On the conditional joint probability distributions of phase-type under the mixture of finite-state absorbing Markov jump processes

B.A. Surya*
School of Mathematics and Statistics
Victoria University of Wellington, New Zealand
12 May 2018

Abstract

This paper presents some new results on the conditional joint probability distributions of phase-type under the mixture of right-continuous Markov jump processes with absorption on the same finite state space $S$ moving at different speeds, where the mixture occurs at a random time. Such mixture was first proposed by Frydman [19] and Frydman and Schuermann [18] as a generalization of the mover-stayer model of Blumen et al. [16], and was recently extended by Surya [33]. When conditioning on all previous and current information $\mathcal{F}_{t,i} = \mathcal{F}_t \cup \{X_t = i\}$, with $\mathcal{F}_t = \{X_s, 0 < s \leq t\}$ and $i \in S$, of the mixture process $X$, distributional identities are explicit in terms of the Bayesian updates of switching probability, the likelihoods of observing the sample paths, and the intensity matrices of the underlying Markov processes, despite the fact that the mixture itself is non-Markov. They form non-stationary function of time and have the ability to capture heterogeneity and path dependence. When the underlying processes move at the same speed, in which case the mixture reduces to a simple Markov jump process, these features are removed, and the distributions coincide with that of given by Neuts [28] and Assaf et al. [8]. Furthermore, when conditioning on $\mathcal{F}_{t-}$ and no exit to the absorbing set has been observed at time $t$, the distributions are given explicitly in terms of an additional Bayesian updates of probability distribution of $X$ on $S$. Examples are given to illustrate the main results.

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

Keywords: Markov jump processes, mixture of Markov jump processes, conditional multivariate phase-type distributions, competing risks

*School of Mathematics and Statistics, Victoria University of Wellington, Gate 6 Kelburn PDE, Wellington 6140, New Zealand. Email address: budhi.surya@vuw.ac.nz
1 Introduction

Markov chain has been one among the most important probabilistic tools in modeling complex stochastic systems evolutions. It has been widely used in variety of applications across various fields such as, among others, in ecology ([11]), finance ([24], and Jarrow et al. [23]), marketing ([13] and Pfeifer and Carraway [29]), etc. The phase-type model describes the lifetime distribution of an absorbing Markov chain. It was first introduced in univariate form by Neuts [28] in 1975 as the generalization of Erlang distribution. It has dense property, which can approximate any distribution of positive random variables arbitrarily well, and has closure property under finite convex mixtures and convolutions. When jumps distribution of compound Poisson process is modelled by phase-type model, it results in a dense class of Lévy processes, see Asmussen [5]. The advantage of working under phase-type distribution is that it allows some analytically tractable results in applications. To mention some, in option pricing (Asmussen et al. [4]), actuarial science (Albrecher and Asmussen [6], Rolski et al. [31], Zadeh et al. [36], in survival analysis (Aalen [2], Aalen and Gjessing [1]), in queueing theory (Chakravarthy and Neuts [17], Asmussen [5]), in reliability theory (Assaf and Levikson [9], Okamura and Dohi [34]).

The phase-type distribution $F$ is expressed in terms of a Markov jump process $\{X_t\}_{t \geq 0}$ with finite state space $S = E \cup \{\Delta\}$, where for some integer $m \geq 1$, $E = \{i : i = 1, ..., m\}$ and $\Delta$ represent respectively the transient and absorbing states. The lifetime of the Markov process and its distribution are defined by $\tau = \inf \{t \geq 0 : X_t = \Delta\}$ and $F(t) = \mathbb{P}\{\tau > t\}$. (1.1)

In view of credit risk applications, the state space $S$ represents the possible credit classes, with 1 being the highest (Aaa in Moody’s rankings) and $m$ being the lowest (C in Moody’s rankings), whilst the absorbing state $\Delta$ represents bankruptcy, D. The distribution $\pi_k$ represents the proportion of homogeneous bonds in the rating $k$. We refer to [24] and [23] and literature therein for details.

Unless stated otherwise, we assume for simplicity that the initial probability $\pi$ of starting in any of the $m+1$ phases has zero mass on the absorbing state $\Delta$, i.e., $\pi_{\Delta} = 0$, so that $\mathbb{P}\{\tau > 0\} = 1$. We also refer to $\Delta$ as the $(m + 1)$th element of the state space $S$, i.e., $\Delta = m + 1$. The speed at which the Markov process moves along the state space $S$ is described by an intensity matrix $Q$. This matrix has block partition according to the process moving in the transient state $E$ and in the absorbing state $\Delta$, which admits the following block-partitioned form:

$$Q = \begin{pmatrix} A & -A1 \\ 0 & 0 \end{pmatrix},$$

with $1 = (1, ..., 1)^{\top}$, as the rows of the intensity matrix $Q$ sums to zero. That is to say that the entry $q_{ij}$ of the matrix $Q$ satisfies the following properties:

$$q_{ii} \leq 0, \quad q_{ij} \geq 0, \quad \sum_{j \neq i} q_{ij} = -q_{ii} = q_{i}, \quad (i, j) \in S.$$ (1.3)
As $-\mathbf{A}$ is a non-negative vector, (1.3) implies that $\mathbf{A}$ to be a negative definite matrix, i.e., $\mathbf{1}^\top \mathbf{A} \mathbf{1} < 0$. The matrix $\mathbf{A}$ is known as the phase generator matrix of $\mathbf{Q}$. The absorption is certain if and only if $\mathbf{A}$ is nonsingular, see Neuts [27].

Following Theorem 3.4 and Corollary 3.5 in [5] and by the homogeneity of $X$, the transition probability matrix $\mathbf{P}(t)$ of $X$ over the period of time $(0, t)$ is

$$\mathbf{P}(t) = \exp(\mathbf{Q} t), \quad t \geq 0. \quad (1.4)$$

The entry $q_{ij}$ has probabilistic interpretation: $1/(-q_{ii})$ is the expected length of time that $X$ remains in state $i \in E$, and $q_{ij}/q_{ii}$ is the probability that when a transition out of state $i$ occurs, it is to state $j \in S$, $i \neq j$. The representation of the distribution $\mathbf{F}$ is uniquely specified by $(\mathbf{\pi}, \mathbf{A})$. We refer among others to Neuts [27] and Asmussen [5] for details. Following [27] and Proposition 4.1 [5],

$$\mathbf{F}(t) = \mathbf{\pi}^\top \mathbf{e}^{\mathbf{A} t} \mathbf{1} \quad \text{and} \quad f(t) = -\mathbf{\pi}^\top \mathbf{e}^{\mathbf{A} t} \mathbf{A} \mathbf{1}. \quad (1.5)$$

The extension of (1.5) to multivariate form was proposed by Assaf et al. [8] and later by Kulkarni [25]. Following [8], let $\Gamma_1, \ldots, \Gamma_n$ be nonempty stochastically closed subsets of $S$ such that $\bigcap_{k=1}^n \Gamma_k$ is a proper subset of $S$. ($\Gamma_1 \subset S$ is said to be stochastically closed if once $X$ enters $\Gamma_1$, it never leaves.) We assume without loss of generality that $\bigcap_{k=1}^n \Gamma_k$ consists of only the absorbing state $\Delta$, i.e., $\bigcap_{k=1}^n \Gamma_k = \Delta$.

Since $\Gamma_k$ is stochastically closed, necessarily $q_{ij} = 0$ if $i \in \Gamma_k$ and $j \in \Gamma_k^c$.

The first time until absorption of $X$ in the set $\Gamma_k$ is defined by

$$\tau_k := \inf\{t \geq 0 : X_t \in \Gamma_k\}. \quad (1.6)$$

The joint distribution $\overline{F}$ of $\{\tau_k\}$ is called the multivariate phase type distribution, see [8]. Let $t_{i_n} \geq \cdots \geq t_{i_1} \geq 0$ be the ordering of $(t_1, \ldots, t_n) \in \mathbb{R}_+^n$. Following [8],

$$\overline{F}(t_1, \ldots, t_n) = \mathbb{P}\{\tau_1 > t_1, \ldots, \tau_n > t_n\} = \mathbf{\pi}^\top \left( \prod_{k=1}^n \exp \left( \mathbf{A}(t_{i_k} - t_{i_{k-1}}) \right) \right) \mathbf{H}_k \mathbf{1}, \quad (1.7)$$

where $\mathbf{H}_k$ is $(m \times m)$ diagonal matrix whose $i$th diagonal element, for $i = 1, \ldots, m$, equals 1 when $i \in \Gamma_k^c$ and is zero otherwise. Again, as before we assume $\mathbf{\pi}$ has zero mass on $\Delta$ and $\pi_i \neq 0$ for $i \in \bigcup_{k=1}^n \Gamma_k^c$ implying that $\mathbb{P}\{\tau_1 > 0, \ldots, \tau_n > 0\} = 1$.

The multivariate distribution (1.7) has found various applications, e.g., in modeling credit default contagion (Herbertsson [22], Bielecki et al. [14]), in modeling aggregate loss distribution in insurance (Berdel and Hipp [12], Asimit and Jones [3] and Willmot and Woo [35]), and in Queueing theory (Badila et al. [10]).

Due to spatial homogeneity of the underlying Markov process, the distributions (1.5) & (1.7) have stationary property and have therefore no ability to capture heterogeneity and available information of its past. In the recent empirical works of Frydman and Schuerermann [18], it was found that bonds of the same credit rating, represented by the state space of the Markov process, can move at different speeds to other ratings. Furthermore, the inclusion of past credit ratings
B.A. Surya improves out-of-sample prediction of the Nelson-Aalen estimate of credit default intensity. These empirical findings suggest that the dynamics of credit rating can be represented by a mixture $X$ of Markov jump processes moving with different speeds (intensity matrices), where the mixture itself is non-Markov. In his recent work, Surya [33] extended the mixture model [18] and gave explicit distributional identities of the mixture process. However, the analyses performed in [18] and [33] were based on knowing the initial and current states of the process. We extend the results by relaxing this assumption. For this purpose, we define by $G_t = F_t - \{X_t \neq \Delta\}$ the set of all previous information and knowing that there is no exit to absorbing set $\{\Delta\}$ has been observed at time $t$, i.e., $G_t = \bigcup_{i \in E} F_{t,i}$. Conditional on $F_{t,i}$ and $G_t$, we derive explicit formula for the joint distributions

$$F_i(t_1, \ldots, t_n) = \mathbb{P}\{\tau_1 > t_1, \ldots, \tau_n > t_n | F_{t,i}\}$$

under the mixture process $X$, with $1 \leq n \in \mathbb{N}$, $i \in S$ and $0 \leq t \leq \min\{t_n, \ldots, t_1\}$.

In view of credit risk applications [15], the quantity $F_i(t_1, \ldots, t_n)$ describes the joint distribution of exit times $\{\tau_k\}$ (1.6), due to cause-specific exits (default, prepayment, calling back, etc), of $i$-rated bonds, conditional on the credit rating history up to the current time $t$, whilst the function $F_t(t_1, \ldots, t_n)$ determines the joint distribution of the bonds’ exit times across credit ratings viewed at the time $t$. In the framework of competing risks (see for instance Pintilie [30]), for the observed exit time $\tau := \min\{\tau_1, \ldots, \tau_n\}$ and reason of exit $\xi = \arg\min\{\tau_1, \ldots, \tau_n\}$, the probability $\mathbb{P}\{t \leq \tau < s, \xi = 1 | F_{t,i}\}$ determines the proportion of $i$-rated bonds exiting by type 1 from the pool within $s-t$ period of time, whilst $\mathbb{P}\{t \leq \tau < s, \xi = 1 | G_t\}$ represents the percentage of bonds exiting by type 1.

The organization of this paper is as follows. We discuss in Section 2 the Markov mixture process in details. Section 3 presents some preliminaries, which extend further the results in [18] and [33]. The main contributions of this paper are presented in Section 4. Some explicit examples of the results are discussed in Section 5, in which we show that the exit times $\{\tau_k\}$ are independent under the Markov model, but not for the mixture model. Section 6 concludes this paper.

### 2 Mixture of Markov jump processes

Throughout the remaining of this paper we denote by $X = \{X_t^{(\phi)}, t \geq 0\}$ the mixture process, which is a continuous-time stochastic process defined as a mixture of two Markov jump processes $X^{(\phi)} = \{X_t^{(\phi)}, t \geq 0\}$ and $X^{(1)} = \{X_t^{(1)}, t \geq 0\}$, whose intensity matrices are given respectively by $Q$ and $G$. We assume that the underlying processes $X^{(0)}$ and $X^{(1)}$ are right-continuous. The two processes are defined on the same finite state space $S$. It is defined following [33] by

$$X = \begin{cases} X^{(0)}, & \phi = 0 \\ X^{(1)}, & \phi = 1, \end{cases}$$

(2.1)
On the conditional joint probability distributions of phase-type

where the variable $\phi$ represents the speed regimes, assumed to be unobservable.

Markov mixture process is a generalization of mover-stayer model, a mixture of two discrete-time Markov chains proposed by Blumen et al. [16] in 1955 to model population heterogeneity in jobs labor market. In the mover-stayer model [16], the population of workers consists of stayers (workers who always stay in the same job category, $Q = 0$) and movers (workers who move according to a stationary Markov chain with intensity matrix $G$). Estimation of the mover-stayer model was discussed in Frydman [20]. Frydman [19] extended the model to a mixture of two continuous-time Markov chains moving with intensity matrices $Q$ and $G = \Psi Q$, where $\Psi$ is a diagonal matrix. Frydman and Schuermann [18] later used the result to model the dynamics of firms’ credit ratings. As empirically shown in [18], there is strong evidence to suggest that firms of the same credit rating can move at different speeds to other credit ratings, a feature that lacks in the Markov model. Further distributional properties and identities of the mixture process were given in [33], in particular in the presence of absorbing states.

For a given initial state $i_0$, there is a separate mixing distribution defined by

$$s_{i_0} = \mathbb{P}\{\phi = 1|X_0 = i_0\} \quad \text{and} \quad 1 - s_{i_0} = \mathbb{P}\{\phi = 0|X_0 = i_0\}, \quad (2.2)$$

with $0 \leq s_{i_0} \leq 1$. The quantity $s_{i_0}$ has the interpretation as the portion of firms with initial rating $i_0$ that evolve according to the process $X^{(1)}$, whilst $1 - s_{i_0}$ is the proportion that propagates according to $X^{(0)}$. In general, $X^{(0)}$ and $X^{(1)}$ have different expected length of time the process occupies a state $i$, i.e., $1/q_i \neq 1/g_i$, but under [18] both processes have the same probability of leaving the state $i \in E$ to state $j \in S$, $i \neq j$, i.e. $q_{ij}/q_i = g_{ij}/g_i$. Note that we have used $g_i$ and $q_{ij}$ to denote negative of the $i$th diagonal element and the $(i, j)$ entry of $G$, respectively. Thus, depending on whether $0 = \psi_i := [\Psi]_{i,i}, 0 < \psi_i < 1, \psi_i > 1$ or...
$\psi_i = 1$, $X^{(1)}$ never moves out of state $i$ (the mover-stayer model), moves out of state $i$ at lower rate, higher rate or at the same rate, subsequently, than that of $X^{(0)}$. If $\psi_i = 1$, for all $i \in S$, $X$ reduces to a simple Markov jump process $X^{(0)}$.

Figure 1 illustrates the transition of $X^{(2.1)}$ between states $J_1$ and $J_2$. When $X$ is observed in state $J_1$, it would stay in the state for an exponential period of time with intensity $q_{j_1}$ or $g_{j_1}$ before moving to state $J_2$ with probability $q_{j_1,j_2}/q_{j_1}$ or $g_{j_1,j_2}/g_{j_1}$ depending on whether it is driven by the Markov chain $X^{(0)}$ or $X^{(1)}$.

The main feature of the mixture process $X^{(2.1)}$ is that unlike its component $X^{(0)}$ and $X^{(1)}$, $X$ does not have the Markov property; future development of its state depends on its past information. The section below discusses this further.

3 Preliminaries

Recall that the process $X^{(2.1)}$ repeatedly changes its speed randomly in time either at the rate $Q$ or $G$. The speed regime, which is represented by the variable $\phi$, is however not directly observable; we can not classify from which regime the observed process $X$ came from. But, it can be identified based on past realizations of the process. We have denoted by $F_{t,i}$ all previous information about $X$ prior to time $t > 0$, and by $F_{t,i} = F_{t-} \cup \{X_t = i\}$, $i \in S$. The set $F_{t-}$ may contain full observation, partial information or maybe nothing about the past of $X$.

The likelihoods of observing the past realization $F_{t,j}$ of $X$ under $X^{(1)}$ and $X^{(0)}$ conditional on knowing its initial state $i$ are defined respectively by

$$L_i^{Q}(t) := P\{F_{t,j} | \phi = 0, X_0 = i\} = \prod_{k \in S} e^{-q_{k}T_k} \prod_{j \neq k, j \in S} (g_{kj})^{N_{kj}},$$

$$L_i^{G}(t) := P\{F_{t,j} | \phi = 1, X_0 = i\} = \prod_{k \in S} e^{-q_{k}T_k} \prod_{j \neq k, j \in S} (g_{kj})^{N_{kj}},$$

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

3.1 Bayesian updates of switching probability

The Bayesian updates of switching probability $s_j(t)$ of $X^{(2.1)}$ is defined by

$$s_j(t) = P\{\phi = 1 | F_{t,j}\}, \quad j \in S, \quad t \geq 0. \quad (3.2)$$

It represents the proportion of those in state $j$ moving according to Markov process $X^{(1)}$. Note that $s_j(0) = s_j$. Denote by $S(t)$ a diagonal matrix defined by

$$\bar{S}(t) = \begin{pmatrix} S(t) & 0 \\ 0 & S_{m+1}(t) \end{pmatrix}, \quad (3.3)$$

with $S(t) = \text{diag}(s_1(t), s_2(t), ..., s_m(t))$, representing switching probability matrix of $X$. For $t = 0$, in which case $F_{t,j} = \{X_0 = j\}$, we write $\bar{S} := \bar{S}(0), S := S(0)$. 
Depending on the availability of the past information of \( X \), the elements \( s_j(t), j \in \mathbb{S} \), of the switching probability matrix \( \bar{S}(t) \) [13] is given below.

**Proposition 3.1** Let \( \pi \) be the initial probability distribution of \( X \) [2.1] on the state space \( \mathbb{S} \). Define by \( L^Q(t) \) and \( L^G(t) \) the likelihood matrices whose \((i, j)\) elements \([L^G(t)]_{i,j}\) and \([L^Q(t)]_{i,j}\) are given by (3.7). Then, for \( j \in \mathbb{S} \) and \( t \geq 0 \),

\[
s_j(t) = \frac{\pi^T \bar{S} L^G(t) e_j}{\pi^T [\bar{S} L^G(t) + (I - \bar{S}) L^Q(t)] e_j}.
\] (3.4)

To be more precise, depending on availability of information set \( \mathcal{F}_{t-} \), we have

(i) Under full observation \( \mathcal{F}_{t-} = \{X_s, 0 < s \leq t-\} \) and conditional on knowing the initial state \( i_0 \), i.e., \( \pi = e_{i_0} \), \( s_j(t) \) simplifies further to

\[
s_j(t) = \frac{s_{i_0} L^G_{i_0,j}(t)}{s_{i_0} L^G_{i_0,j}(t) + (1 - s_{i_0}) L^Q_{i_0,j}(t)}.
\]

(ii) In case of \( \mathcal{F}_{t-} = \emptyset \), \( L^Q(t) = e^{Qt} \) and \( L^G(t) = e^{Gt} \), and therefore

\[
s_j(t) = \frac{\pi^T \bar{S} e^{Gt} e_j}{\pi^T [\bar{S} e^{Gt} + (I - \bar{S}) e^{Qt}] e_j}.
\]

(iii) In case of \( \mathcal{F}_{t-} = \emptyset \) and conditional on knowing the initial state \( i_0 \),

\[
s_j(t) = \frac{e_{i_0}^T \bar{S} e^{Gt} e_j}{e_{i_0}^T [\bar{S} e^{Gt} + (I - \bar{S}) e^{Qt}] e_j}.
\]

The expression (3.4) for \( s_j(t) \) generalizes the result of Lemma 3.1 in [33]. It follows from (3.4) that when the underlying Markov processes \( X^{(1)} \) and \( X^{(0)} \) move at the same speed, i.e., \( G = Q \), we see that \( s_j(t) = 1 \) for all \( j \in \mathbb{S} \) and \( t \geq 0 \), implying that the observed process \( X \) is driven by the Markov chain \( X^{(1)} \).

**Proof** [Proposition 3.1] By the law of total probability and the Bayes’ formula,

\[
P\{\mathcal{F}_{t,j}, \phi = 1\} = \sum_{i \in \mathbb{S}} P\{X_0 = i\} P\{\phi = 1|X_0 = i\} P\{\mathcal{F}_{t,j}|\phi = 1, X_0 = i\}
\]

\[
= \sum_{i \in \mathbb{S}} \pi_i \times s_i \times L^G_{i,j}(t) = \pi^T \bar{S} L^G(t) e_j.
\]

Similarly, one will obtain after applying the same method as above that

\[
P\{\mathcal{F}_{t,j}, \phi = 0\} = \pi^T (I - \bar{S}) L^Q(t) e_j.
\]

The claim in (3.4) is finally established on account of the Bayes’ formula:

\[
s_j(t) = \frac{P\{\mathcal{F}_{t,j}, \phi = 1\}}{P\{\mathcal{F}_{t,j}, \phi = 1\} + P\{\mathcal{F}_{t,j}, \phi = 0\}}. \quad \square
\]
3.2 Conditional transition probability matrix

The following result on the transition probability matrix of $X$ plays an important role in deriving explicit identities for the joint probability distributions $P(t,s)$.

**Theorem 3.2 (Surya [33])** For any $s \geq t \geq 0$, the conditional transition probability matrix $P(t,s)_{i,j} := P\{X_s = j | F_{t,i}\}$ of the mixture process $X$ is given by

$$P(t,s) = \overline{S}(t)e^{G(s-t)} + \left[I - \overline{S}(t)\right]e^{Q(s-t)}. \quad (3.5)$$

It is clear from (3.5) that $X$ does not inherit the Markov property of the underlying processes $X'(t)$ and $X'(0)$, i.e., future development of $X$ is determined by its past information $F_{t,i}$ through its likelihoods. To be more precise, when we set $G = Q$ in (3.5), $P(t,s) = e^{Q(s-t)}$, i.e., $X$ is a simple Markov jump process.

4 Probability distributions of phase-type

This section presents the main results of this paper on the joint probability distributions of lifetime $\tau_k$ under the mixture process $X$ (2.1), conditional on the information set $F_{t,i}$ and $G_t$. We assume that $X$ is defined on the finite state space $S = E \cup \{\Delta\}$, where $E = \{1, 2, ..., m\}$ and $\Delta$ are transient and absorbing states, subsequently. We first discuss univariate $G_t$-conditional distribution $\tau$ (1.1) of $X$. To motivate and illustrate the main results on the multivariate distributions (1.8), we consider the bivariate case in some details. Throughout the remaining of this paper, define intensity matrices $G$ and $Q$ respectively by

$$G = \begin{pmatrix} B & -B \mathbb{1} \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} A & -A \mathbb{1} \\ 0 & 0 \end{pmatrix}. \quad (4.1)$$

The following results on block partition of the transition probability matrix $P(t,s)$ (3.5) and exponential matrix will be used to derive the conditional distributions (1.8) in closed form. We refer to Proposition 3.7 in [33] for details.

**Lemma 4.1** Let the phase generator matrix $A$ be nonsingular. Then,

$$e^Q = \begin{pmatrix} e^A & \mathbb{1} - e^A \mathbb{1} \\ 0 & 1 \end{pmatrix}. \quad (4.2)$$

**Proposition 4.2** The transition probability matrix (3.5) has block partition:

$$P(t,s) = \begin{pmatrix} P_{11}(t,s) & P_{12}(t,s) \\ 0 & 1 \end{pmatrix}, \quad (4.3)$$

where the matrix entries $P_{11}(t,s)$ and $P_{12}(t,s)$ are respectively defined by

$$P_{11}(t,s) = S(t)e^{B(s-t)} + \left[I - S(t)\right]e^{A(s-t)}$$

$$P_{12}(t,s) = S(t)\left[I - e^{B(s-t)}\right]\mathbb{1} + \left[I - S(t)\right]\left(I - e^{A(s-t)}\right)\mathbb{1}.$$

Furthermore, in the sequel below we denote by $\pi(t)$ the time-$t$ probability distribution of $X$ on the state space $S$, whose $i$th element $\pi_i(t)$ is defined by

$$\pi_i(t) = P\{X_t = i | G_t\}, \quad \text{for} \ i \in E, \quad \text{and} \quad \pi_i(t) = 0, \quad \text{for} \ i \in \Delta. \quad (4.4)$$
4.1 Bayesian updates of probability distribution $\pi$

The following proposition gives the distribution $\pi(t)$ of $X$ on $E$ at time $t \geq 0$.

**Proposition 4.3** For a given $\pi_0$, whereas the cases $(\text{ii})$ and $(\text{iii})$ follow on account of (4.2) and that $\pi_\Delta = 0$. □

$$\pi_j(t) = \frac{\pi^\top \{ SL^G(t) + (I - S)L^Q(t) \} e_j}{\pi^\top \{ SL^G(t) + (I - S)L^Q(t) \} 1}. \quad (4.5)$$

To be more precise, depending on availability of information set $\mathcal{F}_{t-}$, we have

(i) Under full observation $\mathcal{F}_{t-} = \{ X_s, 0 < s \leq t- \}$ and conditional on knowing the initial state $i_0$, i.e., $\pi = \pi_{i_0}$, $\pi_j(t)$ simplifies further to

$$\pi_j(t) = \frac{s_{i_0}L^G_{i_0,j}(t) + (1 - s_{i_0})L^Q_{i_0,j}(t)}{\sum_{j \in E} \left( s_{i_0}L^G_{i_0,j}(t) + (1 - s_{i_0})L^Q_{i_0,j}(t) \right)}.$$

(ii) In case of $\mathcal{F}_{t-} = \emptyset$, $L^Q(t) = e^Qt$ and $L^G(t) = e^Gt$, and therefore

$$\pi_j(t) = \frac{\pi^\top \{ Se^Bt + (I - S)e^{A^t} \} e_j}{\pi^\top \{ Se^Bt + (I - S)e^{A^t} \} 1}.$$

(iii) In case of $\mathcal{F}_{t-} = \emptyset$ and conditional on knowing the initial state $i_0$,

$$\pi_j(t) = \frac{e_j^\top \{ Se^Bt + (I - S)e^{A^t} \} e_j}{e_j^\top \{ Se^Bt + (I - S)e^{A^t} \} 1}.$$

It follows that $0 < \pi_j(t) < 1$, $\sum_{j \in E} \pi_j(t) = 1$ for $t \geq 0$, and $\pi = \pi(0)$.

**Proof** The proof follows from applying the law of total probability and the Bayes’ formula for conditional probability. By applying the latter, we have

$$\mathbb{P}\{\mathcal{F}_{t,j}, \phi = 1, X_0 = i\} = \mathbb{P}\{X_0 = i\} \mathbb{P}\{\phi = 1|X_0 = i\} \mathbb{P}\{\mathcal{F}_{t,j}|\phi = 1, X_0 = i\}$$

$$= \pi_i \times s_i \times L^G_{i,j}(t).$$

By the same approach, $\mathbb{P}\{\mathcal{F}_{t,j}, \phi = 0, X_0 = i\} = \pi_i \times (1 - s_i) \times L^Q_{i,j}(t)$. Hence,

$$\mathbb{P}\{\mathcal{F}_{t,j}, X_0 = i\} = \mathbb{P}\{\mathcal{F}_{t,j}, \phi = 1, X_0 = i\} + \mathbb{P}\{\mathcal{F}_{t,j}, \phi = 0, X_0 = i\}$$

$$= \pi_i \times s_i \times L^G_{i,j}(t) + \pi_i \times (1 - s_i) \times L^Q_{i,j}(t).$$

Therefore, we have by the above and applying the law of total probability that

$$\mathbb{P}\{\mathcal{F}_{t,j}\} = \sum_{i \in S} \mathbb{P}\{\mathcal{F}_{t,j}, X_0 = i\}$$

$$= \pi^\top \{ SL^G(t) + [I - S]L^Q(t) \} e_j.$$

The result (4.5) is established by the Bayes’ rule and the law of total probability,

$$\pi_j(t) = \mathbb{P}\{X_t = j|\mathcal{G}_t\} = \frac{\mathbb{P}\{\mathcal{F}_{t,j}\}}{\sum_{k \in E} \mathbb{P}\{\mathcal{F}_{t,k}\}},$$

whereas the cases (ii) and (iii) follow on account of (4.2) and that $\pi_\Delta = 0$. □

The result of Proposition 4.3 gives an additional feature to the distributional properties of the mixture of Markov jump processes discussed in [33] and [18].
4.2 Univariate conditional phase-type distributions

In this section we derive an explicit identity for \( F_{t|s}(s) = \mathbb{P}\{\tau > s|\mathcal{F}_{t|s}\} \), for \( s \geq t \geq 0 \), which extends the conditional probability (see Theorem 4.1 in [33]):

\[
F_{t|s}(s) = \mathbb{P}\{\tau > s|\mathcal{F}_{t|s}\} = e^t \left( S(t) e^{B(s-t)} + \left[ I - S(t) \right] e^{A(s-t)} \right) \mathbb{1}. \tag{4.6}
\]

**Lemma 4.4** The \( \mathcal{G}_t \)-conditional distribution \( F_t(s) \) is given for \( s \geq t \geq 0 \) by

\[
F_t(s) = \pi^\top (t) \left( S(t) e^{B(s-t)} + \left[ I - S(t) \right] e^{A(s-t)} \right) \mathbb{1}. \tag{4.7}
\]

**Proof** As \( \tau \) is the time until absorption of \( X(t) \), by the law of total probability,

\[
\mathbb{P}\{\tau > s|\mathcal{G}_t\} = \sum_{j,k \in E} \left( \mathbb{P}\{X_s = j, X_t = k, \phi = 1|X_t \neq \Delta, \mathcal{F}_{t-}\} \right.
\]

\[
+ \mathbb{P}\{X_s = j, X_t = k, \phi = 0|X_t \neq \Delta, \mathcal{F}_{t-}\} \right) .
\]

On account that \( \{X_t = k\} \subset \{X_t \neq \Delta\} \), for \( k \in E \), by the Bayes’ formula,

\[
\mathbb{P}\{X_s = j, X_t = k, \phi = 1|X_t \neq \Delta, \mathcal{F}_{t-}\} = \mathbb{P}\{X_t = k|X_t \neq \Delta, \mathcal{F}_{t-}\} \mathbb{P}\{\phi = 1|X_t = k, \mathcal{F}_{t-}\}
\]

\[
\mathbb{P}\{X_s = j|\phi = 1, X_t = k, \mathcal{F}_{t-}\} = \pi_k (t) s_k (t) e_k^\top e_j^\top.
\]

Applying similar arguments for the above derivation, one can obtain

\[
\mathbb{P}\{X_s = j, X_t = k, \phi = 0|X_t \neq \Delta, \mathcal{F}_{t-}\} = \pi_k (t) (1 - s_k (t)) e_k^\top e_j^\top.
\]

The claim in (4.7) is established by applying the transition matrix (4.3). \( \square \)

**Remark 4.5** Following the two identities (4.7) and (4.6), we can conclude that

\[
\mathbb{P}\{\tau > s|\mathcal{G}_t\} = \sum_{i \in E} \pi_i (t) \mathbb{P}\{\tau > s|\mathcal{F}_{t|s}\}. \tag{4.8}
\]

Following the same approach discussed in [28], [27] and [33], the density function \( f_t(s) \) of \( \tau \), its Laplace transform and nth moment are presented below.

**Theorem 4.6** The \( \mathcal{G}_t \)-conditional density \( f_t(s) \) of \( \tau \) is given for \( s \geq t \geq 0 \) by

\[
f_t(s) = -\pi^\top (t) \left( S(t) e^{B(s-t)} B + \left[ I - S(t) \right] e^{A(s-t)} A \right) \mathbb{1}. \tag{4.9}
\]

(i) The Laplace transform \( \hat{F}_t[\lambda] = \int_0^\infty e^{-\lambda u} f_t(t + u) du \) is given by

\[
\hat{F}_t[\lambda] = -\pi^\top (t) \left( S(t) (\lambda I - B)^{-1} B + \left[ I - S(t) \right] (\lambda I - A)^{-1} A \right) \mathbb{1}.
\]

(ii) The \( \mathcal{G}_t \)-conditional nth moment, for \( n = 0, 1, \ldots \), of \( \tau \) is given by

\[
\mathbb{E}\{\tau^n|\mathcal{G}_t\} = (-1)^n n! \pi^\top (t) \left( S(t) B^{-n} + \left[ I - S(t) \right] A^{-n} \right) \mathbb{1}.
\]
Setting \( B = A \), in which case the mixture process is driven by \( X^{(0)} \), the above results coincide with that of given in [28] and Proposition 4.1 in [5] for \( t = 0 \).

The following results summarize the dense and closure properties under finite convex mixtures and finite convolutions of \( F_t(s) \). They can be established using matrix analytic approach [27]. See for e.g. Theorems 4.12 and 4.13 in [33].

**Theorem 4.7** The phase-type distribution \( F_t(s) \) is closed under finite convex mixtures and convolutions, and forms a dense class of distributions on \( \mathbb{R}_+ \).

### 4.3 Bivariate conditional phase-type distributions

As in the univariate case, we consider the mixture process \( X \) on the finite state space \( S = E \cup \{ \Delta \} \). Following [8], let \( \Gamma_1 \) and \( \Gamma_2 \) be two nonempty stochastically closed subsets of \( S \) such that \( \Gamma_1 \cap \Gamma_2 = \Delta \) and the absorption into \( \Delta \) is certain, i.e., the generator matrices \( A \) and \( B \) need to be nonsingular. As \( \Gamma_k, k = 1, 2 \), is stochastically closed set, necessarily \( [Q]_{i,j} = 0 = [G]_{i,j} \) if \( i \in \Gamma_k \) and \( j \in \Gamma^c_k \).

We denote by \( \pi \) the initial probability vector on \( S \) such that \( \pi_\Delta = 0 \). We shall assume that \( \pi_i \neq 0 \) if \( i \in \Gamma_1 \cup \Gamma_2^c \) implying \( \mathbb{P}\{\tau_1 > 0, \tau_2 > 0\} = 1 \). As before, \( F_{t,i} = F_{t-} \cup \{X_t = i\} \) defines all previous and current information of \( X \).

#### 4.3.1 The conditional joint survival function of \( \tau_1 \) and \( \tau_2 \)

The joint distribution of \( \tau_k \) for \( k = 1, 2 \), are given by the following.

**Lemma 4.8** The identity for \( F_{t,i,j} \)-conditional joint distribution \( \overline{F}_{t,i,j}(t_1, t_2) = \mathbb{P}\{\tau_1 > t_1, \tau_2 > t_2 | F_{t,i,j}\} \) of \( \tau_1 \) and \( \tau_2 \) is given for \( t_1, t_2 \geq 0 \) and \( i \in E \) by

\[
\overline{F}_{t,i,j}(t_1, t_2) = \begin{cases} 
\mathbb{P}^{(1)}_{t,i,j}(t_1, t_2) := e_i^T \left( S(t)e^{B(t_2-t)}H_2e^{B(t_1-t)}H_1 \right) 
+ [I - S(t)]e^{A(t_2-t)}H_2e^{A(t_1-t)}H_1 1, & \text{if } t_1 \geq t_2 \geq t \geq 0 \\
\mathbb{P}^{(2)}_{t,i,j}(t_1, t_2) := e_i^T \left( S(t)e^{B(t_1-t)}H_1e^{B(t_2-t)}H_2 \right) 
+ [I - S(t)]e^{A(t_1-t)}H_1e^{A(t_2-t)}H_2 1, & \text{if } t_2 \geq t_1 \geq t \geq 0.
\end{cases}
\]

Note that we have used the notation \( H_k \) to denote a \((m \times m)\)-diagonal matrix whose \( i \)th diagonal element for \( i = 1, 2, ..., m \) equals 1 if \( i \in \Gamma_k \) and 0 otherwise.

**Proof** To begin with let \((t_{i_1}, t_{i_2})\), with \( t_{i_2} \geq t_{i_1} \) be the ordering of \((t_1, t_2)\), with \( t_{i_1} \geq t_{i_0} = t \). Since \( \tau_{i_k}, k = 1, 2 \), is the time until absorption of \( X \) into \( \Gamma_{i_k} \),

\[
\mathbb{P}\{\tau_1 > t_1, \tau_2 > t_2 | \mathcal{F}_{t_{i_0},i}\} = \mathbb{P}\{\tau_{i_1} > t_{i_1}, \tau_{i_2} > t_{i_2} | \mathcal{F}_{t_{i_0},i}\} = \mathbb{P}\{X_{t_{i_1}} \in \Gamma_{i_1}, X_{t_{i_2}} \in \Gamma_{i_2}^c | \mathcal{F}_{t_{i_0},i}\} = \sum_{J_{i_1} \in \Gamma_{i_1}^c} \sum_{J_{i_2} \in \Gamma_{i_2}^c} \mathbb{P}\{X_{t_{i_1}} = J_{i_1}, X_{t_{i_2}} = J_{i_2} | \mathcal{F}_{t_{i_0},i}\}. \quad (4.10)
\]
The probability on the r.h.s of the last equality can be worked out as follows.

\[
P\{X_{t_1} = J_1, X_{t_2} = J_2 | \mathcal{F}_{t_0,i}\} = \sum_{J_1} \sum_{J_2} P\{X_{t_1} = J_1, X_{t_2} = J_2 | \mathcal{F}_{t_0,i}\} = \sum_{J_1} \sum_{J_2} \left[ e_i^T S(t) e^{G(t_1-t_0)} e_{J_1} e_{J_1}^T e^{G(t_2-t_1)} e_{J_2} e_{J_2}^T \right] \]

Note that we have applied the law of total probability and Bayes’ rule for conditional probability in the above equality. Recall that \( P\{X_{t_0} = J_0 | \mathcal{F}_{t_0,i}\} = 1 \) iff \( J_0 = i \) and zero otherwise. In terms of the Bayesian updates \((3.2)\) we have:

\[
\sum_{J} \sum_{J_2} \left[ e_i^T S(t) e^{G(t_1-t_0)} e_{J_1} e_{J_1}^T e^{G(t_2-t_1)} e_{J_2} e_{J_2}^T \right] = \sum_{J} \sum_{J_2} \left[ e_i^T [I - S(t)] e^{Q(t_1-t_0)} e_{J_1} e_{J_1}^T e^{Q(t_2-t_1)} e_{J_2} e_{J_2}^T \right].
\]

Therefore, starting from equation \((4.10)\), we have following the above that

\[
P\{\tau_1 > t_1, \tau_2 > t_2 | \mathcal{F}_{t_0,i}\} = \sum_{J_1} \sum_{J_2} P\{X_{t_1} = J_1, X_{t_2} = J_2 | \mathcal{F}_{t_0,i}\} = \sum_{J} \sum_{J_2} \left[ e_i^T S(t) e^{G(t_1-t_0)} (\sum_{J_1} e_{J_1} e_{J_1}^T) e^{G(t_2-t_1)} (\sum_{J_2} e_{J_2} e_{J_2}^T) \right] \]

leading to the form \( F_{t_1,t_2}(t_1, t_2) \) on account of \( H_{ik} = \sum_{J_{ik} \in \Gamma_i} e_{J_{ik}} e_{J_{ik}}^T \) and \((4.2)\). □

**Proposition 4.9** The distribution \( F_{t_1,t_2} = P\{\tau_1 > t_1, \tau_2 > t_2 | \mathcal{G}_0\} \) is given by

\[
F_{t_1,t_2} = \begin{cases} 
\sum_{i \in E} \pi(t) (S(t) e^{B(t_1-t_0)} H_2 e^{B(t_2-t_1)} H_1 + [I - S(t)] e^{A(t_2-t_1)} H_2 e^{A(t_2-t_1)} H_1) \mathbb{1}, & \text{if } t_1 \geq t_2 \geq t \geq 0 \\
\sum_{i \in E} \pi(t) (S(t) e^{A(t_2-t_1)} H_1 e^{A(t_2-t_1)} H_2 + [I - S(t)] e^{A(t_2-t_1)} H_1 e^{A(t_2-t_1)} H_2) \mathbb{1}, & \text{if } t_2 \geq t_1 \geq t \geq 0 
\end{cases}
\]

**Proof** By \((4.8)\) and law of total probability, \( F_{t_1,t_2} = \sum_{i \in E} \pi(t) F_{i,t_1,t_2} \). □
4.3.2 The conditional joint probability density function

In general, the joint distribution \( \overline{F}_{i,t}(t_1, t_2) \) (resp. \( \overline{F}_t(t_1, t_2) \)) has a singular component \( \overline{F}^{(0)}_{i,t}(t_1, t_2) \) (resp. \( \overline{F}^{(0)}_t(t_1, t_2) \)) on the set \( \{(t_1, t_2) : t_2 = t_1 \} \). The singular component can be obtained by deriving the joint density of \( \tau_1 \) and \( \tau_2 \) and deduce the absolutely continuous and singular parts of the pdf, such as discussed in the theorem below. For non-matrix based bivariate function, see for instance [32].

**Theorem 4.10** Given the joint distribution \( \overline{F}_{i,t}(t_1, t_2) \) of \((\tau_1, \tau_2)\) as specified in Lemma 4.8 the joint probability density \( f_{i,t}(t_1, t_2) \) of \((\tau_1, \tau_2)\) is given by

\[
f_{i,t}(t_1, t_2) = \frac{\partial \overline{F}_{i,t}(t_1, t_2)}{\partial t_2 \partial t_1} = \begin{cases} f_{i,t}^{(1)}(t_1, t_2), & \text{if } t_1 \geq t_2 \geq t \geq 0 \\ f_{i,t}^{(2)}(t_1, t_2), & \text{if } t_2 \geq t_1 \geq t \geq 0 \\ f_{i,t}^{(0)}(t_1, t_1), & \text{if } t_1 = t_2 \geq t \geq 0, \end{cases} \tag{4.11}
\]

where the absolutely continuous components \( f_{i,t}^{(1)}(t_1, t_2) \) and \( f_{i,t}^{(2)}(t_1, t_2) \) are

\[
f_{i,t}^{(1)}(t_1, t_2) = e_t^\top \left( (S(t)e^{B(t_2-t)}[B, H_2]) e^{B(t_1-t)}BH_1 \right.
\]
\[
+ [I - S(t)] e^{A(t_2-t)}[A, H_2] e^{A(t_1-t)}AH_1 \bigg) \mathbb{1},
\]
\[
f_{i,t}^{(2)}(t_1, t_2) = e_t^\top \left( (S(t)e^{B(t_1-t)}[B, H_1]) e^{B(t_2-t)}BH_2 \right.
\]
\[
+ [I - S(t)] e^{A(t_1-t)}[A, H_1] e^{A(t_2-t)}AH_2 \bigg) \mathbb{1},
\]

where the matrix operator \([A, B] = AB - BA\) defines the commutator of \(A\) and \(B\), whilst the singular component part \( f_{i,t}^{(0)}(t_1, t_2) \) is defined by the function

\[
f_{i,t}^{(0)}(t_1, t_1) = e_t^\top \left( S(t)e^{B(t_1-t)} \left( [B, H_2]H_1 + [B, H_1]H_2 - B \right) \right.
\]
\[
+ [I - S(t)] e^{A(t_1-t)} \left( [A, H_2]H_1 + [A, H_1]H_2 - A \right) \bigg) \mathbb{1}.
\]

**Proof** The expression for \( f_{i,t}^{(1)}(t_1, t_2) \) (resp. \( f_{i,t}^{(2)}(t_1, t_2) \)) follows from applying the partial derivative \( \frac{\partial^2}{\partial t_2 \partial t_1} \) to \( \overline{F}_{i,t}^{(1)}(t_1, t_2) \) (resp. to \( \overline{F}_{i,t}^{(2)}(t_1, t_2) \)) taking account that

\[
\frac{d}{dt} (Se^{B_1}A) = SB e^{B_1}A = Se^{B_1}BA. \tag{4.12}
\]

To get \( f_{i,t}^{(0)}(t_1, t_2) \), recall that \( \int_t^\infty e^{Bt}dt = -B^{-1} \) due to \(B\) negative definite, and

\[
1 = \int_t^\infty \int_t^{t_1} f_{i,t}^{(1)}(t_1, t_2) dt_2 dt_1 + \int_t^\infty \int_t^{t_2} f_{i,t}^{(2)}(t_1, t_2) dt_1 dt_2 + \int_t^\infty f_{i,t}^{(0)}(t_1, t_1) dt_1.
\]
Applying Fubini's theorem, the first integral is given after some calculations by
\[
\int_t^\infty \int_t^{t_1} f_{i,t}^{(1)}(t_1, t_2) dt_2 dt_1 = \int_t^\infty \int_t^{t_2} f_{i,t}^{(1)}(t_1, t_2) dt_1 dt_2
\]
\[
= \int_t^\infty \int_t^{t_2} \left( e^{t S(t)} e^{B(t_2-t) [B, H_2]} e^{B(t_1-t_2) BH_1} \right) dt_1 dt_2
\]
\[
+ \int_t^\infty \int_t^{t_2} \left( e^{t S(t)} e^{A(t_2-t) [A, H_2]} e^{A(t_1-t_2) AH_1} \right) dt_1 dt_2
\]
\[
= e^t S(t) B^{-1} [B, H_2] + [I - S(t)] A^{-1} [A, H_2] H_1 1
\]
\[
= - \int_t^\infty e^t S(t) e^{B(t_1-t)} [B, H_2] + [I - S(t)] e^{A(t_1-t)} [A, H_2] H_2 1 dt_1.
\]

Following the same approach, one can show after some calculations that
\[
\int_t^\infty \int_t^{t_2} f_{i,t}^{(2)}(t_1, t_2) dt_2 dt_1 = e^t S(t) B^{-1} [B, H_1] + [I - S(t)] A^{-1} [A, H_1] H_2 1
\]
\[
= - \int_t^\infty e^t S(t) e^{B(t_1-t)} [B, H_1] + [I - S(t)] e^{A(t_1-t)} [A, H_1] H_2 1 dt_1.
\]

The proof is established on account of the following fact that
\[
- \int_t^\infty e^t S(t) e^{B(t_1-t)} B + [I - S(t)] e^{A(t_1-t)} A 1 dt_1 = 1. \quad \square
\]

**Proposition 4.11** For \( t \geq 0 \), the \( G_t \)-conditional density \( f_{i}(t_1, t_2) \) is given by
\[
f_{i}(t_1, t_2) = \frac{\partial^2 F_{i}(t_1, t_2)}{\partial t_2 \partial t_1} = \begin{cases} 
  f_{i}^{(1)}(t_1, t_2), & \text{if } t_1 \geq t_2 \geq t \geq 0 \\
  f_{i}^{(2)}(t_1, t_2), & \text{if } t_2 \geq t_1 \geq t \geq 0 \\
  f_{i}^{(0)}(t_1, t_2), & \text{if } t_1 = t_2 \geq t \geq 0,
\end{cases}
\]

(4.13)

where the absolutely continuous components \( f_{i}^{(1)}(t_1, t_2) \) and \( f_{i}^{(2)}(t_1, t_2) \) are
\[
f_{i}^{(1)}(t_1, t_2) = \pi^t (t) \left( S(t) e^{B(t_2-t)} [B, H_2] e^{B(t_1-t_2)} BH_1 \\
+ [I - S(t)] e^{A(t_2-t)} [A, H_2] e^{A(t_1-t_2)} AH_1 \right) 1,
\]
\[
f_{i}^{(2)}(t_1, t_2) = \pi^t (t) \left( S(t) e^{B(t_1-t)} [B, H_1] e^{B(t_2-t_1)} BH_2 \\
+ [I - S(t)] e^{A(t_1-t)} [A, H_1] e^{A(t_2-t_1)} AH_2 \right) 1,
\]

whilst the singular component \( f_{i}^{(0)}(t_1, t_2) \) is defined by the function
\[
f_{i}^{(0)}(t_1, t_2) = \pi^t (t) \left\{ S(t) e^{B(t_1-t)} \left( [B, H_2] H_1 + [B, H_1] H_2 - B \right) \\
+ [I - S(t)] e^{A(t_1-t)} \left( [A, H_2] H_1 + [A, H_1] H_2 - A \right) \right\} 1.
\]

**Proof** It follows from identity (4.8) that \( f_{i}(t_1, t_2) = \sum_{i=1}^m \pi_i(t) f_{i,t}(t_1, t_2) \). \( \square \)
Remark 4.12 By setting $B = A$ in (4.12) and (4.13) and taking the limit $t \to 0$ in the latter, we arrive at the bivariate distribution given in Assaf et al. [8].

Corollary 4.13 The singular component of $F_{t,t}(t_1, t_2)$ and $F_{t}(t_1, t_2)$ are
\[
F_{t,t}^{(0)}(t_1, t_1) = e_i^T \left\{ S(t) e^{B(t_1-t)} B^{-1} \left( B - [B, H_2] H_1 - [B, H_1] H_2 \right) \right. \\
+ \left. [I - S(t)] e^{A(t_1-t)} A^{-1} \left( A - [A, H_2] H_1 - [A, H_1] H_2 \right) \right\} \mathbb{1}
\]
\[
F_{t}^{(0)}(t_1, t_1) = \pi^T(t) \left\{ S(t) e^{B(t_1-t)} B^{-1} \left( B - [B, H_2] H_1 - [B, H_1] H_2 \right) \right. \\
+ \left. [I - S(t)] e^{A(t_1-t)} A^{-1} \left( A - [A, H_2] H_1 - [A, H_1] H_2 \right) \right\} \mathbb{1}.
\]

Hence, the singular component of $F_{t,t}(t_1, t_2)$ and $F_{t}(t_1, t_2)$ is zero if and only if $[A]_{i,j} = 0 = [B]_{i,j}$ for $i \in \Gamma^c_1 \cap \Gamma^c_2$ and $j = \Delta$, which is equivalent to imposing:
\[
A - [A, H_2] H_1 - [A, H_1] H_2 = 0 = B - [B, H_2] H_1 - [B, H_1] H_2. \quad (4.14)
\]

4.3.3 The conditional joint Laplace transform of $\tau_1$ and $\tau_2$

In order to compute the $F_{t,t}$—conditional moment $\mathbb{E}\{\tau_1 \tau_2 | F_{t,t}\}$, it is therefore convenient to study the $F_{t,i}$—conditional joint Laplace transform of $\tau_1$ and $\tau_2$:
\[
\Psi_{t,i}(\lambda_1, \lambda_2) := \mathbb{E}\{e^{-\lambda_1 \tau_1 - \lambda_2 \tau_2} | F_{t,t}\} \quad \text{for } i \in E.
\]

Theorem 4.14 The $F_{t,i}$—conditional joint Laplace transform $\Psi_{t,i}(\lambda_1, \lambda_2)$ of the exit times $\tau_1$ and $\tau_2$ is given for $\lambda_1, \lambda_2 \geq 0$, $t \geq 0$ and $i \in E$ by
\[
\Psi_{t,i}(\lambda_1, \lambda_2) = e_i^T \left\{ S(t) \left( (\lambda_1 + \lambda_2) I - B \right)^{-1} \left( [B, H_2] (\lambda_1 I - B)^{-1} B H_1 \\
+ [B, H_1] (\lambda_2 I - B)^{-1} B H_2 + [B, H_2] H_1 + [B, H_1] H_2 - B \right) + \right. \\
\left. [I - S(t)] \left( (\lambda_1 + \lambda_2) I - A \right)^{-1} \left( [A, H_2] (\lambda_1 I - A)^{-1} A H_1 \\
+ [A, H_1] (\lambda_2 I - A)^{-1} A H_2 + [A, H_2] H_1 + [A, H_1] H_2 - A \right) \right\} \mathbb{1}.
\]

Proof To start with, recall that for $i \in E$, $f_{i,i}(t_1, t_2) = 0$ for $t_1, t_2 < t$. Therefore,
\[
\Psi_{t,i}(\lambda_1, \lambda_2) = \int_0^\infty \int_0^\infty e^{-\lambda_1 u_1} e^{-\lambda_2 u_2} f_{i,t}(t + u_1, t + u_2) du_1 du_2 \\
= \int_0^\infty \int_0^{u_1} e^{-\lambda_1 u_1} e^{-\lambda_2 u_2} f_{i,t}^{(1)}(t + u_1, t + u_2) du_2 du_1 \\
+ \int_0^\infty \int_0^{u_2} e^{-\lambda_1 u_1} e^{-\lambda_2 u_2} f_{i,t}^{(2)}(t + u_1, t + u_2) du_1 du_2 \\
+ \int_0^\infty e^{-(\lambda_1 + \lambda_2) u_1} f_{i,t}^{(0)}(t + u_1, t + u_1) du_1.
\]

The proof is established after applying Fubini’s theorem to each integral. □
By the law of total probability and Bayes’ rule we have the following result.

**Proposition 4.15** The \( G_t \)-conditional joint Laplace transform \( \Psi_t(\lambda_1, \lambda_2) := \mathbb{E}\{ e^{-\lambda_1 \tau_1 - \lambda_2 \tau_2} \mid G_t \} \) of the exit times \( \tau_1 \) and \( \tau_2 \) is given for \( \lambda_1, \lambda_2, t \geq 0 \) by

\[
\Psi_t(\lambda_1, \lambda_2) = \pi^T(t) \left\{ S(t) \left( (\lambda_1 + \lambda_2) I - B \right)^{-1} \left[ B, H_2 \right] (\lambda_1 I - B)^{-1} B H_1 \right. \\
+ \left. [B, H_1] (\lambda_2 I - B)^{-1} B H_2 + [B, H_2] H_1 + [B, H_1] H_2 - B \right) \\
+ \left[ I - S(t) \right] \left( (\lambda_1 + \lambda_2) I - A \right)^{-1} \left[ [A, H_2] (\lambda_1 I - A)^{-1} A H_1 \right. \\
+ \left. [A, H_1] (\lambda_2 I - A)^{-1} A H_2 + [A, H_2] H_1 + [A, H_1] H_2 - A \right) \} \mathbb{1}.
\]

Following the joint Laplace transform (4.15), we obtain the joint moments:

\[
\mathbb{E}\{ \tau_1^n \tau_2^m \mid \mathcal{F}_{t,i} \} = (-1)^{m+n} \frac{\partial^{m+n}}{\partial \lambda_1^m \partial \lambda_2^n} \Psi_{t,i}(\lambda_1, \lambda_2) \Big|_{\lambda_1=0, \lambda_2=0}.
\]

**Example 4.16** The conditional joint moments \( \mathbb{E}\{ \tau_1 \tau_2 \mid G_t \} \) is given by

\[
\mathbb{E}\{ \tau_1 \tau_2 \mid G_t \} = 2! \pi^T(t) \left\{ S(t) B^{-2} + [I - S(t)] A^{-2} \right\} \mathbb{1} \\
+ \pi^T(t) \left\{ S(t) B^{-2} \left[ [B, H_1] B^{-1} H_2 + [B, H_2] B^{-1} H_1 \right) \\
+ \left[ I - S(t) \right] A^{-2} \left[ [A, H_1] A^{-1} H_2 + [A, H_2] A^{-1} H_1 \right) \} \mathbb{1}.
\]

### 4.4 Multivariate conditional phase-type distributions

The extension to multivariate case follows similar approach to the bivariate one. Let \( \Gamma_1, ..., \Gamma_n \) be nonempty stochastically closed subsets of \( S \) such that \( \cap_{k=1}^n \Gamma_k \) is a proper subset of \( S \). Without loss of generality, we assume that \( \cap_{k=1}^n \Gamma_k = \Delta \).

Since \( \Gamma_k \) is stochastically closed, we necessarily assume that \( g_{ij} = 0 \) (and therefore \( g_{ij} = 0 \) if \( i \in \Gamma_k \) and \( j \in \Gamma_k^c \) for \( k \in \{1, ..., n \} \), and \( \pi_i \neq 0 \) whenever \( i \in \cup_{k=1}^n \Gamma_k^c \).

Furthermore, denote by \( \tau_k \) the first entry time of \( X \) in the set \( \Gamma_k \) defined in (1.6). To formulate the joint distribution of \( \tau_k \), let \( (t_{i_1}, ..., t_{i_n}) \) be the time ordering of \( (t_1, ..., t_n) \) in \( \mathbb{R}_+^n \), where \( (i_1, ..., i_n) \) is a permutation of \( (1, 2, ..., n) \). Subsequently, we define by \( j_{i_k} \in \Gamma_k^c \) the state that \( X \) occupies at time \( t = t_{i_k} \).

**Lemma 4.17** Let \( t_{i_1} \geq ... \geq t_{i_l} \geq t_{i_0} = t \geq 0 \) be the time ordering of \( (t_1, ..., t_n) \) in \( \mathbb{R}_+^n \). The joint distribution of the exit times \( \tau \) (1.6) is given by

\[
\mathbb{P}_{t,i}(t_{i_1}, ..., t_{i_n}) = \mathbb{P}\{ \tau_{i_1} > t_{i_1}, ..., \tau_{i_n} > t_{i_n} \mid \mathcal{F}_{t,i} \}
\]

\[
= e^T \left\{ S(t) \prod_{k=1}^n e^{B(t_{i_k} - t_{i_{k-1}})} H_{i_k} + [I - S(t)] \prod_{k=1}^n e^{A(t_{i_k} - t_{i_{k-1}})} H_{i_k} \right\} \mathbb{1},
\]

where \( H_k \) is an \( (m \times m) \) diagonal matrix whose \( i \)th element \( [H_k]_{i,i} = \mathbb{1}(i \in \Gamma_k^c) \).
Proof Following similar arguments of the proof in bivariate case, we obtain

\[ P\{ \tau_1 > t_1, \ldots, \tau_n > t_n | F_{t_0,i} \} = P\{ \tau_{i_1} > t_{i_1}, \ldots, \tau_{i_n} > t_{i_n} | F_{t_0,i} \} \]
\[ = \sum_{j_{i_1} \in \Gamma_i} \ldots \sum_{j_n \in \Gamma_i} P\{ X_{t_{i_0}} = j_{i_0}, X_{t_{i_1}} = j_{i_1}, \ldots, X_{t_{i_n}} = j_{i_n} | F_{t_0,i} \}. \] (4.17)

By Bayes’ theorem for conditional probability and the law of total probability,

\[ P\{ X_{t_{i_0}} = j_{i_0}, X_{t_{i_1}} = j_{i_1}, \ldots, X_{t_{i_n}} = j_{i_n} | F_{t_0,i} \} \]
\[ = P\{ X_{t_{i_0}} = J_{i_0} | F_{t_0,i} \} \times P\{ \phi = 1 | X_{t_{i_0}} = J_{i_0}, F_{t_0,i} \} \]
\[ \times P\{ X_{t_{i_1}} = J_{i_1} | \phi = 1, X_{t_{i_0}} = J_{i_0}, F_{t_0,i} \} \]
\[ \vdots \]
\[ \times P\{ X_{t_{i_n}} = J_{i_n} | \phi = 1, X_{t_{i_{n-1}}} = J_{i_{n-1}}, \ldots, X_{t_{i_0}} = J_{i_0}, F_{t_0,i} \} \]
\[ + P\{ X_{t_{i_0}} = J_{i_0} | F_{t_0,i} \} \times P\{ \phi = 0 | X_{t_{i_0}} = J_{i_0}, F_{t_0,i} \} \]
\[ \times P\{ X_{t_{i_1}} = J_{i_1} | \phi = 0, X_{t_{i_0}} = J_{i_0}, F_{t_0,i} \} \]
\[ \vdots \]
\[ \times P\{ X_{t_{i_n}} = J_{i_n} | \phi = 0, X_{t_{i_{n-1}}} = J_{i_{n-1}}, \ldots, X_{t_{i_0}} = J_{i_0}, F_{t_0,i} \}. \]

Recall that \( P\{ X_{t_{i_0}} = J_{i_0} | F_{t_0,i} \} = 1 \) iff \( J_{i_0} = i \) and 0 otherwise. In terms of (3.2),

\[ P\{ X_{t_{i_0}} = j_{i_0}, X_{t_{i_1}} = j_{i_1}, \ldots, X_{t_{i_n}} = j_{i_n} | F_{t_0,i} \} \]
\[ = e_i^T S(t) e^{G(t_{i_1}-t_{i_0})} e_{j_{i_1}} e_{j_{i_1}}^T e^{G(t_{i_2}-t_{i_1})} \ldots e_{j_{i_{n-1}}} e_{j_{i_{n-1}}}^T e^{G(t_{i_n}-t_{i_{n-1}})} e_{j_{i_n}} e_{j_{i_n}}^T 1 \]
\[ + e_i^T (I - S(t)) e^{Q(t_{i_1}-t_{i_0})} e_{j_{i_1}} e_{j_{i_1}}^T e^{Q(t_{i_2}-t_{i_1})} \ldots e_{j_{i_{n-1}}} e_{j_{i_{n-1}}}^T e^{Q(t_{i_n}-t_{i_{n-1}})} e_{j_{i_n}} e_{j_{i_n}}^T 1. \]

Therefore, starting from equation (4.17) we have following the above that

\[ P\{ \tau_{i_1} > t_{i_1}, \ldots, \tau_{i_n} > t_{i_n} | F_{t_0,i} \} \]
\[ = e_i^T S(t) e^{G(t_{i_1}-t_{i_0})} \left( \sum_{J_{i_1} \in \Gamma_i} e_{J_{i_1}} e_{J_{i_1}}^T \right) e^{G(t_{i_2}-t_{i_1})} \]
\[ \ldots \left( \sum_{J_{i_{n-1}} \in \Gamma_{i_{n-1}}} e_{J_{i_{n-1}}} e_{J_{i_{n-1}}}^T \right) e^{G(t_{i_n}-t_{i_{n-1}})} \left( \sum_{J_{i_n} \in \Gamma_{i_n}} e_{J_{i_n}} e_{J_{i_n}}^T \right) 1 \]
\[ + e_i^T (I - S(t)) e^{Q(t_{i_1}-t_{i_0})} \left( \sum_{J_{i_1} \in \Gamma_{i_1}} e_{J_{i_1}} e_{J_{i_1}}^T \right) e^{Q(t_{i_2}-t_{i_1})} \]
\[ \ldots \left( \sum_{J_{i_{n-1}} \in \Gamma_{i_{n-1}}} e_{J_{i_{n-1}}} e_{J_{i_{n-1}}}^T \right) e^{Q(t_{i_n}-t_{i_{n-1}})} \left( \sum_{J_{i_n} \in \Gamma_{i_n}} e_{J_{i_n}} e_{J_{i_n}}^T \right) 1, \]

leading to \( F_{t_{i_1},i_1, \ldots, i_n} \) on account of the fact that \( H_{j_k} = \sum_{J_{i_k} \in \Gamma_{i_k}} e_{J_{i_k}} e_{J_{i_k}}^T \) and after applying the block partition (1.2) for exponential matrices \( e^G \) and \( e^Q \). □
Notice that the distribution forms a non-stationary function of time with the ability to capture heterogeneity and path dependence when conditioning on all previous and current information $F_{t_i}$. These features are removed when $B = A$, in which case, the result reduces to the multivariate phase-type distribution [8].

**Proposition 4.18** Let $t_{i_0} \geq \cdots \geq t_{i_1} \geq t_{i_0} = t \geq 0$ be the time ordering of $(t_1, \ldots, t_n) \in R^+_n$. The $G_t$-conditional joint distribution of $\tau_k (1, \ldots, \tau_n)$ is given by

$$F_i(t_{i_1}, \ldots, t_{i_n}) = \mathbb{P}\{\tau_{i_1} > t_{i_1}, \ldots, \tau_{i_n} > t_{i_n} | G_t\} = \pi^T(t) \left\{ S(t) \prod_{k=1}^{n} e^{B(t_{i_k} - t_{i_{k-1}})} H_{i_k} + [I - S(t)] \prod_{k=1}^{n} e^{A(t_{i_k} - t_{i_{k-1}})} H_{i_k} \right\} \mathbb{I}. \quad (4.18)$$

**Proof** It follows from (4.8) that $F_i(t_{i_1}, \ldots, t_{i_n}) = \sum_{i=1}^{m} \pi_i(t) F_{i,t}(t_{i_1}, \ldots, t_{i_n})$. □

**Corollary 4.19** Set $B = A$ and $t = 0$ in (4.18). The joint distribution of $\{\tau_k\}$,

$$\mathbb{P}\{\tau_{i_1} > t_{i_1}, \ldots, \tau_{i_n} > t_{i_n}\} = \pi^T \left( \prod_{k=1}^{n} e^{A(t_{i_k} - t_{i_{k-1}})} H_{i_k} \right) \mathbb{I}, \quad (4.19)$$

coincides with the unconditional multivariate phase-type distribution [8].

The absolutely continuous component of the distribution $F_{i,t}(t_{i_1}, \ldots, t_{i_n})$ (resp. $F_i(t_{i_1}, \ldots, t_{i_n})$) has a density given by the following theorem.

**Theorem 4.20** Let $t_{i_0} \geq \cdots \geq t_{i_1} \geq t_{i_0} = t \geq 0$ be the time ordering of $(t_1, \ldots, t_n) \in R^+_n$. The conditional joint density function of $\tau_k (1, \ldots, \tau_n)$ is given by

$$f_{i,t}(t_{i_1}, \ldots, t_{i_n}) = (-1)^n e_{i}^T \left\{ S(t) \left( \prod_{k=1}^{n-1} e^{B(t_{i_k} - t_{i_{k-1}})[B, H_{i_k}]} \right) e^{B(t_n - t_{i_n})} BH_{i_n} \right.$$

$$\left. + (I - S(t)) \left( \prod_{k=1}^{n-1} e^{A(t_{i_k} - t_{i_{k-1}})[A, H_{i_k}]} \right) e^{A(t_n - t_{i_n})} AH_{i_n} \right\} \mathbb{I}.$$ 

$$f_{i}(t_{i_1}, \ldots, t_{i_n}) = (-1)^n \pi^T(t) \left\{ S(t) \left( \prod_{k=1}^{n-1} e^{B(t_{i_k} - t_{i_{k-1}})[B, H_{i_k}]} \right) e^{B(t_n - t_{i_n})} BH_{i_n} \right.$$

$$\left. + (I - S(t)) \left( \prod_{k=1}^{n-1} e^{A(t_{i_k} - t_{i_{k-1}})[A, H_{i_k}]} \right) e^{A(t_n - t_{i_n})} AH_{i_n} \right\} \mathbb{I}.$$ 

**Proof** The proof follows from taking partial derivative to $F_{i,t}(t_{i_1}, \ldots, t_{i_n})$:

$$f_{i,t}(t_{i_1}, \ldots, t_{i_n}) = (-1)^n \frac{\partial^n F_{i,t}}{\partial t_{i_n} \cdots \partial t_{i_1}} (t_{i_1}, \ldots, t_{i_n}).$$

To justify the claim, we use induction argument. For this purpose, recall that

$$\prod_{k=1}^{n} e^{B(t_{i_k} - t_{i_{k-1}})} H_{i_k} = e^{B(t_{i_1} - t_{i_0})} H_{i_1} e^{B(t_{i_2} - t_{i_1})} H_{i_2} \prod_{k=3}^{n} e^{B(t_{i_k} - t_{i_{k-1}})} H_{i_k}. \quad (4.20)$$
Hence, by (4.12) and applying integration by part as we did before, we have
\[
\frac{\partial}{\partial t_{i_1}} \prod_{k=1}^{n} e^{B(t_{i_k} - t_{i_{k-1}})H_{i_k}} = e^{B(t_{i_1} - t_{i_0})}[B, H_{i_1}] \prod_{k=2}^{n} e^{B(t_{i_k} - t_{i_{k-1}})H_{i_k}},
\]
from which the second order partial derivative \(\frac{\partial^2}{\partial t_{i_2} \partial t_{i_1}}\) of (4.20) is given by
\[
\frac{\partial^2}{\partial t_{i_2} \partial t_{i_1}} \prod_{k=1}^{n} e^{B(t_{i_k} - t_{i_{k-1}})H_{i_k}} = e^{B(t_{i_1} - t_{i_0})}[B, H_{i_1}] \frac{\partial}{\partial t_{i_2}} \left( \prod_{k=2}^{n} e^{B(t_{i_k} - t_{i_{k-1}})H_{i_k}} \right)
= e^{B(t_{i_1} - t_{i_0})}[B, H_{i_1}] e^{B(t_{i_2} - t_{i_1})}[B, H_{i_2}] \left( \prod_{k=3}^{n} e^{B(t_{i_k} - t_{i_{k-1}})H_{i_k}} \right).
\]
After \((n-1)\) steps of taking the partial derivative, one can show that
\[
\frac{\partial^{n-1}}{\partial t_{i_n} \ldots \partial t_{i_1}} \prod_{k=1}^{n} e^{B(t_{i_k} - t_{i_{k-1}})H_{i_k}} = \left( \prod_{k=1}^{n-1} e^{B(t_{i_k} - t_{i_{k-1}})H_{i_k}} \right) e^{B(t_{i_n} - t_{i_{n-1}})H_{i_n}}.
\]
The claim is established on account of (4.12) and the fact that
\[
\frac{\partial^n}{\partial t_{i_n} \ldots \partial t_{i_1}} \prod_{k=1}^{n} e^{B(t_{i_k} - t_{i_{k-1}})H_{i_k}} = e_1^T \left\{ S(t) \frac{\partial^n}{\partial t_{i_n} \ldots \partial t_{i_1}} \left( \prod_{k=1}^{n} e^{B(t_{i_k} - t_{i_{k-1}})H_{i_k}} \right) \right\}
+ \left[ I - S(t) \frac{\partial^n}{\partial t_{i_n} \ldots \partial t_{i_1}} \left( \prod_{k=1}^{n} e^{A(t_{i_k} - t_{i_{k-1}})H_{i_k}} \right) \right] \mathbb{1}.
\]
The proof for the \(G_i\)-conditional joint density \(f_t(t_{i_1}, \ldots, t_{i_n})\) follows from
\[
f_t(t_{i_1}, \ldots, t_{i_n}) = \sum_{i=1}^{m} \pi_i(t) \frac{\partial^n}{\partial t_{i_n} \ldots \partial t_{i_1}} \mathcal{F}_{i,t}(t_{i_1}, \ldots, t_{i_n}).
\]
However, due to complexity of the joint distributions, the singular component of \(\mathcal{F}_{i,t}(t_{i_1}, \ldots, t_{i_n})\) (resp. \(\mathcal{F}_t(t_{i_1}, \ldots, t_{i_n})\)) is more complicated to get in closed form.

Following (4.16) and (4.18), we see that the distributions are presented in terms of a generalized mixture of the multivariate phase-type distributions [8].

They are uniquely characterized by the probability \(\pi\) of starting the mixture process \(X\) (2.1) on the state space \(\mathbb{S}\), the speed of the process, which is represented by the phase generator matrices \(B\) and \(A\), and by the switching probability matrix \(S\). They coincide with [8] when the process never repeatedly changes the speed, i.e., \(B = A\) and when sending \(t\) to zero. As in the univariate case, the multivariate distributions have closure and dense properties, which can be established in similar ways to the univariate analogs using matrix analytic approach [8]. We refer among others to [27], [9], [21] and [31] for the Markov model, and to [33] for the mixture model. As a result, we have the following.

**Theorem 4.21 (Closure and dense properties)** The multivariate probability distribution (4.18) forms a dense class of distributions on \(\mathbb{R}_+^n\), which is closed under finite convex mixtures and finite convolutions.
5 Some explicit examples

This section discusses some explicit examples of the main results presented in Section 4, particularly on the bivariate distributions. Using the closed form density functions (4.11) and (4.13), we discuss the mixtures of exponential distributions, Marshall-Olkin exponential distributions, and their generalization.

Example 5.1 (Mixture of exponential distribution) Consider the mixture process \( X \) defined on the state space \( \mathbb{S} = \{1, 2, 3\} \cup \{\Delta\} \) with stochastically closed sets \( \Gamma_1 = \{2, \Delta\} \) and \( \Gamma_2 = \{3, \Delta\} \). Assume that the speed of the mixture process is represented by the following phase generator matrices:

\[
B = \begin{pmatrix}
-(b_1 + b_2) & b_1 & b_2 \\
0 & -b_2 & 0 \\
0 & 0 & -b_1
\end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix}
-(a_1 + a_2) & a_1 & a_2 \\
0 & -a_2 & 0 \\
0 & 0 & -a_1
\end{pmatrix}.
\]

It is straightforward to derive from the state space representation that

\[
H_1 = \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{pmatrix} \quad \text{and} \quad H_2 = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}.
\]

After some calculations, the matrices \([A, H_k]\) and \(AH_k\), \(k = 1, 2\), are given by

\[
[A, H_1] = \begin{pmatrix}
0 & -a_1 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix} \quad \text{and} \quad [A, H_2] = \begin{pmatrix}
0 & 0 & -a_2 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

\[
AH_1 = \begin{pmatrix}
-(a_1 + a_2) & 0 & a_2 \\
0 & 0 & 0 \\
0 & 0 & -a_1
\end{pmatrix} \quad \text{and} \quad AH_2 = \begin{pmatrix}
-(a_1 + a_2) & a_1 & 0 \\
0 & 0 & -a_2 \\
0 & 0 & 0
\end{pmatrix}.
\]

Similarly defined for \([B, H_k]\) and \(BH_k\), for \( k = 1, 2\). Set the matrix \( S = \text{diag}(p_1, p_2, p_3) \), with \( 0 < p_k < 1 \), for \( k = 1, 2, 3 \), whilst the initial probability \( \pi \) has mass one on the state 1, i.e., \( \pi = e_1 \). It is straightforward to check that the condition (4.14) is clearly satisfied implying that the joint density function (4.17) has zero singular component. Hence, following (4.17) we have

\[
f_{\tau_1, \tau_2}(t_1, t_2) = p_1 b_1 e^{-b_1 t_1} b_2 e^{-b_2 t_2} + (1 - p_1) a_1 e^{-a_1 t_1} a_2 e^{-a_2 t_2},
\]

for \( t_1, t_2 \geq 0 \). The marginal distribution of \( \tau_1 \) and \( \tau \) are given respectively by

\[
f_{\tau_1}(t_1) = p_1 b_1 e^{-b_1 t_1} + (1 - p_1) a_1 e^{-a_1 t_1},
\]

\[
f_{\tau_2}(t_2) = p_1 b_2 e^{-b_2 t_2} + (1 - p_1) a_2 e^{-a_2 t_2}.
\]

Hence, clearly, as \( f_{\tau_1, \tau_2}(t_1, t_2) \neq f_{\tau_1}(t_1) f_{\tau_2}(t_2) \), it follows that the exit times \( \tau_1 \) and \( \tau_2 \) are not independent under the mixture model. They are independent if
and only if \( a_1 = b_1 = b_2 = a_2 \), in which case the mixture corresponds to a simple Markov jump process. See the example on p. 691 in [8] and p. 59 in [21].

Furthermore, when conditioning on the information set \( \mathcal{F}_{t,i} \) with \( i = 1 \), the conditional joint density function \( f_{1,t}(t_1, t_2) \) is given for \( t_1, t_2 \geq t \geq 0 \) by

\[
f_{1,t}(t_1, t_2) = s_1(t)e^{(b_1+b_2)t}b_1e^{-b_1t_1}b_2e^{-b_2t_2} + (1 - s_1(t))e^{(a_1+a_2)t}a_1e^{-a_1t_1}a_2e^{-a_2t_2}, \quad (5.1)
\]

where the switching probability \( s_1(t) \) is defined for all \( t \geq 0 \) by

\[
s_1(t) = \frac{p_1e^{-(b_1+b_2)t}}{p_1e^{-(b_1+b_2)t} + (1-p_1)e^{-(a_1+a_2)t}}.
\]

Observe that, on the event \( \{ \min\{\tau_1, \tau_2\} > t \} \), one can check that \( s_1(t) \to 0 \) (resp. 1) as \( t \to \infty \) if \( b_1 + b_2 > \) (resp. <) \( a_1 + a_2 \), implying that the mixture process moves at the slower speed \( A \) (resp. \( B \)) in the long run as a Markov process.

Given that \( \Gamma_1 \cap \Gamma_2 = \{1\} \), we have \( \tau_1(t) = 1 \) for all \( t \geq 0 \). Hence, the density function \( f_1(t_1, t_2) \) has therefore the same expression as (5.1).

**Example 5.2 (Mixture of Marshall-Olkin distribution)** Consider a mixture process \( \tilde{X} \) with the same state space \( \mathcal{S} \) and stochastically closed sets \( \Gamma_1 \) and \( \Gamma_2 \) as defined above. Let the speed of the mixture process be given by

\[
B = \begin{pmatrix}
-(b_1 + b_2 + b_3) & b_1 & b_2 \\
0 & -(b_2 + b_3) & 0 \\
0 & 0 & -(b_1 + b_3)
\end{pmatrix}
\]

\[
A = \begin{pmatrix}
-(a_1 + a_2 + a_3) & a_1 & a_2 \\
0 & -(a_2 + a_3) & 0 \\
0 & 0 & -(a_1 + a_3)
\end{pmatrix}.
\]

Set the matrix \( S = \text{diag}(p_1, p_2, p_3) \), with \( 0 < p_k < 1 \), for \( k = 1, 2, 3 \), and whilst the initial distribution has mass one on the state 1, i.e., \( \pi_1 = 1 \). Following (4.14), the joint density \( f_{1,t}(t_1, t_2) \) has singular part on the set \( \{(t_1, t_2) : t_2 = t_1\} \). By the same approach as above, we have following Theorem 4.10 and Corollary (1.13):

\[
f_{1,t}^{(1)}(t_1, t_2) = s_1(t)b_2(b_1 + b_3)e^{-b_1(t_1-t)}e^{-b_2(t_2-t)}e^{-b_3(t_1-t_2)} + (1 - s_1(t))a_2(a_1 + a_3)e^{-a_1(t_1-t)}e^{-a_2(t_2-t)}e^{-a_3(t_1-t_2)}
\]

\[
f_{1,t}^{(2)}(t_1, t_2) = s_1(t)b_1(b_2 + b_3)e^{-b_1(t_1-t)}e^{-b_2(t_2-t)}e^{-b_3(t_2-t_1)} + (1 - s_1(t))a_1(a_2 + a_3)e^{-a_1(t_1-t)}e^{-a_2(t_2-t)}e^{-a_3(t_2-t_1)}
\]

\[
f_{1,t}^{(0)}(t_1, t_1) = s_1(t)b_2e^{-(b_1+b_2+b_3)(t_1-t)} + (1 - s_1(t))a_3e^{-(a_1+a_2+a_3)(t_1-t)}
\]

whereas the switching probability \( s_1(t) \) is given for all \( t \geq 0 \) by

\[
s_1(t) = \frac{p_1e^{-(b_1+b_2+b_3)t}}{p_1e^{-(b_1+b_2+b_3)t} + (1-p_1)e^{-(a_1+a_2+a_3)t}}
\]
In order to take advantage to the structure of the generator matrices, let

\[
B = \begin{pmatrix}
B_{11} & B_{12} & B_{13} \\
0 & B_{22} & 0 \\
0 & 0 & B_{33}
\end{pmatrix}
\quad \text{and} \quad
A = \begin{pmatrix}
A_{11} & A_{12} & A_{13} \\
0 & A_{22} & 0 \\
0 & 0 & A_{33}
\end{pmatrix}.
\] (5.2)

The generator matrices \(A\) and \(B\) are nonsingular if and only if \(A_{11}, A_{22}, A_{33}, B_{11}, B_{22}\) and \(B_{33}\) are all nonsingular. The matrices \(H_1\) and \(H_2\) are given by

\[
H_1 = \begin{pmatrix}
I & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & I
\end{pmatrix}
\quad \text{and} \quad
H_2 = \begin{pmatrix}
I & 0 & 0 \\
0 & I & 0 \\
0 & 0 & 0
\end{pmatrix}.
\]

After some calculations the matrix \([A, H_k]\) and \(AH_k, k = 1, 2\), are given by

\[
[A, H_1] = \begin{pmatrix}
0 & -A_{12} & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\quad \text{and} \quad
[A, H_2] = \begin{pmatrix}
0 & 0 & -A_{13} \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

\[
AH_1 = \begin{pmatrix}
A_{11} & 0 & A_{13} \\
0 & 0 & 0 \\
0 & 0 & A_{33}
\end{pmatrix}
\quad \text{and} \quad
AH_2 = \begin{pmatrix}
A_{11} & A_{12} & 0 \\
0 & A_{22} & 0 \\
0 & 0 & 0
\end{pmatrix}.
\]

Similarly defined for \([B, H_k]\) and \(BH_k, k = 1, 2\). A rather long calculations using infinite series representation of exponential matrix shows following \((5.11)\),

\[
f_{i,t}^{(1)}(t_1, t_2) = \begin{cases}
    f_{i,t}^{(1)}(t_1, t_2), & \text{if } t_1 \geq t_2 \geq t \geq 0 \\
    f_{i,t}^{(2)}(t_1, t_2), & \text{if } t_2 \geq t_1 \geq t \geq 0 \\
    f_{i,t}^{(0)}(t_1, t_1), & \text{if } t_1 = t_2 \geq t \geq 0,
\end{cases}
\] (5.3)

where the absolutely continuous parts \(f_{i,t}^{(1)}(t_1, t_2)\) and \(f_{i,t}^{(2)}(t_1, t_2)\) are given by

\[
f_{i,t}^{(1)}(t_1, t_2) = -e_i^T \left\{ S_{11}(t) e^{B_{11}(t_2-t)} B_{13} e^{B_{33}(t_1-t_2)} B_{33} \\
+ [I - S_{11}(t)] e^{A_{11}(t_2-t)} A_{13} e^{A_{33}(t_1-t_2)} A_{33} \right\} 1,
\]

\[
f_{i,t}^{(2)}(t_1, t_2) = -e_i^T \left\{ S_{11}(t) e^{B_{11}(t_1-t)} B_{12} e^{B_{22}(t_2-t_1)} B_{22} \\
+ [I - S_{11}(t)] e^{A_{11}(t_1-t)} A_{12} e^{A_{22}(t_2-t_1)} A_{22} \right\} 1,
\]

whereas the singular component \(f_{i,t}^{(0)}(t_1, t_1)\) is defined by the function

\[
f_{i,t}^{(0)}(t_1, t_1) = -e_i^T \left\{ S_{11}(t) e^{B_{11}(t_1-t)} \left( B_{11} + B_{12} + B_{13} \right) + [I - S_{11}(t)] e^{A_{11}(t_1-t)} \left( A_{11} + A_{12} + A_{13} \right) \right\} 1.
\]

Note that \(S_{11}(t)\) denotes the switching probability matrix of \(X\) on \(\Gamma_1^c \cap \Gamma_2^c\).
Thus, the distribution \( f_{t,t}(t_1, t_2) \) has no singular component \( f_{t,t}^{(0)}(t_1, t_2) \) iff

\[
A_{11} + A_{12} + A_{13} = 0 = B_{11} + B_{12} + B_{13}.
\]

Denote by \( \alpha \) the restriction of the probability distribution \( \pi \) on the set \( \Gamma_1 \cap \Gamma_2 \) such that \( \pi = (\alpha, 0) \). The Bayesian updates \( \pi(t) \) on \( \Gamma_1 \cap \Gamma_2 \) is defined by \( \alpha(t) \).

The conditional joint density \( f_t(t_1, t_2) \) of the exit times \( \tau_1 \) and \( \tau_2 \) is given by

\[
f_t(t_1, t_2) = \begin{cases} 
  f_t^{(1)}(t_1, t_2), & \text{if } t_1 \geq t_2 \geq t \geq 0 \\
  f_t^{(2)}(t_1, t_2), & \text{if } t_2 \geq t_1 \geq t \geq 0 \\
  f_t^{(0)}(t_1, t_1), & \text{if } t_1 = t_2 \geq t \geq 0,
\end{cases}
\] (5.4)

where the subdensity functions \( f_t^{(1)}(t_1, t_2) \), \( f_t^{(2)}(t_1, t_2) \) and \( f_t^{(0)}(t_1, t_2) \) are

\[
f_t^{(1)}(t_1, t_2) = -\alpha^\top(t) \left\{ \begin{array}{c} S_{11}(t) e^{B_{11} (t_2 - t)} B_{13} e^{B_{13} (t_1 - t)} B_{33} \\
[I - S_{11}(t)] e^{A_{11} (t_2 - t)} A_{13} e^{A_{33} (t_1 - t)} A_{33} \end{array} \right\} \mathbb{1},
\]

\[
f_t^{(2)}(t_1, t_2) = -\alpha^\top(t) \left\{ \begin{array}{c} S_{11}(t) e^{B_{11} (t_1 - t)} B_{12} e^{B_{12} (t_2 - t)} B_{22} \\
[I - S_{11}(t)] e^{A_{11} (t_1 - t)} A_{12} e^{A_{22} (t_2 - t)} A_{22} \end{array} \right\} \mathbb{1},
\]

\[
f_t^{(0)}(t_1, t_1) = -\alpha^\top(t) \left\{ \begin{array}{c} S_{11}(t) e^{B_{11} (t_1 - t)} (B_{11} + B_{12} + B_{13}) \\
[I - S_{11}(t)] e^{A_{11} (t_1 - t)} (A_{11} + A_{12} + A_{13}) \end{array} \right\} \mathbb{1}.
\]

The marginal distribution of \( \tau_k \), \( k = 1, 2 \), can be established in the same approach for the univariate case. It is given for \( s \geq t \geq 0 \) and \( i \in \Gamma_1 \cap \Gamma_2 \) by

\[
\mathbb{P}\{\tau_k > s | F_{t,i}\} = e_t^\top \left( S_{11}(t) e^{B_{11} (s-t)} + [I - S_{11}(t)] e^{A_{11} (s-t)} \right) \mathbb{1},
\] (5.5)

\[
\mathbb{P}\{\tau_k > s | G_{t}\} = \pi^\top(t) \left( S_{11}(t) e^{B_{11} (s-t)} + [I - S_{11}(t)] e^{A_{11} (s-t)} \right) \mathbb{1},
\] (5.6)

where the phase-generator matrices \( B^{(k)} \) and \( A^{(k)} \), for \( k = 1, 2 \), are defined by

\[
B^{(1)} = \begin{pmatrix} B_{11} & B_{13} \\ 0 & B_{33} \end{pmatrix} \quad \text{and} \quad A^{(1)} = \begin{pmatrix} A_{11} & A_{13} \\ 0 & A_{33} \end{pmatrix},
\]

\[
B^{(2)} = \begin{pmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{pmatrix} \quad \text{and} \quad A^{(2)} = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}.
\]
6 Conclusions

We have introduced a generalization of the multivariate phase-type distributions under the mixture of absorbing Markov jump processes moving at different speeds on the same finite state space. Such mixture process was proposed in [18], and was discussed in further details in [33]. The new distributions form non-stationary function of time and have the ability to capture heterogeneity and the past information of the process, when conditioning on its past observation. The attribution of path dependence is due to the non-Markov property of the process. Identities are presented explicit in terms of the Bayesian updates of switching probability and the probability of starting the process in any of the transient phases at any given time, the likelihoods, and the intensity matrices of the underlying processes, despite the fact that the mixture itself is non-Markov. When the underlying Markov processes move at the same speed, in which case the mixture becomes a simple Markov jump process, heterogeneity and path dependence are removed from the identities, and the distributions reduce to [8].

The results presented in this paper can be extended in a natural way to the mixture of a finite number of absorbing Markov jump processes moving at different speeds. Given their availability in explicit form and fine properties, the new distributions should be able to offer appealing features for applications.

References

[1] Aalen, O.O. and Gjessing, H.K. (2001). Understanding the shape of the hazard rate: a process point of view. Statistical Science, 16, 1-22.

[2] Aalen, O.O. (1995). Phase type distributions in survival analysis. Scandinavian Journal of Statistics, 22, 447-463.

[3] Asimit, A.V. and Jones, B.L. (2007). Extreme behavior of multivariate phase-type distributions. Insurance: Mathematics and Economics, 41, 223-233.

[4] Asmussen, S., Avram, F. and Pistorius, M.R. (2004). Russian and American put options under exponential phase-type Lévy models. Stochastic Processes and Applications, 109, 79-111.

[5] Asmussen, S. (2003). Applied Probability and Queues, 2nd Edition, Springer.

[6] Albrecher, H. and Asmussen, S. (2010). Ruin Probabilities, 2nd Edition, World Scientific.

[7] Asmussen, S., Nerman, O. and Olsson, M. (1996). Fitting phase-type distribution via the EM algorithm. Scandinavian Journal of Statistics, 23, 419-441.
On the conditional joint probability distributions of phase-type

[8] Assaf, D., Langberg, N.A., Savits, T.H. and Shaked, M. (1984). Multivariate phase-type distributions. Operations Research, 32, 688-702.

[9] Assaf, D. and Levikson, B. (1982). Closure of phase type distributions under operations arising in reliability theory. Annals of Probability, 10, 265-269.

[10] Badila, E.S. Boxma, O.J. and Resing, J.A.C. (2014). Document Queues and risk processes with dependencies. Stochastic Models, 30, 390-419.

[11] Balzter, H. (2000). Markov chain models for vegetation dynamics. Ecological Modelling, 126, pp. 139-154.

[12] Berdel, J. and Hipp, C. (2011). Convolutions of multivariate phase-type distributions. Insurance: Mathematics and Economics, 48, 374-377.

[13] Berger, P.D. and Nasr, N.I. (1998). Customer lifetime value: Marketing models and applications. Journal of Interactive Marketing, 12, 17-30.

[14] Bielecki, T., Crépey, S., Herbertsson, A. (2013) Markov chain models of portfolio credit risk. Oxford Handbook of Credit Derivatives, 327-382.

[15] Bielecki, T. and Rutkowski, M. (2002). Credit Risk: Modeling, Valuation and Hedging. Springer.

[16] Blumen, I., Kogan, M. and McCarthy, P.J. (1955). The industrial mobility of labor as a probability process. Cornell Studies of Industrial and Labor Relations, Vol. 6, Ithaca, N.Y., Cornell University Press.

[17] Chakravarthy, S.R. and Neuts, M.F. (2014). Analysis of a multi-server queueing model with MAP arrivals of regular customers and phase type arrivals of special customers. Simulation Modelling Practice and Theory, 43, 79-95.

[18] Frydman, H. and Schuermann, T. (2008). Credit rating dynamics and Markov mixture models, Journal of Banking and Finance, 32, 1062-1075.

[19] Frydman, H. (2005). Estimation in the mixture of Markov chains moving with different speeds. Journal of the American Statistical Association, 100, 1046-1053.

[20] Frydman, H. (1984). Maximum likelihood estimation in the mover-stayer model, Journal of the American Statistical Association, 79, 632-638.

[21] He, Q.M. (2014). Fundamental of Matrix-Analytic Methods. Springer.

[22] Herbertsson, A. (2011). Modelling default contagion using multivariate phase-type distributions, Review of Derivative Research, 14, 1-36.
[23] Jarrow, R., Lando, D., and Turnbull, S. (1997). A Markov model for the terms structure of credit risk spreads, *Review of Financial Derivatives*, **10**, 481-523.

[24] Jarrow, R. and Turnbull, S. (1995). Pricing derivatives on financial securities subject to credit risk, *Journal of Finance*, **50**, 53-86.

[25] Kulkarni, V.G. (1989). New class of multivariate phase type distributions, *Operations Research*, **37**, 151-158.

[26] Lee, S.C.K. and Lin, X.S. (2012). Modelling dependent risks with multivariate Erlang mixtures, *ASTIN Bulletin*, **42**, 153-180.

[27] Neuts, M.F. (1981). *Matrix-Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore.

[28] Neuts, M.F. (1975). Probability distributions of phase-type. In *Liber Amicorum Prof. Emeritus H. Florin*, 173-206, University of Louvain, Belgium.

[29] Pfeiffer, P. E. and Carraway, R. L. (2000). Modelling customer relationship as Markov chains. *Journal of Interactive Marketing*, **14**, 43-55.

[30] Pintilie, M. (2006). *Competing Risks: A Practical Perspective*. John Wiley & Sons, Canada.

[31] Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J. (1998). *Stochastic Processes for Insurance and Finance*, Willey.

[32] Sarhan, A.M. and Balakrishnan, N. (2007). A new class of bivariate distributions and its mixture. *Journal of Multivariate Analysis*, **98**, 1508-1527.

[33] Surya, B. A. (2018). Distributional properties of the mixture of continuous-time absorbing Markov chains moving at different speeds, *Stochastic Systems*, **8**, 29-44.

[34] Okamura, H. and Dohi, T. (2016). Phase-type software reliability model: parameter estimation algorithms with grouped data. *Annals of Operations Research*, **244**, 177-208.

[35] Willmot, G.E. and Woo, J-K. (2014). On some properties of a class of multivariate Erlang mixtures with insurance applications, *ASTIN Bulletin*, **45**, 151-173.

[36] Zadeh, A.H., Jones, B.L. and Stanford, D.A. (2014). The use of phase-type models for disability insurance calculations, *Scandinavian Actuarial Journal*, **8**, 714-728.