Operations Research Letters 48 (2020) 233–239

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Fitting discrete phase-type distribution from censored and truncated observations with pre-specified hazard sequence

Stephane Barde [a], Young Myoung Ko [b], Hayong Shin [a,*]

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

[b] Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, Gyeongbuk, 37673, Republic of Korea

ARTICLE INFO

Article history:
Received 13 October 2018
Received in revised form 13 December 2019
Accepted 25 February 2020
Available online 7 March 2020

Keywords:
Life-time data
Maximum likelihood estimator
Phase-type distribution
Left-truncated & right-censored
Hazard sequence
Markov decision process

ABSTRACT

Phase-type distribution allows approximation of non-Markovian models, which permits to analyze complex systems under Markovian deterioration. In addition, reliability data is often composed of truncated and censored observations. This paper presents a novel approach that fits a restricted class of discrete phase-type distribution through pre-specified hazard sequence from incomplete observations. Numerical results are shown using Balakrishnan's mimicked power transformers dataset. Furthermore, it can be used to fit transition probabilities of maintenance optimization's Markov decision process models from incomplete reliability data.

1. Introduction

Phase-type distributions are widely used to approximate lifetime distributions because one can analyze a system under Markovian deterioration, where the computation of the reliability or the availability becomes tractable [17]. They are also able to approximate general distribution since fitting a general phase-type distribution corresponds to an automatic model-selection within a large class of distribution [1].

In addition, reliability data are typically censored and truncated, i.e. the exact failure times are not always known. For example, some units have to be removed and inspected before knowing whether they have failed or not [9,11,12]. Thus, when fitting a lifetime model from such reliability data, one should take into account these incomplete observations.

Hence, many studies focus on estimating general phase-type distribution under incomplete reliability data. For instance, Olsson (1996) [16] proposes a differential-equation based Expectation–Maximization (EM) algorithm to fit general phase-type distribution from right censored and interval censored observations.

However, fitting general phase-type distributions from incomplete observations when using EM algorithm turns out to be computationally very expensive because it is not scalable to the number of phases, where a large number of phases is often needed for accurate approximation [2,14,15], and depends heavily on the initial values [2]. In addition, the use of general phase-type distribution or its canonical form is in general a non-linear optimization problem, which contains local optima and saddle points [2]. Thus, Thummler et al. (2006) [2] propose an EM algorithm for fitting hyper-Erlang distribution, which is a subclass of general phase-type distributions from incomplete data traces. Similarly, this paper's main objective is to overcome the curse of dimensionality by linking a phase-type distribution to a pre-specified hazard function.

Since the Weibull distribution is widely used in reliability engineering, we propose to fit a restricted class of discrete phase-type (DPH) distribution by linking to a pre-specified hazard sequence that has a similar form than the discrete Weibull hazard sequence [18] from censored and truncated observations. The use of a pre-specified hazard sequence permits to be computationally cheap, and prevents overfitting the reliability data.

This paper is organized as follows. In Section 2, the motivating application is given by explaining the power transmission transformers dataset. In Section 3, we describe the proposed lifetime model. In Section 4, numerical results are given. In Section 5, a conclusion and directions of future works are provided.

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

https://doi.org/10.1016/j.orl.2020.02.009
0167-6377/

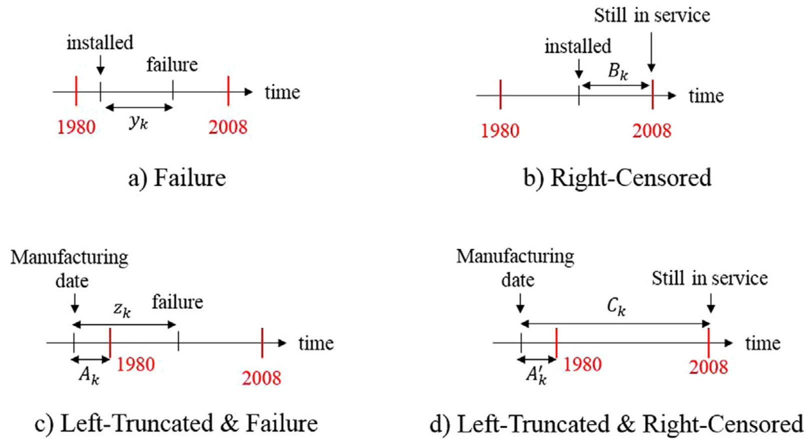

Fig. 1. Power transformers data description.

2. Motivating application

Hong et al. (2009) [8] analyzed a power transmission transformers dataset generated by an energy company that began keeping records from 1980, where the company has no information on units installed or failed before 1980. The data was analyzed in 2008, where units still in service at that time were considered as right-censored. The lifetime data of interest in this paper is the power transmission transformer dataset from [3], where they mimicked it for numerical results' purposes. Fig. 1 depicts four different cases of how power transformers data is generated.

- Case (a): let us denote n_F as the number of realized failure observations and let us denote y_k the realized failure time of kth observation for $k = 1, \ldots, n_F$, which corresponds to a unit that is installed after 1980 and failed before 2008.
- Case (b): let us denote n_B as the number of right-censored observations and let us denote B_k the right-censored level of kth observation for $k = 1, \ldots, n_B$, which refers to a unit that is installed after 1980 but it is still in service in 2008, i.e. it did not fail yet.
- Case (c): let us denote n_{TF} as the number of left-truncated and failure observations, let A_k be the left-truncated level of kth truncated observation for $k = 1, \ldots, n_{TF}$ and let z_k be the failure time of kth truncated observation for $k = 1, \ldots, n_{TF}$. Since the installation date of the failed transformer is unknown because it was installed before 1980, the manufacturing date is used as a proxy to compute the left-truncated level A_k and its failure time z_k.
- Case (d): let us denote n_C as the number of left-truncated and right-censored observations, let A'_k be the left-truncated level of kth truncated observation for $k = 1, \ldots, n_C$ and let C_k be the right-censored level of kth truncated observation for $k = 1, \ldots, n_C$. Similarly to case (c), the manufacturing date is used as a proxy of the installation date to compute the left-truncated level A'_k and its right-censored time C_k.

The reason that the power transformers data generation is divided into four different cases is for the derivation of the likelihood function presented in the next section. For instance, the difference between case a (failure) and case c (left-truncated and failure) is that a failure observation y_k is sampled from a distribution with probability density function (pdf) $g(\cdot)$, whereas a left-truncated failure observation z_k is sampled from a truncated distribution with pdf $\frac{g(\cdot)}{1-G(A_k)}$, where $G(\cdot)$ is the cumulative distribution function (CDF). Similar reasoning can be applied to the second and fourth cases. The dataset of the power transmission transformers is denoted as $\mathcal{D}_{PTD}$, where

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

3. Model description

3.1. Overview

Let $\{D_i\}_{i\geq 0}$ be a discrete-time Markov chain (DTMC) with state space $\mathcal{S} = \{1, \ldots, p, p+1\}$, where the states $1, \ldots, p$ are transient and the state $p + 1$ is absorbing. The DTMC is defined with transition probability matrix $\mathbf{P}$, which has the following form $\mathbf{P} = \begin{bmatrix} \mathbf{T} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$, where $\mathbf{T}$ is the transition matrix between transient states, $\mathbf{t}$ is the vector of probabilities of jumping to the absorbing state, and $\mathbf{0}$ is a vector composed of only zeros. Let $\boldsymbol{\pi}$ be the vector of initial probabilities, and let the stopping time $Y = \inf\{i \geq 1 \mid D_i = p + 1\}$ be the time to absorption. Then, Y has a discrete phase-type (DPH) distribution with parameters $\boldsymbol{\pi}$ and $\mathbf{T}$, but not $\mathbf{t}$ because $\mathbf{t} = \mathbf{e} - \mathbf{T}\mathbf{e}$, where $\mathbf{e}$ is a vector composed of only ones. The probability mass function (pmf) of the DPH distribution is

$$f(y; \boldsymbol{\pi}, \mathbf{T}) = \boldsymbol{\pi}'\mathbf{T}^{y-1}\mathbf{t} \text{ for } y > 0. \tag{2}$$

The CDF of the DPH distribution is

$$F(y; \boldsymbol{\pi}, \mathbf{T}) = 1 - \boldsymbol{\pi}'\mathbf{T}^{y}\mathbf{e}. \tag{3}$$

Since estimating all these probabilities are computationally very expensive, some special cases of the sub-stochastic matrix $\mathbf{T}$ are used for estimation purpose. For instance, Fig. 2 shows a DTMC that is a discrete analogue of the Coxian distribution, which is widely used for approximation [10,16]. Its transition matrix of transient probabilities becomes

$$\mathbf{T} = \begin{bmatrix} t_{1,1} & t_{1,2} & 0 & \cdots & 0 \\ 0 & t_{2,2} & t_{2,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & t_{p,p} \end{bmatrix}, \tag{4}$$

and the initial probability vector is $\boldsymbol{\pi} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$. Then, the number of parameters to be estimated becomes $2p - 1$.

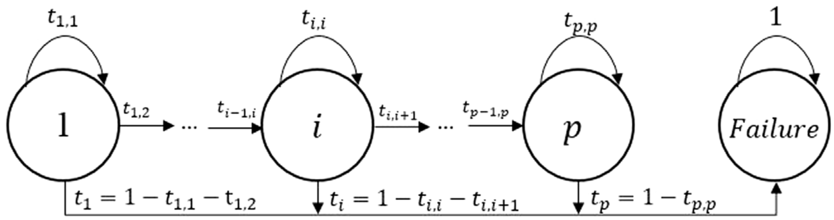

Fig. 2. Diagram of a Markov chain widely used for approximation.

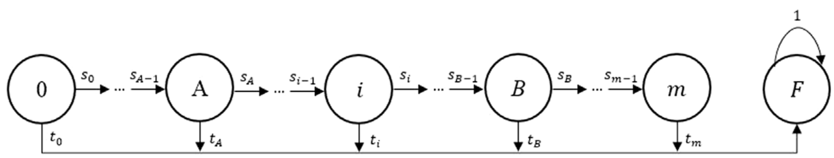

Fig. 3. Diagram of a Markov chain representing an aging process.

3.2. Proposed restricted DPH

We propose a phase-type aging model with DTMC described in Fig. 3 with state-space $\mathcal{X} = \{0, \ldots, m, F\}$, where the states $0, \ldots, m$ are transient and the state F is absorbing. Let s_i be the probability to survive at phase i, whereas let t_i be the probability to fail at phase i, where $t_i = 1 - s_i$. The phase m denotes the maximum age of the system, i.e. $t_m = 1$. Let A be the observed lowest left-truncated level, and let B be the observed largest right-censored level. The parameters of this model consist of $(t_i)_{i=0,\ldots,m}$ and m, but only parameters $(t_i)_{i=A+1,\ldots,B}$ can be directly estimated from left-truncated and right-censored observations via maximum likelihood (ML) method. In other words, $\hat{t}_i = 0$ for $i = 0, \ldots, A, B+1, \ldots, m$ by maximizing its likelihood function, which are undesired estimates. Hence, an interpolation method is needed to estimate the parameters $(t_i)_{i=0,\ldots,A}$ and $(t_i)_{i=B+1,\ldots,m}$. Therefore, the proposed interpolation method is to link the DPH distribution to a pre-specified hazard sequence.

The advantage of using the Markov chain representing an aging process as described in Fig. 3 is that the DPH has a simple hazard sequence $h(\cdot)$, which is given as

$$h(i) = \mathbb{P}(Y = i \mid Y \geq i) = t_i.$$

For the design of the pre-specified hazard sequence, the function $h(\cdot)$ is non-decreasing since we are interested in a wear-out phase system. Hence, one needs to fit a hazard sequence $h(\cdot)$ to a dataset $\mathcal{D}$ such that

$$0 = h(0) \leq \cdots \leq h(i) \leq \cdots \leq h(m) = t_m = 1.$$

The Weibull distribution is extensively used for modeling lifetime distribution, which is a powerful model to analyze a single component system, but since it exhibits non-Markovian deterioration, it makes difficult to analyze complex systems. The idea of this paper is to use a hazard sequence $h(\cdot)$ that has the same characteristics than the one of the discrete Weibull distribution (III) [18], but with Markovian deterioration. It has scale parameter $0 < \lambda < 1$ and shape parameter $\mu \geq 1$. Its hazard sequence is given by $h_W(y; \lambda, \mu) = -\log(\lambda)\, y^{\mu-1} \propto y^{\mu-1}$. Hence, we propose the following hazard sequence

$$h(i) = a i^{\mu-1}.$$

Since $h(m) = a m^{\mu-1} = 1$, then $a = 1/m^{\mu-1}$. Therefore, we have the following hazard sequence

$$h(i) = \left(\frac{i}{m}\right)^{\mu-1} \text{ for } i = 0, \ldots, m. \tag{5}$$

Hence, the proposed restricted DPH distribution has now only two parameters μ and m, which are independent to the number of phases. From (2), the pmf of the DPH linked to a general hazard sequence $h(\cdot)$ is

$$f(y; h(\cdot)) = \prod_{i=0}^{y-1} [1 - h(i)]\, h(y) \text{ for } y > 0. \tag{6}$$

By using the proposed hazard sequence (5), the pmf of DPH (μ, m) is

$$f(y; \mu, m) = \prod_{i=0}^{y-1} \left[1 - \left(\frac{i}{m}\right)^{\mu-1}\right] \left(\frac{y}{m}\right)^{\mu-1} \text{ for } y > 0. \tag{7}$$

The CDF of the DPH linked to a general hazard sequence $h(\cdot)$ is

$$F(y; h(\cdot)) = 1 - \prod_{i=0}^{y} [1 - h(i)] \text{ for } y > 0. \tag{8}$$

By using the proposed hazard sequence (5), the CDF of DPH (μ, m) is

$$F(y; \mu, m) = 1 - \prod_{i=0}^{y} \left[1 - \left(\frac{i}{m}\right)^{\mu-1}\right] \text{ for } y > 0. \tag{9}$$

Even though the hazard sequence of the discrete Weibull distribution is used, one can use other parametric hazard sequences satisfying one needs such as bathtub shaped ones, where the pmf & CDF can be computed from (6) and (8) without loss of generality.

3.3. Likelihood

The maximum likelihood method is used to estimate the parameters of the DPH distribution from left-truncated and right-censored observations (1). Let $L(\mathbf{T};\mathcal{D}_{PTD})$ be the likelihood function of a DPH distribution with respect to the power transformers data $\mathcal{D}_{PTD}$. Since failure, censored, truncated failure and truncated censored samples are independent and identically distributed with respect to Y, we have from (2) and (3)

$$\begin{aligned} L(\mathbf{T};\mathcal{D}_{PTD}) &= \left[\prod_{k=1}^{n_F}\mathbb{P}(Y=y_k)\right]\left[\prod_{k=1}^{n_B}\mathbb{P}(Y>B_k)\right] \\ &\quad \left[\prod_{k=1}^{n_{TF}}\frac{\mathbb{P}(Y=z_k)}{\mathbb{P}(Y>A_k)}\right]\left[\prod_{k=1}^{n_C}\frac{\mathbb{P}(Y>C_k)}{\mathbb{P}\left(Y>A'_k\right)}\right] \\ &= \left[\prod_{k=1}^{n_F} f(y_k)\right]\left[\prod_{k=1}^{n_B}1-F(B_k)\right]\left[\prod_{k=1}^{n_{TF}}\frac{f(z_k)}{1-F(A_k)}\right] \\ &\quad \left[\prod_{k=1}^{n_C}\frac{1-F(C_k)}{1-F\left(A'_k\right)}\right]. \end{aligned} \tag{10}$$

Then, from (10), its log-likelihood is given as

$$\begin{aligned} l(\mathbf{T};\mathcal{D}_{PTD}) &= \sum_{k=1}^{n_F}\log(f(y_k)) + \sum_{k=1}^{n_B}\log(1-F(B_k)) \\ &\quad + \sum_{k=1}^{n_{TF}}\log\left(\frac{f(z_k)}{1-F(A_k)}\right) + \sum_{k=1}^{n_C}\log\left(\frac{1-F(C_k)}{1-F\left(A'_k\right)}\right). \end{aligned} \tag{11}$$

From (11), (7), and (9), the log-likelihood function of the DPH distribution linked to the proposed hazard sequence (5) is given as

$$\begin{aligned} l(\mu,m;\mathcal{D}_{PTD}) &= \sum_{k=1}^{n_F}\left[\sum_{i=0}^{y_k-1}\log\left(1-\left(\frac{i}{m}\right)^{\mu-1}\right)\right. \\ &\quad \left. + (\mu-1)\log\left(\frac{y_k}{m}\right)\right] \\ &\quad + \sum_{k=1}^{n_B}\left[\sum_{i=0}^{B_k}\log\left(1-\left(\frac{i}{m}\right)^{\mu-1}\right)\right] \\ &\quad + \sum_{k=1}^{n_{TF}}\left[\sum_{i=A_k+1}^{z_k-1}\log\left(1-\left(\frac{i}{m}\right)^{\mu-1}\right)\right. \\ &\quad \left. + (\mu-1)\log\left(\frac{z_k}{m}\right)\right] \\ &\quad + \sum_{k=1}^{n_C}\left[\sum_{i=A_k+1}^{C_k}\log\left(1-\left(\frac{i}{m}\right)^{\mu-1}\right)\right]. \end{aligned} \tag{12}$$

Let B_{max} be the largest observed value in the dataset $\mathcal{D}_{PTD}$. Then, $B_{max} = \max\{\max_k y_k, \max_k B_k, \max_k C_k\}$. The parameters $\hat{\mu}$ and $\hat{m}$ can be estimated by applying the direct method (DM) [7] such as $\hat{m} = \operatorname{argmax}_{m>B_{max}}\left(\max_\mu l(\mu,m;\mathcal{D}_{PTD})\right)$. The estimation of m consists of a model-selection procedure as m is the number of phases. We can estimate m by iteratively maximizing $l(\mu,m;\mathcal{D}_{PTD})$ using the quasi-Newton method for $m = B_{max}+\Delta, B_{max}+2\Delta, \ldots$, and choose the one that achieves the largest likelihood value, where Δ can be interpreted as the time interval between two inspection times.

4. Numerical results

4.1. Overview

All the numerical computations are performed by using MATLAB software. Numerical results are provided from the mimicked power transformers dataset described in [3], where they propose an EM algorithm for fitting the Weibull distribution from left-truncated and right-censored observations, which is compared to the proposed method. They use a sample size of 100 with truncation percentage 40%, where the observations are sampled from a Weibull distribution with parameters $\lambda = 35$ and $\mu = 3$. This paper also compares between the DPH distribution with sub-stochastic matrix $\mathbf{T}$ defined in (4) and the proposed restricted DPH distribution linked to the hazard sequence (5). The optimization of the log-likelihood function $l(\mathbf{T};\mathcal{D}_{PTD})$ (11) is as follows

$$\begin{array}{lrcl} \min\limits_{\mathbf{T}} & -l(\mathbf{T};\mathcal{D}_{PTD}) & & \\ \text{s.t.} & \mathbf{Te} & \preceq & 1 \\ & \mathbf{T} & \succeq & 0 \end{array}, \tag{13}$$

where one can solve the mathematical program via sequential quadratic programming (SQP) method [5,13].

4.2. Evaluation metric

For evaluating how well a distribution is fitted, Balakrishnan & Mitra (2012) [3] use the mean squared error (MSE) between the estimated parameters by their proposed EM algorithm and the parameters of the original distribution since the same distribution is used. However, in this paper, different distributions are compared, thus the Jensen-Shanon divergence (D_{JS}) is used, which is a symmetric and smoothed version of the Kullback–Leibler divergence (D_{KL}) [6]. 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 with sample space Ω_P and with cardinality $|\Omega_P|$. Since the original distribution is continuous, its pmf is computed through discretization of its CDF. Let Q be its pmf with sample space Ω_Q and with cardinality $\left|\Omega_Q\right|$. The KL-divergence is computed as $D_{KL}(P \parallel Q) = -\sum_i P(i)\log\left(\frac{Q(i)}{P(i)}\right)$. This measure may be inappropriate for our problem since $|\Omega_P|$ and $\left|\Omega_Q\right|$ may have different cardinality, which may cause inappropriate values. For example, if $|\Omega_P| > \left|\Omega_Q\right|$, then there is a j such that $P(j) > 0$ and $Q(j) = 0$, which implies that $P(j)\log\left(\frac{Q(j)}{P(j)}\right) = \infty$. In addition, if $|\Omega_P| < \left|\Omega_Q\right|$, then there is a j such that $P(j) = 0$ and $Q(j) > 0$, which implies that $P(j)\log\left(\frac{Q(j)}{P(j)}\right) = 0$, but one may want to penalize it. The Jensen-Shanon divergence overcomes these issues for the problem treated in this paper. 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

$$D_{JS}(P \parallel Q) = \frac{1}{2}D_{KL}(P \parallel M) + \frac{1}{2}D_{KL}(Q \parallel M). \tag{14}$$

4.3. Power transmission transformers data

The data consists of $n_F = 15$, $n_B = 45$, $n_{TF} = 35$, and $n_C = 5$. Fig. 4(a) shows the log-likelihood $l\left(\hat{\mathbf{T}};\mathcal{D}_{PTD}\right)$ (LL) in dash–dot blue

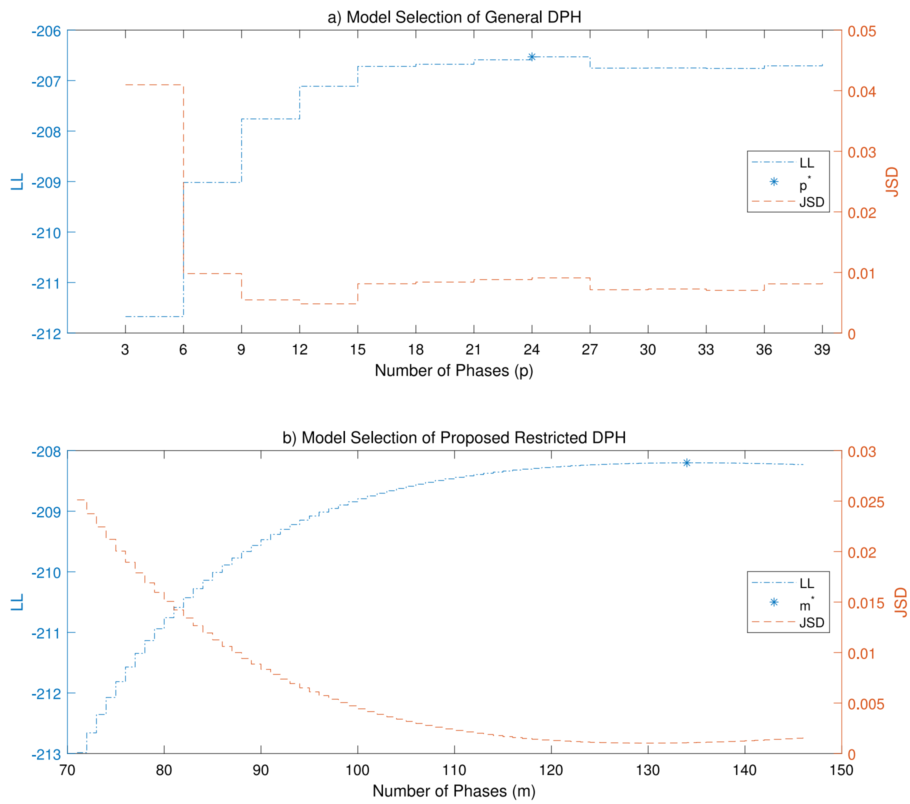

Fig. 4. Model selection of general and proposed DPH. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

lines, and the Jensen-Shanon divergence $D_{JS}\left(\text{DPH}\left(\hat{\mathbf{T}}\right) \parallel \text{Weibull}\,(35,3)\right)$ (JSD) in dashed red line against the number of phases p for the general DPH. One can observe that the log-likelihood is maximized at $\hat{p} = 24$ by solving (13) using MATLAB SQP Solver, where eight different initial points are used to handle local optima, whereas the $D_{JS}\left(\text{DPH}\left(\hat{\mathbf{T}}\right) \parallel \text{Weibull}\,(35,3)\right)$ is minimized at $p^* = 12$ by using (14). One can notice that there is overfitting as the JSD for $p = 24$ is 0.0048, which is lower than the one for $p = 12$, which is valued at 0.0091, where their respective LL is -207.11 and -206.53. Hence, one can distinguish that when fitting a general phase-type distribution, a model-selection procedure is needed, which takes into account model complexity [1,16]. The detailed results of the general DPH are given in Appendix. Fig. 4(b) shows the log-likelihood $l\left(\hat{\mu}, m; \mathcal{D}_{PTD}\right)$ (LL) in dash–dot blue line, and the Jensen-Shanon divergence $D_{JS}\left(\text{DPH}\left(\hat{\mu}, m\right) \parallel \text{Weibull}\,(35,3)\right)$ (JSD) in dashed red line against the parameter m. One can observe that the log-likelihood is maximized at $\hat{m} = 134$ by solving (12), whereas the $D_{JS}\left(\text{DPH}\left(\hat{\mu}, m\right) \parallel \text{Weibull}\,(35,3)\right)$ is minimized at $m^* = 129$ by using (7) and (14). The estimate $\hat{m}$ is a little bit over-estimated compared to the corresponding m^*, which represents the true estimate value for the given data. This may be due to the repartition of the samples being left-skewed and also due to not enough observations on the right-censored part. Also, note that as the sample size increases, $\hat{m}$ will get closer to m^* as ML method is used. In addition, one can observe that the curvature of LL and of D_{JS} is quite similar, which shows that even though the estimate $\hat{m}$ is a little bit over-estimated, the $D_{JS}\left(\text{DPH}\left(\hat{\mu}, m\right) \parallel \text{Weibull}\,(35,3)\right)$ value does not change much whether $\hat{m}$ or m^* is used.

In Fig. 5, the graph (a) shows a stacked histogram of failure observations in dark blue, of right-censored observations in light blue, of left-truncated failure observations in green, and of left-truncated right-censored observations in yellow. It also shows the original distribution in a bold solid black line, the fitted Weibull distribution by the EM algorithm proposed by Balakrishnan & Mitra (2012) [3] in a dashed blue line, the fitted proposed restricted DPH in a bold dashed red line, the fitted general DPH with 12 phases in a bold dash–dot magenta line, and the fitted general DPH with 24 phases in a bold dotted green line. The histogram shows that the data is composed of few failure observations and they are all left-skewed; hence, the use of the censored and truncated observations are needed for a relevant estimation. The estimated parameters are $\hat{m} = 134$ and $\hat{\mu} = 2.859$ by maximizing (12) using quasi-Newton method. The Jensen Shanon divergence for the proposed method is $D_{JS}\left(\text{DPH}\left(\hat{\mu}, \hat{m}\right) \parallel \text{Weibull}\,(35,3)\right) = 0.00118$. The estimated parameters of the Weibull distribution by the EM algorithm are $\hat{\lambda} = 34.36$ and $\hat{\mu} = 2.924$, where $D_{JS}\left(\text{Weibull}\left(\hat{\lambda}, \hat{\mu}\right) \parallel \text{Weibull}\,(35,3)\right) = 0.00046$. The Jensen Shanon divergence for the DPH with 12 phases is $D_{JS}\left(\text{DPH}\left(\hat{\mathbf{T}}\right) \parallel \text{Weibull}\,(35,3)\right) = 0.0048$, whereas the one with 24 phases is $D_{JS}\left(\text{DPH}\left(\hat{\mathbf{T}}\right) \parallel \text{Weibull}\,(35,3)\right) = 0.0091$. One can graphically observes that the general DPH for both phases overfits the data as the estimated ones are bi-modal, whereas the proposed DPH linked to the pre-specified hazard sequence has a similar shape than the original one. This can be expected as the proposed restricted DPH has only two parameters, whereas the general DPH with 12 phases has 23 parameters, and the one with 24 phases has 47 parameters. The graph (b) shows the original Weibull hazard rate in a bold solid black line, the hazard rate of the fitted Weibull by EM algorithm in a solid blue line, the fitted DPH's hazard sequence in a bold solid red stairs, the fitted general DPH's hazard sequence with 12 phases in a bold dash–dot magenta stairs, and the fitted general DPH's hazard sequence

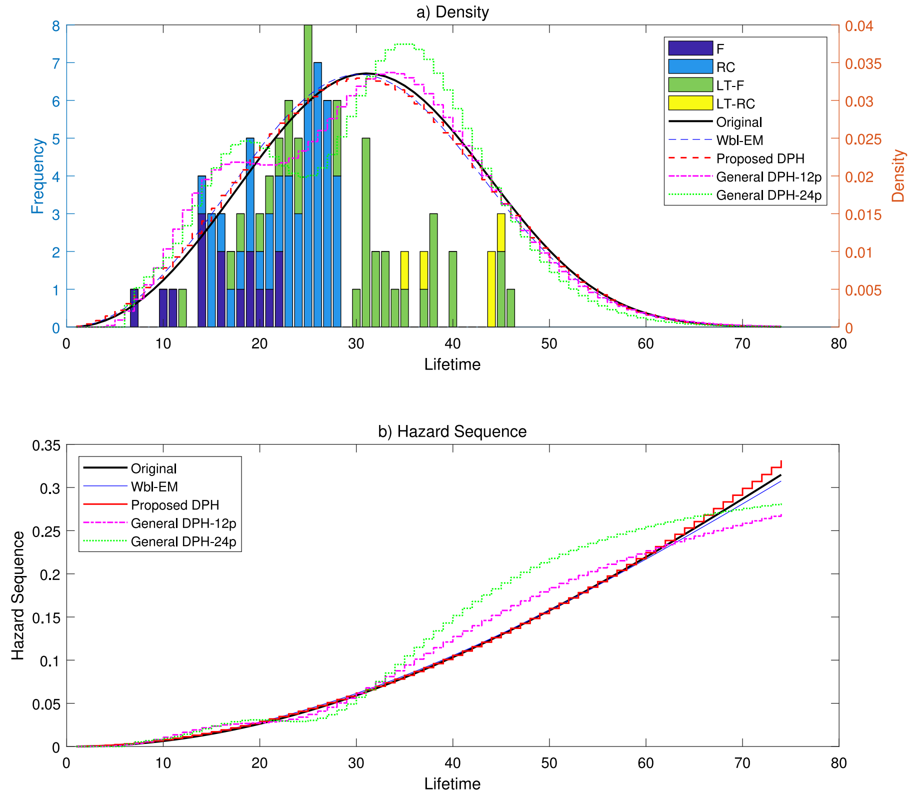

Fig. 5. Approximation of power transmission transformers data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with 24 phases in a bold dotted green stairs. The fitted DPH's hazard sequence is closer or similar to the original one than the fitted Weibull's one up to around 60 years, but then the other method behaves better. Moreover, even though larger error can be seen from 70 years, the error in terms of D_{JS} is low because it corresponds to the right tail, which is practically near 0. Both methods achieve similar result. However, the shape of the general DPH distributions differs from the original one due to overfitting. Therefore, the fitted proposed restricted DPH achieves a better performance than the fitted general DPH as it is less susceptible to skewed data as the general DPH overfits the data easily when one considers only the LL metric. In addition, the fitted restricted DPH allows to analyze multi-component systems with different Weibull distributions easily thanks to its Markovian deterioration, where the proposed method is able to fit transition probabilities of maintenance optimization MDP models from incomplete observations.

5. Conclusion

This paper proposes a ML method to fit a DPH distribution linked to a pre-specified hazard sequence from left-truncated and right-censored observations, where the pre-specified hazard sequence has a similar form than the discrete Weibull distribution's hazard sequence. By doing so, the proposed DPH has only two parameters that are independent to the number of phases, which overcomes the curse of dimensionality of general phase-type distribution. For the experiment with the power transformers data, the method presented in this paper is compared to the Balakrishan & Mitra's EM algorithm, and the fitted general DPH distributions via DM, where similar performance is shown with the fitted Weibull distribution, but achieves better result when compared to the general DPH as it is less prone to overfitting. In addition, the fitted proposed restricted DPH is more suited for modeling complex systems thanks to its Markovian deterioration as it uses an aging process.

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. (2016) [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) we are also working on bathtube hazard sequence for the proposed DPH since it is important to incorporate the full life-cycle of a system. (iii) Finally, one can extend the proposed method to continuous phase type distribution, but we believe that it will be challenging to solve it efficiently.

Acknowledgments

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. Detailed analysis of general DPH

Table A.1 shows the log-likelihood, the Jensen-Shanon divergence, and its computation time in second against the number of phases used. In addition, as the log-likelihood function (11) may possess several local optima, eight different initial points are used to solve (13), and the one that achieves the best LL value is chosen for each phase p.

Table A.1
Analysis of general DPH.

# of phases	LL	JSD	CT (s)
3	−211.68	0.0410	1
6	−209.02	0.0098	11
9	−207.76	0.0055	32
12	−207.11	0.0048	48
15	−206.72	0.0081	130
18	−206.68	0.0084	186
21	−206.59	0.0088	292
24	−206.53	0.0091	468
27	−206.75	0.0071	1123
30	−206.75	0.0073	2084
33	−206.76	0.0070	3697
36	−206.71	0.0081	5770
39	−206.67	0.0085	8465

References

[1] S. Asmussen, O. Nerman, M. Olsson, Fitting phase-type distributions via the EM algorithm, Scand. J. Stat. (1996) 419–441.
[2] P.B. Axel Thummler, M. Telek, A novel approach for phase-type fitting with the EM algorithm, IEEE Trans. Dependable Secure Comput. 3 (3) (2006) 245–258.
[3] N. Balakrishnan, D. Mitra, Left truncated and right censored Weibull data and likelihood inference with an illustration, Comput. Statist. 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] P.T. Boggs, J.W. Tolle, Sequential quadratic programming, Acta Numer. 4 (1995) 1–51.
[6] D.M. Endres, J.E. Schindelin, A new metric for probability distributions, IEEE Trans. Inform. Theory (2003).
[7] 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.
[8] Y. Hong, W.Q. Meeker, J.D. McCalley, et al., Prediction of remaining life of power transformers based on left truncated and right censored lifetime data, Ann. Appl. Stat. 3 (2) (2009) 857–879.
[9] A.K. Jardine, A.H. Tsang, Maintenance, Replacement, and Reliability: Theory and Applications, CRC press, 2005.
[10] Y.M. Ko, E. Byon, Reliability evaluation of large-scale systems with identical units, IEEE Trans. Reliab. 64 (1) (2015) 420–434.
[11] W.Q. Meeker, L.A. Escobar, Statistical Methods for Reliability Data, John Wiley & Sons, 2014.
[12] D. Mitra, Likelihood Inference for Left Truncated and Right Censored Lifetime Data (Ph.D. thesis), McMaster University, 2013.
[13] J. Nocedal, S. Wright, Numerical Optimization, Springer Science & Business Media, 2006.
[14] H. Okamura, T. Dohi, K.S. Trivedi, A refined EM algorithm for PH distributions, Perform. Eval. 68 (10) (2011) 938–954.
[15] H. Okamura, T. Dohi, K.S. Trivedi, Improvement of expectation–maximization algorithm for phase-type distributions with grouped and truncated data, Appl. Stoch. Models Bus. Ind. 29 (2) (2013) 141–156.
[16] M. Olsson, Estimation of phase-type distributions from censored data, Scand. J. Stat. (1996) 443–460.
[17] J.E. Ruiz-Castro, R. Pérez-Ocón, G. Fernández-Villodre, Modelling a reliability system governed by discrete phase-type distributions, Reliab. Eng. Syst. Saf. 93 (11) (2008) 1650–1657.
[18] M. Szymkowiak, M. Iwińska, Characterizations of discrete Weibull related distributions, Statist. Probab. Lett. 111 (2016) 41–48.