

Article

EM Algorithm for Mixture Distributions Model with Type-I Hybrid Censoring Scheme

Tzong-Ru Tsai ^{1,†} , Yuhlong Lio ^{2,*,†}  and Wei-Chen Ting ^{1,†}

¹ Department of Statistics, Tamkang University, Tamsui, New Taipei City 251301, Taiwan; tzongru@gms.tku.edu.tw (T.-R.T.); 608650189@gms.tku.edu.tw (W.-C.T.)

² Department of Mathematical Sciences, University of South Dakota, Vermillion, SD 57069, USA

* Correspondence: Yuhlong.lio@usd.edu

† These authors contributed equally to this work.

Abstract: An expectation–maximization (EM) likelihood estimation procedure is proposed to obtain the maximum likelihood estimates of the parameters in a mixture distributions model based on type-I hybrid censored samples when the mixture proportions are unknown. Three bootstrap methods are applied to construct the confidence intervals of the model parameters. Monte Carlo simulations are conducted to evaluate the performance of the proposed methods. Simulation results show that the proposed methods can perform well to obtain reliable point and interval estimation results. Three examples are used to illustrate the applications of the proposed methods.

Keywords: bootstrap method; EM algorithm; maximum likelihood estimation; mixture distributions model; Monte Carlo simulation



Citation: Tsai, T.-R.; Lio, Y.; Ting, W.-C. EM Algorithm for Mixture Distributions Model with Type-I Hybrid Censoring Scheme. *Mathematics* **2021**, *9*, 2483. <https://doi.org/10.3390/math9192483>

Academic Editor: Christophe Chesneau

Received: 31 August 2021
Accepted: 29 September 2021
Published: 4 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The finite mixture probability models have been applied to model heterogeneity in a dataset. For example, Peel and McLachlan [1] proposed a mixture of t distributions with the application of clustering multivariate data that contains atypical observations. For more information, the reader may refer to the book by McLachlan and Peel [2]. Reliability inference is one of the important issues in industrial management to maintain the good quality of productions. In many practical cases, source items are coming from different suppliers, which have been certified, to support the company's production line. Hence, it is imperative to investigate the mixture type probability model for life testing to deal with the heterogeneity from different resources in industrial reliability analysis. For example, Sultan et al. [3] explored the mixture of two inverse Weibull distributions and provided some important properties, and Razali and Salih [4] provided a mixture of two Weibull distributions to analyze the lifetimes of electronic components. Ruhi et al. [5] investigated the mixture probability model using Weibull distributions as components and provided a case study.

Recent technology advancement has successfully enhanced product reliability. Spending a long time collecting complete lifetime samples from life tests to conduct reliability inference becomes unrealistic under the considerations of the scheduled test time and budget. Different censoring schemes have been adopted in the literature for life testing to conduct reliability inferences; see Balakrishnan and Aggarwala [6] and Balakrishnan and Cramer [7] for comprehensive information. Among all censoring schemes developed in industry life testing, the type-I and type-II censoring schemes have earned more attention because of easy implementation. Place n items on a life test at the same initial time and denote all failure times by X_1, X_2, \dots, X_n , respectively. Without loss of generality, let the initial time be 0. The life test continues up to the scheduled time τ_1 for the type-I censoring scheme; while the life test under the type-II censoring scheme is carried out until a preassigned number, r ($\leq n$), of failures is observed. Let $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ be

the ordered lifetimes of the sample $\{X_1, X_2, \dots, X_n\}$. Then the type-II censoring scheme terminates the life test at lifetime $X_{(r)}$. Mendenhall and Hader [8] proposed approximate maximum likelihood estimators (MLEs) of the mixture exponential distributions based on a type II censored sample; while Jones and Ashour [9] proposed Bayesian estimation method for the mixture exponential distribution parameters to compare with the MLE results. We use the term MLE to denote maximum likelihood estimator/estimate here and after. Amin [10] proposed a maximum likelihood estimation method for the mixture generalized Rayleigh (GR) distribution under the type-I censoring scheme.

Epstein [11] presented the type-I hybrid censoring (TIHC) scheme where the life test is ceased at a random time $\tau = \min\{X_{(r)}, \tau_1\}$ and Childs et al. [12] considered the type-II hybrid censoring scheme where the life test is stopped at a random time $\tau^* = \max\{X_{(r)}, \tau_1\}$. Balakrishnan and Kundu [13] provided comprehensive discussions regarding the applications of hybrid censoring schemes. More reliability applications about using hybrid censoring schemes can be found in [14,15]. Based on our best knowledge, no literature deals with the expectation–maximization MLE (EM-MLE) estimation methods for the mixture distributions model utilizing TIHC samples. In this study, we focus on the applications of using the TIHC scheme for the reliability inferences of the mixture distributions model. Let random variable D indicate the number of failures by the time τ_1 for the TIHC scheme, then,

$$D = \begin{cases} r & \text{if } X_{(r)} \leq \tau_1, \\ m & \text{if } X_{(m)} \leq \tau_1 < X_{(m+1)} \leq \dots \leq X_{(r)}. \end{cases}$$

Moreover, let τ denote the censored time of the TIHC scheme; it can be shown that:

$$\tau = \begin{cases} X_{(r)} & \text{if } X_{(r)} \leq \tau_1, \\ \tau_1 & \text{if } \tau_1 < X_{(r)}. \end{cases}$$

Hence, the TIHC sample with the scheduled r and τ can be denoted by

$$X = \{X_1, X_2, \dots, X_D, rep(\tau, n - D)\},$$

where $rep(\tau, n - D)$ indicates $n - D$ items withdrawn at time τ . The EM algorithm is simple to use, and hence the EM algorithm has been widely applied to obtain the MLEs of the model parameters for censoring data and missing data. When the mixture proportions and the data are type-I hybrid censored, it is not easy to obtain the MLE of the model parameters. If the mixture proportions are known, the EM algorithm surely has a better performance than the cases for which the mixture proportions are unknown.

The rest of this article will be structured as follows. Section 2 addresses the likelihood function using the TIHC sample from the finite mixture distribution. The EM algorithm to find the MLEs of the model parameters and survival function, as well as three bootstrap procedures for confidence intervals, are presented in Section 2. In Section 3, a simulation study will be conducted under the mixture model of two-component distributions with Weibull, generalized exponential (GE) and GR distributions. The simulation results will be discussed in Section 3. Moreover, one numerical TIHC sample is used for illustration. Section 4 addresses applications by using three examples and Section 5 provides concluding remarks.

2. Maximum Likelihood Estimation

Let the probability density function (PDF) of lifetimes be $f(x|\theta)$ and the cumulative distribution function (CDF) be $F(x|\theta)$, where θ is the vector of parameters. Let the realizations of $X_{(1)}, X_{(2)}, \dots, X_{(D)}$ be denoted by $x_{(1)}, x_{(2)}, \dots, x_{(D)}$. The likelihood function based on $x = \{x_{(1)}, x_{(2)}, \dots, x_{(D)}, rep(\tau, n - D)\}$ can be represented by:

$$L(\theta|x) = \frac{n!}{(n - D)!} \prod_{i=1}^D f(x_{(i)}|\theta) \{S(\tau|\theta)\}^{n-D}, \tag{1}$$

and the log-likelihood function can be presented as:

$$\ell(\theta) = \ln\left(\frac{n!}{(n-D)!}\right) + \sum_{i=1}^D \ln\left(f(x_{(i)}|\theta)\right) + (n-D)\ln(S(\tau|\theta)), \tag{2}$$

where $S(\tau|\theta) = 1 - F(\tau|\theta)$.

In practical applications, the lifetime quality of items could not be consistent. For example, if the items are provided by multiple suppliers and the lifetime quality of items from those suppliers are inconsistent. In the aforementioned situation, a mixture distributions model can be a good candidate for characterizing the lifetimes of items.

2.1. Likelihood Function

Let the lifetimes be taken from a mixture distributions model with k groups and the PDFs and CDFs of the k groups be $g_j(x|\theta_j)$ and $G_j(x|\theta_j)$ for $j = 1, 2, \dots, k$. The PDF and CDF of the mixture distributions model can be represented by:

$$f(x|\Theta) = \sum_{j=1}^k \delta_j g_j(x|\theta_j) \tag{3}$$

and

$$F(x|\Theta) = \sum_{j=1}^k \delta_j G_j(x|\theta_j), \tag{4}$$

respectively, where $0 \leq \delta_j \leq 1$ is the proportion parameter to indicate the likelihood of the random variable x from the j th distribution, $\sum_{j=1}^k \delta_j = 1$ and $\Theta = (\theta_1, \theta_2, \dots, \theta_k, \delta_1, \delta_2, \dots, \delta_k)$. Replacing $f(x|\Theta)$ and $F(x|\Theta)$ with Equations (3) and (4), Equations (1) and (2) can be respectively represented by:

$$L(\Theta|x) \propto \prod_{i=1}^D \left(\sum_{j=1}^k \delta_j g_j(x_i|\theta_j)\right) \left(1 - \sum_{j=1}^k \delta_j G_j(\tau|\theta_j)\right)^{n-D} \tag{5}$$

and

$$\ell(\Theta) \propto \sum_{i=1}^D \ln\left(\sum_{j=1}^k \delta_j g_j(x_i|\theta_j)\right) + (n-D)\ln\left(1 - \sum_{j=1}^k \delta_j G_j(\tau|\theta_j)\right). \tag{6}$$

The normal equation is obtained as $\nabla\ell(\Theta) \equiv \mathbf{0}$, where $\nabla\ell(\Theta)$ is a gradient of $\ell(\Theta)$ with respect to Θ , and $\mathbf{0}$ is the zero vector with the same dimension of $\nabla\ell(\Theta)$. The MLE $\hat{\Theta}$ of Θ is the solution of $\nabla\ell(\Theta) \equiv \mathbf{0}$. Because of the complicity of the normal equation, the solution procedure is not tractable. The following EM algorithm is suggested to find the MLE $\hat{\Theta}$ of Θ .

2.2. EM Algorithm

For $i = 1, 2, \dots, n$, let ϵ_i be the indicator of the lifetime of the i th tested item being censored or not and defined as:

$$\epsilon_i = \begin{cases} 1 & \text{if } X_i < \tau, \\ 0 & \text{if } X_i \geq \tau. \end{cases}$$

If X_i is taken from the j th distribution, then the contribution to the joint PDF can be presented by:

$$f_i(\Theta|x_i) = \delta_j [g_j(x_i|\theta_j)]^{\epsilon_i} [1 - G_j(\tau|\theta_j)]^{1-\epsilon_i}, \quad i = 1, 2, \dots, n.$$

Define latent variable by $y_{ij} = 1$ if x_i is from the j th distribution; otherwise, $y_{ij} = 0$ for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, n$. Then, the likelihood function of Equation (5) and log-likelihood function of Equation (6) can be respectively represented as follows:

$$L(\Theta|\mathbf{x}) \propto \prod_{i=1}^n \prod_{j=1}^k \left\{ \delta_j [g_j(x_i|\boldsymbol{\theta}_j)]^{\epsilon_i} [1 - G_j(\tau|\boldsymbol{\theta}_j)]^{1-\epsilon_i} \right\}^{y_{ij}} \tag{7}$$

and

$$\ell(\Theta) \propto \sum_{i=1}^n \sum_{j=1}^k y_{ij} \{ \ln(\delta_j) + \epsilon_i \ln(g_j(x_i|\boldsymbol{\theta}_j)) + (1 - \epsilon_i) \ln(1 - G_j(\tau|\boldsymbol{\theta}_j)) \}. \tag{8}$$

The following two steps of the EM algorithm can be iteratively used to obtain the MLEs of the model parameters.

E-step: The conditional expected value of $\ell(\Theta)|\mathbf{x}$ is calculated and labeled by:

$$\begin{aligned} H(\Theta, \boldsymbol{\eta}, \boldsymbol{\delta}) &= E[\ell(\Theta)|\mathbf{x}] \\ &= \sum_{i=1}^n \sum_{j=1}^k \eta_{ij} \{ \ln(\delta_j) + \epsilon_i \ln(g_j(x_i|\boldsymbol{\theta}_j)) + (1 - \epsilon_i) \ln(1 - G_j(\tau|\boldsymbol{\theta}_j)) \}, \end{aligned} \tag{9}$$

where $\eta_{ij} = E[y_{ij}|\mathbf{x}]$ indicates the expected value of the i th sample observation from the j th distribution under right censored at τ , $\boldsymbol{\eta} = \{\eta_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, k\}$ and $\boldsymbol{\delta} = \{\delta_j, j = 1, 2, \dots, k\}$. During the E-step, for each i in $\{1, 2, \dots, n\}$, η_{ij} can be derived by utilizing the conditional probability concept and the result is given as:

$$\eta_{ij} = \frac{\delta_j [\epsilon_i g_j(x_i|\boldsymbol{\theta}_j) + (1 - \epsilon_i)(1 - G_j(\tau|\boldsymbol{\theta}_j))]}{\sum_{m=1}^k \delta_m [\epsilon_i g_m(x_i|\boldsymbol{\theta}_m) + (1 - \epsilon_i)(1 - G_m(\tau|\boldsymbol{\theta}_m))]}, \quad j = 1, 2, \dots, k. \tag{10}$$

M-step: There are two steps given as follows:

Step 1: Find δ_{j_0} by taking partial derivative of $H(\Theta, \boldsymbol{\eta}, \boldsymbol{\delta})$ with respect to δ_{j_0} :

$$\frac{\partial H(\Theta, \boldsymbol{\eta}, \boldsymbol{\delta})}{\partial \delta_{j_0}} = \sum_{i=1}^n \eta_{ij_0} \frac{1}{\delta_{j_0}} - \sum_{i=1}^n \eta_{ik} \frac{1}{\delta_k} \tag{11}$$

for $1 \leq j_0 < k$. Let $\frac{\partial H(\Theta, \boldsymbol{\eta}, \boldsymbol{\delta})}{\partial \delta_{j_0}} = 0$, we obtain

$$\delta_k \sum_{i=1}^n \eta_{ij_0} = \delta_{j_0} \sum_{i=1}^n \eta_{ik}. \tag{12}$$

Taking summation for the two sides of Equation (12), we have:

$$\delta_k \sum_{j_0=1}^{k-1} \sum_{i=1}^n \eta_{ij_0} = \sum_{j_0=1}^{k-1} \delta_{j_0} \sum_{i=1}^n \eta_{ik},$$

and then we obtain

$$\delta_k (n - \sum_{i=1}^n \eta_{ik}) = (1 - \delta_k) \sum_{i=1}^n \eta_{ik}. \tag{13}$$

Based on Equation (13), we obtain

$$\delta_k = \frac{1}{n} \sum_{i=1}^n \eta_{ik}. \tag{14}$$

Plug Equation (14) into Equation (12) to get

$$\delta_{j_0} = \frac{1}{n} \sum_{i=1}^n \eta_{ij_0}, \quad \text{for } 1 \leq j_0 < k. \tag{15}$$

Step 2: Find the MLE $\hat{\Theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$ by

$$\hat{\Theta} = \arg \max_{\theta_1, \theta_2, \dots, \theta_k} H(\Theta, \eta, \delta).$$

The EM algorithm is implemented based on the E-step and M-step until convergence as follows:

At step s ,

E-step: Update η_{ij} by:

$$\eta_{ij}^{(s)} = \frac{\delta_j^{(s-1)} [\epsilon_i g_j(x_i | \theta_j^{(s-1)}) + (1 - \epsilon_i)(1 - G_j(\tau | \theta_j^{(s-1)}))]}{\sum_{m=1}^k \delta_m^{(s-1)} [\epsilon_i g_m(x_i | \theta_m^{(s-1)}) + (1 - \epsilon_i)(1 - G_m(\tau | \theta_m^{(s-1)}))],}$$

for $j = 1, 2, \dots, k$ and $i = 1, 2, \dots, n$.

M-step: Update δ_j and Θ , respectively, by

$$\delta_j^{(s)} = \frac{1}{n} \sum_{i=1}^n \eta_{ij}^{(s)}, \quad j = 1, 2, \dots, k$$

and

$$\Theta^{(s)} = \arg \max_{\theta_1, \theta_2, \dots, \theta_k} H(\Theta^{(s-1)}, \eta^{(s)}, \delta^{(s)}).$$

2.3. Applications

Three common used lifetime distributions of Weibull, GE and GR are used for illustration. Without loss of generality, it is assumed that the mixture distributions model has two components. For simplification, let $H \equiv H(\Theta, \eta, \delta)$ here and after.

2.3.1. The Mixture Weibull Distributions

Consider two Weibull distributions as the members for mixture, the PDF and CDF are respectively given by:

$$g_j(x | \theta_j) = \frac{\alpha_j}{\lambda_j} \left(\frac{x}{\lambda_j}\right)^{\alpha_j - 1} e^{-\left(\frac{x}{\lambda_j}\right)^{\alpha_j}}, \quad x > 0,$$

and

$$G_j(x | \theta_j) = 1 - e^{-\left(\frac{x}{\lambda_j}\right)^{\alpha_j}}, \quad x > 0,$$

where α_j is shape parameter, λ_j is scale parameter and $j = 1, 2$. Hence,

$$H = \sum_{i=1}^n \sum_{j=1}^2 \eta_{ij} \left(\ln(\delta_j) + \epsilon_i \left(\ln(\alpha_j) - \alpha_j \ln(\lambda_j) + (\alpha_j - 1) \ln(x_i) - \left(\frac{x_i}{\lambda_j}\right)^{\alpha_j} \right) - (1 - \epsilon_i) \left(\frac{\tau}{\lambda_j}\right)^{\alpha_j} \right),$$

$$\frac{\partial H}{\partial \alpha_j} = \sum_{i=1}^n \eta_{ij} \left[\epsilon_i \left(\frac{1}{\alpha_j} + \ln\left(\frac{x_i}{\lambda_j}\right) - \left(\frac{x_i}{\lambda_j}\right)^{\alpha_j} \ln\left(\frac{x_i}{\lambda_j}\right) \right) - (1 - \epsilon_i) \left(\frac{\tau}{\lambda_j}\right)^{\alpha_j} \ln\left(\frac{\tau}{\lambda_j}\right) \right], \quad (16)$$

$$\frac{\partial H}{\partial \lambda_j} = \sum_{i=1}^n \eta_{ij} \left[\epsilon_i \left(-\frac{\alpha_j}{\lambda_j} + \alpha_j x_i^{\alpha_j} (\lambda_j)^{-\alpha_j - 1} \right) + \alpha_j (1 - \epsilon_i) \tau^{\alpha_j} (\lambda_j)^{-\alpha - 1} \right], \quad (17)$$

and

$$\delta_j = \frac{1}{n} \sum_{i=1}^n \eta_{ij}, \quad j = 1, 2.$$

The MLEs of α_j and λ_j , $j = 1, 2$ are the solutions of the system by letting the Equations (16) and (17) be 0.

2.3.2. The Mixture GE Distributions

Consider two GE distributions for mixture, the PDF and CDF are respectively given by

$$g_j(x|\theta_j) = \alpha_j \lambda_j e^{-x\lambda_j} (1 - e^{-x\lambda_j})^{\alpha_j - 1}, \quad x > 0,$$

and

$$G_j(x|\theta_j) = (1 - e^{-x\lambda_j})^{\alpha_j}, \quad x > 0,$$

where α_j is shape parameter, λ_j is scale parameter and $j = 1, 2$. Hence,

$$H = \sum_{i=1}^n \sum_{j=1}^2 \eta_{ij} \left(\ln(\delta_j) + \epsilon_i \left(\ln(\alpha_j) + \ln(\lambda_j) - \lambda_j x_i + (\alpha_j - 1) \ln(1 - e^{-x_i \lambda_j}) \right) + (1 - \epsilon_i) \ln(1 - (1 - e^{-\tau \lambda_j})^{\alpha_j}) \right),$$

$$\frac{\partial H}{\partial \alpha_j} = \sum_{i=1}^n \eta_{ij} \left[\epsilon_i \left(\frac{1}{\alpha_j} + \ln(1 - e^{-x_i \lambda_j}) \right) - (1 - \epsilon_i) \frac{(1 - e^{-\tau \lambda_j})^{\alpha_j} \ln(1 - e^{-\tau \lambda_j})}{1 - (1 - e^{-\tau \lambda_j})^{\alpha_j}} \right], \quad (18)$$

$$\begin{aligned} \frac{\partial H}{\partial \lambda_j} &= \sum_{i=1}^n \eta_{ij} \left[\epsilon_i \left(\frac{1}{\lambda_j} - x_i + \frac{(\alpha_j - 1)x_i e^{-x_i \lambda_j}}{1 - e^{-x_i \lambda_j}} \right) - (1 - \epsilon_i) \frac{\tau \alpha_j e^{-\tau \lambda_j} (1 - e^{-\tau \lambda_j})^{\alpha_j - 1}}{1 - (1 - e^{-\tau \lambda_j})^{\alpha_j}} \right], \end{aligned} \quad (19)$$

and

$$\delta_j = \frac{1}{n} \sum_{i=1}^n \eta_{ij}, \quad j = 1, 2.$$

The MLEs of α_j and λ_j , $j = 1, 2$ are the solutions of the system by letting the Equations (18) and (19) to 0.

2.3.3. The Mixture GR Distributions

Consider two GR distributions for mixture, the PDF and CDF are respectively given by

$$g_j(x|\theta_j) = 2\alpha_j \lambda_j^2 x e^{-x^2 \lambda_j^2} (1 - e^{-x^2 \lambda_j^2})^{\alpha_j - 1}, \quad x > 0,$$

and

$$G_j(x|\theta_j) = (1 - e^{-x^2 \lambda_j^2})^{\alpha_j}, \quad x > 0,$$

where α_j is shape parameter, λ_j is scale parameter and $j = 1, 2$. Hence,

$$H = \sum_{i=1}^n \sum_{j=1}^2 \eta_{ij} \left(\ln(\delta_j) + \epsilon_i \left(\ln(2x_i \alpha_j) + 2 \ln(\lambda_j) - x_i^2 \lambda_j^2 + (\alpha_j - 1) \ln(1 - e^{-x_i^2 \lambda_j^2}) \right) + (1 - \epsilon_i) \ln(1 - (1 - e^{-\tau^2 \lambda_j^2})^{\alpha_j}) \right),$$

$$\frac{\partial H}{\partial \alpha_j} = \sum_{i=1}^n \eta_{ij} \left[\epsilon_i \left(\frac{1}{\alpha_j} + \ln(1 - e^{-x_i^2 \lambda_j^2}) \right) - (1 - \epsilon_i) \frac{(1 - e^{-\tau^2 \lambda_j^2})^{\alpha_j} \ln(1 - e^{-\tau^2 \lambda_j^2})}{1 - (1 - e^{-\tau^2 \lambda_j^2})^{\alpha_j}} \right], \quad (20)$$

$$\frac{\partial H}{\partial \lambda_j} = \sum_{i=1}^n \eta_{ij} \left[\epsilon_i \left(\frac{2}{\lambda_j} - 2x_i^2 \lambda_j + \frac{2\lambda_j(\alpha_j - 1)x_i^2 e^{-x_i^2 \lambda_j^2}}{1 - e^{-x_i^2 \lambda_j^2}} \right) - (1 - \epsilon_i) \frac{\tau^2 \alpha_j e^{-\tau^2 \lambda_j^2} 2\lambda_j (1 - e^{-\tau^2 \lambda_j^2})^{\alpha_j - 1}}{1 - (1 - e^{-\tau^2 \lambda_j^2})^{\alpha_j}} \right], \tag{21}$$

and

$$\delta_j = \frac{1}{n} \sum_{i=1}^n \eta_{ij}, \quad j = 1, 2.$$

The MLEs of α_j and λ_j , $j = 1, 2$ are the solutions of the system by letting the Equations (20) and (21) be 0.

Analogously, the results of mixture Weibull, mixture GE and mixture GR distributions can be extended to the mixture distributions with two different distributions; that is, the mixture Weibull and GE distributions, the mixture Weibull and GR distribution and the mixture GR and GE distributions. We skip the details here to save pages.

2.4. Maximum Likelihood Estimate of Survival Function

From Equation (4), the survival function at given time $x_0 > 0$ is defined by

$$S(x_0) \equiv S(x_0|\Theta) = 1 - F(x_0|\Theta) = \sum_{j=1}^k \delta_j (1 - G_j(x_0|\theta_j)). \tag{22}$$

Let $\hat{\Theta}$ denote the MLE of Θ by using the proposed EM algorithm. Then, the MLE of $S(x_0)$ can be obtained by the plug-in method and be labeled by:

$$\hat{S}(x_0) = \sum_{j=1}^k \hat{\delta}_j (1 - G_j(x_0|\hat{\theta}_j)).$$

It is difficult to obtain the confidence intervals of Θ and $S(x_0)$ based on the sampling distribution of $\hat{\Theta}$ and $\hat{S}(x_0)$. To overcome this difficulty, the parametric bootstrap percentile (PBP) procedure and two bootstrap correction methods are proposed to obtain approximate confidence intervals of Θ and $S(x_0)$, respectively.

(a) PBP procedure

Step 1. Under the TIHC scheme with r and τ_1 , a TIHC sample,

$$x = \{x_{(1)}, x_{(2)}, \dots, x_{(D)}, rep(\tau, n - D)\},$$

is collected from a mixture distribution;

Step 2. The MLEs $\hat{\Theta}$ based on TIHC sample from Step 1 is obtained by utilizing the proposed EM algorithm and the MLE $\hat{S}(x_0)$ is obtained via the plug-in method;

Step 3. A bootstrap TIHC sample with r and τ_1 is drawn from the same mixture distribution with the parameters Θ substituted by $\hat{\Theta}$. Let the bootstrap TIHC sample be denoted by $x^* = \{x_{(1)}^*, x_{(2)}^*, \dots, x_{(D)}^*, rep(\tau, n - D)\}$;

Step 4. The MLEs $\hat{\Theta}^*$ and $\hat{S}^*(x_0)$ are derived based on the bootstrap TIHC sample x^* from Step 3;

Step 5. Repeat Step 3 and Step 4 B times, where B is a given large positive number. Label all MLEs by $\hat{\Theta}_j^*$ and $\hat{S}_j^*(x_0)$ for $j = 1, 2, \dots, B$;

Step 6. Let $\hat{\omega}_j$, $j = 1, 2, \dots, B$ be a B bootstrap MLEs obtained in Step 5 for a parameter ω considered. The empirical distribution, \hat{G}_ω^B , of $\{\hat{\omega}_j, j = 1, 2, \dots, B\}$ can be obtained. Given $0 < \gamma < 1$, the γ th empirical quantile, $(\hat{G}_\omega^B)^{-1}(\gamma)$, is defined as the $\lfloor \gamma B \rfloor$ th order statistic, $\hat{\omega}_{(\lfloor \gamma B \rfloor)}^*$, of $\hat{\omega}_j$, $j = 1, 2, \dots, B$, where $\lfloor \gamma B \rfloor$ is the greatest positive integer less than or equal to γB . Then an $1 - \gamma$ bootstrap percentile confidence interval of ω can be obtained by $((\hat{G}_\omega^B)^{-1}(\frac{\gamma}{2}), (\hat{G}_\omega^B)^{-1}(1 - \frac{\gamma}{2}))$.

More information regarding PBP procedure, readers may refer to the books by Efron [16], Efron and Tibshirani [17] and Shao and Tu [18]. The two bootstrap correction methods are presented as follows:

- (b) **Hybrid bootstrap percentile (HBP) procedure** Shao and Tu [18] proposed the HBP procedure is provided as follows:

Step 1. Repeat Step 1 to Step 5 of the PBP procedure. Let $\hat{\omega}$ be the MLE utilizing the original TIHC sample. Moreover, let $\hat{\omega}_j^*$, $j = 1, 2, \dots, B$ be the B bootstrap MLEs of ω ;

Step 2. Let \hat{H}_B be the empirical distribution of $\hat{\phi}_j$, which is defined by $\hat{\phi}_j = \sqrt{n}(\hat{\omega}_j^* - \hat{\omega})$, $j = 1, 2, \dots, B$;

Step 3. Let $\hat{H}_B^{-1}(\gamma)$ be the γ th quantile of \hat{H}_B , where $0 < \gamma < 1$. An $1 - \gamma$ HBP confidence interval for ω can be obtained as the following interval:

$$\left(\hat{\omega} - n^{-1/2}\hat{H}_B^{-1}\left(1 - \frac{\gamma}{2}\right), \hat{\omega} - n^{-1/2}\hat{H}_B^{-1}\left(\frac{\gamma}{2}\right)\right).$$

The second bootstrap correction method that can be found in the books by Efron and Tibshirani [17], as well as in Shao and Tu [18], is addressed as follows:

- (c) **Bootstrap bias-corrected percentile (BCP) procedure**

Step 1. Implement the Step 1 of the HBP procedure. Let the empirical distribution, \hat{G}_ω^B , be constructed based on the bootstrap MLEs $\hat{\omega}_j^*$, $j = 1, 2, \dots, B$;

Step 2. An $1 - \gamma$ bootstrap BCP confidence interval for parameter ω can be obtained by:

$$\left((\hat{G}_\omega^B)^{-1}\left(\Phi\left[Z_{\frac{\gamma}{2}} + 2\Phi^{-1}\left(\hat{G}_\omega^B(\hat{\omega})\right)\right]\right), (\hat{G}_\omega^B)^{-1}\left(\Phi\left[Z_{1-\frac{\gamma}{2}} + 2\Phi^{-1}\left(\hat{G}_\omega^B(\hat{\omega})\right)\right]\right)\right),$$

where $\Phi(\cdot)$ is standard normal distribution CDF and $\Phi(Z_\gamma) = \gamma$.

3. Simulation Study

To evaluate the accuracy of the proposed EM algorithm to obtain the MLEs of model parameters using the censored sample under the TIHC scheme with r and τ_1 , an intensive simulation study will be conducted in this section. The simulation study will utilize three distributions that include Weibull, GE and GR distributions as the baseline component distributions for the two-component mixture distributions model; that is, $k = 2$. All computation work is obtained via using R codes. R is an open software and can be freely downloaded. Population parameters $(\alpha_1, \lambda_1, \alpha_2, \lambda_2) = (2, 1.5, 4, 2)$, mixture proportions $(\delta_1, \delta_2) = (0.6, 0.4)$ and $(0.75, 0.25)$, as well as the pre-required failure numbers $r = 0.6n$ and $0.75n$, are considered in this simulation study for the sample sizes of $n = 500$ and 1000 . τ_1 is taken as the maximum of the 85th percentiles of the two distributions for mixture; that is, $\tau_1 = \max\{G_1^{-1}(0.85|\theta_1), G_2^{-1}(0.85|\theta_1)\}$.

For each set of simulation inputs, the simulation study was conducted 1000 iterations. The maximal iteration for the convergence of EM algorithm is 50 and $B = 10,000$ is used for bootstrap sampling. To evaluate the quality of the MLE $\hat{\omega}$ of parameter ω , relative bias $rBias = \frac{Bias}{\omega}$ and relative squared root of mean square error $rsMSE = \frac{sMSE}{\omega}$ are used, where:

$$Bias = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\omega} - \omega)$$

and

$$sMSE = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} (\hat{\omega} - \omega)^2}.$$

The simulation results for the MLEs of the model parameters are displaced in Tables 1–3. Tables 1–3 show that almost all cells of $rBias$ and $rsMSE$ are reduced when either sample size n or r increases. For the Weibull mixture distributions model, the high survival function

value with small failure time input is underestimated by MLE and two other lower survival function values with large failure time inputs are overestimated by MLEs. For the GE mixture distributions model, the low survival function value with large failure time input is overestimated by MLE and the other two higher survival function values with two smaller failure time inputs are underestimated by MLEs. For the GR mixture distributions model, the survival function at median with the failure time input at median is overestimated by MLE and two other survival function values with either high failure time input or low failure time input are underestimated by MLEs. Generally, except for the small survival function values, the other two higher survival function values can be well estimated by MLEs with small rBias and rsMSE.

Table 1. Mixture Weibull distributions model with $\lambda_1 = 1.5, \alpha_1 = 2, \lambda_2 = 2, \alpha_2 = 4$ and $S(x_0) = 0.1, 0.5, 0.9$ for $(x_{01}, x_{02}, x_{03}) = (2.13, 1.098, 0.444)$.

n	r	δ		$\hat{\lambda}_1$	$\hat{\alpha}_1$	$\hat{\lambda}_2$	$\hat{\alpha}_2$	$\hat{S}(x_{01})$	$\hat{S}(x_{02})$	$\hat{S}(x_{03})$
500	0.6n	0.6	rBias	0.187	0.039	-0.106	-0.408	0.512	0.039	-0.043
			rsMSE	0.189	0.073	0.108	0.409	0.520	0.048	0.045
		0.75	rBias	0.119	0.007	-0.150	-0.433	0.438	0.061	-0.024
			rsMSE	0.123	0.056	0.151	0.434	0.457	0.066	0.027
	0.8n	0.6	rBias	0.151	0.092	-0.122	-0.376	0.380	0.053	-0.029
			rsMSE	0.153	0.107	0.123	0.377	0.385	0.059	0.031
		0.75	rBias	0.097	0.040	-0.159	-0.413	0.322	0.061	-0.017
			rsMSE	0.100	0.063	0.160	0.413	0.334	0.067	0.020
1000	0.6n	0.6	rBias	0.187	0.035	-0.106	-0.410	0.514	0.038	-0.043
			rsMSE	0.188	0.056	0.107	0.411	0.518	0.042	0.044
		0.75	rBias	0.119	0.004	-0.150	-0.435	0.440	0.060	-0.025
			rsMSE	0.121	0.040	0.151	0.435	0.449	0.063	0.026
	0.8n	0.6	rBias	0.151	0.088	-0.122	-0.378	0.381	0.052	-0.029
			rsMSE	0.152	0.096	0.122	0.379	0.384	0.055	0.030
		0.75	rBias	0.097	0.038	-0.159	-0.414	0.322	0.060	-0.017
			rsMSE	0.098	0.051	0.160	0.415	0.328	0.063	0.019

Table 2. Mixture GE distributions model with $\lambda_1 = 1.5, \alpha_1 = 2, \lambda_2 = 2, \alpha_2 = 4$ and $S(x_0) = 0.1, 0.5, 0.9$ for $(x_{01}, x_{02}, x_{03}) = (1.9, 0.862, 0.305)$.

n	r	δ		$\hat{\lambda}_1$	$\hat{\alpha}_1$	$\hat{\lambda}_2$	$\hat{\alpha}_2$	$\hat{S}(x_{01})$	$\hat{S}(x_{02})$	$\hat{S}(x_{03})$
500	0.6n	0.6	rBias	0.026	0.164	-0.129	-0.340	0.080	-0.002	-0.003
			rsMSE	0.078	0.197	0.141	0.345	0.164	0.039	0.013
		0.75	rBias	0.014	0.095	-0.141	-0.382	0.058	-0.003	-0.001
			rsMSE	0.073	0.135	0.152	0.386	0.156	0.039	0.013
	0.8n	0.6	rBias	0.045	0.180	-0.120	-0.330	0.043	-0.009	-0.002
			rsMSE	0.076	0.206	0.128	0.334	0.124	0.037	0.013
		0.75	rBias	0.026	0.105	-0.135	-0.375	0.030	-0.008	-0.001
			rsMSE	0.065	0.137	0.142	0.378	0.121	0.037	0.013
1000	0.6n	0.6	rBias	0.022	0.157	-0.132	-0.344	0.083	-0.001	-0.003
			rsMSE	0.056	0.175	0.138	0.346	0.132	0.027	0.010
		0.75	rBias	0.010	0.090	-0.144	-0.385	0.061	-0.001	-0.001
			rsMSE	0.051	0.112	0.149	0.386	0.119	0.028	0.009
	0.8n	0.6	rBias	0.041	0.174	-0.122	-0.333	0.046	-0.008	-0.002
			rsMSE	0.060	0.188	0.127	0.335	0.096	0.027	0.009
		0.75	rBias	0.023	0.101	-0.137	-0.377	0.033	-0.007	-0.001
			rsMSE	0.048	0.118	0.141	0.378	0.091	0.026	0.009

Table 3. Mixture GR distributions model with $\lambda_1 = 1.5, \alpha_1 = 2, \lambda_2 = 2, \alpha_2 = 4$ and $S(x_0) = 0.1, 0.5, 0.9$ for $(x_{01}, x_{02}, x_{03}) = (1.074, 0.708, 0.429)$.

n	r	δ		$\hat{\lambda}_1$	$\hat{\alpha}_1$	$\hat{\lambda}_2$	$\hat{\alpha}_2$	$\hat{S}(x_{01})$	$\hat{S}(x_{02})$	$\hat{S}(x_{03})$
500	0.6n	0.6	rBias	0.081	0.208	-0.101	-0.289	-0.076	0.001	-0.003
			rsMSE	0.090	0.237	0.105	0.296	0.157	0.040	0.013
	0.75	rBias	0.050	0.126	-0.125	-0.341	-0.062	<0.001	-0.002	
		rsMSE	0.063	0.160	0.128	0.345	0.153	0.040	0.013	
	0.8n	0.6	rBias	0.078	0.191	-0.111	-0.296	-0.061	0.005	-0.004
			rsMSE	0.085	0.215	0.113	0.302	0.132	0.037	0.013
1000	0.6n	0.6	rBias	0.079	0.200	-0.103	-0.293	-0.072	0.002	-0.003
			rsMSE	0.083	0.215	0.105	0.297	0.121	0.028	0.010
	0.75	rBias	0.049	0.120	-0.126	-0.344	-0.060	0.002	-0.002	
		rsMSE	0.056	0.139	0.127	0.346	0.115	0.028	0.009	
	0.8n	0.6	rBias	0.076	0.184	-0.112	-0.300	-0.057	0.006	-0.004
			rsMSE	0.079	0.196	0.113	0.302	0.101	0.027	0.010
0.75	rBias	0.045	0.107	-0.136	-0.349	-0.045	0.006	-0.003		
	rsMSE	0.051	0.124	0.137	0.351	0.096	0.027	0.009		

To evaluate the interval estimator, the proposed bootstrap procedures of PBP, HBP and BCP described in Section 2.4 with the $B = 10,000$ bootstrap sample are used. For each simulation run, the empirical distributions of the MLEs of the distribution parameters and survival functions are constructed based on 10,000 bootstrap samples. Moreover, the confidence intervals of the distribution parameters and survival function are obtained based on the bootstrap methods of PBP, HBP and BCP. The quality of the PBP, HBP and BCP bootstrap methods is evaluated based on the coverage probability (CP) with 1000 confidence intervals. Part of the simulated CPs for 90% confidence intervals of survival function values based on sample size $n = 1000$ and $r = 0.8n$ are given as follows:

1. For the mixture Weibull distributions model with $\lambda_1 = 1.5, \alpha_1 = 2, \lambda_2 = 2, \alpha_2 = 4, \delta = 0.6$ and $S(x_0) = 0.5$ at $x_0 = 1.098$, the $CP = 0.894, 0.895$ and 0.897 by using PBP, HBP and BCP procedures, respectively. Figure 1 shows that the lengths of the obtained PBP confidence interval is more consistent than that of the obtained HBP and BCP confidence intervals. The mean length of 1000 confidence intervals based on the PBP, HBP and BCP methods are 0.0256, 0.0286 and 0.0218, respectively.
2. For the mixture GE distributions model with $\lambda_1 = 1.5, \alpha_1 = 2, \lambda_2 = 2, \alpha_2 = 4, \delta = 0.6$ and $S(x_0) = 0.5$ at $x_0 = 0.862$, the $CP = 0.898, 0.897$ and 0.899 by using PBP, HBP and BCP procedures, respectively. Figure 2 shows that the lengths of the obtained PBP confidence interval are more consistent than those of the obtained HBP and BCP confidence intervals. The mean lengths of 1000 confidence intervals based on the PBP, HBP and BCP methods are 0.0425, 0.0449 and 0.0358, respectively.
3. For the mixture GR distributions model with $\lambda_1 = 1.5, \alpha_1 = 2, \lambda_2 = 2, \alpha_2 = 4, \delta = 0.6$ and $S(x_0) = 0.5$ at $x_0 = 0.708$, the $CP = 0.894, 0.894$ and 0.896 by using PBP, HBP and BCP procedures, respectively. Figure 3 shows that the lengths of the obtained PBP confidence interval is more consistent than that of the obtained HBP and BCP confidence intervals. The mean lengths of 1000 confidence intervals based on the PBP, HBP and BCP methods are 0.0424, 0.045 and 0.0357, respectively.

Based on the simulation results, the PBP, HBP and BCP methods are competitive, and the PBP method slightly outperforms the other two due to the length of the PBP confidence interval is shorter than that of the other two bootstrap methods.

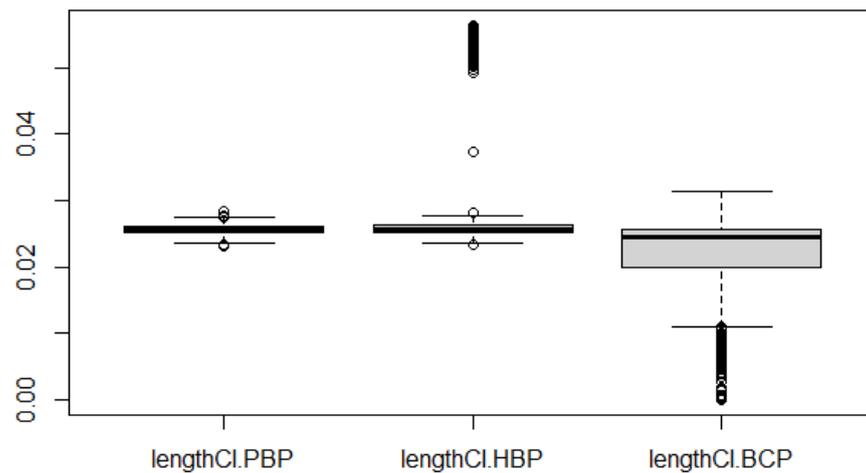


Figure 1. The box plots of bootstrap confidence intervals based on the PBP, HBP and BCP methods for the mixture Weibull distributions.

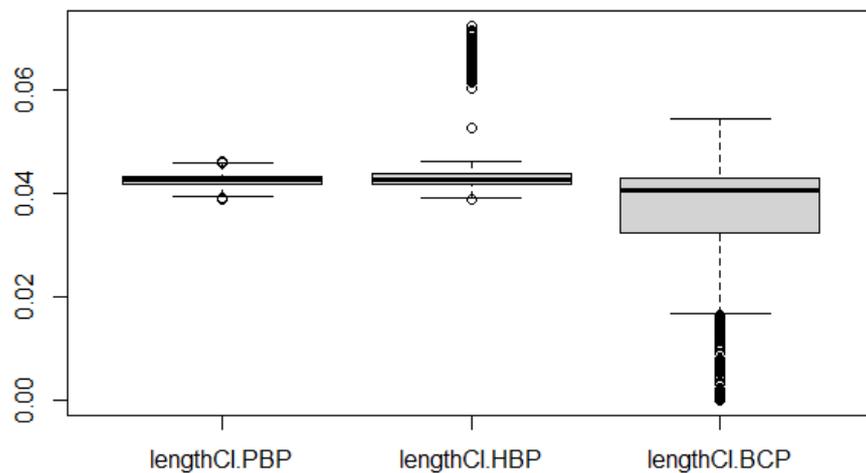


Figure 2. The box plots of bootstrap confidence intervals based on the PBP, HBP and BCP methods for the mixture GE distributions.

The simulation study is also conducted to evaluate the PDF and CDF for the three mixture distributions models, which include the mixture model with GR and GE distribution, the mixture model with GR and Weibull distributions, and the mixture model with Weibull and GE distributions. Additional simulation parameter inputs include population parameters $(\alpha_1, \lambda_1, \alpha_2, \lambda_2) = (2, 1, 5, 1)$ and mixture proportion, $(\delta_1, \delta_2) = (0.2, 0.8), (0.50, 0.50)$. We skip the outputs to save pages. Generally, all the empirical PDFs and CDFs are close to their true function curves.

A simulated example is used to show the applications of the proposed EM-MLE method. The two datasets in Table 4 present the strength of carbon fibers measured in GPa for single carbon fibers and impregnated 1000-carbon fiber tows; see Murthy et al. [19] Kundu and Raqab [20]. These two datasets were originally reported by Badar and Pries [21] and later studied by Kundu and Raqab [20]. The first and second datasets contain 63 and 69 strength measurements, respectively, which are tested under tension at gauge lengths of 10 mm and 20 mm. The histograms and empirical density plots are given in Figure 4.

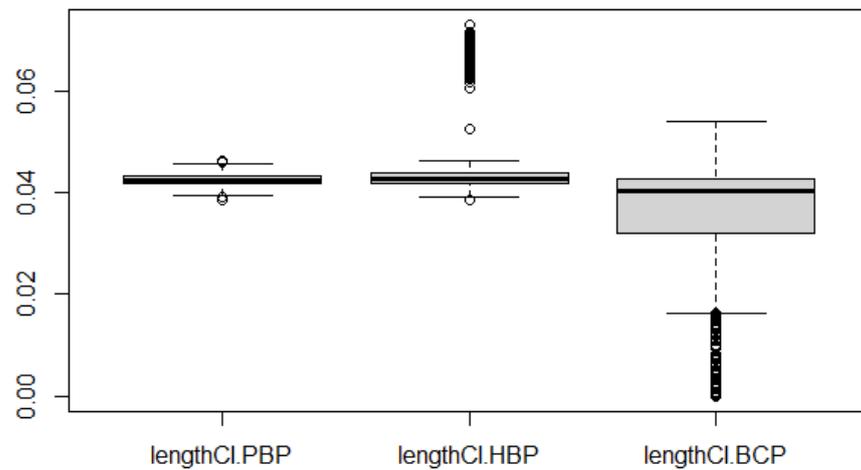


Figure 3. The box plots of bootstrap confidence intervals based on the PBP, HBP and BCP methods for the mixture GR distributions.

The Kolmogorov–Smirnov test is used to check the goodness-of-fit of using the Weibull distribution to model these two datasets. The first dataset provides the MLEs $\hat{\lambda}_1 = 2.651$ and $\hat{\alpha}_1 = 5.502$, and the second dataset provides the MLEs $\hat{\lambda}_2 = 3.31$ and $\hat{\alpha}_2 = 5.048$. Replacing the parameters with their MLEs, we obtain the Kolmogorov–Smirnov distances $D = 0.067$ with p -value = 0.965 for the first dataset and $D = 0.09$ with p -value = 0.775 for the second dataset. The testing results indicate that the Weibull distribution can characterize these two datasets well.

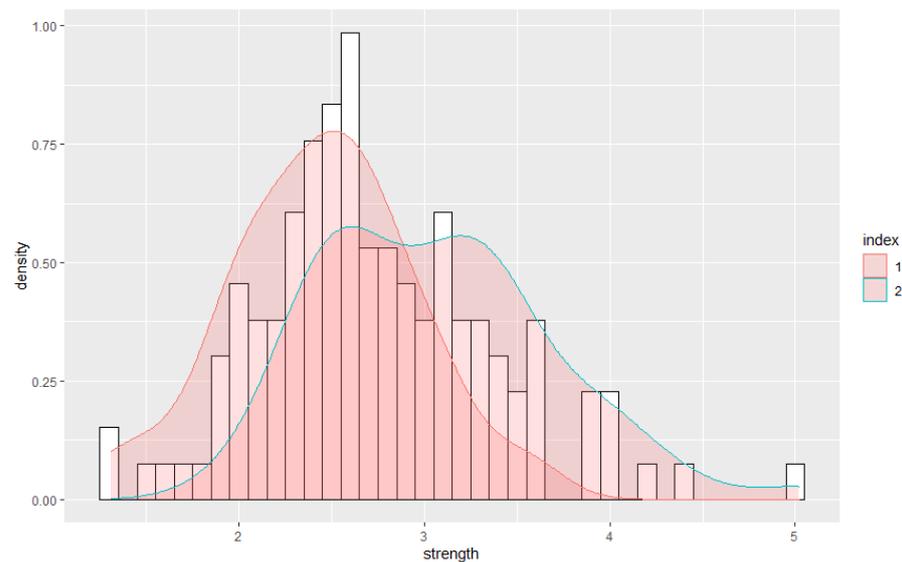


Figure 4. The histograms and empirical density plots of the strength of carbon fibers from the two datasets in Table 4.

Table 4. The strength measurements of carbon fibers.

Gague length of 10 mm								
1.901	2.132	2.203	2.228	2.257	2.35	2.361	2.396	2.397
2.445	2.454	2.474	2.518	2.522	2.525	2.532	2.575	2.614
2.616	2.618	2.624	2.659	2.675	2.738	2.74	2.856	2.917
2.928	2.937	2.937	2.977	2.996	3.03	3.125	3.139	3.145
3.22	3.223	3.235	3.243	3.264	3.272	3.294	3.332	3.346
3.377	3.408	3.435	3.493	3.501	3.537	3.554	3.562	3.628
3.852	3.871	3.886	3.971	4.024	4.027	4.225	4.395	5.02
Gague length of 20 mm								
1.312	1.314	1.479	1.552	1.7	1.803	1.861	1.865	1.944
1.958	1.966	1.997	2.006	2.021	2.027	2.055	2.063	2.098
2.14	2.179	2.224	2.24	2.253	2.27	2.272	2.274	2.301
2.301	2.359	2.382	2.382	2.426	2.434	2.435	2.478	2.49
2.511	2.514	2.535	2.554	2.566	2.57	2.586	2.629	2.633
2.642	2.648	2.684	2.697	2.726	2.77	2.773	2.8	2.809
2.818	2.821	2.848	2.88	2.954	3.012	3.067	3.084	3.09
3.096	3.128	3.233	3.433	3.585	3.585			

These two datasets are not TIHC samples. Assuming a future pool of these two datasets could be coming from a mixture distributions model with Weibull($\alpha_1 = 5.502, \lambda_1 = 2.651$) and Weibull($\alpha_2 = 5.048, \lambda_2 = 3.315$) and $\delta = 0.6$. We regenerate a TIHC sample with $r = 0.8n$ and $n = 500, \tau_1 = 1.95$, and $x_0 = 3.76$. We are interested in estimating the parameters of the mixture model of two Weibull distributions and the survival function at 0.5(or at $x_0 = 3.76$). Utilizing the proposed EM algorithm for the maximum likelihood estimation method and the regenerated sample. The generated TIHC sample is given in Table 5. The obtained MLEs are $\hat{\lambda}_1 = 2.672, \hat{\alpha}_1 = 5.435, \hat{\lambda}_2 = 3.306, \hat{\alpha}_2 = 5.261, \hat{\delta} = 0.593$ and the estimated survival function is 0.498. Using the three proposed bootstrap methods with 5000 iterations to construct the bootstrap empirical distribution of the survival, we obtain the confidence intervals of (0.482, 0.521) for the parametric PBP procedure, (0.474, 0.512) for the HBC procedure and (0.473, 0.513) for the bootstrap BCP procedure.

Table 5. The generated TIHC numerical sample.

1.099	1.822	2.045	2.247	2.382	2.476	2.614	2.723	2.921	3.044
1.109	1.828	2.054	2.248	2.384	2.482	2.619	2.726	2.927	3.051
1.167	1.839	2.054	2.25	2.384	2.487	2.62	2.738	2.929	3.056
1.241	1.841	2.054	2.251	2.388	2.492	2.628	2.741	2.934	3.06
1.255	1.844	2.067	2.258	2.388	2.493	2.63	2.745	2.935	3.067
1.26	1.853	2.067	2.26	2.391	2.494	2.634	2.747	2.936	3.069
1.293	1.856	2.073	2.268	2.392	2.495	2.639	2.759	2.937	3.075
1.314	1.862	2.076	2.269	2.393	2.497	2.639	2.765	2.938	3.093
1.35	1.865	2.077	2.269	2.396	2.509	2.639	2.769	2.94	3.102
1.365	1.867	2.078	2.278	2.4	2.511	2.642	2.773	2.941	3.102
1.413	1.873	2.092	2.279	2.404	2.518	2.642	2.774	2.943	3.102
1.417	1.877	2.097	2.283	2.407	2.52	2.643	2.777	2.95	3.105
1.452	1.886	2.1	2.284	2.408	2.53	2.645	2.778	2.959	3.114
1.465	1.888	2.102	2.286	2.413	2.532	2.648	2.794	2.96	3.122
1.486	1.894	2.102	2.296	2.414	2.534	2.653	2.799	2.961	3.125
1.498	1.895	2.108	2.297	2.416	2.538	2.656	2.808	2.964	3.13
1.505	1.904	2.116	2.301	2.417	2.54	2.658	2.813	2.964	3.133
1.529	1.912	2.121	2.302	2.419	2.545	2.659	2.814	2.967	3.135
1.547	1.916	2.122	2.31	2.424	2.547	2.666	2.822	2.967	3.148
1.572	1.918	2.125	2.313	2.424	2.551	2.669	2.832	2.975	3.148
1.611	1.924	2.133	2.316	2.424	2.552	2.67	2.833	2.985	3.149
1.629	1.925	2.137	2.319	2.424	2.571	2.672	2.835	2.986	3.149
1.648	1.948	2.137	2.32	2.429	2.572	2.673	2.849	2.989	3.15

Table 5. *Cont.*

1.669	1.955	2.138	2.321	2.431	2.575	2.677	2.852	2.992	3.151
1.688	1.975	2.157	2.325	2.433	2.577	2.685	2.854	2.996	3.151
1.699	1.976	2.173	2.328	2.433	2.579	2.686	2.859	2.997	3.151
1.729	1.977	2.174	2.328	2.439	2.58	2.688	2.869	2.997	3.154
1.733	1.991	2.176	2.333	2.441	2.582	2.691	2.873	2.998	3.161
1.749	1.994	2.177	2.335	2.444	2.584	2.694	2.878	3.001	3.171
1.75	1.994	2.182	2.338	2.444	2.585	2.695	2.882	3.003	3.174
1.768	2.001	2.183	2.354	2.445	2.586	2.696	2.883	3.006	3.187
1.78	2.003	2.205	2.355	2.446	2.586	2.699	2.884	3.007	3.19
1.787	2.004	2.209	2.356	2.446	2.589	2.704	2.887	3.011	3.192
1.788	2.005	2.213	2.36	2.451	2.592	2.706	2.888	3.021	3.196
1.798	2.009	2.219	2.364	2.452	2.593	2.711	2.904	3.026	3.196
1.798	2.014	2.224	2.369	2.453	2.595	2.712	2.904	3.026	3.212
1.799	2.016	2.233	2.369	2.461	2.596	2.715	2.906	3.028	3.222
1.804	2.016	2.235	2.37	2.465	2.599	2.718	2.912	3.03	3.224
1.808	2.021	2.236	2.374	2.468	2.603	2.72	2.913	3.038	3.232
1.811	2.027	2.244	2.379	2.47	2.614	2.722	2.914	3.039	3.234

4. Illustrative Examples

Three practical datasets will be used for the further investigation on the mixture modeling and the illustration of the proposed EM-MLE algorithm. To save space, only the first example dataset is reported in Table 6. Table 6 has 20 electronic failure times that were used by Razali and Salih [4] for the mixture distribution based on two Weibull distributions where one has a positive support with a positive left boundary and the other has the whole positive support. The second dataset regarding 153 aircraft windshield failure times in terms of 1000 h was originally given in Blischke and Murthy [19]. This dataset was also used by Ruhi et al. [5] for the twofold Weibull mixture model. The third dataset contains 600 lifetimes of adapters which are used in the liquid-crystal display (LCD) assembling monitors. Tsai et al. [22] has studied this dataset and their suggested model of two Weibull mixture distributions can characterize this dataset well.

4.1. Example 1

Table 6 has 20 electronic failure times that were used by Razali and Salih [4] for the mixture distribution based on two Weibull distributions, where one has a positive support with a positive left boundary and the other has the whole positive support. In this example, the mixture model of Weibull and GE distributions is used to model TIHC samples that can be generated from 20 electronic failure times with $r = 0.5n, 0.6n, 0.75n, 0.8n, 0.9n, n$ with $n = 20$, and a large value of τ_1 is used to allow all observations are not censored.

Table 6. Twenty electronic failure times.

0.03	0.12	0.22	0.35	0.73	0.79	1.25	1.41	1.52	1.79
1.80	1.94	2.38	2.40	2.87	2.99	3.14	3.17	4.72	5.09

Given a TIHC sample generated by the TIHC with each given r and τ_1 , the MLEs of the mixture distribution parameters were obtained through the proposed EM algorithm. Then the corresponding six mixture distribution PDFs using six different sets of MLEs as the parameter inputs were obtained and drawn in Figure 5.

All six MLEs of mixture PDFs are very close except at $x_0 = 2$, where the estimated PDFs using smaller r values have higher peak. Moreover, the 95% confidence intervals of $S(x_0)$ are reported in Table 7 based on the PBP, HBP and BCP procedures, where the LB and UB denote the lower and upper bounds of a bootstrap confidence interval, respectively. All PBP, HBP and BCP confidence intervals at each line of given r are closed. Moreover, two boundaries from the PBP and BCP procedures are closed for each given value of r among three procedures.

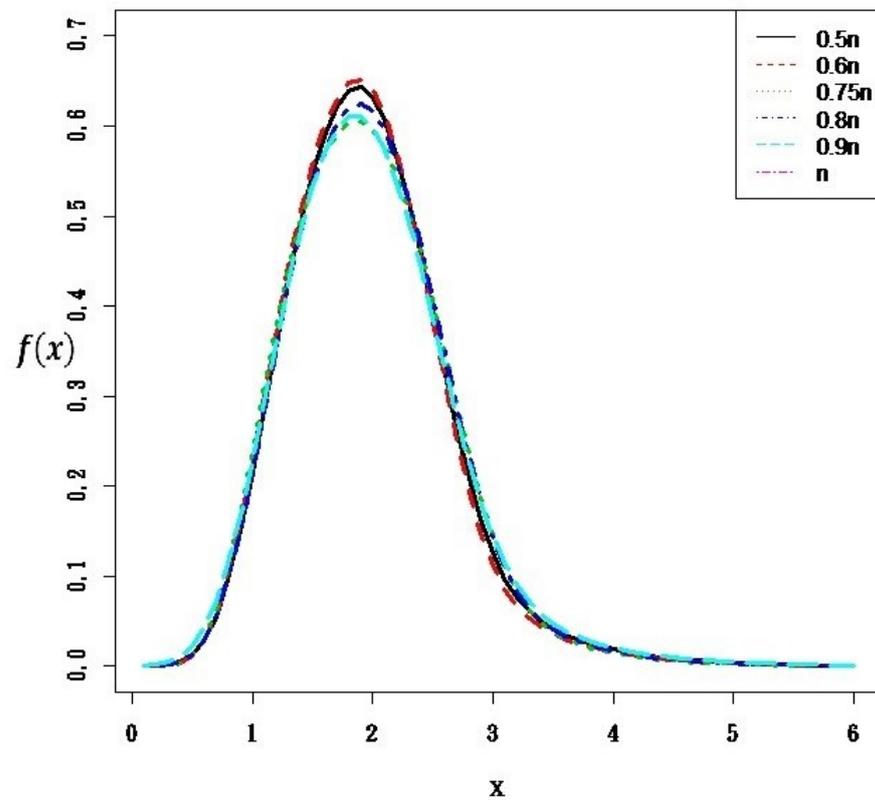


Figure 5. The MLEs of Mixture PDFs of Example 1.

Table 7. 95% Confidence intervals of the survival function at $x_0 = 2$.

r	$S(x_0)$	PBP		HBP		BCP	
		LB	UB	LB	UB	LB	UB
$0.5n$	0.4436	0.4170	0.4625	0.4247	0.4702	0.4150	0.4612
$0.6n$	0.4496	0.4326	0.4606	0.4386	0.4666	0.4319	0.4548
$0.75n$	0.4178	0.4080	0.4412	0.3943	0.4275	0.4088	0.4471
$0.8n$	0.4409	0.4300	0.4533	0.4287	0.4519	0.4300	0.4533
$0.9n$	0.4542	0.4508	0.4562	0.4522	0.4576	0.4533	0.4562
n	0.4489	0.4463	0.4512	0.4465	0.4514	0.4463	0.4512

4.2. Example 2

In this example, aircraft windshield failure times will be used to check the mixture model through the three proposed distributions. Similar to the procedure used by Ruhi et al. [5], we separated 153 observations into two groups; one group has 88 observations and the other group has 65 observations. Ruhi et al. [5] checked Weibull distribution as good-of-fit for both groups. The Kolmogorov–Smirnov test is used to further check the goodness-of-fit of using GE distribution to model these two groups. The first group provides the MLEs $\hat{\lambda}_1 = 0.7593$ and $\hat{\alpha}_1 = 3.648$, and the second group provides the MLEs $\hat{\lambda}_2 = 0.7024$ and $\hat{\alpha}_2 = 1.9458$. Replacing the parameters by their MLEs, we obtain the Kolmogorov–Smirnov distances $D = 0.11287$ with p -value = 0.2122 for the first group and $D = 0.14015$ with p -value = 0.1413 for the second group. The testing results indicate that the GE distribution can characterize these two groups well.

In this study, the mixture model with Weibull and GE distributions is suggested to model TIHC samples that were generated from these 153 observations with $r = 0.5n, 0.6n, 0.75n, 0.8n, 0.9n, n, n = 153$ and $\tau_1 = 2$. Given a TIHC sample generated by the TIHC with each given r and τ_1 , the MLEs of mixture distribution parameters were obtained through the proposed EM algorithm. Then the corresponding six mixture

distribution PDFs using the calculated six different sets of MLEs as the parameter inputs were obtained.

The 95% confidence intervals of $S(t_0)$ are reported in Table 8 based on the PBP, HBP and BCP procedures, where the LB and UB denote the lower and upper bounds of a bootstrap confidence interval, respectively. Table 8 shows that all estimated values for $r \geq \frac{n}{2}$ are very closed to the corresponding estimated results by using all 153 observations. Moreover, the 95% confidence intervals by BP and BC are very closed and have shorter lengths than the 95% confidence intervals by HB.

Table 8. 95% Confidence intervals of survival function at $x_0 = 2$ for windshield.

r	$S(t_0)$	PBP		HBP		BCP	
		LB	UB	LB	UB	LB	UB
$0.5n$	0.6011	0.5949	0.6067	0.5946	0.6073	0.5949	0.6067
$0.6n$	0.6277	0.6077	0.6497	0.6057	0.6477	0.6076	0.6496
$0.75n$	0.6336	0.6079	0.6643	0.6028	0.6592	0.6088	0.6649
$0.8n$	0.6248	0.6083	0.6400	0.6095	0.6412	0.6077	0.6395
$0.9n$	0.6123	0.6064	0.6212	0.6035	0.6182	0.6076	0.6213
n	0.6218	0.6085	0.6372	0.6063	0.6351	0.6085	0.6372

4.3. Example 3

This example was based on a past statistical consultation project for a LCD manufacturer in Asia. The manufacturer’s information is not disclosed here due to confidentiality. We refer to this manufacturer as the company. The dataset is regarding the failure time of a type of video graphics array (VGA) adapter used for the liquid-crystal display (LCD) assembling monitors. The work done by Tsai et al. [22] was suggested and the company continued to purchase adapters from two suppliers and started to record the exact failure time in terms of hours and the originality of each adapter. In this example, the dataset that contains a sample of $n = 600$ failure times in terms of days, with their originality, are considered. Therefore, all 600 failure times can be separated into two groups. The first group has 266 observations and the second group has 334 observations. The Kolmogorov–Smirnov test is used to check the goodness-of-fit of using the Weibull distribution to model these two groups. The first group provides the MLEs $\hat{\lambda}_1 = 58.18$ and $\hat{\alpha}_1 = 2.027$, and the second group provides the MLEs $\hat{\lambda}_2 = 39.62$ and $\hat{\alpha}_2 = 4.145$. Replacing the parameters with their MLEs, we obtain the Kolmogorov–Smirnov distances $D = 0.0000+$ with p -value > 0.999 for the first group and $D = 0.0000+$ with p -value > 0.999 for the second group. The testing results indicate that the Weibull distribution can characterize the dataset of two groups well.

In this study, the mixture model with Weibull and Weibull distributions is suggested to model TIHC samples, which were generated based on these 600 observations. The TIHC scheme is $r = 0.8n = 480$ and $\tau_1 = 80$. Based on the generated sample, the MLEs of mixture distribution parameters were obtained by $\hat{\lambda}_1 = 58.479$, $\hat{\alpha}_1 = 2.01$, $\hat{\lambda}_2 = 39.396$ and $\hat{\alpha}_2 = 3.978$. The 95% confidence intervals of $S(t_0) = 0.5$ were also obtained and reported in Table 9 based on the PBP, HBP and BCP procedures. Table 9 shows that all 95% bootstrap confidence intervals cover the true survival function.

Table 9. 95% confidence intervals of the survival function at $x_0 = 39$ for adapters.

r	$S(x_0)$	PBP		HBP		BCP	
		LB	UB	LB	UB	LB	UB
$0.8n$	0.5129	0.4804	0.5374	0.4745	0.5315	0.4774	0.5350

5. Conclusions and Remarks

In the real world, source items could come from different suppliers to support a company’s production line. The heterogeneity from different resources may cause inconsis-

tendencies in the production quality. In these circumstances, it would be more reasonable to utilize a mixture distributions model to characterize the product's lifetime quality. Moreover, it is very difficult to collect the entire lifetime sample from the life test because of the finite test schedule and restricted budget. Hence, the TIHC scheme with the given failure number r and life test schedule τ_1 has been proposed for collecting lifetime data from numerous practitioners. In this work, the EM algorithm has been proposed to find the MLEs of the mixture distribution model parameters and the survival function. Three different bootstrap procedures—the PBP, HBC and BCP procedures—are provided to establish the empirical distribution of the MLE of the mixture distribution model parameters and the survival function based on TIHC samples, and to obtain the confidence intervals. The parameter estimation for the mixture distributions model requires a large sample; even the mixture proportions are known with a complete sample. The TIHC scheme renders the sample censored and incomplete. Hence, the proposed EM-MLE requests a very large sample in order to obtain good quality MLEs for the model parameters. This is a drawback of the proposed EM-MLE method, and how to reduce the sample size can be left for future study.

All existing studies on the mixture distributions model have been focused on the mixture of common distributions but with different distribution parameters. In the current work, we extended the EM-MLE method to the mixture distribution model with different distributions as members. The approximate MLE approach proposed by Mendenhall and Hader [8] is a very interesting procedure for mixture distribution modeling. This will be the focus of a future research project. Meanwhile, the Bayesian estimation approach and some other censoring schemes might provide novel future research potential to add to the mixture distributions model. These two topics will be studied in the future.

Author Contributions: Data curation, W.-C.T.; Funding acquisition, T.-R.T.; Investigation, T.-R.T. and Y.L.; Methodology, T.-R.T. and W.-C.T.; Project administration, T.-R.T.; Resources, T.-R.T. and Y.L.; Software, T.-R.T. and W.-C.T.; Supervision, T.-R.T. and Y.L.; Writing—original draft, T.-R.T. and Y.L.; Writing—review & editing, T.-R.T. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported by the grant of Ministry of Science and Technology, Taiwan MOST 110-2221-E-032-034-MY2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data can be found in Section 4.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Peel, D.; McLachlan, G.J. Robust Mixture modelling using the t distribution. *Stat. Comput.* **2000**, *10*, 339–348. [[CrossRef](#)]
2. McLachlan, G.J.; Peel, D. *Finite Mixture Models*; Wiley: Hoboken, NJ, USA, 2000.
3. Sultan, K.S.; Ismail, M.A.; Al-Moisheer, A.S. Mixture of two inverse Weibull distributions: Properties and estimation. *Comput. Stat. Data Anal.* **2007**, *51*, 5377–5387. [[CrossRef](#)]
4. Razali, A.M.; Salih, A.A. Combining two Weibull Distribution Using a Mixing Parameter. *Eur. J. Sci. Res.* **2009**, *31*, 296–305.
5. Ruhi, S.; Sarker, S.; Karim, M.R. Mixture models for analyzing product reliability data: A case study. *Springer Plus* **2015**, *4*, 634. [[CrossRef](#)] [[PubMed](#)]
6. Balakrishnan, N.; Aggarwala, R. *Progressive Censoring: Theory, Methods, and Applications*; Birkhauser: Boston, MA, USA, 2000.
7. Balakrishnan, N.; Cramer, E. *The Art of Progressive Censoring*; Birkhauser: Boston, MA, USA, 2014.
8. Mendenhall, W.; Hader, R.J. Estimation of Parameters of Mixed Exponentially Distribution Failure Time Distributions from Censored Life Test Data. *Biometrika* **1958**, *45*, 504–520. [[CrossRef](#)]
9. Jones, P.W.; Ashour, S.K. Bayesian Estimation of the Parameters of the Mixed Exponential distribution from censored samples. *Biometrical* **1976**, *18*, 633–637. [[CrossRef](#)]
10. Amin, E.A. Maximum likelihood estimation of the mixed generalized Rayleigh distribution from Type I censored samples. *Appl. Math. Sci.* **2011**, *5*, 2753–2764.
11. Epstein, B. Truncated Life Tests in the Exponential Case. *Ann. Math. Stat.* **1954**, *25*, 555–564. [[CrossRef](#)]

12. Childs, A.; Chandrasekar, B.; Balakrishnan, N.; Kundu, D. Exact likelihood inference exact likelihood inference based on type-I and type-II hybrid censored samples from the exponential distribution. *Ann. Inst. Stat. Math.* **2003**, *55*, 319–330. [[CrossRef](#)]
13. Balakrishnan, N.; Kundu, D. Hybrid censoring: Models, inferential results and applications. *Comput. Stat. Data Anal.* **2013**, *57*, 166–209. [[CrossRef](#)]
14. Tian, Y.; Zhu, Q.; Tian, M. Estimation for mixed exponential distributions under type-II progressively hybrid censored samples. *Comput. Stat. Data Anal.* **2015**, *89*, 85–96. [[CrossRef](#)]
15. Tian, Y.; Yang, A.; Li, E.; Tian, M. Parameters estimation for mixed generalized inverted exponential distributions with type-II progressive hybrid censoring. *Hacet. J. Math. Stat.* **2018**, *47*, 1023–1039. [[CrossRef](#)]
16. Efron, B. *The Jackknife, the Bootstrap, and Other Resampling Plans*; SIAM: Philadelphia, PA, USA, 1982.
17. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman & Hall: New York, NY, USA, 1993.
18. Shao, J.; Tu, D. *The Jackknife and Bootstrap*; Springer Series in Statistics; Springer: New York, NY, USA, 2012.
19. Murthy, D.N.P.; Xie, M.; Jiang, R. *Weibull Model*; Wiley: New York, NY, USA, 2004.
20. Kundu, D.; Raqab, M.Z. Estimation of $R = P(Y < X)$ for three-parameter Weibull distribution. *Stat. Probab. Lett.* **2009**, *79*, 1839–1846.
21. Badar, M.G.; Priest, A.M. Statistical aspects of fiber and bundle strength in hybrid composites. In *Progress in Science and Engineering Composites*; Hayashi, T., Kawata, K., Umekawa, S., Eds.; ICCM-IV: Tokyo, Japan, 2018; pp. 1129–1136.
22. Tsai, T.-R.; Ng, H.K.T.; Pham, H.; Lio, Y.; Chiang, J.-Y. Reliability inference for VGA adapter from dual suppliers based on contaminated type-I interval-censored data. *Qual. Reliab. Eng. Int.* **2019**, *35*, 2297–2313. [[CrossRef](#)]