



General EM algorithm for fitting non-monotone hazard functions from truncated and censored observations

Stephane Barde^a, Young Myoung Ko^c, Hayong Shin^{b,*}

^a Innovation Center, Samsung Electronics, Hwasung, Gyeonggi-do, 18448, Republic of Korea

^b Department of Industrial & Systems Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

^c Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Gyeongbuk, 37673, Republic of Korea

ARTICLE INFO

Article history:

Received 26 May 2021

Received in revised form 17 February 2022

Accepted 3 July 2022

Available online 11 July 2022

Keywords:

EM algorithm

Sequential quadratic programming

Phase-type distribution

Left-truncated & right-censored

Hazard function

ABSTRACT

Recently, many researchers focused on modeling non-monotonic hazard functions such as bath-tube and hump shapes. However, most of their estimation methods are focused on complete observations. Since reliability data are typically censored and truncated, a general EM algorithm is proposed, which can fit any of those complex hazard functions. The proposed EM algorithm is analyzed by fitting well-known 4-parameter hazard functions, where its performance is compared by their specific direct methods through extensive Monte Carlo simulations.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Hazard rates are useful tools in reliability analysis and decision-making [15,20]. As monotonic hazard functions such as the Weibull one have been heavily studied in the past [17], many researchers focused on modeling non-monotonic hazard functions recently. For instance, the hump-shaped hazard rate, which has an increasing then decreasing rate, is a popular model for electronic components that exhibits a decreasing hazard rate, but reaches a maximum hazard rate early in life. The log-normal distribution is often used for such systems [17]. In addition, the bath-tube shaped hazard rate, which has decreasing then increasing rate, is used to model the complete life cycle of a system. The bath-tube shape has been studied extensively by extending the forms of the Weibull distribution by adding a new parameter such as the flexible Weibull model [7] or by introducing several new parameters [1] such as the discrete additive Weibull distribution (DAddW) [8], the generalized modified Weibull distribution (GMW) [11], and the exponentiated generalized gamma distribution (EGG) [12].

However, reliability data are typically censored and truncated, i.e. the exact failure times are not always known. Right-censored data are frequently observed because of the frequent need to ana-

lyze life test data before all units are failed, and also due to the practice of preventive maintenance (PM) in the past [17]. Also, burn-in is a commonly used technique to improve the reliability of a system that exhibits high hazard rate during the infant mortality period due to defects during manufacturing processes [20]. Thus, when analyzing reliability data, left-truncation often occurs due to burn-in periods with distribution having bath-tube shaped hazard rate. Hence, when estimating the parameters of a hazard function, one has to take into account this incompleteness of the reliability data. The maximum likelihood estimation (MLE) method for incomplete data problems can be carried out by the Expectation-Maximization (EM) algorithm [16]. Recently, the usage of such tool was only applied to well-known simple distributions: Mitra (2013) focuses on parameter estimation of lognormal, Weibull, and Gamma distributions [18]. However, there is not such implementation for more complex hazard functions that exhibits non-monotone shapes, especially the bath-tube one, which is intensively applied in survival analysis.

Hence, the goal of this paper is to provide a general EM algorithm that can fit any desired hazard function. Barde et al. proposed an interesting approach, where the parameters of a restricted Discrete Phase-type (DPH) distribution linked to a pre-specified hazard function was estimated via the direct method (DM) [5]. They use the DPH distribution because it is able to approximate any general distribution since fitting a general phase-type distribution corresponds to an automatic model-selection within a large class of distribution [2]. Hence, one can estimate

* Corresponding author.

E-mail addresses: sbace@alumni.kaist.ac.kr (S. Barde), youngko@postech.ac.kr (Y.M. Ko), hyshin@kaist.ac.kr (H. Shin).

the parameters of a desired hazard function by using the sufficient statistics of the restricted DPH distribution. Thus, this paper extends the previous work by developing an EM algorithm that can fit any pre-specified hazard function from left-truncated and right-censored data. The proposed EM algorithm is analyzed via extensive Monte Carlo simulations by comparing the three well-known 4-parameter distributions recently developed: DAddW [8], GMW [11], and EGG [12] distributions.

This paper is organized as follows. In section 2, the problem description is described. In section 3, the proposed methodology is provided. In section 4, numerical results are given. In section 5, a conclusion and directions of future works are provided.

2. Problem description

As discussed previously, reliability data are composed of incomplete observations. Censoring time occurs when units are followed until the end of a survival analysis, where units that survive at that point are right-censored as their actual failure times fall to the right of this censoring point. Note that this can be also due to PM practice in the past. Such censoring is also termed as type-I censoring in the literature [17,18]. Also, truncation occurs when observations are greater than a specific lowerbound, i.e. all the values that are below the threshold level are never observed, which is a common form of truncation for lifetime data. For example, most products are tested for some pre-specified number of hours before being in the market. During the testing periods, some units may have failed, but it is not known to us, which leads to a left-truncated dataset [17,18].

Let us denote n_F as the number of realized failure observations and y_k as the realized failure time of the k^{th} observation for $k = 1, \dots, n_F$. Let n_B be the number of right-censored observations and B_k be the right-censored level of the k^{th} observation for $k = 1, \dots, n_B$. Let us denote n_{TF} as the number of left-truncated failure observations, let A_k be the left-truncated level of the k^{th} truncated observation for $k = 1, \dots, n_{TF}$ and let z_k be the failure time of the k^{th} truncated observation for $k = 1, \dots, n_{TF}$. Let us denote n_C as the number left-truncated and right-censored observations, let A'_k be the left-truncated level of k^{th} truncated observation for $k = 1, \dots, n_C$ and let C_k be the right-censored level of k^{th} truncated observation for $k = 1, \dots, n_C$.

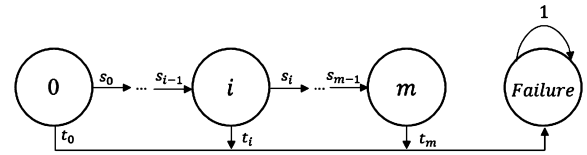
The dataset composed of left-truncated and right-censored observations is denoted as \mathcal{D}_{LTRC} , where

$$\mathcal{D}_{LTRC} = \left\{ (y_k)_{k=1, \dots, n_F}, (B_k)_{k=1, \dots, n_B}, (Z_k)_{k=1, \dots, n_{TF}}, (A_k)_{k=1, \dots, n_{TF}}, (C_k)_{k=1, \dots, n_C}, (A'_k)_{k=1, \dots, n_C} \right\}. \tag{1}$$

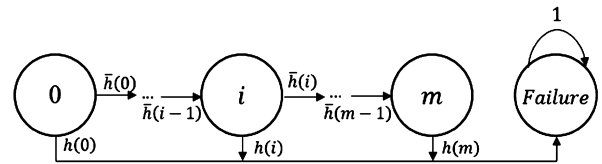
3. Methodology

3.1. Restricted class of DPH distribution

Let the hazard function of interest be defined as $h(\cdot; \theta_U)$ with parameter θ_U . The estimation of the proposed hazard sequence is based on the specific DTMC representing the degradation of a system presented in Fig. 1a proposed by [5]. Let $\{X_k\}_{k \geq 0}$ be the Markov chain representing the degradation process. Its states $0, \dots, m$ are transient, and its failure state F is recurrent. We assume that the system always renews at phase 0 (“as good as new”), i.e. $\mathbb{P}(X_0 = 0) = 1$. Let $s_i = \mathbb{P}(X_{k+1} = i + 1 | X_k = i)$ be the probability to survive at phase i , whereas let $t_i = \mathbb{P}(X_{k+1} = F | X_k = i)$ be the probability to fail at phase i , where $t_i = 1 - s_i$. Let the stopping time $Y = \inf\{k \geq 1 | X_k = F\}$ be the time to absorption. Then, Y has a DPH distribution with parameters $(t_i, s_i)_{i=0, \dots, m}$ and number of phases m . It possesses a simple hazard sequence $h(\cdot)$, which is



(a) Diagram of Restricted DPH via Markov Chain



(b) Diagram of Restricted DPH linked to a hazard function $h(\cdot)$

Fig. 1. Diagrams of a Markov chain representing the degradation process.

$$h(i) = \mathbb{P}(Y = i | Y \geq i) = t_i. \tag{2}$$

Therefore, the DPH distribution depends solely on the hazard sequence $h(\cdot)$ as described in Fig. 1b, where $\bar{h}(i) = 1 - h(i)$. In addition, the number of phases m has a physical interpretation: the maximum age of the system. Thus, $t_m = 1$ for non-decreasing and bath-tube shaped hazard functions, whereas $t_m = 0$ for hump shaped hazard function, i.e. $h(m; \theta_U) \in \{0, 1\}$.

The probability mass function (PMF) of the restricted DPH by a hazard sequence $h(\cdot; \theta_U)$ is

$$f(y; \theta_U) = \prod_{i=0}^{y-1} [1 - h(i; \theta_U)] h(y; \theta_U) \text{ for } y > 0. \tag{3}$$

The cumulative density function (CDF) of the restricted DPH by a hazard sequence $h(\cdot; \theta_U)$ is

$$F(y; \theta_U) = 1 - \prod_{i=0}^y [1 - h(i; \theta_U)] \text{ for } y > 0. \tag{4}$$

3.2. Proposed EM algorithm

This paper maximizes the likelihood function of the DPH distribution restricted by any hazard function instead of the likelihood function of the hazard function. The main advantages of this approach are that the derivation of the EM algorithm becomes straightforward, and the computation of the Q-function [16] does not depend solely to the hazard function, but can be applied to any parametric function $h(\cdot; \theta_U)$.

We define the following sufficient statistics [14] of the restricted DPH distribution by the DTMC defined in Fig. 1b. Let N_i be the number of transitions from phase i to phase $i + 1$ for $i = 0, \dots, m$, and let F_i be the number of jumps to the failure (absorbing) state F for $i = 0, \dots, m$. Note that there is one more sufficient statistic that represents the number of times that it starts from phase i , but since the system always starts at phase 0, we omitted it. Then, the likelihood of complete data is

$$L_C(\theta_U) = \prod_{i=0}^m [1 - h(i; \theta_U)]^{N_i} \prod_{i=0}^m h(i; \theta_U)^{F_i}. \tag{5}$$

Then, the log-likelihood of complete data is

$$l_C(\theta_U) = \sum_{i=0}^m N_i \log(1 - h(i; \theta_U)) + \sum_{i=0}^m F_i \log(h(i; \theta_U)). \tag{6}$$

3.2.1. Left-truncated and right-censored data

Let \mathcal{D} be either \mathcal{D}_{RC} or \mathcal{D}_{LTRC} . Then, one needs to compute the conditional expectation of $l_C(\theta)$ given the observed data \mathcal{D} , known as the Q-function, which depends on θ given $\theta_U^{(l)}$, to derive the expectation step [16], where the superscript l denotes the l^{th} EM iteration. Then, we have

$$Q(\theta; \theta_U^{(l)}) = \mathbb{E}_{\theta_U^{(l)}} [l_C(\theta) | \mathcal{D}] = \sum_{i=0}^m \mathbb{E}_{\theta_U^{(l)}} [N_i | \mathcal{D}] \log(1 - h(i; \theta)) + \sum_{i=0}^m \mathbb{E}_{\theta_U^{(l)}} [F_i | \mathcal{D}] \log(h(i; \theta)). \tag{7}$$

Note that we use the notation $\mathbb{E}_{\theta_U^{(l)}} [l_C(\theta) | \mathcal{D}]$ to denote $\mathbb{E}_{Y \sim \text{DPH}(\theta_U^{(l)})} [l_C(\theta) | \mathcal{D}]$, where $\text{DPH}(\theta_U^{(l)})$ has the PMF defined by (3).

Let $Q_{LTRC}(\theta; \theta_U^{(l)})$ be the Q-function for the left-truncated and right-censored dataset \mathcal{D}_{LTRC} at the l^{th} EM iteration. Then, we have

$$Q_{LTRC}(\theta; \theta_U^{(l)}) = \sum_{k=1}^{n_F} \frac{h(y_k - 1; \theta_U^{(l)})}{h(y_k; \theta_U^{(l)})} \log(h(y_k; \theta)) + \sum_{k=1}^{n_F} 1_{\{y_k \geq 2\}} \sum_{i=0}^{y_k-2} (1 - h(i; \theta_U^{(l)})) \log(1 - h(i; \theta)) + \sum_{k=1}^{n_B} \sum_{i=0}^{B_k} (1 - h(i; \theta_U^{(l)})) \log(1 - h(i; \theta)) + \sum_{k=1}^{n_{TF}} \frac{h(z_k - 1; \theta_U^{(l)})}{h(z_k; \theta_U^{(l)})} \log(h(z_k; \theta)) + \sum_{k=1}^{n_{TF}} 1_{\{z_k \geq 2\}} \sum_{i=A_k+1}^{z_k-2} (1 - h(i; \theta_U^{(l)})) \log(1 - h(i; \theta)) + \sum_{k=1}^{n_C} \sum_{i=A'_k+1}^{C_k} (1 - h(i; \theta_U^{(l)})) \log(1 - h(i; \theta)). \tag{8}$$

The derivation of Q_{LTRC} is given in Appendix A. The maximization step can be computed as follows:

$$\theta_U^{(l+1)} = \operatorname{argmax}_{\theta \in \Omega_U} Q_{RC}(\theta; \theta_U^{(l)}), \tag{9}$$

which is subject to $h(m; \theta_U) \in \{0, 1\}$, and may not have a closed form. In addition, it is a constraint optimization, which can be solved by using the sequential quadratic programming (SQP) method [19,9]. Hence, the proposed EM algorithm is similar to the EM gradient method [16], as the numerical gradient and hessian of $Q_{RC}(\theta; \theta_U^{(l)})$ are used to estimate the desired parameters.

3.2.2. Model selection (m)

There is a need of a procedure to select the number of phases m of the proposed approach. Let K be the number of parameters in θ_U , i.e. $K = |\theta_U|$. By looking at the most widely used model selection methods such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) are based on the trade-off between the goodness-of-fit of the model measured in log-likelihood

value at the estimated point $\hat{\theta}_U$, i.e. $l(\hat{\theta}_U)$ and the simplicity of the model measured as a function of the number of parameters, i.e. $g(K)$ [10]. Thus, the information criterion (IC) can be expressed as $IC = l(\hat{\theta}_U) + g(K)$. Our work's goal is the estimation of θ_U for a pre-defined $h(\cdot; \theta_U)$, but the number of parameters $|\theta_U|$ does not change as m varies. Therefore, one can select the best phase \hat{m} with respect to only the log-likelihood function, i.e.

$$\hat{m} = \operatorname{argmax}_m \left\{ \max_{\theta_U \in \Omega_U} l(\theta_U; m, \mathcal{D}) \right\}. \tag{10}$$

4. Numerical results

4.1. Overview

All the numerical computations are performed by using MATLAB software. We analyze the proposed EM algorithm by fitting three well-known 4-parameter hazard functions, which are

- DAddW: the discrete distribution has the following PMF

$$f(x) = q_1^x q_2^{x^\gamma} - q_1^{(x+1)^\theta} q_2^{(x+1)^\gamma}, \tag{11}$$

where $0 < q_1, q_2 < 1$ and $\theta, \gamma > 0$ [8].

- GMW: Let us define two functions $\Upsilon(x; \theta_U)$ and $\Psi(x; \theta_U)$ where $\theta_U = (\alpha, \beta, \gamma, \lambda)$ such as

$$\Upsilon(x; \theta_U) = \alpha \beta x^{\gamma-1} (\gamma + \lambda x) \exp(\lambda x - \alpha x^\gamma \exp(\lambda x)), \tag{12}$$

and

$$\Psi(x; \theta_U) = 1 - \exp(-\alpha x^\gamma \exp(\gamma x)). \tag{13}$$

Then, the continuous distribution has the following PDF

$$f(x) = \frac{\Upsilon(x; \theta_U)}{\Psi(x; \theta_U)^{1-\beta}}, \tag{14}$$

where $\alpha, \beta > 0$ and $\lambda, \gamma \geq 0$ [11].

- EGG: Let us define the incomplete gamma ratio function as $\tilde{\gamma}(k, x) = \frac{\gamma(k, x)}{\Gamma(k)}$, where $\gamma(k, x) = \int_0^x w^{k-1} e^{-w} dw$ and $\Gamma(\cdot)$ is the gamma function. Then, the continuous distribution has the following PDF

$$f(x) = \frac{\lambda \beta}{\alpha \Gamma(k)} \left(\frac{x}{\alpha}\right)^{\beta k-1} e^{-\left(\frac{x}{\alpha}\right)^\beta} \left[\tilde{\gamma}\left(k, \left(\frac{x}{\alpha}\right)^\beta\right) \right]^{\lambda-1} \tag{15}$$

where $\alpha, \beta, \gamma, k > 0$ [12].

The proposed approach is compared to specific direct methods of each distribution by extensive Monte Carlo simulations for left-truncated and right-censored data (LTRCD). Each simulation experiment is based on two non-monotone distributions:

- Lognormal distribution, which has a hump-shaped hazard rate, where we apply *Lognormal*(3, 0.8).
- Modified Weibull (MW) distribution [20], which has a bathtub shaped hazard rate. *MW*(0.01512, 0.0876, 0.389) is used, which is a fitted model to a failure data composed of 18 devices [20]. The MW's pdf is

$$f_{MW}(y; \lambda, \alpha, \beta) = \lambda \beta \left(\frac{y}{\alpha}\right)^{\beta-1} e^{-\lambda \alpha \left(1 - e^{\left(\frac{y}{\alpha}\right)^\beta} + \left(\frac{y}{\alpha}\right)^\beta\right)}. \tag{16}$$

Different datasets are sampled from each original distribution to evaluate the proposed methods. The data generation process for RCD and LTRCD experiments is similar to [3] simulation settings.

The stopping criterion used for the proposed EM algorithm is as follows:

$$\left\| \max_{\hat{\theta}^*} Q(\hat{\theta}^*; \hat{\theta}^{(l)}) - Q(\hat{\theta}^{(l)}; \hat{\theta}^{(l-1)}) \right\|_2 < \epsilon, \tag{17}$$

where we choose $\epsilon = 10^{-2}$.

4.2. Evaluation metric

For evaluating how well a distribution is fitted, it is common practice to use the mean squared error (MSE) between the estimated parameters and the parameters of the original distribution [3] because the same distribution is compared. However, this paper compares different distributions. Thus, the Jensen-Shanon divergence (D_{JS}) is used, which is a symmetric and smoothed version of the Kullback-Leibler divergence (D_{KL}) [13]. Since a discrete distribution is compared to a continuous one, the fitted DPH can be considered as a piecewise constant continuous distribution, but the computation of D_{KL} and D_{JS} does not have a close form. Hence, this paper computes them through discretization. Let P be the pmf of the fitted DPH. Since the original distribution is continuous, its pmf is computed through discretization of its cumulative density function (cdf). Let Q be its pmf. The KL-divergence is computed as $D_{KL}(P \parallel Q) = -\sum_i P(i) \log\left(\frac{Q(i)}{P(i)}\right)$. Let $M = \frac{1}{2}(P + Q)$, which is a discrete mixture distribution between the fitted DPH and the discretized original distribution. Then, we have

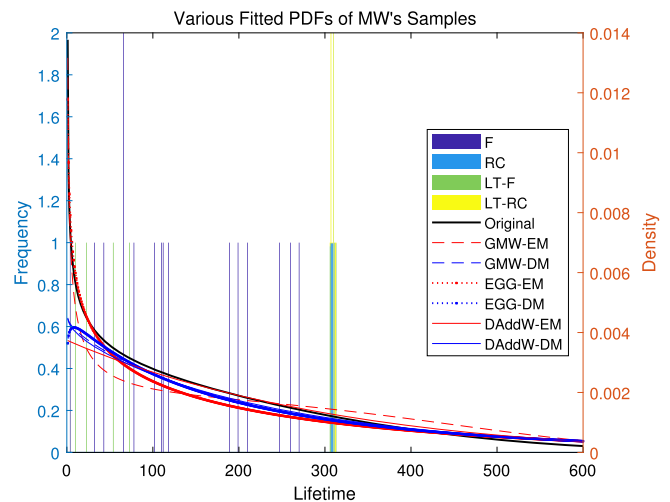
$$D_{JS}(P \parallel Q) = \frac{1}{2}D_{KL}(P \parallel M) + \frac{1}{2}D_{KL}(Q \parallel M). \tag{18}$$

4.3. Monte Carlo simulations

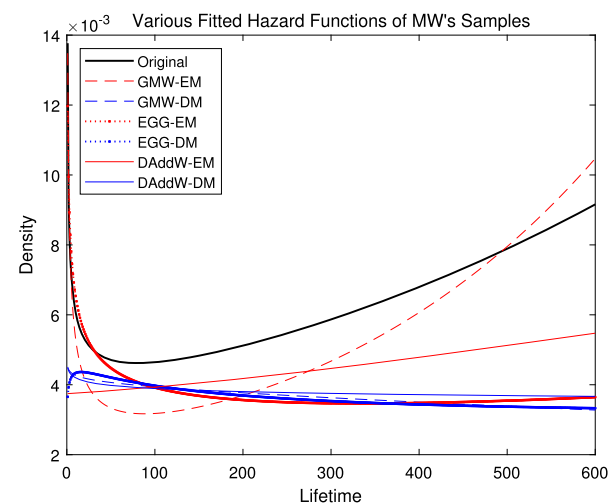
In this section, the proposed EM algorithm is analyzed by comparing to the specific direct method for each distribution (DAddW, GMW, EGG). 30 different datasets are sampled from each original distribution (lognormal, and modified Weibull) to evaluate the proposed method. Parameters' estimation is based on two different data-sizes: 30 and 100. A small dataset is used to exhibit the small data-size in reliability analysis, whereas a larger dataset is also used, which is similar to the power transformer case [3]. JSD is the performance metric used to evaluate the goodness-of-fit of the proposed approach, which is used to see how well the parameters are estimated compared to the original distribution. One of the main differences between DM and the proposed EM algorithm is the likelihood functions. DM is based on the PDF/PMF and CDF of the distribution in question, whereas the expected likelihood function for complete observations for the proposed EM algorithm is based on the sufficient statistics of the restricted DPH linked to the hazard function in question.

4.3.1. Graphical analysis

The dataset sampled from the MW distribution consists of $n_F = 16$, $n_{TF} = 5$, $n_B = 6$ and $n_C = 3$. In Fig. 2a, the graph shows a stacked histogram of failure observations in dark blue, right-censored observations in light blue, left-truncated failure observations in green, and left-truncated right-censored observations in yellow. The original distribution is represented in a solid black line denoted as "Original". The fitted GMW distribution via the proposed EM algorithm denoted as "GMW-EM" is shown by a red dashed line, whereas the one fitted via the DM method using (14) and its survival function presented in [11] is shown by a blue dashed line. The fitted EGG distribution via the proposed EM algorithm denoted as "EGG-EM" is shown by a red dashed-dotted line, whereas the one fitted via the DM method using (15) and its survival function presented in [12] is shown by a blue dashed-dotted



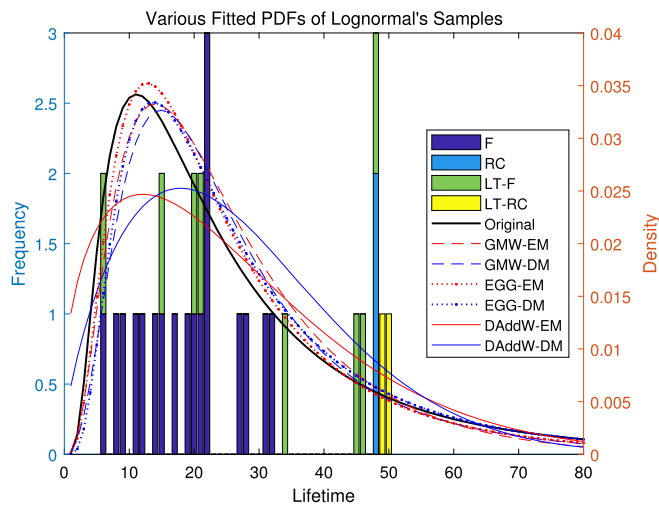
(a) Fitted PDFs from bath-tube shaped scarce data



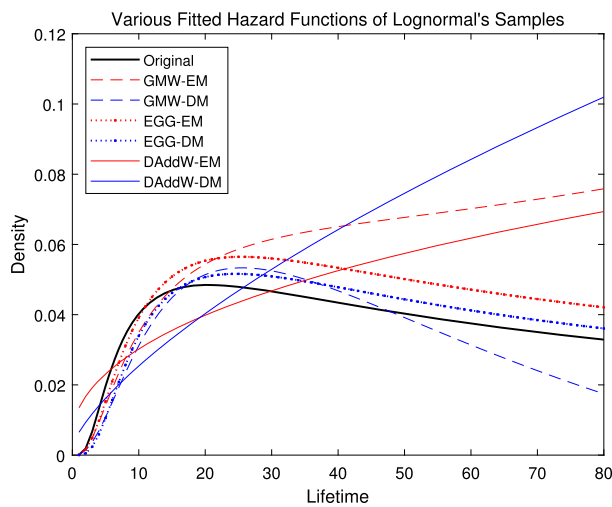
(b) Fitted HFs from bath-tube shaped scarce data

Fig. 2. Graphical results of the various models for bath-tube shaped dataset. (For interpretation of the colors in the figures, the reader is referred to the web version of this article.)

line. The fitted DAddW distribution via the proposed EM algorithm denoted as "DAddW-EM" is shown by a red solid line, whereas the one fitted via the DM method using (11) and its survival function presented in [8] is shown by a blue solid line. Even though the samples are scarce, the histogram shows that the use of censored and left-truncated censored observations is indispensable for a relevant estimation as it helps with the tail estimation. All the six PDFs exhibit a similar shape than the original one except for the "EGG-DM", which has a highly left-skewed shape. The graph 2b shows the original MW hazard rate in a solid black line, the hazard rates of the fitted GMW distribution via the proposed EM algorithm in a red dashed line and via the DM method using (14) and its survival function presented in [11] in a blue dashed line, the hazard rates of the fitted EGG distribution via the proposed EM algorithm in a red dashed-dotted line and via the DM method using (15) and its survival function presented in [12] in a blue dashed-dotted line, and the hazard rates of the fitted DAddW distribution via the proposed EM algorithm in a red solid line and via the DM method using (11) and its survival function presented in [8] in a blue solid line. Among the six fitted hazard functions, only "GMW-EM" and "EGG-EM" have the correct shape, where "GMW-EM" has



(a) Fitted PDFs from hump shaped scarce data



(b) Fitted HFs from hump shaped scarce data

Fig. 3. Graphical results of the various models for hump shaped dataset.

the lowest JSD performance value of 0.0096. It outperforms the DM method “GMW-DM” for this dataset, which has a value of 0.0179. Similarly, for both the other two models, the proposed EM algorithm performs better than the specific direct method of EGG and DAddW distributions. For the EGG distribution, the JSD values are 0.0169 and 0.0196 for the proposed EM algorithm and its specific DM respectively, and for the DAddW distribution, the JSD values are 0.0126 and 0.0175 for the proposed EM algorithm and its specific DM respectively.

The dataset sampled from the lognormal distribution consists of $n_F = 18$, $n_{TF} = 8$, $n_B = 2$ and $n_C = 2$. In Fig. 3a, the graph shows a stacked histogram of failure observations in dark blue, right-censored observations in light blue, left-truncated failure observations in green, and left-truncated right-censored observations in yellow. The original distribution is represented in a solid black line denoted as “Original”. The fitted GMW distribution via the proposed EM algorithm denoted as “GMW-EM” is shown by a red dashed line, whereas the one fitted via the DM method using (14) and its survival function presented in [11] is shown by a blue dashed line. The fitted EGG distribution via the proposed EM algorithm denoted as “EGG-EM” is shown by a red dashed-

dotted line, whereas the one fitted via the DM method using (15) and its survival function presented in [12] is shown by a blue dashed-dotted line. The fitted DAddW distribution via the proposed EM algorithm denoted as “DAddW-EM” is shown by a red solid line, whereas the one fitted via the DM method using (11) and its survival function presented in [8] is shown by a blue solid line. Similarly, the histogram shows that the use of censored and left-truncated censored observations is indispensable for a relevant estimation as it helps with the tail estimation. All the six PDFs exhibit a similar shape than the original one except for “DAddW-EM” and “DAddW-DM”, which have much lower mode. The graph 3b shows the original lognormal hazard rate in a solid black line, the hazard rates of the fitted GMW distribution via the proposed EM algorithm in a red dashed line and via the DM method using (14) and its survival function presented in [11] in a blue dashed line, the hazard rates of the fitted EGG distribution via the proposed EM algorithm in a red dashed-dotted line and via the DM method using (15) and its survival function presented in [12] in a blue dashed-dotted line, and the hazard rates of the fitted DAddW distribution via the proposed EM algorithm in a red solid line and via the DM method using (11) and its survival function presented in [8] in a blue solid line. “GMW-DM”, “EGG-EM”, and “EGG-DM” have the correct shape, where “EGG-EM” has the lowest JSD performance value of 0.0027. It outperforms the DM method “EGG-DM” for this dataset, which has a value of 0.0044. For the GMW distribution, even though “GMW-EM” fails to capture the correct shape, both methods achieve a similar performance of 0.0108 and 0.0100 respectively. Finally, “DAddW-EM” and “DAddW-DM” achieve much worse performance of 0.0199 and 0.0244 respectively as the DAddW distribution is incapable of modeling concave shape hazard function.

4.3.2. Cross-validation analysis

In this section, the proposed EM algorithm is analyzed via extensive Monte Carlo simulations to the specific DM of each hazard function (DAddW, GMW, and EGG). As the experiments are based on 30 replications, the statistics of the performance metrics are its mean and its standard deviation (std), and the number of correct fitted shapes among the 30 replications.

Table 1 depicts the statistics of the proposed EM algorithm and its specific DM for DAddW, GMW, and EGG models for datasets sampled from a MW distribution. The proposed approach shows similar or better performance for both data-sizes of 30 and 100, especially for the GMW distribution. For the DAddW model, one can notice that the JSD’s mean has the same value of 0.015 for both the proposed EM and DM approaches. As the data-size increases to 100, one can observe a significant improvement of JSD’s mean performance to 0.0074 and 0.0083, and of the number of correct fitted shapes from 13 to 22 among 30 for EM and DM approaches respectively due to additional information. Similar observation can be seen with the EGG model, where the JSD’s mean performance improves from 0.02 and 0.0214 to 0.013 and 0.0113 for both EM and DM approaches respectively. However, the proposed EM algorithm outperforms the specific DM for the GMW model as the JSD’s mean performance is 0.0133 against 0.0265 and 0.0087 and 0.01 for data-size 30 and 100 respectively. The reason behind this is the proposed EM algorithm selects the correct shape more often than the DM approach, which can also be observed for the other two models.

Table 2 shows the statistics of the proposed EM algorithm and its specific DM for DAddW, GMW, and EGG models for datasets sampled from a Lognormal distribution. The proposed approach shows similar or better performance for both data-sizes of 30 and 100, especially for the GMW distribution. For the DAddW model, one can notice that the JSD’s mean has a similar value of 0.0305

Table 1
Analysis of proposed EM algorithm for MW Dist.

Original Dist.	MW Dist.					
# of samples	30					
Pre-defined HF	DAddW		GMW		EGG	
Stats	EM	DM	EM	DM	EM	DM
Mean (JSD)	0.0153	0.0152	0.0133	0.0265	0.0200	0.0214
Std (JSD)	0.0105	0.0105	0.0122	0.0208	0.0099	0.0112
# of correct shapes	13	13	17	7	13	10
# of samples	100					
Pre-defined HF	DAddW		GMW		EGG	
Stats	EM	DM	EM	DM	EM	DM
Mean (JSD)	0.0074	0.0083	0.0087	0.0100	0.0130	0.0113
Std (JSD)	0.0046	0.0035	0.0096	0.0076	0.0033	0.0044
# of correct shapes	24	23	22	18	24	23

Table 2
Analysis of proposed EM algorithm for Lognormal Dist.

Original Dist.	Lognormal Dist.					
# of samples	30					
Pre-defined HF	DAddW		GMW		EGG	
Stats	EM	DM	EM	DM	EM	DM
Mean (JSD)	0.0305	0.0291	0.0196	0.0298	0.0149	0.0151
Std (JSD)	0.0117	0.0082	0.0150	0.0175	0.0146	0.0124
# of correct shapes	0	0	9	1	13	13
# of samples	100					
Pre-defined HF	DAddW		GMW		EGG	
Stats	EM	DM	EM	DM	EM	DM
Mean (JSD)	0.0249	0.0232	0.0096	0.0148	0.0063	0.0047
Std (JSD)	0.0043	0.0030	0.0069	0.0096	0.0053	0.0036
# of correct shapes	0	0	19	10	25	28

and 0.0291 for the proposed EM and DM approaches respectively. As the data-size increases to 100, one can also observe an improvement of JSD’s mean performance to 0.0249 and 0.0232, which is not significant. This is due to the fact that the DAddW distribution is incapable of expressing the hump shape pattern, which is why the number of correct fitted shapes are 0. For the EGG model, the JSD’s mean performance improves significantly from 0.0149 and 0.0151 to 0.0063 and 0.0047 for both EM and DM approaches respectively. However, the proposed EM algorithm outperforms the specific DM for the GMW model as the JSD’s mean performance is 0.0196 against 0.0298 and 0.0096 and 0.0148 for data-size 30 and 100 respectively. The reason behind this is the proposed EM algorithm selects the correct shape more often than the DM approach, which can also be observed for the other two models.

5. Conclusion

This paper proposed a new general EM algorithm to fit any desired hazard function by using the sufficient statistics of a restricted DPH distribution linked to the desired hazard function from left-truncated and right-censored observations. From the introduced approach, the derivation of the EM algorithm becomes straightforward, and the computation of the Q-function does not depend solely to a specific hazard function, but can be applied to any of them. The proposed EM algorithm was analyzed by comparing to the specific DM of DAddW, GMW, and EGG 4-parameter distributions through extensive Monte Carlo simulations. We observed that the proposed approach showed similar or better performance with respect to the JSD performance metric for both data-sizes of 30 and 100, especially for the GMW distribution. This was due to the fact that the proposed approach selects the correct shape more often than the DM approach.

Areas of further research are: (i) we are currently working on a model-based Reinforcement Learning approach to find an effective maintenance decision-making of a heterogeneous multi-component system such as the one presented by Barde et al. (2019) [4], where the transition probabilities of the model can be fitted through the method presented in this paper from censored and/or truncated observations. (ii) One can look for optimal burn-in periods of a complex multi-component system possessing a bath-tube shape hazard function. (iii) One can extend the proposed method to continuous phase type distribution, but we believe that it will be challenging to solve it efficiently. (iv) Finally, one can extend the proposed methodology and the work of Bebbington et al. into estimating useful periods such as optimal burn-in or wear-out periods directly from the data [6].

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science and ICT (2017R1A2B4006290).

Appendix A. Derivation of Q_{LTRC}

All the conditional expectations are derived from (7).

- Computation of $\mathbb{E}_{\theta_U^{(i)}} [F_{y_k} | Y = y_k]$

Let the sufficient statistic F_i be defined as

$$F_i = 1_{\{X_{y_{k-1}=i, X_{y_k}=F\}} \tag{A.1}$$

Then, we have $\mathbb{E}_{\theta_U^{(i)}} [F_{y_k} | Y = y_k] = \mathbb{P}(X_{y_{k-1}} = i, X_{y_k} = F | Y = y_k) = \frac{\mathbb{P}(X_{y_k}=F | X_{y_{k-1}}=i) \mathbb{P}(X_{y_{k-1}}=i)}{\mathbb{P}(Y=y_k)}$, where each term is derived

as $\mathbb{P}(Y = y_k) = \prod_{i=0}^{y_k-1} [1 - h(i; \theta_U^{(l)})] h(y_k; \theta_U^{(l)})$, $\mathbb{P}(X_{y_k} = F | X_{y_k-1} = i) = h(i; \theta_U^{(l)})$, and $\mathbb{P}(X_{y_k-1} = i) = \begin{cases} \prod_{j=0}^{y_k-1} [1 - h(j; \theta_U^{(l)})] & \text{if } i = y_k - 1 \\ 0 & \text{otherwise} \end{cases}$.

Therefore, we have

$$\mathbb{E}_{\theta_U^{(l)}} [F_{y_k} | Y = y_k] = \frac{h(y_k - 1; \theta_U^{(l)})}{h(y_k; \theta_U^{(l)})}. \tag{A.2}$$

- Computation of $\mathbb{E}_{\theta_U^{(l)}} [F_{B_k} | Y > B_k]$

It is straightforward to see that $\mathbb{E}_{\theta_U^{(l)}} [F_{B_k} | Y > B_k] = 0$ because failure is not observed since $Y > B_k$.

- Computation of $\mathbb{E}_{\theta_U^{(l)}} [N_i | Y = y_k]$

Let the sufficient statistic N_i be defined as

$$N_i = 1_{\{y_k \geq 2\}} \sum_{j=0}^{y_k-2} 1_{\{X_j=i, X_{j+1}=i+1\}}. \tag{A.3}$$

Then, we have $\mathbb{E}_{\theta_U^{(l)}} [N_i | Y = y_k] = 1_{\{y_k \geq 2\}} \sum_{j=0}^{y_k-2} \mathbb{P}(X_j = i, X_{j+1} = i + 1 | Y = y_k)$. By Bayes' theorem, for a given j , we have

$$\begin{aligned} \mathbb{P}(X_j = i, X_{j+1} = i + 1 | Y = y_k) &= \frac{\mathbb{P}(Y = y_k | X_{j+1} = i + 1) \mathbb{P}(X_{j+1} = i + 1 | X_j = i)}{P(Y = y_k)} \\ &\times \mathbb{P}(X_j = i). \end{aligned}$$

Since $\mathbb{P}(X_j = i) = \prod_{v=0}^j [1 - h(v; \theta_U^{(l)})]$ only if $j = i$, one gets $\mathbb{E}_{\theta_U^{(l)}} [N_{y_k} | Y = y_k] = 1_{\{y_k \geq 2\}} \mathbb{P}(X_i = i, X_{i+1} = i + 1 | Y = y_k)$. Each term can be computed as $\mathbb{P}(Y = y_k | X_{i+1} = i + 1) = \prod_{j=i+1}^{y_k-1} [1 - h(j; \theta_U^{(l)})] h(y_k; \theta_U^{(l)})$ and $\mathbb{P}(X_{i+1} = i + 1 | X_i = i) = 1 - h(i; \theta_U^{(l)})$. Therefore, we have

$$\mathbb{E}_{\theta_U^{(l)}} [N_i | Y = y_k] = 1_{\{y_k \geq 2\}} (1 - h(i; \theta_U^{(l)})). \tag{A.4}$$

- Computation of $\mathbb{E}_{\theta_U^{(l)}} [N_i | Y > B_k]$

Let the sufficient statistic N_i be defined as

$$N_i = \sum_{j=0}^{B_k} 1_{\{X_j=i, X_{j+1}=i+1\}}. \tag{A.5}$$

Then, we have $\mathbb{E}_{\theta_U^{(l)}} [N_i | Y > B_k] = \sum_{j=0}^{B_k} \mathbb{P}(X_j = i, X_{j+1} = i + 1 | Y > B_k)$. By Bayes' theorem, for a given j , we have

$$\begin{aligned} \mathbb{P}(X_j = i, X_{j+1} = i + 1 | Y > B_k) &= \frac{\mathbb{P}(Y > B_k | X_{j+1} = i + 1) \mathbb{P}(X_{j+1} = i + 1 | X_j = i)}{P(Y > B_k)} \\ &\times \mathbb{P}(X_j = i). \end{aligned}$$

Each term can be computed as $\mathbb{P}(Y > B_k | X_{i+1} = i + 1) = \prod_{j=i+1}^{B_k} [1 - h(j; \theta_U^{(l)})]$ and $\mathbb{P}(Y > B_k) = \prod_{j=0}^{B_k} [1 - h(j; \theta_U^{(l)})]$. Therefore, we have

$$\mathbb{E}_{\theta_U^{(l)}} [N_i | Y > B_k] = (1 - h(i; \theta_U^{(l)})). \tag{A.6}$$

- Computation of $\mathbb{E}_{\theta_U^{(l)}} [F_{y_k} | Y = y_k]$

Let the sufficient statistic F_i be defined as

$$F_i = 1_{\{X_{z_k-1}=i, X_{z_k}=F\}}. \tag{A.7}$$

Then, we have $\mathbb{E}_{\theta_U^{(l)}} [F_{z_k} | Y = z_k] = \mathbb{P}(X_{z_k-1} = i, X_{z_k} = F | Y = z_k) = \frac{\mathbb{P}(X_{z_k}=F | X_{z_k-1}=i) \mathbb{P}(X_{z_k-1}=i)}{\mathbb{P}(Y=z_k)}$, where each term is derived as $\mathbb{P}(Y = z_k) = \prod_{j=A_k+1}^{z_k-1} [1 - h(j; \theta_U^{(l)})] h(z_k; \theta_U^{(l)})$ and $\mathbb{P}(X_{z_k-1} = i) = \prod_{j=A_k+1}^{z_k-1} [1 - h(j; \theta_U^{(l)})]$ only if $i = z_k - 1$. Therefore, we have

$$\mathbb{E}_{\theta_U^{(l)}} [F_i | Y = z_k] = \frac{h(z_k - 1; \theta_U^{(l)})}{h(z_k; \theta_U^{(l)})}. \tag{A.8}$$

- Computation of $\mathbb{E}_{\theta_U^{(l)}} [F_i | Y > C_k]$

$\mathbb{E}_{\theta_U^{(l)}} [F_i | Y > C_k] = 0$ because left-truncated failures are not observed.

- Computation of $\mathbb{E}_{\theta_U^{(l)}} [N_i | Y = z_k]$

Let the sufficient statistic N_i be defined as

$$N_i = 1_{\{z_k \geq 2\}} \sum_{j=A_k+1}^{z_k-2} 1_{\{X_j=i, X_{j+1}=i+1\}}. \tag{A.9}$$

Then, we have $\mathbb{E}_{\theta_U^{(l)}} [N_i | Y = z_k] = 1_{\{z_k \geq 2\}} \sum_{j=A_k+1}^{z_k-2} \mathbb{P}(X_j = i, X_{j+1} = i + 1 | Y = z_k)$. By Bayes' theorem, for a given j , we have

$$\begin{aligned} \mathbb{P}(X_j = i, X_{j+1} = i + 1 | Y = z_k) &= \frac{\mathbb{P}(Y = z_k | X_{j+1} = i + 1) \mathbb{P}(X_{j+1} = i + 1 | X_j = i)}{P(Y = z_k)} \\ &\times \mathbb{P}(X_j = i). \end{aligned}$$

Since $\mathbb{P}(X_j = i) = \prod_{v=A_k+1}^j [1 - h(v; \theta_U^{(l)})]$ only if $j = i$, we finally have

$$\mathbb{E}_{\theta_U^{(l)}} [N_i | Y = z_k] = 1_{\{z_k \geq 2\}} \mathbb{P}(X_i = i, X_{i+1} = i + 1 | Y = z_k). \tag{A.10}$$

- Computation of $\mathbb{E}_{\theta_U^{(l)}} [N_i | Y > C_k]$

Let the sufficient statistic N_i be defined as

$$N_i = \sum_{j=A'_k+1}^{C_k} 1_{\{X_j=i, X_{j+1}=i+1\}}. \tag{A.11}$$

Then, we have $\mathbb{E}_{\theta_U^{(l)}} [N_i | Y > C_k] = \sum_{j=A'_k+1}^{C_k} \mathbb{P}(X_j = i, X_{j+1} = i + 1 | Y > C_k)$. By Bayes' theorem, for a given j , we have

$$\begin{aligned} \mathbb{P}(X_j = i, X_{j+1} = i + 1 | Y > C_k) &= \frac{\mathbb{P}(Y > C_k | X_{j+1} = i + 1) \mathbb{P}(X_{j+1} = i + 1 | X_j = i)}{P(Y > C_k)} \\ &\times \mathbb{P}(X_j = i). \end{aligned}$$

Each term can be computed as $\mathbb{P}(Y > C_k | X_{i+1} = i + 1) = \prod_{j=i+1}^{C_k} [1 - h(j; \theta_U^{(l)})]$, $\mathbb{P}(Y > C_k) = \prod_{j=A'_k+1}^{C_k} [1 - h(j; \theta_U^{(l)})]$ and $\mathbb{P}(X_i = i) = \prod_{j=A'_k+1}^{C_k} [1 - h(j; \theta_U^{(l)})]$. Therefore, we have

$$\mathbb{E}_{\theta_U^{(l)}} [N_i | Y > C_k] = (1 - h(i; \theta_U^{(l)})). \tag{A.12}$$

References

- [1] S.J. Almalki, S. Nadarajah, Modifications of the Weibull distribution: a review, *Reliab. Eng. Syst. Saf.* 124 (2014) 32–55.
- [2] S. Asmussen, O. Nerman, M. Olsson, Fitting phase-type distributions via the em algorithm, *Scand. J. Stat.* (1996) 419–441.
- [3] N. Balakrishnan, D. Mitra, Left truncated and right censored Weibull data and likelihood inference with an illustration, *Comput. Stat. Data Anal.* 56 (12) (2012) 4011–4025.
- [4] S.R. Barde, S. Yacout, H. Shin, Optimal preventive maintenance policy based on reinforcement learning of a fleet of military trucks, *J. Intell. Manuf.* 30 (1) (2019) 147–161.
- [5] S. Barde, Y.M. Ko, H. Shin, Fitting discrete phase-type distribution from censored and truncated observations with pre-specified hazard sequence, *Oper. Res. Lett.* 48 (3) (2020) 233–239.
- [6] M. Bebbington, C.-D. Lai, R. Zitikis, Useful periods for lifetime distributions with bathtub shaped hazard rate functions, *IEEE Trans. Reliab.* 55 (2) (2006) 245–251.
- [7] M. Bebbington, C.-D. Lai, R. Zitikis, A flexible Weibull extension, *Reliab. Eng. Syst. Saf.* 92 (6) (2007) 719–726.
- [8] M. Bebbington, C.-D. Lai, M. Wellington, R. Zitikis, The discrete additive Weibull distribution: a bathtub-shaped hazard for discontinuous failure data, *Reliab. Eng. Syst. Saf.* 106 (2012) 37–44.
- [9] P.T. Boggs, J.W. Tolle, Sequential quadratic programming, *Acta Numer.* 4 (1995) 1–51.
- [10] K.P. Burnham, *A Practical Information-theoretic Approach*, Springer, 1998.
- [11] J.M. Carrasco, E.M. Ortega, G.M. Cordeiro, A generalized modified Weibull distribution for lifetime modeling, *Comput. Stat. Data Anal.* 53 (2) (2008) 450–462.
- [12] G.M. Cordeiro, E.M. Ortega, G.O. Silva, The exponentiated generalized gamma distribution with application to lifetime data, *J. Stat. Comput. Simul.* 81 (7) (2011) 827–842.
- [13] D.M. Endres, J.E. Schindelin, A new metric for probability distributions, *IEEE Trans. Inf. Theory* 49 (7) (2003) 1858–1860.
- [14] L.J.R. Esparza, B.F. Nielsen, M. Bladt, Maximum likelihood estimation of phase-type distributions, Ph.D. thesis, Technical University of Denmark (DTU), 2011.
- [15] C. Lai, M. Xie, D. Murthy, A modified Weibull distribution, *IEEE Trans. Reliab.* 52 (1) (2003) 33–37.
- [16] G. McLachlan, T. Krishnan, *The EM Algorithm and Extensions*, vol. 382, John Wiley & Sons, 2007.
- [17] W.Q. Meeker, L.A. Escobar, *Statistical Methods for Reliability Data*, John Wiley & Sons, 2014.
- [18] D. Mitra, *Likelihood inference for left truncated and right censored lifetime data*, Ph.D. thesis, McMaster University, 2013.
- [19] J. Nocedal, S. Wright, *Numerical Optimization*, Springer Science & Business Media, 2006.
- [20] M. Xie, Y. Tang, T.N. Goh, A modified Weibull extension with bathtub-shaped failure rate function, *Reliab. Eng. Syst. Saf.* 76 (3) (2002) 279–285.