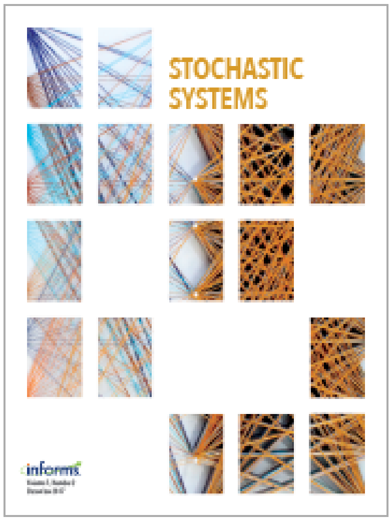


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Distributional Properties of the Mixture of Continuous-Time Absorbing Markov Chains Moving at Different Speeds

B. A. Surya

To cite this article:
B. A. Surya (2018) Distributional Properties of the Mixture of Continuous-Time Absorbing Markov Chains Moving at Different Speeds. Stochastic Systems 8(1):29-44. https://doi.org/10.1287/stsy.2017.0007

informs
http://pubsonline.informs.org/journal/stsy/

STOCHASTIC SYSTEMS
Vol. 8, No. 1, March 2018, pp. 29–44
ISSN 1946-5238 (print), ISSN 1946-5238 (online)

Distributional Properties of the Mixture of Continuous-Time Absorbing Markov Chains Moving at Different Speeds

B. A. Surya[a]

[a] School of Mathematics and Statistics, Victoria University of Wellington, Wellington 6140, New Zealand

Contact: budhi.surya@vuw.ac.nz, http://orcid.org/0000-0002-9041-8608 (BAS)

Received: April 4, 2017
Revised: September 4, 2017
Accepted: November 28, 2017
Published Online in Articles in Advance: February 5, 2018

MSC2010 Subject Classification: 60J20, 60J27, 60J28, 62N99

https://doi.org/10.1287/stsy.2017.0007

Abstract. This paper analyses the mixture of two continuous-time Markov chains with absorption on the same state space moving at different speeds, where the mixture occurs at a random time. Variety of associated distributional properties of the Markov mixture process are discussed, for example the transition matrix, the distribution of its lifetime and the corresponding forward intensity. Identities are given explicitly in terms of the Bayesian updates of switching probability and the intensity matrices of the underlying Markov chains despite the fact that the mixture process is not Markovian. They form nonstationary functions of time and duration and have two appealing features: the ability to capture heterogeneity and path dependence when conditioning on the available information (either full or partial) about the past history of the process until current time. In particular, the unconditional lifetime distribution forms a generalized mixture of phase-type distributions. The distribution has dense and closure properties under finite convex mixtures and finite convolutions. When the underlying Markov chains move at the same speed, in which case the mixture process reduces to a simple Markov chain, the heterogeneity and path dependence are removed, and the lifetime has the usual phase-type distribution, Neuts [Neuts MF (1975) Probability distributions of phase-type. *Liber Amicorum Prof. Emiritus H. Florin* (University of Louvain, Belgium), 173–206]. Some numerical examples are given to illustrate the main results.

Funding: This research is financially supported by Victoria University PBRF [Research Grants 212885 and 214168] for which the author is grateful.

Keywords: Markov chains • Markov mixture processes • generalized phase-type distributions • forward intensity • residual occupation time

1. Introduction

It has been well documented that Markov chain has been one of the most important probabilistic models in the analysis of complex stochastic systems evolution. The lifetime distribution of a finite-state absorbing Markov chain, known as the phase-type distribution, see Neuts (1975, 1981), forms a dense class of distributions on $\mathbb{R}_+$ which can approximate any distribution of positive random variables arbitrarily well. It has found a number of interesting applications across various fields. The advantage of working under phase-type distribution is that it allows some analytically tractable results in applications: e.g., in actuarial science (Rolski et al. 1998, Albrecher and Asmussen 2010, Lin and Liu 2007, Lee and Lin 2010), option pricing (Asmussen et al. 2004, Rolski et al. 1998), queueing theory (Badila et al. 2014, Chakravarthy and Neuts 2014, Buchholz et al. 2014, Breuer and Baum 2005, Asmussen 2003), reliability theory (Assaf and Levikson 1982, Okamura and Dohi 2015), and in survival analysis (Aalen 1995, Aalen and Gjessing 2001). When jumps distribution of a compound Poisson process is modeled by phase-type distribution, it results in a dense class of Lévy processes, Asmussen et al. (2004).

In credit risk, Jarrow et al. (1997) used Markov chain to model the dynamics of credit rating transition for defaultable bonds. Under the Markov model, bonds of the same credit rating, represented by the state space of the Markov chain, move at the same speed to other credit ratings. However, in their empirical study, Frydman and Schuermann (2008) found that firms of the same credit rating can move at different speeds to other credit ratings, a feature that is lacking in the Markov model. This empirical finding suggests that each pool of credit ratings consists of two sub-classes of bonds; one moving with higher speed than the other. In addition, it was shown in Frydman and Schuermann (2008) that the incorporation of past information of credit rating helps improve the out-of-sample prediction of the Nelson-Aalen estimate of cumulative intensity of the bonds.

The empirical study performed in Frydman and Schuermann (2008) was based on the Markov mixture model proposed in Frydman (2005).

The Markov mixture model (Frydman 2005) was introduced as a generalization of the mover-stayer model proposed in 1955 by Blumen et al. (1955). The mover-stayer model (Blumen et al. 1955) was used to describe labour mobility of low-productivity workers' tendency to move out of their jobs according to a Markov chain, as opposed to those with high-productivity, who tend to stay in their jobs. As shown in Frydman and Schuermann (2008), the Markov mixture process does not have the Markov property. We revisit the mixture model Frydman and Schuermann (2008) for continuous-time finite-state absorbing Markov chains.

The organization of this paper is as follows. We discuss in Section 2 the mixture process for further details. Section 3 presents some preliminary results on the Bayesian updates of switching probability, the transition matrix and its partition, which are derived based on extending the identities (7) and (11) in Frydman and Schuermann (2008) explicitly in terms of the intensity matrices of the underlying Markov chains. The main results concerning the lifetime distribution and the forward intensity are given in Section 4. We also discuss in Section 4 occupation time and residual lifetime for the mixture process. Section 5 presents some numerical examples on married-and-divorced problem, for which divorced intensity data is available for comparison for every marriage cohort in different years, see Aalen et al. (2008, p. 222). Numerical outcomes show that for a given state both intensity and distribution can take different shapes in duration for different cohorts under the mixture model. In particular, they reveal similar resemblances of the intensity data given in Aalen et al. (2008). Moreover, the residual lifetime has the U-bend shape of lifetime found in *The Economist* (2010) article. Section 6 concludes this paper.

2. Mixture of Absorbing Markov Chains

In this paper, we consider $(X_t)_{t\geq 0}$ as a mixture of two Markov jump processes $(X_t^Q)_{t\geq 0}$ and $(X_t^G)_{t\geq 0}$ both of which propagate on the same state space $\mathbb{S} = E \cup \{\Delta\}$, where $E = \{1, 2, \ldots, m\}$ is transient state, while $\Delta = m+1$ is the absorbing state. The intensity matrices $\mathbf{Q}$ and $\mathbf{G}$ for X^Q and X^G are given by

$$\mathbf{Q} = \begin{pmatrix} \mathbf{T} & \boldsymbol{\delta} \\ \mathbf{0} & 0 \end{pmatrix} \qquad \text{and} \qquad \mathbf{G} = \begin{pmatrix} \boldsymbol{\Psi}\mathbf{T} & \boldsymbol{\Psi}\boldsymbol{\delta} \\ \mathbf{0} & 0 \end{pmatrix}, \tag{1}$$

where $\mathbf{T}$ is $(m \times m)$-nonsingular matrix, while $\boldsymbol{\delta}$ is $(m \times 1)$–vector. The $(m \times m)$-matrix $\boldsymbol{\Psi}$ is as such that $\mathbf{G}$ satisfies the property of intensity matrix. Both phase-generator matrix $\mathbf{T}$ and exit intensity $\boldsymbol{\delta}$ satisfy the equality constraint:

$$\mathbf{T}\mathbf{1} + \boldsymbol{\delta} = \mathbf{0}. \tag{2}$$

As $\mathbf{1}^\top \boldsymbol{\delta} > 0$, it implies following (2) that the matrix $\mathbf{T}$ is negative definite.

Frydman and Schuermann (2008) proposed the use of mixture model for the estimation of credit rating dynamics of risky bonds. In their work, intensity matrix $\mathbf{G}$ was taken in a simpler form where $\boldsymbol{\Psi}$ was given by diagonal matrix

$$\boldsymbol{\Psi} = \mathrm{diag}(\psi_1, \psi_2, \ldots, \psi_m). \tag{3}$$

Furthermore, they considered the absorbing state Δ corresponding to censored events such as bond issuing firms being in default, the bonds are withdrawn from being rated, expiration of the bonds, or calling back of the bonds, etc.

Markov mixture process is a generalization of mover-stayer model, a mixture of two discrete-time Markov chains which was introduced by Blumen et al. (1955) in 1955 to model population heterogeneity in labour mobility. In the mover-stayer model (Blumen et al. 1955), the population of workers consists of stayers (workers who always stay in the same job category ($\psi_i = 0$)) and movers (workers who move according to a stationary Markov chain with intensity matrix $\mathbf{Q}$). The estimation of mover-stayer model was given by Frydman (1984). The extension to a mixture of a finite number of continuous-time Markov chains moving at different speeds ($\psi_i \neq 0$) was proposed by Frydman (2005). Frydman and Schuermann (2008) later on used the mixture process to model the dynamics of firms' credit ratings. As empirically shown in Frydman and Schuermann (2008), there is strong evidence that firms of the same credit rating move at different speeds to other ratings, a feature that is lacking in the Markov model.

The mixture is defined conditionally on a given initial state. We denote by $\phi = \mathbf{1}_{\{X = X^G\}}$ random variable having value 1 when $X = X^G$ and 0 otherwise,

$$X = \begin{cases} X^G, & \phi = 1 \\ X^Q, & \phi = 0. \end{cases} \tag{4}$$

We see following (4) that the mixture process X is defined in terms of the underlying Markov chains, X^G and X^Q, and the speed regime ϕ.

Note that when the random variable ϕ is independent of the Markov chains X^G and X^Q having Bernoulli distribution with probability of success s_{i_0} (5), (4) reduces to a regime switching model with changing intensity matrices $\mathbf{Q}$ and $\mathbf{G}$.

In the sequel below, unless stated otherwise, we denote by $\tilde{\boldsymbol{\pi}}^\top = (\boldsymbol{\pi}^\top, \pi_{m+1})$ an initial distribution of the mixture process X on $\mathbb{S}$, where $\boldsymbol{\pi}$ is an E-vector.

From credit risk applications (Jarrow et al. 1997), Frydman (2005) and Frydman and Schuermann (2008) suggest in view of (4) that there are two sub-classes of bond populations each of which moving according to its own Markov chains, one w.r.t X^Q and the other according to X^G. For initial state $X_0 = i_0 \in E$, there is a separate mixing distribution s_{i_0} defined by

$$s_{i_0} = \mathbb{P}\{\phi = 1 \mid X_0 = i_0\}, \tag{5}$$

with $0 \le s_{i_0} \le 1$. The quantity s_{i_0} has the interpretation as being the proportion of population with initial state i_0 moving according to X^G, while $1 - s_{i_0}$ is the proportion that propagates according to X^Q. It follows from (1) that in general X^Q and X^G are different in rates at which the process leaves the states, i.e., $q_i \neq g_i$, but under (3) both of them have the same transition probability of leaving from state $i \in E$ to state $j \in \mathbb{S}$, $i \neq j$, i.e., $q_{ij}/q_i = g_{ij}/g_i$. Note that we have used q_i and g_i to denote the absolute values of diagonal elements of $\mathbf{Q}$ and $\mathbf{G}$, respectively. Thus, depending on the value of ψ_i (3), X^G never moves out of state i when $\psi_i = 0$ (mover-stayer model), moves out of state i at lower rate when $0 < \psi_i < 1$ and moves at higher rate when $\psi_i > 1$ (or the same for $\psi_i = 1$) than X^Q. If $\psi_i = 1$, for all $i = 1, \ldots, m$, X simplifies to a simple Markov process.

The main feature of the Markov mixture process X (4) is that unlike its components X^Q and X^G, X does not inherit the Markovian property; the conditional distribution of future state of X thus depends on its past information. Empirical evidence of this fact was given by Frydman and Schuermann (2008). It was shown in Frydman and Schuermann (2008) that the incorporation of past information helps improve the out-of-sample prediction of the Nelson-Aalen estimate of the cumulative intensity rate for risky bonds. This path dependence property certainly gives favourable feature in applications. Motivated by these findings, we derive the distribution of lifetime (time until absorption τ of X) and introduce forward intensity associated with the distribution. This intensity determines the rate of occurrence of future events given currently available information of the process. Also, we derive the occupation time of X in any state $i \in \mathbb{S}$ as well as its residual lifetime.

Below we present some preliminaries needed to establish the main results.

3. Preliminaries

Recall that the mixture model (4) contains two regimes (processes) of different speeds $\mathbf{Q}$ and $\mathbf{G}$. The regime is however not directly observable, but can be identified based on past information (Frydman 2005, Frydman and Schuermann 2008). To incorporate past information, we denote by $\mathcal{F}_{t-} = \{X(s), 0 \le s \le t-\}$ the available realization of X prior to time t, and by $\mathcal{F}_{i,t} = \mathcal{F}_{t-} \cup \{X_t = i\}$ all previous and current information of X. The set $\mathcal{F}_{t-}$ may contain complete, partial or maybe no information about X.

The likelihoods of observing the past realization $\mathcal{F}_{i,t}$ of X under X^Q and X^G conditional on the initial state $X_0 = i_0$ are given respectively, following Frydman and Schuermann (2008), by

$$\begin{aligned} L^Q &:= \mathbb{P}\{\mathcal{F}_{i,t} \mid \phi = 0, X_0 = i_0\} = \prod_{k \in E} e^{-q_k T_k} \prod_{j \neq k,\, j \in \mathbb{S}} (q_{kj})^{N_{kj}}, \\ L^G &:= \mathbb{P}\{\mathcal{F}_{i,t} \mid \phi = 1, X_0 = i_0\} = \prod_{k \in E} e^{-g_k T_k} \prod_{j \neq k,\, j \in \mathbb{S}} (g_{kj})^{N_{kj}}, \end{aligned} \tag{6}$$

where in the both expressions we have denoted subsequently by T_k and N_{kj} the total time the process spent in state k for $\mathcal{F}_{i,t}$, $k \in \mathbb{S}$, and the number of transitions from state k to state j, with $j \neq k$, observed in $\mathcal{F}_{i,t}$; whereas q_{ij} and g_{ij} represent the (i,j)-entry of the intensity matrices $\mathbf{Q}$ and $\mathbf{G}$, respectively.

Throughout the remainder of this paper, we define quantity $s_i(t)$ by

$$s_i(t) = \mathbb{P}\{\phi = 1 \mid \mathcal{F}_{i,t}\}, \quad i \in \mathbb{S}, \quad t \ge 0. \tag{7}$$

Note that in the case $t = 0$, for which $\mathcal{F}_{i,t} = \{X_0 = i_0\}$, $s_i(0) = s_{i_0}$ (5). It specifies the Bayesian update of the probability of switching the rate at which the process leaves the state i; given that the process ends in state i at time t and all its prior realizations $\mathcal{F}_{t-}$, the proportion of those moving according to X^G is $s_i(t)$. Without loss of generality, similar to s_{i_0}, $s_i(t)$ can be thought of as the time-t proportion of population in state i that moves w.r.t. Markov chain X^G.

Next, we denote by $\mathbf{S}_n(t)$ diagonal matrix, with $n \in \{m, m+1\}$, defined by

$$\mathbf{S}_n(t) = \mathrm{diag}(s_1(t), s_2(t), \ldots, s_n(t)), \quad \text{with } \mathbf{S}_n := \mathbf{S}_n(0). \tag{8}$$

Depending on the availability of the past information $\mathcal{F}_{i,t}$ of X, the elements $s_i(t)$ of the information matrix $\mathbf{S}_m(t)$ (8) are given following (7) below.

Lemma 3.1. *For a given $t \geq 0$ and any state $i, i_0 \in \mathbb{S}$, we have*

(i) *in case of full information $\mathcal{F}_{i,t} = \{X(s), 0 \leq s \leq t-\} \cup \{X(t) = i\}$,*

$$s_i(t) = \frac{s_{i_0} L^G}{s_{i_0} L^G + (1 - s_{i_0}) L^Q}, \tag{9}$$

where s_{i_0}, L^G and L^Q are defined respectively in (5) and (6).

(ii) *in case of limited information $\mathcal{F}_{i,t} = \{X_0 = i_0\} \cup \{X(t) = i\}$,*

$$s_i(t) = \frac{\mathbf{e}_{i_0}^\top \mathbf{S}_{m+1} e^{\mathbf{G}t} \mathbf{e}_i}{\mathbf{e}_{i_0}^\top (\mathbf{S}_{m+1} e^{\mathbf{G}t} + [\mathbf{I}_{m+1} - \mathbf{S}_{m+1}] e^{\mathbf{Q}t}) \mathbf{e}_i}. \tag{10}$$

(iii) *in case of limited information $\mathcal{F}_{i,t} = \{X(t) = i\}$,*

$$s_i(t) = \frac{\tilde{\boldsymbol{\pi}}^\top \mathbf{S}_{m+1} e^{\mathbf{G}t} \mathbf{e}_i}{\tilde{\boldsymbol{\pi}}^\top (\mathbf{S}_{m+1} e^{\mathbf{G}t} + [\mathbf{I}_{m+1} - \mathbf{S}_{m+1}] e^{\mathbf{Q}t}) \mathbf{e}_i}. \tag{11}$$

Proof. The proof of (9) is given in Frydman and Schuermann (2008, p. 1074), while (11) and (10) are the matrix notation of the corresponding quantity $s_i(t)$ given in Frydman and Schuermann (2008, p. 1074). □

Below we give the value of $s_i(t)$ as $t \to \infty$ in case of limited information.

Proposition 3.2. *Let $\mathbf{T}$ ($\boldsymbol{\Psi}\mathbf{T}$) have distinct eigenvalues $\varphi_j^{(1)}$ ($\varphi_j^{(2)}$), $j = 1, \ldots, m$, with $\varphi_{p_k}^{(k)} = \max\{\varphi_j^{(k)}, j = 1, \ldots, m\}$ and $p_k = \arg\max_j \{\varphi_j^{(k)}\}$, $k = 1, 2$. Then,*

$$\lim_{t \to \infty} s_i(t) = \begin{cases} \begin{cases} 1, & \text{if } \varphi_{p_2}^{(2)} > \varphi_{p_1}^{(1)} \\ 0, & \text{if } \varphi_{p_2}^{(2)} < \varphi_{p_1}^{(1)} \\ \dfrac{\boldsymbol{\pi}^\top \mathbf{S}_m L_{p_2}(\boldsymbol{\Psi}\mathbf{T}) \mathbf{e}_i}{\boldsymbol{\pi}^\top (\mathbf{S}_m L_{p_2}(\boldsymbol{\Psi}\mathbf{T}) + (\mathbf{I}_m - \mathbf{S}_m) L_{p_1}(\mathbf{T})) \mathbf{e}_i}, & \text{if } \varphi_{p_2}^{(2)} = \varphi_{p_1}^{(1)} \end{cases} & \text{for } i \in E \\ \dfrac{\boldsymbol{\pi}^\top \mathbf{S}_m \mathbf{1}_m}{1 - \pi_{m+1}}, \quad \text{for } i = m+1, \end{cases} \tag{12}$$

where $L_{p_1}(\mathbf{T})$ and $L_{p_2}(\boldsymbol{\Psi}\mathbf{T})$ are the Lagrange interpolation coefficients given by

$$L_{p_1}(\mathbf{T}) = \prod_{j=1,\, j \neq p_1}^{m} \frac{\mathbf{T} - \varphi_j^{(1)} \mathbf{I}_m}{\varphi_{p_1}^{(1)} - \varphi_j^{(1)}} \quad \text{and} \quad L_{p_2}(\boldsymbol{\Psi}\mathbf{T}) = \prod_{j=1,\, j \neq p_2}^{m} \frac{\boldsymbol{\Psi}\mathbf{T} - \varphi_j^{(2)} \mathbf{I}_m}{\varphi_{p_2}^{(2)} - \varphi_j^{(2)}}. \tag{13}$$

Proof. Consider the matrix partition (16) for $e^{\mathbf{Q}t}$ and $e^{\mathbf{G}t}$. Since $\mathbf{T}$ and $\boldsymbol{\Psi}\mathbf{T}$ are negative definite, it is known that their eigenvalues have negative real part, and that among them there is a dominant eigenvalue (with the smallest absolute value of the real part) which is unique and real (see O'Cinneide 1990, p. 27). Then, following Theorem 2 in Apostol (1969), we have for intensity matrices $\mathbf{T}$ ($\boldsymbol{\Psi}\mathbf{T}$), with m distinct eigenvalues $\varphi_1^{(j)}, \ldots, \varphi_m^{(j)}$, for $j = 1, 2$, that $\exp(\mathbf{T}t)$ has the form:

$$\exp(\mathbf{T}t) = \sum_{k=1}^{m} \exp(\varphi_k^{(1)} t) L_k(\mathbf{T}), \tag{14}$$

similarly defined for $\exp(\boldsymbol{\Psi}\mathbf{T}t)$. Applying (14) to (11) we arrive at (12). □

Recall that under (3), the dominance of eigenvalues of $\mathbf{T}$ over $\mathbf{\Psi T}$ holds when $\psi_i > 1$, for $i = 1, \ldots, m$, in which case $\lim_{t\to\infty} s_i(t) = 0$, or 1 otherwise, under limited information. As $s_i(t)$ depends on past realization of X, we can not in general specify its limiting value under full information for transient states i.

Assumption 3.3. *We assume throughout that* $\mathbf{\Psi}$ *is a nonsingular matrix.*

The $\mathscr{F}_t$-conditional transition matrix $\{P_{ij}(t,s) := \mathbb{P}\{X_s = j \mid \mathscr{F}_{i,t}\}: i, j \in \mathbb{S}\}$ constitutes the basic building block for deriving the distribution of lifetime of X and its forward intensity. Using Lemma 3.1, this conditional transition matrix is given below as an adaptation in matrix notation of Equation (11) on p. 1074 of Frydman and Schuermann (2008).

Theorem 3.4. *For any* $s \geq t \geq 0$, *the* $\mathscr{F}_t$*-conditional transition matrix is*

$$\mathbf{P}(t,s) = \mathbf{S}_{m+1}(t)e^{\mathbf{G}(s-t)} + [\mathbf{I}_{m+1} - \mathbf{S}_{m+1}(t)]e^{\mathbf{Q}(s-t)}. \tag{15}$$

Proof. Applying the law of total probability along with Bayes' theorem, we have

$$\begin{aligned}\mathbb{P}\{X_s = j \mid \mathscr{F}_{i,t}\} &= \mathbb{P}\{X_s = j, \phi = 1 \mid \mathscr{F}_{i,t}\} + \mathbb{P}\{X_s = j, \phi = 0 \mid \mathscr{F}_{i,t}\}\\ &= \mathbb{P}\{X_s = j \mid \phi = 1, \mathscr{F}_{i,t}\}\mathbb{P}\{\phi = 1 \mid \mathscr{F}_{i,t}\} + \mathbb{P}\{X_s = j \mid \phi = 0, \mathscr{F}_{i,t}\}\mathbb{P}\{\phi = 0 \mid \mathscr{F}_{i,t}\}\\ &= p_{ij}^{G}(s-t)s_i(t) + p_{ij}^{Q}(s-t)(1 - s_i(t)),\end{aligned}$$

Denote by $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)^\top$ $(m+1)\times 1$-vector with 1 at the ith element,

$$\mathbb{P}\{X_s = j \mid \mathscr{F}_{i,t}\} = \mathbf{e}_i^\top(\mathbf{S}_{m+1}(t)e^{\mathbf{G}(s-t)} + [\mathbf{I}_{m+1} - \mathbf{S}_{m+1}(t)]e^{\mathbf{Q}(s-t)})\mathbf{e}_j. \quad \square$$

It is clear following (15) that, unless $\mathbf{\Psi} = \mathbf{I}_m$, X does not inherit the Markov and stationary property of X^G and X^Q, i.e., future development of X is determined by its past information $\mathscr{F}_{i,t}$ through its likelihoods and the age (t).

The result generalizes transition matrix for simple Markov process.

Corollary 3.5. *Let* $\mathbf{\Psi} = \mathbf{I}_m$. *Then, the transition probability of* X *is given by*

$$\mathbf{P}(t,s) = e^{\mathbf{Q}(s-t)},$$

implying that the mixture process is a simple homogeneous Markov chain.

Lemma 3.6. *Let phase-generator matrix* $\mathbf{T}$ *of* $\mathbf{Q}$ *be nonsingular. Then,*

$$\exp(\mathbf{Q}) = \begin{pmatrix} \exp(\mathbf{T}) & \mathbf{T}^{-1}(\exp(\mathbf{T}) - \mathbf{I})\boldsymbol{\delta} \\ \mathbf{0} & 1 \end{pmatrix}, \tag{16}$$

Proof. Given the representation of intensity matrix $\mathbf{Q}$, for $n = 0, 1, 2, \ldots$,

$$\mathbf{Q}^0 = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}, \qquad \mathbf{Q}^2 = \begin{pmatrix} \mathbf{T}^2 & \mathbf{T}^{-1}\mathbf{T}^2\boldsymbol{\delta} \\ \mathbf{0} & 0 \end{pmatrix}, \ldots, \mathbf{Q}^n = \begin{pmatrix} \mathbf{T}^n & \mathbf{T}^{-1}\mathbf{T}^n\boldsymbol{\delta} \\ \mathbf{0} & 0 \end{pmatrix}.$$

Therefore, following the series representation of exponential matrix $\exp(\mathbf{Q})$,

$$\exp(\mathbf{Q}) = \sum_{n=0}^{\infty} \frac{\mathbf{Q}^n}{n!} = \begin{pmatrix} \displaystyle\sum_{n=0}^{\infty} \frac{\mathbf{T}^n}{n!} & \displaystyle\mathbf{T}^{-1}\sum_{n=1}^{\infty}\frac{\mathbf{T}^n}{n!}\boldsymbol{\delta} \\ \mathbf{0} & 1 \end{pmatrix}. \quad \square$$

Proposition 3.7. *The conditional transition matrix* (15) *has the partition*

$$\mathbf{P}(t,s) = \begin{pmatrix} \mathbf{F}_{11}(t,s) & \mathbf{F}_{12}(t,s) \\ \mathbf{0} & 1 \end{pmatrix}, \tag{17}$$

where the matrix entries $\mathbf{F}_{11}(t,s)$ *and* $\mathbf{F}_{12}(t,s)$ *are defined by*

$$\begin{aligned}\mathbf{F}_{11}(t,s) &= \mathbf{S}_m(t)e^{\mathbf{\Psi T}(s-t)} + (\mathbf{I}_m - \mathbf{S}_m(t))e^{\mathbf{T}(s-t)},\\ \mathbf{F}_{12}(t,s) &= \mathbf{S}_m(t)(\mathbf{\Psi T})^{-1}(e^{\mathbf{\Psi T}(s-t)} - \mathbf{I}_m)\mathbf{\Psi}\boldsymbol{\delta} + (\mathbf{I}_m - \mathbf{S}_m(t))\mathbf{T}^{-1}(e^{\mathbf{T}(s-t)} - \mathbf{I}_m)\boldsymbol{\delta}.\end{aligned}$$

Proof. The proof follows from applying Lemma 3.6 to $e^{\mathbf{Q}t}$ and $e^{\mathbf{G}t}$, and by partitioning matrices $\mathbf{I}_{m+1}$ and $\mathbf{S}_{m+1}$ (8) into block diagonal matrices. $\square$

4. Lifetime Distributions of the Mixture Process

This section presents the main results of this paper. First, we derive the distribution of lifetime τ of the mixture process X until absorption, i.e.,

$$\tau = \inf\{t > 0 \colon X_t = \Delta\}.$$

The conditional distribution is given based on available information of X. The unconditional distribution follows by taking into account initial distribution of X in each state. Secondly, we study the dense and closure properties of the distribution based on its Laplace transform and matrix analytic approach (Neuts 1981). Afterwards, we discuss conditional forward intensity and occupation time of X.

4.1. Conditional Distribution

Theorem 4.1. *For a given $t \geq 0$, the $\mathcal{F}_{i,t}$-conditional lifetime distribution $\bar{F}_i(t,s) := \mathbb{P}\{\tau > s \mid \mathcal{F}_{i,t}\}$ of X and its density $f_i(t,s)$ are given for any $s \geq t$ by*

$$\begin{aligned} \bar{F}_i(t,s) &= \mathbf{e}_i^\top(\mathbf{S}_m(t)e^{\mathbf{\Psi T}(s-t)} + (\mathbf{I}_m - \mathbf{S}_m(t))e^{\mathbf{T}(s-t)})\mathbf{1}_m, \\ f_i(t,s) &= \mathbf{e}_i^\top(\mathbf{S}_m(t)e^{\mathbf{\Psi T}(s-t)}\mathbf{\Psi} + (\mathbf{I}_m - \mathbf{S}_m(t))e^{\mathbf{T}(s-t)})\boldsymbol{\delta}, \end{aligned} \tag{18}$$

for any initial state $i \in E$ that X started in at time t and is zero for $i = m+1$.

Under full information, we observe that the distribution has path dependence on its past information through the likelihood functions of the mixture process. This is due to the fact that the entry $\{s_i(t)\}$ of the diagonal matrix $\mathbf{S}_m(t)$ is given by (9). However, when the information is limited to only the current (and initial) state, the distribution forms a nonstationary function of time t and duration $d = s - t$, in which case $\{s_i(t)\}$ is given by (10)–(11). In both cases, we see that the distribution has the ability to capture heterogeneity represented by the speed matrix $\mathbf{\Psi}$ and the age (t). These properties are removed when $\mathbf{\Psi} = \mathbf{I}_m$.

Proof. As τ is the time until absorption of X, we have by (15)

$$\bar{F}_i(t,s) = \sum_{j=1}^{m} \mathbb{P}\{X_s = j \mid \mathcal{F}_{i,t}\} = \sum_{j=1}^{m} \mathbf{e}_i^\top \mathbf{P}(t,s)\mathbf{e}_j = \mathbf{e}_i^\top \mathbf{P}(t,s)(\mathbf{1}_{m+1} - \mathbf{e}_{m+1}).$$

On account of the fact that $\mathbf{1}_{m+1} - \mathbf{e}_{m+1} = (\mathbf{1}_m^\top, 0)^\top$, the claim for distribution $\bar{F}_i(t,s)$ follows on account of (17). The density $f_i(t,s)$ is obtained by taking partial derivative of $\bar{F}_i(t,s)$ w.r.t s and then applying the condition (2). □

Observe following the above proof that the $(m+1)$th elements of $\mathbf{S}_{m+1}$ and $\tilde{\boldsymbol{\pi}}$ do not play any role in getting the final result (18). So, necessarily we can set $s_{m+1} = 0$ and $\pi_{m+1} = 0$, which we will later use in a numerical example.

4.2. Unconditional Distribution

Setting $t = 0$ in the main result (18) and summing up over all $i \in E$ for each initial distribution $\mathbb{P}\{X_0 = i\}$, we obtain the unconditional lifetime distribution $\bar{F}(t) = \mathbb{P}\{\tau > t\}$ and density $f(t) = -\bar{F}'(t)$ for the mixture process X.

Theorem 4.2. *The unconditional lifetime distribution $\bar{F}(t)$ and its density $f(t)$ are given for any $t \geq 0$ by*

$$\bar{F}(t) = \boldsymbol{\pi}^\top(\mathbf{S}_m e^{\mathbf{\Psi T}t} + (\mathbf{I}_m - \mathbf{S}_m)e^{\mathbf{T}t})\mathbf{1}_m, \quad f(t) = \boldsymbol{\pi}^\top(\mathbf{S}_m e^{\mathbf{\Psi T}t}\mathbf{\Psi} + (\mathbf{I}_m - \mathbf{S}_m)e^{\mathbf{T}t})\boldsymbol{\delta}. \tag{19}$$

Proof. The proof is based on similar arguments employed before to establish (18) and by conditioning X on its initial position X_0. Using the result in (18),

$$\bar{F}(t) = \sum_{i=1}^{m+1}\sum_{j=1}^{m} \mathbb{P}\{X_0 = i\}\mathbb{P}\{X_t = j \mid X_0 = i\} = (\boldsymbol{\pi}^\top, \pi_{m+1})\mathbf{P}(0,t)(\mathbf{1}_{m+1} - \mathbf{e}_{m+1}),$$

where we have used $(\boldsymbol{\pi}^\top, \pi_{m+1})^\top = \sum_{i=1}^{m+1} \mathbb{P}\{X_0 = i\}\mathbf{e}_i$, while π_{m+1} is the mass of initial distribution on the absorbing state. Our claim for survival function $\bar{F}(t)$ is complete on account of (17). On recalling (2), we have $f(t) = -\bar{F}'(t)$. □

Remark 4.3. Observe that the unconditional distribution (19) is specified by parameters $(\boldsymbol{\pi}, \mathbf{T}, \mathbf{\Psi}, \mathbf{S}_m)$, while the characterization of the conditional distribution (18) is given by replacing $\boldsymbol{\pi}$ by $\mathbf{e}_i$ and $\mathbf{S}_m$ by $\mathbf{S}_m(t)$. This in turn extends the phase-type distribution (Neuts 1975), which is specified by $(\boldsymbol{\pi}, \mathbf{T})$. Throughout the remainder of this paper, we refer to $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \mathbf{\Psi}, \mathbf{S}_m)$ the lifetime distribution of the mixture process. Setting $\mathbf{S}_m = \alpha\mathbf{I}_m$, with $0 \leq \alpha \leq 1$, in (19) leads to the convex mixture of two phase-type distributions $\mathrm{PH}(\boldsymbol{\pi}, \mathbf{\Psi T})$ and $\mathrm{PH}(\boldsymbol{\pi}, \mathbf{T})$,

$$\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \mathbf{\Psi}, \alpha\mathbf{I}_m) = \alpha\mathrm{PH}(\boldsymbol{\pi}, \mathbf{\Psi T}) + (1-\alpha)\mathrm{PH}(\boldsymbol{\pi}, \mathbf{T}).$$

Thus, (19) is a generalized mixture of the phase-type distributions.

Corollary 4.4. *By setting* $\boldsymbol{\Psi} = \mathbf{I}_m$ *in* (19), *the distribution reduces to Neuts* (1975).

As the phase-type distributions $\mathrm{PH}(\boldsymbol{\pi}, \mathbf{T})$ and $\mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{\Psi}\mathbf{T})$ represent the lifetime distribution (19) in special case, i.e., $\mathrm{PH}(\boldsymbol{\pi}, \mathbf{T}) := \mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \mathbf{I}_m, \mathbf{S}_m)$ and $\mathrm{PH}(\boldsymbol{\pi}, \boldsymbol{\Psi}\mathbf{T}) := \mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{I}_m)$, we will show in Section 4.5 that $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$ is closed under finite convex mixtures and finite convolutions.

4.3. Laplace Transform of the Lifetime Distributions

The theorem below gives the Laplace transform of (19) and its nth moment.

Theorem 4.5. *Let F be the lifetime distribution* $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$ (19). *Then,*

(i) *the Laplace transform* $\hat{F}[\theta] = \int_0^\infty e^{-\theta u} f(u)\, du$ *is given by*

$$\hat{F}[\theta] = \boldsymbol{\pi}^\top (\mathbf{S}_m(\theta \mathbf{I}_m - \boldsymbol{\Psi}\mathbf{T})^{-1}\boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m)(\theta\mathbf{I}_m - \mathbf{T})^{-1})\boldsymbol{\delta}. \tag{20}$$

(ii) *the nth moment, for $n = 0, 1, \ldots,$ of τ is given by*

$$\mathbb{E}\{\tau^n\} = (-1)^n n! \boldsymbol{\pi}^\top (\mathbf{S}_m(\boldsymbol{\Psi}\mathbf{T})^{-n} + (\mathbf{I}_m - \mathbf{S}_m)\mathbf{T}^{-n})\mathbf{1}_m. \tag{21}$$

Remark 4.6. As the conditional distribution (18) is a function of two time-variables t and s, with $s \geq t$, we can reparameterize it in terms of initial time t and remaining time $v := s - t \geq 0$ such that, e.g., density can be rewritten as

$$f_i(t, v) = \mathbf{e}_i^\top (\mathbf{S}_m(t) e^{\boldsymbol{\Psi}\mathbf{T}v}\boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m(t))e^{\mathbf{T}v})\boldsymbol{\delta},$$

whose Laplace transform $\hat{F}_t^{(i)}[\theta] = \int_0^\infty e^{-\theta v} f_i(t, v)\, dv$ is given below for any fixed $t \geq 0$ and $i \in E$. The result is obtained in the same way as before to get (20),

$$\hat{F}_t^{(i)}[\theta] = \mathbf{e}_i^\top (\mathbf{S}_m(t)(\theta\mathbf{I}_m - \boldsymbol{\Psi}\mathbf{T})^{-1}\boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m(t))(\theta\mathbf{I}_m - \mathbf{T})^{-1})\boldsymbol{\delta}. \tag{22}$$

So, there is similarity between the Laplace transform (20) and (22). For (22) the initial distribution $\boldsymbol{\pi}$ has mass one on state i and $\mathbf{S}_m$ has dependence on t.

Proof. As the sub-intensity matrices $\mathbf{T}$ and $\boldsymbol{\Psi}$ are nonsingular, then following Lemma 8.2.5 in Rolski et al. (1998) $\theta\mathbf{I}_m - \mathbf{T}$ and $\theta\mathbf{I}_m - \boldsymbol{\Psi}\mathbf{T}$ are nonsingular for each $\theta \geq 0$ and whose entries are all rational functions of $\theta \geq 0$. Furthermore, for all $\theta \geq 0$,

$$\int_0^\infty e^{-\theta u} e^{\mathbf{B}u}\, du = (\theta\mathbf{I}_m - \mathbf{B})^{-1},$$

for any nonsingular matrix $\mathbf{B}$. The Laplace transform of (19) is therefore

$$\begin{aligned}\hat{F}[\theta] &= \int_0^\infty e^{-\theta u}(\boldsymbol{\pi}^\top(\mathbf{S}_m e^{\Psi T u}\boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m)e^{\mathbf{T}u})\boldsymbol{\delta})\, du \\ &= \boldsymbol{\pi}^\top \mathbf{S}_m \left(\int_0^\infty e^{-\theta u} e^{\Psi T u}\, du\right)\boldsymbol{\Psi}\boldsymbol{\delta} + \boldsymbol{\pi}^\top(\mathbf{I}_m - \mathbf{S}_m)\left(\int_0^\infty e^{-\theta u} e^{\mathbf{T}u}\, du\right)\boldsymbol{\delta} \\ &= \boldsymbol{\pi}^\top \mathbf{S}_m(\theta\mathbf{I}_m - \boldsymbol{\Psi}\mathbf{T})^{-1}\boldsymbol{\Psi}\boldsymbol{\delta} + \boldsymbol{\pi}^\top(\mathbf{I}_m - \mathbf{S}_m)(\theta\mathbf{I}_m - \mathbf{T})^{-1}\boldsymbol{\delta}.\end{aligned}$$

Following the same Lemma 8.2.5 of Rolski et al. (1998), we have for all $\theta \geq 0$ and $n \in \mathbb{N}$,

$$\frac{d^n}{d\theta^n}(\theta\mathbf{I}_m - \mathbf{B})^{-1} = (-1)^n n! (\theta\mathbf{I}_m - \mathbf{B})^{-(n+1)},$$

for any nonsingular matrix $\mathbf{B}$. On account of this, the nth derivative of $\hat{F}[\theta]$ is

$$\begin{aligned}\frac{d^n}{d\theta^n}\hat{F}[\theta] &= \boldsymbol{\pi}^\top\left(\mathbf{S}_m \frac{d^n}{d\theta^n}(\theta\mathbf{I}_m - \boldsymbol{\Psi}\mathbf{T})^{-1}\boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m)\frac{d^n}{d\theta^n}(\theta\mathbf{I}_m - \mathbf{T})^{-1}\right)\boldsymbol{\delta} \\ &= (-1)^n n! \boldsymbol{\pi}^\top(\mathbf{S}_m(\theta\mathbf{I}_m - \boldsymbol{\Psi}\mathbf{T})^{-(n+1)}\boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m)(\theta\mathbf{I}_m - \mathbf{T})^{-(n+1)})\boldsymbol{\delta}.\end{aligned}$$

The claim is established by setting $\theta = 0$ and taking into account (2). $\square$

When $\boldsymbol{\Psi} = \mathbf{I}_m$, (20)–(21) reduces to the results given by Theorem 8.2.5 in Rolski et al. (1998):

$$\hat{F}[\theta] = \boldsymbol{\pi}^\top(\theta\mathbf{I}_m - \mathbf{T})^{-1}\boldsymbol{\delta} \qquad \text{and} \qquad \mathbb{E}\{\tau^n\} = (-1)^n n! \boldsymbol{\pi}^\top \mathbf{T}^{-n}\mathbf{1}_m. \tag{23}$$

4.4. Matrix Representation of the Lifetime Distributions

This section discusses matrix representation of distributions $\mathrm{GPH}(\boldsymbol{\pi},\mathbf{T},\boldsymbol{\Psi},\mathbf{S}_m)$ (18)–(19), in a similar way it describes phase-type distributions $\mathrm{PH}(\boldsymbol{\pi},\mathbf{T})$ (Neuts 1975), with the purpose of establishing the closure properties of the distributions.

Lemma 4.7. *The distribution* $\mathrm{GPH}(\boldsymbol{\pi},\mathbf{T},\boldsymbol{\Psi},\mathbf{S}_m)$ *has* $\mathrm{PH}(\tilde{\boldsymbol{\pi}},\tilde{\mathbf{T}})$ *representation,*

$$\tilde{\boldsymbol{\pi}}^\top = ((\mathbf{S}_m\boldsymbol{\pi})^\top, ((\mathbf{I}_m-\mathbf{S}_m)\boldsymbol{\pi})^\top) \qquad \text{and} \qquad \tilde{\mathbf{T}} = \begin{pmatrix} \boldsymbol{\Psi}\mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} \end{pmatrix}. \tag{24}$$

Remark 4.8. Based on the Laplace transform (22), the same conclusion is reached as in Lemma 4.7 for the conditional distribution (18) for which case

$$\tilde{\boldsymbol{\pi}}^\top = ((\mathbf{S}_m(t)\mathbf{e}_i)^\top, ((\mathbf{I}_m-\mathbf{S}_m(t))\mathbf{e}_i)^\top) \qquad \text{and} \qquad \tilde{\mathbf{T}} = \begin{pmatrix} \boldsymbol{\Psi}\mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} \end{pmatrix}. \tag{25}$$

Remark 4.9. Note that even though the two distributions $\mathrm{GPH}(\boldsymbol{\pi},\mathbf{T},\boldsymbol{\Psi},\mathbf{S}_m)$ and $\mathrm{PH}(\tilde{\boldsymbol{\pi}},\tilde{\mathbf{T}})$ have identical form, they have different underlying processes. The Markov chain $\{M_t\colon t\geq 0\}$ associated with $\mathrm{PH}(\tilde{\boldsymbol{\pi}},\tilde{\mathbf{T}})$ is constructed by catenating the states of the underlying Markov chains X^G and X^Q of the Markov mixture process X (4) of $\mathrm{GPH}(\boldsymbol{\pi},\mathbf{T},\boldsymbol{\Psi},\mathbf{S}_m)$. That is to say that the transient state E of M is given by $E=E_1\cup E_2$, where $E_1=\{1,\ldots,m\}$ and $E_2=\{m+1,\ldots,2m\}$, with initial distribution $\tilde{\boldsymbol{\pi}}$, moving with speed $\tilde{\mathbf{T}}$. Recall that under Markov process such construction for the lifetime distribution (18) can not determine the element $s_i(t)$ of the matrix $\mathbf{S}_m(t)$ in the distribution. This is in contrast to the Markov mixture model. Denote by $\tilde{\boldsymbol{\pi}}_1$ and $\tilde{\boldsymbol{\pi}}_2$ the distributions of the Markov chain M on the transient states E_1 and E_2, i.e., $\tilde{\boldsymbol{\pi}}^\top=(\tilde{\boldsymbol{\pi}}_1^\top,\tilde{\boldsymbol{\pi}}_2^\top)$. The corresponding distribution $\boldsymbol{\pi}$ of the Markov mixture process is determined following (24) by $\boldsymbol{\pi}=\tilde{\boldsymbol{\pi}}_1+\tilde{\boldsymbol{\pi}}_2$. As $\mathbf{S}_m$ is a diagonal matrix, each element s_i, $i=1,2,\ldots,m$, of $\mathbf{S}_m$ is given by $s_i=\tilde{\pi}_1^{(i)}/(\tilde{\pi}_1^{(i)}+\tilde{\pi}_2^{(i)})$, where $\tilde{\pi}_1^{(i)}$ and $\tilde{\pi}_2^{(i)}$ are the i-th element of $\tilde{\boldsymbol{\pi}}_1$ and $\tilde{\boldsymbol{\pi}}_2$, provided that $\pi^{(i)}=\tilde{\pi}_1^{(i)}+\tilde{\pi}_2^{(i)}\neq 0$. Otherwise, s_i can be arbitrarily chosen. Recall following (24) that since $\tilde{\mathbf{T}}=\mathrm{diag}(\boldsymbol{\Psi}\mathbf{T},\mathbf{T})$ and $\boldsymbol{\Psi}\mathbf{T}$ and $\mathbf{T}$ are square matrices, we have $e^{\tilde{\mathbf{T}}t}=\mathrm{diag}(e^{\boldsymbol{\Psi}\mathbf{T}t},e^{\mathbf{T}t})$. Therefore, following (24) we have $\bar{F}(t)=\tilde{\boldsymbol{\pi}}^\top e^{\tilde{\mathbf{T}}t}\mathbf{1}$ and $f(t)=\tilde{\boldsymbol{\pi}}^\top e^{\tilde{\mathbf{T}}t}\tilde{\boldsymbol{\delta}}$, which are just the same results (19).

Proof. Following the Laplace transform (20), define $\boldsymbol{\pi}_1=\mathbf{S}_m\boldsymbol{\pi}$, $\boldsymbol{\delta}_1=\boldsymbol{\Psi}\boldsymbol{\delta}$, $\boldsymbol{\pi}_2=(\mathbf{I}_m-\mathbf{S}_m)\boldsymbol{\pi}$ and $\boldsymbol{\delta}_2=\boldsymbol{\delta}$. Respectively, the pairs $(\boldsymbol{\pi}_1,\boldsymbol{\delta}_1)$ and $(\boldsymbol{\pi}_2,\boldsymbol{\delta}_2)$ define the initial proportion of population and their exit rate to absorbing state moving at speeds $\boldsymbol{\Psi}\mathbf{T}$ and $\mathbf{T}$. Having defined these, we can rearrange (20) as:

$$\hat{F}[\theta]=\boldsymbol{\pi}_1^\top(\theta\mathbf{I}_m-\boldsymbol{\Psi}\mathbf{T})^{-1}\boldsymbol{\delta}_1+\boldsymbol{\pi}_2^\top(\theta\mathbf{I}_m-\mathbf{T})^{-1}\boldsymbol{\delta}_2=(\boldsymbol{\pi}_1^\top\ \ \boldsymbol{\pi}_2^\top)\begin{pmatrix}(\theta\mathbf{I}_m-\boldsymbol{\Psi}\mathbf{T})^{-1} & \mathbf{0}\\ \mathbf{0} & (\theta\mathbf{I}_m-\mathbf{T})^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{\delta}_1\\ \boldsymbol{\delta}_2\end{pmatrix}.$$

Notice that we have taken advantage of $\mathbf{S}_m$ and $\mathbf{I}_m-\mathbf{S}_m$ being symmetric matrices. On recalling that $\theta\mathbf{I}_m-\boldsymbol{\Psi}\mathbf{T}$ and $\theta\mathbf{I}_m-\mathbf{T}$ are square nonsingular matrices, the above block diagonal matrix can be further simplified as follows

$$\begin{pmatrix}(\theta\mathbf{I}_m-\boldsymbol{\Psi}\mathbf{T})^{-1} & \mathbf{0}\\ \mathbf{0} & (\theta\mathbf{I}_m-\mathbf{T})^{-1}\end{pmatrix}=\begin{pmatrix}\theta\mathbf{I}_m-\boldsymbol{\Psi}\mathbf{T} & \mathbf{0}\\ \mathbf{0} & \theta\mathbf{I}_m-\mathbf{T}\end{pmatrix}^{-1}.$$

To arrive at the representation of (20) in the framework of Laplace transform (23) of the existing phase-type distribution, define the following matrices

$$\mathbf{I}_{2m}=\begin{pmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_m\end{pmatrix} \qquad \text{and} \qquad \tilde{\mathbf{T}}=\begin{pmatrix}\boldsymbol{\Psi}\mathbf{T} & \mathbf{0}\\ \mathbf{0} & \mathbf{T}\end{pmatrix}.$$

As a result of having defined the two matrices $\mathbf{I}_{2m}$ and $\tilde{\mathbf{T}}$, the Laplace transform (20) of the lifetime distribution (19) is finally given by

$$\hat{F}[\theta]=\tilde{\boldsymbol{\pi}}^\top(\theta\mathbf{I}_{2m}-\tilde{\mathbf{T}})^{-1}\tilde{\boldsymbol{\delta}}, \tag{26}$$

which falls in the form of (23) with $\tilde{\boldsymbol{\pi}}^\top=(\boldsymbol{\pi}_1^\top,\boldsymbol{\pi}_2^\top)$ and $\tilde{\boldsymbol{\delta}}=(\boldsymbol{\delta}_1^\top,\boldsymbol{\delta}_2^\top)^\top$. $\square$

Below we give two examples of the lifetime distribution GPH($\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m$) which encloses known distributions such as Erlang distribution and its mixture.

Example 4.10 (Erlang Distribution). Let F be GPH($\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m$) distribution with characteristics $\boldsymbol{\pi} = (1,0,0)^\top$ and matrices $\mathbf{T}$, $\boldsymbol{\Psi}$ and $\mathbf{S}_m$ given by

$$\mathbf{T} = \begin{pmatrix} -\beta & \beta & 0 \\ 0 & -\beta & \beta \\ 0 & 0 & -\beta \end{pmatrix}, \qquad \boldsymbol{\Psi} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \mathbf{S}_m = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then, following Theorem 4.2, F has the Erlang distribution Erl(m, β):

$$F(t; m, \beta) = 1 - \sum_{k=0}^{m-1} \frac{(\beta t)^k}{k!} e^{-\beta t}, \tag{27}$$

with $m = 3$, whose density function $f(t; m, \beta) = F'(t; m, \beta)$ is given by

$$f(t; m, \beta) = \frac{1}{(m-1)!} t^{m-1} \beta^m e^{-\beta t}. \tag{28}$$

Example 4.11 (Mixture of Erlang Distributions). Let F be GPH($\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m$) distribution with $\boldsymbol{\pi} = (1,0,0)^\top$ and matrices $\mathbf{T}$, $\boldsymbol{\Psi}$ and $\mathbf{S}_m$ given by

$$\mathbf{T} = \begin{pmatrix} -\beta_2 & \beta_2 & 0 \\ 0 & -\beta_2 & \beta_2 \\ 0 & 0 & -\beta_2 \end{pmatrix}, \qquad \boldsymbol{\Psi} = \begin{pmatrix} \beta_1/\beta_2 & 0 & 0 \\ 0 & \beta_1/\beta_2 & 0 \\ 0 & 0 & \beta_1/\beta_2 \end{pmatrix}, \qquad \mathbf{S}_m = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix}.$$

By Lemma 4.7, GPH($\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m$) has identical distribution PH($\tilde{\boldsymbol{\pi}}, \tilde{\mathbf{T}}$) with

$$\tilde{\boldsymbol{\pi}} = \begin{pmatrix} \alpha \\ 0 \\ 0 \\ 1-\alpha \\ 0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \tilde{\mathbf{T}} = \begin{pmatrix} -\beta_1 & \beta_1 & 0 & 0 & 0 & 0 \\ 0 & -\beta_1 & \beta_1 & 0 & 0 & 0 \\ 0 & 0 & -\beta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta_2 & \beta_2 & 0 \\ 0 & 0 & 0 & 0 & -\beta_2 & \beta_2 \\ 0 & 0 & 0 & 0 & 0 & -\beta_2 \end{pmatrix},$$

which is found to be the familiar form of the mixture of Erlang distributions:

$$F(t) = \alpha F(t; m, \beta_1) + (1-\alpha) F(t; m, \beta_2),$$

where $F(t; m, \beta_i)$, for $i = 1, 2$ and $m = 3$, is the Erlang distribution (27).

4.5. Closure and Dense Properties

In this section we prove that the lifetime distribution (19) (and 18) is closed under finite convolutions and finite convex mixtures, i.e., the sum and convex combinations of finitely many GPH($\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m$) distributed random variables have distributions which again belong to the same class. The proof is based on Theorem 4.5 and the matrix representation (24). Furthermore, taking account of closure properties, we show by adapting arguments in Rolski et al. (1998) and Tijms (1994) that GPH($\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m$) forms a dense class of distributions on $\mathbb{R}_+$. The closure and dense properties emphasize the additional importance of the distribution.

Theorem 4.12 (Closure Property). *The distribution* GPH($\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m$) *possesses closure property under finite convex mixtures and finite convolutions.*

Proof. To start with, let $F_k(\cdot)$ be the distribution of GPH($\boldsymbol{\pi}_k, \mathbf{T}_k, \boldsymbol{\Psi}_k, \mathbf{S}_{m_k}$) with m_k transient states, $k = 1, 2$. Assume that $m_1 \neq m_2$ and let $m = \mathrm{lcm}(m_1, m_2)$, the least common multiple of m_1 and m_2. Define $\tilde{m}_1 = m/m_1$, $\tilde{m}_2 = m/m_2$ and

$$\tilde{\boldsymbol{\pi}}_k^\top = (\boldsymbol{\pi}_k^\top \mathbf{S}_{m_k}, \boldsymbol{\pi}_k^\top (\mathbf{I}_{m_k} - \mathbf{S}_{m_k})) \qquad \text{and} \qquad \tilde{\mathbf{T}}_k = \begin{pmatrix} \boldsymbol{\Psi}_k \mathbf{T}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_k \end{pmatrix}.$$

We first show that the convex mixture $F(\cdot) := pF_1(\cdot) + (1-p)F_2(\cdot)$, with $0 \le p \le 1$, has $\text{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S})$ distribution with the following representation:

$$\boldsymbol{\pi} = \frac{p}{\tilde{m}_1}\mathbf{1}_{\tilde{m}_1} \otimes \tilde{\boldsymbol{\pi}}_1 + \frac{(1-p)}{\tilde{m}_2}\mathbf{1}_{\tilde{m}_2} \otimes \tilde{\boldsymbol{\pi}}_2, \qquad \mathbf{S}\boldsymbol{\pi} = \frac{p}{\tilde{m}_1}\mathbf{1}_{\tilde{m}_1} \otimes \tilde{\boldsymbol{\pi}}_1, \qquad \mathbf{T} = \mathbf{I}_{\tilde{m}_2} \otimes \tilde{\mathbf{T}}_2, \qquad \boldsymbol{\Psi} = (\mathbf{I}_{\tilde{m}_1} \otimes \tilde{\mathbf{T}}_1)(\mathbf{I}_{\tilde{m}_2} \otimes \tilde{\mathbf{T}}_2)^{-1},$$

where we have denoted by $\otimes$ the usual Kronecker product. The fact that $\mathbf{S}$ is a diagonal matrix, whose elements can be found from the first two equations. To establish the claim, recall on account of (24) that $\text{GPH}(\boldsymbol{\pi}_k, \mathbf{T}_k, \boldsymbol{\Psi}_k, \mathbf{S}_{m_k})$ has identical distribution with $\text{PH}(\tilde{\boldsymbol{\pi}}_k, \tilde{\mathbf{T}}_k)$. By Theorem 8.2.7 in Rolski et al. (1998), we know that the mixture $p\text{PH}(\tilde{\boldsymbol{\pi}}_1, \tilde{\mathbf{T}}_1) + (1-p)\text{PH}(\tilde{\boldsymbol{\pi}}_2, \tilde{\mathbf{T}}_2)$ has $\text{PH}(\boldsymbol{\alpha}, \mathbf{B})$ distribution with $\boldsymbol{\alpha}^\top = (p\tilde{\boldsymbol{\pi}}_1^\top, (1-p)\tilde{\boldsymbol{\pi}}_2^\top)$ and $\mathbf{B} = \text{diag}(\tilde{\mathbf{T}}_1, \tilde{\mathbf{T}}_2)$. Notice that $\tilde{\mathbf{T}}_1$ and $\tilde{\mathbf{T}}_2$ have different phases, for which reason one can not concatenate the underlying Markov chains of $\text{PH}(\tilde{\boldsymbol{\pi}}_1, \tilde{\mathbf{T}}_1)$ and $\text{PH}(\tilde{\boldsymbol{\pi}}_2, \tilde{\mathbf{T}}_2)$ to form a Markov mixture process X. However, since the matrix representation of phase-type distribution is not unique, see O'Cinneide (1989), $\text{PH}(\boldsymbol{\alpha}, \mathbf{B})$ has identical distribution with $\text{PH}(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{B}})$ with $\tilde{\boldsymbol{\alpha}}^\top = (\boldsymbol{\pi}^\top\mathbf{S}, \boldsymbol{\pi}^\top(\mathbf{I}-\mathbf{S}))$ and $\tilde{\mathbf{B}} = \text{diag}(\boldsymbol{\Psi}\mathbf{T}, \mathbf{T})$. Under $\text{PH}(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{B}})$ representation, there is a corresponding mixture process X constructed by concatenating two Markov chains $\{X_1(t)\colon t \ge 0\}$ and $\{X_2(t)\colon t \ge 0\}$ moving respectively with phase matrix $\mathbf{T}$ and $\boldsymbol{\Psi}\mathbf{T}$ of equal dimension. Lemma 24 concludes the claim.

When $m_1 = m_2$, the mixture $F(\cdot)$ has $\text{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S})$ distribution with

$$\boldsymbol{\pi} = p\tilde{\boldsymbol{\pi}}_1 + (1-p)\tilde{\boldsymbol{\pi}}_2, \qquad \mathbf{S}\boldsymbol{\pi} = p\tilde{\boldsymbol{\pi}}_1, \qquad \mathbf{T} = \tilde{\mathbf{T}}_2, \qquad \text{and} \qquad \boldsymbol{\Psi} = \tilde{\mathbf{T}}_1\tilde{\mathbf{T}}_2^{-1}.$$

Next, we show that the convolution $G(\cdot) = F_1 * F_2(\cdot)$ has $\text{GPH}(\boldsymbol{\pi}, \mathbf{T}, \mathbf{I}, \mathbf{I}/2)$ distribution with

$$\boldsymbol{\pi}^\top = (\tilde{\boldsymbol{\pi}}_1^\top, (1-\tilde{\boldsymbol{\pi}}_1^\top\mathbf{1})\tilde{\boldsymbol{\pi}}_2^\top) \qquad \text{and} \qquad \mathbf{T} = \begin{pmatrix} \tilde{\mathbf{T}}_1 & -\tilde{\mathbf{T}}_1\mathbf{1}\tilde{\boldsymbol{\pi}}_2^\top \\ \mathbf{0} & \tilde{\mathbf{T}}_2 \end{pmatrix}.$$

By Theorem 8.2.6 in Rolski et al. (1998), we have $\text{PH}(\tilde{\boldsymbol{\pi}}_1, \tilde{\mathbf{T}}_1) * \text{PH}(\tilde{\boldsymbol{\pi}}_2, \tilde{\mathbf{T}}_2) = \text{PH}(\boldsymbol{\alpha}, \mathbf{B})$ with $\boldsymbol{\alpha} = \boldsymbol{\pi}$ and $\mathbf{B} = \mathbf{T}$. However, $\text{PH}(\boldsymbol{\alpha}, \mathbf{B})$ has identical form with $\text{PH}(\tilde{\boldsymbol{\alpha}}, \tilde{\mathbf{B}})$, where $\tilde{\boldsymbol{\alpha}}^\top = (\boldsymbol{\pi}^\top\mathbf{I}/2, \boldsymbol{\pi}^\top\mathbf{I}/2)$ and $\tilde{\mathbf{B}} = \text{diag}(\mathbf{T}, \mathbf{T})$. Lemma 24 concludes the claim.

Applying similar arguments, the results above extend in natural way to convex mixtures and convolutions of finitely many GPH distributions, in the same way applies to PH distributions, see e.g., Rolski et al. (1998), we leave it to the reader. □

Theorem 4.13 (Dense Property). *The distribution* $\text{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$ *forms a dense class of distributions on* $\mathbb{R}_+$.

Proof. The proof follows from adapting the approach used in Tijms (1994) and Rolski et al. (1998). Let F be any right continuous distribution on $\mathbb{R}_+$ with $F(0) = 0$. For a fixed $t \ge 0$, assume that F is continuous at t. Consider the following approximation of F:

$$F_n(t) = \sum_{j=1}^{\infty} p_j(n)F(t; j, n), \tag{29}$$

where $F(t; j, n)$ is the Erlang distribution (27) with $\beta = n$ and $p_j(n)$ is

$$p_j(n) = F(j/n) - F((j-1)/n).$$

It is clear that $\sum_{j=1}^{\infty} p_j(n) = 1$. Thus, (29) can be seen as an infinite convex mixture of Erlang distributions. On account of (27), we have following (29)

$$\begin{aligned} F_n(t) &= \sum_{j=1}^{\infty}(F(j/n) - F((j-1)/n))\sum_{k=j}^{\infty} e^{-nt}\frac{(nt)^k}{k!} = \sum_{k=0}^{\infty} e^{-nt}\frac{(nt)^k}{k!}\sum_{j=1}^{k}(F(j/n) - F((j-1)/n)) \\ &= \sum_{k=0}^{\infty} e^{-nt}\frac{(nt)^k}{k!}F(k/n) = \mathbb{E}\{F(X_n(t)/n)\}, \end{aligned}$$

where $X_n(t)$ is a Poisson random variable with intensity nt. Since F is a bounded function, i.e., $\sup_t F(t) \le 1$, and that F is continuous at t, for each $\epsilon > 0$ we choose $\delta > 0$ such that $|F(x) - F(t)| \le \epsilon/2$ whenever $|x - t| \le \delta$. By doing so,

$$\begin{aligned} \mathbb{E}\{|F(X_n(t)/n) - F(t)|\} &= \mathbb{E}\big\{|F(X_n(t)/n) - F(t)|; |X_n(t)/n - t| \le \delta\big\} + \mathbb{E}\big\{|F(X_n(t)/n) - F(t)|; |X_n(t)/n - t| > \delta\big\} \\ &\le \frac{\epsilon}{2} + 2\mathbb{P}\{|X_n(t)/n - t| > \delta\} \le \frac{\epsilon}{2} + 2\frac{1}{\delta^2}\operatorname{Var}(X_n(t)/n) \le \frac{\epsilon}{2} + \frac{2t}{\delta^2 n}, \end{aligned}$$

where we have applied Chebyshev's inequality in the second last inequality and the fact that $\mathbb{E}\{X_n(t)/n\} = t$. Thus, by choosing n big enough so that $2t/(\delta^2 n) \le \epsilon/2$,

$$|F_n(t) - F(t)| \le \epsilon,$$

i.e., $\lim_{n\to\infty} F_n(t) = F(t)$. Next, consider the truncated version of (29): $\tilde{F}_n(t) := \sum_{j=1}^{n^2} p_j(n)F(t;j,n)$. By the above and the monotone convergence theorem, we have $\lim_{n\to\infty} |\tilde{F}_n(t) - F_n(t)| = 0$ for each $t \ge 0$. The proof is complete by recalling Theorem 4.12 and Example 4.11 that a mixture of Erlang distributions belongs to the class of distributions $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$. The case $F(0) \neq 0$ is handled by adding up the term $F(0)\delta_0$ in (29), where δ_0 is Dirac measure at $t = 0$. □

4.6. Forward Intensity

In survival and event history analysis (see e.g., Aalen et al. 2008), it has been common practice to associate an intensity function $\lambda(t)$ with the stopping time τ. This intensity determines the rate of occurrence of an event within the period of time $(t, t+dt]$ given information set $\mathcal{F}_{i,t}$ available at time t. Quite recently, Duan et al. (2012) introduced the notion of forward intensity which to some extent plays a similar role to forward rate in finance. However, as stated in Duan et al. (2012) the forward intensity can only be calculated with the exact knowledge of the underlying Markov process. Motivated by this work, we discuss forward intensity defined for any $s \ge t$ by

$$\lambda_i(t,s)\,ds = \mathbb{P}\{s < \tau \le s + ds \mid \tau > s, \mathcal{F}_{i,t}\}, \tag{30}$$

which corresponds to the usual intensity $\lambda_i(t)$ (see Aalen et al. 2008) when we set $s = t$. Following (30), the forward intensity $\lambda_i(t,s)$ determines the rate of occurrence of a future event in $(s, s+ds]$ based on the information $\mathcal{F}_{i,t}$ available at time t.

In the theorem below we give an explicit expression for the forward intensity.

Theorem 4.14. *For a given $s \ge t \ge 0$ and information $\mathcal{F}_{i,t}$, with any $i \in E$,*

$$\lambda_i(t,s) = \frac{\mathbf{e}_i^\top(\mathbf{S}_m(t)e^{\boldsymbol{\Psi}\mathbf{T}(s-t)}\boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m(t))e^{\mathbf{T}(s-t)})\boldsymbol{\delta}}{\mathbf{e}_i^\top(\mathbf{S}_m(t)e^{\boldsymbol{\Psi}\mathbf{T}(s-t)} + (\mathbf{I}_m - \mathbf{S}_m(t))e^{\mathbf{T}(s-t)})\mathbf{1}_m}, \tag{31}$$

based on which the distribution $\bar{F}_i(t,s)$ and density $f_i(t,s)$ (18) are given by

$$\bar{F}_i(t,s) = \exp\left(-\int_t^s \lambda_i(t,u)\,du\right), \qquad f_i(t,s) = \lambda_i(t,s)\exp\left(-\int_t^s \lambda_i(t,u)\,du\right). \tag{32}$$

Proof. As τ is absolutely continuous, the proof follows from (30) and (18). □

However, when we set $s = t$ in (31), we obtain intensity rate $\lambda_i(t)$ for occurrence of an event within interval $(t, t+dt]$. This intensity is given below.

Corollary 4.15. *Given $i \in E$ and past information $\mathcal{F}_{i,t}$, we have for any $t \ge 0$,*

$$\lambda_i(t) = \mathbf{e}_i^\top(\mathbf{I}_m + \mathbf{S}_m(t)(\boldsymbol{\Psi} - \mathbf{I}_m))\boldsymbol{\delta}. \tag{33}$$

As it plays a central role in survival analysis, it is interesting to see that the inclusion of past information of X, particularly when the information $\mathcal{F}_{i,t}$ is fully available, on the intensity λ is now made possible under the mixture model. The long-term rate of $\lambda_i(t)$ can be worked out using the results of Proposition 3.2.

Observe that the intensities (31) and (33) are both dependent on past realization $\mathcal{F}_{i,t}$ of X. Below we give an explicit expression for the baseline intensity $\alpha(t)dt = \mathbb{P}\{t < \tau \le t + dt \mid \tau > t\}$ of the unconditional distribution (19).

Corollary 4.16. *For any $t \ge 0$, the baseline intensity function of τ is given by*

$$\alpha(t) = \frac{\boldsymbol{\pi}^\top(\mathbf{S}_m e^{\boldsymbol{\Psi}\mathbf{T}t}\boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m)e^{\mathbf{T}t})\boldsymbol{\delta}}{\boldsymbol{\pi}^\top(\mathbf{S}_m e^{\boldsymbol{\Psi}\mathbf{T}t} + (\mathbf{I}_m - \mathbf{S}_m)e^{\mathbf{T}t})\mathbf{1}_m}, \tag{34}$$

based on which the distribution $\bar{F}(t)$ and density $f(t)$ (19) are given by

$$\bar{F}(t) = \exp\left(-\int_0^t \alpha(u)\,du\right), \qquad f(t) = \alpha(t)\exp\left(-\int_0^t \alpha(u)\,du\right). \tag{35}$$

Proof. Following similar arguments for the proof of Theorem 4.14, it can be shown that $\alpha(t) = f(t)/\bar{F}(t)$, where $f(t)$ and $\bar{F}(t)$ are given by (19). □

Below we give the long-run estimate of the forward intensity (31).

Proposition 4.17. *Let* $\mathbf{T}$ ($\mathbf{\Psi T}$) *have distinct eigenvalues* $\varphi_j^{(1)}$ ($\varphi_j^{(2)}$), $j = 1, \dots, m$ *with* $\varphi_{p_k}^{(k)} = \max\{\varphi_1^{(k)}, \dots, \varphi_m^{(k)}\}$ *and* $p_k = \arg\max_j\{\varphi_j^{(k)}\}$, $k = 1, 2$. *Then, for any* $t \ge 0$ *and* $i \in E$,

$$\lim_{s\to\infty} \lambda_i(t,s) = \begin{cases} \dfrac{\mathbf{e}_i^\top \mathbf{S}_m(t) L_{p_2}(\mathbf{\Psi T})\mathbf{\Psi}\boldsymbol{\delta}}{\mathbf{e}_i^\top \mathbf{S}_m(t) L_{p_2}(\mathbf{\Psi T})\mathbf{1}_m}, & \varphi_{p_2}^{(2)} > \varphi_{p_1}^{(1)} \\ \dfrac{\mathbf{e}_i^\top (\mathbf{I}_m - \mathbf{S}_m(t)) L_{p_1}(\mathbf{T})\boldsymbol{\delta}}{\mathbf{e}_i^\top (\mathbf{I}_m - \mathbf{S}_m(t)) L_{p_1}(\mathbf{T})\mathbf{1}_m}, & \varphi_{p_2}^{(2)} < \varphi_{p_1}^{(1)} \\ \dfrac{\mathbf{e}_i^\top (\mathbf{S}_m(t) L_{p_2}(\mathbf{\Psi T})\mathbf{\Psi} + (\mathbf{I}_m - \mathbf{S}_m(t)) L_{p_1}(\mathbf{T}))\boldsymbol{\delta}}{\mathbf{e}_i^\top (\mathbf{S}_m(t) L_{p_2}(\mathbf{\Psi T}) + (\mathbf{I}_m - \mathbf{S}_m(t)) L_{p_1}(\mathbf{T}))\mathbf{1}_m}, & \varphi_{p_2}^{(2)} = \varphi_{p_1}^{(1)}, \end{cases} \tag{36}$$

where $L_{p_1}(\mathbf{T})$ *and* $L_{p_2}(\mathbf{\Psi T})$ *are the Lagrange interpolation coefficients* (13).

Proof. Since generator matrices $\mathbf{T}$ and $\mathbf{\Psi T}$ are both negative definite, the proof follows from applying similar arguments to the proof of Proposition 3.2. □

Observe that the long-term rate depends on the age (t) and the past information $\mathscr{F}_{i,t}$ of X. This feature is absence under the Markov model ($\mathbf{\Psi} = \mathbf{I}_m$).

4.7. Residual Occupation Time

The last quantity we are interested in is the expected length of time X spends in a given state, a generalization of residual lifetime expectancy. For this purpose, denote by $T_j(t)$ the total time X spends in state $j \in \mathbb{S}$ up to time t defined by

$$T_j(t) = \int_0^t \mathbf{1}_{\{X_s = j\}}\, ds.$$

This quantity appears as T_k in the likelihood function (6). Given available information $\mathscr{F}_{i,t}$, the expected length of time X spends in $j \in \mathbb{S}$ within $[t,s]$ is

$$\bar{T}_j(t,s) = \mathbb{E}\{T_j(s) - T_j(t) \mid \mathscr{F}_{i,t}\}. \tag{37}$$

Theorem 4.18. *For a given* $s \ge t \ge 0$ *and past information* $\mathscr{F}_{i,t}$, *we have*

$$\bar{T}_j(t,s) = \begin{cases} \mathbf{e}_i^\top \bar{\mathbf{F}}_{11}(t,s)\mathbf{e}_j, & \forall\, i,j \in E \\ \mathbf{e}_i^\top \bar{\mathbf{F}}_{12}(t,s)\mathbf{e}_{m+1}, & \forall\, i \in E \\ 0, & \forall\, j \in E,\ \ i = m+1 \\ s - t, & \text{for } i, j = m+1, \end{cases} \tag{38}$$

where the matrices $\bar{\mathbf{F}}_{11}(t,s)$ *and* $\bar{\mathbf{F}}_{12}(t,s)$ *are defined by*

$$\bar{\mathbf{F}}_{11}(t,s) = \mathbf{S}_m(t)[\mathbf{\Psi T}]^{-1}(e^{\mathbf{\Psi T}(s-t)} - \mathbf{I}_m) + (\mathbf{I}_m - \mathbf{S}_m(t))[\mathbf{T}]^{-1}(e^{\mathbf{T}(s-t)} - \mathbf{I}_m),$$
$$\bar{\mathbf{F}}_{12}(t,s) = \mathbf{S}_m(t)[(s-t)\mathbf{I}_m - [\mathbf{\Psi T}]^{-1}(e^{\mathbf{\Psi T}(s-t)} - \mathbf{I}_m)]\mathbf{1}_m + (\mathbf{I}_m - \mathbf{S}_m(t))[(s-t)\mathbf{I}_m - [\mathbf{T}]^{-1}(e^{\mathbf{T}(s-t)} - \mathbf{I}_m)]\mathbf{1}_m.$$

Proof. The proof follows from applying Theorem 3.4 to obtain

$$\begin{aligned} \bar{T}_j(t,s) &= \int_t^s \mathbb{P}\{X_u = j \mid \mathscr{F}_{i,t}\}\, du = \int_t^s \mathbf{e}_i^\top (\mathbf{S}_{m+1}(t) e^{\mathbf{G}(u-t)} + [\mathbf{I}_{m+1} - \mathbf{S}_{m+1}(t)] e^{\mathbf{Q}(u-t)})\mathbf{e}_j\, du \\ &= \mathbf{e}_i^\top \left(\int_t^s \begin{pmatrix} \mathbf{F}_{11}(t,u) & \mathbf{F}_{12}(t,u) \\ \mathbf{0} & 1 \end{pmatrix} du \right) \mathbf{e}_j, \end{aligned}$$

where the last equality is due to Proposition 3.7, which proves the claim. □

Below we give the expected residual lifetime $R_i(t)$ of X defined by

$$R_i(t) = \mathbb{E}\{\tau - t \mid \tau > t, \mathscr{F}_{i,t}\}.$$

Corollary 4.19. *For a given $t \geq 0$ and past information $\mathcal{F}_{i,t}$, we have*

$$R_i(t) = -\mathbf{e}_i^\top (\mathbf{S}_m(t)(\mathbf{\Psi T})^{-1} + (\mathbf{I}_m - \mathbf{S}_m(t))\mathbf{T}^{-1})\mathbf{1}_m, \tag{39}$$

for any initial state $i \in E$ of X at time t and is zero for $i = m+1$.

5. Numerical Examples

We consider marriage and divorce problem under the mixture model (4). The marriage breakup is due to divorce. We are interested in the lifetime distribution of marriage until a breakup is observed. The transient states are given by $E = \{N, M, S, W\}$, while $\Delta = \{D\}$ is the absorbing state. We denote by N for never married, M for married, W for widowed, S for separated, and D for divorced. For simplicity, we label $N = 1$, $M = 2$, $S = 3$, $W = 4$ and $D = 5$. The transition rate from never married to married is given by q_{12}, married to widowed by q_{24} and widowed to married by q_{42}. We assume that married couples may get separated after marriage which occurs at a rate of q_{23}, and eventually getting divorced at the rate of q_{35}. We also allow divorced to occur at rate q_{25} from married.

The composition of individuals in each state is denoted by initial vector $\boldsymbol{\pi}$. The corresponding intensity matrix $\mathbf{Q}$ is given by (40). Heterogeneity of individuals are described by the speed at which they move along the states. We assume that there are two groups of individuals: one moving with intensity matrix $\mathbf{T}$, while the other with $\mathbf{\Psi T}$, with $\mathbf{\Psi}$ given by (3). At initial time, the composition of these groups in each state is given by diagonal matrix $\mathbf{S}_m$ (8).

$$\mathbf{Q} = \begin{pmatrix} -q_{12} & q_{12} & 0 & 0 & 0 \\ 0 & -(q_{23}+q_{24}+q_{25}) & q_{23} & q_{24} & q_{25} \\ 0 & 0 & -(q_{34}+q_{35}) & q_{34} & q_{35} \\ 0 & q_{42} & 0 & -q_{42} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{40}$$

For the purpose of simulation, we set the following values for the mixture model. We set the elements of intensity matrix $\mathbf{T}$ by $q_{12} = 0.95$, $q_{24} = 0.05$, $q_{34} = 0.1$, $q_{23} = 0.25$, $q_{25} = 0.07$, $q_{42} = 0.85$, and $q_{35} = 0.5$. Heterogeneity of the population is described by matrix $\mathbf{S}_m = \mathrm{diag}(0.5, 0.5, 0.5, 0.5, 0.5)$, i.e., there are 50% of the population in each state moving with intensity matrix $\mathbf{T}$, while the other 50% moving at rate $\mathbf{\Psi T}$ with $\mathbf{\Psi} = \mathrm{diag}(0.25, 0.25, 0.25)$. The initial distribution of population is given by $\boldsymbol{\pi} = (0.5, 0.3, 0.1)^\top$; 50% never married, 30% married, 10% widowed and the rest is being in separated status. Based on these values, we compute and compare intensity rate $\lambda_i(t,s)$ (31), $\lambda_i(t)$ (33) and baseline intensity $\alpha(t)$ (34) of getting divorced for individuals who have been married ($\mathbf{e}_i = (0,1,0,0)^\top$) for t period of time. Heterogeneity is removed by setting $\mathbf{\Psi}$ to be the identity matrix $\mathbf{I}$. We assume that we have only the current status of marriage (limited information on $\mathcal{F}_{i,t}$). The results are presented in Figures 1 and 2. We observe in the case there is no heterogeneity in the population that the divorced intensity is always increasing in duration, regardless of the age of marriage t. A closer look at Figure 2 suggests that the forward divorced intensity $\lambda(t,d)$ (31), with $d = s - t$, is the same as the baseline intensity $\alpha(t)$. Moreover, the instantaneous divorced intensity $\lambda_i(t)$ is not affected by the marriage age t, i.e., divorced intensity $\lambda_i(t)$ is just a constant.

In contrast, for a heterogeneous population we notice from Figure 1 that the divorced intensity $\lambda(t,d)$ (31) changes its shape w.r.t the marriage age t.[1] As we see from Figure 1, those who have been married for $t = 10$ periods of time have the lowest divorced intensity and those who just got married for $t = 0.01$ periods of time have the highest divorced intensity of all. In between is the divorced intensity for those who have been married for $t = 4$ periods of time.

From these curves we notice that the newly married and those who have been married for few periods face increasing divorced intensity for the next 3.6 time periods of their marriage. Especially at the beginning, the increment is rather steep. After approximately 3.6 periods of time, the marriage is getting stable for which the divorced rate decreases to the long-term rate 0.0644 given by (36). Even though the divorced rate is relatively low for older couples, it tends to increase at a faster rate at the beginning and then gradually reaching up to the long-term rate. The instantaneous divorced intensity $\lambda(t)$ confirms the above phenomenon that the intensity tends to increase for newly married, and then decreases to a long-term rate 0.0175 (by (12) and (33)) indicating that the relationship gets more stable by the increase of marriage age t. The baseline intensity $\alpha(t)$ presented in Figure 2 also exhibits similar type of behaviour.

In terms of survival probability, we see from Figure 1 that in both models (with or without heterogeneity) the marriage survival probability decreases as the marriage duration d increases. This is due to the fact that both intensity matrices $\mathbf{T}$ and $\mathbf{\Psi T}$ are negative definite. The only main differences we observe is that the survival

Figure 1. Forward divorced intensity $\lambda_i(t,d)$ (31) and survival distribution $\bar{F}_i(t,d)$ (18) as functions of marriage duration $d = s - t$ for a given age $t > 0$. The two graphs on top correspond to the case where the population is homogeneous, i.e., $\mathbf{\Psi} = \mathbf{I}$, whereas the other two below for the case where there is heterogeneity in population; two different groups moving at different speeds, i.e., $\mathbf{\Psi} \neq \mathbf{I}$.

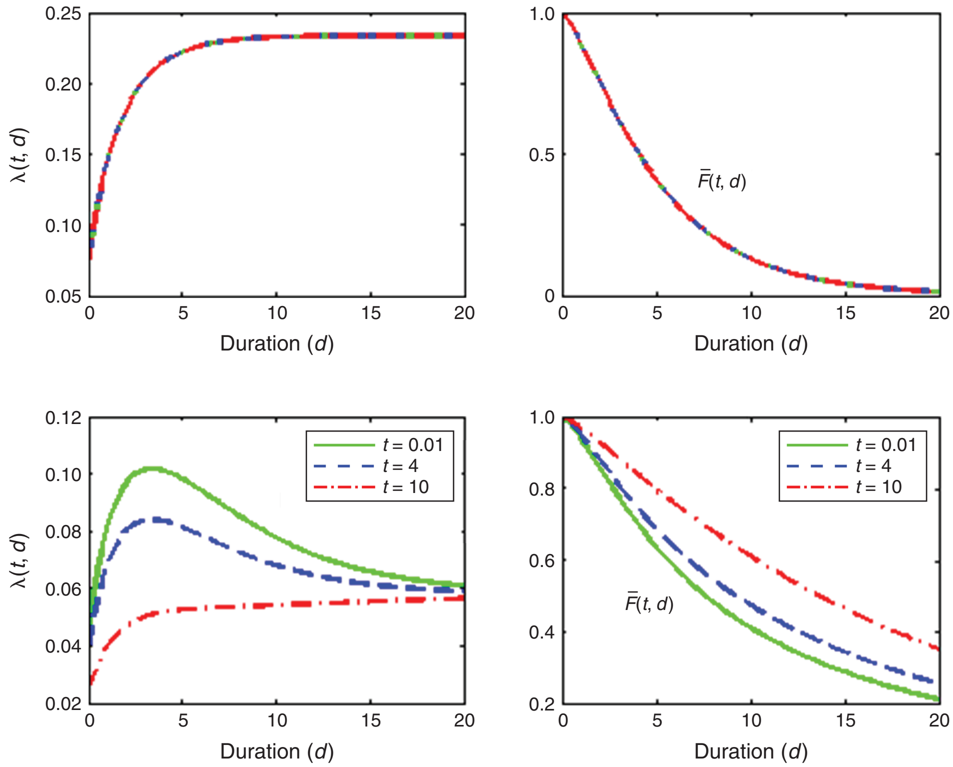


Figure 2. Intensity rate $\lambda_i(t)$ (33) and baseline function $\alpha(t)$ (34) as functions of the marriage age t. The two graphs on top correspond to the case where the population is homogeneous, i.e., $\mathbf{\Psi} = \mathbf{I}$, whereas the other two below for the case where there is heterogeneity; two groups moving at different speeds, i.e., $\mathbf{\Psi} \neq \mathbf{I}$.

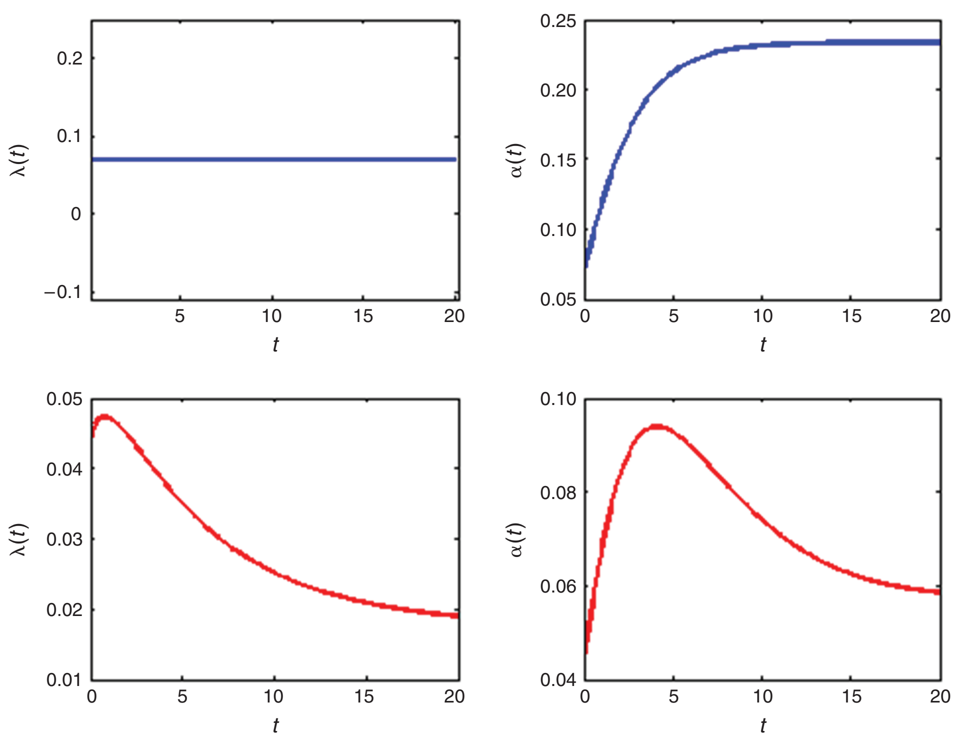
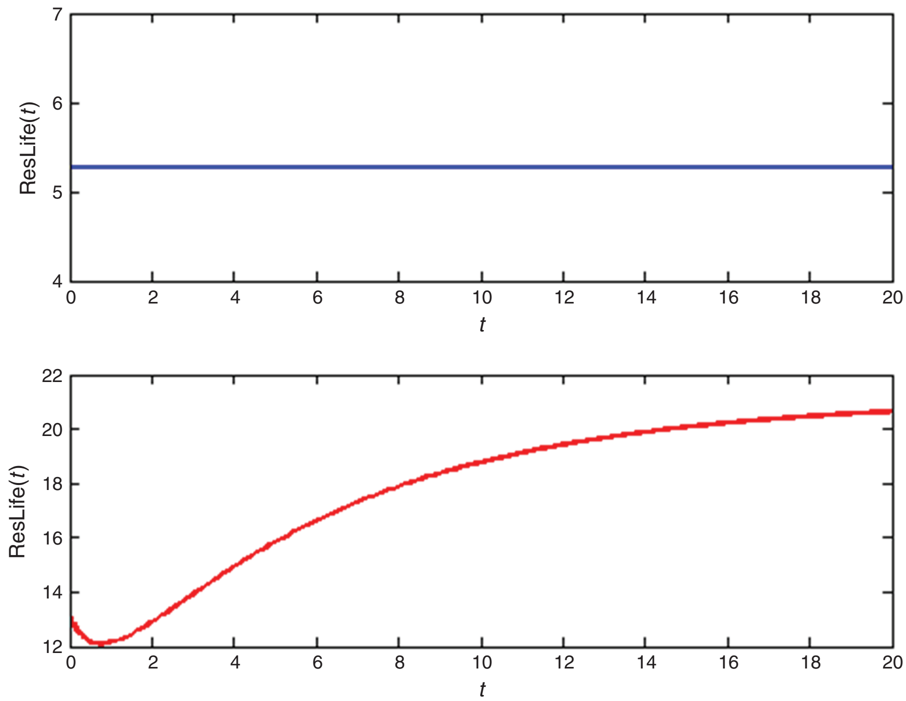


Figure 3. Residual lifetime (39) of the marriage against the marriage age t. The graph on top correspond to the case where the population is homogeneous, i.e., $\mathbf{\Psi} = \mathbf{I}$, whereas the one below is for the case where there is heterogeneity—two groups moving at different speeds, i.e., $\mathbf{\Psi} \neq \mathbf{I}$—and has U-bend shape.



probability goes to zero faster under homogeneity, as seen by the strict increasing divorced intensity, than under heterogeneity, see Figure 1. This is due to $\psi_i < 1$ for all $i = 1, \ldots, 4$. As a result, the Markov model implies that every marriage gets divorced in the long run, something which does not necessarily occur in practice. In contrast to this observation, we see under the Markov mixture model (with heterogeneity) that the survival probability does not decay to zero at the end of observation period; marriage would last in longer run with a positive probability. Moreover, we can investigate the effect of the marriage age t to the marriage survival probability. We observe that for the same marriage duration, the survival probability increases by the age, i.e., the older the age of marriage, the higher the probability that the marriage would survive.

Even though the above observations are based on a numerical study on married-and-divorced model, the results from the mixture model are able to provide sound explanations to the data depicted in Figure 5.3 in Aalen et al. (2008, p. 222).

Figure 3 depicts expected residual lifetime of marriage. We observe that the residual lifetime under homogeneity (the Markov model) is not affected by the age of marriage t, i.e., it is just flat. In contrast to this observation, we notice that the residual lifetime exhibits a U-bend shape curve under heterogeneity (the mixture model), i.e., the residual lifetime is decreasing for newly married, and then increasing as the age of marriage gets older. These numerical figures reveal the U-bend shape of lifetime discussed in the article from *The Economist* (2010).

6. Conclusions

We have extended the Markov mixture model Frydman (2005) and Frydman and Schuermann (2008) to the mixture of two continuous-time Markov chains with absorption on the same state space moving at different speeds. From the viewpoint of credit risk applications (Frydman and Schuermann 2008), the Markov mixture process represents the credit rating dynamics of defaultable bonds, whereby bonds of the same credit rating can move at different speeds to other credit ratings. In general, the mixture process does not have the Markov property and thus, given the current state, its future development does depend on its past history. We discussed variety of associated distributional properties of the process. Identities concerning the transition matrix, the lifetime distribution and the forward intensity are explicit in terms of the Bayesian updates of switching probability and the intensity matrices of the underlying Markov chains.

The identities form nonstationary functions of time and duration and have the ability to capture heterogeneity and path dependence when conditioning on the currently available information (either full or partial) about the past history of the mixture process. The unconditional lifetime distribution forms a generalized mixture of phase-type distributions. Using matrix analytic approach (Rolski et al. 1998 and Tijms 1994), we showed the dense and closure properties under finite convex mixtures and finite convolutions of the distribution. When the underlying Markov chains move at the same speed, the nonstationarity, heterogeneity and path dependence are removed from the identities, whereas the lifetime has the usual phase-type distribution (Neuts 1975).

The results presented in this paper can be extended to a mixture of finite number of continuous-time absorbing Markov chains. Given their availability in explicit forms and the ability to capture path dependence and heterogeneity, we believe that the results should be able to offer appealing features for applications.

Acknowledgments
The author would like to thank a number of anonymous referees and associate editors for their useful suggestions and comments that improved the presentation of this paper. The author acknowledges the support and hospitality provided by the Center for Applied Probability of Columbia University during his academic visit in May 2014 where he saw the work of Frydman and Schuermann (2008) and started to work on the problem. He thanks Professor Karl Sigman for the invitation.

Endnote
[1] It exhibits similar type of behavior to Figure 9 in Aalen (1995) for different initial states.

References
Aalen OO (1995) Phase type distributions in survival analysis. *Scand. J. Stat.* 22(4):447–463.
Aalen OO, Gjessing HK (2001) Understanding the shape of the hazard rate: A process point of view. *Stat. Sci.* 16(1):1–22.
Aalen OO, Borgan Ø, Gjessing HK (2008) *Survival and Event History Analysis: A Process Point of View* (Springer, New York).
Albrecher H, Asmussen S (2010) *Ruin Probabilities*, 2nd ed. (World Scientific, Singapore).
Apostol TM (1969) Explicit formulas for the exponential matrix e^{tA}. *Am. Math. Mon.* 76(3):289–292.
Asmussen S (2003) *Applied Probability and Queues*, 2nd ed. (Springer, New York).
Asmussen S, Avram F, Pistorius MR (2004) Russian and American put options under exponential phase-type Lévy models. *Stoch. Proc. Appl.* 109(1):79–111.
Assaf D, Levikson B (1982) Closure of phase type distributions under operations arising in reliability theory. *Ann. Probab.* 10(1):265–269.
Badila ES, Boxma OJ, Resing JAC (2014) Document queues and risk processes with dependencies. *Stoch. Models* 30(3):390–419.
Blumen I, Kogan M, McCarthy PJ (1955) The industrial mobility of labor as a probability process. *Cornell Stud. Ind. Labor Relat.*, Vol. 6 (Cornell University Press, Ithaca, NY).
Breuer L, Baum D (2005) *An Introduction to Queueing Theory and Matrix-Analytic Methods* (Springer, Dordrecht, Netherlands).
Buchholz P, Kriege J, Felko I (2014) *Input Modeling with Phase-Type Distributions and Markov Models: Theory and Applications* (Springer International, Cham, Switzerland).
Chakravarthy SR, Neuts MF (2014) Analysis of a multi-server queueing model with MAP arrivals of regular customers and phase type arrivals of special customers. *Simul. Model. Pract. Th.* 43:79–95.
Cheng H-W, Yau S-T (1997) More explicit formulas for exponential matrix. *Linear Algebra Appl.* 262:131–163.
Crowder M (2001) *Classical Competing Risks* (Chapman & Hall, Boca Raton, FL).
Duan JC, Sun J, Wang T (2012) Multiperiod corporate default prediction—A forward intensity approach. *J. Econometrics* 170(1):191–209.
Economist, The (2010) Age and happiness: The U-bend of life. *The Economist*, (December 16th), http://www.economist.com/node/17722567.
Frydman H (1984) Maximum likelihood estimation in the mover-stayer model. *J. Am. Stat. Assoc.* 79(387):632–638.
Frydman H (2005) Estimation in the mixture of Markov chains moving with different speeds. *J. Am. Stat. Assoc.* 100(471):1046–1053.
Frydman H, Schuermann T (2008) Credit rating dynamics and Markov mixture models. *J. Bank. Financ.* 32(6):1062–1075.
Jarrow RA, Lando D, Turnbull SM (1997) A Markov model for the term structure of credit risk spreads. *Rev. Financ. Stud.* 10(2):481–523.
Kalbfleisch JD, Prentice RL (2002) *The Statistical Analysis of Failure Time Data*, 2nd ed. (John Wiley & Sons, Hoboken, NJ).
Klugman SA, Panjer HH, Willmot GE (2012) *Loss Models: From Data to Decisions*, Vol. 715 (John Wiley & Sons, Hoboken, NJ).
Lee SCK, Lin XS (2010) Modeling and evaluating insurance losses via mixtures of Erlang distributions. *N. Am. Actuar. J.* 14(1):107–130.
Lin XS, Liu X (2007) Markov aging process and phase-type law of mortality. *N. Am. Actuar. J.* 11(4):92–109.
Neuts MF (1975) Probability distributions of phase-type. *Liber Amicorum Prof. Emiritus H. Florin* (University of Louvain, Belgium), 173–206.
Neuts MF (1981) *Matrix-Geometric Solutions in Stochastic Models* (Johns Hopkins University Press, Baltimore).
O'Cinneide C (1989) On non-uniqueness of representations of phase-type distributions. *Commun. Statist.-Stochastic Models* 5(2):247–259.
O'Cinneide C (1990) Characterization of phase-type distributions. *Commun. Statist.-Stochastic Models* 6(1):1–57.
Okamura H, Dohi T (2015) Phase-type software reliability model: Parameter estimation algorithms with grouped data. *Ann. Oper. Res.* 244(1):177–208.
Rolski T, Schmidli H, Schmidt V, Teugels J (1998) *Stochastic Processes for Insurance and Finance* (John Willey & Sons, Chichester, UK).
Tijms H (1994) *Stochastic Models: An Algorithm Approach* (John Wiley & Sons, Chichester, UK).