

## Research Article

# An Expectation Maximization Algorithm to Model Failure Times by Continuous-Time Markov Chains

Qihong Duan,<sup>1</sup> Zhiping Chen,<sup>1</sup> and Dengfu Zhao<sup>2</sup>

<sup>1</sup> Department of Mathematics, Faculty of Science, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

<sup>2</sup> School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

Correspondence should be addressed to Dengfu Zhao, dfzhao@263.net

Received 5 June 2010; Revised 28 July 2010; Accepted 29 July 2010

Academic Editor: Ming Li

Copyright © 2010 Qihong Duan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In many applications, the failure rate function may present a bathtub shape curve. In this paper, an expectation maximization algorithm is proposed to construct a suitable continuous-time Markov chain which models the failure time data by the first time reaching the absorbing state. Assume that a system is described by methods of supplementary variables, the device of stage, and so on. Given a data set, the maximum likelihood estimators of the initial distribution and the infinitesimal transition rates of the Markov chain can be obtained by our novel algorithm. Suppose that there are  $m$  transient states in the system and that there are  $n$  failure time data. The devised algorithm only needs to compute the exponential of  $m \times m$  upper triangular matrices for  $O(nm^2)$  times in each iteration. Finally, the algorithm is applied to two real data sets, which indicates the practicality and efficiency of our algorithm.

## 1. Introduction

Among many quantitative analysis methods of system reliability, the state space representation has often been employed by reliability engineers, for example, [1–5]. The state space representation has been recognized by industrial IEC61511 standards [6]. This method assumes that the structure of the system is modelled with a continuous-time Markov chain (CTMC). In the state space representation, nodes represent states of the system and arcs represent the transitions between nodes. This method is well adapted to study the reliability of various systems and allows an exact analysis of their probability of failure.

In the state space representation method, the time interval between two consecutive transitions is a random variable of an exponential distribution. However, many studies have presented counterexamples. For some mechanical and electronic components, the failure rate function of time to failure (TTF) has a bathtub curve: a monotonically decreasing

function initially, eventually becoming a constant, and finally changing to a monotonically increasing function after sufficient time elapses [7]. Because existing approaches do not provide a perfect solution to deal with this class of systems, several researchers have explored models with nonexponential distributions (see [8–10] and references therein). Among many approaches treating TTF with nonexponential distributions, the extended Markov-model [11] is recommendable. In the extended Markov-model, an operation state is divided into substates with different levels of failure rates, which result in a nonconstant failure rate of the operation state. That is, in these models, the bug of failure rate functions is fixed within the framework of the state space representation (see [12, 13] and references therein for some applications).

Methods of supplementary variables [14] and the device of stages [15] are two classical approaches of extended Markov-models. These two approaches allow one to model a wide range of experimental probability density functions through a set of additional stages in series or in parallel. For supplementary variables with given parameters, a formula for evaluating the occurrence of failures is derived in [16]. In fact, with suitable parameters, additional stages can approximate the behavior of many types of distributions in practice. Reference [17] is a recent application of these approaches.

Assume that the structure of a system is established. That is, the nodes and arcs of the system are given. For an experimental data set on the TTFs of the system, the bottleneck of the above two approaches is the estimation of parameters for the nodes and arcs. Since the system is modelled with a CTMC, the TTFs are phase-type distributed random variables. The problem is similar to phase-type fitting methods [18]. However, traditional phase-type fitting methods give the maximum likelihood estimators of the minimal representation [18], which may not correspond to the system with the expected structure. Moreover, from the standpoint of computational complexity, traditional methods are not suitable for the system with many paths.

Inspired by the theory of continuous-time Markov chains observed in continuous-time channels (see [19, 20]), we propose a new expectation maximization (EM) algorithm for data of TTFs in this paper. In Section 2 we specify the framework and hidden variables for the system. In Section 3 we derive several conditional probabilities and conditional expectations, which will be used in the EM algorithm. In Section 4 we present the new EM algorithm. In Section 5, we provide numerical results. Section 6 concludes the paper.

## 2. System Definition and Hidden Variables

Assume that a system is modelled by a CTMC  $X := \{X_t : t \geq 0\}$  with transient states  $1, 2, \dots, m$ , and one absorbing state  $m + 1$ . The infinitesimal transition rates  $Q = (q_{ij})$  of  $X$  satisfy

$$q_{m+1,m+1} = 0, \quad q_{ii} < 0, \quad \text{for } i \neq m. \quad (2.1)$$

The quantities  $q_{ij}, i \neq m + 1$ , are unknown. However, the characteristic matrix  $C = (c_{ij})$ ,

$$c_{ij} = \begin{cases} 1, & q_{ij} \neq 0, i \neq j, \\ 0, & \text{otherwise,} \end{cases} \quad (2.2)$$

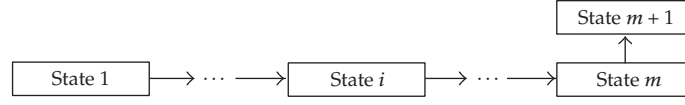


Figure 1: The state transition diagram of a device of stages.

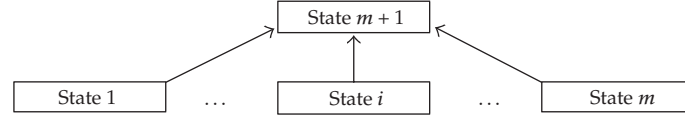


Figure 2: The state transition diagram of a parallel system.

which represents the structure of the system, is the known. Assume that the characteristic  $C$  is an upper triangular matrix, that is,  $c_{ij} = 0$  for  $i < j$ . We suppose that the initial probability function  $\pi = (\pi_1, \pi_2, \dots, \pi_{m+1})$  satisfies  $\pi_{m+1} = 0$  and  $\pi_1, \pi_2, \dots, \pi_m$  are unknown.

*Example 2.1* (The device of stages). The system has a state transition diagram shown in Figure 1. The characteristic matrix  $C = (c_{ij})$  of this system is

$$c_{ij} = 1, \quad \text{for } j = i + 1, \quad c_{ij} = 0, \quad \text{for } j \neq i + 1. \quad (2.3)$$

Hence  $C$  is an upper triangular matrix.

*Example 2.2* (The parallel system). The system has a state transition diagram shown in Figure 2. The characteristic matrix  $C = (c_{ij})$  of this system is

$$c_{ij} = 1, \quad \text{for } j = m + 1, \quad c_{ij} = 0, \quad \text{for } j \neq m + 1. \quad (2.4)$$

Hence  $C$  is also an upper triangular matrix.

Like the above two examples, the structures of many practical systems can be represented by upper triangular characteristic matrixes.

We define a path, the sequence of nodes visited from an initial node to the absorbing node  $m + 1$ . The  $k$ th path,  $r_k = (n_1, n_2, \dots, n_s)$ , is characterized by a set of nodes  $n_1, n_2, \dots, n_s$ , representing the nodes visited before absorption, where  $c_{n_i n_{i+1}} = 1$  for  $i = 1, 2, \dots, s - 1$  and  $c_{n_s, m+1} = 1$ . Since the characteristic matrix  $C$  and hence the infinitesimal transition rates  $Q$ , are upper triangular matrixes, any node will never appear again in a path. So the length of a path is no more than  $m + 1$ , and hence the number of paths is finite.

According to the above discussion, similarly to part b) of Theorem 5.4 in [21], we can provide the following noncanonical construction of a Markov chain with the transition rates  $Q$  and the initial distribution  $\pi$ , respectively. The hidden variables, which comprise the noncanonical construction, will serve the EM algorithm in Section 4.

**Lemma 2.3.** For  $k = 1, 2, \dots, n$ , let  $T_1^k, T_2^k, \dots, T_m^k$  be a sequence of exponentially distributed random variables with parameters  $-q_{11}, -q_{22}, \dots, -q_{mm}$ , respectively. For  $k = 1, 2, \dots, n$ , let  $J_0^k, J_1^k, J_2^k, \dots, J_m^k$  be a sequence of random variables with discrete distributions described by probability functions

$$q_i(j) = P(J_i^k = j) = \begin{cases} 0, & i = j, i \neq 0, \\ \frac{q_{ij}}{-q_{ii}}, & i \neq j, i \neq 0, \\ \pi_j, & i = 0. \end{cases} \quad (2.5)$$

Assume also that the sequences  $\{J_0^k, J_j^k, T_j^k, j = 1, 2, \dots, m, k = 1, 2, \dots, n\}$  comprise an independent family. Define

$$\begin{aligned} Y_0^k &= J_0^k, & Y_s^k &= J_{Y_{s-1}^k}^k, & \text{for } s \geq 1, Y_{s-1}^k &\neq m+1, \\ X_t^k &= Y_0^k, & 0 \leq t &< T_{Y_0^k}^k, \\ X_t^k &= Y_{s+1}^k, & \text{for } T_{Y_0^k}^k + T_{Y_1^k}^k + \dots + T_{Y_s^k}^k &\leq t < T_{Y_0^k}^k + T_{Y_1^k}^k + \dots + T_{Y_{s+1}^k}^k. \end{aligned} \quad (2.6)$$

Then the constructed process  $X^k := \{X_t^k : t \geq 0\}$  is Markovian with the initial distribution  $\pi$  and the infinitesimal transition rates  $Q$ . Moreover,  $X^k, k = 1, 2, \dots, n$  are independent.

For a sample of  $J := (J_0^k, J_1^k, J_2^k, \dots, J_m^k)$ , Lemma 2.3 indicates that there is a unique path corresponding to the discrete-time embedded Markov chain  $(Y_0^k, Y_1^k, \dots)$ . We denote this path as  $\mathcal{R}(J)$ . Write  $j \in \mathcal{R}(J)$  if the node  $j$  appears in the path  $\mathcal{R}(J)$ . Given a path  $\mathcal{R}(J)$ , define the first node  $\mathcal{F}_J(0) = J_0^k$  for this path, and the following node  $\mathcal{F}_J(j) = J_j^k$  for any node  $j \in \mathcal{R}(J)$ .

A sample  $(\tau^1, \tau^2, \dots, \tau^n)$  of the TTFs for a system is a family of independent identically distributed random variables. Precisely, the stochastic processes  $X^k, k = 1, 2, \dots, n$ , are independent, which have the same infinitesimal transition rates  $Q$  and the same initial probability function  $\pi$ . Then define

$$\tau^k = \inf\{t \geq 0 : X_t^k = m+1\}. \quad (2.7)$$

Using the above notations, we have

$$\tau^k = \sum_{i \in \mathcal{R}(J)} T_i^k, \quad J = (J_0^k, J_1^k, J_2^k, \dots, J_m^k). \quad (2.8)$$

Based on the observations  $\tau^1, \tau^2, \dots, \tau^n$  and the known characteristic matrix  $C$ , our problem is to give estimators  $\hat{\pi}$  and  $\hat{Q}$  for the unknown parameters.

### 3. Conditional Probabilities and Conditional Expectations

To establish an EM algorithm for our problem, some conditional probabilities and conditional expectations will be calculated in this section. From now on, we will omit the superscript  $k$

of the stochastic processes when the clarity is being held. Firstly, we recall the cumulative distribution function of the TTF  $\tau$ ,

$$F_\tau(t) = P(\tau \leq t) = p_{m+1}(t). \quad (3.1)$$

Here  $p(t) = (p_1(t), p_2(t), \dots, p_{m+1}(t))$  is the probability function of  $X_t$ , which satisfies the Kolmogorov's forward equation  $p'(t) = p(t)Q$ . From the solution to the equation in exponential form, we have

$$p_{m+1}(t) = \pi \exp(tQ)1_D^T, \quad (3.2)$$

where the vector  $1_D^T$  of dimensions  $m + 1$  is the transpose of  $1_D = (0, 0, \dots, 0, 1)$ . So the probability density function of the TTF  $\tau$  is

$$f_\tau(t) = \pi \exp(tQ)Q1_D^T. \quad (3.3)$$

The above methodology can be used to calculate other probabilities and expectations in this section.

**Lemma 3.1.** *For the above noncanonical construction of  $X$  in Lemma 2.3, we have for  $1 \leq i \leq m$ ,*

$$P(i \in \mathcal{R}(J_0, J_1, \dots, J_m) \mid \tau = t) = \frac{1}{f_\tau(t)} \pi^{(i)} \exp(tQ_i)Q_i 1_P^T, \quad \text{for } t \geq 0. \quad (3.4)$$

Here  $Q_i = (q'_{rs})$  is a matrix of dimensions  $(m + 2) \times (m + 2)$  determined as follows:

$$q'_{rs} = \begin{cases} q_{rs}, & r < i, s \leq i, \\ \sum_{t=i+1}^{m+1} q_{rt}, & r < i, s = m + 2, \\ q_{rs}, & i \leq r \leq m + 1, s \leq m + 1, \\ 0, & \text{otherwise.} \end{cases} \quad (3.5)$$

$1_P = (0, 0, \dots, 0, 1, 0)$  and  $\pi^{(i)} = (\pi_1^{(i)}, \pi_2^{(i)}, \dots, \pi_{m+2}^{(i)})$  are vectors of dimensions  $m + 2$ ,

$$\pi_s^{(i)} = \begin{cases} \pi_s, & s \leq i, \\ 0, & i < s < m + 1, \\ \sum_{t=i+1}^{m+1} \pi_t, & s = m + 2. \end{cases} \quad (3.6)$$

*Proof.* Define the events

$$A_t = \{i \in \mathcal{R}(J_0, J_1, \dots, J_m), \tau \leq t\}, \quad t \geq 0. \quad (3.7)$$

Considering all possible realizations of  $J = (J_0, J_1, \dots, J_m)$ , it follows from (2.8) that

$$P(A_t) = \sum_{\mathcal{R}(J) \ni i} P \left( J_0 = j_0, J_s = \mathcal{F}(J, s), s \in \mathcal{R}(J), \sum_{s \in \mathcal{R}(J)} T_s \leq t \right). \quad (3.8)$$

Since the characteristic matrix  $C$  is an upper triangular matrix, the numbers of nodes along a path are monotonic increasing. Then if  $J = (j_0, j_1, \dots, j_m)$  satisfies  $\mathcal{R}(J) \ni i$ , we have  $j_0 \leq i$  and

$$\mathcal{F}(J, s) \leq i, \quad \text{for } s \in \mathcal{R}(J), s < i. \quad (3.9)$$

Considering another stochastic process  $X'$  with the infinitesimal transition rates  $Q_i$  and the initial probability function  $\pi^{(i)}$ , we can deduce a noncanonical construction consisting of  $J'_0, J'_1, \dots, J'_m, T'_1, T'_2, \dots, T'_m$ . It follows from (3.9) that

$$P(A_t) = \sum_{\mathcal{R}(J') \ni i} P \left( J'_0 = j_0, J'_s = \mathcal{F}(J', s), s \in \mathcal{R}(J'), \sum_{s \in \mathcal{R}(J')} T_s \leq t \right). \quad (3.10)$$

For a path  $J'$  of the stochastic process  $X'$ , the property  $\mathcal{R}(J') \ni i$  means that the end of the path  $J'$  is the absorbing state  $m + 1$ . Similarly to (3.2), we have

$$P(A_t) = P(X'_t = m) = \pi^{(i)} \exp(tQ_i) \mathbf{1}_P^T. \quad (3.11)$$

The result follows from the above equation and

$$P(A_t | \tau = t) = \lim_{\epsilon \downarrow 0} \frac{P(A_{t+\epsilon}) - P(A_t)}{P(t < \tau \leq t + \epsilon)}. \quad (3.12)$$

□

Similarly to the discussion in Lemma 3.1, we have the following Lemma.

**Lemma 3.2.** *For the noncanonical construction of  $X$  in Lemma 2.3,  $1 \leq i \leq m$ , given  $i \in \mathcal{R}(J_0, J_1, \dots, J_m)$ , the conditional density function of  $\tau$  is*

$$f_\tau(t | i \in \mathcal{R}(J)) = \frac{1}{p'_i} \pi^{(i)} \exp(tQ_i) Q_i \mathbf{1}_P^T, \quad \text{for } t \geq 0, \quad (3.13)$$

where  $J = (J_0, J_1, \dots, J_m)$  and  $p'_i = P(i \in \mathcal{R}(J))$ .

$p'_i$  in Lemma 3.2 can be expressed through the 1-step transition probability matrix of the embedded Markov chain of the continuous-time Markov chain with the infinitesimal transition rates  $Q_i$ . However, since  $p'_i$  will be eliminated in the further propositions, it is not necessary to present the expression.

**Theorem 3.3.** For the noncanonical construction of  $X$  in Lemma 2.3,  $1 \leq i \leq m$ ,  $1 \leq j \leq m+1$ , and  $t \geq 0$ , we have

$$P(j = \mathcal{F}(J, i), i \in \mathcal{R}(J_0, J_1, \dots, J_m) \mid \tau = t) = \begin{cases} 0, & c_{i,j} = 0, \\ \frac{1}{f_\tau(t)} \pi^{(i)} \exp(tQ_{ij}) Q_{ij} 1_P^T, & c_{i,j} \neq 0. \end{cases} \quad (3.14)$$

Here  $1_P$  and  $\pi^{(i)}$  are the same as those in Lemma 3.1.  $Q_{ij} = (\tilde{q}_{rs})$  is a matrix of dimensions  $(m+2) \times (m+2)$  determined as follows:

$$\tilde{q}_{rs} = \begin{cases} q_{rs}, & r \neq i, \\ q_{is}, & r = i, s = i \text{ or } j, \\ \sum_{t=i+1}^{m+1} q_{it} - q_{ij}, & r = i, s = m+2, \\ 0, & \text{otherwise.} \end{cases} \quad (3.15)$$

*Proof.* If  $c_{i,j} = 0$ , it is obvious that  $P(i \in \mathcal{R}(J), j = \mathcal{F}(J, i), \tau \leq t) = 0$ . Otherwise, by constructing an auxiliary stochastic process with the infinitesimal transition rates  $Q_{ij}$  and the initial probability function  $\pi^{(i)}$ , we can obtain, similarly to the proof of Lemma 3.1, that

$$P(i \in \mathcal{R}(J), j = \mathcal{F}(J, i), \tau \leq t) = \pi^{(i)} \exp(tQ_{ij}) 1_P^T, \quad (3.16)$$

where  $J = (J_0, J_1, \dots, J_m)$ . The required result can then be derived in the similar way as that in Lemma 3.1.  $\square$

**Theorem 3.4.** For the noncanonical construction of  $X$  in Lemma 2.3 and  $i \leq m$ ,  $t \geq 0$ , we have

$$E[T_i \mid i \in \mathcal{R}(J), \tau = t] = -\frac{q_{ii} \exp(q_{ii}t)}{\pi^{(i)} \exp(tQ_i) Q_i 1_P^T} \times \left( t \pi_{m+1}^{(i,T)} + \pi^{(i,T)} \int_0^t (t-u) \exp(u(M_i - q_{ii}I)) du M_i 1_P^T \right), \quad (3.17)$$

where  $J = (J_0, J_1, \dots, J_m)$ .  $I$  is an identity matrix and  $M_i = (M_{rs})$  is a matrix of dimensions  $(m+2) \times (m+2)$  determined as follows:

$$M_{rs} = \begin{cases} q_{rs}, & r < i, s < i, \\ -\frac{q_{ri}q_{i,s}}{q_{ii}}, & r < i, i < s \leq m+1, \\ \sum_{t=i+1}^{m+1} q_{rt}, & r < i, s = m+2, \\ q_{rs}, & i < r \leq m+1, i < s \leq m+1, \\ 0, & \text{otherwise.} \end{cases} \quad (3.18)$$

$\pi^{(i,T)} = (\pi_1^{(i,T)}, \pi_2^{(i,T)}, \dots, \pi_{m+2}^{(i,T)})$  is vectors of dimensions  $m + 2$  determined as follows:

$$\pi_s^{(i,T)} = \begin{cases} \pi_s, & s < i, \\ 0, & s = i, \\ -\frac{\pi_i q_{is}}{q_{ii}}, & i < s \leq m + 1, \\ \sum_{t=i+1}^{m+1} \pi_t, & s = m + 2. \end{cases} \quad (3.19)$$

*Proof.* It follows from (2.8) that given  $i \in \mathcal{R}(J)$ ,

$$\tau = T_i + \sum_{j \in \mathcal{R}(J), j \neq i} T_j, \quad (3.20)$$

where  $T_i$  and  $T^i = \sum_{j \in \mathcal{R}(J), j \neq i} T_j$ , are independent random variables. Given  $i \in \mathcal{R}(J)$ , the conditional density function of  $T_i$  is

$$f_i(t) = -q_{ii} \exp(q_{ii}t), \quad \text{for } t \geq 0. \quad (3.21)$$

Given  $i \in \mathcal{R}(J)$ , suppose that the conditional density function of  $T^i$  is  $g_i(t)$ , where  $g_i(t) = 0$  for  $t < 0$ . Since the Jacobian of the transformation  $(x, y) \rightarrow (x, x + y)$  is 1, the joint density of  $T_i, \tau$  is given by  $f(t_1, t_2) = f_i(t_1)g_i(t_2 - t_1)$ . Then for  $t \geq 0$ , we have

$$\begin{aligned} E[T_i \mid i \in \mathcal{R}(J), \tau = t] &= \frac{1}{f_\tau(t \mid i \in \mathcal{R}(J))} \int_0^t u f_i(u) g_i(t - u) du \\ &= \frac{1}{f_\tau(t \mid i \in \mathcal{R}(J))} \int_0^t (t - u) f_i(t - u) g_i(u) du. \end{aligned} \quad (3.22)$$

Here the conditional density function  $f_\tau(t \mid i \in \mathcal{R}(J))$ , given  $i \in \mathcal{R}(J)$ , is determined through (3.13) in Lemma 3.2.

As to the cumulative distribution function of  $T^i$ , given  $i \in \mathcal{R}(J)$ , we consider another stochastic process  $X'$  with the infinitesimal transition rates  $M_i$  and the initial probability function  $\pi^{(i,T)}$ . It follows from Lemma 2.3 that  $X'$  has a noncanonical construction consisting of  $J'_0, J'_1, \dots, J'_m, T'_1, T'_2, \dots, T'_m$ . Similarly to the discussion in Lemma 3.1, we have

$$\begin{aligned} &P(T^i \leq t, i \in \mathcal{R}(J)) \\ &= P\left(\sum_{j \in \mathcal{R}(J'), j \neq i} T_j \leq t, \text{the end of the path } J' \text{ is the absorbing state } m + 1\right), \end{aligned} \quad (3.23)$$



where  $J' = (J'_0, J'_1, \dots, J'_m)$ . Hence

$$P(T^i \leq t, i \in \mathcal{R}(J)) = \boldsymbol{\pi}^{(i,T)} \exp(tM_i) \mathbf{1}_P^T. \quad (3.24)$$

Then it follows from (3.13) and (3.24) that

$$P(T^i \leq t \mid i \in \mathcal{R}(J)) = \frac{1}{p'_i} \boldsymbol{\pi}^{(i,T)} \exp(tM_i) \mathbf{1}_P^T. \quad (3.25)$$

Since the probability in (3.25) may jump at  $t = 0$ , the conditional density function corresponding to (3.24) is

$$g_i(t) = \frac{1}{p'_i} \boldsymbol{\pi}_{m+1}^{(i,T)} \delta(t) + \frac{1}{p'_i} \boldsymbol{\pi}^{(i,T)} \exp(tM_i) M_i \mathbf{1}_P^T, \quad \text{for } t \geq 0, \quad (3.26)$$

where  $\delta(t)$  is the Dirac function.  $p'_i = P(i \in \mathcal{R}(J))$  is the same as that in Lemma 3.2. Then the integral in (3.22) becomes

$$\begin{aligned} \int_0^t (t-u) f_i(t-u) g_i(u) du &= -\frac{q_{ii}}{p'_i} t \exp(q_{ii}t) \boldsymbol{\pi}_{m+1}^{(i,T)} \\ &\quad - \frac{q_{ii}}{p'_i} \exp(q_{ii}t) \boldsymbol{\pi}^{(i,T)} \int_0^t (t-u) \exp(u(M_i - q_{ii}I)) du M_i \mathbf{1}_P^T. \end{aligned} \quad (3.27)$$

The desired result follows from the above equation and Lemma 3.2.  $\square$

*Remark 3.5.* The integral in (3.17) can be efficiently computed by using an approach developed by Van Loan [22]. To apply that approach here, we first compute the exponential of the following  $3(m+2) \times 3(m+2)$  block triangular matrix:

$$\exp \left[ t \begin{pmatrix} O & I & O \\ O & O & I \\ O & O & (M_i - q_{ii}I) \end{pmatrix} \right] = \begin{pmatrix} I & tI & H(t) \\ O & I & G(t) \\ O & O & F(t) \end{pmatrix}. \quad (3.28)$$

Then the upper right block  $H(t)$  of this matrix gives the integral  $\int_0^t (t-u) \exp(u(M_i - q_{ii}I)) du$ . Moreover, the matrix exponentials can be efficiently performed by using the diagonal Padé approximation with repeated squaring, as recommended in [22, 23].

*Remark 3.6.* Sometimes, the denominator of (3.17) is 0. For example, under some characteristic matrices, a node  $i$  may never be reached. Since the estimation of  $T_i$  is nonsense under this case, we can assign any value to  $E[T_i \mid i \in \mathcal{R}(J), \tau = t]$ . We suggest  $-1/q_{ii}$ .

The following corollary can be deduced from the above theorems and the tower property of conditional expectations.

**Corollary 3.7.** For the noncanonical construction of  $X$  in Lemma 2.3 and  $1 \leq i \leq m, 1 \leq j \leq m+1, t \geq 0$ , we have

$$\begin{aligned} E[T_i | \tau = t] &= p_i E[T_i | i \in \mathcal{R}(J), \tau = t] - \frac{1-p_i}{q_{ii}}, \\ P(J_i = j | \tau = t) &= P(j = \mathcal{F}(J, i), i \in \mathcal{R}(J) | \tau = t) - \frac{(1-p_i)q_{ij}}{q_{ii}}, \\ P(J_0 = j | \tau = t) &= \frac{1}{f_\tau(t)} \pi_i e_i \exp(tQ) Q 1_D^T. \end{aligned} \quad (3.29)$$

Here  $e_i$  is the unit basis vectors  $(0, 0, \dots, 1, 0, \dots, 0, 0)$  with "1" in the  $i$ th position.  $J = (J_0, J_1, \dots, J_m)$  and  $p_i = P(i \in \mathcal{R}(J_0, J_1, \dots, J_m) | \tau = t)$ .

*Proof.* The first equation follows from Theorem 3.4 and the tower property of conditional expectations. The second equation follows from

$$P(J_i = j | \tau = t) = P(j = \mathcal{F}(J, i) | i \in \mathcal{R}(J) \tau = t) p_i + P(J_i = j) (1 - p_i). \quad (3.30)$$

As to the third equation, we have

$$P(J_0 = j, \tau \leq t) = P(J_0 = j) P(\tau \leq t | J_0 = j). \quad (3.31)$$

From the solution to the Kolmogorov's forward equation in the exponential form, we obtain

$$P(\tau \leq t | J_0 = j) = \pi_i e_i \exp(tQ) Q 1_D^T. \quad (3.32)$$

Similarly to (3.12), we can see that the third equation follows from the above two equations and (3.2).  $\square$

#### 4. EM Algorithm

The EM algorithm is a two-step iterative process for computing the maximum likelihood estimate. This process is terminated when some imposed measure of convergence for the sequence of estimators is satisfied. A filter-based form of the EM algorithm for a Markov chain was presented in [24] and was developed in [25]. Here we review firstly the filter-based EM framework.

Let  $\theta$  index a given family of probability density functions  $f_\theta$  for some hidden random variable  $X_i$ , where  $\theta$  is an element of a parameter space  $\Theta$ . Then based on a hidden complete data  $X_1, \dots, X_n$ , a likelihood  $\mathcal{L}$  can be written as

$$\mathcal{L}(\theta) = \prod_{k=1}^n f_\theta(X_k). \quad (4.1)$$

Let a  $\sigma$ -field  $\mathcal{F} = \sigma\{X_1, \dots, X_n\}$ . For a random variable  $Y \in \mathcal{F}$ ,  $E^{\theta^*} Y$  denotes the expectation of  $Y$  with assumption that  $\theta^*$  is the parameter of  $X_i$ s. Suppose another  $\sigma$ -field  $\mathcal{Y} \subset \mathcal{F}$ .

Two iterative steps in the EM algorithm are as follows.

(1) Expectation step: fix  $\theta^* = \theta_t$ , then compute  $Q(\cdot, \theta^*)$ , where

$$Q(\theta, \theta^*) = E^{\theta^*} [\log(\mathcal{L}(\theta)) | \mathcal{Y}]. \quad (4.2)$$

(2) Maximization step: Set  $\theta_{t+1} = \hat{\theta}$  as a best solution to the following optimization problem:

$$\max_{\hat{\theta} \in \Theta} Q(\hat{\theta}, \theta^*). \quad (4.3)$$

Under this framework, we propose an EM algorithm which attempts to find a maximum likelihood estimate (MLE) of the parameter  $\theta = (\pi, Q)$ . Here  $\mathcal{Y}$  is a  $\sigma$ -field generated from a sequence  $(\tau^1, \tau^2, \dots, \tau^n)$  of TTFs. As for the hidden complete data, we choose  $T_1^k, T_2^k, \dots, T_m^k$  and  $J_0^k, J_1^k, J_2^k, \dots, J_m^k$ , where  $k = 1, 2, \dots, n$ . As showed in Lemma 2.3, no information is gained or lost by considering stochastic processes  $X^k$ ,  $k = 1, 2, \dots, n$ , as hidden processes instead of  $T$  and  $J$ . However, our choice of the hidden random variables results in a simple E-step, and it leads to an explicit M-step for estimating the parameters.

In this case, the complete data likelihood  $\mathcal{L}$  can be written as

$$\mathcal{L} = \prod_{k=1}^n \left( \prod_{i=0}^m q_i(J_i^k) \prod_{i=1}^m (-q_{ii}) \exp(q_{ii} T_i^k) \right). \quad (4.4)$$

The factors  $q_i(J_i^k)$  in (4.4) represent the probability functions of the hidden random variables  $J_i^k$ . The factors  $-q_{ii} \exp(q_{ii} T_i^k)$  represent the density functions of  $T_i^k$ . Taking the logarithm of (4.4) yields

$$\log \mathcal{L} = \sum_{k=1}^n \left( \sum_{i=0}^m \sum_{j=1}^m \log(q_i(j)) I(J_i^k = j) + \sum_{i=1}^m (\log(-q_{ii}) + q_{ii} T_i^k) \right), \quad (4.5)$$

where  $I(J_i^k = j) = 1$  if  $J_i^k = j$  and otherwise  $I(J_i^k = j) = 0$ .

A new parameter estimate, say  $\hat{\theta} = (\hat{\pi}, \hat{Q})$ , is obtained as the parameter which maximizes the expected value of (4.5) given  $(\tau^1, \tau^2, \dots, \tau^n)$  and a current parameter estimate  $\theta^* = (\pi, Q)$ . The constrained maximization gives us the following new estimate:

$$\begin{aligned} \hat{q}_i(j) &= \frac{1}{n} \sum_{k=1}^n P^{\theta^*} (J_i = j | \tau = \tau^k), \\ \hat{q}_{ii} &= - \frac{n}{\sum_{k=1}^n E^{\theta^*} (T_i | \tau = \tau^k)}, \end{aligned} \quad (4.6)$$

**Table 1:** Failure time data from [26].

0.1	0.2	1	1	1	1	1	2	3	6
7	11	12	18	18	18	18	18	21	32
36	40	45	46	47	50	55	60	63	63
67	67	67	67	72	75	79	82	82	83
84	84	84	85	85	85	85	85	86	86

where the conditional probabilities and the conditional expectations are given in Corollary 3.7. Then it follows from (2.5) that

$$\hat{\pi}_j = \hat{q}_0(j), \quad \hat{q}_{ij} = -\hat{q}_{ii}\hat{q}_i(j), \quad \text{for } i \neq j, \quad i \neq 0. \quad (4.7)$$

These expressions constitute the M-step of our EM algorithm.

## 5. Two Applications

In this section, we present two examples to illustrate the results derived in the previous sections.

The first data set is a widely used sample of 50 components taken from [26], which possess a bath-shaped failure rate property. These data consist of times to failure of 50 devices put on life test at time 0. The data is shown in Table 1.

Using our EM algorithm with  $m = 3$ , which means that there are transient states 1, 2, 3, and one absorbing state 4, we obtain

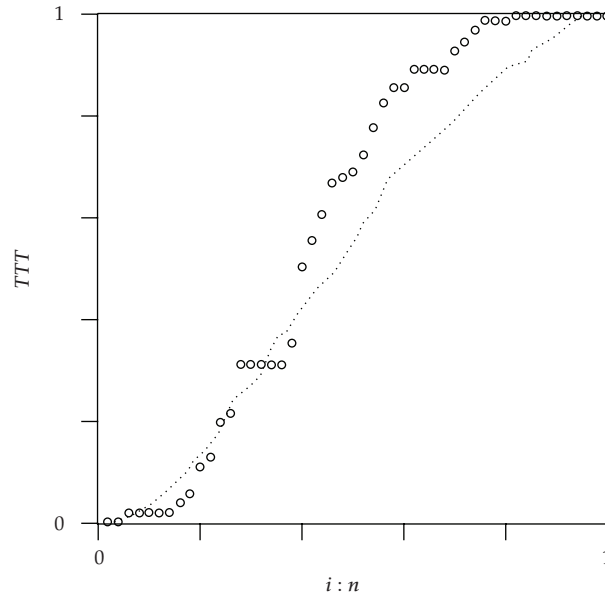
$$\hat{\pi} = (0.000007, 0.124512, 0.875480, 0.0),$$

$$\hat{Q} = \begin{pmatrix} -1.007004 & 1.007004 & 0.0 & 0.0 \\ 0.0 & -1.064019 & 0.011295 & 1.052724 \\ 0.0 & 0.0 & -0.019270 & 0.019270 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix}. \quad (5.1)$$

In many applications of the failure rate function, the total time on test (TTT) plot [26] was used to show the efficiency of the selected model. The TTT plot is obtained by plotting

$$G\left(\frac{r}{n}\right) = \frac{(\sum_{i=1}^r T_{i:n} + (n-r)T_{r:n})}{\sum_{i=1}^n T_{i:n}} \quad (5.2)$$

against  $r/n$ . Here  $r = 1, \dots, n$  and  $T_{i:n}$ ,  $i = 1, \dots, n$ , are the order statistics of the sample. To check the efficiency of the result, Monte Carlo simulations have also been carried out. We generated 500 independent TTFs, with the estimated parameters  $\hat{\pi}, \hat{Q}$ . Due to the sensitivity of TTT plot to  $T_{n:n}$ , the simulated TTFs are truncated by the maximum of the data set. The TTT plots for the first data set and the truncated simulated TTFs are shown in Figure 3.



**Figure 3:** The TTT plots for the first data set and the simulated TTFs. The circles represent the real data and the solid points represent the simulated results with the estimated parameters.

Another real set of lifetime data set, taken from [27], is the failure data collected during unit testing. During a unit testing phase, the number of failures in a certain time interval (of equal length) is recorded. The total number of units tested is equal to 311. The data is given in Table 2.

Using our EM algorithm with  $m = 3$ , we obtain

$$\hat{\pi} = (0.998642, 0.001358, 0.0, 0.0) ,$$

$$\hat{Q} = \begin{pmatrix} -0.172957 & 0.165593 & 0.007363 & 0.0 \\ 0.0 & -1.164560 & 1.159626 & 0.004934 \\ 0.0 & 0.0 & -4.767049 & 4.767049 \\ 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix}. \tag{5.3}$$

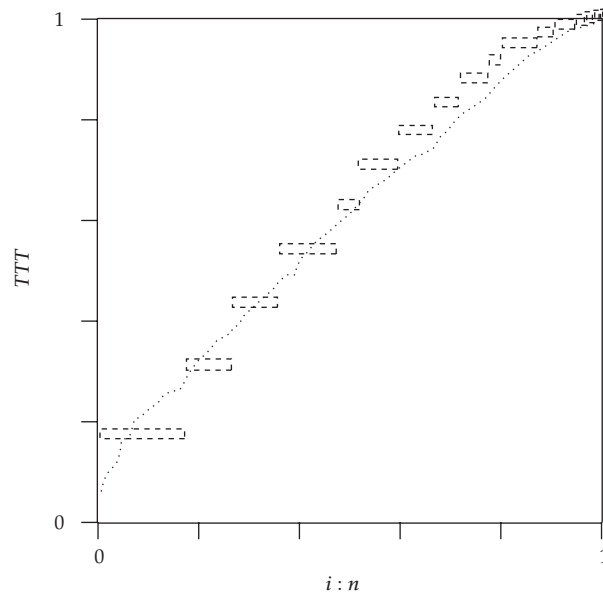
The TTT plots for this data set and the Monte Carlo simulated TTFs are given in Figure 4.

### 6. Conclusions

One great advantage with the extended Markov-models is that these models can tackle operation states of nonexponential distributions under the framework of the state space representation. Based on the theories of continuous-time Markov chains and EM algorithms, this paper deals with the problem of parameters estimation for extended Markov-models. Given a data set with  $n$  failure time data, for a system with  $m$  transient states, the proposed EM algorithm only needs to compute the exponential of  $m \times m$  upper triangular matrices for  $O(nm^2)$  times in each iteration. The effectiveness of the proposed EM algorithm is well demonstrated through two real world examples.

**Table 2:** Failure time data from [27].

Time interval	Number of failures	Time interval	Number of failures
1	53	2	29
3	29	4	36
5	13	6	25
7	22	8	16
9	18	10	8
11	22	12	11
13	13	14	5
15	5	16	4
17	1	18	1

**Figure 4:** The TTT plots for the second data set and the simulated TTFs. The dashed boxes represent the real data and the solid points represent the simulated results with the estimated parameters.

Two real applications show the practicality and efficiency of our new EM algorithm. Nevertheless, one theoretical limitation about EM algorithms is that its iterations sequence converges to a stationary point of the likelihood. Only in the case that the likelihood function is convex, the global convergence from any initial value can be ensured. However, our experience with many examples is against this theoretical drawback. In few practical cases, the stationary points are saddle points. As a simple perturbation can lead the algorithm to diverge from the saddle point, this phenomenon is thus not a problem in practice. However, to avoid that the limit point is only a local maximum of the likelihood, the initial value must be carefully selected in some cases. We suggest to use moments of related random variables to construct the initial estimates.

Of course, it will be interesting to investigate conditions ensuring the convergence to the global maximum of the likelihood. This is left for future research.

## Acknowledgments

Dengfu Zhao and Qihong Duan acknowledge the funding of the National Natural Science Foundation of China (Grant number 50977073) and the Fundamental Research Funds for the Central Universities. Zhiping Chen and Qihong Duan acknowledge the funding of the National Natural Science Foundation of China (Grant number 70971109).

## References

- [1] C. K. Pil, M. Rausand, and J. Vatn, "Reliability assessment of reliquefaction systems on LNG carriers," *Reliability Engineering and System Safety*, vol. 93, pp. 1345–1353, 2008.
- [2] S. Y. Chen, C. Y. Yao, G. Xiao, Y. S. Ying, and W. L. Wang, "Fault detection and prediction of clocks and timers based on computer audition and probabilistic neural networks," in *Proceedings of the 8th International Workshop on Artificial Neural Networks (IWANN '05)*, vol. 3512 of *Lecture Notes on Computer Science*, pp. 952–959, 2005.
- [3] S. Y. Chen, Y. F. Li, and J. W. Zhang, "Vision processing for realtime 3-D data acquisition based on coded structured light," *IEEE Transactions on Image Processing*, vol. 17, no. 2, pp. 167–176, 2008.
- [4] X. Zhao and L. Cui, "On the accelerated scan finite Markov chain imbedding approach," *IEEE Transactions on Reliability*, vol. 58, no. 2, pp. 383–388, 2009.
- [5] S. Y. Chen, Y. F. Li, Q. Guan, and G. Xiao, "Real-time three-dimensional surface measurement by color encoded light projection," *Applied Physics Letters*, vol. 89, no. 11, article no. 111108, 2006.
- [6] "IEC61511. Functional safety—safety instrumented systems for the process industry sector," Geneva, Switzerland, IEC, 2004.
- [7] M. Xie, Y. Tang, and T. N. Goh, "A modified Weibull extension with bathtubshaped failure rate function," *Reliability Engineering and System Safety*, vol. 76, pp. 279–285, 2002.
- [8] A. Pievatolo, E. Tironi, and I. Valadé, "Semi-Markov processes for power system reliability assessment with application to uninterruptible power supply," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1326–1333, 2004.
- [9] H. R. Guo, H. Liao, W. Zhao, and A. Mettas, "A new stochastic model for systems under general repairs," *IEEE Transactions on Reliability*, vol. 56, no. 1, pp. 40–49, 2007.
- [10] M. Li, "Fractal time series—a tutorial review," *Mathematical Problems in Engineering*, vol. 2010, Article ID 157264, 26 pages, 2010.
- [11] J. Endrenyi, G. J. Anders, and A. M. Leite Da Suva, "Probabilistic evaluation of the effect of maintenance on reliability—an application," *IEEE Transactions on Power Systems*, vol. 13, no. 2, pp. 576–583, 1998.
- [12] M. Li and W. Zhao, "Representation of a stochastic traffic bound," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 9, pp. 1368–1372, 2010.
- [13] M. Li and S. C. Lim, "Modeling network traffic using generalized Cauchy process," *Physica A*, vol. 387, no. 11, pp. 2584–2594, 2008.
- [14] D. R. Cox, "The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 51, pp. 433–441, 1955.
- [15] C. Singh, R. Billinton, and S. Y. Lee, "The method of stages for non-Markov models," *IEEE Transactions on Reliability*, vol. 26, no. 2, pp. 135–137, 1977.
- [16] Y. Lam, "Calculating the rate of occurrence of failures for continuous-time markov chains with application to a two-component parallel system," *Journal of the Operational Research Society*, vol. 46, pp. 528–536, 1995.
- [17] E. A. Oliveira, A. C.M. Alvim, and P. Melo, "Unavailability analysis of safety systems under aging by supplementary variables with imperfect repair," *Annals of Nuclear Energy*, vol. 32, no. 2, pp. 241–252, 2005.
- [18] A. Bobbio, A. Horváth, M. Scarpa, and M. Telek, "Acyclic discrete phase type distributions: properties and a parameter estimation algorithm," *Performance Evaluation*, vol. 54, no. 1, pp. 1–32, 2003.
- [19] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models*, vol. 29 of *Applications of Mathematics*, Springer, New York, NY, USA, 1995.

- [20] R. J. Elliott, Z. P. Chen, and Q. H. Duan, "Insurance claims modulated by a hidden Brownian marked point process," *Insurance: Mathematics & Economics*, vol. 45, no. 2, pp. 163–172, 2009.
- [21] R. N. Bhattacharya and E. C. Waymire, *Stochastic Processes with Applications*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1990.
- [22] C. F. Van Loan, "Computing integrals involving the matrix exponential," *IEEE Transactions on Automatic Control*, vol. 23, no. 3, pp. 395–404, 1978.
- [23] F. Carbonell, J. C. Jiménez, and L. M. Pedroso, "Computing multiple integrals involving matrix exponentials," *Journal of Computational and Applied Mathematics*, vol. 213, no. 1, pp. 300–305, 2008.
- [24] R. J. Elliott, "New finite-dimensional filters and smoothers for noisily observed Markov chains," *IEEE Transactions on Information Theory*, vol. 39, no. 1, pp. 265–271, 1993.
- [25] R. J. Elliott and W. P. Malcolm, "Discrete-time expectation maximization algorithms for Markov-modulated Poisson processes," *IEEE Transactions on Automatic Control*, vol. 53, no. 1, pp. 247–256, 2008.
- [26] M. V. Aarset, "How to identify bathtub hazard rate," *IEEE Transactions on Reliability*, vol. 36, no. 1, pp. 106–108, 1987.
- [27] M. Xie and C. D. Lai, "Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function," *Reliability Engineering and System Safety*, vol. 52, no. 1, pp. 87–93, 1996.