# A New Algorithm for MLE with Interval Censored Data 

Conghua CHENG ${ }^{1, *}$, Jinyuan CHEN ${ }^{2}$<br>1. School of Mathematics and Computation Science, Zhanjiang Normal University, Guangdong 524048. P. R. China;<br>2. School of Mathematics and Statistics, Lanzhou University, Gansu 730000, P. R. China


#### Abstract

In this paper, we study the two-parameter maximum likelihood estimation (MLE) problem for the GE distribution with consideration of interval data. In the presence of interval data, the analytical forms for the restricted MLE of the parameters of GE distribution do not exist. Since interval data is kind of incomplete data, the EM algorithm can be applied to compute the MLEs of the parameters. However the EM algorithm could be less effective. To improve effectiveness, an equivalent lifetime method is employed. The two methods are discussed via simulation studies.


Keywords GE distribution; EM algorithm; equivalent method; MLE; interval data.
MR(2010) Subject Classification 62F10; 62N01

## 1. Introduction

The two parameter GE distribution was introduced by Gupta and Kundu [1]. The GE distribution, density function and survival function have the following forms:

$$
\begin{gather*}
F(x)=\left(1-e^{-\lambda x}\right)^{\alpha}  \tag{1}\\
f(x)=\alpha \lambda\left(1-e^{-\lambda x}\right)^{\alpha-1} e^{-\lambda x}  \tag{2}\\
R(x)=1-\left(1-e^{-\lambda x}\right)^{\alpha} \tag{3}
\end{gather*}
$$

where $\lambda, \alpha, x>0, \lambda$ and $\alpha$ are the scale parameter and shape parameter of GE distribution respectively. GE distribution with parameters $\lambda$ and $\alpha$ is denoted by $\operatorname{GE}(\alpha, \lambda)$. The GE distribution has lots of interesting properties which can be referred to Gupta and Kundu [2, 3]. In recent years, it is observed that the two-parameter $\operatorname{GE}(\alpha, \lambda)$ can be used quite effectively in analyzing many lifetime data, particularly in place of two-parameter gamma and two-parameter Weibull distributions. It is observed that in many situations GE distribution provides better fit than Weibull distribution, gamma distribution or log-normal distribution [4-7].

GE distribution has been studied in many literatures and has lots of applications in the field other than lifetime distributions [8-11]. Many papers considered the parameter estimation of the

[^0]GE distribution for complete sample case in the literature. Mitra and Kundu [12] analyzed the GE distribution with left censored data. Pradhan and Kundu [13] considered statistical inference of the parameters of the GE distribution in presence of progressive censoring. However, few papers considered the estimation of the GE distribution with interval data, which is a common form of data in lifetime analysis or reliability theory.

In this paper, we study the estimations of the shape parameter $\alpha$ and the scale parameter $\lambda$ in GE distribution with consideration of interval data, right censored data, including complete data. In presence of the interval data and the right censored data, regular MLE method can be applied, but there are no analytical solutions for the parameters $\alpha$ and $\lambda$. The EM algorithm can be applied to solve the estimation problem of GE distribution with the three types of data. However the EM algorithm could be less effective. We apply an equivalent lifetime method to the problems. The equivalent lifetime method was proposed by Tan [14]. This method of estimation for exponential distribution and Weibull distribution has been successfully applied by Tan [15].

The rest of this paper is organized as follows. In Section 2, we discuss the MLEs. In Section 3, we introduce the equivalent lifetime method for MLEs of the parameters. Simulation study and some discussions are given in Section 4.

## 2. Maximum likelihood estimation

With $n$ observed data $x_{1}, x_{2}, \ldots, x_{n}, m$ right censored data $c_{1}, c_{2}, \ldots, c_{m}$ and $l$ interval data $\left(a_{1}, b_{1}\right),\left(a_{2}, b_{2}\right), \ldots,\left(a_{l}, b_{l}\right)$, the log likelihood function is

$$
\begin{align*}
\log (L)= & \sum_{i=1}^{n} \log f\left(x_{i}\right)+\sum_{i=1}^{m} \log R\left(c_{i}\right)+\sum_{i=1}^{l} \log \left(F\left(b_{i}\right)-F\left(a_{i}\right)\right) \\
= & n \log \alpha+n \log \lambda+(\alpha-1) \sum_{i=1}^{n} \log \left(1-e^{-\lambda x_{i}}\right)-\lambda \sum_{i=1}^{n} x_{i}+ \\
& \sum_{i=1}^{m} \log \left(1-\left(1-e^{-\lambda c_{i}}\right)^{\alpha}\right)+\sum_{i=1}^{l} \log \left(\left(1-e^{-\lambda b_{i}}\right)^{\alpha}-\left(1-e^{-\lambda a_{i}}\right)^{\alpha}\right) \tag{4}
\end{align*}
$$

Taking derivatives with respect to parameters $\alpha$ and $\lambda$ respectively and setting them to be zero, we get the log likelihood equations

$$
\begin{align*}
\frac{\partial \log (L)}{\partial \alpha}= & \frac{n}{\alpha}+\sum_{i=1}^{n} \log \left(1-e^{-\lambda x_{i}}\right)-\sum_{i=1}^{m} \frac{\left(1-e^{-\lambda c_{i}}\right)^{\alpha} \log \left(1-e^{-\lambda c_{i}}\right)}{1-\left(1-e^{-\lambda c_{i}}\right)^{\alpha}}+ \\
& \sum_{i=1}^{l} \frac{\left(1-e^{-\lambda b_{i}}\right)^{\alpha} \log \left(1-e^{-\lambda b_{i}}\right)-\left(1-e^{-\lambda a_{i}}\right)^{\alpha} \log \left(1-e^{-\lambda a_{i}}\right)}{\left(1-e^{-\lambda b_{i}}\right)^{\alpha}-\left(1-e^{-\lambda a_{i}}\right)^{\alpha}}=0  \tag{5}\\
\frac{\partial \log (L)}{\partial \lambda}= & \frac{n}{\lambda}+(\alpha-1) \sum_{i=1}^{n} \frac{x_{i} e^{-\lambda x_{i}}}{1-e^{-\lambda x_{i}}}-\sum_{i=1}^{n} x_{i}-\sum_{i=1}^{m} \frac{\alpha c_{i}\left(1-e^{-\lambda c_{i}}\right)^{\alpha-1} e^{-\lambda c_{i}}}{1-\left(1-e^{-\lambda c_{i}}\right)^{\alpha}}+ \\
& \sum_{i=1}^{l} \frac{\alpha a_{i}\left(1-e^{-\lambda a_{i}}\right)^{\alpha} e^{-\lambda a_{i}}-\alpha b_{i}\left(1-e^{-\lambda b_{i}}\right)^{\alpha} e^{-\lambda b_{i}}}{\left(1-e^{-\lambda b_{i}}\right)^{\alpha}-\left(1-e^{-\lambda a_{i}}\right)^{\alpha}}=0 \tag{6}
\end{align*}
$$

In the presence of the right censored data and the interval data, from (5) and (6), we can see the
log likelihood equations are so complex that the two parameters are highly correlative. We can no longer obtain the MLE of $\alpha$ and $\lambda$ in an analytical form. Since interval data and censored data are kind of incomplete data [16], we propose to use EM algorithm to compute the MLEs of $\alpha$ and $\lambda$. There are some obstacles in the EM algorithm. First, it is well known that the EM algorithm may converge slowly even in some seemingly innocuous problems. It will be seen when we do some simulation studies in Section 4. Another issue is that in order to obtain the estimations of the parameters, we have to solve a boring equation which involves some boring integrals. Therefore, it may be less effective and efficient in computation. In order to overcome those obstacles, we use an equivalent method in the following section.

## 3. Maximum likelihood estimation via equivalent method

The equivalent method combines the GE-to-Exponential transformation and the equivalent lifetime method. First, we give an algorithm to estimate the failure rate of exponential distribution and a transformation with GE random variable to exponential random variable. Then, we solve the estimation problem for GE distribution through an equivalent method.

### 3.1 MLE of failure rate in exponential distribution

For exponential distribution $F_{E}(x)=1-e^{-\mu x}$, with $n$ complete data $x_{1}, x_{2}, \ldots, x_{n}, m$ right censored data $c_{1}, c_{2}, \ldots, c_{m}$, it is well known that the MLE of $\mu$ is

$$
\begin{equation*}
\widehat{\mu}=\frac{n}{\sum_{h=1}^{n} x_{h}+\sum_{j=1}^{m} c_{j}} . \tag{7}
\end{equation*}
$$

In addition, there are $l$ interval data $\left(a_{1}, b_{1}\right),\left(a_{2}, b_{2}\right), \ldots,\left(a_{l}, b_{l}\right)$. The $l$ pseudo values $x_{i}^{*}(\mu)$, $i=1,2, \ldots, l$ can be used to replace the intervals, where, $x_{i}^{*}(\mu)$ s are the conditional MTTF, that is,

$$
x_{i}^{*}(\mu)=E\left(X_{i} \mid a_{i} \leq X_{i} \leq b_{i}\right)=\frac{a_{i} R_{E}\left(a_{i}\right)-b_{i} R_{E}\left(b_{i}\right)}{R_{E}\left(a_{i}\right)-R_{E}\left(b_{i}\right)}+\frac{1}{\mu},
$$

where $X_{i}$ is an exponential random variable following $F_{E}(x)$ and $R_{E}(x)=e^{-\mu x}$.
Then, the estimation of failure rate $\mu$ is

$$
\begin{equation*}
\widehat{\mu}=\frac{n+l}{\sum_{h=1}^{n} x_{h}+\sum_{j=1}^{m} c_{j}+\sum_{i=1}^{l} x_{i}^{*}(\mu)} . \tag{8}
\end{equation*}
$$

We can use an iterative method to get $\widehat{\mu}$ with certain precision. There is an algorithm to estimate failure rate from interval data, right censored data and complete data from exponential distribution as follows:

Algorithm 1 MLE of failure rate for the exponential distribution with interval data, right censored data and complete data.

With $n$ complete data $x_{1}, x_{2}, \ldots, x_{n}, m$ right censored data $c_{1}, c_{2}, \ldots, c_{m}$ and $l$ interval data $\left(a_{1}, b_{1}\right),\left(a_{2}, b_{2}\right), \ldots,\left(a_{l}, b_{l}\right)$ from an exponential distribution, the following procedure is executed:
(i) Initialize a value of $\widehat{\mu}$.
(ii) Calculate equivalent failure time $x_{i}^{*}$ for failure interval $\left(a_{i}, b_{i}\right), i=1,2, \ldots, l$,
(iii) Update failure rate

$$
\widehat{\mu^{\prime}}=\frac{n+l}{\sum_{i=1}^{n} x_{i}+\sum_{i=1}^{m} c_{i}+\sum_{i=1}^{l} x_{i}^{*}(\widehat{\mu})}
$$

(iv) If $\widehat{\mu^{\prime}}=\widehat{\mu}$ with certain precision, stop. Otherwise, let $\widehat{\mu}=\widehat{\mu^{\prime}}$ and go to step (ii).

### 3.2 The equivalent method to MLE of GE distribution

It is well known that a random variable $X$ follows the $\operatorname{GE}(\alpha, \lambda)$ distribution with pdf $f(x ; \alpha, \lambda)$, then the random variable $Y=-\log \left(1-e^{-\lambda X}\right)$ follows the exponential distribution with failure rate $\alpha$. From the discussion, we have known that: since the function $-\log \left(1-e^{-\lambda x}\right)$ is a decreasing function, for a given data set data $x_{1}, x_{2}, \ldots, x_{n} ; c_{1}, c_{2}, \ldots, c_{m} ;\left(a_{1}, b_{1}\right),\left(a_{2}, b_{2}\right)$, $\ldots,\left(a_{l}, b_{l}\right)$ that follow $\operatorname{GE}(\alpha, \lambda)$, the complete data are transformed to $-\log \left(1-e^{-\lambda x_{1}}\right),-\log (1-$ $\left.e^{-\lambda x_{2}}\right), \ldots,-\log \left(1-e^{-\lambda x_{n}}\right)$; the right censored data are transformed to $\left(0,-\log \left(1-e^{-\lambda c_{1}}\right)\right)$, $\left(0,-\log \left(1-e^{-\lambda c_{2}}\right)\right), \ldots,\left(0,-\log \left(1-e^{-\lambda c_{m}}\right)\right)$; the interval data are transformed to $(-\log (1-$ $\left.\left.e^{-\lambda b_{1}}\right),-\log \left(1-e^{-\lambda a_{1}}\right)\right),\left(-\log \left(1-e^{-\lambda b_{2}}\right),-\log \left(1-e^{-\lambda a_{2}}\right)\right), \ldots,\left(-\log \left(1-e^{-\lambda b_{l}}\right),-\log (1-\right.$ $\left.e^{-\lambda a_{l}}\right)$ ), if the left end points are not zeros; otherwise, they will be transformed to right censored data $-\log \left(1-e^{-\lambda b_{i}}\right)$ s. The derived data follow an exponential distribution with failure rate $\alpha$. We have known that the derived data follow an exponential distribution, then the estimation failure rate is (8).

We have transformed the GE data to exponential data. Assuming the value of $\lambda$ is known, using (8), we can estimate the exponential failure rata $\alpha$ which is a function of $\lambda$. We know that estimation of failure rate is very fast and always converges by using the equivalent failure and equivalent failure time. Once we have the estimation of $\widehat{\alpha}=\alpha(\lambda)$, substituting $\widehat{\alpha}=\alpha(\lambda)$ into the log likelihood function (4), we can get the MLE for $\lambda$.

Now we state the above procedure as the following algorithm.
Algorithm 2 Two-parameter MLE of the GE distribution with interval data.
With $n$ failure terminated data $x_{1}, x_{2}, \ldots, x_{n}, m$ right censored data $c_{1}, c_{2}, \ldots, c_{m}$ and $l$ interval data $\left(a_{1}, b_{1}\right),\left(a_{2}, b_{2}\right), \ldots,\left(a_{l}, b_{l}\right)$, we present the following algorithm:
(i) Select a value of $\lambda$.
(ii) Transform $x_{1}, x_{2}, \ldots, x_{n} ; c_{1}, c_{2}, \ldots, c_{m} ;\left(a_{1}, b_{1}\right),\left(a_{2}, b_{2}\right), \ldots,\left(a_{l}, b_{l}\right)$ to $-\log \left(1-e^{-\lambda x_{1}}\right)$, $-\log \left(1-e^{-\lambda x_{2}}\right), \ldots,-\log \left(1-e^{-\lambda x_{n}}\right) ;\left(0,-\log \left(1-e^{-\lambda c_{1}}\right)\right),\left(0,-\log \left(1-e^{-\lambda c_{2}}\right)\right), \ldots,(0,-\log (1-$ $\left.\left.e^{-\lambda c_{m}}\right)\right) ;\left(-\log \left(1-e^{-\lambda b_{1}}\right),-\log \left(1-e^{-\lambda a_{1}}\right)\right),\left(-\log \left(1-e^{-\lambda b_{2}}\right),-\log \left(1-e^{-\lambda a_{2}}\right)\right), \ldots,(-\log (1-$ $\left.\left.e^{-\lambda b_{l}}\right),-\log \left(1-e^{-\lambda a_{l}}\right)\right)$, or right censored data $-\log \left(1-e^{-\lambda b_{i}}\right) \mathrm{s}$.
(iii) Estimate failure rate $\alpha$ from the derived data $-\log \left(1-e^{-\lambda x_{1}}\right),-\log \left(1-e^{-\lambda x_{2}}\right), \ldots$, $-\log \left(1-e^{-\lambda x_{n}}\right) ;\left(0,-\log \left(1-e^{-\lambda c_{1}}\right)\right),\left(0,-\log \left(1-e^{-\lambda c_{2}}\right)\right), \ldots,\left(0,-\log \left(1-e^{-\lambda c_{l}}\right)\right) ;(-\log (1-$ $\left.\left.e^{-\lambda b_{1}}\right),-\log \left(1-e^{-\lambda a_{1}}\right)\right),\left(-\log \left(1-e^{-\lambda b_{2}}\right),-\log \left(1-e^{-\lambda a_{2}}\right)\right), \ldots,\left(-\log \left(1-e^{-\lambda b_{l}}\right),-\log (1-\right.$ $\left.e^{-\lambda a_{l}}\right)$ ), or right censored data $-\log \left(1-e^{-\lambda b_{i}}\right) \mathrm{s}$, using Algorithm 1.
(iv) Substitute $\alpha(\lambda)$ into (4), and calculate the log likelihood function.
(v) Repeat steps (i)-(iv) until the maximum of (iv) is reached, output $\lambda$ and corresponding $\alpha$.

## 4. Simulation study

In this section, to see the overall performance of the method, we perform some numerical experiments to demonstrate the method for several parameter values in different sample size. As a reference, we apply the traditional EM algorithm to estimate the parameters. We will see the estimation accuracy of the methods in different situations. All the computations are implemented by using MATLAB.

### 4.1 Generate interval data sets

We set sample sizes as small $n=50$, moderate $n=100$, large $n=200$ and $\lambda=0.5$. Regarding the shape parameter $\alpha$, we choose six typical values as $\alpha=1,2$ and 3. Particularly, $\alpha=1$, represent the exponential distribution.

In this study, we consider three types of data: complete data, interval data and right censored data. Let $p$ be the percentage of complete data in a data set. Then $1-p$ is the percentage of interval data and right censored data in a data set. To check the impact of the uncertainty on estimation accuracy, we select 4 values of $p$ as $0,0.3,0.7$ and 1 . In the 0 case, there are no complete data in a data set. In the 0.3 and 0.7 case, the percentages of complete data are 0.3 and 0.7. In the 1 case, there are all complete data.

With the combination of different settings for $\alpha$, in every case we generate a data set using the following steps:
(i) Generate an observation $U$ from the uniform distribution $U(0,1)$.
(ii) Calculate $X=-\frac{1}{\lambda} \log \left(1-U^{\frac{1}{\alpha}}\right)$. Then $X$ is a sample from the GE distribution $\operatorname{GE}(\alpha, \lambda)$.
(iii) Set the intervals as $[0,1),[1,2),[2,3),[3,4),[4,5),[5,6),[6,7),[7,8),[8, \infty)$. Generate a $U(0,1)$ observation $V$ independent of $U$. If $v<p, X$ is classified as a complete sample. If $v \geq p$ and $X$ is located in a finite interval, $X$ is classified as an interval sample, otherwise, it is a right censored sample.
(iv) Repeat steps (i)-(iii) until the sample size is reached.

### 4.2 Parameter estimations and discussions

In each case we replicate the calculation 1000 times and get the average estimations of $\alpha$ and $\lambda$. The setting ML is calculated from (4) with the setting parameter value. All of the results are listed in Table 1. When the sample size is 200 and $p=0$, we provide the histograms of the simulation data and the fitted densities functions in Figures 1 using the equivalent method.

|  |  |  |  | equivalent method |  |  | EM algorithm |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| size | $\alpha$ | $p$ | setting ML | ML | $\hat{\alpha}$ | $\hat{\lambda}$ | $\hat{\lambda}$ | $\hat{\alpha}$ | ML |
| $n=50$ | 1.0 | $p=0.0$ | -83.3963 | -82.4419 | 1.0825 | 0.5371 | 0.5236 | 1.0633 | -81.9377 |
|  |  | $p=0.3$ | -84.1432 | -83.1623 | 1.0731 | 0.5283 | 0.5239 | 1.0784 | -83.8403 |
|  |  | $p=0.7$ | -84.5979 | -83.6294 | 1.0692 | 0.5254 | 0.5230 | 1.0634 | -83.6852 |
|  |  | $p=1.0$ | -84.9173 | -83.9374 | 1.0585 | 0.5222 | 0.5213 | 1.0555 | -83.9877 |
|  | 2.0 | $p=0.0$ | -98.2168 | -97.3185 | 2.0324 | 0.5114 | 0.5216 | 2.1750 | -98.7875 |
|  |  | $p=0.3$ | -98.6579 | -97.7197 | 2.1023 | 0.5184 | 0.5288 | 2.2067 | -98.4807 |
|  |  | $p=0.7$ | -99.4924 | -98.4835 | 2.1307 | 0.5191 | 0.5178 | 2.1552 | -99.1984 |
|  |  | $p=1.0$ | -100.3204 | -99.3063 | 2.1362 | 0.5159 | 0.5166 | 2.1384 | -99.2330 |
|  | 3.0 | $p=0.0$ | -101.4066 | -100.4910 | 3.0883 | 0.5157 | 0.5177 | 3.2867 | -103.6868 |
|  |  | $p=0.3$ | -102.2739 | -101.4027 | 3.1347 | 0.5162 | 0.5183 | 3.2595 | -103.6275 |
|  |  | $p=0.7$ | -103.6778 | -102.6377 | 3.2145 | 0.5196 | 0.5136 | 3.2127 | -103.9751 |
|  |  | $p=1.0$ | -105.0988 | -104.1014 | 3.2416 | 0.5134 | 0.5153 | 3.2270 | -103.9149 |
| $n=80$ | 1.0 | $p=0.0$ | -167.0553 | -166.1567 | 1.0313 | 0.5209 | 0.5137 | 1.0406 | -167.0422 |
|  |  | $p=0.3$ | -167.6069 | -166.7014 | 1.0298 | 0.5178 | 0.5159 | 1.0400 | -167.3922 |
|  |  | $p=0.7$ | -169.1165 | -168.1189 | 1.0280 | 0.5110 | 0.5110 | 1.0310 | -168.4670 |
|  |  | $p=1.0$ | -169.3164 | -168.2284 | 1.0350 | 0.5151 | 0.5117 | 1.0234 | -168.1549 |
|  | 2.0 | $p=0.0$ | -195.8735 | -194.9820 | 2.0031 | 0.5118 | 0.5116 | 2.0727 | -198.4385 |
|  |  | $p=0.3$ | -196.9629 | -195.9912 | 2.0252 | 0.5114 | 0.5116 | 2.0842 | -198.6556 |
|  |  | $p=0.7$ | -198.4974 | -197.4922 | 2.0564 | 0.5118 | 0.5094 | 2.0706 | -199.0234 |
|  |  | $p=1.0$ | -199.3088 | -198.3580 | 2.0754 | 0.5132 | 0.5079 | 2.0693 | -199.3006 |
|  | 3.0 | $p=0.0$ | -202.6538 | -201.8891 | 2.9306 | 0.5063 | 0.5120 | 3.1863 | -208.0535 |
|  |  | $p=0.3$ | -205.0362 | -204.1488 | 2.9542 | 0.5039 | 0.5098 | 3.1190 | -208.1668 |
|  |  | $p=0.7$ | -207.4354 | -206.4655 | 3.0671 | 0.5082 | 0.5115 | 3.1435 | -207.9823 |
|  |  | $p=1.0$ | -209.5895 | -208.5227 | 3.1428 | 0.5093 | 0.5100 | 3.1342 | -208.2657 |
|  | 3.5 | $p=0.0$ | -203.4895 | -202.7987 | 3.4299 | 0.5085 | 0.5121 | 3.7070 | -210.4784 |
|  |  | $p=0.3$ | -206.2178 | -205.3851 | 3.4596 | 0.5073 | 0.5091 | 3.6460 | -210.8267 |
|  |  | $p=0.7$ | -209.2520 | -208.2883 | 3.6125 | 0.5106 | 0.5109 | 3.6957 | -210.7667 |
|  |  | $p=1.0$ | -212.4449 | -211.4314 | 3.6401 | 0.5066 | 0.5075 | 3.6478 | -211.2825 |
| $n=200$ | 1.0 | $p=0.0$ | -334.8853 | -334.0348 | 1.0067 | 0.5105 | 0.5076 | 1.0245 | -336.7433 |
|  |  | $p=0.3$ | -336.3680 | -335.3974 | 1.0027 | 0.5058 | 0.5083 | 1.0158 | -336.1827 |
|  |  | $p=0.7$ | -337.3694 | -336.3915 | 1.0116 | 0.5070 | 0.5059 | 1.0103 | -336.9715 |
|  |  | $p=1.0$ | -338.7647 | -337.8155 | 1.0131 | 0.5054 | 0.5065 | 1.0134 | -337.4768 |
|  | 2.0 | $p=0.0$ | -392.1438 | -391.3173 | 1.9536 | 0.5048 | 0.5066 | 2.0489 | -398.4290 |
|  |  | $p=0.3$ | -394.5352 | -393.6400 | 1.9827 | 0.5051 | 0.5037 | 2.0384 | -399.3595 |
|  |  | $p=0.7$ | -397.7724 | -396.8234 | 2.0116 | 0.5044 | 0.5081 | 2.0491 | -398.0801 |
|  |  | $p=1.0$ | -399.7609 | -398.7501 | 2.0408 | 0.5059 | 0.5059 | 2.0330 | -398.5579 |
|  | 3.0 | $p=0.0$ | -405.5254 | -404.8504 | 2.8751 | 0.5032 | 0.5075 | 3.0923 | -416.8965 |
|  |  | $p=0.3$ | -408.6492 | -407.8302 | 2.9560 | 0.5069 | 0.5040 | 3.0634 | -418.0014 |
|  |  | $p=0.7$ | -414.9834 | -414.0240 | 3.0041 | 0.5038 | 0.5061 | 3.0725 | -417.2950 |
|  |  | $p=1.0$ | -418.7975 | -417.7791 | 3.0661 | 0.5050 | 0.5048 | 3.0668 | -417.8325 |

Table 1 Summary of parameter estimates in simulations


Figure 1 The histograms of simulation data and fitted density functions via equivalent method
From the results of Table 1 and Figure 1, we observe the following:
(i) The histograms of the simulation data and the fitted density functions in Figure 1 indicate that the estimations provide a good fittings for the simulative data sets.
(ii) The simulation results show that the two methods are robust and effective to deal with interval data with respect to $\alpha$ and $\lambda$.
(iii) For fixed $p$, as the sample sizes grow, the accuracy of the estimations of $\alpha$ is improved as expected.
(iv) The estimations of $\lambda$ are accurate in all cases. Certainly, the accuracy is improved as the sample sizes increase. When $n=200$ and $\alpha=3$, the estimations are $0.5032,0.5069,0.5038$, 0.5050 and $0.5075,0.5040,0.5061,0.5048$ corresponding to equivalent method and EM algorithm for 4 values of $p$.
(v) The $\log$ likelihood values are almost equal to the setting ML calculated from the setting parameter values, which indicates that the methods are feasible.
(vi) The estimations of $\alpha$ and $\lambda$ via equivalent method are more accurate than the estimations via EM algorithm when the sample is small or moderate. However, when the sample size is large, the conclusion is opposite.
(vii) In computations, the more loss the information has, the more time the convergence of the EM algorithm will take. We find that the time to converge in $p=0$ case is much longer than that in $p=1$ case. So we recommend to use the equivalent method to obtain the parameter estimations when the sample size is large.

## References

[1] R. D. GUPTA, D. KUNDU. Generalized exponential distributions. Aust. N. Z. J. Stat., 1999, 41(2): 173-188.
[2] R. D. GUPTA, D. KUNDU. Exponentiated exponential family: an alternative to gamma and Weibull. Biom. J., 2001, 43(1): 117-130.
[3] R. D. GUPTA, D. KUNDU. Generalize exponential distributions: statistical inferences. J. Stat. Theory Appl., 2002, 1(2): 101-118.
[4] R. D. GUPTA, D. KUNDU. Discriminating between Weibull and generalized exponential distributions. Comput. Statist. Data Anal., 2003, 43(2): 179-196
[5] R. D. GUPTA, D. KUNDU. Closeness of gamma and generalized exponential distribution. Comm. Statist. Theory Methods, 2003, 32(4): 705-721.
[6] R. D. GUPTA, D. KUNDU. Discriminating between gamma and generalized exponential distributions. J. Stat. Comput. Simul., 2004, 74(2): 107-121.
[7] D. KUNDU, R. D. GUPTA, A. MANGLICk. Discriminating between the log-normal and generalized exponential distribution. J. Statist. Plann. Inference, 2005, 127(1-2): 213-227.
[8] R. D. GUPTA, D. KUNDU. Generalized exponential distributions: different methods of estimation. J. Statist. Comput. Simulation, 2001, 69(4): 315-337.
[9] D. KUNDU, R. D. GUPTA. Estimation of $P(Y<X)$ for generalized exponential distribution. Metrika, 2005, 61(3): 291-308.
[10] D. KUNDU, R. D. GUPTA. Generalized exponential distribution: Bayesian estimations. Comput. Statist. Data Anal., 2008, 52(4): 1873-1883.
[11] A. H. ABDEL-HAMID, E. K. AL-HUSSAINI. Estimation in step-stress accelerated life tests for the exponentiated exponential distribution with type-I censoring. Comput. Statist. Data Anal., 2009, 53(4): 1328-1338.
[12] S. MITRA, D. KUNDU. Analysis of left censored data from the generalized exponential distribution. J. Stat. Comput. Simul., 2008, 78(7-8): 669-679.
[13] B. PRADHAN, D. KUNDU. On progressively censored generalized exponential distribution. TEST, 2009, 18(3): 497-515,
[14] Z. B. TAN. An equivalence based approach to dealing with uncertainty in maximum Likelihood estimation. in: IEEE international conference on information reuse and integration, 2005, 8: 15-17.
[15] Z. B. TAN. A new approach to MLE of Weibull distribution with interval data. Reli. Enge. Syst. Safe., 2009, 94: 394-403.
[16] H. K. T. NG, P. S. CHAN, N. BALAKRISHNAN. Estimation of parameters from progressively censored data using EM algorithm. Comput. Statist. Data Anal., 2002, 39(4): 371-386.


[^0]:    Received December 4, 2010; Accepted February 10, 2012
    Supported by the Natural Science Foundation of Guangdong Province (Grant No. S2012040007369), the Distinguished Young Talents in Higher Education of Guangdong (Grant No. 2012LYM_0089) and the National Natural Science Foundation of China (Grant No. 71171103).
    *Corresponding author
    E-mail address: conghua.cheng@yahoo.com.cn (Conghua CHENG); jinyuan.chen@yahoo.com.cn (Jinyuan CHEN)

