The Mixture of Markov Jump Processes: Monte Carlo Method and the EM Estimation

H. Frydman[*] and B.A. Surya[†]

Department of IOMS-Statistics

New York University Stern School of Business

and

School of Mathematics and Statistics

Victoria University of Wellington

27 November 2018

Abstract

This paper discusses tractable development and statistical estimation of a continuous time stochastic process with a finite state space having non-Markov property. The process is formed by a finite mixture of right-continuous Markov jump processes moving at different speeds on the same finite state space, whereas the speed regimes are assumed to be unobservable. The mixture was first proposed by Frydman [6] in 2005 and recently generalized in Surya [11] and [12], in which distributional properties and explicit identities of the process are given in its full generality. The contribution of this paper is two fold. First, we present Monte Carlo method for constructing the process and show distributional equivalence between the simulated process and the actual process. Secondly, we perform statistical inference on the distribution parameters of the process. Under complete observation of the sample paths, maximum likelihood estimates are given in explicit form in terms of sufficient statistics of the process. Estimation under incomplete observation is performed using the EM algorithm. The estimation results completely characterize the process in terms of the initial probability of starting the process in any phase of the state space, intensity matrices of the underlying Markov jump processes, and the switching probability matrix of the process. Some numerical examples are given to test the performance of the developed method. The proposed

---

[*]NYU Leonard N. Stern School of Business, Kaufman Management Center 44 West Fourth Street, 8-55 New York, NY 10012, USA. Email address: hfrydman@stern.nyu.edu

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

estimation generalizes the existing statistical inferences for the Markov model [1], the mover-stayer model [7] and the Markov mixture model [6].

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

Keywords: Markov jump process, Markov mixture process, mover-stayer model, statistical inference, maximum likelihood, EM algorithm

1   The mixture of Markov jump processes

Throughout the remaining of this paper we denote by $X = \{X^{(\phi)}(t), t \geq 0\}$ a Markov mixture process, which is a continuous-time stochastic process defined as a finite mixture of independent Markov jump processes $X^{(m)} = \{X^{(m)}(t) : t \geq 0\}$, with $m = 1, \ldots, M$, whose intensity matrices are given by $\{\mathbf{Q}^{(m)}\}$. We assume that the underlying Markov processes $\{X^{(m)}\}$ have right-continuous sample paths, defined on the same state space $\mathbb{S} = \{1, \ldots, p\}$. It is defined by

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

where the variable ϕ represents the speed regimes, assumed to be unobservable. This is to say that when the realization of the mixture process X is observed, we do not know from which speed regime ϕ the observed process came from.

More conveniently, we can represent X in terms of the underlying processes $\{X^{(m)}\}$ as follows. Define an indicator variable $\Phi^{(m)} = \mathbb{1}_{\{\phi=m\}}$, which will be used later for the estimation, see (3.1). Notice that $\sum_{m=1}^{M} \Phi^{(m)} = 1$. Thus,

$$X(t) = \sum_{m=1}^{M} \Phi^{(m)} X^{(m)}(t) \quad \text{for } t \geq 0. \tag{1.1}$$

It is clear that X (1.1) represents a finite mixture of Markov processes $X^{(m)}$, and that the random variable $\Phi^{(m)}$ may in general depend on the realization of X as $\Phi^{(m)} = 1$ if and only if $X = X^{(m)}$. This implies that the conditional probability $\mathbb{P}\{\Phi^{(m)} = 1 | X(s), 0 \leq s \leq t\}$ depends on the past realizations of the process. Empirical evidence of this fact can be found in [5]. We refer to [11] and [12] for further distributional properties and explicit identities of the mixture process, in particular in the presence of stochastically closed (absorbing) sets.

The entry $\{q_{ij}^{(m)} : i, j = 1, \ldots, p\}$ of matrix $\mathbf{Q}^{(m)}$ satisfies the properties:

$$q_{ii}^{(m)} \leq 0, \;\; q_{ij}^{(m)} \geq 0, \;\; \sum_{j \neq i} q_{ij}^{(m)} = -q_{ii}^{(m)} = q_i^{(m)}, \quad (i, j) \in \mathbb{S}. \tag{1.2}$$

For a given initial state $i_0 \in \mathbb{S}$, there is a separate mixing probability

$$s_{i_0}^{(m)} = \mathbb{P}\{\phi = m | X_0 = i_0\} \quad \text{with} \quad \sum_{m=1}^{M} s_{i_0}^{(m)} = 1, \tag{1.3}$$
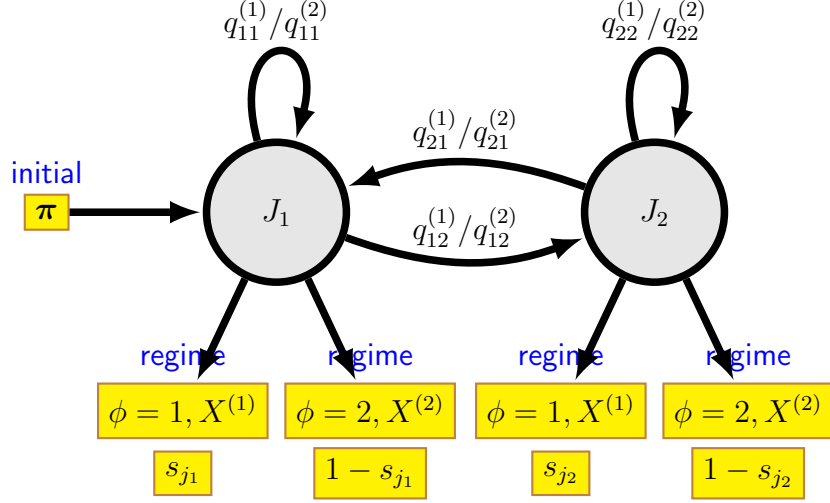
Figure 1: State diagram of the Markov mixture process (1.1) with $m = 2$.

and $0 \le s_{i_0}^{(m)} \le 1$. The quantity $s_{i_0}^{(m)}$ has the interpretation as the proportion of population with initial state i_0 evolving w.r.t to $X^{(m)}$. In general, $X^{(k)}$ and $X^{(l)}$, $k \ne l$, have different expected length of occupation time of a state i, i.e., $1/q_i^{(k)} \ne 1/q_i^{(l)}$, and have different probability of leaving state $i \in E$ to state $j \in \mathbb{S}$, $j \ne i$, i.e. $q_{ij}^{(k)}/q_i^{(k)} \ne q_{ij}^{(l)}/q_i^{(l)}$. Note that we have used $q_i^{(m)}$ and $q_{ij}^{(m)}$ to denote the negative of the ith diagonal element and the (i, j) entry of $\mathbf{Q}^{(m)}$.

Figure 1 illustrates the transition of X for the mixture of two Markov jump processes moving from state J_1 to J_2, and vice versa. When X is observed in state J_1, it would stay in the state for an exponential period of time with intensity $q_{j_1}^{(1)}$ or $q_{j_1}^{(2)}$ before moving to J_2 with probability $q_{j_1,j_2}^{(1)}/q_{j_1}^{(1)}$ or $q_{j_1,j_2}^{(2)}/q_{j_1}^{(2)}$, depending on whether it is either driven by the underlying Markov process $X^{(1)}$ or $X^{(2)}$.

Markov mixture process is a generalization of mover-stayer model, a mixture of two discrete-time Markov chains proposed in 1955 by Blumen et al [3] to model population heterogeneity in jobs labor market. In the mover-stayer model [3], the population of workers consists of stayers (workers who always stay in the same job category, $\mathbf{Q}^{(1)} = \mathbf{0}$) and movers (workers who move to other job according to a stationary Markov chain with intensity matrix $\mathbf{Q}^{(2)}$). Estimation of the mover-stayer model was discussed in Frydman [7]. Frydman [6] generalized the model to a finite mixture of Markov chains moving with different speeds. Frydman and Schuermann [5] later on used the result for the mixture of two Markov jump processes moving with intensity matrices $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)} = \mathbf{\Psi}\mathbf{Q}^{(1)}$, i.e., $q_{ij}^{(2)} = \psi_i q_{ij}^{(1)}$, where $\mathbf{\Psi}$ is a diagonal matrix, to model the dynamics of firms' credit ratings. Depending on whether $0 = \psi_i := [\mathbf{\Psi}]_{i,i}$, $0 < \psi_i < 1$, $\psi_i > 1$ or $\psi_i = 1$, $X^{(2)}$ 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 $X^{(1)}$. If $\psi_i = 1$, for all $i \in \mathbb{S}$, X reduces to a simple Markov jump process $X^{(1)}$.

However, the mixture model considered in [6] and [5] is restricted to the case in which the underlying Markov processes $X^{(1)}$ and $X^{(2)}$ have the same probability of leaving a state to another different state. That is $q_{ij}^{(2)}/q_i^{(2)} = q_{ij}^{(1)}/q_i^{(1)}$, $j \neq i$.

The mixture process X has appealing features that, unlike its underlying process $X^{(m)}$, the mixture itself lacks the Markov property; future development of its state depends on its past information and the current time. We refer to [5], [11] and [12] for further distributional properties of the mixture process.

The transition probability matrix $\mathbf{P}(t)$ of X is given following [6] and [11] by

$$\mathbf{P}(t) = \sum_{m=1}^{M} \mathbf{S}^{(m)} e^{\mathbf{Q}^{(m)} t} \quad \text{with} \quad \sum_{m=1}^{M} \mathbf{S}^{(m)} = \mathbf{I}, \tag{1.4}$$

for all $t \geq 0$, where $\mathbf{I}$ is $(p \times p)-$identity matrix, whereas $\mathbf{S}^{(m)}$ denotes a $(p \times p)-$diagonal matrix, representing the switching probability matrix of X, i.e.,

$$\mathbf{S}^{(m)} = \begin{pmatrix} s_1^{(m)} & \mathbf{0} \\ \mathbf{0} & s_p^{(m)} \end{pmatrix}. \tag{1.5}$$

It is clear from (1.4) and (1.5) that the distribution of the mixture process X (1.1) is uniquely characterized by the variables $\{q_{ij}^{(m)}\}$, $\{q_i^{(m)}\}$, respectively representing the off and diagonal elements of $\mathbf{Q}^{(m)}$, the element $\{s_i^{(m)}\}$ of switching probability matrix $\mathbf{S}$, and the probability distribution $\{\pi_i\}$ of starting X in any state $i \in \mathbb{S}$. Furthermore, when we set $\mathbf{Q}^{(m)} = \mathbf{Q}$, all underlying Markov processes move at the same speed $\mathbf{Q}$, X becomes just a $\mathbf{Q}-$Markov process.

2 Construction of the mixture process

This section discusses construction of the mixture process X (1.1), which can be used to generate the sample paths of the mixture process using Monte Carlo method developed by adapting the approach of Sections 2.1 and 5.1 in Resnick [9]. The simulated sample paths will later be used to solve the inverse problem of estimating the distribution parameters of the process, given its full or incomplete observation of the sample paths, which is the subject of Sections 3 and 4.

2.1 Finite mixture of Markov chains

To start with, let U_0, U, V, and W be independent uniform $[0, 1]$ random variables. Introduce a discrete-time Markov chain $Z^{(m)} = \{Z_n^{(m)} : n \in \mathbb{N}_+\}$, living on the same finite state space $\mathbb{S}$. It is defined as the corresponding embedded Markov chain for the Markov process $X^{(m)}$. The transition probability matrix of $Z^{(m)}$ is specified by $(p \times p)-$matrix $\mathbf{\Pi}^{(m)}$ whose $(i, j)-$element is defined by

$$\pi_{ij}^{(m)} = \begin{cases} q_{ij}^{(m)}/q_i^{(m)}, & j \neq i \\ 0, & j = i. \end{cases} \tag{2.1}$$

In the mixture model [6], each embedded Markov chain $Z^{(m)}$ was assumed to have the same transition probability matrix $\mathbf{\Pi}$, i.e., $\mathbf{\Pi}^{(m)} = \mathbf{\Pi}$ for all $m = 1, \dots, M$.

Assume X chooses its initial state $X_0 = i_0$ randomly with probability $\boldsymbol{\pi}$:

$$X_0 = \sum_{k=1}^{p} k \mathbb{1}_{[\sum_{i=1}^{k-1} \pi_i, \sum_{i=1}^{k} \pi_i)}(U_0), \tag{2.2}$$

where, we set $\sum_{i=1}^{0} \pi_i = 0$. Applying similar idea, the speed regime $\phi = m$ of X can be selected randomly, given initial state $X_0 = i_0$, at probability $s_{i_0}^{(m)}$ using

$$\phi = \sum_{m=1}^{M} m \mathbb{1}_{[\sum_{k=1}^{m-1} s_{i_0}^{(k)}, \sum_{k=1}^{m} s_{i_0}^{(k)})}(U). \tag{2.3}$$

In the sequel below we denote respectively by $\{V_n\}$ and $\{W_n\}$ n independent copies of the random variables V and W independent of U_0 and U. The result below gives a Monte Carlo construction of the discrete-time Markov chain $Z_n^{(m)}$.

Lemma 2.1 *For a given $m = 1, \dots, M$, the process $\{Y_n : n \in \mathbb{N}_+\}$ defined by*

$$Y_{n+1} = \sum_{k=1}^{p} k \mathbb{1}_{[\sum_{j=1}^{k-1} [\mathbf{\Pi}^{(m)}]_{Y_n, j}, \sum_{j=1}^{k} [\mathbf{\Pi}^{(m)}]_{Y_n, j})}(V_{n+1}),$$
$$Y_0 = X_0 \text{ a.s.,} \tag{2.4}$$

forms the Markov chain $\{Z_n^{(m)}\}$ with transition probability matrix $\mathbf{\Pi}^{(m)}$ (2.1).

Proof By applying the Bayes' formula and the law of total probability,

$$\mathbb{P}\{Y_{n+1} = i_{n+1} | Y_0 = i_0\} = \sum_{i_1 \in \mathbb{S}} \cdots \sum_{i_n \in \mathbb{S}} \mathbb{P}\{Y_{n+1} = i_{n+1}, Y_n = i_n, \dots, Y_1 = i_1 | Y_0 = i_0\}$$
$$= \sum_{i_1 \in \mathbb{S}} \cdots \sum_{i_n \in \mathbb{S}} \mathbb{P}\{Y_1 = i_1 | Y_0 = i_0\} \mathbb{P}\{Y_2 = i_2 | Y_1 = i_1, Y_0 = i_0\}$$
$$\times \cdots \times \mathbb{P}\{Y_n = i_n | Y_{n-1} = i_{n-1}, \dots, Y_0 = i_0\}$$
$$\times \mathbb{P}\{Y_{n+1} = i_{n+1} | Y_n = i_n, \dots, Y_0 = i_0\}. \tag{2.5}$$

On account that $\{V_n\}$ is a series of independent random variables independent of U_0, $\mathbb{P}\{Y_k = i_k | Y_{k-1} = i_{k-1}, \dots, Y_0 = i_0\} = \mathbb{P}\{Y_k = i_k | Y_{k-1} = i_{k-1}\}$. Thus,

$$\mathbb{P}\{Y_k = i_k | Y_{k-1} = i_{k-1}, \dots, Y_0 = i_0\} = \mathbf{e}_{i_{k-1}}^{\top} \mathbf{\Pi}^{(m)} \mathbf{e}_{i_k}, \tag{2.6}$$

for $k = 1, \dots, n+1$. As $\sum_{i_k \in \mathbb{S}} \mathbf{e}_{i_k} \mathbf{e}_{i_k}^{\top} = \mathbf{I}_{p \times p}$, we have following (2.6) and (2.5),

$$\mathbb{P}\{Y_{n+1} = i_{n+1} | Y_0 = i_0\} = \sum_{i_1 \in \mathbb{S}} \cdots \sum_{i_n \in \mathbb{S}} \mathbf{e}_{i_0}^{\top} \mathbf{\Pi}^{(m)} \mathbf{e}_{i_1} \mathbf{e}_{i_1}^{\top} \mathbf{\Pi}^{(m)} \dots \mathbf{\Pi}^{(m)} \mathbf{e}_{i_n} \mathbf{e}_{i_n}^{\top} [\mathbf{\Pi}^{(m)}] \mathbf{e}_{i_{n+1}}$$
$$= \mathbf{e}_{i_0}^{\top} \mathbf{\Pi}^{(m)} \Big(\sum_{i_1 \in \mathbb{S}} \mathbf{e}_{i_1} \mathbf{e}_{i_1}^{\top} \Big) \mathbf{\Pi}^{(m)} \dots \mathbf{\Pi}^{(m)} \Big(\sum_{i_n \in \mathbb{S}} \mathbf{e}_{i_n} \mathbf{e}_{i_n}^{\top} \Big) [\mathbf{\Pi}^{(m)}] \mathbf{e}_{i_{n+1}}$$
$$= \mathbf{e}_{i_0}^{\top} [\mathbf{\Pi}^{(m)}]^{n+1} \mathbf{e}_{i_{n+1}},$$

showing that $\{Y_n\}$ forms the Markov chains with transition matrix $\mathbf{\Pi}^{(m)}$. $\square$

Next, define a Bernoulli random variable $\Phi^{(m)} = \mathbb{1}_{\{\phi=m\}}$ and the function:

$$F(i, \Phi, V) = \sum_{k=1}^{p} \sum_{m=1}^{M} k\Phi^{(m)} \mathbb{1}_{[\sum_{j=1}^{k-1}[\mathbf{\Pi}^{(m)}]_{i,j}, \sum_{j=1}^{k}[\mathbf{\Pi}^{(m)}]_{i,j})}(V), \qquad (2.7)$$

where we set $\sum_{j=1}^{0}[\mathbf{\Pi}^{(m)}]_{i,j} = 0$ for all $i = 1, \ldots, p$, $m = 1, \ldots, M$ and $n \in \mathbb{N}$. Following (2.7), consider a finite mixture of Markov chains $\{Z_n\}$ defined by

$$Z_{n+1} = F(Z_n, \mathbf{\Phi}, V_{n+1}) \quad \text{with} \quad Z_0 = X_0 \text{ a.s.,} \qquad (2.8)$$

where $\mathbf{\Phi}$ denotes $(1 \times M)-$vector of speed regime, i.e., $\mathbf{\Phi} = (\Phi^{(1)}, \ldots, \Phi^{(M)})$.

Proposition 2.2 *The representation (2.8) yields the mixture of Markov chains*

$$Z_n = \sum_{m=1}^{M} \Phi^{(m)} Z_n^{(m)}, \quad with \ Z_0 = X_0, \qquad (2.9)$$

where $Z_n^{(m)}$ satisfies the recursive equation $Z_{n+1} = F(Z_n, \Phi^{(m)} = 1, V_{n+1})$, see eqn. (2.4), while the n-step transition probability matrix of $\{Z_n\}$ is defined by

$$\mathbf{P}^{(n)} = \sum_{m=1}^{M} \mathbf{S}^{(m)} \left[\mathbf{\Pi}^{(m)}\right]^n. \qquad (2.10)$$

Proof The representation (2.9) follows from (2.7) and (2.8) and applying the Fubini's principle to (2.7). By the Bayes formula and law of total probability,

$$\begin{aligned}
\left[\mathbf{P}^{(n)}\right]_{i,j} &= \mathbb{P}\{Z_n = j | Z_0 = i\} \\
&= \sum_{m=1}^{M} \mathbb{P}\{\Phi^{(m)} = 1 | Z_0 = i\} \mathbb{P}\{Z_n = j | \Phi^{(m)} = 1, Z_0 = i\} \\
&= \sum_{m=1}^{M} \mathbb{P}\{\Phi^{(m)} = 1 | Z_0 = i\} \mathbb{P}\{Z_n^{(m)} = j | Z_0^{(m)} = i\} \\
&= \sum_{m=1}^{M} s_i^{(m)} \left[\left[\mathbf{\Pi}^{(m)}\right]^n\right]_{i,j} = \mathbf{e}_i^{\top} \sum_{m=1}^{M} \mathbf{S}^{(m)} \left[\mathbf{\Pi}^{(m)}\right]^n \mathbf{e}_j,
\end{aligned}$$

where the last equality follows from taking account of Lemma 2.1, leading to the establishment of identity (2.10), given $\mathbf{S}^{(m)}$ is a $(p \times p)-$diagonal matrix. $\square$

Remark 2.3 *It is worth noticing that if the underlying Markov chains $\{Z_n^{(m)} : m = 1, \ldots, M\}$ have the same transition probability matrix $\mathbf{\Pi}$, which is the case for the mixture model [6], one can show that $\{Z_n\}$ (2.8) has the same distribution as $\{Y_n\}$ (2.4). The latter is used in [9] to generate the Markov chains $\{Y_n\}$.*
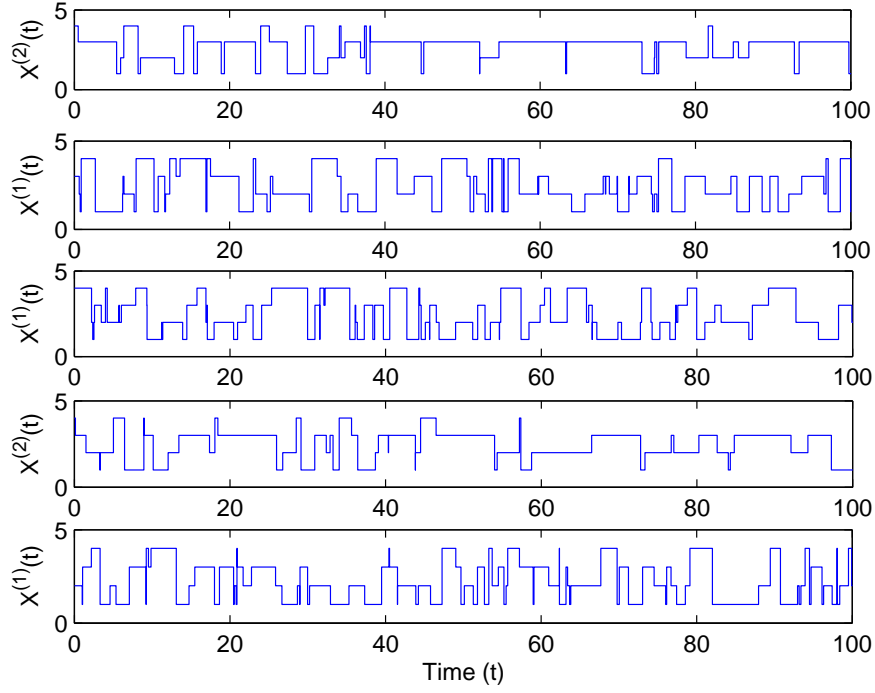
Figure 2: Sample paths of the mixture of Markov jump process X (1.1)

2.2   Finite mixture of Markov jump processes

The epoch time $\{T_n\}$ of the mixture process X (1.1) is defined by

$$T_{n+1} = T_n - \sum_{k=1}^{M} \Phi^{(k)} \frac{\log W_n}{q_{Z_n}^{(k)}}, \quad \text{with } T_0 = 0 \text{ a.s.} \tag{2.11}$$

Theorem 2.4 *Define a continuous-time stochastic process* $X = \{X(t) : t \geq 0\}$,

$$X(t) = \sum_{n=0}^{\infty} Z_n \mathbb{1}_{[T_n, T_{n+1})}(t) \text{ with } X(0) = X_0. \tag{2.12}$$

Then, the stochastic process X *has the following properties:*

(i) *it has the representation of the mixture process (1.1) with*

$$X^{(m)}(t) := \sum_{n=0}^{\infty} Z_n^{(m)} \mathbb{1}_{[T_n, T_{n+1})}(t), \tag{2.13}$$

(ii) *and has the same distribution (1.4) as the mixture process* X *(1.1).*

Figure 2 displays Monte Carlo simulation of the sample paths of X (2.12). As we can see, at every fixed point of time $t \geq 0$ and state $i \in \mathbb{S}$ the observed paths is comprised of a mixture of two Markov jump processes $X^{(1)}(t)$ and $X^{(2)}(t)$.

In order to establish the theorem, the following results are required.

Lemma 2.5 *Define the following transition probability matrix* $\mathbf{P}_{i,j}^{(m)}(t)$:

$$\mathbf{P}_{i,j}^{(m)}(t) := \mathbb{P}\{X(t) = j | \Phi^{(m)} = 1, X(0) = i\}. \qquad (2.14)$$

Then, for all $t \geq 0$, the function $t \to \mathbf{P}_{i,j}^{(m)}(t)$ solves the integral equation:

$$\mathbf{P}_{i,j}^{(m)}(t) = e^{-q_i^{(m)}t}\delta_{ij} + \int_0^t q_i^{(m)} e^{-q_i^{(m)}u} \sum_{l \neq i} \pi_{i,l}^{(m)} \mathbf{P}_{l,j}^{(m)}(t-u) du. \qquad (2.15)$$

Proof Since T_1 is the first jump time of X, the number of steps n until the Markov chain $Z^{(m)}$ makes a jump from state i to state $j \neq i$ has geometric distribution with probability $\mathbf{e}_i^\top [1 - \mathbf{\Pi}^{(m)}]^{n-1} \mathbf{\Pi}^{(m)} \mathbf{e}_j$. By independence of W_0,

$$\mathbb{P}\{X(t) = j, T_1 > t | \Phi^{(m)} = 1, X(0) = i\}$$
$$= \sum_{n=1}^\infty \mathbb{P}\{Z_n = j, T_1 > t | \Phi^{(m)} = 1, Z_0 = i\}$$
$$= \sum_{n=1}^\infty \mathbb{P}\left\{Z_n^{(m)} = j, -\frac{\log W_0}{q_i^{(m)}} > t \Big| Z_0^{(m)} = i\right\}$$
$$= \sum_{n=1}^\infty e^{-q_i^{(m)}t} \mathbf{e}_i^\top (1 - \mathbf{\Pi}^{(m)})^{n-1} \mathbf{\Pi}^{(m)} \mathbf{e}_j$$
$$= e^{-q_i^{(m)}t}\delta_{ij},$$

with $\delta_{ij} = \mathbf{e}_i^\top \mathbf{e}_j$, provided the transition probability matrix $\mathbf{\Pi}^{(m)}$ is invertible.

Furthermore, following (2.11) the epoch time T_n of X can be singled out as

$$T_n = -\sum_{k=0}^{n-1} \sum_{m=1}^M \Phi^{(m)} \frac{\log W_k}{q_{Z_k}^{(m)}} \quad \text{with} \quad T_0 = 0 \quad \text{a.s.}$$

Again, given that T_1 is the first jump time of X moving from state i to $j \neq i$,

we have by independence of W_0, Bayes' formula and Fubini's principle that

$$\mathbb{P}\{X(t) = j, T_1 \leq t | \Phi^{(m)} = 1, X(0) = i\}$$

$$= \sum_{l \neq i} \mathbb{P}\left\{X(t) = j, -\frac{\log W_0}{q_i^{(m)}} \leq t, Z_1 = l \Big| \Phi^{(m)} = 1, Z_0 = i\right\}$$

$$= \sum_{l \neq i} \int_0^t \mathbb{P}\left\{-\frac{\log W_0}{q_i^{(m)}} \in du \Big| \Phi^{(m)} = 1, Z_0 = i\right\} \mathbb{P}\left\{Z_1 = l | \Phi^{(m)} = 1, Z_0 = i\right\}$$

$$\times \mathbb{P}\left\{X(t) = j \Big| \Phi^{(m)} = 1, -\frac{\log W_0}{q_i^{(m)}} = u, Z_1 = l, Z_0 = i\right\}$$

$$= \int_0^t q_i^{(m)} e^{-q_i^{(m)} u} \sum_{l \neq i} \pi_{i,l}^{(m)} \mathbb{P}\left\{X(t) = j \Big| \Phi^{(m)} = 1, -\frac{\log W_0}{q_i^{(m)}} = u, Z_1 = l, Z_0 = i\right\} du.$$

The proof is accomplished once we have shown that the following identity holds:

$$\mathbb{P}\left\{X(t) = j \Big| \Phi^{(m)} = 1, -\frac{\log W_0}{q_i^{(m)}} = u, Z_1 = l, Z_0 = i\right\} = \mathbf{P}_{l,j}^{(m)}(t - u). \quad (2.16)$$

To this end, recall that the conditional probability in (2.16) can be simplified as

$$\mathbb{P}\left\{X(t) = j \Big| \Phi^{(m)} = 1, -\frac{\log W_0}{q_i^{(m)}} = u, Z_1 = l, Z_0 = i\right\}$$

$$= \sum_{n=1}^{\infty} \mathbb{P}\left\{Z_n = j, T_n \leq t < T_{n+1} \Big| \Phi^{(m)} = 1, -\frac{\log W_0}{q_i^{(m)}} = u, Z_1 = l, Z_0 = i\right\}$$

$$= \sum_{n=1}^{\infty} \mathbb{P}\left\{Z_n = j, -\sum_{k=1}^{n-1} \frac{\log(W_k)}{q_{Z_k}^{(m)}} \leq t - u < -\sum_{k=1}^{n} \frac{\log(W_k)}{q_{Z_k}^{(m)}} \Big| \Phi^{(m)} = 1, Z_1 = l, Z_0 = i\right\}$$

$$= \sum_{n=1}^{\infty} \mathbb{P}\left\{Z_{n-1} = j, -\sum_{k=1}^{n-1} \frac{\log(W_k)}{q_{Z_{k-1}}^{(m)}} \leq t - u < -\sum_{k=1}^{n} \frac{\log(W_k)}{q_{Z_{k-1}}^{(m)}} \Big| \Phi^{(m)} = 1, Z_0 = l\right\},$$

where in the last equality we have used the fact that conditional on $\Phi^{(m)} = 1$, the discrete-time mixture process Z_n is a Markov chain moving according to $Z_n^{(m)}$, which by Lemma 2.1 has the memoryless property. Next, define new indexes $n' = n - 1$, $k' = k - 1$, and $W_k' = W_{k'+1}$. By doing so, we can rewrite

$$\sum_{n=1}^{\infty} \mathbb{P}\left\{Z_{n-1} = j, -\sum_{k=1}^{n-1} \frac{\log(W_k)}{q_{Z_{k-1}}^{(m)}} \leq t - u < -\sum_{k=1}^{n} \frac{\log(W_k)}{q_{Z_{k-1}}^{(m)}} \Big| \Phi^{(m)} = 1, Z_0 = l\right\}$$

$$= \sum_{n'=0}^{\infty} \mathbb{P}\left\{Z_{n'} = j, -\sum_{k'=0}^{n'-1} \frac{\log(W_k')}{q_{Z_{k'}}^{(m)}} \leq t - u < -\sum_{k'=0}^{n'} \frac{\log(W_k')}{q_{Z_{k'}}^{(m)}} \Big| \Phi^{(m)} = 1, Z_0 = l\right\}$$

$$= \mathbf{P}_{l,j}^{(m)}(t - u).$$

The integral equation (2.15) is obtained by invoking the law of total probability and the Bayes' formula for conditional probability to (2.14) to get

$$
\begin{aligned}
\mathbf{P}_{i,j}^{(m)}(t) =& \mathbb{P}\{X(t) = j, T_1 > t | \Phi^{(m)} = 1, X(0) = i\} \\
& + \mathbb{P}\{X(t) = j, T_1 \leq t | \Phi^{(m)} = 1, X(0) = i\}.
\end{aligned}
$$

The final result (2.15) is obtained by collecting the two pieces together.   □

Proposition 2.6 *The integral equation (2.15) has an explicit solution:*

$$
\mathbf{P}_{i,j}^{(m)}(t) = \mathbf{e}_i^\top e^{\mathbf{Q}^{(m)}t} \mathbf{e}_j, \quad i.e., \quad \mathbf{P}^{(m)}(t) = e^{\mathbf{Q}^{(m)}t}. \tag{2.17}
$$

Proof On account of the fact that $\sup_t \mathbf{P}_{i,j}^{(m)}(t) \leq 1$ and $\sum_{l \neq i} \pi_{i,l}^{(m)} = 1$, the function inside the integral in (2.15) is uniformly bounded. Thus, the integral is a continuous function of t, and therefore so is the function $\mathbf{P}_{i,j}^{(m)}(t)$. Hence, as a result, $\mathbf{P}_{i,j}^{(m)}(t)$ is absolutely continuous w.r.t. to Lebesgue measure dt, i.e., it is continuously differentiable. Applying change of variable $s = t - u$, we obtain

$$
\mathbf{P}_{i,j}^{(m)}(t) = e^{-q_i^{(m)}t}\left(\delta_{ij} + q_i^{(m)} \int_0^t e^{q_i^{(m)}s} \sum_{l \neq i} \pi_{i,l}^{(m)} \mathbf{P}_{l,j}^{(m)}(s)ds\right).
$$

As the function $\mathbf{P}_{i,j}^{(m)}(t)$ is continuously differentiable, we have

$$
\begin{aligned}
\frac{d}{dt}\mathbf{P}_{i,j}^{(m)}(t) =& -q_i^{(m)} e^{-q_i^{(m)}t}\left(\delta_{ij} + q_i^{(m)} \int_0^t e^{q_i^{(m)}s} \sum_{l \neq i} \pi_{i,l}^{(m)} \mathbf{P}_{l,j}^{(m)}(s)ds\right) \\
& + e^{-q_i^{(m)}t}\left(q_i^{(m)} e^{q_i^{(m)}t} \sum_{l \neq i} \pi_{i,l}^{(m)} \mathbf{P}_{l,j}^{(m)}(t)\right) \\
=& -q_i^{(m)}\mathbf{P}_{i,j}^{(m)}(t) + q_i^{(m)} \sum_{l \neq i} \pi_{i,l}^{(m)} \mathbf{P}_{l,j}^{(m)}(t) \\
=& \sum_{l \in \mathbb{S}} \mathbf{Q}_{i,l}^{(m)} \mathbf{P}_{l,j}^{(m)}(t).
\end{aligned}
$$

As $\mathbf{P}^{(m)}(0) = \mathbf{I}$, it is straightforward to check that the linear systems of equation has the unique solution (2.17) for the transition matrix $\mathbf{P}^{(m)}(t)$ for $t \geq 0$.   □

Proof of Theorem 2.4

The claim is established by the law of total probability and Bayes' formula for conditional probability taking account of Lemma 2.5 and Proposition 2.6, i.e.,

$$
\begin{aligned}
[\mathbf{P}(t)]_{i,j} =& \mathbb{P}\{X(t) = j | X(0) = i\} \\
=& \sum_{m=1}^{M} \mathbb{P}\{\Phi^{(m)} = 1 | X(0) = i\}\, \mathbb{P}\{X(t) = j | \Phi^{(m)} = 1, X(0) = i\} \\
=& \sum_{m=1}^{M} s_i^{(m)} \mathbf{P}_{i,j}^{(m)}(t) = \mathbf{e}_i^\top \sum_{m=1}^{M} \mathbf{S}^{(m)} \mathbf{P}^{(m)}(t)\mathbf{e}_j.
\end{aligned}
$$

Inserting the expression of Z_n (2.9) in (2.12) yields the mixture (1.1). $\square$

Pseudo algorithm for the simulation

(1) **Step 1** Choose initial state $X_0 = i_0$ with the probability $\boldsymbol{\pi}$ using (2.2).

(2) **Step 2** On the chosen initial state i_0, select the speed regime $\phi = m$ with the probability $\mathbf{s}_{i_0} = (s_{i_0}^{(1)}, \ldots, s_{i_0}^{(M)})$ using (2.3). Then, set $\Phi^{(m)} = 1$.

(3) **Step 3** Make transition from time zero to a specified final time $T > 0$ according to the selected Markov jump process $\{X_t^{(m)}\}_{t \geq 0}$ starting from the initial state i_0 by applying the recursive equations (2.8)-(2.12).

3   Estimation with complete information

Statistical estimation of $\{q_{ij}^{(m)}\}$, $\{q_i^{(m)}\}$, and $\{s_i^{(m)}\}$ was discussed in [6] for special structure of intensity matrix $\mathbf{Q}^{(m)}$, where it is assumed that $\mathbf{Q}^{(m)} = \boldsymbol{\Psi}^{(m)} \mathbf{Q}$ with $\boldsymbol{\Psi}^{(m)} = \mathrm{diag}(\psi_1^{(m)}, \ldots, \psi_p^{(m)})$ for $1 \leq m \leq M - 1$ and $\boldsymbol{\Psi}^{(M)} = \mathbf{I}$, i.e., $q_{ij}^{(m)} = \psi_i^{(m)} q_{ij}$ implying that $\pi_{ij}^{(m)} = \pi_{ij}$ for all $m = 1, \ldots, M$, see (2.1). This is equivalent to imposing the condition on the embedded Markov chains $Z_n^{(m)}$ to have the same transition probability matrix $\boldsymbol{\Pi}$, i.e., $\boldsymbol{\Pi}^{(m)} = \boldsymbol{\Pi}$ for all m.

This paper attempts to generalize the estimation method [6] for inferring the distribution parameters $\{q_{ij}^{(m)}\}$, $\{q_i^{(m)}\}$, and $\{s_i^{(m)}\}$ of X for a general structure of $\mathbf{Q}^{(m)}$ as well as to estimate $\{\pi_i\}$, which was not discussed in [6]. Importantly, as opposed to the EM estimator $\widehat{s}_i^{(m)}$ given in [6], our estimate for $s_i^{(m)}$ sums to one, i.e., $\sum_{m=1}^M \widehat{s}_i^{(m)} = 1$, for $i = 1, \ldots, p$, which is due to the constraint (1.3). The results can be used to recover the estimation for the restricted mixture [6].

3.1   Maximum likelihood estimation

This section discusses maximum likelihood estimation of the distribution of X based on complete observations of X, where the underlying process driving the evolution of X is known. To start with, we assume that N independent realizations $\{X_k\}$ of X are observed continuously on the time interval $[0, T]$ with

$0 < T < \infty$. For notational convenience, we use the following conventions:

$$\Phi_k^{(m)} = \mathbb{1}_{\{X_k = X^{(m)}\}}$$

$$B_i^{(k)} = \mathbb{1}_{\{X_k(0)=i\}}$$

$$N_{ij}^{(k)} = \sum_{l=1}^{N} \mathbf{1}_{\{X_k(lh)=j, X_k((l-1)h)=i\}}, \text{ with } h = T/N, \ N \in \mathbb{N}$$

$$N_i^{(k)} = \sum_{j \neq i} N_{ij}^{(k)}$$ (3.1)

$$Z_i^{(k)} = \int_0^T \mathbb{1}_{\{X_k(u)=i\}} du,$$

for $i, j = 1, \ldots, p$, $k = 1, \ldots, N$ and $m = 1, \ldots, M$. More precisely, $N_{ij}^{(k)}$ counts the (number) of transitions of X_k from state i to state $j \neq i$, $N_i^{(k)}$ counts the number of transitions from state i, whereas $Z_i^{(k)}$ represents the occupation time of X_k in state i when X_k is observed on the interval $[0, T]$, while $B_i^{(k)}$ counts the number of realizations of X_k starting in state i at time zero. Also, note that

$$\sum_{m=1}^{M} \Phi_k^{(m)} = 1 \quad \text{and} \quad \sum_{i=1}^{p} B_i^{(k)} = 1, \quad \text{for } k = 1, \ldots, N,$$ (3.2)

which in turn implies that

$$\sum_{i=1}^{p} \sum_{k=1}^{N} \sum_{m=1}^{M} \Phi_k^{(m)} B_i^{(k)} = N.$$ (3.3)

Under complete information, the statistics (3.1) are assumed to be available for the maximum likelihood estimation of the distribution parameters of X (1.1). In the sequel below we denote by $\boldsymbol{\theta} = (\boldsymbol{\pi}, \mathbf{Q}^{(m)}, \mathbf{S}^{(m)})$ and $f_{\boldsymbol{\theta}}(X_k, \Phi_k^{(m)})$ the joint probability density function of the observations X_k and $\Phi_k^{(m)}$, indicator variable which provides information on which underlying process that drives X_k.

Suppose that X_k chooses its initial state i_k to start with randomly at probability π_{i_k}. On account that the bivariate process $(X_k, \Phi_k^{(m)})$ is Markovian, it follows from applying the Bayes' formula for conditional probability that

$$f_{\boldsymbol{\theta}}(X_k, \Phi_k^{(m)}) = f_{\boldsymbol{\theta}}(X_k(0) = i_k) f_{\boldsymbol{\theta}}(\Phi_k^{(m)} | X_k(0) = i_k)$$
$$\times f_{\boldsymbol{\theta}}(X_k | \Phi_k^{(m)}, X_k(0) = i_k)$$

$$= (\pi_{i_k} s_{i_k}^{(m)})^{B_{i_k}^{(k)}} f_{\boldsymbol{\theta}}(X_k | \Phi_k^{(m)}, X_k(0) = i_k).$$ (3.4)

Conditional on knowing $\Phi_k^{(m)}$, $f_{\boldsymbol{\theta}}(X_k | \Phi_k^{(m)}, X_k(0) = i_k)$ represents the likelihood function of observing the sample paths of X_k under the Markov process $X^{(m)}$

for which the likelihood is given by (e.g. Albert [1], Basawa and Rao [2]):

$$f_{\boldsymbol{\theta}}\big(X_k | \Phi_k^{(m)}, X_k(0) = i_k\big) = \prod_{i=1}^{p} \prod_{j \neq i}^{p} \big(q_{ij}^{(m)}\big)^{N_{ij}^{(k)}} e^{-q_{ij}^{(m)} Z_i^{(k)}}.$$

Hence, following (3.4), the likelihood contribution $L_k^{(m)}$ of $(X_k, \Phi_k^{(m)})$ is given by

$$L_k^{(m)} = \prod_{i=1}^{p} (s_i^{(m)} \pi_i)^{B_i^{(k)}} \prod_{i=1}^{p} \prod_{j \neq i}^{p} \big(q_{ij}^{(m)}\big)^{N_{ij}^{(k)}} e^{-q_{ij}^{(m)} Z_i^{(k)}}.$$

The likelihood contribution of all realizations $\{X_k\}$ is therefore given by

$$L = \prod_{k=1}^{N} \prod_{m=1}^{M} \big[L_k^{(m)}\big]^{\Phi_k^{(m)}}. \tag{3.5}$$

Notice that the likelihood (3.5) reduces to the one given in [1] and [2] when $s_i^{(m)} = 1$ for all $i \in \mathbb{S}$ and m, and $M = 1$, which implies $\Phi_k^{(m)} = 1$ and $q_{ij}^{(m)} = q_{ij}$ for all $i, j \in \mathbb{S}$. Equivalently in terms of the log-likelihood function, we have

$$
\begin{aligned}
\log L &= \sum_{k=1}^{N} \sum_{m=1}^{M} \Phi_k^{(m)} \log L_k^{(m)} \\
&= \sum_{k=1}^{N} \sum_{m=1}^{M} \Phi_k^{(m)} \Big[\sum_{i=1}^{p} B_i^{(k)} \log \big(s_i^{(m)} \pi_i\big) \\
&\qquad + \sum_{i=1}^{p} \sum_{j \neq i}^{p} N_{ij}^{(k)} \log q_{ij}^{(m)} - \sum_{i=1}^{p} \sum_{j \neq i}^{p} q_{ij}^{(m)} Z_i^{(k)} \Big].
\end{aligned}
\tag{3.6}
$$

Recall that the probability $\{\pi_i\}$ and $\{s_i^{(m)}\}$ satisfy the constraint, see (1.3):

$$\sum_{i}^{p} \pi_i = 1 \quad \text{and} \quad \sum_{m=1}^{M} s_i^{(m)} = 1, \ \text{ for all } i = 1, \ldots, p. \tag{3.7}$$

The maximum likelihood estimators for π_i, $q_{ij}^{(m)}$, $q_i^{(m)}$, and $s_i^{(m)}$ are given explicitly in terms of the statistics (3.1). The results are summarized below.

Theorem 3.1 *The maximum likelihood estimates of π_i, $q_{ij}^{(m)}$, $q_i^{(m)}$, and $s_i^{(m)}$ are*

$$\widehat{\pi}_{N,i} = \frac{1}{N} \sum_{k=1}^{N} B_i^{(k)}, \tag{3.8}$$

$$\widehat{q}_{N,ij}^{(m)} = \frac{\sum_{k=1}^{N} \Phi_k^{(m)} N_{ij}^{(k)}}{\sum_{k=1}^{N} \Phi_k^{(m)} Z_i^{(k)}}, \tag{3.9}$$

$$\widehat{q}_{N,i}^{(m)} = \frac{\sum_{k=1}^{N} \Phi_k^{(m)} N_i^{(k)}}{\sum_{k=1}^{N} \Phi_k^{(m)} Z_i^{(k)}}, \tag{3.10}$$

$$\widehat{s}_{N,i}^{(m)} = \frac{\sum_{k=1}^{N} \Phi_k^{(m)} B_i^{(k)}}{\sum_{k=1}^{N} B_i^{(k)}}. \tag{3.11}$$

Proof To find the estimators $\widehat{q}_{N,ij}^{(m)}$, $\widehat{q}_{N,i}^{(m)}$, $\widehat{s}_{N,i}^{(m)}$ and $\widehat{\pi}_{N,i}$ of the parameters distribution of the mixture process X, we introduce the Lagrangian function:

$$\mathcal{L} = \log L - \lambda \Big(\sum_{i=1}^{p} \pi_i - 1 \Big) - \sum_{i=1}^{p} \gamma_i \Big(\sum_{m=1}^{M} s_i^{(m)} - 1 \Big),$$

where λ and γ_i are the corresponding Lagrange multipliers of the constraint (3.7). Applying the first order Euler condition w.r.t π_i to the Lagrangian $\mathcal{L}$,

$$\frac{\partial \mathcal{L}}{\partial \pi_i} = \sum_{k=1}^{N} \sum_{m=1}^{M} \Phi_k^{(m)} \frac{B_i^{(k)}}{\pi_i} - \lambda = 0 \implies \lambda = N,$$

taking the note on (3.2), (3.3) and (3.7). The estimator $\widehat{\pi}_{N,i}$ is given by (3.8).

To get the estimator $\widehat{q}_{N,ij}^{(m)}$, we set the following Euler equation:

$$\frac{\partial \mathcal{L}}{\partial q_{ij}^{(m)}} = \sum_{k=1}^{N} \Phi_k^{(m)} \Big(\frac{N_{ij}^{(k)}}{q_{ij}^{(m)}} - Z_i^{(k)} \Big) = 0,$$

solving which for $q_{ij}^{(m)}$ gives the estimator $\widehat{q}_{N,ij}^{(m)}$ defined by (3.9).

Given that $q_i^{(m)}$ satisfies (1.2), we have following (3.1), $\widehat{q}_{N,i}^{(m)} = \sum_{j \neq i} \widehat{q}_{N,ij}^{(m)}$.

Finally, we set the following Euler equation for $s_i^{(m)}$:

$$\frac{\partial \mathcal{L}}{\partial s_i^{(m)}} = \sum_{k=1}^{N} \Phi_k^{(m)} \frac{B_i^{(k)}}{s_i^{(m)}} - \gamma_i = 0 \implies \gamma_i = \sum_{k=1}^{N} B_i^{(k)},$$

on account of (3.2) and (3.7). Hence, the estimator $\widehat{s}_i^{(m)}$ is given by (3.11). $\square$

3.1.1   Restricted mixture of Markov jump processes

The mixture model and its EM estimation were first discussed in [6] for a special class of mixture process in which case the element $q_{ij}^{(m)}$ of the intensity matrix $\mathbf{Q}^{(m)}$ is defined by $q_{ij}^{(m)} = \psi_i^{(m)} q_{ij}$ implying that each underlying Markov jump process has the same probability of leaving a state, i.e., $\pi_{ij}^{(m)} = \pi_{ij}$. For simplicity, we set following [6] $\psi^{(M)} = 1$. As a result, the maximum likelihood estimators of q_{ij} and $\psi_i^{(m)}$, for $m = 1, \ldots, M - 1$, $j \neq i$, are given following (3.6) and (3.7) by

$$\widehat{q}_i = \frac{\sum_{k=1}^{N} \Phi_k^{(M)} N_i^{(k)}}{\sum_{k=1}^{N} \Phi_k^{(M)} Z_i^{(k)}}, \tag{3.12}$$

$$\widehat{\psi}_i^{(m)} = \frac{\sum_{k=1}^{N} \Phi_k^{(m)} N_i^{(k)}}{\widehat{q}_i \sum_{k=1}^{N} \Phi_k^{(m)} Z_i^{(k)}}, \tag{3.13}$$

$$\widehat{\pi}_{ij} = \frac{\sum_{k=1}^{N} N_{ij}^{(k)}}{\sum_{k=1}^{N} N_i^{(k)}},$$

$$\widehat{q}_{ij} = \widehat{\pi}_{ij} \widehat{q}_i,$$

$$\widehat{q}_{ij}^{(m)} = \widehat{\psi}_i^{(m)} \widehat{q}_{ij}, \tag{3.14}$$

while the estimators $\widehat{s}_i^{(m)}$ and $\widehat{\pi}_i$ are the same as (3.8) and (3.11), respectively.

3.2   Consistency of the MLE estimators

To establish consistency of the MLE estimators (3.8)-(3.11), the following results are required. For convenience, we write $\Phi^{(m)} := \Phi_1^{(m)}$, $N_{ij} := N_{ij}^{(1)}$ and $Z_i := Z_i^{(1)}$.

Lemma 3.2 *For given $i, j = 1, \ldots, p$ and $m = 1, \ldots, M$, we have*

$$\mathbb{E}\{\Phi^{(m)} N_{ij}\} = q_{ij}^{(m)} \int_0^T \mathbb{P}\{X(u) = i, \Phi^{(m)} = 1\} du, \tag{3.15}$$

$$\mathbb{E}\{\Phi^{(m)} Z_i\} = \int_0^T \mathbb{P}\{X(u) = i, \Phi^{(m)} = 1\} du. \tag{3.16}$$

Proof Recall that $\Phi^{(m)} N_{ij}$ counts the number of transition of X (1.1), which is driven by the underlying Markov process $X^{(m)}$, between state i and j over the period of time $[0, T]$. The proof of (3.15) can follow similar approach to the proof of Theorem 5.1(a) in [1]. To be more precise, since $\Phi^{(m)} N_{ij} =$

$\sum_{k=1}^{N} \mathbf{1}_{\{X(kh)=j, X((k-1)h)=i, \Phi^{(m)}=1\}}$, then by the Bayes' formula we obtain,

$$\mathbb{E}\{\Phi^{(m)} N_{ij}\} = \sum_{k=1}^{N} \mathbb{P}\{X(kh) = j, X((k-1)h) = i, \Phi^{(m)} = 1\}$$

$$= \sum_{k=1}^{N} \mathbb{P}\{X(kh) = j | X((k-1)h) = i, \Phi^{(m)} = 1\} \mathbb{P}\{X((k-1)h) = i, \Phi^{(m)} = 1\}$$

$$= q_{ij}^{(m)} \sum_{k=1}^{N} \mathbb{P}\{X((k-1)h) = i, \Phi^{(m)} = 1\} h + o(h)$$

$$\xrightarrow{N \to \infty} q_{ij}^{(m)} \int_{0}^{T} \mathbb{P}\{X(u) = i, \Phi^{(m)} = 1\} du,$$

where the limit is due to Lebesgue dominated convergence theorem. (3.16) follows given that $\mathbb{E}\{\int_{0}^{T} |\mathbb{1}_{\{X(u)=i, \Phi^{(m)}=1\}}| du\} \leq T < \infty$, by which the claim follows from applying Fubini's theorem to the expectation $\mathbb{E}\{\int_{0}^{T} \mathbb{1}_{\{X(u)=i, \Phi^{(m)}=1\}} du\}$. $\qquad \square$

Thanks to the results of Lemma 3.2, consistency of the estimators follows.

Theorem 3.3 *If for $i = 1, \ldots, p$ and $t \geq 0$, $\mathbb{P}\{X(t) = i\} > 0$, then*

$$\lim_{N \to \infty} \widehat{q}_{N,ij}^{(m)} = q_{ij}^{(m)}, \qquad \lim_{N \to \infty} \widehat{q}_{N,i}^{(m)} = q_{i}^{(m)},$$

$$\lim_{N \to \infty} \widehat{s}_{N,i}^{(m)} = s_{i}^{(m)}, \qquad \lim_{N \to \infty} \widehat{\pi}_{N,i} = \pi_{i}.$$

with probability one, for $i, j = i, \ldots, p$ and $m = 1, \ldots, M$.

Proof The proof is based on applying the law of large numbers and continuous mapping theorem applied to independent paired observations $(X_k, \Phi_k^{(m)})$.

For convenience, we write $B_i := B_i^{(1)}$ and $N_i := N_i^{(1)}$. Recall that

$$\mathbb{E}\{B_i\} = \mathbb{P}\{X(0) = i\} \quad \text{and} \quad \mathbb{E}\{\Phi^{(m)} B_i\} = \mathbb{P}\{X(0) = i, \Phi^{(m)} = 1\}.$$

Furthermore, if $\mathbb{P}\{X(t) = i\} > 0$ for $t \geq 0$ and $i \in \mathbb{S}$, then by Lemma 3.2,

$$\lim_{N \to \infty} \widehat{q}_{N,ij}^{(m)} = \lim_{N \to \infty} \frac{N^{-1} \sum_{k=1}^{N} \Phi_k^{(m)} N_{ij}^{(k)}}{N^{-1} \sum_{k=1}^{N} \Phi_k^{(m)} Z_i^{(k)}} = \frac{\mathbb{E}\{\Phi^{(m)} N_{ij}\}}{\mathbb{E}\{\Phi^{(m)} Z_i\}} = q_{ij}^{(m)},$$

which in turn implies following (1.2) that $\lim_{N \to \infty} \widehat{q}_{N,i}^{(m)} = q_i^{(m)}$. Moreover, we have

$$\lim_{N \to \infty} \widehat{s}_{N,i}^{(m)} = \lim_{N \to \infty} \frac{N^{-1} \sum_{k=1}^{N} \Phi_k^{(m)} B_i^{(k)}}{N^{-1} \sum_{k=1}^{N} B_i^{(k)}} = \frac{\mathbb{E}\{\Phi^{(m)} B_i\}}{\mathbb{E}\{B_i\}} = s_i^{(m)},$$

where the last equality is due to applying the Bayes' formula for conditional probability, i.e., $\mathbb{P}\{\Phi^{(m)} = 1 | X(0) = i\} = \mathbb{P}\{X(0) = i, \Phi^{(m)} = 1\} / \mathbb{P}\{X(0) = i\}$.

Finally, we have $\lim_{N \to \infty} N^{-1} \sum_{k=1}^{N} B_i^{(k)} = \mathbb{E}\{B_i\} = \pi_i$. All limits hold with probability one, due to the law of large numbers. These justify the claim. $\qquad \square$

4 Estimation with incomplete information

4.1 The EM algorithm

Note that under complete observations, the estimators $\widehat{q}_{N,ij}^{(m)}$, $\widehat{q}_{N,i}^{(m)}$, and $\widehat{s}_{N,i}^{(m)}$ of the distribution parameters of the process X (1.1) are given based on knowing the random variable $\Phi_k^{(m)}$ which provides information about the underlying process that drives the kth realization of X. Under incomplete information, where we only know the realizations $\{\mathbf{X}_k\}$ of X, we need to replace $\Phi_k^{(m)}$ by the corresponding estimator $\widehat{\Phi}_k^{(m)}$ based on the observed sample $\{\mathbf{X}_k\}$ of the process.

For this purpose, we apply the EM algorithm, see Dempster et al. [4] and McLachlan and Krishnan [8] for more details. The first step of the iteration, the E−step, consists of calculating the conditional expectation of the sufficient statistics $\Phi_k^{(m)}N_{ij}^{(k)}$, $\Phi_k^{(m)}N_i^{(k)}$, $\Phi_k^{(m)}Z_i^{(k)}$, and $\Phi_k^{(m)}B_i^{(k)}$, given the sample $\{\mathbf{X}_k\}$. Then, in the M−step, the log-likelihood (3.6) is maximized, using the conditional expectation of the sufficient statistics as its observed value. The new estimates of the parameters are given by replacing the statistics in the estimators (3.8)-(3.11) by their corresponding conditional expectations evaluated at the E−step.

The EM algorithm is given below, which follows by an adaptation of [6].

(1) **Step 1**. Choose initial values of the distribution parameters $\pi_{0,i}$, $q_{0,ij}^{(m)}$, $q_{0,i}^{(m)}$ and $s_{0,i}^{(m)}$ for $i, j = 1, \ldots, p$, and $m = 1, ..., M$, all denoted by a vector $\boldsymbol{\theta}_0$.

> **Remark 4.1** *Note that the estimator $\widehat{\pi}_i$ of the distribution π_i does not get updated at each iteration. It is estimated separately by $\widehat{\pi}_{N,i}$ (3.8).*

(2) **Step 2 (E-step)** For the kth realization $\mathbf{X}_k$ of X, recall that

$$
\begin{aligned}
f_{\boldsymbol{\theta}_0}(\Phi_k^{(m)} = 1, \mathbf{X}_k) =& f_{\boldsymbol{\theta}_0}(X_k(0) = i) f_{\boldsymbol{\theta}_0}(\Phi_k^{(m)} = 1 | X_k(0) = i) \\
& \times f_{\boldsymbol{\theta}_0}(\mathbf{X}_k | \Phi_k^{(m)} = 1, X_k(0) = i) \\
=& \big(\pi_{0,i} s_{0,i}^{(m)}\big)^{B_i^{(k)}} f_{\boldsymbol{\theta}_0}(\mathbf{X}_k | \Phi_k^{(m)} = 1, X_k(0) = i) \\
=& \big(\pi_{0,i} s_{0,i}^{(m)}\big)^{B_i^{(k)}} \prod_{i=1}^{p} \prod_{j \neq i}^{p} \big(q_{0,ij}^{(m)}\big)^{N_{ij}^{(k)}} e^{-q_{0,ij}^{(m)} Z_i^{(k)}}.
\end{aligned}
$$

The estimate $\widehat{\Phi}_k^{(m)}$ of the random variable $\Phi_k^{(m)}$ is defined by

$$
\widehat{\Phi}_k^{(m)} = \mathbb{E}_{\boldsymbol{\theta}_0}\{\Phi_k^{(m)} | \mathbf{X}_k\}.
$$

Given that $\Phi_k^{(m)}$ is a Bernoulli random variable, $\widehat{\Phi}_k^{(m)}$ is given by

$$
\begin{aligned}
\widehat{\Phi}_k^{(m)} &= f_{\boldsymbol{\theta}_0}(\Phi_k^{(m)} = 1 | \mathbf{X}_k) \\
&= \frac{f_{\boldsymbol{\theta}_0}(\Phi_k^{(m)} = 1, \mathbf{X}_k)}{\sum_{m=1}^{M} f_{\boldsymbol{\theta}_0}(\Phi_k^{(m)} = 1, \mathbf{X}_k)} \\
&= \frac{\prod_{i=1}^{p} (\pi_{0,i} s_{0,i}^{(m)})^{B_i^{(k)}} \prod_{i=1}^{p} \prod_{j \neq i}^{p} (q_{0,ij}^{(m)})^{N_{ij}^{(k)}} e^{-q_{0,ij}^{(m)} Z_i^{(k)}}}{\sum_{m=1}^{M} \prod_{i=1}^{p} (\pi_{0,i} s_{0,i}^{(m)})^{B_i^{(k)}} \prod_{i=1}^{p} \prod_{j \neq i}^{p} (q_{0,ij}^{(m)})^{N_{ij}^{(k)}} e^{-q_{0,ij}^{(m)} Z_i^{(k)}}}.
\end{aligned}
$$

Remark 4.2 *It is straightforward to check that the estimate $\widehat{\Phi}_k^{(m)}$ satisfies, for each iteration k, $\sum_{m=1}^{M} \widehat{\Phi}_k^{(m)} = 1$, improving that of given in [6].*

Then, for $i, j = 1, \ldots, p$ and $m = 1, \ldots, M$, compute the conditional expectation of the sufficient statistics $\Phi_k^{(m)} N_{ij}^{(k)}$, $\Phi_k^{(m)} N_i^{(k)}$, $\Phi_k^{(m)} Z_i^{(k)}$, and $\Phi_k^{(m)} B_i^{(k)}$, given the sample $\{\mathbf{X}_k\}$. Note that the random variables $N_{ij}^{(k)}$, $N_i^{(k)}$, and $B_i^{(k)}$ are all adapted to the information set generated by $\{\mathbf{X}_k\}$.

$$
\mathbb{E}_{\boldsymbol{\theta}_0}\{\Phi_k^{(m)} N_{ij}^{(k)} | \mathbf{X}_k\} = \widehat{\Phi}_k^{(m)} N_{ij}^{(k)}
$$

$$
\mathbb{E}_{\boldsymbol{\theta}_0}\{\Phi_k^{(m)} N_i^{(k)} | \mathbf{X}_k\} = \widehat{\Phi}_k^{(m)} N_i^{(k)}
$$

$$
\mathbb{E}_{\boldsymbol{\theta}_0}\{\Phi_k^{(m)} Z_i^{(k)} | \mathbf{X}_k\} = \widehat{\Phi}_k^{(m)} Z_i^{(k)}
$$

$$
\mathbb{E}_{\boldsymbol{\theta}_0}\{\Phi_k^{(m)} B_i^{(k)} | \mathbf{X}_k\} = \widehat{\Phi}_k^{(m)} B_i^{(k)}.
$$

(3) **Step 3 (M-step)** Compute the new values $\pi_{1,i}$, $q_{1,ij}^{(m)}$, $q_{1,i}^{(m)}$ and $s_{1,i}^{(m)}$ for $i, j = 1, \ldots, p$, and $m = 1, \ldots, M$, using (3.9), (3.10), and (3.11) by

$$
q_{1,ij}^{(m)} = \frac{\sum_{k=1}^{N} \widehat{\Phi}_k^{(m)} N_{ij}^{(k)}}{\sum_{k=1}^{N} \widehat{\Phi}_k^{(m)} Z_i^{(k)}},
$$

$$
q_{1,i}^{(m)} = \frac{\sum_{k=1}^{N} \widehat{\Phi}_k^{(m)} N_i^{(k)}}{\sum_{k=1}^{N} \widehat{\Phi}_k^{(m)} Z_i^{(k)}},
$$

$$
s_{1,i}^{(m)} = \frac{\sum_{k=1}^{N} \widehat{\Phi}_k^{(m)} B_i^{(k)}}{\sum_{k=1}^{N} B_i^{(k)}}.
$$

Notice that we have replaced the random variable $\Phi_k^{(m)}$ by its respective estimate $\widehat{\Phi}_k^{(m)}$. Stack all the updated estimates into a new vector $\boldsymbol{\theta}_1$.

Remark 4.3 *For the restricted mixture, the new update $s_{1,i}^{(m)}$ is the same as above. However, the updates $\psi_{1,i}^{(m)}$ and $q_{1,ij}$ are given for $j \neq i$ by*

$$q_{1,i} = \frac{\sum_{k=1}^{N} \widehat{\Phi}_k^{(M)} N_i^{(k)}}{\sum_{k=1}^{N} \widehat{\Phi}_k^{(M)} Z_i^{(k)}},$$

$$\psi_{1,i}^{(m)} = \frac{\sum_{k=1}^{N} \widehat{\Phi}_k^{(m)} N_i^{(k)}}{q_{1,i} \sum_{k=1}^{N} \widehat{\Phi}_k^{(m)} Z_i^{(k)}},$$

$$\pi_{1,ij} = \frac{\sum_{k=1}^{N} N_{ij}^{(k)}}{\sum_{k=1}^{N} N_i^{(k)}},$$

$$q_{1,ij} = \pi_{1,ij} q_{1,i},$$

$$q_{1,ij}^{(m)} = \psi_{1,i}^{(m)} q_{1,ij}.$$

(4) **Step 4** Stop if the convergence criterion is achieved. Otherwise, return to Step 1 by replacing $q_{0,ij}^{(m)}$, $q_{0,i}^{(m)}$ and $s_{0,i}^{(m)}$ for $i, j = 1, \ldots, p$, and $m = 1, ..., M$, correspondingly by $q_{1,ij}^{(m)}$, $q_{1,i}^{(m)}$ and $s_{1,i}^{(m)}$ for $i, j = 1, \ldots, p$, and $m = 1, ..., M$.

For example, the convergence criterion is satisfied when the difference between the updated value $\boldsymbol{\theta}_1$ of each parameter and its previous value $\boldsymbol{\theta}_0$ is less than a specified small positive number, say ϵ, i.e., $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\| \leq \epsilon$.

5   Numerical examples

To test the proposed statistical estimation method, we use Monte Carlo simulation to generate sample paths of the mixture process (1.1) for a predetermined parameter values. Based on the simulated sample paths of the process, we attempt to re-estimate the distribution parameter values using the EM algorithm.

For simplicity, we consider a mixture of two Markov jump processes $X^{(1)}$ and $X^{(2)}$ on the finite state space $\mathbb{S} = \{1, 2, 3, 4\}$. Parameter values set for the simulation, which include the initial probability distribution $\boldsymbol{\pi}$ and the speed regime switching probabilities $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$, are presented in the Table 1.

State (i)	π_i	$s_i^{(1)}$	$s_i^{(2)}$
1	0.1	0.4	0.6
2	0.2	0.5	0.5
3	0.3	0.6	0.4
4	0.4	0.7	0.3

Table 1: Initial and switching probability $\boldsymbol{\pi}$, $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$, respectively.

Intensity matrices of the Markov processes $X^{(1)}$ and $X^{(2)}$ are given resp. by

$$\mathbf{Q}^{(1)} = \begin{pmatrix} -0.9632 & 0.3312 & 0.3127 & 0.3193 \\ 0.3758 & -0.9101 & 0.3660 & 0.1683 \\ 0.4107 & 0.4378 & -0.9994 & 0.1509 \\ 0.2675 & 0.1543 & 0.2010 & -0.6228 \end{pmatrix},$$

and

$$\mathbf{Q}^{(2)} = \begin{pmatrix} -0.9519 & 0.4795 & 0.3039 & 0.1685 \\ 0.0746 & -0.5499 & 0.2335 & 0.2418 \\ 0.1913 & 0.0667 & -0.2752 & 0.0172 \\ 0.1627 & 0.1663 & 0.2614 & -0.5904 \end{pmatrix}.$$

5.1   The EM estimation

We generate $N = 100,000$ independent sample paths of the mixture process X making transitions on the interval $[0, T]$, with $T = 100$. Simulation results on five randomly selected sample paths are displayed in Figure 2. The initial parameter values for $\boldsymbol{\pi}_0$ is set to be equal to the estimate $\widehat{\boldsymbol{\pi}}$, see Table 2 below, while the switching probabilities $\mathbf{s}_0^{(1)}$ and $\mathbf{s}_0^{(2)}$ are chosen randomly on $[0, 1]$, whereas the intensity matrices $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ are set to be equal to the intensity matrix

$$\mathbf{Q} = \begin{pmatrix} -0.9533 & 0.3723 & 0.3074 & 0.2736 \\ 0.2530 & -0.7609 & 0.3084 & 0.1995 \\ 0.2723 & 0.2060 & -0.5456 & 0.0672 \\ 0.2320 & 0.1557 & 0.2164 & -0.6041 \end{pmatrix}, \tag{5.1}$$

assuming that $\{\mathbf{X}_k\}$ was Markov process. Based on the sample paths, we obtain:

State (i)	$\widehat{\pi}_i$	$\widehat{s}_i^{(1)}$	$\widehat{s}_i^{(2)}$
1	0.1023	0.3938	0.6062
2	0.2045	0.4824	0.5176
3	0.2990	0.5300	0.4700
4	0.3942	0.6778	0.3222

Table 2: Estimates of initial and switching probability $\boldsymbol{\pi}$, $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$, resp.

It is straightforward to check that $\widehat{s}_i^{(1)} + \widehat{s}_i^{(2)} = 1$ for all $i = 1, \ldots, p$, see (1.3).

The estimate of the intensity matrix $\mathbf{Q}^{(1)}$ of the Markov process $X^{(1)}$ is

$$\widehat{\mathbf{Q}}^{(1)} = \begin{pmatrix} -0.9517 & 0.3268 & 0.3092 & 0.3157 \\ 0.3793 & -0.9114 & 0.3624 & 0.1697 \\ 0.4086 & 0.4371 & -0.9961 & 0.1503 \\ 0.2650 & 0.1520 & 0.1996 & -0.6166 \end{pmatrix},$$

whereas the estimate of the intensity matrix $\mathbf{Q}^{(2)}$ of $X^{(2)}$ is found to be

$$\widehat{\mathbf{Q}}^{(2)} = \begin{pmatrix} -0.9574 & 0.4856 & 0.3030 & 0.1687 \\ 0.0748 & -0.5486 & 0.2323 & 0.2415 \\ 0.1898 & 0.0661 & -0.2728 & 0.0169 \\ 0.1585 & 0.1639 & 0.2537 & -0.5760 \end{pmatrix}.$$

From the EM estimation outcomes, we observe that the estimates are reasonably close enough to the true values of the distribution parameters.

5.1.1   Estimation based on the restricted mixture

The EM estimation of the distribution parameters is based on the restricted model $\mathbf{Q}^{(1)} = \mathbf{\Psi}\mathbf{Q}^{(2)}$, with $\mathbf{\Psi} = \text{diag}(\psi_1, \ldots, \psi_p)$. The initial condition for $\mathbf{Q}^{(2)}$ is defined by the matrix $\mathbf{Q}$ (5.1), whilst ψ_i, the initial distribution π_i, $s_i^{(1)}$ and $s_i^{(2)}$, for $i = 1, \ldots, p$, are all chosen randomly on the unit interval $[0, 1]$.

State (i)	$\widehat{\pi}_i$	$\widehat{s}_i^{(1)}$	$\widehat{s}_i^{(2)}$
1	0.1023	0.6059	0.3941
2	0.2045	0.5191	0.4809
3	0.2990	0.4700	0.5300
4	0.3942	0.3223	0.6777

Table 3: Estimates of initial and switching probability $\boldsymbol{\pi}$, $\mathbf{s}^{(1)}$ and $\mathbf{s}^{(2)}$, resp.

It is straightforward to check that $\widehat{s}_i^{(1)} + \widehat{s}_i^{(2)} = 1$ for all $i = 1, \ldots, p$, see (1.3). The estimate of the intensity matrix $\mathbf{Q}^{(1)}$ of the Markov process $X^{(1)}$ is

$$\widehat{\mathbf{Q}}_{\text{res}}^{(1)} = \begin{pmatrix} -0.9577 & 0.3741 & 0.3088 & 0.2748 \\ 0.1825 & -0.5489 & 0.2225 & 0.1439 \\ 0.1362 & 0.1031 & -0.2729 & 0.0336 \\ 0.2212 & 0.1484 & 0.2063 & -0.5759 \end{pmatrix},$$

whereas the estimate of the intensity matrix $\mathbf{Q}^{(2)}$ of $X^{(2)}$ is found to be

$$\widehat{\mathbf{Q}}_{\text{rest}}^{(2)} = \begin{pmatrix} -0.9515 & 0.3717 & 0.3069 & 0.2730 \\ 0.3031 & -0.9114 & 0.3694 & 0.2389 \\ 0.4973 & 0.3763 & -0.9963 & 0.1228 \\ 0.2369 & 0.1589 & 0.2209 & -0.6167 \end{pmatrix},$$

with the speed reference variable $\widehat{\mathbf{\Psi}} = \text{diag}(1.0065, 0.6022, 0.2739, 0.9339)$ satisfying the constraint $\widehat{\mathbf{Q}}^{(1)} = \widehat{\mathbf{\Psi}}\widehat{\mathbf{Q}}^{(2)}$. As we can see, the EM estimations for the unrestricted mixture outperform that of for the restricted mixture model [6].

In the section below a statistical test is performed to compare the statistical significance of the Markov model against the Markov mixture model, and the restricted mixture model [6] against the unrestricted model at a certain level.

5.2   Likelihood ratio test

To test the hypothesis $H_0 : q_{i,j}^{(m)} = q_{ij}$ for $i, j = 1, \ldots, p$ and $m = 1, \ldots, M$ that the simulated processes is driven by a Markov jump process against the alternative hypothesis $H_1 : q_{i,j}^{(m)} \neq q_{ij}$ that it is a mixture of M Markov jump processes, we apply the likelihood ratio test by adapting the one presented in [6]. The test statistic is described as follows. Under the H_0 hypothesis, the likelihood of observing $N-$independent realizations of the sample paths of X is given by

$$L_{\text{Markov}}(\boldsymbol{\pi}, \mathbf{Q}) = \prod_{k=1}^{N} f_{\boldsymbol{\theta}_0}(\mathbf{X}_k) = \prod_{k=1}^{N} \prod_{i=1}^{p} (\pi_i)^{B_i^{(k)}} \prod_{i=1}^{p} \prod_{j \neq i}^{p} (q_{ij})^{N_{ij}^{(k)}} e^{-q_{ij} Z_i^{(k)}},$$

where $\mathbf{Q}$ denotes the intensity matrix of a Markov jump process and $Z_i^{(k)}$ is the total time the $k-$th realization X_k stays in state i. For the mixture model, the likelihood contribution of observing the sample paths $\mathbf{X}_k$ is given by $f_{\boldsymbol{\theta}_1}(\mathbf{X}_k) = \sum_{m=1}^{M} f_{\boldsymbol{\theta}_1}(\mathbf{X}_k, \Phi_k^{(m)})$. Thus, the likelihood under the mixture model is given by

$$L_{\text{Mixture}}\big(\boldsymbol{\pi}, \mathbf{Q}^{(m)}, \mathbf{S}^{(m)}, m = 1, \ldots, M\big) = \prod_{k=1}^{N} f_{\boldsymbol{\theta}_1}(\mathbf{X}_k)$$

$$= \prod_{k=1}^{N} \Big(\sum_{m=1}^{M} \prod_{i=1}^{p} (s_i^{(m)} \pi_i)^{B_i^{(k)}} \prod_{i=1}^{p} \prod_{j \neq i}^{p} (q_{ij}^{(m)})^{N_{ij}^{(k)}} e^{-q_{ij}^{(m)} Z_i^{(k)}} \Big), \qquad (5.2)$$

where the product is for all realizations. The likelihood ratio statistic is given by

$$\boldsymbol{\Lambda}_1 = \frac{L_{\text{Markov}}(\widehat{\boldsymbol{\pi}}, \widehat{\mathbf{Q}})}{L_{\text{Mixture}}\big(\widehat{\boldsymbol{\pi}}, \widehat{\mathbf{Q}}^{(m)}, \widehat{\mathbf{S}}^{(m)}, m = 1, \ldots, M\big)}, \qquad (5.3)$$

where $\widehat{\boldsymbol{\pi}}, \widehat{\mathbf{Q}}, \widehat{\mathbf{Q}}^{(m)}$ and $\widehat{\mathbf{S}}^{(m)}$, with $m = 1, \ldots, M$, are the MLEs of $\boldsymbol{\pi}, \mathbf{Q}, \mathbf{Q}^{(m)}$ and $\mathbf{S}^{(m)}$, with $m = 1, \ldots, M$. Notice that as the two likelihoods share the same $\boldsymbol{\pi}$, the term $\prod_{k=1}^{N} \prod_{i=1}^{p} (\pi_i)^{B_i^{(k)}}$ gets canceled out from the likelihood ratio statistic. The statistic (5.3) is calculated for $\mathbf{Q}^{(m)} = \boldsymbol{\Psi}\mathbf{Q}$ and for the unrestricted model.

To test the significance of the restricted mixture model against the unrestricted mixture, we consider the following likelihood ratio test statistic

$$\boldsymbol{\Lambda}_2 = \frac{L_{\text{RestMixture}}\big(\widehat{\boldsymbol{\pi}}, \widehat{\mathbf{Q}}, \boldsymbol{\Psi}^{(m)}, \widehat{\mathbf{S}}^{(m)}, m = 1, \ldots, M\big)}{L_{\text{Mixture}}\big(\widehat{\boldsymbol{\pi}}, \widehat{\mathbf{Q}}^{(m)}, \widehat{\mathbf{S}}^{(m)}, m = 1, \ldots, M\big)}. \qquad (5.4)$$

Given that the entries of intensity matrices $\mathbf{Q}$, $\mathbf{Q}^{(m)}$ and $\mathbf{S}^{(m)}$ respectively satisfy the constraint (1.2) and (1.3), it is known by standard theory that under the null hypothesis $-2 \ln \boldsymbol{\Lambda}_1$ has χ_d^2-distribution with d.f. $d = p^2(m-1)$. Similarly, with the same arguments, under the null hypothesis $H_0 : \psi_i^{(m)} = 1, 1 \leq i \leq p, 1 \leq m \leq M-1$ versus $H_1 :$ at least one $\psi_i^{(m)} \neq 1, 1 \leq i \leq p, 1 \leq m \leq M-1$, the test statistic $-2 \ln \boldsymbol{\Lambda}_2$ has χ_d^2-distribution with d.f. $d = p(p-1)(m-1)$.

On account that (5.2) admits no closed-form solution for the MLE estimates of $\boldsymbol{\pi}$, $\mathbf{Q}^{(m)}$ and $\mathbf{S}^{(m)}$, with $m = 1, \ldots, M$, we therefore necessarily use the EM estimates in the test statistic (5.3) given that the estimation results are reasonably close enough to the actual parameter values of the mixture distribution.

Based on the MLE estimations, the likelihood ratio statistic $-2 \ln \boldsymbol{\Lambda}_1$ for comparing the Markov model against the alternative restricted mixture model and unrestricted mixture model has in each case the observe value $7.0408e + 05$ and $1.1220e + 06$, respectively. On the other hand, we have $-2 \ln \boldsymbol{\Lambda}_2 = 4.1795e + 05$. Each alternative is found to be significant at the level $\alpha = 5\%$. We therefore conclude that unrestricted mixture model is found to be statistically more significant at the level $\alpha = 5\%$ than the Markov and restricted mixture models.

6   Conclusions

We have developed tractable construction of a continuous-time stochastic process based on a finite mixture of right-continuous Markov jump processes moving at different speeds on the same finite state space. As discussed in more details in Frydman and Schuermann [5] and Surya ([11], [12]), that unlike the underlying Markov processes the mixture itself lacks stationarity and the Markov property.

Monte Carlo method for simulating the process was discussed along with proving distributional equivalence between the simulated process and the theoretical corresponding process. Maximum likelihood estimation was presented for complete and incomplete information. Under complete information, consistent estimators of the distribution parameters were obtained in closed form in terms of sufficient statistics of the process. The EM estimation was proposed for incomplete information knowing only the sample paths of the process.

Based on Monte Carlo simulation, the EM estimations for the unrestricted mixture were shown to be close enough to the actual value of the distribution parameters, and is found to be statistically significant based on the likelihood ratio test statistic at the level $\alpha = 5\%$ compared to the Markov model and restricted mixture model [6]. The results presented in this paper offer appealing features for various applications, for instance in estimating the distribution of first exit time to absorbing state of the mixture process, see for e.g. Surya [10].

7   Acknowledgments

References

[1] Albert A. (1961). Estimating the infinitesimal generator of a continuous time, finite state Markov process, *Ann. Math. Statist.*, **38**, p.727-753.

[2] Basawa, I.V. and Rao, B.L.S. (1980). *Statistical Inference for Stochastic Processes.* Academic Press, London.

[3] Blumen, I., Kogan, M. and McCarthy, P.J. (1955). The industrial mobility of labor as a probability process. *Cornell Stud. Ind. Labor Relat.*, Vol. 6, Ithaca, N.Y., Cornell University Press.

[4] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser.*, **39**, p.1-38.

[5] Frydman, H. and Schuermann, T. (2008). Credit rating dynamics and Markov mixture models, *J. Bank. Financ.*, **32**, 1062-1075.

[6] Frydman, H. (2005). Estimation in the mixture of Markov chains moving with different speeds. *J. Am. Stat. Assoc.*, **100**, 1046-1053.

[7] Frydman, H. (1984). Maximum likelihood estimation in the mover-stayer model, *J. Am. Stat. Assoc.*, **79**, 632-638.

[8] McLachlan, G.J. and Krishnan, T. (1997). *The EM Algorithm and Extensions.* John Wiley & Sons, Inc.

[9] Resnick, S. I. (2002). *Adventures in Stochastic Processes.* Birkh´auser.

[10] Surya, B. A. (2018). The EM estimation of the distribution of first exit time to absorbing state of the mixture of Markov jump processes. *Preprint.*

[11] Surya, B. A. (2018). Conditional joint probability distributions of first exit times to overlapping absorbing sets of the mixture of Markov jump processes. *Submitted for publication.*

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