



Received: 04 October 2016  
Accepted: 20 December 2016  
First Published: 28 February 2017

\*Corresponding author: Xiaogang Dong,  
College of Basic Science, Changchun  
University of Technology, Changchun  
130012, China  
E-mail: [dongxiaogang@ccut.edu.cn](mailto:dongxiaogang@ccut.edu.cn)

Reviewing editor:  
Yueqing Hu, Fudan University School of  
Life Sciences, China

Additional information is available at  
the end of the article

## STATISTICS | RESEARCH ARTICLE

# Parameter estimations for mixed generalized exponential distribution based on progressive type-I interval censoring

Chunjie Wang<sup>1</sup>, Shuying Wang<sup>1,2</sup>, Dehui Wang<sup>2</sup>, Chunjing Li<sup>1,2</sup> and Xiaogang Dong<sup>1\*</sup>

**Abstract:** This paper considers the estimation of parameters based on a progressively type-I interval-censored data from a mixed generalized exponential distribution. The maximum likelihood estimation is used but an analytic form cannot be obtained. The EM algorithm is applied to obtain the maximum likelihood estimates. The performance of the estimates is judged by a simulating study and a real data is presented to illustrate the method of estimation developed here.

**Subjects:** Science; Mathematics & Statistics; Statistics & Probability; Statistics; Mathematical Statistics; Statistical Computing

**Keywords:** progressive type-I interval censoring; mixed generalized exponential distribution; maximum likelihood estimation; EM algorithm

### ABOUT THE AUTHORS

Chunjie Wang obtained her PhD degree in Probability and Mathematics Statistics from the School of Mathematics, Jilin University, Changchun, China. She is currently working in Changchun University of Technology (CCUT), Changchun, China, as an associate professor. Her research interests include Survival analysis, Mathematical Statistics.

Shuying Wang is currently doing her PhD from Jilin University, Changchun, China. Her research interests include Mathematical Statistics and Survival analysis.

Dehui Wang is currently working as professor in the School of Mathematics in Jilin University, Changchun, China. His research interests include Mathematical Statistics, Time Series, and Survival Analysis.

Chunjie Li is currently doing her PhD from Jilin University, Changchun, China. Her research interests include Mathematical Statistics and Survival Analysis.

Xiaogang Dong obtained his PhD degree in Statistics from Jilin University, Changchun, China. He is currently working in Changchun University of Technology (CCUT), Changchun, China, as a professor. His research interests include High frequency data analysis, Survival analysis, Mathematical Statistics, and Applied Statistics.

### PUBLIC INTEREST STATEMENT

Progressive type-I interval censoring data in survival analysis and reliability analysis are a kind of interested and popular data type. This work is modeling for this kind of data with a more flexible Mixed Generalized Exponential Distribution including the exponential distribution and Weibull distribution. And the result of this paper can apply to the lifetime analysis in clinical trial, engineering and so on.

## 1. Introduction

In the life testing, reliability studies and survival analysis, it is extremely common that items are lost or removed from experiments before failure. The most popular censoring schemes are type-I and type-II censoring. Under type-I censoring, the test is continued until a pre-specified time. In type-II censoring, the test is continued up to a pre-specified number of failures. Progressive type-II censoring is an extension of the traditional type-II censoring scheme and it has been studied such as Balakrishnan and Aggarwala (2000), Balakrishnan, Kannan, Lin, and Wu (2004) and so on. Sometimes it is impossible to observe the whole life test process continuously due to time and cost constraint. Only the number of failures is observed in an interval instead of observing the failure time exactly and this is called interval censoring. As the mixture of the interval censoring and progressive censoring, Aggarwala (2001) introduced progressive type-I interval censoring scheme and developed the statistical inference for the exponential distribution based on progressively type-I interval-censored data. Ng and Wang (2009) discussed statistical inference for Weibull distribution under progressive type-I interval censoring. They have compared with different estimation methods for the parameters of Weibull distribution via simulation. Lin, Wu, and Balakrishnan (2009) considered determination of optimum life testing plan with progressively type-I interval-censored data from the log-normal distribution. Chen and Lio (2010) considered parameter estimations for generalized exponential distribution under progressive type-I interval censoring. They have obtained the estimates of unknown parameters using EM algorithm, midpoint approximation method and method of moments. Further, Lio, Chen, and Tsai (2011) presented the parameter estimations of the generalized Rayleigh distribution at the same censoring scheme. Peng and Yan (2011) considered the MLEs and moments estimators of the parameters for gamma distribution based on progressive type-I interval censoring and compared with the biases and mean square errors through simulation. Pradhan and Gijo (2013) considered inference for the unknown parameters of log-normal distribution based on progressively type-I interval-censored data through both frequency and Bayesian approaches.

In most simple life tests, the experimental data come from one population and have only one type of failure. In fact, a mixed distribution can provide a flexible candidate to depict the time to failure. Especially in the analysis of biased life data, the failure hazard in the beginning is quite high and the failure rate is decreased or constant as the age increases. On this occasion, population is not homogeneous but is made up from sub-populations. There are many researches for the mixed distribution, such as McClean (1986), Soliman (2006), Tian, He, and Chen (2008), Tian, Tian and Chen (2012), Tian, Tian, and Zhu (2014), Tian, Zhu and Tian (2013) and so on. Nevertheless, to our best knowledge, no published papers address maximum likelihood estimation of the mixed generalized exponential distribution (MGED) under progressive type-I interval censoring. In this work, we consider the estimation for parameters of the MGED based on progressively type-I interval-censored data. This paper gives the likelihood equations of the unknown parameters. It is observed that there is no closed form of MLEs. Therefore, it is suggested to use the EM algorithm to compute the MLEs and present the performance of the estimates by simulations.

The rest of the paper is organized as follows. The model and data description are provided in Section 2. The maximum likelihood estimators of the unknown parameters are obtained in Section 3. The EM algorithm procedure is given in Section 4. The performance of the estimators is investigated via simulation in Section 5. A data-set is analyzed in Section 6. Section 7 includes some conclusion remarks.

## 2. The description of the model and data

Consider the MGED with  $m$  components with its probability density function, cumulative distribution function and hazard function as follows

$$f(t; P, \alpha, \lambda) = \sum_{k=1}^m P_k \alpha_k \lambda_k (1 - e^{-\lambda_k t})^{\alpha_k - 1} e^{-\lambda_k t}, \quad t > 0 \quad (2.1)$$

$$F(t;P, \alpha, \lambda) = \sum_{k=1}^m p_k (1 - e^{-\lambda_k t})^{\alpha_k}, \quad t \geq 0 \tag{2.2}$$

$$h(t;P, \alpha, \lambda) = \frac{\sum_{k=1}^m p_k \alpha_k \lambda_k (1 - e^{-\lambda_k t})^{\alpha_k - 1} e^{-\lambda_k t}}{1 - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t})^{\alpha_k}}, \quad t \geq 0 \tag{2.3}$$

where  $P = (p_1, p_2, \dots, p_{m-1})$ ,  $\alpha = (\alpha_1, \dots, \alpha_m)$ ,  $\lambda = (\lambda_1, \dots, \lambda_m)$ ,  $0 < p_k < 1$ ,  $k = 1, \dots, m-1$ ,  $p_m = 1 - \sum_{k=1}^{m-1} p_k$ ,  $\alpha_k > 0$  are the shape parameters,  $\lambda_k > 0$  are the scale parameters, the number of the parameters is  $3m - 1$ .

Then, progressive type-I interval censoring scheme is as follows. Let  $n$  identical items be placed on a test at time  $t_0 = 0$  and pre-specify  $N$  inspection times  $t_1 < t_2 < \dots < t_N$ , where  $t_N$  is the scheduled termination time of the experiment. Suppose  $X_i$  is the failure number within the interval of  $(t_{i-1}, t_i]$ ,  $Y_i$  is the survival items numbers at time  $t_i$  and  $R_i$  is the number of items randomly removed from the surviving items at time  $t_i$ ,  $R_1, R_2, \dots, R_N$  can be pre-specified by the percentage of the remaining surviving items. Therefore, for the pre-specified percentage  $p'_i$ ,  $R_i = \lfloor Y_i p'_i \rfloor$ ,  $i = 1, 2, \dots, N$ ,  $p'_N = 1$ . Alternatively  $R_1, R_2, \dots, R_m$  can also be pre-specified positive integers. In this case, there must be at least  $R_i$  units available for removal. The proportions of surviving units to be removed at the monitoring and censoring points have been given in the simulation. And the progressively type-I interval-censored data are given by  $\{X_i, R_i, t_i\}$ ,  $i = 1, 2, \dots, N$ , where the sample size  $n = \sum_{i=1}^N (X_i + R_i)$ . If  $R_i = 0$ ,  $i = 1, 2, \dots, N - 1$ , then progressive type-I interval censoring is reduced to the traditional interval-censored data,  $X_1, X_2, \dots, X_N, X_{N+1} = R_N$ .

### 3. Maximum likelihood estimation

Given a progressively type-I interval-censored data,  $\{X_i, R_i, t_i\}$ ,  $i = 1, 2, \dots, N$  which sample size is  $n$ , from a lifetime distribution with distribution function  $F(T, \theta)$  and  $\theta$  is the parameter vector. Then, the likelihood function is (Aggarwala, 2001)

$$\begin{aligned} L(\theta) &\propto [F(t_1, \theta)]^{X_1} [1 - F(t_1, \theta)]^{R_1} \times [F(t_2, \theta) - F(t_1, \theta)]^{X_2} [1 - F(t_2, \theta)]^{R_2} \\ &\times \dots \times [F(t_N, \theta) - F(t_{N-1}, \theta)]^{X_N} [1 - F(t_N, \theta)]^{R_N} \\ &= \prod_{i=1}^N [F(t_i, \theta) - F(t_{i-1}, \theta)]^{X_i} [1 - F(t_i, \theta)]^{R_i} \end{aligned} \tag{3.1}$$

where  $t_0 = 0$ . We can obtain the MLEs of the parameters by maximizing the likelihood function in (3.1).

Given the MGED by the Equation (2.2), the above likelihood function can be specified as follows:

$$L(P, \alpha, \lambda) \propto \prod_{i=1}^N \left[ \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k} - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_{i-1}})^{\alpha_k} \right]^{X_i} \left[ 1 - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k} \right]^{R_i}$$

And the log-likelihood function is

$$\begin{aligned} l(P, \alpha, \lambda) = \ln L(P, \alpha, \lambda) &= \text{constant} + \sum_{i=1}^N \left\{ X_i \ln \left[ \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k} \right. \right. \\ &\left. \left. - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_{i-1}})^{\alpha_k} \right] + R_i \ln \left[ 1 - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k} \right] \right\} \end{aligned}$$

Taking derivatives of the log-likelihood function with respect to  $p_k, \alpha_k, \lambda_k$  to zero, the following score equations are obtained:

$$\sum_{i=1}^N \frac{X_i [(1 - e^{-\lambda_k t_i})^{\alpha_k} - (1 - e^{-\lambda_k t_{i-1}})^{\alpha_k}]}{\left[ \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k} - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_{i-1}})^{\alpha_k} \right]} - \sum_{i=1}^N \frac{R_i (1 - e^{-\lambda_k t_i})^{\alpha_k}}{1 - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k}} = 0$$

$$\sum_{i=1}^N \left\{ \frac{X_i [p_k (1 - e^{-\lambda_k t_i})^{\alpha_k} \ln(1 - e^{-\lambda_k t_i}) - p_k (1 - e^{-\lambda_k t_{i-1}})^{\alpha_k} \ln(1 - e^{-\lambda_k t_{i-1}})]}{\left[ \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k} - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_{i-1}})^{\alpha_k} \right]} \right\}$$

$$- \sum_{i=1}^N \left[ \frac{R_i p_k (1 - e^{-\lambda_k t_i})^{\alpha_k} \ln(1 - e^{-\lambda_k t_i})}{1 - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k}} \right] = 0$$

$$\sum_{i=1}^N \left\{ \frac{X_i [p_k \alpha_k t_i (1 - e^{-\lambda_k t_i})^{\alpha_k - 1} e^{-\lambda_k t_i} - p_k \alpha_k t_{i-1} (1 - e^{-\lambda_k t_{i-1}})^{\alpha_k - 1} e^{-\lambda_k t_{i-1}}]}{\sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k} - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_{i-1}})^{\alpha_k}} \right\}$$

$$- \sum_{i=1}^N \left[ \frac{R_i p_k \alpha_k t_i (1 - e^{-\lambda_k t_i})^{\alpha_k - 1} e^{-\lambda_k t_i}}{1 - \sum_{k=1}^m p_k (1 - e^{-\lambda_k t_i})^{\alpha_k}} \right] = 0$$

Obviously, the equations are quite complex and there is no closed form for the solutions to the above equations and the EM algorithm introduced as follows to find the MLEs of  $p_k, \alpha_k, \lambda_k, k = 1, 2, \dots, m$ .

#### 4. EM algorithm

The Expectation Maximization (EM) Algorithm (Dempster, Laird, & Rubin, 1977) is a useful tool to estimate the parameters of the distribution based on an incomplete data. Here, some notations are similar to that in the Tian et al. (2014). In order to estimate the MGED models under progressive type-I interval censoring, EM algorithm is applied to estimate the unknown parameters. Suppose  $\tau_1, \tau_2, \dots, \tau_n$  are  $n$  independent identically distributed (iid) samples from the MGED model and denote

$$f_{ki} = \alpha_k \lambda_k (1 - e^{-\lambda_k \tau_i})^{\alpha_k - 1} e^{-\lambda_k \tau_i} \quad f_i = \sum_{k=1}^m p_k f_{ki}$$

$$S_{ki} = 1 - (1 - e^{-\lambda_k \tau_i})^{\alpha_k} \quad S_i = \sum_{k=1}^m p_k S_{ki}$$

where  $k = 1, 2, \dots, m, i = 1, 2, \dots, n$ .

Here, the sample  $\tau_p, i = 1, 2, \dots, n$  is divided into two components that consist of  $\tau_{ij}$  and  $\tau_{ij}^*$ . Let  $\tau_{ij}, j = 1, 2, \dots, X_i$  be the lifetimes within interval  $(t_{i-1}, t_i]$  and  $\tau_{ij}^*, j = 1, 2, \dots, R_i$  be the lifetimes for those withdrawn items at  $t_i$  for  $i = 1, 2, \dots, N$  and the number of the  $\tau_{ij}$  and  $\tau_{ij}^*$  are  $\sum_{i=1}^N X_i$  and  $\sum_{i=1}^N R_i$ , respectively.

Introduce an indicator vector  $I_{ij} = (I_{ij1}, I_{ij2}, \dots, I_{ijm})$  of  $\tau_{ij}$ ,  $I_{ij'} = (I_{ij'1}, I_{ij'2}, \dots, I_{ij'm})$  of  $\tau_{ij'}$ ,  $I_{ijk}^*(I_{ij'k}^*)$  is a binary random variable only taking the value 1 if  $\tau_{ij}(\tau_{ij'})$  comes the  $k$ -th component, and 0 otherwise for  $k = 1, 2, \dots, m$ . Also denote  $I = (I_{ij}, I_{ij'}), i = 1, 2, \dots, N, j = 1, 2, \dots, X_i, j' = 1, 2, \dots, R_i$  as an indicator matrix composed of  $n$  indicator vectors of all life variables  $\tau_{ij}$  and  $\tau_{ij'}$ . Evidently random vector  $I_{ij} = (I_{ij1}, I_{ij2}, \dots, I_{ijm})$  and  $I_{ij'} = (I_{ij'1}, I_{ij'2}, \dots, I_{ij'm})$  follows the multinomial distribution. However, we may not know which component the variate comes from. Namely,  $I$  cannot be observed, thus regard it as the missing data in the EM algorithm. In the following paper, denote  $I_{ij}^{(1)} = (I_{ij1}^{(1)}, I_{ij2}^{(1)}, \dots, I_{ijm}^{(1)})$ ,  $I_{ij'}^{(2)} = (I_{ij'1}^{(2)}, I_{ij'2}^{(2)}, \dots, I_{ij'm}^{(2)})$  as the indicator vectors of the data  $\tau_{ij}$  and  $\tau_{ij'}$ , respectively.

As to the the lifetimes  $\tau_{ij}$  within interval  $(t_{i-1}, t_i]$ , the joint density of  $\tau_{ij}$  and  $I_{ij}^{(1)}$  is  $g(\tau_{ij}, I_{ij}^{(1)} | P, \alpha, \lambda) = \prod_{k=1}^m [p_k(S_{ki-1} - S_{ki})]^{I_{ijk}^{(1)}}$ . Given  $\tau_{ij}$ , the conditional density of  $I_{ij}^{(1)}$  is  $p(I_{ijk}^{(1)} = 1 | \tau_{ij}, P, \alpha, \lambda) = \frac{p_k(S_{ki-1} - S_{ki})}{(S_{i-1} - S_i)}, k = 1, 2, \dots, m$ .

$$\text{Where, } S_{ki} = 1 - (1 - e^{-\lambda_k t_i})^{\alpha_k}, S_{ki-1} = 1 - (1 - e^{-\lambda_k t_{i-1}})^{\alpha_k}, S_i = \sum_{k=1}^m p_k S_{ki}, S_{i-1} = \sum_{k=1}^m p_k S_{ki-1}$$

For the right-censored data  $\tau_{ij'}$ , the joint density of  $\tau_{ij'}$  and  $I_{ij'}^{(2)}$  is given by  $g(\tau_{ij'}, I_{ij'}^{(2)} | P, \alpha, \lambda) = \prod_{k=1}^m [p_k S_{ki}]^{I_{ij'k}^{(2)}}$ . Given  $\tau_{ij'}$ , the conditional density of  $I_{ij'}^{(2)}$  is  $p(I_{ij'k}^{(2)} = 1 | \tau_{ij'}, P, \alpha, \lambda) = \frac{p_k S_{ki}}{S_i}, k = 1, 2, \dots, m$ . Where  $S_{ki} = 1 - (1 - e^{-\lambda_k t_i})^{\alpha_k}, S_i = \sum_{k=1}^m p_k S_{ki}$ .

In Section 2, denote the progressively type-I interval-censored sample as  $\{X_i, R_i, t_i\}, i = 1, 2, \dots, N$ . In the progressive type-I interval censoring experiments, we can only observe the failure numbers  $X_i$  within the intervals  $(t_{i-1}, t_i]$  and  $R_i$ , the number of the censored items withdrawn at the censoring time  $t_i$  for  $i = 1, 2, \dots, N$ . Then, the observed values can be simply denoted as  $Y = (t_1, t_2, \dots, t_N, X_1, X_2, \dots, X_N, R_1, R_2, \dots, R_N)$ . However, the true failure time within the interval  $(t_{i-1}, t_i]$  denoted as  $\tau_{ij}, i = 1, 2, \dots, N, j = 1, 2, \dots, X_i$  and the true failure time of censored units at  $t_i$  denoted as  $\tau_{ij'}, i = 1, 2, \dots, N, j' = 1, 2, \dots, R_i$  can not be observed in life experiments. Consequently,  $\tau = (\tau_{ij}, \tau_{ij'}; i = 1, 2, \dots, N, j = 1, 2, \dots, X_i, j' = 1, 2, \dots, R_i)$  can be regarded as missing data. All the missing data can be denoted  $(\tau, I)$  and all the complete data can be denoted as  $W = (\tau, I, Y)$ . The following procedure will give the MLEs of all unknown parameters via the EM algorithm consisting of two steps: E-step and M-step.

First of all, the likelihood function of the MGED model under the complete data  $W$  is given by

$$l(P, \alpha, \lambda | W) = \prod_{i=1}^N \left\{ \prod_{j=1}^{X_i} \prod_{k=1}^m [p_k \alpha_k \lambda_k (1 - e^{-\lambda_k \tau_{ij}})^{\alpha_k - 1} e^{-\lambda_k \tau_{ij}}]^{I_{ijk}^{(1)}} \times \prod_{j'=1}^{R_i} \prod_{k=1}^m [p_k \alpha_k \lambda_k (1 - e^{-\lambda_k \tau_{ij'}})^{\alpha_k - 1} e^{-\lambda_k \tau_{ij'}}]^{I_{ij'k}^{(2)}} \right\}$$

The log-likelihood function of the complete data is

$$\ln l(P, \alpha, \lambda | W) = \sum_{i=1}^N \sum_{k=1}^m \left\{ \sum_{j=1}^{X_i} I_{ijk}^{(1)} \left[ \ln(p_k \alpha_k \lambda_k) + (\alpha_k - 1) \ln(1 - e^{-\lambda_k \tau_{ij}}) - \lambda_k \tau_{ij} \right] + \sum_{j'=1}^{R_i} I_{ij'k}^{(2)} \left[ \ln(p_k \alpha_k \lambda_k) + (\alpha_k - 1) \ln(1 - e^{-\lambda_k \tau_{ij'}}) - \lambda_k \tau_{ij'} \right] \right\}$$

Given initial values  $P^{(0)}, \alpha^{(0)}, \lambda^{(0)}$  of the unknown parameter vectors, we can obtain parameter estimates of model (2.1) based on EM algorithm via the following two steps. Of course, estimation performance has a great relationship with the choice of the initial values. Generally, different initial values may be lead to different convergence rate. In the simulation, it is suggest to choose some groups of initial values to compare the estimation results.

E-step: suppose the  $(h - 1)$ -th iteration values are  $P^{(h-1)}$ ,  $\alpha^{(h-1)}$ ,  $\lambda^{(h-1)}$ , then the  $Q$  function of the  $h$ -th iteration is obtained by

$$\begin{aligned}
 Q(P, \alpha, \lambda | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, Y) &= E \left[ \ln l(P, \alpha, \lambda | W) | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, Y \right] = \sum_{i=1}^N \\
 &\sum_{k=1}^m E \left\{ \sum_{j=1}^{X_i} I_{ijk}^{(1)} \left[ \ln(p_k \alpha_k \lambda_k) + (\alpha_k - 1) \ln(1 - e^{-\lambda_k \tau_{ij}}) - \lambda_k \tau_{ij} \right] \right. \\
 &\left. + \sum_{j=1}^{R_i} I_{ij'k}^{(2)} \left[ \ln(p_k \alpha_k \lambda_k) + (\alpha_k - 1) \ln(1 - e^{-\lambda_k \tau_{ij}^*}) - \lambda_k \tau_{ij}^* \right] \right\} \\
 &= \sum_{i=1}^N \sum_{k=1}^m \left\{ \sum_{j=1}^{X_i} \ln(p_k \alpha_k \lambda_k) E(I_{ijk}^{(1)}) + (\alpha_k - 1) \sum_{j=1}^{X_i} E \left[ \ln(1 - e^{-\lambda_k \tau_{ij}}) I_{ijk}^{(1)} \right] \right. \\
 &\left. - \lambda_k \sum_{j=1}^{X_i} E(\tau_{ij} \cdot I_{ijk}^{(1)}) + \sum_{j=1}^{R_i} \ln(p_k \alpha_k \lambda_k) E(I_{ij'k}^{(2)}) + (\alpha_k - 1) \sum_{j=1}^{R_i} E \left[ \ln(1 - e^{-\lambda_k \tau_{ij}^*}) I_{ij'k}^{(2)} \right] - \lambda_k \sum_{j=1}^{R_i} E(\tau_{ij}^* \cdot I_{ij'k}^{(2)}) \right\} \\
 &= \sum_{i=1}^N \sum_{k=1}^m \left\{ \sum_{j=1}^{X_i} \ln(p_k \alpha_k \lambda_k) E[E(I_{ijk}^{(1)} | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, \tau, Y)] + (\alpha_k - 1) \right. \\
 &\sum_{j=1}^{X_i} E \left[ E(\ln(1 - e^{-\lambda_k \tau_{ij}}) I_{ijk}^{(1)} | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, \tau, Y) \right] \\
 &- \lambda_k \sum_{j=1}^{X_i} E[E(\tau_{ij} \cdot I_{ijk}^{(1)} | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, \tau, Y)] + \sum_{j=1}^{R_i} \ln(p_k \alpha_k \lambda_k) E \left[ E(I_{ij'k}^{(2)} | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, \tau, Y) \right] \\
 &\left. + (\alpha_k - 1) \sum_{j=1}^{R_i} E \left[ E(\ln(1 - e^{-\lambda_k \tau_{ij}^*}) I_{ij'k}^{(2)} | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, \tau, Y) \right] \right. \\
 &\left. - \lambda_k \sum_{j=1}^{R_i} E \left[ E(\tau_{ij}^* \cdot I_{ij'k}^{(2)} | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, \tau, Y) \right] \right\} = \sum_{i=1}^N \sum_{k=1}^m \left\{ \sum_{j=1}^{X_i} \ln(p_k \alpha_k \lambda_k) E \left[ a_{kij}^{(h-1)}(\tau_{ij}) \right] + (\alpha_k - 1) \sum_{j=1}^{X_i} E \left[ \ln(1 - e^{-\lambda_k \tau_{ij}}) a_{kij}^{(h-1)}(\tau_{ij}) \right] \right. \\
 &\left. - \lambda_k \sum_{j=1}^{X_i} E \left[ \tau_{ij} \cdot a_{kij}^{(h-1)}(\tau_{ij}) \right] + \sum_{j=1}^{R_i} \ln(p_k \alpha_k \lambda_k) E \left[ b_{kij'}^{(h-1)}(\tau_{ij}^*) \right] + (\alpha_k - 1) \sum_{j=1}^{R_i} E \left[ \ln(1 - e^{-\lambda_k \tau_{ij}^*}) b_{kij'}^{(h-1)}(\tau_{ij}^*) \right] - \lambda_k \sum_{j=1}^{R_i} E \left[ \tau_{ij}^* \cdot b_{kij'}^{(h-1)}(\tau_{ij}^*) \right] \right\} = \sum_{i=1}^N \\
 &\sum_{k=1}^m \left\{ \sum_{j=1}^{X_i} \ln(p_k \alpha_k \lambda_k) a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1}) + (\alpha_k - 1) \sum_{j=1}^{X_i} E \left[ \ln(1 - e^{-\lambda_k \tau_{ij}}) a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1}) \right] - \lambda_k \sum_{j=1}^{X_i} E \left[ \tau_{ij} \cdot a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1}) \right] \right. \\
 &\left. + R_i \cdot \ln(p_k \alpha_k \lambda_k) \cdot b_{ki}^{(h-1)}(t_i) + (\alpha_k - 1) \sum_{j=1}^{R_i} E \left[ \ln(1 - e^{-\lambda_k \tau_{ij}^*}) b_{ki}^{(h-1)}(t_i) \right] - \lambda_k \sum_{j=1}^{R_i} E \left[ \tau_{ij}^* \cdot b_{ki}^{(h-1)}(t_i) \right] \right\}
 \end{aligned}$$

where,

$$a_{kij}^{(h-1)}(\tau_{ij}) = \frac{p_k^{(h-1)} (S_{ki-1}^{(h-1)}(t_{i-1}) - S_{ki}^{(h-1)}(t_i))}{(S_{i-1}^{(h-1)}(t_{i-1}) - S_i^{(h-1)}(t_i))} \hat{=} a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1})$$

$$b_{kij'}^{(h-1)}(\tau_{ij}^*) = \frac{p_k^{(h-1)} \cdot S_{ki}^{(h-1)}(t_i)}{S_i^{(h-1)}(t_i)} \hat{=} b_{ki}^{(h-1)}(t_i)$$

$$S_{ki-1}^{(h-1)}(t_{i-1}) = 1 - (1 - e^{-\lambda_k^{(h-1)} t_{i-1}})^{\alpha_k^{(h-1)}}$$

$$S_{ki}^{(h-1)}(t_i) = 1 - (1 - e^{-\lambda_k^{(h-1)} t_i})^{\alpha_k^{(h-1)}}$$

$$S_{i-1}^{(h-1)}(t_{i-1}) = \sum_{k=1}^m p_k^{(h-1)} S_{ki-1}^{(h-1)}(t_{i-1})$$

$$S_i^{(h-1)}(t_i) = \sum_{k=1}^m p_k^{(h-1)} S_{ki}^{(h-1)}(t_i)$$

$$k = 1, 2, \dots, m, i = 1, 2, \dots, N, j = 1, 2, \dots, X_j, j' = 1, 2, \dots, R_j$$

Then,

$$\begin{aligned} Q(P, \alpha, \lambda | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, Y) &= \sum_{i=1}^N \sum_{k=1}^m \left[ \sum_{j=1}^{X_i} \ln(p_k \alpha_k \lambda_k) a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1}) \right. \\ &+ (\alpha_k - 1) \sum_{j=1}^{X_i} \int_{t_{i-1}}^{t_i} \ln(1 - e^{-\lambda_k x}) a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1}) p_{ij}(x) dx - \lambda_k \sum_{j=1}^{X_i} \int_{t_{i-1}}^{t_i} x \cdot a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1}) p_{ij}(x) dx + R_i \cdot \ln(p_k \alpha_k \lambda_k) b_{ki}^{(h-1)}(t_i) \\ &+ (\alpha_k - 1) \sum_{j=1}^{R_i} \int_{t_i}^{\infty} \ln(1 - e^{-\lambda_k x}) b_{ki}^{(h-1)}(t_i) p_{ij'}(x) dx - \lambda_k \sum_{j=1}^{R_i} \int_{t_i}^{\infty} x \cdot b_{ki}^{(h-1)}(t_i) p_{ij'}(x) dx \left. \right] \\ &= \sum_{i=1}^N \sum_{k=1}^m \left[ X_i \ln(p_k \alpha_k \lambda_k) \Delta 1_{kij}^{(h-1)} + (\alpha_k - 1) X_i \Delta 2_{kij}^{(h-1)} - \lambda_k X_i \Delta 3_{kij}^{(h-1)} + R_i \cdot \ln(p_k \alpha_k \lambda_k) \Delta 4_{ki}^{(h-1)} \right. \\ &+ (\alpha_k - 1) R_i \Delta 5_{kij'}^{(h-1)} - \lambda_k R_i \Delta 6_{kij'}^{(h-1)} \left. \right] \\ &= \sum_{i=1}^N \sum_{k=1}^m \left[ (X_i \Delta 1_{kij}^{(h-1)} + R_i \Delta 4_{ki}^{(h-1)}) \ln(p_k \alpha_k \lambda_k) + (\alpha_k - 1) (X_i \Delta 2_{kij}^{(h-1)} + R_i \Delta 5_{kij'}^{(h-1)}) - \lambda_k (X_i \Delta 3_{kij}^{(h-1)} + R_i \Delta 6_{kij'}^{(h-1)}) \right] \end{aligned}$$

where

$$\Delta 1_{k,i,j-1}^{(h-1)} = a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1}), \quad \Delta 2_{kij}^{(h-1)} = \int_{t_{i-1}}^{t_i} \ln(1 - e^{-\lambda_k x}) a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1}) p_{ij}(x) dx$$

$$\Delta 3_{kij}^{(h-1)} = \int_{t_{i-1}}^{t_i} x a_{k,i,j-1}^{(h-1)}(t_i, t_{i-1}) p_{ij}(x) dx, \quad \Delta 4_{ki}^{(h-1)} = b_{ki}^{(h-1)}(t_i)$$

$$\Delta 5_{kij'}^{(h-1)} = \int_{t_i}^{\infty} \ln(1 - e^{-\lambda_k x}) b_{ki}^{(h-1)}(t_i) p_{ij'}(x) dx, \quad \Delta 6_{kij'}^{(h-1)} = \int_{t_i}^{\infty} x b_{ki}^{(h-1)}(t_i) p_{ij'}(x) dx$$

$$p_{ij}(x | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, Y) = \frac{\sum_{k=1}^m p_k^{(h-1)} \alpha_k^{(h-1)} \lambda_k^{(h-1)} (1 - e^{-\lambda_k^{(h-1)} x})^{\alpha_k^{(h-1)} - 1} e^{-\lambda_k^{(h-1)} x}}{\sum_{k=1}^m p_k^{(h-1)} (1 - e^{-\lambda_k^{(h-1)} t_i})^{\alpha_k^{(h-1)}} - \sum_{k=1}^m p_k^{(h-1)} (1 - e^{-\lambda_k^{(h-1)} t_{i-1}})^{\alpha_k^{(h-1)}}},$$

$x \in (t_{i-1}, t_i)$ , and

$$p_{ij'}(x | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, Y) = \frac{\sum_{k=1}^m p_k^{(h-1)} \alpha_k^{(h-1)} \lambda_k^{(h-1)} (1 - e^{-\lambda_k^{(h-1)} x})^{\alpha_k^{(h-1)} - 1} e^{-\lambda_k^{(h-1)} x}}{1 - \sum_{k=1}^m p_k^{(h-1)} (1 - e^{-\lambda_k^{(h-1)} t_i})^{\alpha_k^{(h-1)}}},$$

$x \in (t_i, +\infty)$ .

M-step: we maximize the approximate Q function numerically in E-step with respect to unknown parameters  $P, \alpha, \lambda$  to update estimates which are denoted as  $P^{(h)}, \alpha^{(h)}, \lambda^{(h)}$ .

In order to obtain the estimates of unknown parameters more conveniently in M-step, use  $\tilde{\Delta}2_{kij}^{(h-1)}, \tilde{\Delta}5_{kij'}^{(h-1)}$  instead of  $\Delta 2_{kij}^{(h-1)}, \Delta 5_{kij'}^{(h-1)}$ , respectively. Here,  $\tilde{\Delta}2_{kij}^{(h-1)}, \tilde{\Delta}5_{kij'}^{(h-1)}$  have the following expressions:

$$\tilde{\Delta}2_{kij}^{(h-1)} = \int_{t_{i-1}}^{t_i} \ln(1 - e^{-\lambda_k^{(h-1)}x}) a_{k,i,i-1}^{(h-1)}(t_i, t_{i-1}) p_{ij}(x) dx$$

$$\tilde{\Delta}5_{kij'}^{(h-1)} = \int_{t_i}^{\infty} \ln(1 - e^{-\lambda_k^{(h-1)}x}) b_{ki}^{(h-1)}(t_i) p_{ij'}(x) dx$$

Then the above Q function is:

$$Q(P, \alpha, \lambda | P^{(h-1)}, \alpha^{(h-1)}, \lambda^{(h-1)}, Y) \approx \sum_{i=1}^N \sum_{k=1}^m \left[ (X_i \Delta 1_{k,i,i-1}^{(h-1)} + R_i \Delta 4_{ki}^{(h-1)}) \ln(p_k \alpha_k \lambda_k) \right. \\ \left. + (\alpha_k - 1)(X_i \tilde{\Delta}2_{kij}^{(h-1)} + R_i \tilde{\Delta}5_{kij'}^{(h-1)}) - \lambda_k (X_i \Delta 3_{kij}^{(h-1)} + R_i \Delta 6_{kij'}^{(h-1)}) \right]$$

Then, take the derivatives of unknown parameters and let

$$\frac{\partial Q}{\partial \alpha_k} \approx \sum_{i=1}^N \left[ \frac{1}{\alpha_k} (X_i \Delta 1_{k,i,i-1}^{(h-1)} + R_i \Delta 4_{ki}^{(h-1)}) + X_i \tilde{\Delta}2_{kij}^{(h-1)} + R_i \tilde{\Delta}5_{kij'}^{(h-1)} \right] = 0 \quad k = 1, \dots, m \tag{4.1}$$

$$\frac{\partial Q}{\partial \lambda_k} \approx \sum_{i=1}^N \left[ \frac{1}{\lambda_k} (X_i \Delta 1_{k,i,i-1}^{(h-1)} + R_i \Delta 4_{ki}^{(h-1)}) - (X_i \Delta 3_{kij}^{(h-1)} + R_i \Delta 6_{kij'}^{(h-1)}) \right] = 0 \quad k = 1, \dots, m \tag{4.2}$$

$$\frac{\partial Q}{\partial p_k} = \frac{1}{p_k} \sum_{i=1}^N (X_i \Delta 1_{k,i,i-1}^{(h-1)} + R_i \Delta 4_{ki}^{(h-1)}) - \frac{1}{1 - \sum_{j=1}^m p_j} \sum_{i=1}^N (X_i \Delta 1_{m,i,i-1}^{(h-1)} + R_i \Delta 4_{mi}^{(h-1)}) = 0 \quad k = 1, \dots, m-1. \tag{4.3}$$

Solve the Equations (4.1) and (4.2), there are

$$\hat{\alpha}_k^{(h)} \approx - \frac{\sum_{i=1}^N (X_i \Delta 1_{k,i,i-1}^{(h-1)} + R_i \Delta 4_{ki}^{(h-1)})}{\sum_{i=1}^N (X_i \tilde{\Delta}2_{kij}^{(h-1)} + R_i \tilde{\Delta}5_{kij'}^{(h-1)})}, \quad k = 1, \dots, m \tag{4.4}$$

$$\hat{\lambda}_k^{(h)} \approx \frac{\sum_{i=1}^N (X_i \Delta 1_{k,i,i-1}^{(h-1)} + R_i \Delta 4_{ki}^{(h-1)})}{\sum_{i=1}^N (X_i \Delta 3_{kij}^{(h-1)} + R_i \Delta 6_{kij'}^{(h-1)})}, \quad k = 1, \dots, m \tag{4.5}$$

Solve the Equation (4.3), there is



**Table 1. Average biases, MSE(s) of estimators for scheme 1 when the sample sizes are  $n = 60, 120, 180, 300, 500$**

$n$		$\hat{\rho}_1$	$\hat{\alpha}_1$	$\hat{\lambda}_1$	$\hat{\alpha}_2$	$\hat{\lambda}_2$
60	Bias	-0.0598	0.0293	-0.0384	0.0597	0.1282
	MSE	0.0237	0.0029	0.0046	0.0054	0.1485
120	Bias	-0.0942	0.0507	-0.0483	0.0738	0.1209
	MSE	0.0197	0.0034	0.0042	0.0061	0.0802
180	Bias	-0.1083	0.0598	-0.0527	0.0793	0.1283
	MSE	0.0189	0.0041	0.0038	0.0067	0.0575
300	Bias	-0.1148	0.0623	-0.0539	0.0819	0.1339
	MSE	0.0177	0.0042	0.0037	0.0069	0.0423
500	Bias	-0.1169	0.0666	-0.0559	0.0836	0.1310
	MSE	0.0163	0.0046	0.0036	0.0071	0.0327

**Table 2. Average biases, MSE(s) of estimators for scheme 2 when the sample sizes are  $n = 60, 120, 180, 300, 500$**

$n$		$\hat{\rho}_1$	$\hat{\alpha}_1$	$\hat{\lambda}_1$	$\hat{\alpha}_2$	$\hat{\lambda}_2$
60	Bias	-0.0899	0.0363	-0.0427	0.0654	0.1428
	MSE	0.0015	0.0002	0.0002	0.0029	0.0098
120	Bias	-0.0791	0.0459	-0.0472	0.0716	0.1371
	MSE	0.0008	0.0002	0.0002	0.0003	0.0040
180	Bias	-0.1091	0.0613	-0.0554	0.0780	0.1507
	MSE	0.0009	0.0002	0.0002	0.0003	0.0025
300	Bias	-0.1096	0.0681	-0.0557	0.0825	0.1034
	MSE	0.0008	0.0002	0.0002	0.0004	0.0012
500	Bias	-0.1126	0.0660	-0.0498	0.08726	0.1505
	MSE	0.0007	0.0002	0.0001	0.0004	0.0019

**Table 3. Average biases, MSE(s) of estimators for scheme 3 when the sample sizes are  $n = 60, 120, 180, 300, 500$**

$n$		$\hat{\rho}_1$	$\hat{\alpha}_1$	$\hat{\lambda}_1$	$\hat{\alpha}_2$	$\hat{\lambda}_2$
60	Bias	-0.0106	-0.0156	-0.0205	-0.0180	0.0406
	MSE	0.0209	0.0033	0.0032	0.0016	0.1572
120	Bias	-0.0217	-0.0055	-0.0230	-0.0148	0.0504
	MSE	0.0109	0.0017	0.0021	0.0007	0.0904
180	Bias	-0.0314	-0.0008	-0.0245	-0.0135	0.0433
	MSE	0.0085	0.0011	0.0016	0.0005	0.0583
300	Bias	-0.0319	0.0004	-0.0255	-0.0132	0.0563
	MSE	0.0058	0.0007	0.0013	0.0003	0.0382
500	Bias	-0.0322	0.0020	-0.0252	-0.0126	0.0653
	MSE	0.0036	0.0004	0.0009	0.0003	0.2480

**Table 4. Average biases, MSE(s) of estimators for scheme 4 when the sample sizes are  $n = 60, 120, 180, 300, 500$**

$n$		$\hat{\rho}_1$	$\hat{\alpha}_1$	$\hat{\lambda}_1$	$\hat{\alpha}_2$	$\hat{\lambda}_2$
60	Bias	-0.0018	-0.0198	-0.0171	-0.0276	0.0231
	MSE	0.0209	0.0043	0.0032	0.0025	0.1468
120	Bias	-0.0096	-0.0113	-0.0169	-0.0250	0.0553
	MSE	0.0112	0.0022	0.0016	0.0011	0.0839
180	Bias	-0.0155	-0.0062	-0.0191	-0.0229	0.0542
	MSE	0.0070	0.0013	0.0012	0.0008	0.0594
300	Bias	-0.0164	-0.0028	-0.0190	-0.0203	0.0708
	MSE	0.0048	0.0009	0.0008	0.0006	0.0365
500	Bias	-0.0203	-0.0007	-0.0200	-0.0192	0.0694
	MSE	0.0029	0.0005	0.0007	0.0005	0.0247

$$\begin{aligned}
 & p_k \sum_{i=1}^N (X_i \Delta 1_{m,i,j-1}^{(h-1)} + R_i \Delta 4_{mi}^{(h-1)}) + \sum_{j=1}^{m-1} p_j \cdot \sum_{i=1}^N (X_i \Delta 1_{k,i,j-1}^{(h-1)} + R_i \Delta 4_{ki}^{(h-1)}) \\
 &= \sum_{i=1}^N (X_i \Delta 1_{k,i,j-1}^{(h-1)} + R_i \Delta 4_{ki}^{(h-1)}), \quad k = 1, 2, \dots, m-1 \tag{4.6}
 \end{aligned}$$

From Equation (4.6), obtain the  $h$ -th iteration values of parameters  $p_1, \dots, p_{m-1}$  which are the solutions of the linear equation group denoted by  $AP = b$ , where  $P, A, b$  are given respectively as follows

$$\begin{aligned}
 & P = (p_1, p_2, \dots, p_{m-1})^T \\
 & A_{m-1} = (a_{ls}), \quad a_{ls} = \begin{cases} \sum_{i=1}^N (X_i \Delta 1_{l,i,j-1}^{(h-1)} + R_i \Delta 4_{li}^{(h-1)}) + \sum_{i=1}^N (X_i \Delta 1_{m,i,j-1}^{(h-1)} + R_i \Delta 4_{mi}^{(h-1)}), & l = s \\ \sum_{i=1}^N (X_i \Delta 1_{l,i,j-1}^{(h-1)} + R_i \Delta 4_{li}^{(h-1)}), & l \neq s \end{cases} \\
 & b = \left( \sum_{i=1}^N (X_i \Delta 1_{1,i,j-1}^{(h-1)} + R_i \Delta 4_{1i}^{(h-1)}), \dots, \sum_{i=1}^N (X_i \Delta 1_{m-1,i,j-1}^{(h-1)} + R_i \Delta 4_{m-1i}^{(h-1)}) \right)^T
 \end{aligned}$$

Since  $\sum_{i=1}^N (X_i \Delta 1_{l,i,j-1}^{(h-1)} + R_i \Delta 4_{li}^{(h-1)}) > 0, \quad l = 1, 2, \dots, m-1$ , it is easy to testify that the rank of matrix  $A$  is  $m-1$ , i.e.  $A$  is an invertible matrix. Therefore, the unique solution of parameter vector  $P$  of the  $h$ -th iteration in the  $M$ -step is obtained

$$\hat{p}^{(h)} = (p_1^{(h)}, p_2^{(h)}, \dots, p_{m-1}^{(h)})^T = A^{-1}b \tag{4.7}$$

From the above Equations (4.4), (4.5) and (4.7), we can update  $\hat{\rho}^{(h)}, \hat{\alpha}^{(h)}, \hat{\lambda}^{(h)}$  by repeating E-step and M-step till the total error of all estimated parameters approaches the supposed restraint.

### 5. Simulation study

The purpose of simulation study is to investigate the performance of the estimates for the MGED parameters in modeling progressive type-I interval censoring lifetime data. Here, we use some similar algorithm steps proposed in Aggarwala (2001). The simulation is conducted in R language. To be self-contained, the algorithm is re-produced as follows.

**Table 5. Survival time in interval form for the patients who underwent operations in connection with a type of malignant disease**

Interval	$d_j$	$R_i$
[0, 1)	90	0
[1, 2)	76	0
[2, 3)	51	0
[3, 4)	25	12
[4, 5)	20	5
[5, 6)	7	9
[6, 7)	4	9
[7, 8)	1	3
[8, 9)	3	5
[9, 10)	2	5
[10, $\infty$ )	47	0

Firstly, generate the numbers,  $X_p$ , of failed items in each subinterval  $(t_{i-1}, t_i]$ ,  $i = 1, 2, \dots, N$ , from a sample of size  $n$  putting on life testing at time  $t_i = 0$ . A progressively type-I interval-censored data,  $(X_p, R_p, t_i)$ ,  $i = 1, 2, \dots, N$ , from MGED which has distribution (2.2) can be generated using the fact that

$$X_1 \sim \text{Binom}[n, F(T_1)] \tag{5.1}$$

and let  $X_0 = 0$  and  $R_0 = 0$  for  $i = 1, 2, \dots, N$ ,

$$\begin{aligned}
 &X_i | X_{i-1}, \dots, X_1, R_{i-1}, \dots, R_1 \\
 &\sim r\text{Binom} \left[ n - \sum_{j=1}^{i-1} (X_j + R_j), \frac{F(T_i) - F(T_{i-1})}{1 - \sum_{j=1}^{i-1} [F(T_j) - F(T_{j-1})]} \right] \\
 &= r\text{Binom} \left[ n - \sum_{j=1}^{i-1} (X_j + R_j), \frac{F(T_i) - F(T_{i-1})}{1 - F(T_{i-1})} \right] \tag{5.2}
 \end{aligned}$$

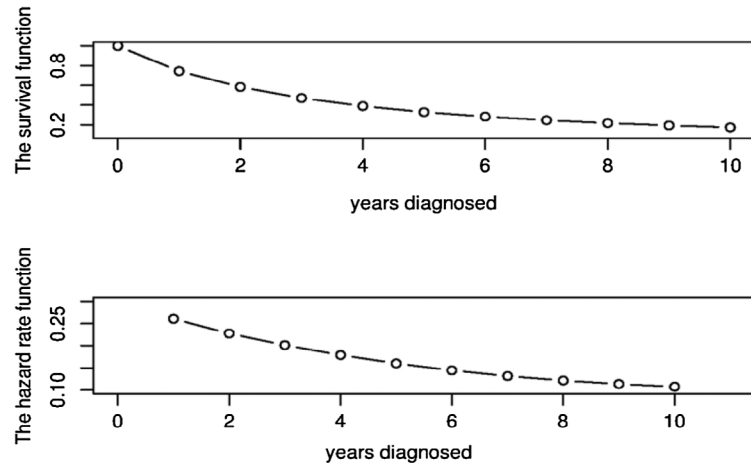
$$R_i = \text{floor} \left[ p'_i \times \left( n - \sum_{j=1}^{i-1} (X_j + R_j) - X_i \right) \right] \tag{5.3}$$

where floor() returns the largest integer not greater than the argument in R language and  $0 = t_0 < t_1 < \dots < t_N < \infty$  are pre-scheduled times.

In this paper, we just consider the MGED model with two components under the progressive type-I interval censoring scheme. Suppose that the true values of model parameters are  $p_1 = 0.6$ ,  $\alpha_1 = 0.9$ ,  $\lambda_1 = 0.13$ ,  $\alpha_2 = 1$ ,  $\lambda_2 = 1.3$ , while the initial values are taken as  $p_1^{(0)} = 0.4$ ,  $\alpha_1^{(0)} = 0.8$ ,  $\lambda_1^{(0)} = 0.1$ ,  $\alpha_2^{(0)} = 0.9$ ,  $\lambda_2^{(0)} = 1$ . Each replication of the simulation generates a progressively type-I interval-censored data of size  $n = 60, 120, 180, 300, 500$  with  $N = 10$ , pre-specified inspection times  $t_1 = 1, t_2 = 2, t_3 = 3, t_4 = 4, t_5 = 5, t_6 = 6, t_7 = 7, t_8 = 8, t_9 = 9$  and  $t_{10} = 10$  and  $t_{10} = 10$  as the time to terminate the experiment.

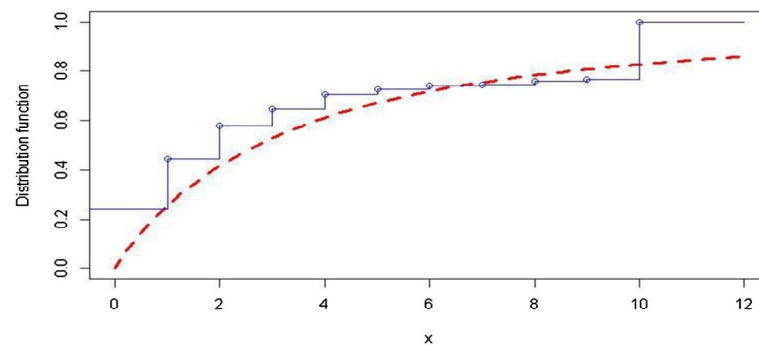
In this paper, we consider the following progressive interval censoring

**Figure 1. The fitted survival function and hazard rate function of the real data.**



**Figure 2. The empirical and fitted distribution function.**

Note: Fold line represents the empirical distribution function, dotted line represents the fitted distribution function.



**Schemes:**

- Scheme 1:  $p'_{(1)} = (0.05, 0.05, 0.05, 0.05, 0.1, 0.1, 0.1, 0.1, 0.1, 1)$
- Scheme 2:  $p'_{(2)} = (0.05, 0.05, 0.05, 0.05, 0.1, 0.1, 0, 0, 0, 1)$
- Scheme 3:  $p'_{(3)} = (0, 0, 0, 0.1, 0.05, 0.1, 0.15, 0.05, 0.05, 1)$
- Scheme 4:  $p'_{(4)} = (0, 0, 0, 0, 0, 0, 0, 0, 0, 1)$

where censoring in  $p'_{(1)}$  in lighter for the first four intervals and heavier for the next five intervals. The censoring pattern is reversed in  $p'_{(2)}$ ,  $p'_{(3)}$  is no censoring in the first three interval and then some censoring afterwards.  $p'_{(4)}$  is the conventional interval censoring where no removals prior to the experiment termination.

Repeat the simulation experiments 1,000 times for the sample sizes  $n = 60, 120, 180, 300, 500$ . Denote the  $h$ -th estimates as  $\Theta^{(h)} = (p_1^{(h)}, \alpha_1^{(h)}, \lambda_1^{(h)}, \alpha_2^{(h)}, \lambda_2^{(h)})$ ,  $h = 1, \dots, s$ ,  $s = 1,000$ . The final average bias and mean square errors (MSEs) of the estimates are given respectively as  $\text{Bias}(\theta_j) = (1/s) \sum_{h=1}^s (\hat{\theta}_j^{(h)} - \theta_j)$  and  $\text{MSE}(\theta_j) = (1/s) \sum_{h=1}^s (\hat{\theta}_j^{(h)} - \theta_j)^2$ , where  $\hat{\theta}_j$  is the  $j$ -th coordinate of the unknown parameter vector  $\theta$ . The computation results for four different censoring schemes are illustrated in Tables 1–4, respectively.

From Tables 1–4, we can see that the EM algorithm is very effective in dealing with parameter estimation of the MGED for the given four sampling schemes under the progressive type-I interval censoring. The number of the specified inspection is fixed. For the given one scheme, it is obvious that the simulation results are changed as  $n$  increases, but the range is very small. It is easy to understand the phenomenon that the estimation of the unknown parameters is biased slightly when

the number of the items withdrawn from the experiment is increasing. It is obvious that the simulation results are mainly decided by the sample size  $n$  and the censoring proportion. For different censoring schemes, the smaller censoring proportion we design in life testing, the better estimation results we will get. Therefore, the scheme 4 presents the most precise estimation. Further, for fixed  $N$  as  $n$  increases, on the whole, the biases, the MSEs decrease.

### 6. Data analysis

In order to illustrate the effectiveness of the model and algorithm, we analyze a real data-set below. A data-set which describe the survival times for surgery of a group of 374 patients who underwent operations in connection with a type of malignant disease (Berkson & Gage, 1950; Lawless, 2003). The data are given in Table 5.

According to the data above,  $d_j$ s in the first five intervals are significantly different with the last intervals and think that the data may not come from the homogenous population. Thus, consider two components of the MGED (1.1) to analyze this data-set. Through the method discussed above, we can obtain that the parameter estimates are  $\hat{\rho}_1 = 0.4213$ ,  $\hat{\alpha}_1 = 0.9216$ ,  $\hat{\lambda}_1 = 0.0886$ ,  $\hat{\alpha}_2 = 0.9118$ ,  $\hat{\lambda}_2 = 0.4017$ , respectively. And the survival function and the hazard rate function of the model (1.1) are as follow:

$$\hat{S}(x) = 1 - [0.4213 \times (1 - e^{-0.0886x})^{0.9216} + 0.5787 \times (1 - e^{-0.4017x})^{0.9118}], \quad x \geq 0 \quad (6.1)$$

$$\hat{h}(x) = \frac{0.0344 \times (1 - e^{-0.0886x})^{-0.0784} e^{-0.0886x} + 0.2119 \times (1 - e^{-0.4017x})^{-0.0882} e^{-0.4017x}}{1 - [0.4213 \times (1 - e^{-0.0886x})^{0.9216} + 0.5787 \times (1 - e^{-0.4017x})^{0.9118}]}, \quad x \geq 0 \quad (6.2)$$

From (6.1) and (6.2), we can get the fitted survival function and hazard rate function of the real data as in Figure 1. For Figure 1, the hazard rate function is monotone decreasing function. There is great change in the first few years and then tends to be gentle after eight years. Therefore, for this disease, the risk rate of patients after operation in the short time is higher. In order to avoid the recurrence in the recovery stage after the operation, we suggest to check regularly.

In order check the validity of the model, we adopt the Kolmogorov-Smirnov goodness-of-fit test statistic for the fitted distribution  $\hat{F}(x)$ . Define the maximum distance in Kolmogorov-Smirnov goodness-of-fit test:

$$D_n(F) = \sup_{0 \leq x < \infty} |\hat{F}_n(x) - F(x|\hat{\theta})|, \quad (6.3)$$

as the distance between the empirical distribution  $\hat{F}_n(x)$ , of the given complete data-set and the fitted distribution function  $F(x|\hat{\theta})$  with  $\hat{\theta}$  as the MLE of unknown parameter vector  $\theta$ . When a progressively type-I interval-censored data are given, the empirical distribution is replaced by the following (6.4) in the formula  $D_n(F)$ .

$$\hat{F}(x_j) = 1 - \prod_{j=1}^i (1 - \hat{\rho}_j), \quad i = 1, \dots, N \quad (6.4)$$

where  $\hat{\rho}_j = \frac{d_j}{n - \sum_{k=0}^{j-1} d_k - \sum_{k=0}^{j-1} R_k}$ ,  $j = 1, \dots, N$ .

Fit the data-set in Table 5, we obtain the K-S distance 0.2406. So it is reasonable to say that the MGED provides a good fit for the given data-set in Table 5. Figure 2 is given to compare the empirical distribution function with the fitted distribution function.

## 7. Conclusion and remarks

This article considers the estimation of parameters based on a progressively type-I interval-censored data from a MGED. Because the data in this paper are not complete, the EM algorithm is used to obtain the maximum likelihood estimates. Then the performance of the estimates is presented by a simulating study in different setups and gets effective results. Lastly, a surgery data of a group of 374 patients who underwent operations in connection with a type of malignant disease are presented to illustrate the method of estimation developed here.

### Funding

This research was partially supported by National Science Foundation of China (NSFC) [grant numbers 11301037, 11671054 and 11571051] and The Education Department of Jilin Province, "13th Five-Year" project planning 2016317.

### Author details

Chunjie Wang<sup>1</sup>  
E-mail: [cjwang2014@126.com](mailto:cjwang2014@126.com)  
Shuying Wang<sup>1,2</sup>  
E-mail: [wangshuying0601@163.com](mailto:wangshuying0601@163.com)  
Dehui Wang<sup>2</sup>  
E-mail: [wangdh@jlu.edu.cn](mailto:wangdh@jlu.edu.cn)  
Chunjing Li<sup>1,2</sup>  
E-mail: [lichunjing@ccut.edu.cn](mailto:lichunjing@ccut.edu.cn)  
Xiaogang Dong<sup>1</sup>  
E-mail: [dongxiaogang@ccut.edu.cn](mailto:dongxiaogang@ccut.edu.cn)

<sup>1</sup> College of Basic Science, Changchun University of Technology, Changchun 130012, China.

<sup>2</sup> College of Mathematics, Jilin University, Changchun, China.

### Citation information

Cite this article as: Parameter estimations for mixed generalized exponential distribution based on progressive type-I interval censoring, Chunjie Wang, Shuying Wang, Dehui Wang, Chunjing Li & Xiaogang Dong, *Cogent Mathematics* (2017), 4: 1280913.

### References

- Aggarwala, R. (2001). Progressive interval censoring: Some mathematical results with applications to inference. *Communications in Statistics-Theory and Methods*, 30, 1921–1935.  
<http://dx.doi.org/10.1081/STA-100105705>
- Balakrishnan, N., & Aggarwala, R. (2000). *Progressive Censoring*. Boston, MA: Birkhauser.  
<http://dx.doi.org/10.1007/978-1-4612-1334-5>
- Balakrishnan, N., Kannan, N., Lin, C. T., & Wu, S. J. (2004). Inference for the extreme value distribution under progressive Type-II censoring. *Journal of Statistical Computation and Simulation*, 74, 25–45.  
<http://dx.doi.org/10.1080/0094965031000105881>
- Berkson, J., & Gage, R. P. (1950). Calculation of survival rates for cancer. *Proceedings of the Staff Meetings Mayo Clinic*, 25, 270–286.
- Chen, D. G., & Lio, Y. L. (2010). Parameter estimations for generalized exponential distribution under progressive type-I interval censoring. *Computational Statistics and Data Analysis*, 54, 1581–1591.  
<http://dx.doi.org/10.1016/j.csda.2010.01.007>
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, 39, 1–38.
- Lawless, J. F. (2003). *Statistical models and methods for lifetime data* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Lin, C. T., Wu, S. J. S., & Balakrishnan, N. (2009). Planning life tests with progressively type-I interval censored data from the lognormal distribution. *Journal of Statistical Planning and Inference*, 139, 54–61.  
<http://dx.doi.org/10.1016/j.jspi.2008.05.016>
- Lio, Y. Z., Chen, D. G., & Tsai, T. R. (2011). Parameter estimations for generalized Rayleigh distribution under progressively type-I interval censored data. *Open Journal of Statistics*, 1, 46–57.  
<http://dx.doi.org/10.4236/ojs.2011.12006>
- McClellan, S. (1986). Estimation for the mixed exponential distribution using grouped follow-up data. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 35, 31–37.
- Ng, H., & Wang, Z. (2009). Statistical estimation for the parameters of Weibull distribution based on progressively type-I interval censored sample. *Journal of Statistical Computation and Simulation*, 79, 145–159.  
<http://dx.doi.org/10.1080/00949650701648822>
- Peng, X., & Yan, Z. (2011). Parameter estimations with gamma distribution based on progressive type-I interval censoring. *IEEE International Conference on Computer Science and Automation Engineering*, 1, 449–453.
- Pradhan, B., & Gijo, E. V. (2013). *Parameter estimation of lognormal distribution under progressive type-I interval censoring* (Technical Report No. SQCOR-2013-02).
- Soliman, A. A. (2006). Estimators for the finite mixture of Rayleigh model based on progressively censored data. *Communications in Statistics-Theory and Methods*, 35, 803–820.  
<http://dx.doi.org/10.1080/03610920500501379>
- Tian, Y., He, J., & Chen, P. (2008). Mixed generalized exponential with censored data. *Journal of Chongqing Institute of Technology (Natural Science)*, 22, 79–81.
- Tian, Y. Z., Tian, M. Z. I., & Chen, P. (2012). Parameter estimation for a mixture of generalized exponential distribution under grouped and right-censored samples. *Chinese Journal of Applied Probability and Statistics*, 28, 561–571.
- Tian, Y. Z., Tian, M. Z., & Zhu, Q. Q. (2014). Estimating a finite mixed exponential distribution under progressively type-II censored data. *Communications in Statistics-Theory and Methods*, 43, 3762–3776.  
<http://dx.doi.org/10.1080/03610926.2012.752843>
- Tian, Y. Z., Zhu, Q. Q., & Tian, M. Z. (2013). Inference for mixed generalized exponential distribution under progressively type-II censored samples. *Journal of Applied Statistics*, 41, 660–676.



© 2017 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



***Cogent Mathematics* (ISSN: 2331-1835) is published by Cogent OA, part of Taylor & Francis Group.**

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at [www.CogentOA.com](http://www.CogentOA.com)**

