Hacettepe Journal of Mathematics and Statistics
Volume 45 (5) (2016), 1605 – 1620

Comparing of some estimation methods for parameters of the Marshall-Olkin generalized exponential distribution under progressive Type-I interval censoring

Hamzeh Torabi*† , Masumeh Forooghi ‡ and Hossein Nadeb §

Abstract

In this paper, we estimate the parameters of the Marshall-Olkin generalized exponential distribution under progressive Type-I interval censoring based on maximum likelihood, moment method and probability plot. A simulation study is conducted to compare these estimates in terms of mean squared errors and biases. Finally, these estimate methods are applied to a real data set based on patients with breast cancer in order to demonstrate the applicabilities.

Keywords: EM algorithm, Generalized exponential distribution, Maximum likelihood estimate, Method of moments, Type-I interval censoring.

2000 AMS Classification: 65C05, 62N01, 62N02.

Received : 06.12.2014 *Accepted* : 28.07.2015 *Doi* : 10.15672/HJMS.20158112561

1. Introduction

Aggarwala [2001] introduced Type-I interval and progressive censoring and developed the statistical inference for the exponential distribution based on progressively Type-I interval censored data. Ng and Wang [2009] introduced the concept of progressive Type-I interval censoring to the Weibull distribution and compared many different estimation methods for two parameters in the Weibull distribution via simulation.

*Department of Statistics, Yazd University, 89175-741, Yazd, IRAN,
Email: htorabi@yazd.ac.ir
†Corresponding Author.
‡Department of Statistics, Yazd University, 89175-741, Yazd, IRAN,
Email: forughi87@gmail.com
§Department of Statistics, Yazd University, 89175-741, Yazd, IRAN,
Email: honadeb@yahoo.com

The generalized exponential (GE) distribution has the following probability density function (pdf)

$$f(t;\alpha,\lambda) = \alpha\lambda(1-e^{-\lambda t})^{\alpha-1}e^{-\lambda t},$$

where $t > 0$, $\alpha > 0$ and $\lambda > 0$. The cumulative distribution and the hazard rate function of the GE distribution are as follows:

$$F(t;\alpha,\lambda) = (1-e^{-\lambda t})^{\alpha},$$

and

$$h(t;\alpha,\lambda) = \frac{\alpha\lambda(1-e^{-\lambda t})^{\alpha-1}e^{-\lambda t}}{1-(1-e^{-\lambda t})^{\alpha}},$$

where $t > 0$; The GE distribution was introduced by Gupta and Kunda [2001]. Recently, Chen and Lio [2010] introduced the concept of progressive Type-I interval censoring for the generalized exponential distribution and compared many different estimation methods for the parameters of the distribution via a simulation study.

The Marshall-Olkin generalized exponential (MOGE) distribution was first proposed by Marshall and Olkin [1997] and extensively discussed by Alice and Jose [1999]. The PDF of the MOGE distribution with the parameters λ and α is

$$f(t;\alpha,\lambda) = \frac{\alpha\lambda e^{-\lambda t}}{(1-(1-\alpha)e^{-\lambda t})^2}, \quad t > 0, \quad 0 < \alpha \leq 1, \lambda > 0. \tag{1.1}$$

Also, the distribution function and the hazard rate function of the MOGE distribution are as follows:

$$F(t;\alpha,\lambda) = \frac{1-e^{-\lambda t}}{1-(1-\alpha)e^{-\lambda t}}, \tag{1.2}$$

and

$$h(t;\alpha,\lambda) = \frac{\lambda}{1-(1-\alpha)e^{-\lambda t}},$$

where $t > 0$. Note that if $\alpha = 1$, the MOGE distribution reduces to the conventional exponential distribution. Plots of the density functions, distribution functions and hazard rate functions for different values of α and λ are given in figures 1, 2 and 3, respectively.

The first two moments and variance of the MOGE distribution are given by

$$E[T] = \frac{\alpha\log(\alpha)}{(\alpha-1)\lambda}, \tag{1.3}$$

$$E[T^2] = \frac{2\alpha\text{PolyLog}[2,1-\alpha]}{(1-\alpha)\lambda^2}, \tag{1.4}$$

$$Var[T] = \frac{-\alpha\Big(\alpha\log(\alpha)^2 + 2(\alpha-1)\text{PolyLog}[2,1-\alpha]\Big)}{(\alpha-1)^2\lambda^2},$$

where

$$\text{PolyLog}[2,1-\alpha] = \sum_{k=1}^{\infty}\frac{(1-\alpha)^k}{k^2}.$$

In this paper, we study the maximum likelihood estimates, estimates via moment methods and estimates via probability plot for two parameters of the MOGE distribution under the progressive Type-I interval censoring. Section 2 introduces the progressive Type-I interval censoring for the MOGE distribution. In Section 3, some methods for parameters estimation are given. In Section 4, a simulation study is conducted to compare the performances of these estimation methods based on the mean squared error

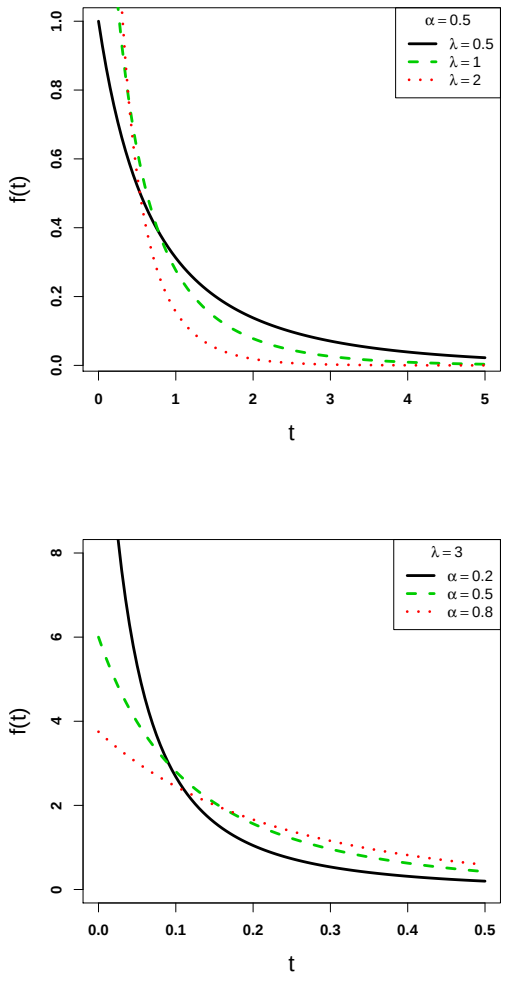

Figure 1. Plots of density functions for different values of α and λ

(MSE) and bias. Finally, a numerical example for a real data set is considered and some discussions and conclusions are given.

2. Progressively Type-I interval censored data

Suppose that n items are placed on a life testing problem simultaneously at time $t_0 = 0$ under inspection at m pre-specified times $t_1 < t_2 < \ldots < t_m$ where t_m is the scheduled time to terminate the experiment. At the ith inspection time, t_i, the number, X_i, of failures within $(t_{i-1}, t_i]$ is recorded and R_i surviving items are randomly removed from the life testing, for $i = 1, \ldots, m$. It is obvious that the number of surviving items at the time t_i is $Y_i = n - \sum_{j=1}^{i} X_j - \sum_{j=1}^{i-1} R_j$. Since Y_i is a random variable and the exact

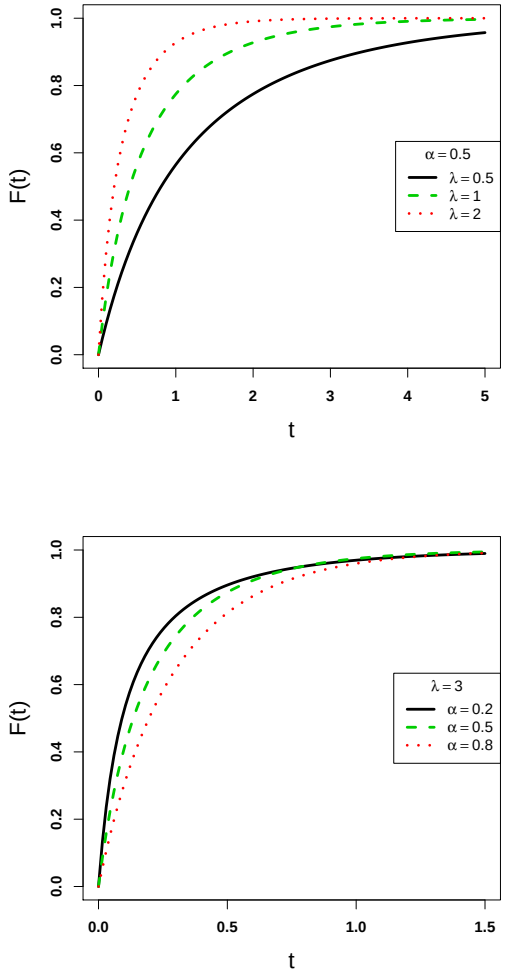

Figure 2. Plots of distributions functions for different values of α and λ

number of items withdrawn should not be greater than Y_i at time schedule t_i, R_i could be determined by the pre-specified percentage of the remaining surviving units at t_i for given $i = 1, 2, \ldots, m$. Also, given pre-specified percentage values, $p_1, \ldots, p_{m-1}$ and $p_m = 1$, for withdrawing at $t_1 < t_2 < \ldots < t_m$, respectively, $R_i = \lfloor p_i y_i \rfloor$ at each inspection time t_i where $i = 1, 2, \ldots, m$. Therefore, a progressively Type-I interval censored sample can be denoted as (X_i, R_i, t_i), $i = 1, 2, \ldots m$, where sample size is $n = \sum_{i=1}^{m} (X_i + R_i)$. Note that if $R_i = 0$, $i = 1, 2, \ldots, m-1$, then the progressively Type-I interval censored sample is a Type-I interval censored sample, $X_1, X_2, \ldots, X_m, X_{m+1} = R_m$.

Let a progressively Type-I interval censored sample be collected as described above, beginning with a random sample of n units with a continuous life time distribution

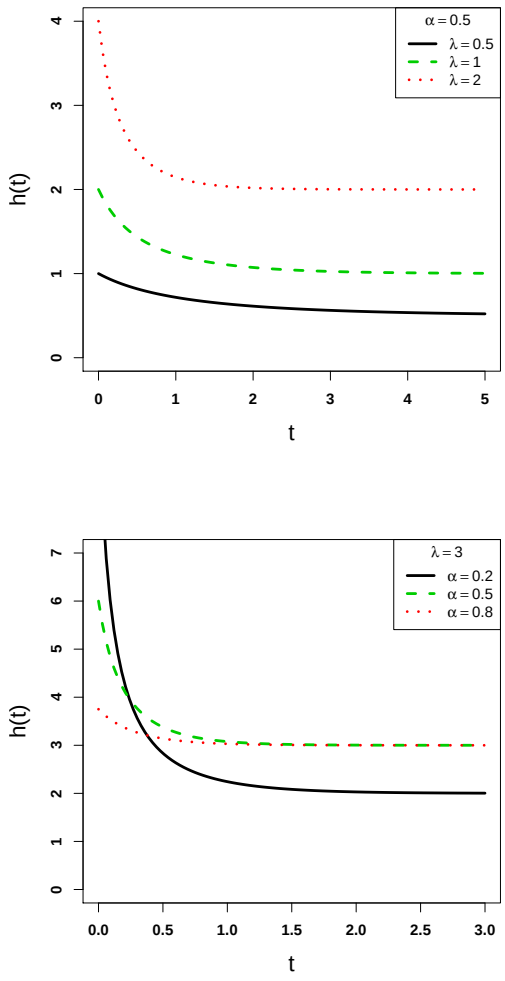

Figure 3. Plots of hazard functions for different values of α and λ

function $F(.;\boldsymbol{\theta})$. Then, based on the observed data, the likelihood function will be as follows:

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^{m} \left[F(t_i;\boldsymbol{\theta}) - F(t_{i-1};\boldsymbol{\theta})\right]^{X_i} \left[1 - F(t_i;\boldsymbol{\theta})\right]^{R_i}.$$

3. Some parameter estimation methods

In this section, we give some estimation methods for the parameters of the MOGE distribution.

3.1. Maximum likelihood estimation. Suppose a progressive Type-I interval censored sample is collected for the MOGE distribution. Using (1.2), the likelihood function is

$$L(\alpha,\lambda) \propto \prod_{i=1}^{m}\left[\frac{1-e^{-\lambda t_i}}{1-(1-\alpha)e^{-\lambda t_i}}-\frac{1-e^{-\lambda t_{i-1}}}{1-(1-\alpha)e^{-\lambda t_{i-1}}}\right]^{X_i}\left[\frac{\alpha e^{-\lambda t_i}}{1-(1-\alpha)e^{-\lambda t_i}}\right]^{R_i},$$

and the log-likelihood function is

$$\begin{aligned}\ell(\alpha,\lambda) \quad \propto \quad & \sum_{i=1}^{m} X_i \log\left[\frac{1-e^{-\lambda t_i}}{1-(1-\alpha)e^{-\lambda t_i}}-\frac{1-e^{-\lambda t_{i-1}}}{1-(1-\alpha)e^{-\lambda t_{i-1}}}\right] \\ & +\sum_{i=1}^{m} R_i \log\left[\frac{\alpha e^{-\lambda t_i}}{1-(1-\alpha)e^{-\lambda t_i}}\right].\end{aligned}$$

Hence, we have the following log-likelihood equations:

$$\begin{cases}\dfrac{\partial \ell(\alpha,\lambda)}{\partial \alpha}=0,\\ \dfrac{\partial \ell(\alpha,\lambda)}{\partial \lambda}=0.\end{cases} \tag{3.1}$$

The MLEs of α and λ cannot be obtained in a closed form by solving equations (3.1) and they must be calculated using a numerical method. Since there is no closed form for MLEs, a mid-point approximation and the EM algorithm are introduced as follows for finding the MLEs of α and λ.

3.1.1. *Mid-point approximation method.* The midpoint estimators based on progressively Type-I interval censoring can be obtained by assuming that X_i failures occurred at the center of the interval, $m_i = \frac{t_{i-1}+t_i}{2}$, and the R_i censored items withdrawn at the censoring time t_i. Then log-likelihood function from the MOGE distribution can be specified as follows:

$$\begin{aligned}\log(L^{\circ}) \quad \propto \quad & \sum_{i=1}^{m}\left[X_i \log(f(m_i;\alpha,\lambda))+R_i\log(1-F(t_i;\alpha,\lambda))\right] \\ = \quad & n\log\alpha+\log\lambda\sum_{i=1}^{m}X_i-\lambda\sum_{i=1}^{m}(X_i m_i+R_i t_i) \\ & -2\sum_{i=1}^{m}X_i\log(1-(1-\alpha)e^{-\lambda m_i})-\sum_{i=1}^{m}R_i\log(1-(1-\alpha)e^{-\lambda t_i}).\end{aligned}$$

Therefore, the maximum likelihood estimate of α, $\hat{\alpha}$, and the maximum likelihood estimate of λ, $\hat{\lambda}$, are the solution of the sequel equations:

$$\frac{n}{\hat{\alpha}} = 2\sum_{i=1}^{m}X_i\frac{e^{-\hat{\lambda}m_i}}{1-(1-\hat{\alpha})e^{-\hat{\lambda}m_i}}+\sum_{i=1}^{m}R_i\frac{e^{-\hat{\lambda}t_i}}{1-(1-\hat{\alpha})e^{-\hat{\lambda}t_i}}, \tag{3.2}$$

and

$$\begin{aligned}\frac{\sum_{i=1}^{m}X_i}{\hat{\lambda}} \quad = \quad & \sum_{i=1}^{m}(X_i m_i+R_i t_i)+2(1-\hat{\alpha})\sum_{i=1}^{m}X_i m_i\frac{e^{-\hat{\lambda}m_i}}{1-(1-\hat{\alpha})e^{-\hat{\lambda}m_i}} \\ & +(1-\hat{\alpha})\sum_{i=1}^{m}R_i t_i\frac{e^{-\hat{\lambda}t_i}}{1-(1-\hat{\alpha})e^{-\hat{\lambda}t_i}}.\end{aligned} \tag{3.3}$$

There is no closed form for the solutions of (3.2) and (3.3), thus an iterative numerical method is needed to obtain the parameter estimates, i.e., $\hat{\alpha}$ and $\hat{\lambda}$.

3.1.2. *EM algorithm.* The EM algorithm is applicable to obtain the maximum likelihood estimator of the parameters and useful in a variety of incomplete-data problems where algorithms such as the Newton-Raphson method may sometimes be complicated. On each iteration of the EM algorithm, there are two steps called E-step and the M-step: Let $y_{ij}, j = 1, 2, \ldots, X_i$, be the survival times within subinterval $(t_{i-1}, t_i]$ and $z_{ij}, j = 1, 2, \ldots, R_i$, be the survival times for those withdrawn items at t_i for $i = 1, 2, 3, \ldots, m$, then the log-likelihood, $\log(L^*)$, for the complete lifetimes of n items from the MOGE distribution is given as follows:

$$\begin{aligned}\log(L^*) &\propto \sum_{i=1}^{m}\left[\sum_{j=1}^{X_i}\log(f(y_{ij},\theta)) + \sum_{j=1}^{R_i}\log(f(z_{ij},\theta))\right] \\ &= n(\log\alpha + \log\lambda) - \lambda\sum_{i=1}^{m}\left[\sum_{j=1}^{X_i} y_{ij} + \sum_{j=1}^{R_i} z_{ij}\right] \\ &\quad -2\sum_{i=1}^{m}\left[\sum_{j=1}^{X_i}\log(1-(1-\alpha)e^{-\lambda y_{ij}}) + \sum_{j=1}^{R_i}\log(1-(1-\alpha)e^{-\lambda z_{ij}})\right].\end{aligned} \tag{3.4}$$

Taking the derivative with respective to α and λ, respectively, on (3.4), likelihood equations are obtained by

$$\frac{n}{\alpha} = 2\sum_{i=1}^{m}\left[\sum_{j=1}^{X_i}\frac{e^{-\lambda y_{ij}}}{(1-(1-\alpha)e^{-\lambda y_{ij}})} + \sum_{j=1}^{R_i}\frac{e^{-\lambda z_{ij}}}{(1-(1-\alpha)e^{-\lambda z_{ij}})}\right],$$

and

$$\begin{aligned}\frac{n}{\lambda} &= 2\sum_{i=1}^{m}\left[\sum_{j=1}^{X_i}\frac{(1-\alpha)y_{ij}e^{-\lambda y_{ij}}}{(1-(1-\alpha)e^{-\lambda y_{ij}})} + \sum_{j=1}^{R_i}\frac{(1-\alpha)z_{ij}e^{-\lambda z_{ij}}}{(1-(1-\alpha)e^{-\lambda z_{ij}})}\right] \\ &\quad +\sum_{i=1}^{m}\left[\sum_{j=1}^{X_i} y_{ij} + \sum_{j=1}^{R_i} z_{ij}\right].\end{aligned}$$

The EM-algorithm has the following steps:
Step 1. Given initial estimates of α and λ , say $\alpha^{(0)}$ and $\lambda^{(0)}$;
Step 2. In the kth iteration, the E-step requires to compute

$$E_{1i} = E_{\hat{\alpha}^{(k)},\ \hat{\lambda}^{(k)}}\left[Y\middle|Y\in[t_{i-1},t_i)\right],$$

$$E_{2i} = E_{\hat{\alpha}^{(k)},\ \hat{\lambda}^{(k)}}\left[Y\middle|Y\in[t_i,\infty)\right],$$

$$E_{3i} = E_{\hat{\alpha}^{(k)},\ \hat{\lambda}^{(k)}}\left[\frac{e^{-\hat{\lambda}^{(k)}Y}}{1-(1-\hat{\alpha}^{(k)})e^{-\hat{\lambda}^{(k)}Y}}\middle|Y\in[t_{i-1},t_i)\right],$$

$$E_{4i} = E_{\hat{\alpha}^{(k)},\ \hat{\lambda}^{(k)}}\left[\frac{e^{-\hat{\lambda}^{(k)}Y}}{1-(1-\hat{\alpha}^{(k)})e^{-\hat{\lambda}^{(k)}Y}}\middle|Y\in[t_{i-1},\infty)\right],$$

$$E_{5i} = E_{\hat{\alpha}^{(k)},\ \hat{\lambda}^{(k)}}\left[\frac{Ye^{-\hat{\lambda}^{(k)}Y}}{1-(1-\hat{\alpha}^{(k)})e^{-\hat{\lambda}^{(k)}Y}}\middle|Y\in[t_{i-1},t_i)\right],$$

and

$$E_{6i} = E_{\hat{\alpha}^{(k)},\ \hat{\lambda}^{(k)}}\left[\frac{Ye^{-\hat{\lambda}^{(k)}Y}}{1-(1-\hat{\alpha}^{(k)})e^{-\hat{\lambda}^{(k)}Y}}\middle| Y \in [t_{i-1},\infty)\right],$$

where Y is a random variable which has the MOGE distribution density function (1.1).
Step 3. The M-step maximize the likelihood function. Based on the likelihood equations for complete data, we can obtain the estimates

$$\hat{\alpha}^{(k+1)} = \frac{n}{2\sum_{i=1}^{m}\left[\sum_{j=1}^{X_i} E_{3i} + \sum_{j=1}^{R_i} E_{4i}\right]},$$

and

$$\hat{\lambda}^{(k+1)} = \frac{n}{\sum_{i=1}^{m}\left[\sum_{j=1}^{X_i} E_{1i} + 2(1-\hat{\alpha}^{(k+1)})E_{5i} + \sum_{j=1}^{R_i} E_{2i} + 2(1-\hat{\alpha}^{(k+1)})E_{6i}\right]};$$

Step 4. Setting $k = k+1$, the MLEs of α and λ can be obtained by repeating the E-step and M-step until convergence occurs.

Note that numerical integration methods are required to compute the above conditional expectations in Step 2.

3.2. Method of moments. Let Y be a random variable which has the MOGE distribution density function (1.1). The kth moment of a doubly truncated generalized exponential distribution in the interval (a,b) where $0 < a < b$ is given by

$$E_{\alpha,\ \lambda}\left[Y^k\middle| Y \in [a,b)\right] = \frac{\int_a^b y^k f(y;\alpha,\lambda)dy}{F(b;\alpha,\lambda) - F(a;\alpha,\lambda)}$$

Equating the sample moment to the corresponding population moment up to the second order, the following equations can be used to find the estimates of moment method:

$$E[Y] = \frac{1}{n}\left[\sum_{i=1}^{m} X_i E_{\alpha,\lambda}[Y|\, Y \in [t_{i-1},t_i)] + R_i E_{\alpha,\lambda}[Y|Y \in [t_{i-1},\infty)]\right],$$

and

$$\begin{aligned} E[Y^2] &= \frac{1}{n}\left[\sum_{i=1}^{m} X_i E_{\alpha,\lambda}[Y^2\big|\, Y \in [t_{i-1},t_i)]\right] \\ &\quad + \left[\sum_{i=1}^{m} R_i E_{\alpha,\lambda}[Y^2\big|\, Y \in [t_{i-1},\infty)]\right]. \end{aligned}$$

An iterative procedure can be employed to solve the above equations for α and λ as follows:
Step 1. Consider the initial values of α and λ, say $\hat{\alpha}^{(0)}$ and $\hat{\lambda}^{(0)}$ with $k = 0$;
Step 2. In the $k+1$th iteration,
• we compute $E_{\hat{\alpha}^{(k)},\ \hat{\lambda}^{(k)}}[Y|\, Y \in [t_{i-1},t_i)]$ and $E_{\hat{\alpha}^{(k)},\ \hat{\lambda}^{(k)}}\left[Y^2\big|\, Y \in [t_{i-1},t_i)\right]$ and solve the following equation for α, say $\hat{\alpha}^{(k+1)}$:

$$P(\alpha) = \frac{\left[\sum_{i=1}^{m} X_i E_{\alpha,\lambda}[Y|\, Y \in [t_{i-1},t_i)] + R_i E_{\alpha,\lambda}[Y|\, Y \in [t_{i-1},\infty)]\right]^2}{n\sum_{i=1}^{m}\left[X_i E_{\alpha,\lambda}[Y^2|Y \in [t_{i-1},t_i)] + R_i E_{\alpha,\lambda}[Y^2|\, Y \in [t_{i-1},\infty)]\right]},$$

where using (1.3) and (1.4),

$$P(\alpha) = E^2[Y]/E[Y^2] = \frac{\alpha \log^2 \alpha}{2(1-\alpha)\text{PolyLog}[2, 1-\alpha])}.$$

• The solution for α, say $\hat{\alpha}^{(k+1)}$, is obtained through the following equation:

$$\begin{aligned}\frac{\hat{\alpha}^{(k+1)} \log(\hat{\alpha}^{(k+1)})}{(\hat{\alpha}^{(k+1)} - 1)\lambda^{(k+1)}} &= \frac{1}{n}\Big[\sum_{i=1}^{m} X_i E_{\hat{\alpha}^{(k)},\hat{\lambda}^{(k)}}[Y|\, Y \epsilon\, [t_{i-1}, t_i)] \\ &\quad + \sum_{i=1}^{m} R_i E_{\hat{\alpha}^{(k)},\hat{\lambda}^{(k)}}[Y|\, Y \in [t_{i-1}, \infty)]\Big];\end{aligned}$$

Step 3. Checking convergence, if the convergence occurs then the current $\hat{\alpha}^{(k+1)}$ and $\hat{\lambda}^{(k+1)}$ are the estimates of α and λ by the method of moments; otherwise set $k = k + 1$ and go to Step 2.

3.3. Estimation based on probability plot. For progressively Type-I interval censored data, (X_i, R_i, t_i), $i = 1, 2, \ldots, m$, of size n, the distribution function at time t_i can be estimated as

$$\hat{F}(t_i) = 1 - \prod_{j=1}^{i}(1 - \hat{p}_j), \quad i = 1, 2, \ldots, m,$$

where

$$\hat{p}_j = \frac{X_j}{n - \sum_{k=0}^{j-1} X_k - \sum_{k=0}^{j-1} R_k}, \quad j = 1, 2, \ldots, m.$$

From (1.2), we have

$$t = -\frac{1}{\lambda} \log \frac{1 - F(t)}{1 - (1-\alpha)F(t)}.$$

If $\hat{F}(t_i)$ is the estimate of $F(t_i)$, then the estimates of α and λ in the MOGE distribution based on probability plot can be obtained by minimizing $\sum_{i=1}^{m}\Big[t_i + \frac{1}{\lambda}\log\frac{1-\hat{F}(t_i)}{1-(1-\alpha)\hat{F}(t_i)}\Big]^2$ with respect to α and λ.

3.4. Simulation algorithm. In this section, we give A short algorithm for simulating $X_1, X_2, \ldots, X_m$ from a random sample of size n put on life test at time 0 is therefore given below. Let $X_0 = R_0 = 0$; We use the fact that for $i = 1, ..., m$,

$$\begin{aligned}X_i|\, X_{i-1}, \ldots, X_0, R_{i-1}, \ldots, R_0 &\sim Binom\Big(n - \sum_{j=1}^{i-1}(X_j + R_j), \\ &\qquad \frac{F(t_i) - F(t_{i-1})}{1 - F(t_{i-1})}\Big),\end{aligned}$$

and

$$R_i = \left\lfloor p_i\big(n - \sum_{j=1}^{i-1}(X_i + R_i) - X_i\big) \right\rfloor.$$

Hence we can give an algorithm as follows:

Step 1. Set $i = 0$ and let $xsum = rsum = 0$;
Step 2. Next i;
Step 3. If $i = m + 1$, exit the algorithm;
Step 4. Generate X_i as a binomial random variable with parameters $(n - xsum - rsum)$

and $\frac{F(t_i)-F(t_{i-1})}{1-F(t_{i-1})}$;

Step 5. Calculate $R_i^{obs} = \left\lfloor p_i(n - \sum_{j=1}^{i-1}(X_i + R_i) - X_i) \right\rfloor$ or $R_i^{obs} = \min(n - xsum - rsum - X_i, R_i)$, depending upon how the censoring scheme is chosen;
Step 6. Set $xsum = xsum + X_i$, $rsum = rsum + R_i^{obs}$;
Step 7. Go to step 2.

3.5. Simulation schemes. Continuing with our exploration of progressive Type-I interval censoring under the MOGE distribution lifetime models, let us consider a numerical example, and discuss some of the issues which arise. We use the values $t_1 = 5.5$, $t_2 = 10.5$, $t_3 = 15.5$, $t_4 = 20.5$, $t_5 = 25.5$, $t_6 = 30.5$, $t_7 = 40.5$, $t_8 = 50.5$ and $t_9 = 60.5$. The lifetime distribution is the MOGE Type with parameters $(\alpha, \lambda) = (0.5, .06)$, where are the simulation input parameters. To compare the performances of the estimation procedures developed in this paper, we consider the following four progressive interval censoring schemes which are similar to the patterns of simulation schemes used in Aggarwala (2001) and also used in Ng and Wang (2009) and Chen and Lio (2010):

$$\begin{aligned} \boldsymbol{p}_{(1)} &= (.25, .25, .25, .25, .5, .5, .5, .5, 1), \\ \boldsymbol{p}_{(2)} &= (.5, .5, .5, .5, .25, .25, .25, .25, 1), \\ \boldsymbol{p}_{(3)} &= (0, 0, 0, 0, 0, 0, 0, 0, 1), \\ \boldsymbol{p}_{(4)} &= (.25, 0, 0, 0, 0, 0, 0, 0, 1), \end{aligned}$$

where censoring in $\boldsymbol{p}_{(1)}$ is lighter for the first four intervals and heavier for the next four intervals. The censoring pattern is reversed in $\boldsymbol{p}_{(2)}$. $\boldsymbol{p}_{(3)}$ is the conventional interval censoring where no removals prior to the experiment termination and the censoring in $\boldsymbol{p}_{(4)}$ only occurs at the left-most and the right-most. The initial values of α and λ for iterative progresses of MLE, mid-point approximation, EM algorithm, moment method and probability plot are given the same values, which for each simulation run, is randomly generated.

3.6. Simulation results. The result for the 1000 simulation runs by R software is shown in Table 1 and Table 2 and is graphically illustrated in Figures 4 and 5. As the performances among the four censoring schemes, the third scheme $\boldsymbol{p}_{(3)}$ provides the most precise results as seen from "Bias", "SD" (i.e. the standard deviation) and "MSE" (i.e. the mean squared errors) shown in Table 1 and Table 2 from the dispersions of the boxplots shown in the Figures 1 and 2, then followed by the schemes $\boldsymbol{p}_{(4)}$, $\boldsymbol{p}_{(1)}$ and $\boldsymbol{p}_{(2)}$.

4. Real data analysis

A data set which consists of 118 patients with breast cancer treated at the Sadouqi Hospital of Yazd is used for modelling the MOGE distribution; This data set is explored from [4] and summarized in Table 3. In this table, the first column shows 7 pre-assigned time intervals in years which were determined before the experiment, i.e., $[t_{i-1}, t_i)$, $i = 1, ..., 7$. The second column shows the number of patients who are died in the time intervals, i.e., $X_1, ..., X_7$ and finally, the last column is the number of patients who were dropped out from the study at the right end of each time interval; These dropped patients are known to be survived at the right end of each time interval but no follow up. Hence, the last column in Table 3 provides the values of R_i, $i = 1, \ldots, m = 7$.

Scheme		EM	Midpoint	MLE	MME	probpt
1	Median	0.4813	0.9553	0.5468	0.3978	0.4272
2	Median	0.4307	1.0000	0.5624	0.3858	0.4614
3	Median	0.5036	0.7766	0.5640	0.4640	0.5983
4	Median	0.4985	0.5783	0.3517	0.4316	0.5639
1	Mean	0.5080	0.8411	0.5740	0.3982	0.5193
2	Mean	0.4809	0.9020	0.5883	0.3860	0.5328
3	Mean	0.5237	0.7696	0.6223	0.4706	0.6100
4	Mean	0.5265	0.6433	0.3708	0.4318	0.5912
1	Bias	0.0080	0.3411	0.0740	-0.1018	0.0193
2	Bias	-0.0191	0.4020	0.0883	-0.1140	0.0328
3	Bias	0.0237	0.2696	0.1223	-0.0294	0.1100
4	Bias	0.0265	0.1433	-0.1292	-0.0682	0.0912
1	SD	0.2136	0.2024	0.2777	0.0071	0.3376
2	SD	0.2124	0.1829	0.3110	0.0034	0.3382
3	SD	0.1775	0.2217	0.2683	0.0403	0.2746
4	SD	0.1858	0.2966	0.0979	0.0166	0.2829
1	MSE	0.0457	0.1573	0.0826	0.0104	0.1144
2	MSE	0.0455	0.1950	0.1045	0.0130	0.1155
3	MSE	0.0321	0.1219	0.0869	0.0025	0.0875
4	MSE	0.0352	0.1085	0.0263	0.0049	0.0884

Table 1. Estimates of α from 1000 simulations for the five estimation methods and four simulation schemes.

4.1. Model selection. To select a suitable model for the given data set in Table 3, we start with the MOGE distribution. We will fit the MOGE distribution and statistically test whether the MOGE distribution model can be reduced to the exponential (E) distribution model for the given data set in the Table 3.

Fitting the MOGE to the given data, MLE of (α, λ) is

$$(\hat{\alpha}, \hat{\lambda}) = (0.05785, 0.52959),$$

$-2\log L(MOGE) = 137.4273$ and $AIC(MOGE) = 141.4273$. Then we fit the exponential distribution model to the given data set, MLE of λ is $\hat{\lambda} = 2.99422$, $-2\log L(E) = 152.0508$ and $AIC(E) = 154.0508$. Note that $AIC(MOGE) < AIC(E)$ and also, the log-likelihood ratio statistic is

$$-2\log(\Lambda) = (-2\log L(E)) - (-2\log L(MOGE)) = 14.6235,$$

which is greater than $\chi^2_{0.05}(1) = 3.8415$, hence the MOGE distribution provides a better fit for the data at size 0.05; indeed, the p-value of the test is 0.00013!

Additional model fitting to the GE distribution yields the estimated parameters $(\hat{\alpha}, \hat{\lambda}) = (0.19251, 1.03246)$ and $-2\log L(GE) = 138.1842$, so $AIC(GE) = 142.1842$.

4.2. Conclusion. In this paper, three methods to estimate the parameters of the MOGE distribution under progressive Type-I interval censoring have been developed; These methods were maximum likelihood estimation, estimation of method moments and the estimation based on the probability plot. The simulation study in the case of moderate

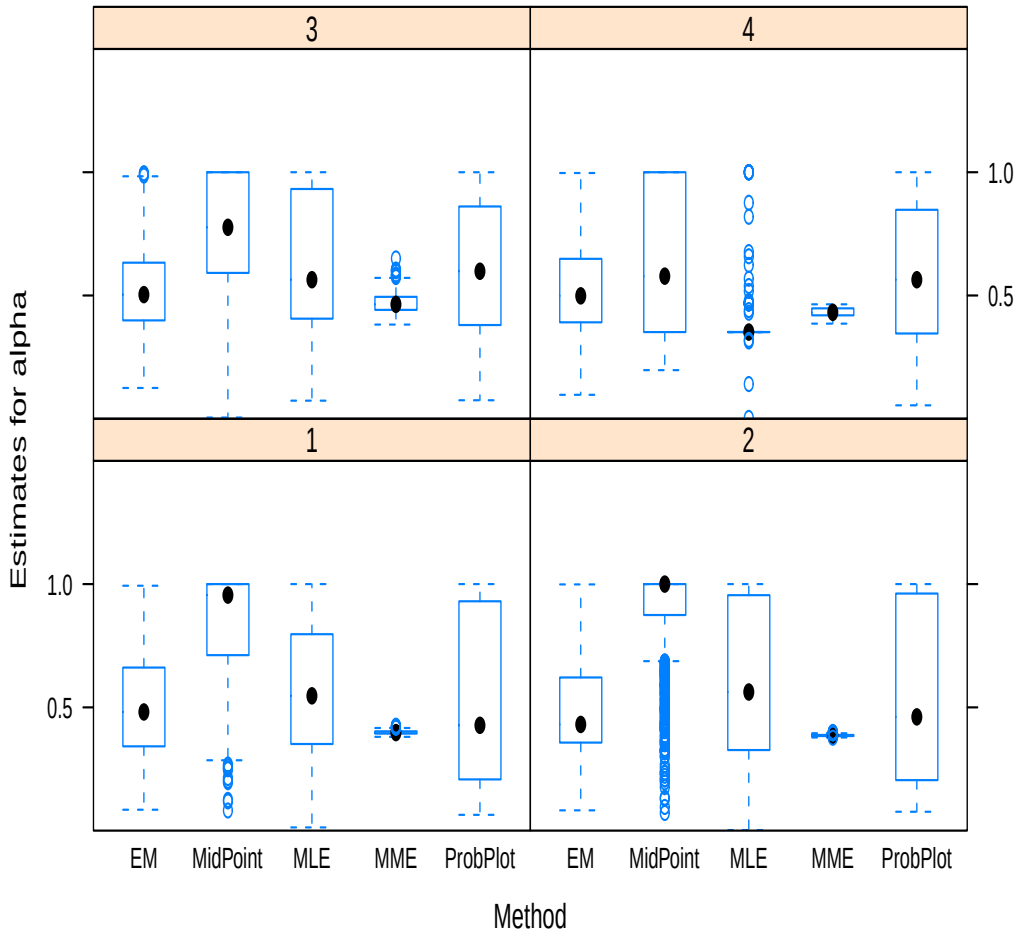

Figure 4. Boxplots for α from 1000 simulations for the five estimation methods and four simulation schemes

large size data set indicated that all these estimators give relatively accurate parameter estimation and the maximum likelihood estimator gives the most precise estimation as summarized in the Table 1 and 2 and Figures 4 and 5. We therefore recommend the "MLE" to be used to estimate the parameters in the MOGE distribution under progressive Type-I interval censoring. In the end of the paper, a real data set based on patients with breast cancer in order to demonstrate the applicabilities was used. Table 4 showed high flexibility of the MOGE distribution to model the data.

Acknowledgment

Scheme		EM	Midpoint	MLE	MME	probpt
1	Median	0.06267	0.08473	0.06333	0.02206	0.05194
2	Median	0.05943	0.09222	0.06375	0.01634	0.05005
3	Median	0.06125	0.06425	0.05359	0.02987	0.06601
4	Median	0.06142	0.04903	0.04344	0.0267	0.06359
1	Mean	0.06625	0.0821	0.06254	0.02236	0.05634
2	Mean	0.07151	0.089	0.08204	0.01659	0.05631
3	Mean	0.06317	0.06403	0.05481	0.03016	0.06446
4	Mean	0.06438	0.05095	0.04309	0.02691	0.06391
1	Bias	0.00625	0.0221	0.00254	-0.03764	-0.00366
2	Bias	0.01151	0.029	0.02204	-0.04341	-0.00369
3	Bias	0.00317	0.00403	-0.00519	-0.02984	0.00446
4	Bias	0.00438	-0.00905	-0.01691	-0.03309	0.00391
1	SD	0.02626	0.01911	0.02387	0.00406	0.0296
2	SD	0.03347	0.03598	0.01973	0.27499	0.0043
3	SD	0.01657	0.01194	0.0136	0.00345	0.01817
4	SD	0.02836	0.01926	0.01341	0.014	0.00374
1	MSE	0.00089	0.00073	0.00085	0.00058	0.00143
2	MSE	0.00113	0.00143	0.00123	0.0761	0.0019
3	MSE	0.00035	0.00028	0.00016	0.00021	9e-04
4	MSE	0.00082	0.00039	0.00026	0.00048	0.00111

Table 2. Estimates of λ from 1000 simulations for the five estimation methods and four simulation schemes.

Intervals	Number of failures	Number of Withdraws
$[0, 0.5)$	99	4
$[0.5, 1.0)$	8	2
$[1.0, 1.5)$	3	0
$[1.5, 2.0)$	1	0
$[2.0, 2.5)$	0	0
$[2.5, 3.0)$	0	0
$[3.0, 3.5)$	0	1

Table 3. Breast cancer survival times.

Distribution	$\hat{\alpha}$	$\hat{\lambda}$	$-2\log\hat{L}$	AIC	BIC	CAIC
E	-	2.99422	152.0508	154.0508	156.8215	154.0853
GE	0.19251	1.03246	138.1842	142.1842	147.7256	142.2885
MOGE	0.05785	0.52959	137.4273	141.4273	146.9687	141.5316

Table 4. Comparison of the E, GE and MOGE distributions

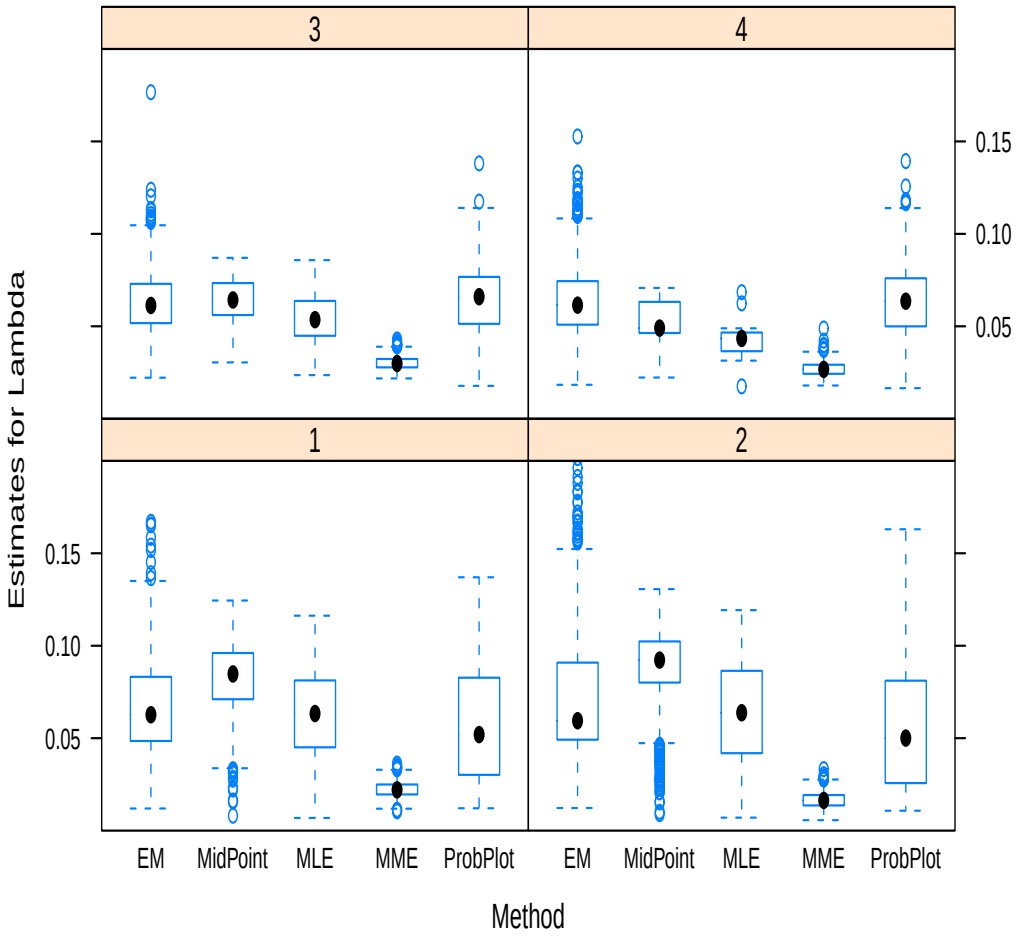

Figure 5. Boxplot for λ from 1000 simulations for the five estimation methods and four simulation schemes

The authors are highly grateful to referees for their valuable comments and suggestions for improving the paper.

References

[1] Aggarwala, R. *Progressively interval censoring Some mathematical results with application to inference*, Communications in Statistics-Theory and Methods, **30**, 1921-1935, 2001.

[2] Alice, T. and Jose, K.K. *On Marshall- Olkin Generalized Exponential Distribution and its Applications*. The Second Annual Conference of the Society of Statistics

and Computer Applications held at St. Thomas College, Pala, Kerala during 28th November to 1st December 1999.

[3] Chen, D.G. and Lio, Y.L. *Parameter estimations for generalized exponential distribution under progressive type-I interval censoring*, Computational Statistics and Data Analysis, **54**, 1581-1591, 2010.

[4] Fallahzadeh, H. *Five Year Survival of Women with Breast Cancer in Sadouqi Hospital of Yazd*. Technical Report, 2014.

[5] Gupta, R.D. and Kundu, D. *Exponentiated exponential distribution: An alternative to gamma and Weibull distributions*, Biometrical Journal, **43**, 117-130, 2001.

[6] Lawless, J. *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York, 1982

[7] Marshall, A.W. and Olkin, I. *A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families*, Biometrika, **84**, 641-652, 1997.

[8] Ng, H. and Wang, Z. *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, 2009.