Flexible Methods for Reliability Estimation Using Aggregate Failure-time Data

SAMIRA KARIMI[1], HAITAO LIAO[1,*] and NENG FAN[2]

[1]*Department of Industrial Engineering, University of Arkansas, Fayetteville, AR 72701, USA*

Email: sakarimi@email.uark.edu, liao@uark.edu

[2]*Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ 85721, USA*

Email: nfan@email.arizona.edu

Abstract

The actual failure times of individual components are usually unavailable in many applications. Instead, only aggregate failure-time data are collected by actual users due to technical and/or economic reasons. When dealing with such data for reliability estimation, practitioners often face challenges of selecting the underlying failure-time distributions and the corresponding statistical inference methods. So far, only the Exponential, Normal, Gamma and Inverse Gaussian distributions have been used in analyzing aggregate failure-time data because these distributions have closed-form expressions for such data. However, the limited choices of probability distributions cannot satisfy extensive needs in a variety of engineering applications. Phase-type (PH) distributions are robust and flexible in modeling failure-time data as they can mimic a large collection of probability distributions of nonnegative random variables arbitrarily closely by adjusting the model structures. In this paper, PH distributions are utilized, for the first time, in reliability estimation based on aggregate failure-time data. A maximum likelihood estimation (MLE) method and a Bayesian alternative are developed. For the MLE method, an expectation-maximization (EM) algorithm is developed for parameter estimation, and the corresponding Fisher information is used to construct the confidence intervals for the quantities of interest. For the Bayesian method, a procedure for performing point and interval estimation is also introduced. Numerical examples show that the proposed PH-based reliability estimation methods are quite flexible and alleviate the burden of selecting a probability distribution when the underlying failure-time distribution is general or even unknown.

Keywords: Aggregate failure-time data; Phase-type distributions; Maximum likelihood estimation; Bayesian method.

1. Introduction

1.1 Background

The accuracy and authenticity of failure-time data play an important role in the reliability analysis of a product. One source of failure-time data is from laboratory life tests. However, a common issue in using laboratory test data is that some of unknown influential factors (e.g., ambient temperature, humidity, corrosive gas, ultraviolet light) exposed by the product in the field may be ignored in the tests. As a result, the outcome of laboratory tests may not be consistent with the behavior of the product's lifetime in the field. Another source of failure-time data is provided by the actual users of the product. Obviously, it is quite valuable to utilize the rich sources of field data for product reliability estimation as such data reflect the behavior of the product under the real usage conditions. For this reason, organizations, such as the U.S. Department of Defense, have collected a large volume of failure data (OREDA, 2009; Mahar et al., 2011; Denson et al., 2014).

One hurdle of using field data is that the exact failure times of individual components are usually unavailable. In many applications, the collected data contains the number of failed components in a single position of a system along with the system's cumulative operating time until the last failure. This type of data is called aggregate failure-time data. Figure 1 shows an example of aggregate data. For a specific component in each system, the user replaces it whenever it fails without recording the actual failure time. Eventually, a data point representing the time from the first installation to the last component failure (e.g., m_2 failures [replacements] in System #2) is reported. Compared to laboratory testing data with actual failure times, the aggregate data is more concise (Chen and Ye, 2017a). To estimate the product reliability from such aggregate data, only a few probability distributions (i.e., Exponential, Normal, Gamma and Inverse Gaussian (IG)) have been used because their closed-form expressions for aggregate data are available. For other widely used probability distributions, such as Weibull and Lognormal, the closed-form expressions are not attainable. Apparently, the limited choices of probability distributions cannot satisfy extensive needs in many engineering applications where only aggregate data are reported. To assist practitioners in using abundant aggregate data, it is necessary to develop a flexible approach and the corresponding statistical inference methods beyond the use of limited probability distributions for reliability estimation.

Phase-type (PH) distributions are robust and flexible in modeling failure-time data as they can mimic a large collection of probability distributions of nonnegative random variables arbitrarily closely by adjusting the model structures. In this paper, PH distributions are utilized to model aggregate data for the first time. A new expectation-maximization (EM) algorithm is developed to obtain the maximum likelihood estimates

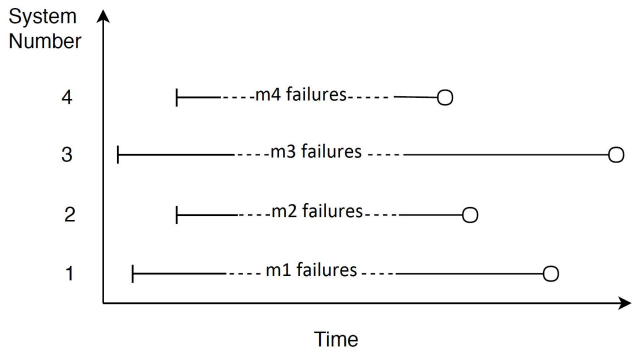

Figure 1: An example of aggregate data: the number of failures (e.g., m_1, ..., m_4) and the time interval from the first installation till the death of the last component.

(MLE) of model parameters. A Bayesian alternative is also introduced to incorporate prior knowledge in parameter estimation. For both methods, the interval estimates for the quantities of interest are derived.

1.2 Related Work

The Exponential distribution has been widely used in reliability for modeling failure-time data. Because of its tractability, aggregate data are often collected and analyzed using this distribution. Coit and Dey (1999) developed an approach to analyze Type II censored data when individual failure times were not available. They also presented a hypothesis test to examine the Exponential distribution assumption and tested their specific data set from a Weibull distribution. Regarding the use of Gamma distribution, Coit and Jin (2000) developed an MLE procedure for handling aggregate failure-time data. A quasi-Newton method was used to find the MLE of model parameters. Chen and Ye (2017a) proposed random effects models based on the Gamma and IG distributions to handle aggregate data. Later, Chen and Ye (2017b) provided a collection of approaches to handle aggregate data using the Gamma and IG distributions. It is worth pointing out that interval estimation of quantity of interest using individual failure-time data has been extensively studied (see Bhaumik et al., 2009), but much less effort has been taken on the analysis of aggregate data. Chen and Ye (2017b) proposed powerful interval estimation algorithms for the Gamma and IG distributions using aggregate data. An extension to the analysis of aggregate lifetime data is the modeling of time-censored aggregate data. This type of data is also abundant, and Chen et al. (2020) proposed models for the analysis of this type of data under a Bayesian framework for Gamma and IG distributions.

When dealing with aggregate data using probability distributions other than the Exponential, Normal, Gamma and IG distributions, an intuitive idea is to perform distribution approximation. To approximate probability distributions for data analysis, extensive studies have been focused mainly on the use of Lognormal distribution (Beaulieu and Rajwani, 2004; Beaulieu and Xie, 2004; Lam and Le-Ngoc, 2007; Mehta et

al., 2007; Cobb et al.; 2012, Asmussen et al., 2016), mixture of Weibull distributions (Bučar et al., 2004; Jin and Gonigunta, 2010; Elmahdy and Aboutahoun, 2013) and the Laplace method (Rizopoulos et al.; 2009, Rue et al., 2009; Asmussen et al., 2016). Moreover, PH distributions are proved to be able to approximate a large collection of probability distributions of non-negative random variables arbitrarily closely. Because of this, a large amount of work has been done on approximating general distributions with PH distributions. The most straightforward method is to match the first k moments of a PH distribution with those of a target distribution. For example, Marie (1980) proposed a moment-matching method using the Coxian distribution for distributions with square coefficient of variation (C^2) greater than 1 and the generalized Erlang distribution for those with C^2 less than 1. Telek and Heindl (2003) matched two-phase acyclic PH distributions with no mass probability at 0 for distributions with $C^2 \geq \frac{1}{2}$. To make approximation more accurate and general, Osogami and Harchol-Balter (2003) and Osogami and Harchol-Balter (2006) proposed an algorithm for mapping a general distribution to a PH distribution by matching the first three moments. Horvath and Telek (2007) proposed an approximation approach for matching the first $2N-1$ moments for an acyclic PH distribution with N phases. Other than moments matching, some studies have been focused on matching the shape of a desired distribution via PH approximation (Starobinski and Sidi (2000), Riska et al. (2004)).

PH distributions have been applied in queueing, healthcare, risk analysis, and reliability. In the area of reliability, Delia and Rafael (2008) modeled a deteriorating system involving both internal and external failures and applied PH distributions to two different repair types. Kharoufeh et al. (2010) introduced a hybrid, degradation-based component reliability model considering environmental effects by PH distribution. Segovia and Labeau (2013) investigated the reliability of a multi-state system subject to internal wear-out and external shocks using a PH distribution. Liao and Guo (2013) modeled accelerate life testing (ALT) data using the Erlang-Coxian distribution. Liao and Karimi (2017) proposed a flexible method for analyzing ALT data using a PH distribution. More recently, Cui and Wu (2019) used PH distributions to model multistate systems with competing failure modes. Li et al. (2019) studied deteriorating structures using PH distributions. Xu et al. (2020) evaluated the reliability of smart meters subject to degradation and shocks based on PH distributions. In the literature, however, PH distributions have never been utilized in modeling aggregate failure-time data. To alleviate the burden of selecting probability distributions and provide a flexible means for data analysis, this paper studies the use of PH distributions in modeling aggregate failure-time data for the first time.

A technical challenge of using PH distributions is model parameter estimation. Asmussen et al. (1996) developed an EM algorithm to obtain the MLE of model parameters. They also used the EM algorithm to minimize information divergence in density approximation. Since the EM algorithm is computationally

intensive, Okamura et al. (2011) proposed a refined EM algorithm to reduce the computational time using uniformization and an improved forward-backward algorithm. As an alternative, under the framework of Bayesian statistics, Bladt et al. (2003) used a Markov chain Monte Carlo (MCMC) method combined with Gibbs sampling for general PH distributions. Watanabe et al. (2012) also presented an MCMC approach to fit PH distributions while using uniformization and backward likelihood computation to reduce the computational time. Ausín et al. (2008) and McGrory et al. (2009) explored two special cases of PH distributions (i.e., Erlang and Coxian) through a Reversible Jump Markov chain Monte Carlo (RJMCMC) method. Yamaguchi et al. (2010), and Okamura et al. (2014) presented variational Bayesian methods to improve the computational efficiency of PH estimation in comparison to MCMC. It is worth pointing out that all of these estimation methods were not developed for aggregate data. In this paper, efforts will be focused on developing a collection of new MLE and Bayesian methods for the analysis of aggregate failure-time data.

1.3 Overview

The remainder of this paper is organized as follows. Section 2 introduces PH distributions. Section 3 provides the statistical procedures of the proposed MLE method, including the EM algorithm and the use of Fisher information for interval estimation. The Bayesian alternative is presented in Section 4 for both parameter and credible interval estimation. In Section 5, numerical examples are provided to illustrate the practical use of the proposed PH-based aggregate data analysis methods. A simulation study shows the strength of PH distribution in dealing with aggregate data from an arbitrarily selected probability distribution, and the coverage probability of the proposed Normal approximate interval estimation method is compared with the one obtained via nonparametric bootstrapping. In addition, a real dataset is also analyzed to demonstrate the practical use of the proposed methods in industrial statistics. Finally, conclusions are drawn in Section 6.

2. PH Distributions

A PH distribution describes the time to absorption of a Continuous-time Markov Chain (CTMC) defined on a finite-state space. Consider a finite-state CTMC $X(t)_{t\geq 0}^{\infty}$ with N transient states and an absorbing state $N+1$, then the CTMC with the specific structure can be described by an infinitesimal generator matrix:

$$\mathbb{Q} = \begin{pmatrix} 0 & \mathbf{0}' \\ \mathbf{S}^0 & \mathbb{S} \end{pmatrix}. \tag{1}$$

where $\mathbf{0}' = [0, ..., 0]$, $\mathbb{S}$ is the subgenerator matrix of the transition rates between the transient states, and $\mathbf{S}^0 = -\mathbb{S}\mathbf{1}$ represents the absorption rates with $\mathbf{1} = [1, ..., 1]^T$ (Buchholz et al., 2014). In particular, the transition rate matrix of an acyclic CTMC can be expressed as:

$$\mathbb{S} = \begin{pmatrix} -\lambda_1 & p_{12}\lambda_1 & p_{13}\lambda_1 & \cdots & p_{1N}\lambda_1 \\ 0 & -\lambda_2 & p_{23}\lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -\lambda_{N-1} & p_{(N-1)N}\lambda_{N-1} \\ 0 & \cdots & \cdots & 0 & -\lambda_N \end{pmatrix}, \tag{2}$$

where $0 \leq p_{ij} \leq 1, i < j, i = 1, 2, ..., N-1, j = 1, 2, ..., N$, and $\sum_{j=1}^{N} p_{ij} \leq 1$.

The probability density function (PDF) and cumulative distribution function (CDF) of PH distribution are:

$$f(t) = \boldsymbol{\pi} e^{\mathbb{S}t}\mathbf{S}^0, \qquad F(t) = 1 - \boldsymbol{\pi} e^{\mathbb{S}t}\mathbf{1}, \tag{3}$$

respectively, where $\boldsymbol{\pi} = [\pi_1, ..., \pi_k, ..., \pi_N]$ is the initial probability vector with $\sum_{k=1}^{N} \pi_k = 1$ that describes the probability of the process being started in each phase.

The most popular PH distributions are the Exponential, Erlang, Hyper-exponential, Hypo-exponential, Hyper-Erlang, and Coxian distributions. Specially, Coxian distribution has been widely used for resolving the non-identifiability problem of PH distributions. Figure 2 shows the CTMC of an N-phase Coxian distribution. The transition rate matrix of an N-phase Coxian distribution is sparse, which has zero p_{ij}'s, except $p_{i(i+1)}$'s for $i = 1, 2, \cdots, N-1$.

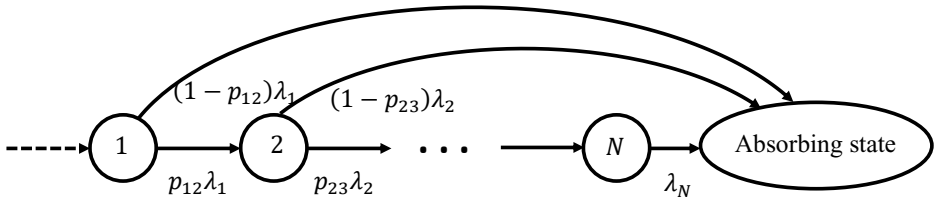

Figure 2: CTMC of an N-Phase Coxian distribution.

In this work, for the purpose of parameter estimation, the following reparameterization is used for the

transition rate matrix of an N-phase Coxian distribution:

$$\mathbb{S} = \begin{pmatrix} -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots & 0 \\ 0 & -(\lambda_2+\mu_2) & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & -(\lambda_{N-1}+\mu_{N-1}) & \lambda_{N-1} \\ 0 & \cdots & \cdots & 0 & -\mu_N \end{pmatrix}. \tag{4}$$

Then, the absorption rate matrix can be expressed as $\mathbf{S}^0 = [\mu_1, \mu_2, ..., \mu_N]^T$. In practice, Coxian distribution emerges as a very flexible distribution while carrying considerably less parameters than general PH distributions. Indeed, the number of parameters in a general PH distribution is $O(N^2)$ while for Coxian, it becomes $O(N)$ which justifies the use of Coxian in practice. Moreover, it can be shown that any acyclic PH distribution can be converted to a Coxian distribution. Because of its flexibility and structural simplicity, Coxian distribution is used in this paper although the proposed methods can be applied to other PH distributions.

3. Maximum Likelihood Estimation

3.1 EM Algorithm for Individual Failure-time Data

An EM algorithm for estimating the parameters of a PH distribution was first proposed by Asmussen et al. (1996). Given each individual failure time, one needs to deal with having a number of unobserved sojourning times in those transient states of CTMC. The likelihood function can be rewritten as:

$$\mathcal{L}((\boldsymbol{\pi}, \mathbb{S})|\boldsymbol{\tau}) = f(z|(\boldsymbol{\pi}, \mathbb{S})) = \prod_{i=1}^{N} \pi(i)^{B_i} \prod_{i=1}^{N} e^{Z_i \mathbb{S}(i,i)} \prod_{i=1}^{N} \prod_{j=1}^{N+1} \mathbb{S}(i,j)^{N_{ij}}, \tag{5}$$

where $\boldsymbol{\tau} = (t_1, t_2, \cdots, t_M)$ contains M observed individual failure times, z represents the complete observation, and B_i, N_{ij} and Z_i are the missing values of the data representing the number of times the Markov process started in phase i, the number of jumps from phase i to phase j, and the total time spent in phase i, respectively, for $i = 1, 2, ..., N$ in an N-phase PH distribution.

Note that this likelihood function is evaluated using the estimated values of the unobserved data obtained in the Expectation step (E-step). To do this, a few statistics are defined in advance:

$$\mathbf{f}_{(\boldsymbol{\pi},\mathbb{S}),t} = \boldsymbol{\pi} e^{\mathbb{S}t}, \qquad \mathbf{b}_{(\boldsymbol{\pi},\mathbb{S}),t} = e^{\mathbb{S}t}\mathbf{S}^0, \qquad \mathbf{F}_{(\boldsymbol{\pi},\mathbb{S}),t} = \int_0^t (\mathbf{f}_{(\boldsymbol{\pi},\mathbb{S}),t-u})^T (\mathbf{b}_{(\boldsymbol{\pi},\mathbb{S}),u})^T du. \tag{6}$$

Then, the conditional expectation of unobserved variables are calculated using the current estimates of model parameters as:

$$E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[B_i] = \frac{1}{M}\sum_{k=1}^{M}\frac{\boldsymbol{\pi}(i)\mathbf{b}_{(\boldsymbol{\pi},\mathbb{S}),t_k}(i)}{\boldsymbol{\pi}\mathbf{b}_{(\boldsymbol{\pi},\mathbb{S}),t_k}}, \tag{7}$$

$$E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[Z_i] = \frac{1}{M}\sum_{k=1}^{M}\frac{\mathbf{F}_{(\boldsymbol{\pi},\mathbb{S}),t_k}(i,i)}{\boldsymbol{\pi}\mathbf{b}_{(\boldsymbol{\pi},\mathbb{S}),t_k}}, \tag{8}$$

$$E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[N_{ij}] = \frac{1}{M}\sum_{k=1}^{M}\frac{\mathbb{S}(i,j)\mathbf{F}_{(\boldsymbol{\pi},\mathbb{S}),t_k}(i,j)}{\boldsymbol{\pi}\mathbf{b}_{(\boldsymbol{\pi},\mathbb{S}),t_k}}, \tag{9}$$

$$E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[N_{in+1}] = \frac{1}{M}\sum_{k=1}^{M}\frac{\mathbf{S}^0(i)\mathbf{f}_{(\boldsymbol{\pi},\mathbb{S}),t_k}(i)}{\boldsymbol{\pi}\mathbf{b}_{(\boldsymbol{\pi},\mathbb{S}),t_k}}, \tag{10}$$

where $i, j = 1, 2, \cdots, N$. In the M-step, the parameters of the distribution are re-estimated using the current estimate of the complete data (Buchholz et al., 2014):

$$\hat{\boldsymbol{\pi}}(i) = E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[B_i], \qquad \hat{\mathbb{S}}(i,j) = \frac{E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[N_{ij}]}{E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[Z_i]},$$
$$\hat{\mathbf{S}}^0(i) = \frac{E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[N_{in+1}]}{E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[Z_i]}, \qquad \hat{\mathbb{S}}(i,i) = -(\hat{\mathbf{S}}^0(i) + \sum_{i\neq j}^{n}\hat{\mathbb{S}}(i,j)). \tag{11}$$

Note that this EM algorithm monotonically improves the likelihood value to achieve the MLE of model parameters. However, it was developed only for individual failure-time data.

3.2 Proposed EM Algorithm for Aggregate Failure-time Data

The previous EM algorithm uses each data point in the E-step to contribute to estimating the unobserved or missing values. As it can be seen, from each data point one value for each Z_i, B_i and N_{ij} can be found for each phase of the distribution, and the mean values of these give the expected values of the variables in the E-step.

The challenge of using aggregate data, however, is that each data point corresponds to PH distributions with different numbers of failures. This causes the underlying distributions for different data points to have different numbers of phases. Unlike individual failure-time data, in this case we have independent but not identically distributed variables. Considering m_k as the number of failed components for data point k, the data point follows a PH distribution with Nm_k phases. As a result, the transition rate matrix for m_k failures is an $(Nm_k) \times (Nm_k)$ matrix. So, it is necessary to determine how and for which phases those variables should be estimated (Karimi et al., 2019).

Primarily, the most important aspects are finding the resulting transition rate matrix for the sum of a number of similar N-phase PH variables and deriving the properties of the resulting distribution to develop an EM algorithm for the case of aggregate data. For some distributions, such as Gamma, this is straightforward. In the case of sum of m similar Gamma variables, the resulting variable will follow a Gamma distribution with shape parameter equal to m times the shape parameter of the single variable. However, this turns out to be a more challenging issue for PH distribution and requires further analysis of the parameters that are in a matrix format.

Clearly, if variable C is the sum of two PH variables A and B, the transition rate matrix of C can be shown as:

$$\mathbb{S}^{(C)} = \begin{pmatrix} \mathbb{S}^{(A)} & \mathbf{S}^0\boldsymbol{\pi}^{(B)} \\ \mathbf{0} & \mathbb{S}^{(B)} \end{pmatrix}, \tag{12}$$

and the initial probability vector becomes $\boldsymbol{\pi}^{(C)} = [\boldsymbol{\pi}^{(A)}, \boldsymbol{\pi}^{(A)}(N+1)\boldsymbol{\pi}^{(B)}]$. The term $\boldsymbol{\pi}^{(A)}(N+1)$ is the probability of process A starting in an absorption state that is considered 0 here, so $\boldsymbol{\pi}^{(C)} = [\boldsymbol{\pi}^{(A)}, \mathbf{0}_{1\times N}]$. When modeling aggregate failure-time data, the sum of m_k similar PH variables has a transition rate matrix consisting of submatrices equal to the single PH variable transition rate matrix and failure vector. The design of these matrices is in the following form:

$$\mathbb{S}^{(new)} = \begin{pmatrix} \mathbb{S} & \mathbf{S}^0\boldsymbol{\pi} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbb{S} & \mathbf{S}^0\boldsymbol{\pi} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & & \ddots & \mathbb{S} & \mathbf{S}^0\boldsymbol{\pi} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbb{S} \end{pmatrix}_{Nm_k \times Nm_k}. \tag{13}$$

In the previous EM algorithm, each element estimated in the E-step is comprised of the mean of M matrices driven from the data set. In case of aggregate data, the sizes of matrices contributing to calculating the missing values are different. As such, a different approach should be developed for model parameter estimation.

More specially, matrices $\mathbf{f}_{(\boldsymbol{\pi},\mathbb{S}),t_k}$, $\mathbf{b}_{(\boldsymbol{\pi},\mathbb{S}),t_k}$ and $\mathbf{F}_{(\boldsymbol{\pi},\mathbb{S}),t_k}$ have different dimensions for different data points. Indeed, $\mathbf{f}_{(\boldsymbol{\pi},\mathbb{S}),t_k}$ is a $1 \times Nm_k$, $\mathbf{b}_{(\boldsymbol{\pi},\mathbb{S}),t_k}$ is an $Nm_k \times 1$ and $\mathbf{F}_{(\boldsymbol{\pi},\mathbb{S}),t_k}$ is an $Nm_k \times Nm_k$ matrix. Consider $\mathbf{f}$ and $\mathbf{b}$ as m_k concatenated $1 \times N$ matrices, each of which referring to one component failure and representing an N-phase Markov process. Each matrix $\mathbf{F}$ includes m_k diagonal $N \times N$ submatrices,

each being equivalent to matrix $\mathbf{F}$ for a single component's failure time. Note that in this method matrix $\mathbf{S}^{0(k)}$ shows the rate that in a corresponding phase a component is moved to the absorption of the m_k'th component. To catch the single component absorption rates, the rates of transition to the phases relative to the other components should be added to $\mathbf{S}^{0(k)}$. The values of $\mathbf{S}^{0(k)}$ that are relative to any component except the m_k'th, are zero. For the proposed EM algorithm, a new absorption rate matrix should be defined, which consists of the individual absorption rates. Each $\mathbf{S}^0\boldsymbol{\pi}$ in Equation (13) contributes to absorption rates as follows:

$$\mathbf{S}^0\boldsymbol{\pi} = \begin{pmatrix} \mathbf{S}^0(1)\boldsymbol{\pi}(1) & \mathbf{S}^0(1)\boldsymbol{\pi}(2) & \cdots & \mathbf{S}^0(1)\boldsymbol{\pi}(N) \\ \mathbf{S}^0(2)\boldsymbol{\pi}(1) & \mathbf{S}^0(2)\boldsymbol{\pi}(2) & \cdots & \mathbf{S}^0(2)\boldsymbol{\pi}(N) \\ \vdots & \vdots & \ddots & \cdots \\ \mathbf{S}^0(N)\boldsymbol{\pi}(1) & \mathbf{S}^0(N)\boldsymbol{\pi}(2) & \cdots & \mathbf{S}^0(N)\boldsymbol{\pi}(N) \end{pmatrix}. \tag{14}$$

Consequently, the absorption rate at phase i of each individual component becomes $\mathbf{d}_i = \sum_{j=1}^{N} \mathbf{S}^0(i)\boldsymbol{\pi}(j)$. Thus, the new absorption vector $\mathbf{d}^{(k)}$ could be constructed and used in the E-step as:

$$\mathbf{d}^{(k)} = [\mathbf{d}_1^{(k)}, \cdots, \mathbf{d}_N^{(k)}, \cdots, \mathbf{d}_1^{(k)}, \cdots, \mathbf{d}_N^{(k)}]_{1\times Nm_k} \tag{15}$$

Note that $\mathbf{d}^{(k)}$ is not the actual absorption rate matrix of data point k, but it is the set of hidden absorption matrices related to each individual component failed in that data point.

The likelihood function for this case can be described as Equation (5) after modifying the definitions for some variables. In particular, $\boldsymbol{\tau} = (t_1, t_2, \cdots, t_M, m_1, m_2, \cdots, m_M)$, and $\mathbb{S}$ represents the single component transition rate matrix. Each single component lifetime is related to one Markov process, and for a data point k, m_k Markov processes occur successively. In addition, the unobserved variables, B_i, Z_i, and N_{ij} for the case of aggregate data are defined as:

B_i: the number of times the Markov process started in phase $lN + i$, $l = 0, \cdots, max(m_k) - 1$.

Z_i: the time that was spent in phase $ln + i$; $\quad l = 0, \cdots, max(m_k) - 1$.

N_{ij}: the number of jumps from phase $lN + i$ to phase $lN + j$; $\quad l = 0, \cdots, max(m_k) - 1$.

Using the above definitions, we can make sure that inside the Markov process of each data point, before reaching the absorption state of the current component, transitions to those phases related to the subsequent components are not allowed. Then, the equations for the E-step are:

$$E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[B_i] \;=\; \frac{1}{M}\sum_{k=1}^{p}\sum_{l=0}^{m_k} \frac{\boldsymbol{\pi}^{(k)}(i+ln)\mathbf{b}_{(\boldsymbol{\pi}^{(k)},\mathbb{S}^{(k)}),t_k}(i+ln)}{\boldsymbol{\pi}^{(k)}\mathbf{b}_{(\boldsymbol{\pi}^{(k)},\mathbb{S}^{(k)}),t_k}}, \tag{16}$$

$$E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[Z_i] = \frac{1}{M}\sum_{k=1}^{p}\sum_{l=0}^{m_k}\frac{\mathbf{F}^{(k)}_{(\boldsymbol{\pi}^{(k)},\mathbb{S}^{(k)}),t_k}(i+ln,i+ln)}{\boldsymbol{\pi}^{(k)}\mathbf{b}_{(\boldsymbol{\pi}^{(k)},\mathbb{S}^{(k)}),t_k}}, \tag{17}$$

$$E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[N_{ij}] = \frac{1}{M}\sum_{k=1}^{p}\sum_{l=0}^{m_k}\frac{\mathbb{S}^{(k)}(i+ln,j+ln)\mathbf{F}^{(k)}_{(\boldsymbol{\pi}^{(k)},\mathbb{S}^{(k)}),t_k}(i+ln,j+ln)}{\boldsymbol{\pi}^{(k)}\mathbf{b}_{(\boldsymbol{\pi}^{(k)},\mathbb{S}^{(k)}),t_k}}, \tag{18}$$

$$E_{(\boldsymbol{\pi},\mathbb{S}),\boldsymbol{\tau}}[N_{in+1}] = \frac{1}{M}\sum_{k=1}^{p}\sum_{l=0}^{m_k}\frac{\mathbf{d}^{(k)}(i+ln)\mathbf{f}^{(k)}_{(\boldsymbol{\pi}^{(k)},\mathbb{S}^{(k)}),t_k}(i+ln)}{\boldsymbol{\pi}^{(k)}\mathbf{b}_{(\boldsymbol{\pi}^{(k)},\mathbb{S}^{(k)}),t_k}}, \tag{19}$$

where p is the number of available data points, $M = \sum_{k=1}^{p} m_k$ is the total number of failures, $\boldsymbol{\pi}$ and $\mathbb{S}$ are the estimated initial probability vector and transition rate matrix of a single component's failure time, $\boldsymbol{\pi}^{(k)}$ and $\mathbb{S}^{(k)}$ are those of m_k components and $i, j = 1, 2, \cdots, N$. Using these E-step equations, the M-step can be performed using the formulas stated in Section 3.1. In summary, the proposed EM algorithm for handling aggregate data is as follows:

(i) Define initial values for the parameters of an N-phase PH distribution.

(ii) Define the proper transition rate and absorption matrices for each data point based on the corresponding number of failures.

(iii) Define the statistics of EM algorithm as in Equation (6) separately for each data point.

(iv) Use Equations (16) - (19) to estimate the unobserved data based on the current parameter estimates.

(v) Use Equation (11) to update the parameter estimates.

(vi) If a stopping criterion (e.g., a certain number of iterations or the difference between the likelihood values of the last two iterations) is met, stop. Otherwise, go to step (iv).

3.3 Model Selection and Setting of Initial Values

To avoid the non-identifiability problem of parameters, we have used Coxian distribution. It can be shown that any general PH distribution can be represented by a Coxian distribution. Using Coxian distribution with ordered diagonal values eliminates the redundancy in parameters. As in the EM algorithm, the initial values should be used for the parameters, we suggest an approach to obtain initial parameter values. Based on our experiments, the algorithm is not highly sensitive to the initial parameter values. As long as the initial values are not chosen such that an extremely low likelihood is attained, the algorithm can find its way to the optimum solution. Although this seems like an easy job, in practice, it can be difficult to obtain a reasonable first guess. As Erlang distribution is a special case of Coxian distribution, the parameter estimate of Erlang distribution can be used as the start point. Recall that in the EM algorithm for PH distributions, if a value is initially set to zero, it will be zero in the ML estimate. To avoid all zero values in the absorption rate matrix,

a small value relative to the optimum λ of Erlang distribution can be assigned to the absorption rates.

Since the convolution of Erlang distribution is tractable, it can be easily used for aggregate data. If the lifetime of each component follows $Erlang(N, \lambda)$, an aggregate data point with m failures follows $Erlang(mN, \lambda)$. Then, parameter λ can be found by maximizing the likelihood function:

$$\mathcal{L}((\pi, \mathbb{S})|\tau) = \prod_{k=1}^{p} \frac{\lambda^{m_k N} t_k^{m_k N-1} e^{-\lambda t_k}}{(m_k N - 1)!} \tag{20}$$

where

$$\mathbb{S} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & 0 \\ 0 & -\lambda & \lambda & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & -\lambda & \lambda \\ 0 & \cdots & \cdots & 0 & -\lambda \end{pmatrix}. \tag{21}$$

When fitting a PH distribution to aggregate data, models with different numbers of phases can be considered. Although it can be shown that increasing the number of phases could potentially improve the likelihood, a model selection method is required to determine the most suitable number of phases in some sense. It is worth pointing out that using Akaike Information Criterion (AIC) may not be effective in selecting a PH distribution, even for Coxian distribution, as the number of phases often grows rapidly in comparison to the likelihood value. In this work, the Maximum a Posteriori (MAP) estimation method with a Laplacian prior is used for the purpose of model selection. Specially, the Laplacian prior is denoted as:

$$p(\boldsymbol{\Theta}|(\boldsymbol{\mu}, \boldsymbol{b})) = \prod_{i=1}^{n} \frac{1}{2b_i} e^{-\frac{|\Theta_i - \mu_i|}{b_i}}, \tag{22}$$

where $\boldsymbol{\Theta}$ contains the parameters of the distribution under test, n equals the number of parameters, and $(\boldsymbol{\mu}, \boldsymbol{b})$ is the vector of Laplacian distribution parameters. With the likelihood function $f(\boldsymbol{\tau}|\boldsymbol{\Theta})$ as Equation (5) with $\boldsymbol{\tau} = (t_1, t_2, \cdots, t_M, m_1, m_2, \cdots, m_M)$, the MAP estimator is $\text{argmax}_{\boldsymbol{\Theta}} f(\boldsymbol{\tau}|\boldsymbol{\Theta}) p(\boldsymbol{\Theta}|(\boldsymbol{\mu}, \boldsymbol{b}))$, and the candidate distribution with an appropriate number of parameters that results in the highest MAP value will be selected.

3.4 ML-based Confidence Interval

In this section, a method for finding the confidence intervals of quantities of interest using Fisher information is presented. It is worth pointing out that Fisher information for PH distribution has been studied in the literature, but it has never been extended and used in dealing with aggregate failure-time data.

Let $\boldsymbol{\Theta} = (\boldsymbol{\pi}, vector(\mathbb{S}))'$ be the vector containing all the parameters to be estimated, $\boldsymbol{\tau}$ be the available aggregate data, and $\ell(\boldsymbol{\Theta}|\boldsymbol{\tau}) = \ln \mathcal{L}(\boldsymbol{\Theta}|\boldsymbol{\tau})$ be the log-likelihood function. The empirical Fisher information matrix can be expressed as:

$$I(\boldsymbol{\Theta}) = -\left[\frac{\partial^2}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}'}\ell(\boldsymbol{\Theta}|\boldsymbol{\tau})\right]. \tag{23}$$

Due to the special structure of PH distribution and the fact that the parameters are masked inside the transition rate matrix, the formation of Fisher information matrix is not straightforward. Bladt et al. (2011) proposed an EM algorithm and a Newton-Raphson method to attain the Fisher information matrix for the parameters of PH distribution. In this paper, we will use the Newton-Raphson method for Fisher information matrix estimation and extend their method to deal with aggregate data. To this end, some of the expressions used in their method need to be updated.

First, the Newton-Raphson method is explained here. The derivative of the log-likelihood function with respect to the vector of parameters is:

$$\frac{\partial\ell(\boldsymbol{\Theta}|\boldsymbol{\tau})}{\partial\boldsymbol{\Theta}} = \sum_{k=1}^{M}\frac{1}{f(t_k|\boldsymbol{\Theta})}\frac{\partial f(t_k|\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}, \tag{24}$$

where $f(\cdot)$ is the PDF. Note that taking the derivative of PDF f with respect to $\boldsymbol{\Theta}$ is an issue, for which the following formulas are created. The parameters of an N-phase PH distributions are $N-1$ elements of $\boldsymbol{\pi}$, non-diagonal elements of $\mathbb{S}$, noted as d_{hn}, and all the elements of $\mathbf{S}^0$, noted as d_h, for $h, n = 1, 2, \cdots, N$. To get started, using uniformization for matrix exponential, we define $c = max\{-d_{hh} : 1 < h < N\}$ and $\mathbf{K} = (1/c)\mathbb{S} + \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Then, we have:

$$e^{\mathbb{S}t} = \sum_{r=0}^{\infty} e^{-ct}\frac{ct^r}{r!}\mathbf{K}^r. \tag{25}$$

Let $\boldsymbol{\Psi}(t) = e^{(\mathbb{S}t)}$. By decomposing $\boldsymbol{\pi}$ into $\sum_{j=1}^{p-1}\pi_j\mathbf{e}_j^\top + (1-\sum_{j=1}^{p-1}\pi_j)\mathbf{e}_N^\top$, we have:

$$\frac{\partial f(t|\boldsymbol{\Theta})}{\partial\pi_h} \quad = \quad \mathbf{e}_h^\top\boldsymbol{\Psi}(t)\mathbf{S}^0 - \mathbf{e}_N^\top\boldsymbol{\Psi}(t)\mathbf{S}^0, \tag{26}$$

$$\frac{\partial f(t|\boldsymbol{\Theta})}{\partial d_{hn}} = \boldsymbol{\pi}\frac{\partial \boldsymbol{\Psi}(t)}{\partial d_{hn}}\mathbf{S}^0, \qquad h \neq n, \tag{27}$$

$$\frac{\partial f(t|\boldsymbol{\Theta})}{\partial t_h} = \boldsymbol{\pi}\boldsymbol{\Psi}(t)\mathbf{e}_h + \boldsymbol{\pi}\frac{\partial \boldsymbol{\Psi}(t)}{\partial d_h}\mathbf{S}^0, \tag{28}$$

where $\mathbf{e}_j$ is a column vector with 1 in the jth place and 0 elsewhere. Then, the problem is reduced to calculating the partial derivatives of $\boldsymbol{\Psi}(t)$:

$$\frac{\partial \boldsymbol{\Psi}(t)}{\partial \Theta_q} = e^{-ct}\sum_{s=0}^{\infty}\frac{(ct)^{s+1}}{(s+1)!}\mathbf{D}_q(s) + \frac{\partial c}{\partial \Theta_q}te^{\mathbb{S}t}(\mathbf{K}-\mathbf{I}), \quad q = 1, 2, \cdots, N^2+N-1, \tag{29}$$

where $\mathbf{D}_q(s) = \partial \mathbf{K}^{s+1}/\partial\Theta_q$. To calculate this, partial derivatives of powers of the transition rate matrix with respect to the parameters are required. Especially, we have:

$$\frac{\partial \mathbb{S}^r}{\partial \Theta_q} = \sum_{k=0}^{r-1}\mathbb{S}^k\frac{\partial \mathbb{S}}{\partial \Theta_q}\mathbb{S}^{r-1-k}, \tag{30}$$

where $[\partial\mathbb{S}/\partial d_{ij}]_{ij} = 1$, $[\partial\mathbb{S}/\partial d_{ij}]_{ii} = -1$, $[\partial\mathbb{S}/\partial d_i]_{ii} = -1$, and the rest of the elements are all 0. Based on these basic definitions, the following results can be obtained:

$$\frac{\partial^2 \mathbb{S}^r}{\partial\Theta_p\partial\Theta_q} = \sum_{k=0}^{r-1}\mathbb{S}^k\frac{\partial \mathbb{S}}{\partial\Theta_q}\frac{\partial \mathbb{S}^{r-1-k}}{\partial\Theta_p} + \mathbb{S}^{r-1-k}\frac{\partial \mathbb{S}^k}{\partial\Theta_p}\frac{\partial \mathbb{S}}{\partial\Theta_q}, \tag{31}$$

$$\frac{\partial^2 e^{\mathbb{S}t}}{\partial\Theta_p\partial\Theta_q} = e^{-ct}\sum_{k=0}^{\infty}\frac{(ct)^{k+1}}{(k+1)!}\frac{\partial^2\mathbf{K}^{k+1}}{\partial\Theta_p\partial\Theta_q} + \frac{\partial c}{\partial\Theta_q}t\Big(e^{\mathbf{T}t}\frac{\partial\mathbf{K}}{\partial\Theta_p} + \frac{\partial e^{\mathbf{T}t}}{\partial\Theta_p}(\mathbf{K}-\mathbf{I})\Big), \tag{32}$$

where $p, q = 1, 2, \cdots, N^2+N-1$. It is worth pointing out that these formulas are used to produce a Fisher information matrix based on individual failure-time data (Bladt et al., 2011). In this paper, these formulas are extended to adapt to aggregate failure-time data.

The second derivative of log-likelihood function is:

$$\frac{\partial^2\ell(\boldsymbol{\Theta}|\boldsymbol{\tau})}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}'} = \sum_{k=1}^{m}\frac{1}{f(t_k|\boldsymbol{\Theta})^2}\left[f(t_k|\boldsymbol{\Theta})\frac{\partial^2 f(t_k|\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}'} - \frac{\partial f(t_k|\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}}\frac{\partial f(t_k|\boldsymbol{\Theta})}{\partial\boldsymbol{\Theta}'}\right]. \tag{33}$$

For each aggregate failure-time data t_k, the corresponding PDF is defined based on the number of failed components in that data point, as shown previously. In other words, we have a different PDF (thus a different transition rate matrix and different number of phases), as given in Equation (13), for each data point as:

$$f_k(t|(\boldsymbol{\pi}, \mathbb{S})) = \boldsymbol{\pi}_k e^{\mathbb{S}_k}\mathbf{S}_k^0. \tag{34}$$

Moreover, $\partial\mathbb{S}/\partial\Theta_q$ should be updated. For data point k with m_k failures, $[\partial\mathbb{S}/\partial d_{hn}]_{uv} = 1$ and $[\partial\mathbb{S}/\partial d_{hn}]_{uu} = -1$, where $h, n = 1, 2, \cdots, N$, $h \neq n$, $u = h + rN$, $v = n + rN$, $r = 0, 1, \cdots, N-1$. The rest of the parameters of this $Nm_k \times Nm_k$ matrix will be 0s. The following example shows the derivative of transition rate matrix with respect to a parameter when $N = 3$ and two cumulative failures:

$$\frac{\partial\mathbb{S}_k}{\partial d_{23}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{35}$$

Note that in this case, the only independent parameters are again the parameters of the PH distribution for a single component, and the total number of parameters is $N^2 + N - 1$. So, for each data point, no matter how many failures it contains, $\mathbf{\Theta}$ always contains $N^2 + N - 1$ model parameters.

After producing the Fisher information matrix, the normal approximation method can be used to attain the confidence intervals of interest. The Wald statistic is defined as:

$$W = (\hat{\mathbf{\Theta}} - \mathbf{\Theta})'[\hat{\mathbf{\Sigma}}_{\hat{\mathbf{\Theta}}}]^{-1}(\hat{\mathbf{\Theta}} - \mathbf{\Theta}), \tag{36}$$

where $\hat{\mathbf{\Theta}}$ is the MLE of $\mathbf{\Theta}$, and $\hat{\mathbf{\Sigma}}_{\hat{\mathbf{\Theta}}}$ is the estimated variance-covariance matrix obtained by taking the inverse of Fisher information matrix. Since $\hat{\mathbf{\Theta}}$ asymptotically follows a multivariate Normal distribution with parameters $\mathbf{\Theta}$ and $\mathbf{\Sigma}$, W follows a Chi-square distribution with degrees of freedom equal to the length of $\mathbf{\Theta}$ (i.e., the number of parameters noted as v). Then, a $100(1-\alpha)\%$ approximate confidence region for $\mathbf{\Theta}$ can be obtained from:

$$(\hat{\mathbf{\Theta}} - \mathbf{\Theta})'[\hat{\mathbf{\Sigma}}_{\hat{\mathbf{\Theta}}}]^{-1}(\hat{\mathbf{\Theta}} - \mathbf{\Theta}) \leq \chi^2_{(1-\alpha;v)}. \tag{37}$$

In certain circumstances, the statistics of Wald and likelihood ratio are equivalent, so that the distribution is exact. For other cases, it can be shown that Wald interval is the quadratic approximation to a likelihood-based confidence region (Meeker and Escobar (1995)). Regarding PH distributions, this is an asymptotic approximate method, and exact pivotal quantities for the parameters are not discussed in the literature. In this paper, we provide the ML confidence interval for each individual parameter by Nor-

mal approximation. For example, for transition rate parameter λ_1, we have $[\underset{\sim}{\lambda_1}, \tilde{\lambda_1}] = \left[\hat{\lambda_1}/V, \hat{\lambda_1} \times V\right]$, where $V = \exp\left(z_{1-\alpha/2}\hat{se}_{\hat{\lambda_1}}/\hat{\lambda_1}\right)$, $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard Normal distribution, and $\hat{se}_{\hat{\lambda_1}} = \sqrt{\widehat{Var}(\hat{\lambda_1})}$.

For practical purposes, the confidence interval for the CDF of failure-time distribution is of interest. Based on the estimated variance-covariance matrix, the confidence interval for the CDF can be obtained as follows. First, the variance of the CDF estimate is calculated via the delta method as:

$$\widehat{Var}(F(t)) = \left[\frac{\partial F}{\partial \mu_1}, \cdots, \frac{\partial F}{\partial \lambda_2}\right] \hat{\mathbf{\Sigma}}_{\hat{\mathbf{\Theta}}} \left[\frac{\partial F}{\partial \mu_1}, \cdots, \frac{\partial F}{\partial \lambda_2}\right]'. \tag{38}$$

Then, the approximate confidence interval for the CDF can be expressed as:

$$[\underset{\sim}{F}, \tilde{F}] = \left[\frac{\hat{F}(t)}{\hat{F}(t) + (1-\hat{F}(t)) \times W}, \frac{\hat{F}(t)}{\hat{F}(t) + (1-\hat{F}(t))/W}\right], \tag{39}$$

where

$$W = \exp\left(\frac{z_{1-\alpha/2}\hat{se}_{\hat{F}}}{\hat{F}(t_e)(1-\hat{F}(t_e))}\right) \text{ and } \hat{se}_{\hat{F}} = \sqrt{\widehat{Var}(\hat{F}(t))}. \tag{40}$$

4. Bayesian Alternative

A Bayesian alternative is also provided in this work for reliability estimation using aggregate data. The studies by Ausín et al. (2008) and McGrory et al. (2009) concentrated on Bayesian methods for Coxian distributions. In this section, we will use the method developed by McGrory et al. (2009) and extend their model to estimate the parameters of Coxian based on aggregate data. Moreover, using the same posterior distribution for Coxian and with the assistance of Metropolis-Hastings algorithm, credible intervals are estimated for the model parameters.

McGrory et al. (2009) introduced a Bayesian formulation for a Coxian distribution with covariates and unknown number of phases. They considered a Gamma prior distribution for each parameter. Here, we will utilize the same model while ignoring covariates. For the transition rate matrix of an N-phase Coxian distribution given in Equation (4), we assume that the prior distributions of model parameters are $\lambda_j \sim Gamma(\alpha_j, \beta_j)$, $j = 1, 2, \cdots, N-1$, and $\mu_j \sim Gamma(\gamma_j, \sigma_j)$, $j = 1, 2, \cdots, N$ (McGrory et al.

(2009)). Then, the posterior distribution can be obtained as:

$$\begin{aligned} p(\mathbf{\Theta}_N, N|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{\Theta}_N, N)p(\mathbf{\Theta}_N|N)p(N) \\ &= \prod_{i=1}^{M} \pi e^{(\mathbb{S}_i y_i)} \mathbf{S}_i^0 \prod_{j=1}^{N-1} \frac{1}{\Gamma(\alpha_i)} \frac{1}{\beta_j^{\alpha_j}} \lambda_j^{\alpha_j - 1} exp(-\frac{\lambda_j}{\beta_j}) \\ &\times \prod_{j=1}^{N} \frac{1}{\Gamma(\gamma_j)} \frac{1}{\sigma_j^{\gamma_j}} \mu_j^{\gamma_j - 1} exp(-\frac{\mu_j}{\sigma_j}) \times p(N). \end{aligned} \tag{41}$$

For the case of aggregate data, λ_j and μ_j are the parameters of Coxian distribution for a single component failure, but $\mathbb{S}_i$ and $\mathbf{S}_i^0$ are those related to data point i based on Equation (13), which have different dimensions for different data points. Note that a subscript N is added to the parameter vector $\mathbf{\Theta}$ to emphasize the number of phases of the current model, which may be adjusted.

RJMCMC (Green (1995)) is a method that enables jumps between models with different dimensions. The algorithm proposed by McGrory et al. (2009) considers three main possibilities with equal probabilities: a fixed dimension update of the parameters, splitting the phase into two or combining two existing phases into one, and birth of a new phase or death of an existing phase. In particular, fixed dimension parameter update is done through a Metropolis-Hastings algorithm. For dimension changing reversible jump moves, some basic definitions are needed. Let the current number of phases be N and the proposed number of phases be N^*. In each jump step, the dimension can only increase or decrease by one unit while satisfying the requirement on the maximum and minimum numbers of phases. u, v, u^* and v^* are auxiliary variables defined to keep the dimensionality of the current and proposed parameter spaces, (Θ_N, u, v) and (Θ_{N^*}, u^*, v^*), respectively. We define:

$$R = \frac{p(\mathbf{y}|\mathbf{\Theta}_{N^*}, N^*)p(\mathbf{\Theta}_{N^*})p(N^*)}{p(\mathbf{y}|\mathbf{\Theta}_N, N)p(\mathbf{\Theta}_N)p(N)} \times \frac{Q_{N^*,N}p(u^*, v^*|N^*, N, \mathbf{\Theta}_{N^*})}{Q_{N,N^*}p(u, v|N, N^*, \mathbf{\Theta}_N)} \times \left| \frac{\partial(\mathbf{\Theta}_{N^*}, u^*, v^*)}{\partial(\mathbf{\Theta}_N, u, v)} \right|, \tag{42}$$

where Q_{N,N^*} is the probability of moving from N to N^*, and the third term is the Jacobian for transformation, which will be addressed later. Then, the probability of accepting a proposed move is $min(R, 1)$. To perform a reasonable mapping, it is ensured that the mean time and probability of absorption in current and proposed phase(s) remain similar, such that:

$$\frac{\mu}{\mu + \lambda} = \frac{\mu_a}{\mu_a + \lambda_a} + \big(\frac{\lambda_a}{\mu_a + \lambda_a} \times \frac{\mu_b}{\mu_b + \lambda_b}\big), \tag{43}$$

$$\frac{\mu}{\mu + \lambda} = \frac{1}{\mu_a + \lambda_a} + \frac{1}{\mu_b + \lambda_b}. \tag{44}$$

For split and birth moves, where one new phase is introduced, μ and λ denote the rates before transformation,

and μ_a, μ_b, λ_a and λ_b denote the rates after the transformation. For combine and death moves, the process will be performed reversely.

Accordingly, for each move we need to find the new parameters, based on Equations (43) and (44), as well as the Jacobian of the transformation, which will be succinctly described here. Note that split and combine moves cannot be applied for the final phase, while birth and death moves are only performed on the final phase.

Via combine move: $(\mu_a, \lambda_a, \mu_b, \lambda_b) \rightarrow (u, v, \mu, \lambda)$, where $u = \mu_a$ and $v = \lambda_a$, we have:

$$\mu = \frac{\mu_a\mu_b + \mu_a\lambda_b + \lambda_a\mu_b}{\mu_a + \lambda_a + \mu_b + \lambda_b}, \tag{45}$$

$$\lambda = \frac{\lambda_a\lambda_b}{\mu_a + \lambda_a + \mu_b + \lambda_b}, \tag{46}$$

$$|J| = \frac{(\mu_a + \lambda_a)^2}{(\mu_a + \lambda_a + \mu_b + \lambda_b)^3}. \tag{47}$$

Via split move: $(u, v, \mu, \lambda) \rightarrow (\mu_a, \lambda_a, \mu_b, \lambda_b)$, again $u = \mu_a$ and $v = \lambda_a$, and u and v should be simulated from $u \sim N_T(2\mu, \sigma^2)$ and $v \sim N_T(2\lambda, \sigma^2)$ truncated at 0. The Jacobian for a split move is the reciprocal of the one for a combine move:

$$\mu_b = \frac{\mu_a^2\lambda + \mu_a\lambda_a\lambda - \lambda_a\mu\mu_a - \lambda_a^2\mu}{\lambda_a(-\mu_a - \lambda_a + \mu + \lambda)}, \tag{48}$$

$$\lambda_b = -\frac{(\mu_a + \lambda_a)^2\lambda}{\lambda_a(-\mu_a - \lambda_a + \mu + \lambda)}. \tag{49}$$

Via death move: $(\mu_a, \lambda_a, \mu_b) \rightarrow (u, v, \mu)$, we have:

$$\mu = \frac{(\mu_a + \lambda_a)\mu_b}{(\mu_b + \mu_a + \lambda_a)}, \tag{50}$$

$$|J| = \frac{(\mu_a + \lambda_a)^2}{(\mu_b + \mu_a + \lambda_a)^2}. \tag{51}$$

Via birth move: $(u, v, \mu) \rightarrow (\mu_a, \lambda_a, \mu_b)$ with u and v being simulated from $u \sim N_T(\mu, \sigma^2)$ and $v \sim N_T(\mu, \sigma^2)$ truncated at 0, the Jacobian is the reciprocal of the expression used for death move and

$$\mu_b = \frac{(u + v)\mu}{u + v - \mu}. \tag{52}$$

For more detailed explanations of the RJMCMC method, readers are referred to McGrory et al. (2009). This method can be used for updating the parameters of Coxian distribution for a single component as a part of sum of a number of variables with the use of posterior distribution stated in Equation (41). As the RJMCMC method jumps between models with different numbers of phases, model selection is automatically

performed within the estimation procedure.

For credible interval estimation, we propose to apply the same Gamma prior distributions for the parameters. The posterior distribution takes the number of phases as constant. Based on the posterior distribution in Equation (41) and using Metropolis-Hastings approach to generate parameter estimates, credible intervals for parameters using quantiles of the generated parameter values can be obtained. Needless to say, this model, if used only for credible interval estimation and not for RJMCMC, can be easily extended to handle general PH distributions.

5. Numerical Examples

5.1 A Simulation Study

5.1.1 Capability of Fitting Different Failure-time Distributions

To demonstrate the capability and flexibility of our proposed methods in reliability estimation, aggregate data from different probability distributions are generated, and Coxian distributions are used to fit the data and compared against the true distributions from which the data are generated. In particular, the ML estimation method is illustrated in this study.

The simulated data are generated from Gamma, IG and Weibull distributions. For each distribution, two cases are considered. The first case considers 6 aggregate data points with vector of number of failures $\mathbf{M} = [2 \quad 9 \quad 8 \quad 8 \quad 6 \quad 5]$, and the second case involves 12 aggregate data points with $\mathbf{M} = [2 \quad 2 \quad 9 \quad 9 \quad 8 \quad 8 \quad 8 \quad 8 \quad 6 \quad 6 \quad 5 \quad 5]$. When fitting the Coxian distributions, only the aggregate data are used. However, if individual failure-times are available, to estimate the model parameters using the ML method, Equations (7(-(11) should be applied. To visualize the estimation capability of the proposed method, the CDF's of the true distribution and the estimated Coxian distribution are shown together in each figure. Moreover, we have saved individual failure times so that the Kaplan-Meier estimate is also calculated and presented in the same figure for comparison.

The distributions are chosen in different ranges for fair comparison. Figure 3 shows the results for the aggregate data generated from $Gamma(2.5, 4)$. The result in the left figure is obtained based on 6 data points involving a total of 38 component failures, and the right figure is based on 12 data points for a total of 76 component failures. One can see that the Coxian distribution can mimic the true distribution quite closely, and as the number of data points increases, the deviation from the true distribution becomes negligible. This result illustrates the flexibility of Phase-type distribution in approximating other distributions. For the IG distribution as illustrated in Figure 4, the data generated from $IG(10, 8)$ is used. Our results show

that the Coxian distribution can also mimic the IG distribution closely. For the two-parameter-Weibull distribution, $Weibull(1, 1.5)$, although the estimated 3-phase Coxian distribution is a little off from the true distribution, the number of phases can be increased to increase the accuracy. As an illustration, a 6-phase Coxian distribution is used to fit the same Weibull data. Figure 6 shows clear improvement by increasing the number of phases. It is worth pointing out that for all the tested distributions, the same number of hidden failures and the same number of simulated data points are used in each case. The flexibility of PH distribution and its capability to handle aggregate data are obvious. The proposed method has potential to be applied for the analysis of aggregate or individual failure-time data when the underlying distribution cannot be conjectured. Moreover, increasing either the number of data points or the number of phases will improve the estimation accuracy of the proposed method. This is particularly favorable for aggregate data, since many probability distributions are intractable for aggregate data.

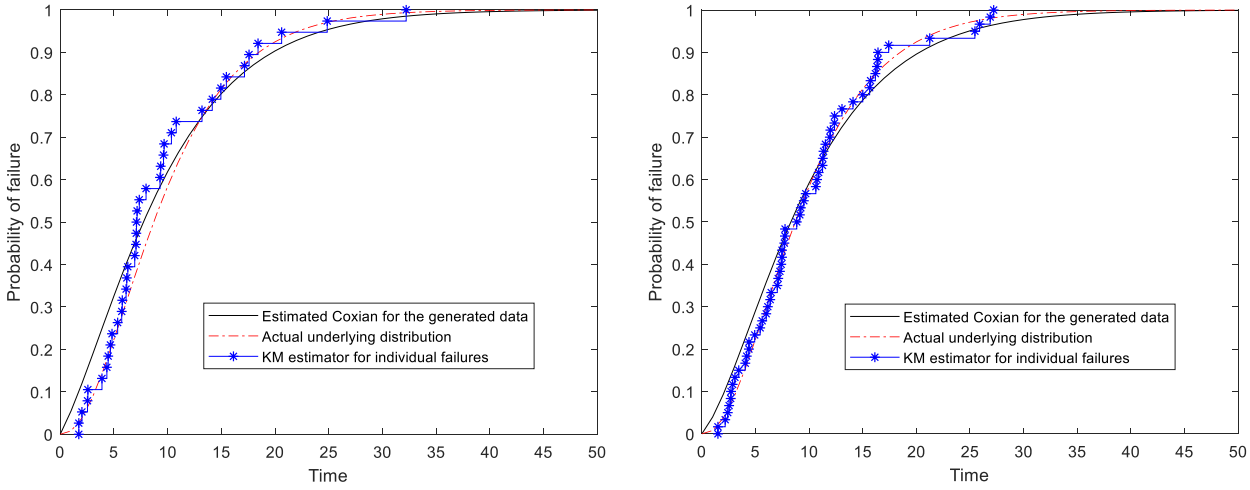

Figure 3: Estimated 3-phase Coxian distribution vs. the real underlying distribution, $Gamma(2.5, 4)$, and Kaplan-Meier estimate. The left figure is the result based on 6 data points, and the right figure is based on 12 aggregate data points.

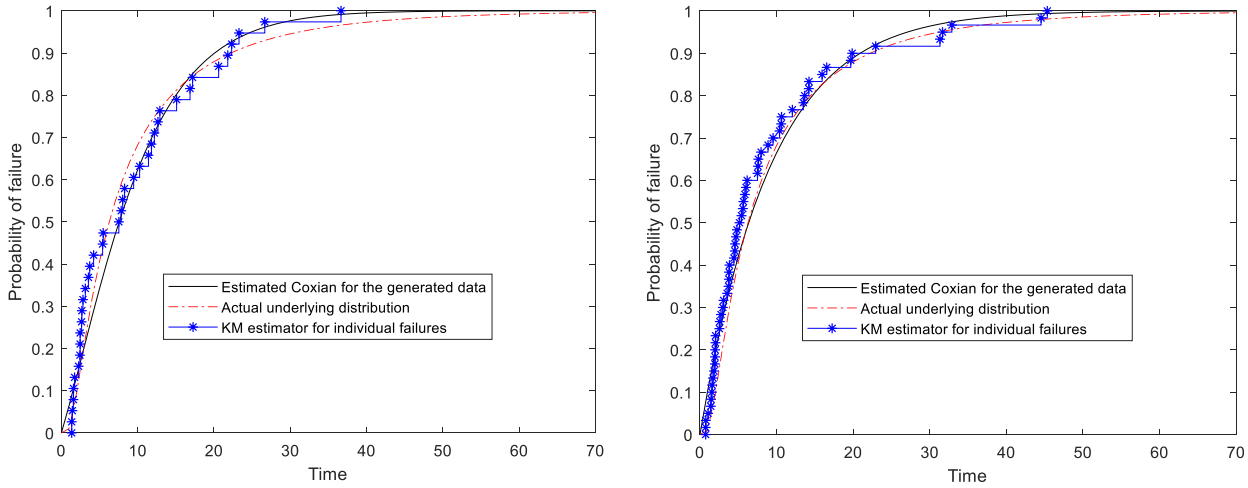

Figure 4: Estimated 3-phase Coxian distribution vs. the real underlying distribution, $IG(10, 8)$, and Kaplan-Meier estimate. The left figure is the result based on 6 data points and the right figure is based on 12 aggregate data points.

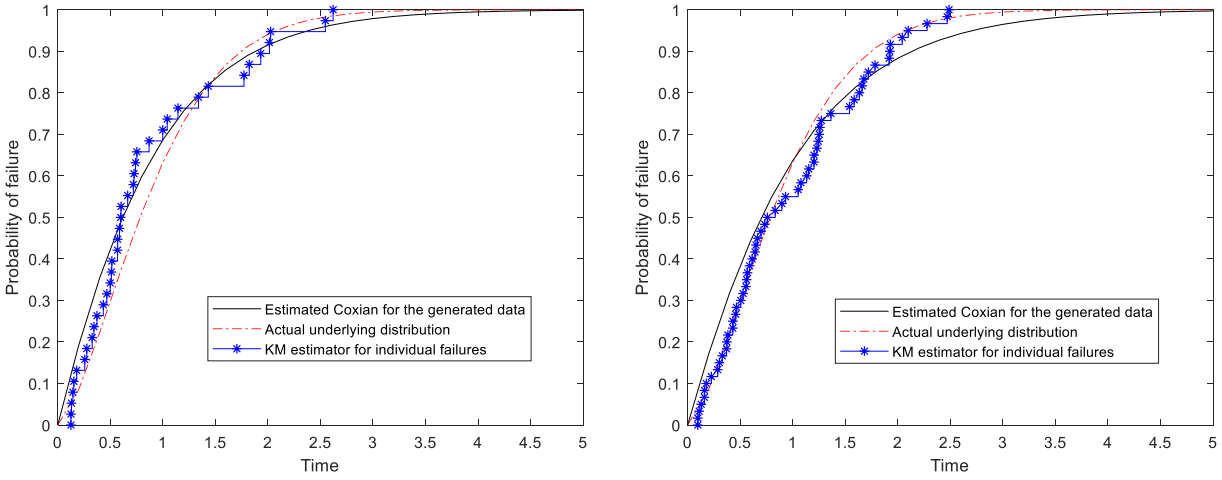

Figure 5: Estimated 3-phase Coxian distribution vs. the real underlying distribution, $Weibull(1, 1.5)$, and Kaplan-Meier estimate. The left figure is the result based on 6 data points and the right figure is based on 12 aggregate data points.

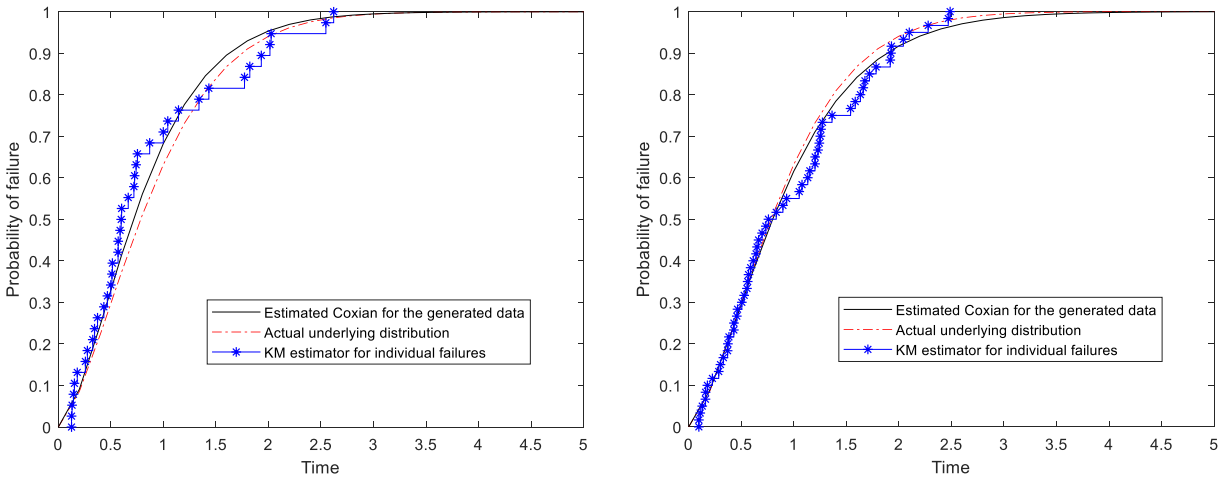

Figure 6: Estimated 6-phase Coxian distribution vs. the real underlying distribution, $Weibull(1, 1.5)$, and Kaplan-Meier estimate. The left figure is the result based on 6 data points and the right figure is based on 12 aggregate data points.

5.1.2 Study on the Coverage Probability of Normal Approximate Conference Interval

Nonparametric bootstrapping, while being widely used in many situations for interval estimation, should be applied carefully. In particular, the coverage probabilities can be significantly lower than the intended confidence level for small to moderate samples. The reason mainly lies behind the resampling of the bootstrap procedure (Schenker, 1985). In case of aggregate data, failure-times are aggregated into one data point, so practically, we are sampling groups of failures, where the groups do not change. As a result, the resampling problem deteriorates for aggregate data, making the coverage probability of the confidence interval even lower. In this section, the coverage probability of confidence interval obtained using the proposed Normal approximation method is studied against the nonparametric bootstrap method. Note that the coverage probability of credible interval obtained using the Bayesian alternative depends on the selection of prior distribution, thus is not studied in this paper.

For illustration, a 3-phase Coxian distribution is used to estimate the CDF of a true failure-time distribution, $Weibull(15, 0.95)$. The study is conducted for cases with 6 and 12 aggregate data points, respectively. For each case, the coverage probabilities of the two methods are estimated based on 5000 simulation runs. Table 1 shows the results, which clearly show that the bootstrap CIs for both cases give a much lower coverage probabilities than expected. On the other hand, the proposed Normal approximation method provides much better coverage for those failure-time percentiles.

Table 1: Coverage probabilities of 90% CIs using normal approximation and non-parametric bootstrap

Percentile	Normal approx. 6 data points	Normal approx. 12 data points	Bootstrap 6 data points	Bootstrap 12 data points
10	0.8246	0.8751	0.1584	0.6634
50	0.9980	0.9990	0.7426	0.6733
90	0.9965	0.9985	0.6040	0.6832

5.2 A Real-world Application

5.2.1 The Data

The Reliability Information Analysis Center (RIAC) is a U.S. DoD center who serves for collecting reliability data of fielded systems. Due to the possibilities and technical obstacles in practice, a large amount of the data are not individual component failure-time data (Coit and Jin (2000)). The reliability data shown in Table 2 is gathered by RIAC from aircraft indicator lights and has been previously studied by Coit and Jin (2000), and Chen and Ye (2017). In this data, 6 systems were observed, and the number of failures and the cumulative operating time up to the last failure for each system was recorded.

Table 2: Aircraft indicator lights failure data

System number	Cumulative operating time (hours)	Number of component failures
1	51000	2
2	194900	9
3	45300	8
4	112400	8
5	104000	6
6	44800	5

For each system k, the cumulative operating time t_k represents the time from the installation of the first component to the failure of the m_k-th component at a certain component position in system k. For this set of data, the reliability of an individual aircraft indicator light is desired.

5.2.2 Reliability Estimation and Model Selection

In this section, the proposed methods are applied on the data set and compared with the three distributions previously studied: Gamma, IG and Normal (Chen and Ye, 2017). The algorithms were run on a computer with Core(TM) i5-6300HQ CPU, 8.00 GB RAM and on Matlab 2017b.

First, the proposed ML estimation method with the new EM algorithm is implemented. Figure 7 illustrates the estimated 3-phase Coxian distribution in comparison to the estimated Gamma, IG and Normal distributions studied by Chen and Ye (2017). While the Normal distribution does not provide a very good

estimate because of high coefficient of variation. The CDF estimate from the Coxian distribution is close to those of Gamma and IG. The computational time of this method is 57.19 seconds with resulting likelihood value of -30.9821.

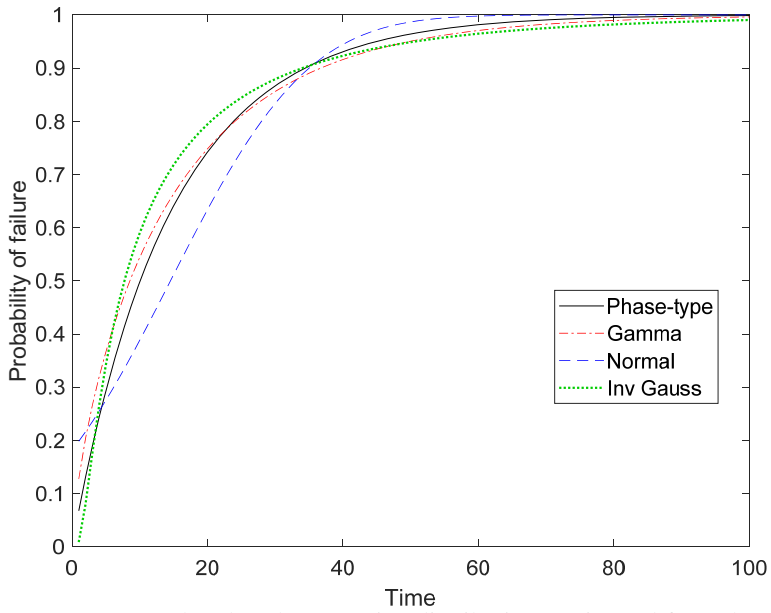

Figure 7: CDFs of Gamma, IG, Normal and 3-phase Coxian distributions estimated from the aggregate aircraft indicator light data

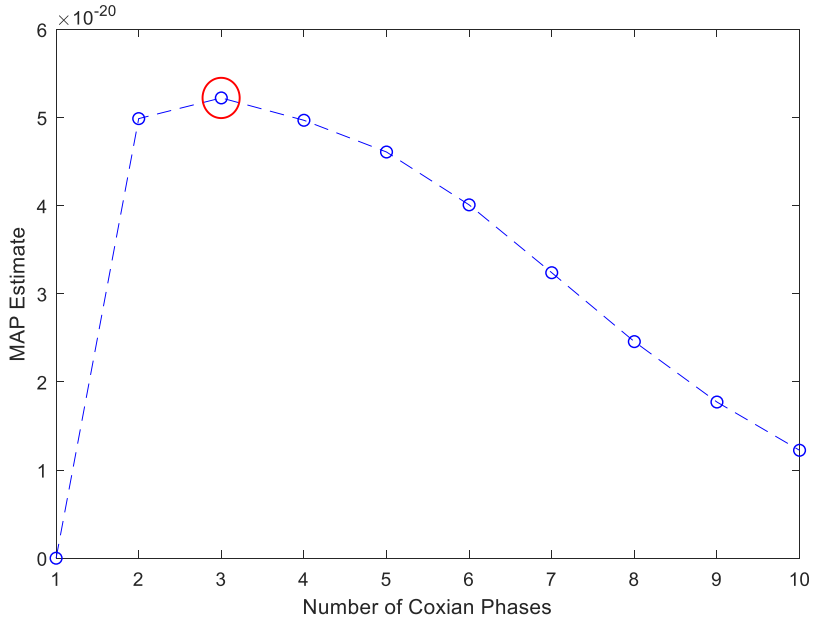

Figure 8: MAP model selection method performed for the data provided in Table 2 over the range of 1-phase through 10-phase Coxian with Laplacian prior distributions with parameters $(0, 1)$. The maximum MAP estimation suggests a 3-phase Coxian.

For the aircraft indicator light data, using $Laplace(0, 1)$ as the prior distribution, Figure 8 shows the MAP estimation result for Coxian distributions with 1 to 10 phases. It is clear that a 3-phase Coxian distribution is suggested. Therefore, the CDF estimate based on the 3-phase Coxian distribution presented in Figure 7 is an adequate estimate.

Regarding the Bayesian alternative, the time elapsed for 100 iterations of RJMCMC algorithm with a 20-iteration Metropolis-Hastings, varies between 45 to 50 seconds depending on the size of matrices that are randomly chosen in the algorithm for calculations. Since RJMCMC algorithm moves forward based on random movements to improve the estimation and the number of times each movement is performed during one implementation is different, it could result in different numbers of phases and transition rate matrices of different sizes. The following two matrices are the estimated transition rate matrices from two different implementations of the algorithm:

$$\mathbb{S}_1 = \begin{pmatrix} -.0674 & 0.0000 & 0 & 0 & 0 & 0 \\ 0 & -0.0233 & 0.0000 & 0 & 0 & 0 \\ 0 & 0 & -0.0072 & 0.0069 & 0 & 0 \\ 0 & 0 & 0 & -0.0001 & 0.0000 & 0 \\ 0 & 0 & 0 & 0 & -0.0516 & 0.0041 \\ 0 & 0 & 0 & 0 & 0 & -0.3830 \end{pmatrix} \text{ (6-phase Coxian),}$$

$$\mathbb{S}_2 = \begin{pmatrix} -0.0664 & 0.0000 & 0 \\ 0 & -0.1525 & 0.1099 \\ 0 & 0 & -0.4036 \end{pmatrix} \text{ (3-phase Coxian).}$$

Clearly, the two matrices are associated with two different Coxian distributions with different numbers of phases. Unlike the MAP method used in MLE, the disadvantage of this automatic model selection method is that it may not result in a unique model. However, as shown in Figure 9, the resulting CDF's obtained from the two implementations are quite close.

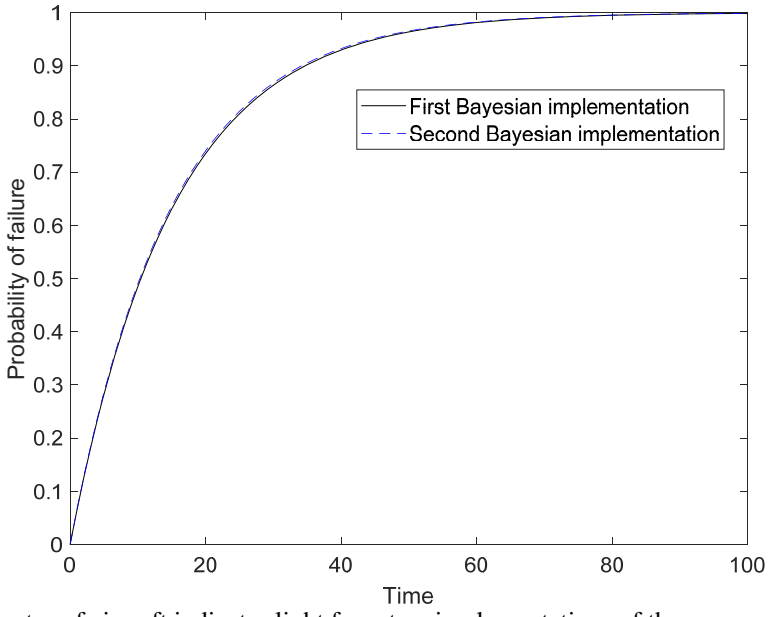

Figure 9: CDF estimates of aircraft indicator light from two implementations of the proposed Bayesian method

5.2.3 Interval Estimation

In this section, the ML confidence intervals (Normal approximation) and Bayesian credible intervals of model parameters and CDF of the 3-phase Coxian distribution are calculated. In particular, the ML confidence interval is found by deriving the Fisher information matrix first followed by calculating the estimated variance covariance matrix as:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\Theta}}} = \begin{pmatrix} 0.0046 & 0.0069 & -0.0052 & -0.0018 & -0.0026 \\ 0.0069 & 0.0113 & -0.0064 & -0.0035 & -0.0043 \\ -0.0052 & -0.0046 & 0.0157 & -0.0084 & -0.0053 \\ -0.0018 & 0.0035 & -0.0084 & 0.0237 & 0.0072 \\ -0.0026 & -0.0043 & -0.0053 & 0.0072 & 0.0206 \end{pmatrix}.$$

Afterwards, the Normal-approximation confidence intervals of model parameters are calculated as addressed in Section 3. Regarding the Bayesian credible intervals of the parameters, the results are obtained based on 1000 Metropolis-Hastings samples. The resulting 90% confidence intervals and credible intervals of model parameters are shown in Table 3. The results of the two methods are relatively close except the upper-bound for λ_1.

Figure 10 shows the 90% confidence interval of CDF estimated using Equation (39). A nonparametric 90% bootstrap confidence interval is also calculated and provided in Figure 11. One can see that the

Table 3: C.I.'s based on the MLE and Bayesian methods

Parameter	ML Estimate	90% MLE C.I.	90% Bayesian C.I.
μ_1	0.0702	(0.0274,0.1130)	(0.0514, 0.0902)
μ_2	0.0431	(0, 0.0300)	(0.0014, 0.0529)
μ_3	0.0823	(0, 0.3768)	(0, 0.3861)
λ_1	0.0121	(0, 0.5992)	(0, 0.0336)
λ_2	0.0392	(0, 0.4316)	(0, 0.3525)

nonparametric bootstrap confidence interval appears to be much narrower than the one from the Normal-approximation alternative. Finally, Figure 12 presents the credible interval of CDF from the Bayesian alternative, which depends on the selection of prior distribution and the sample size.

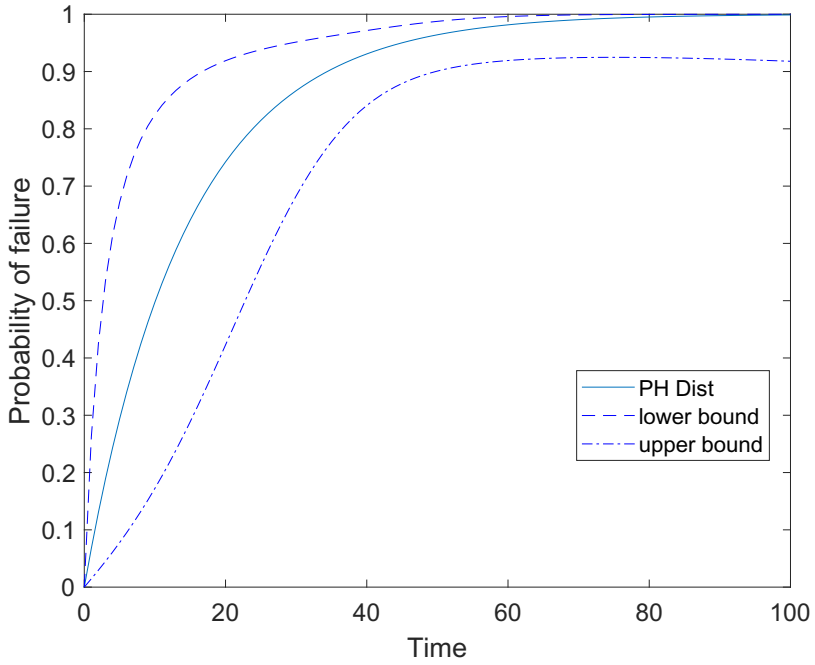

Figure 10: 90% Normal approximate confidence interval of CDF based on the 3-phase Coxian

6. Conclusions and Future Work

Reliability estimation using aggregate data has been studied with only a few probability distributions. This work presents more flexible methods based on PH distributions to deal with such data for the first time. An EM algorithm is developed in this work by exploring the submatrices to utilize aggregate data. An alternative Bayesian method is also introduced to incorporate prior knowledge for parameter estimation. For the MLE method, model selection is performed through an MAP method. For the Bayesian method, model selection is concealed within the estimation procedure. Interval estimations are also obtained for the

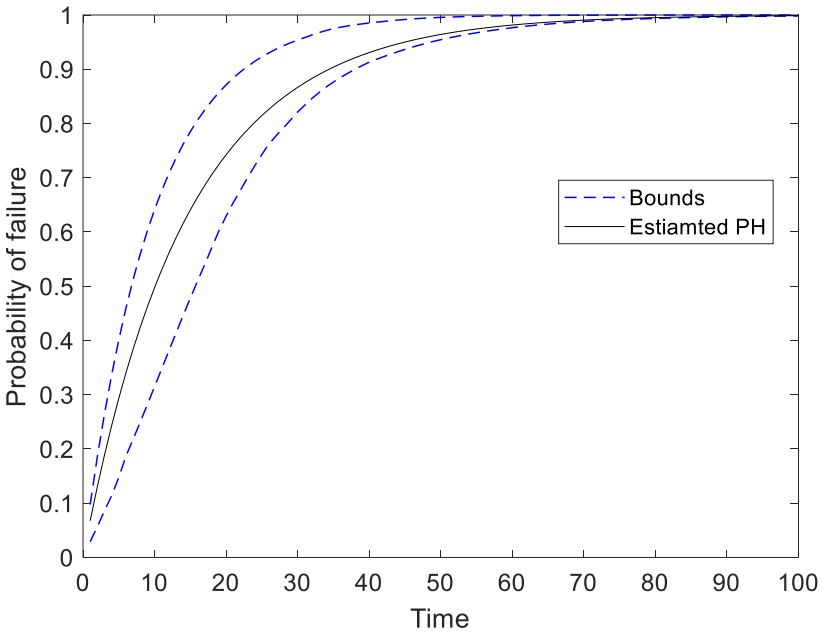

Figure 11: 90% bootstrap confidence interval of CDF based on the 3-phase Coxian

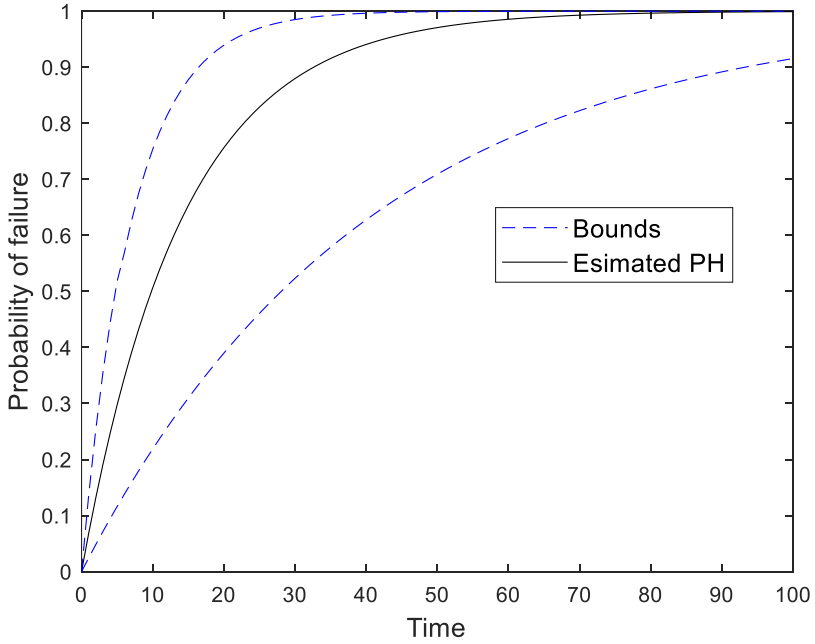

Figure 12: 90% Bayesian credible interval of CDF based on the 3-phase Coxian

two methods. The flexibility of PH distribution for analyzing aggregate data with an arbitrary underlying distribution is explored in a simulation study and the capability of PH distribution is clearly illustrated. In addition, the proposed methods are successfully applied to the real dataset from RIAC. Considering that only a few probability distributions have been utilized for analyzing aggregate data, this work provides

more flexible methods for analyzing aggregate failure-time data. Technically, the new EM algorithm, Fisher information and RJMCMC for PH distribution are used to analyze aggregate data for the first time.

For future work, interval estimation for PH distribution based on generalized pivotal quantity can be studied. Moreover, developing a nonparametric estimator based on aggregate data is a favorable while challenging research topic. Another interesting and common type of field data is time-censored aggregate data. The most common reason for collecting such data is to perform scheduled inspections. For time-censored aggregate data, each data point represents the number of failures in a certain period of time (e.g., during an inspection period). Unlike the aggregate data studied in this paper, each time-censored aggregate time is not recorded at one of failures. The analysis of time-censored aggregate data has recently been discussed by Chen et al. (2020). Bayesian methods were provided for the Gamma, Inverse Gaussian, Weibull and Lognormal distributions. It is worth pointing out that the analysis of time-censored aggregate data through PH distribution has not been discussed in the literature. The authors of this paper have considered this research gap, and both ML and Bayesian estimation methods will be provided in their future work.

Acknowledgments

The authors would like to thank the Editor and two anonymous referees for their valuable comments and suggestions, which significantly improved the quality and presentation of this paper.

Funding

Dr. Liao's research was partly supported by the U.S. National Science Foundation (Grant #CMMI 1635379), and Dr. Fan's research was partly supported by the U.S. National Science Foundation (Grant #CMMI 1634282).

Notes on contributors

Samira Karimi is a Ph.D. student in the Industrial Engineering Department at the University of Arkansas, Fayetteville, Arkansas. She received her B.Sc. (2014) and M.Sc. (2016) in Industrial Engineering from Sharif University of Technology, Tehran, Iran. Her research interests include reliability modeling and statistical analysis. She received 2017 PHM-Harbin Best Paper Award and the 2019 SRE Stan Ofsthun Best Paper Award. She is a student member of IISE and INFORMS.

Dr. Haitao Liao is a Professor and John and Mar Lib White Endowed Systems Integration Chair in the Department of Industrial Engineering at the University of Arkansas - Fayetteville. He received a Ph.D. degree in Industrial and Systems Engineering from Rutgers University in 2004. He also earned M.S. degrees in Industrial Engineering and Statistics from Rutgers University, and a B.S. degree in Electrical Engineering from Beijing Institute of Technology. His research interests include: (1) reliability models, (2) maintenance and service logistics, (3) prognostics, (4) probabilistic risk assessment, and (5) analytics of sensor data. His research has been sponsored by the National Science Foundation, Department of Energy, Nuclear Regulatory Commission, Oak Ridge National Laboratory, and industry. The findings of his group have been published in IISE Transactions, European Journal of Operational Research, Naval Research Logistics, IEEE Transactions on Reliability, IEEE Transactions on Cybernetics, The Engineering Economist, Reliability Engineering & System Safety, etc. He received a National Science Foundation CAREER Award in 2010, IISE William A. J. Golomski Award in 2011, 2014 and 2018, SRE Stan Ofsthun Best Paper Award in 2015 and 2019, and 2017 Alan O. Plait Award for Tutorial Excellence. He is a Fellow of IISE, a member of INFORMS, and a lifetime member of SRE.

Dr. Neng Fan is an associate professor at Department of Systems and Industrial Engineering, and also member of Graduate Interdisciplinary Program (GIDP) in Statistics, and GIDP in Applied Math at the University of Arizona (UA), Tucson, Arizona. He received his bachelor degree in computational mathematics from Wuhan University in China, and master degree in applied mathematics from Nankai University in China. He also received his master and PhD degrees from Department of Industrial and Systems Engineering at University of Florida. Before joining UA, he worked in Los Alamos National Laboratory, Los Alamos, New Mexico and Sandia National Laboratories, Albuquerque, New Mexico. His research focuses on th development of various optimization methodologies, and their applications in energy systems, renewable energy integration, healthcare, sustainable agriculture, and data analytics. His work has been supported by the U.S. National Science Foundation, Department of Energy, Department of Agriculture, and local industry.

References

[1] Asmussen S., Jensen J. L., and Rojas-Nandayapa L. (2016). On the Laplace transform of the lognormal distribution. *Methodology and Computing in Applied Probability,* 18(2), 441-58.

[2] Asmussen, S., Nerman, O., and Olsson, M. (1996). Fitting Phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23(4), 419-441.

[3] Ausín, M. C., Wiper, M. P., and Lillo, R. E. (2008). Bayesian prediction of the transient behaviour and busy period in short-and long-tailed GI/G/1 queueing systems. *Computational Statistics & Data Analysis,* 52(3), 1615-1635.

[4] Beaulieu N. C. and Rajwani F. (2004). Highly accurate simple closed-form approximations to lognormal sum distributions and densities. *IEEE Communications Letters*, 8(12), 709-11.

[5] Beaulieu, N. C. and Xie, Q. (2004). An optimal lognormal approximation to lognormal sum distributions. *IEEE Transactions on Vehicular Technology*, 53(2), 479-489.

[6] Bhaumik, D. K., Kapur, K., and Gibbons, R. D. (2009). Testing parameters of a gamma distribution for small samples. *Technometrics,* 51(3), 326-334.

[7] Bladt, M., Esparza, L. J. R., and Nielsen, B. F. (2011). Fisher information and statistical inference for phase-type distributions. *Journal of Applied Probability,* 48(A), 277-293.

[8] Bladt, M., Gonzalez, A., and Lauritzen, S. L. (2003). The estimation of Phase-type related functionals using Markov chain Monte Carlo methods. *Scandinavian Actuarial Journal,* 2003(4), 280-300.

[9] Bučar, T., Nagode M., and Fajdiga M. (2004). Reliability approximation using finite Weibull mixture distributions. *Reliability Engineering & System Safety*, 84(3), 241-51.

[10] Buchholz, P., Kriege, J., and Felko, I. (2014). *Input Modeling with Phase-Type Distributions and Markov Models: Theory and Applications*. SpringerBriefs in Mathematics.

[11] Chen, P. and Ye, Z. S. (2017a). Random effects models for aggregate lifetime data. *IEEE Transactions on Reliability,* 66(1), 76-83.

[12] Chen, P. and Ye, Z. S. (2017b). Estimation of field reliability based on aggregate lifetime data. *Technometrics,* 59(1), 115-125.

[13] Chen, P., Ye, Z. S., and Zhai, Q. (2020). Parametric analysis of time-censored aggregate lifetime data. *IISE Transactions,* 52(5), 516-527.

[14] Cobb, B. R., Rumi, R., and Salmerŏn, A. (2012). Approximating the distribution of a sum of lognormal random variables. *Statistics and Computing*, 16(3), 293-308.

[15] Coit, D. W. and Dey, K. A. (1999). Analysis of grouped data from field-failure reporting systems. *Reliability Engineering & System Safety,* 65(2), 95-101.

[16] Coit, D. W. and Jin, T. (2000). Gamma distribution parameter estimation for field reliability data with missing failure times. *IIE Transactions,* 32(12), 1161-1166.

[17] Cui, L. and Wu, B. (2019). Extended Phase-type models for multistate competing risk systems. *Reliability Engineering & System Safety*, 181, 1-16.

[18] Delia, M. C. and Rafael, P. O. (2008). A maintenance model with failures and inspection following Markovian arrival processes and two repair modes. *European Journal of Operational Research ,* 186(2), 694-707.

[19] Denson, W., Crowell, W., Jaworski, P., and Mahar, D. (2014). *Electronic Parts Reliability Data 2014*. Reliability Information Analysis Center, Rome, NY, USA.

[20] Elmahdy, E. E. and Aboutahoun, A. W. (2013). A new approach for parameter estimation of finite Weibull mixture distributions for reliability modeling. *Applied Mathematical Modelling*, 37(4), 1800-1810.

[21] Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika,* 82(4), 711-732.

[22] Horvath, A. and Telek, M. (2007). Matching more than three moments with acyclic phase type distributions. *Stochastic Models*, 23(2), 167194.

[23] Jin, T. and Gonigunta, L. S. (2010). Exponential approximation to Weibull renewal with decreasing failure rate. *Journal of Statistical Computation and Simulation*, 80(3), 273-285.

[24] Karimi, S., Liao, H., and Pohl, E. (2019). A robust approach for estimating field reliability using aggregate failure time data. *Proceedings of Annual Reliability and Maintainability Symposium (RAMS),* (1-6), Orlando, FL: IEEE.

[25] Kharoufeh, J. P., Solo, C. J., and Ulukus, M. Y. (2010). Semi-Markov models for degradation-based reliability. *IIE Transactions,* 42(8), 599-612.

[26] Lam, C. L. J. and Le-Ngoc, T. (2007). Log-shifted gamma approximation to lognormal sum distributions. *IEEE Transactions on Vehicular Technology,* 56(4), 2121-2129.

[27] Li, J., Chen, J., and Zhang, X. (2019). Time-dependent reliability analysis of deteriorating structures based on phase-type distributions. *IEEE Transactions on Reliability*, in print.

[28] Liao, H. and Guo, H. (2013). A generic method for modeling accelerated life testing data. *Proceedings of Annual Reliability and Maintainability Symposium (RAMS),* (1-6), Orlando, FL: IEEE.

[29] Liao, H. and Karimi, S. (2017). Comparison study on general methods for modeling lifetime data with covariates. *Prognostics and System Health Management Conference (PHM-Harbin)* (1-5). Harbin: IEEE.

[30] Mahar, D., Fields, W., Reade, J., Zarubin, P., and McCombie, S. (2011). *Nonelectronic Parts Reliability Data*. Reliability Information Analysis Center.

[31] Marie, R. (1980). Calculating equilibrium probabilities for λ(n)/ck/1/n queues. *In Proceedings of the Performance 1980*, 117125.

[32] McGrory, C. A., Pettitt, A. N., and Faddy, M. J. (2009). A fully Bayesian approach to inference for Coxian phase-type distributions with covariate dependent mean. *Computational Statistics & Data Analysis,* 53(12), 4311-4321.

[33] Meeker, W. Q. and Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician,* 49(1), 48-53.

[34] Mehta, N. B., Wu, J., Molisch, A. F., and Zhang, J. (2007). Approximating a sum of random variables with a lognormal. *IEEE Transactions on Wireless Communications*, 6(7), 2690-2699.

[35] Okamura, H., Dohi, T., and Trivedi, K. S. (2011). A refined EM algorithm for PH distributions. *Performance Evaluation,* 68(10), 938-954.

[36] Okamura, H., Watanabe, R., and Dohi, T. (2014). Variational Bayes for phase-type distribution. *Communications in Statistics-Simulation and Computation,* 43(8), 2031-2044.

[37] OREDA (2009). *OREDA offshore Reliability Data Handbook*. Det Norske Veritas (DNV), Høvik, Norway.

[38] Osogami, T. and Harchol-Balter, M. (2003). A closed-form solution for mapping general distributions to minimal PH distributions. *In International Conference on Modelling Techniques and Tools for Computer Performance Evaluation* 200-217, Springer, Berlin, Heidelberg.

[39] Osogami, T. and Harchol-Balter, M. (2006). Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Performance Evaluation*, 63(6), 524-552.

[40] Rizopoulos, D., Verbeke, G., and Lesaffre, E. (2009). Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(3), 637-654.

[41] Segovia, M. C. and Labeau, P. E. (2013). Reliability of a multi-state system subject to shocks using phase-type distributions. *Applied Mathematical Modelling,* 37(7), 4883-4904.

[42] Starobinski, D. and Sidi, M. (2000). Modeling and analysis of power-tail distributions via classical teletraffic methods. *Queueing Systems,* 36(1-3), 243-267.

[43] Telek, M. and Heindl, A. (2003). Matching moments for acyclic discrete and continuous phase-type distributions of second order,. *International Journal of Simulation*, 3(3-4), 47-57.

[44] Watanabe, R., Okamura, H., and Dohi, T. (2012). An efficient MCMC algorithm for continuous PH distributions. *Simulation Conference (WSC), Proceedings of the 2012 Winter* (1-12). Berlin: IEEE.

[45] Xu, D., Xiao, X., and Yu, H. (2020). Reliability evaluation of smart meters under degradation-shock loads based on phase-type distributions. *IEEE Access*, 8, 39734-39746.

[46] Yamaguchi, Y., Okamura, H., and Dohi, T. (2010). A variational Bayesian approach for estimating parameters of a mixture of Erlang distribution. *Communications in Statistics - Theory and Methods*, 39(13), 2333-2350.