150	-0.003191	-0.003799	-0.003742	-0.003425	-0.004119
	180	-0.001603	-0.001771	-0.002047	-0.002225	-0.002329
	200	-0.001215	-0.001572	-0.001226	-0.001329	-0.001750
β_0	40	0.000973	0.000656	-0.000730	0.001555	0.001837
	100	0.000631	0.000812	0.000648	0.000677	0.000875
	150	0.000262	0.000327	0.000285	0.000271	0.000318
	180	0.000144	0.000158	0.000187	0.000254	0.000269
	200	0.000124	0.000131	0.000133	0.000117	0.000170
β_1	40	0.001153	0.000440	0.000753	0.001131	0.001556
	100	0.000452	0.000517	0.000522	0.000401	0.000617
	150	0.000191	0.000237	0.000273	0.000156	0.000371
	180	0.000177	0.000085	0.000120	0.000150	0.000192
	200	0.000068	0.000096	0.000071	0.000085	0.000032
μ	40	0.014605	0.015541	0.017072	0.022402	0.021479
	100	0.008354	0.009777	0.010057	0.010394	0.011282
	150	0.003699	0.004634	0.004752	0.004255	0.005011
	180	0.002058	0.002181	0.002451	0.002819	0.002934
	200	0.001513	0.001332	0.001517	0.001589	0.001022

Table 5: SE values of the parameters with covariate for method (2)

Estimates	n	SE				
		CP				
		10%	20%	30%	40%	50%
α	40	0.004688	0.011414	0.011667	0.016993	0.015980
	100	0.007256	0.007492	0.007731	0.008566	0.008730
	150	0.004823	0.005116	0.005447	0.005402	0.006156
	180	0.004338	0.003323	0.003810	0.004491	0.004841
	200	0.002644	0.003229	0.003502	0.002612	0.002027
β_0	40	0.000899	0.001030	0.001139	0.002961	0.003621
	100	0.000916	0.001109	0.001136	0.000936	0.001175
	150	0.000482	0.000514	0.000535	0.000538	0.000661
	180	0.000463	0.000333	0.000391	0.000608	0.000634
	200	0.000284	0.000302	0.000443	0.000262	0.000190
β_1	40	0.001130	0.001252	0.001741	0.004345	0.006357
	100	0.001055	0.001314	0.001332	0.000871	0.001359
	150	0.000556	0.000712	0.000720	0.000432	0.000791
	180	0.000417	0.000294	0.000340	0.000514	0.000664
	200	0.000255	0.000423	0.000271	0.000319	0.000500
μ	40	0.006898	0.014614	0.015618	0.024107	0.023556
	100	0.006569	0.010086	0.009636	0.011151	0.011580
	150	0.006018	0.006881	0.006818	0.006962	0.007740
	180	0.005900	0.004166	0.004668	0.005847	0.006165
	200	0.003200	0.004089	0.004461	0.003150	0.004090

Table 6: RMSE values of the parameters with covariate for method (2)

Estimates	n	RMSE				
		CP				
		10%	20%	30%	40%	50%
α	40	0.012247	0.017136	0.017745	0.024192	0.022781
	100	0.009824	0.010661	0.011336	0.011961	0.012455
	150	0.005784	0.006598	0.006608	0.006396	0.007406
	180	0.005319	0.003766	0.004326	0.005012	0.005372
	200	0.002944	0.003572	0.003711	0.002930	0.002207
β_0	40	0.001325	0.001221	0.001353	0.003346	0.004061
	100	0.001113	0.001375	0.001407	0.001155	0.001465
	150	0.000548	0.000602	0.000606	0.000622	0.000689
	180	0.000524	0.000368	0.000433	0.000659	0.000688
	200	0.000310	0.000334	0.000463	0.000287	0.000203
β_1	40	0.001615	0.001327	0.001897	0.004490	0.006545
	100	0.001148	0.001412	0.001381	0.000958	0.001426
	150	0.000587	0.000750	0.000752	0.000456	0.000852
	180	0.000453	0.000305	0.000360	0.000536	0.000692
	200	0.000264	0.000434	0.000281	0.000330	0.000534
μ	40	0.016152	0.021333	0.023138	0.032998	0.031878
	100	0.012702	0.013978	0.014382	0.015244	0.016167
	150	0.007171	0.008296	0.008197	0.008160	0.009221
	180	0.006740	0.004703	0.005272	0.006491	0.006828
	200	0.003561	0.004795	0.004312	0.003528	0.006167

Table 7: MLE of Diabetic Retinopathy patients data with 95% and 90% confidence intervals

Estimate	SE	95% Z-test	length	90% Z-test	length
0.6999857	0.0002545	(0.6994869, 0.7004845)	0.0009976	(0.6992671, 0.7004043)	0.0011372
-0.4522179	0.0108751	(-0.4735327, -0.4309031)	-0.0426296	(-0.4751058, -0.4343351)	-0.0407707
-0.3441711	0.0108791	(-0.3654937, -0.3228485)	-0.0426452	(-0.3660656, -0.3262766)	-0.0397890
0.2851697	0.0106155	(0.2643637, 0.3059757)	0.0416120	(0.2627088, 0.3026306)	0.0399218

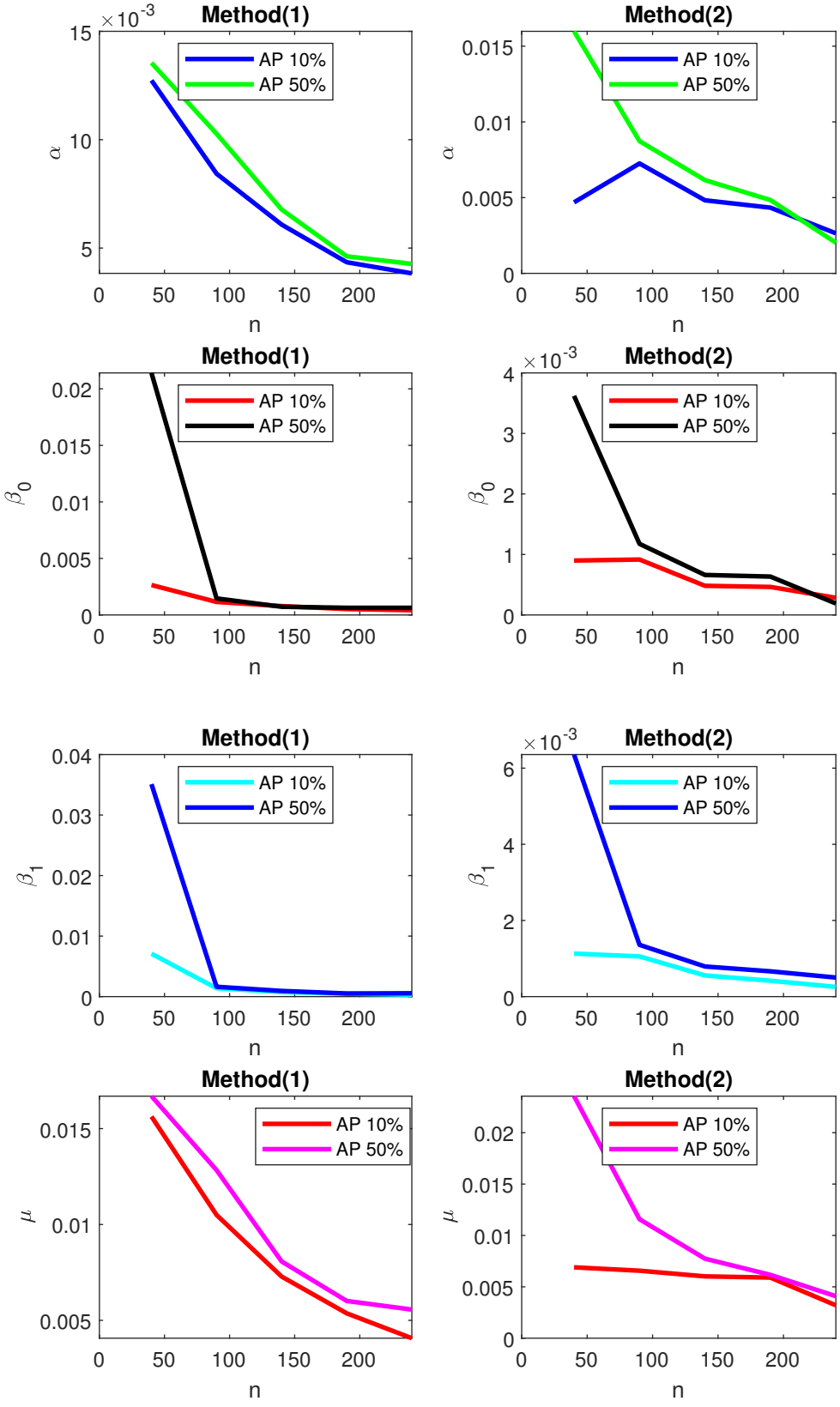

Figure 1: Comparison the SE values of method(1) and method(2)

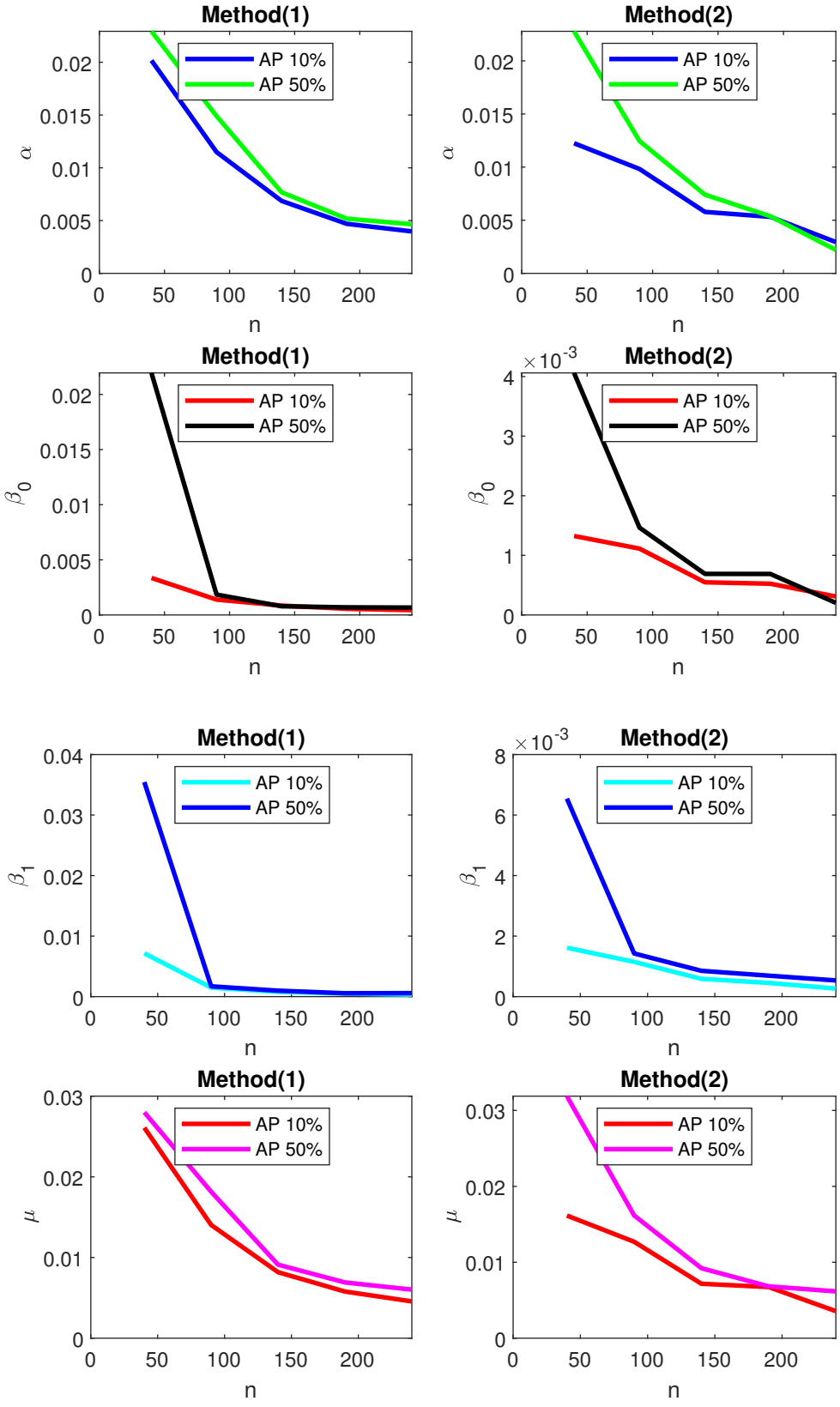

Figure 2: Comparison the RMSE values of method(1) and method(2)

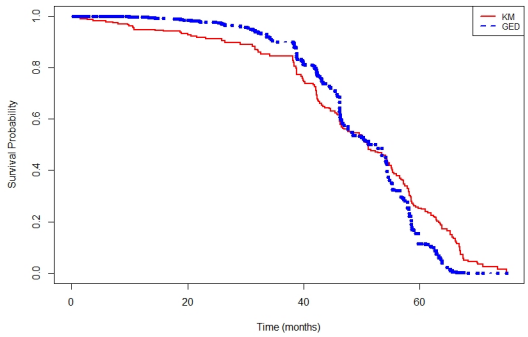

Figure 3: Plot of Kaplan-Meier and GED survival functions with full age

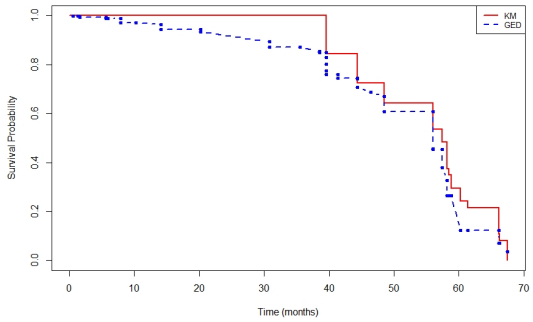

Figure 4: Plot of Kaplan-Meier and GED survival functions, when age>20 years

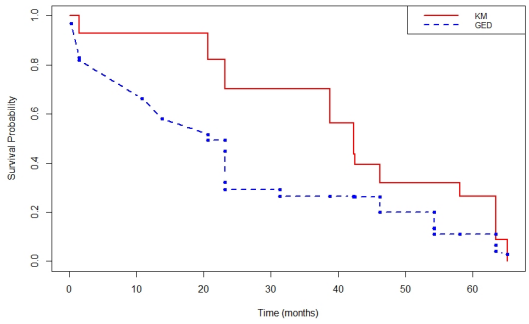

Figure 5: Plot of Kaplan-Meier and GED survival functions, when age≤20 years

6 Conclusions and Recommendations

In this research, the maximum likelihood estimation (MLE) for the three parameters of the Generalized Exponential distribution (GED) with covariate in the presence of Interval-censored data was obtained. It was shown that the bias, SE and RMSE increase substantially when censoring proportion (CP) increases and sample size decreases. Also,it was shown that the values of the estimated parameters of Bias, SE and RMSE in which generated from the method (2) are slightly lower compared to the estimated parameter values in method (1), this indicates that the method (2) produces better estimate of the parameters compared to the method (1). The discussion in this paper was restricted to one type of censored data. Thus, it would be possible to carry out further work to include more censored data as right-censored data. Also, the model could be extended to include a larger number of covariates to see their performance. The Generalized Exponential distribution (GED) could also be extended further to include additional scale parameters, if necessary. Also, we can use other methods to get good performance of the parameters as midpoint imputation method.

Acknowledgment

We would like to thank Universiti Putra Malaysia for financially supporting this research work via the PUTRA GRANT, VOT 9595300.

References

[1] J. Arasan and M. Lunn, *Survival model of a parallel system with dependent failures and time varying covariates*, J. Statist. Plan. Infer. 139 (2009), no. 3, 944–951.

[2] D.G. Chen and Y.L. Lio, *Parameter estimations for generalized exponential distribution under progressive type-I interval censoring*. Comput. Statist. Data Anal. **54** (2010), no. 6, 1581–1591.

[3] R.D. Gupta and D. Kundu, *Theory & methods: Generalized exponential distributions*, Aust. New Zealand J. Statist. **41** (1999), no. 2, 173–188.

[4] R.D. Gupta and D. Kundu, *Exponentiated exponential family: an alternative to gamma and Weibull distributions*, Biometrical J.: J. Math. Meth. Biosci. **43** (2001), no. 1, 117–130.

[5] R.D. Gupta and D. Kundu, *Generalized exponential distribution: different method of estimations*, J. Statist. Comput. Simul. **69** (2001), no. 4, 315–337.

[6] R.D. Gupta and D. Kundu, *Discriminating between Weibull and generalized exponential distributions*, Comput. Statist. Data Anal. **43** (2003), no. 2, 179–196.

[7] E.M. Hashimoto, E.M.M. Ortega, V.G. Cancho and G.M. Cordeiro, *The log-exponentiated Weibull regression model for interval-censored data*, Comput. Statist. Data Anal. **54** (2010), no. 4, 1017–1035.

[8] J. Huang and J.A. Wellner, *Interval censored survival data: a review of recent progress*, Proc. First Seattle Symp. Biostatist., New York, 1997, pp. 123–169.

[9] W.J. Huster, R. Brookmeyer and S.G. Self, *Modelling paired survival data with covariates*, Biometrics **45** (1989), no. 1, 145–156.

[10] J.F. Lawless and D. Babineau, *Models for interval censoring and simulation-based inference for lifetime distributions*, Biometrika **93** (2006), no. 3, 671–686.

[11] Mudholkar, Govind S., and Deo Kumar Srivastava, *Exponentiated Weibull family for analyzing bathtub failure-rate data*, IEEE Trans. Reliab. **42** (1993), no. 2, 299–302.

[12] A.M.N.N. Naslina, A. Jayanthi, Z.H. Syahida and A.M. Bakri, *Assessing the goodness of fit of the gompertz model in the presence of right and interval censored data with covariate*, Aust. J. Statist. **49** (2020), no. 3, 57–71.

[13] S.K. Sahu and K. D. Dipak, *A comparison of frailty and other models for bivariate survival data*, Lifetime Data Anal. **6** (2000), no. 3, 207–228.