Estimation of the Finite Mixture of Markov Jump Processes via the EM Algorithm

Budhi Surya and Halina Frydman

Victoria University of Wellington and New York University

budhi.surya@vuw.ac.nz

Wellington Workshop in Probability and Mathematical Statistics, 5-7/12

This work is motivated by the results in the following papers:

1. S. Asmussen, O. Nerman, & M. Olsson. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, 23:419-441. (Google scholar # citations: 737)
2. H. Frydman & T. Schuermann. (2008). Credit rating dynamics and Markov mixture models. *Journal of Banking and Finance* 32: 1062-1075.
3. H. Frydman. (2005). Estimation in the mixture of Markov chains moving with different speeds. *Journal of the American Statistical Association* 100: 1046-1053.

References:

1. H. Frydman & B. A. Surya. (2019). The finite mixture of Markov jump processes: Monte Carlo method and the EM estimation. *Working paper*.
2. B. A. Surya. (2019). Fitting exit time distributions of the finite mixture of Markov jump processes via the EM algorithm. *Working paper*
3. B. A. Surya. (2018). Distributional Properties of the Mixture of Continuous-Time Absorbing Markov Chains Moving at Different Speeds. *Stochastic Systems*, **8**, p. 29-44.

Outline

Outline

Markov jump processes in practice

Since their introduction by Markov in 1906, the Markov processes have been used in various fields, e.g.,

- **Finance/Credit Risk**: Jarrow & Turnbull [1] , Jarrow et.al. [12].
- **Actuarial science**: (Albrecher & Asmussen [4], Lee & Lin [3], Lin & Liu [4], and Rolski et al. [10]),
- **Option pricing**: (Asmussen et al. [5], Rolski et. al [10]),
- **Queueing theory**: (Badila et al. [9], Chakravarthy & Neuts [4], Buchholz et al. [3], Breuer & Baum [2], Asmussen [6]),
- **Reliability theory**: (Assaf & Levikson [8], Okamura & Dohi [7]),
- **Survival analysis, Biostatistics**: (Aalen [3], Aalen & Gjessing [2]).
- **Ecological modelling**: Balzter [10], etc.
- **Marketing**: Berger & Nasr [11], Pfeifer & Carraway [8].

Remarks

Markov model allows some analytically tractable results in applications.

Outline

Comparison between the Markov and mixture model

The reasons for the popularity of Markov jump (MJ) processes are

1. They have a simple structure
2. Their properties have been well studied
3. They serve as baseline models; the departure from an MJ model may indicate the type of a needed extension.

But

1. In MJ process, each individual evolves according to the same constant intensity matrix. Thus, it does not allow for population heterogeneity.
2. Because of the Markov property, the future evolution of an individual depends only on their current state, but not on their past history.

The mixture of MJ processes, defined below, allows for both population heterogeneity and for taking past history into account.

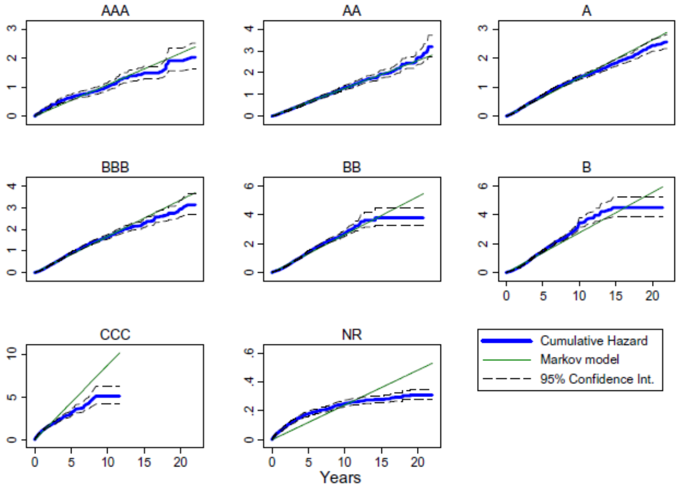


Figure: Nelson-Aalen plots for cumulative hazard rate (credit rating for S& P rated US firms, 1981-2002) under the Markov and mixture model. Source: Frydman & Schuermann, *Journal of Banking and Finance* (2008)

Outline

The finite mixture of Markov jump processes (MJPs)

Let $\{X_t^{(\phi)} : \phi = 1, \ldots, M, t \geq 0\}$ be finite-state right-continuous Markov jump processes with intensity matrices $\mathbf{Q}_\phi$ defined on state space $\mathbb{S} = E \cup \Delta$. Conditionally on $X_0 = i_0 \in E$, the mixture is defined by

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

There is a separate mixing probability for each initial state $X_0 = i_0 \in E$:

$$s_{i_0,m} = \mathbb{P}\{\phi = m | X_0 = i_0\}. \qquad (2)$$

Define Bernoulli r.v. $\Phi_m = \mathbb{1}_{\{\phi=m\}}$. It is clear that $\sum_{m=1}^{M} \Phi_m = 1$. Thus,

$$X_t = \sum_{m=1}^{M} \Phi_m X_t^{(m)}, \quad \text{for } t \geq 0. \qquad (3)$$

Parameters of underlying MJP $X^{(m)}$

Let $E = \{1, \ldots, n\}$ (transient states) and the (absorbing state) Δ be the $(n+1)-$state. The law of each underlying MJP $X^{(m)}$ is characterized by initial distribution $\tilde{\boldsymbol{\pi}} = (\boldsymbol{\pi}^\top, \pi_{n+1})^\top$, and a constant intensity matrix

$$\mathbf{Q}_m = \begin{pmatrix} \mathbf{T}_m & \boldsymbol{\delta}_m \\ \mathbf{0} & 0 \end{pmatrix} \quad \text{with} \quad \boldsymbol{\delta}_m = -\mathbf{T}_m \mathbf{1} > \mathbf{0}, \tag{4}$$

and $\mathbf{1} = (1, \ldots, 1)^\top$. Assume that $\pi_{n+1} = 0$ and $s_{n+1,m} = 0$.

Let $t_{ij,m} = [\mathbf{T}_m]_{i,j}$. By (4), $t_{ii,m} = -(\delta_{i,m} + \sum_{j=1, j\neq i}^{n} t_{ij,m})$.

Equivalently, the $(i,j)-$component of $\mathbf{Q}_m$ is defined for $(i,j) \in \mathbb{S}$ by

$$q_{ij,m} = \begin{cases} \lim_{h\downarrow 0} \frac{1}{h} \mathbb{P}\{X_h = j | \Phi_m = 1, X_0 = i\}, & \text{for } j \neq i \\ \lim_{h\downarrow 0} \frac{1}{h} (1 - \mathbb{P}\{X_h = j | \Phi_m = 1, X_0 = i\}), & \text{for } j = i \end{cases}$$

Parameters of underlying MJP $X^{(m)}$: cont'd

The transition matrix $[\mathbf{P}(t)]_{i,j} := \mathbb{P}\{X_t = j | X_0 = i\}$ of X is given by

$$\mathbf{P}(t) = \sum_{m=1}^{M} \widetilde{\mathbf{S}}_m e^{\mathbf{Q}_m t},$$

where $\widetilde{\mathbf{S}}_m = \mathrm{diag}(\mathbf{S}_m, 0)$ with $\mathbf{S}_m = \mathrm{diag}(s_{1,m}, \dots, s_{n,m})$. See, Frydman [10], Frydman & Schuermann [9] for the mixture on the speed of Markov chains, and Frydman & Surya [8] for a general mixture of Markov chains.

Define the following $i \to j$ transition probability

$$\pi_{ij,m} = \begin{cases} t_{ij,m}/t_{ii,m}, & i \in E, j \in \mathbb{S} \\ 0, & j = i, \text{ with } i,j \in E \\ 1, & j = i, \text{ with } i,j = \Delta \end{cases} \tag{5}$$

Parameters of underlying MJP $X^{(m)}$: cont'd

1. The matrix $\mathbf{\Pi}_m = (p_{ij,m})_{i,j}$ forms the transition probability matrix of the embedded Markov chain $Z_0^{(m)}, Z_1^{(m)}, \ldots, Z_{N-1}^{(m)}, Z_N^{(m)}$ of $X^{(m)}$ with N being the number of jumps until $X_t^{(m)}$ hits the state Δ.
2. The diagonal element $t_{ii,m}$ determines the speed of the MJP $X^{(m)}$. The sojourn time/length of stay W_i of $X^{(m)}$ in a state $i \in E$ has exponential distribution with intensity $t_{ii,m}$. Note that, the larger $t_{ii,m}$, the smaller $\mathbb{E}\{S_i\}$, the faster $X^{(m)}$ moves on the state $\mathbb{S}$.

Following (4) and (5), the two matrices $\mathbf{Q}_m$ and $\mathbf{\Pi}_m$ satisfy

$$\mathbf{Q}_m = \mathbf{\Psi}_m(\mathbf{\Pi}_m - \mathbf{I}),$$

where $\mathbf{\Psi}_m = \mathrm{diag}(t_{11,m}, \ldots, t_{nn,m}, 0)$ and $\mathbf{I}$ is an identity matrix.

Remarks

- *The model was first introduced by Blumen et al. [1] in 1955 as a stochastic model (mover-stayer model) for jobs mobility dynamics. It is a mixture of two discrete-time Markov chains with transition probability matrices specified by* $\mathbf{\Pi}_1 \neq \mathbf{I}$ *and* $\mathbf{\Pi}_2 = \mathbf{I}$.
- *Frydman [10], extended the model to a mixture on the speed of Markov chains, but with* $\mathbf{\Pi}_m = \mathbf{\Pi}$ *for all* $m = 1, \dots, M$*, and provided EM algorithm for the estimation. Frydman & Schuermann [9] applied the result for estimation of credit rating dynamics.*
- *Estimation for a general mixture of MJPs with different* $(\mathbf{\Psi}_m, \mathbf{\Pi}_m)$ *for* $m = 1, \dots, M$ *is proposed in Frydman & Surya [8].*

The above works were not specifically concerned with absorbing Markov chains. We are in particular interested in the distribution of the exit time:

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

State diagram of the finite mixture of MJPs

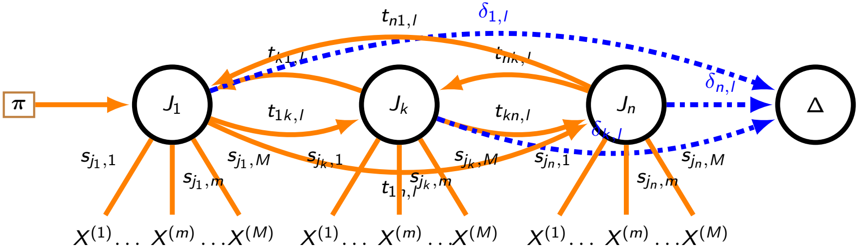
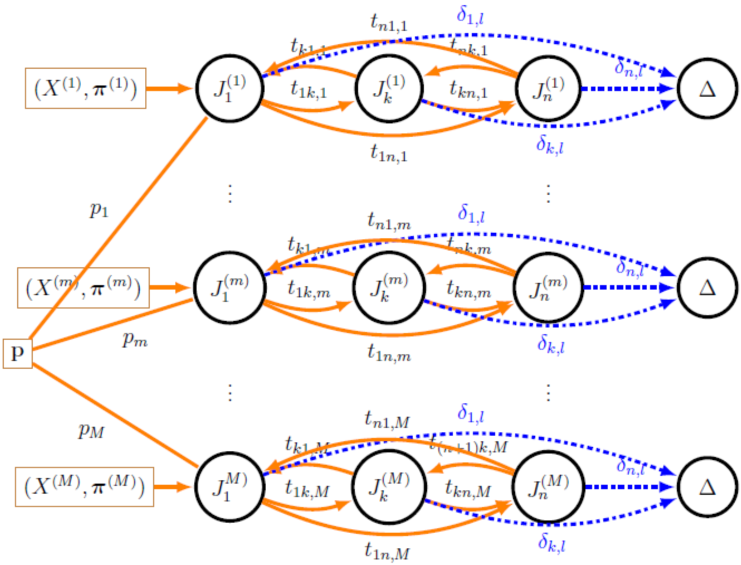


Figure: State diagram of X with M underlying processes $\{X^{(k)}\}$.

Monte Carlo simulation: The process X chooses an initial state $i \in E$ to start with at probability π_i and then selects underlying process $X^{(m)}$ with probability $s_{i,m}$ to make transition on $\mathbb{S}$ at the respective rate $\mathbf{T}_m$.
The law of the mixture process X is characterized by the parameters:

$$\boldsymbol{\theta} := (\boldsymbol{\pi}, \mathbf{T}_m, \mathbf{S}_m : m = 1, \dots, M).$$

Re-parameterization of the finite mixture of MJPs



Re-parameterization of the finite mixture of MJPs: cont'd

Re-parameterization of the mixture process in the second state diagram:

$$p_m \times \pi_i^{(m)} = \mathbb{P}\{\Phi_m = 1, X_0 = i\} = \pi_i \times s_{i,m}.$$

Monte Carlo simulation: The process X selects the regime $X^{(m)}$ with $p_m = \mathbb{P}\{\Phi_m = 1\}$ and chooses a state $i \in E$ to start with at probability $\pi_i^{(m)} = \mathbb{P}\{X_0 = i | \Phi_m = 1\}$ to make transition on $\mathbb{S}$ at the rate $\mathbf{T}_m$.

The law of the mixture process X is characterized by the parameters:

$$\boldsymbol{\theta} := (p_m, \boldsymbol{\pi}^{(m)}, \mathbf{T}_m : m = 1, \ldots, M).$$

The mixture process X (3) reduces to the simple MJP when either

1. intensity matrices $\{\mathbf{Q}_m\}$ are all the same for all $m = 1, \ldots, M$,
2. or $\mathbf{S}_m = \mathbf{I}$, and identity matrix, for all $m = 1, \ldots, M$.
3. The transient state is given by $E = \bigcup_{m=1}^{M} E_m$ with $E_k \cap E_m = \emptyset$.

Mixture process as a concatenation of MJPs

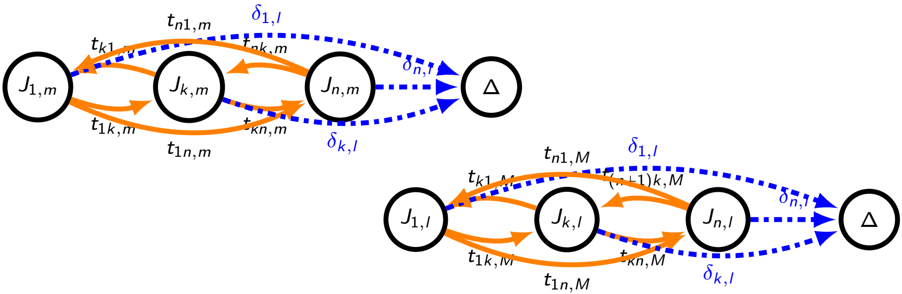


Figure: Concatenation of $\{X^{(m)}\}$. In this case $X^{(m)}$ is the only MJP occupying the state E_m.

$$s_i^{(m)} = \begin{cases} 1, & i \in E_m \\ 0, & i \notin E_m \end{cases}, \quad \pi_i^{(m)} = \begin{cases} \pi_i / p_m, & i \in E_m \\ 0, & i \notin E_m \end{cases}$$

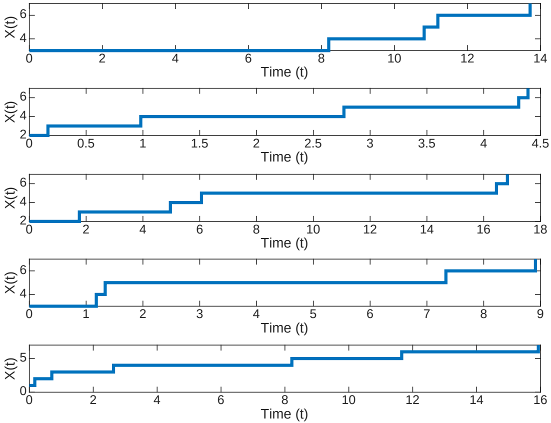


Figure: Mixture of two birth-death processes with $E = \{1, \ldots, 6\}$ and $\Delta = \{7\}$

$$f(t) = \sum_{m=1}^{M} \pi^{\top} \mathbf{S}_m e^{\mathbf{T}_m t} \delta_m.$$

Outline

1. Markov Jump Processes in Practice
2. Comparison Between the Markov and Markov Mixture Model
3. The Finite Mixture of Markov Jump Processes
4. EM Estimation of the Exit Time Distribution
   - Estimation based on the sample paths $\{X_k\}$
   - Estimation based on the exit times $\{\tau_k\}$
   - Estimation of the finite mixture of phase-type distributions
5. Some Numerical Examples
6. Ongoing/future works

Statistics of a sample path

Suppose that we observe $\{X_s : 0 \leq s \leq \tau\}$. Statistics of the observation:

$$
\begin{aligned}
\Phi_m =& \mathbf{1}_{\{X = X^{(m)}\}} \\
B_i =& \mathbf{1}_{\{X_0 = i\}} \\
N_{ij} =& \#\ \text{of times } X \text{ makes an } i \to j \text{ transition, } j \neq i \\
N_i =& \sum_{j=1, j\neq i}^{n} N_{ij} \\
Z_i =& \text{the length of time } X \text{ occupies state } i \in E.
\end{aligned} \tag{6}
$$

Likelihood contribution of observations $\{\Phi_{m,k}, \mathbf{X}_k\}$

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

$$\log L = \sum_{k=1}^{K}\sum_{m=1}^{M}\Phi_{m,k}\Big[\sum_{i=1}^{n} B_{i,k}\log\left(s_{i,m}\pi_i\right) + \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} N_{ij}^{(k)}\log t_{ij,m} - \sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n} t_{ij,m}Z_i^{(k)} - \sum_{i=1}^{n}\delta_{i,m}Z_i^{(k)}\Big], \tag{7}$$

subject to the constraints: $\delta_{i,m} + \sum_{j=1,\, j\neq i}^{n} t_{ij,m} = -t_{ii,m}$,

$$\sum_{i}^{n}\pi_i = 1, \quad \text{and} \quad \sum_{m=1}^{M} s_{i,m} = 1, \ \text{for all } i = 1,\dots,n.$$

Maximum likelihood estimates under complete information

Theorem

MLE of π_i, $t_{ij,m}$, $t_{ii,m}$, $\delta_{i,m}$, and $s_{i,m}$ are given by

$$\widehat{\pi}_i = \frac{B_i}{K}, \quad \widehat{t}_{ij,m} = \frac{N_{ij,m}}{Z_{i,m}}, \quad \widehat{\delta}_{i,m} = \frac{N_{i\Delta,m}}{Z_{i,m}}, \quad \widehat{s}_{i,m} = \frac{B_{i,m}}{B_i},$$

$$\widehat{t}_{ii,m} = -\Big(\widehat{\delta}_{i,m} + \sum_{j=1, j\neq i}^{n} \widehat{t}_{ij,m}\Big)$$

where $N_{ij,m}$, $N_{i\Delta,m}$, $Z_{i,m}$, $B_{i,m}$, and B_i are defined respectively by

$$N_{ij,m} = \sum_{k=1}^{K} \Phi_{m,k} N_{ij}^{(k)}, \quad N_{i\Delta,m} = \sum_{k=1}^{K} \Phi_{m,k} N_{i\Delta}^{(k)},$$

$$Z_{i,m} = \sum_{k=1}^{K} \Phi_{m,k} Z_{i}^{(k)}, \quad B_{i,m} = \sum_{k=1}^{K} \Phi_{m,k} B_{i}^{(k)}, \quad B_i = \sum_{k=1}^{K} B_{i}^{(k)}.$$

Outline

EM estimation of the distribution parameters

1. As parameters of the mixture process X uniquely characterize the distribution of exit time τ, for given N independent observations $\{X_k\}$ of X, estimation of the exit time distribution is performed by maximizing the log-likelihood function of $\{X_k\}$ using EM algorithm.

2. If only independent observation of exit times $\{\tau_k\}$ are available, estimation of the distribution is performed by maximizing the log-likelihood function of the observation using the EM algorithm.

3. In both cases, stopping criterion for convergence after $l-$iteration is

$$||\widehat{F} - F_{\boldsymbol{\theta}^{(l)}}|| < \epsilon, \quad \text{with} \;\; 0 < \epsilon \leq 2,$$

where $\widehat{F}$ is the empirical cdf of the exit times $\{\tau_k\}$, whereas $F_{\boldsymbol{\theta}^{(l)}}$ represents the theoretical cdf of exit time τ under the estimate $\boldsymbol{\theta}^{(l)}$.

Estimation based on the sample paths $\{X_k\}$

(1) **Step 1**.Choose initial values of the distribution parameters π_i, $t_{ij,m}$, and $s_{i,m}$ for $i,j = 1,\dots,n$, and $m = 1,\dots, M$, all denoted by $\boldsymbol{\theta}$.

(2) **Step 2 (E-step)**

$$N_{ij,m}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{\Phi_{m,k} N_{ij}^{(k)} | \mathbf{X}_k\}, \; N_{i\Delta,m}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{\Phi_{m,k} N_{i\Delta}^{(k)} | \mathbf{X}_k\},$$

$$Z_{i,m}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{\Phi_{m,k} Z_{i}^{(k)} | \mathbf{X}_k\}, \; B_{i,m}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{\Phi_{m,k} B_{i}^{(k)} | \mathbf{X}_k\},$$

$$B_{i}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{B_{i}^{(k)} | \mathbf{X}_k\}$$

Note: the results coincide with that of for MJP when $\Phi_{m,k} = 1$ a.s.

Conditional expectations given the sample paths $\{X_k\}$

where the conditional expectation $\mathbb{E}_\theta\{\bullet|\mathbf{X}_k\}$ are specified by

$$\mathbb{E}_\theta\{\Phi_{m,k}N_{ij}^{(k)}|\mathbf{X}_k\} = \widehat{\Phi}_{m,k}N_{ij}^{(k)}, \quad \mathbb{E}_\theta\{\Phi_{m,k}N_{i\Delta}^{(k)}|\mathbf{X}_k\} = \widehat{\Phi}_{m,k}N_{i\Delta}^{(k)},$$

$$\mathbb{E}_\theta\{\Phi_{m,k}Z_{i}^{(k)}|\mathbf{X}_k\} = \widehat{\Phi}_{m,k}Z_{i}^{(k)}, \quad \mathbb{E}_\theta\{\Phi_{m,k}B_{i}^{(k)}|\mathbf{X}_k\} = \widehat{\Phi}_{m,k}B_{i}^{(k)}.$$

The estimate $\widehat{\Phi}_{m,k} := \mathbb{E}\{\Phi_{m,k}|\mathbf{X}_k\}$ of $\Phi_{m,k}$ is determined by

$$\widehat{\Phi}_{m,k} = \frac{\prod_{i=1}^n (\pi_i s_{i,m})^{B_i^{(k)}} \prod_{i=1}^n e^{t_{ii,m}Z_i^{(k)}} \prod_{i=1}^n \prod_{j=1,j\neq i}^{n+1} \big(t_{ij,m}\big)^{N_{ij}^{(k)}}}{\sum_{m=1}^M \prod_{i=1}^n (\pi_i s_{i,m})^{B_i^{(k)}} \prod_{i=1}^n e^{t_{ii,m}Z_i^{(k)}} \prod_{i=1}^n \prod_{j=1,j\neq i}^{n+1} \big(t_{ij,m}\big)^{N_{ij}^{(k)}}}.$$

Note: this is a needed generalization of the MJP.

(3) **Step 3 (M-step)**

$$\widehat{\pi}_i^{(l+1)} = \frac{B_i^{(l+1)}}{N}, \quad \widehat{t}_{ij,m}^{(l+1)} = \frac{N_{ij,m}^{(l+1)}}{Z_{i,m}^{(l+1)}}, \quad \widehat{\delta}_{i,m}^{(l+1)} = \frac{N_{i\Delta,m}^{(l+1)}}{Z_{i,m}^{(l+1)}},$$

$$\widehat{s}_{i,m}^{(l+1)} = \frac{B_{i,m}^{(l+1)}}{B_i^{(l+1)}}, \quad \widehat{t}_{ii,m}^{(l+1)} = -\Big(\widehat{\delta}_{i,m}^{(l+1)} + \sum_{\substack{j=1 \\ j\neq i}}^{n} \widehat{q}_{ij,m}^{(l+1)}\Big).$$

(4) **Step 4** Stop if the convergence criterion is achieved, for e.g.,

$$||\widehat{F} - F_{\boldsymbol{\theta}^{(l+1)}}|| < \epsilon, \quad \text{with } 0 < \epsilon \leq 2.$$

Outline

Estimation based on exit times $\{\tau_k\}$

For the estimation purposes, we use the following expressions

$$N_{ij}(t) := \lim_{\epsilon \downarrow 0} \sum_{k=1}^{[t/\epsilon]-1} \mathbf{1}_{\{X_{k\epsilon}=i, X_{(k+1)\epsilon}=j\}}$$

$$Z_i(t) := \int_0^t \mathbf{1}_{\{X_u=i\}} du.$$

The estimation is performed by minimizing information divergence (Kullback-Leibler information or relative entropy) $I(f,h)$ of the probability density f of complete observation of X with respect to the probability density h of $Y = u(X)$, when X is observed partially. The minimization is completely analogous with the EM algorithm. The approach is an adaptation of the one proposed by Asmussen et al. [7] for MJP.

Estimation based on the exit times $\{\tau_k\}$

(1) **Step 1**. Choose initial values of the distribution parameters π_i, $t_{ij,m}$, and $s_{i,m}$ for $i,j = 1,\dots,n$, and $m = 1,\dots,M$, all denoted by $\boldsymbol{\theta}$.

(2) **Step 2 (E-step)**

$$N_{ij,m}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{\Phi_{m,k} N_{ij}^{(k)} | \tau_k\}, \quad N_{i\Delta,m}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{\Phi_{m,k} N_{i\Delta}^{(k)} | \tau_k\},$$

$$Z_{i,m}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{\Phi_{m,k} Z_i^{(k)} | \tau_k\}, \quad B_{i,m}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{\Phi_{m,k} B_i^{(k)} | \tau_k\},$$

$$B_{i}^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\boldsymbol{\theta}}^{(l)}}\{B_i^{(k)} | \tau_k\}.$$

Conditional expectations given exit times τ_k

where the conditional expectation $\mathbb{E}_{\theta}\{\bullet|\tau_k\}$ are specified by

$$\mathbb{E}_{\theta}\left[\Phi_{m,k}N_{ij}^{(k)}|\tau_k=y\right]=\frac{\int_0^y \boldsymbol{\pi}^\top \mathbf{S}_m e^{\mathbf{T}_m u}\mathbf{e}_i\mathbf{e}_j^\top e^{\mathbf{T}_m(y-u)}\boldsymbol{\delta}_m t_{ij,m}du}{\sum_{m=1}^M \boldsymbol{\pi}^\top \mathbf{S}_m e^{\mathbf{T}_m y}\boldsymbol{\delta}_m}$$

$$\mathbb{E}_{\theta}\left[\Phi_{m,k}N_{i,\Delta}^{(k)}|\tau_k=y\right]=\frac{\boldsymbol{\pi}^\top \mathbf{S}_m e^{\mathbf{T}_m y}\mathbf{e}_i\delta_{i,m}}{\sum_{m=1}^M \boldsymbol{\pi}^\top \mathbf{S}_m e^{\mathbf{T}_m y}\boldsymbol{\delta}_m}$$

$$\mathbb{E}_{\theta}\left[\Phi_{m,k}Z_{i}^{(k)}|\tau_k=y\right]=\frac{\int_0^y \boldsymbol{\pi}^\top \mathbf{S}_m e^{\mathbf{T}_m u}\mathbf{e}_i\mathbf{e}_i^\top e^{\mathbf{T}_m(y-u)}\boldsymbol{\delta}_m du}{\sum_{m=1}^M \boldsymbol{\pi}^\top \mathbf{S}_m e^{\mathbf{T}_m y}\boldsymbol{\delta}_m}$$

$$\mathbb{E}_{\theta}\left[\Phi_{m,k}B_{i}^{(k)}|\tau_k=y\right]=\frac{\pi_i\mathbf{e}_i^\top \mathbf{S}_m e^{\mathbf{T}_m y}\boldsymbol{\delta}_m}{\sum_{m=1}^M \boldsymbol{\pi}^\top \mathbf{S}_m e^{\mathbf{T}_m y}\boldsymbol{\delta}_m}$$

$$\mathbb{E}_{\boldsymbol{\theta}}\big[B_i^{(k)}\big|\tau_k = y\big] = \frac{\sum_{m=1}^{M} \pi_i \mathbf{e}_i^\top \mathbf{S}_m e^{\mathbf{T}_m y} \boldsymbol{\delta}_m}{\sum_{m=1}^{M} \boldsymbol{\pi}^\top \mathbf{S}_m e^{\mathbf{T}_m y} \boldsymbol{\delta}_m}$$

(3) **Step 3 (M-step)**

$$\widehat{\pi}_i^{(l+1)} = \frac{B_i^{(l+1)}}{N}, \quad \widehat{t}_{ij,m}^{(l+1)} = \frac{N_{ij,m}^{(l+1)}}{Z_{i,m}^{(l+1)}}, \quad \widehat{\delta}_{i,m}^{(l+1)} = \frac{N_{i\Delta,m}^{(l+1)}}{Z_{i,m}^{(l+1)}},$$

$$\widehat{s}_{i,m}^{(l+1)} = \frac{B_{i,m}^{(l+1)}}{B_i^{(l+1)}}, \quad \widehat{t}_{ii,m}^{(l+1)} = -\Big(\widehat{\delta}_{i,m}^{(l+1)} + \sum_{\substack{j=1\\ j\neq i}}^{n} \widehat{q}_{ij,m}^{(l+1)}\Big)$$

(4) **Step 4** Stop if the convergence criterion is achieved, for e.g.,

$$||\widehat{F} - F_{\boldsymbol{\theta}^{(l+1)}}|| < \epsilon, \quad \text{with} \;\; 0 < \epsilon \leq 2.$$

Outline

Estimation of the finite mixture of phase-type distributions

By parameterization $p_m \times \pi_j^{(m)} = \pi_j \times s_{j,m}$, the prior distribution of the exit time τ can be rewritten as the mixture of phase-type distributions:

$$f(t) = \sum_{m=1}^{M} p_m \boldsymbol{\pi}_m^\top e^{\mathbf{T}_m t} \boldsymbol{\delta}_m, \quad \text{with} \;\; \boldsymbol{\pi}_m^\top = (\pi_1^{(m)}, \ldots, \pi_n^{(m)}).$$

Note: the transient state $E = \bigcup_{m=1}^{M} E_m$ with $E_i \cap E_j = \emptyset$, for $i \neq j$,

$$\pi_i^{(m)} = 0, \quad \text{for } i \notin E_m.$$

EM algorithm for $\hat{p}_m$ and $\hat{\pi}_m$ based on $\{\mathbf{X}_k\}$

Based on the observation on the sample paths $\{\mathbf{X}_k\}$:

(2) **Step 2 (E-step)**

$$\Phi_m^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\theta}^{(l)}}\{\Phi_{m,k}\big|\mathbf{X}_k\} \text{ with } \mathbb{E}_{\hat{\theta}^{(l)}}\{\Phi_{m,k}\big|\mathbf{X}_k\} = \hat{\Phi}_{m,k}.$$

(2) **Step 3 (M-step)**

$$\hat{p}_m^{(l+1)} = \frac{\Phi_m^{(l+1)}}{N} \quad \text{and} \quad \hat{\pi}_i^{(m)} = \frac{B_{i,m}^{(l+1)}}{\Phi_m^{(l+1)}}.$$

EM algorithm for $\hat{p}_m$ and $\hat{\pi}_m$ based on $\{\tau_k\}$

Based on the observation on the exit times $\{\tau_k\}$:

(2) **Step 2 (E-step)**

$$\Phi_m^{(l+1)} = \sum_{k=1}^{N} \mathbb{E}_{\hat{\theta}^{(l)}}\{\Phi_{m,k} | \tau_k = y\},$$

with

$$\mathbb{E}_{\hat{\theta}}\{\Phi_{m,k} | \tau_k\} = \frac{\hat{p}_m \boldsymbol{\pi}_m^\top e^{\mathbf{T}_m y} \boldsymbol{\delta}_m}{\sum_{m=1}^{M} \hat{p}_m \boldsymbol{\pi}_m^\top e^{\mathbf{T}_m y} \boldsymbol{\delta}_m}.$$

(2) **Step 3 (M-step)**

$$\hat{p}_m^{(l+1)} = \frac{\Phi_m^{(l+1)}}{N} \quad \text{and} \quad \hat{\pi}_i^{(m)} = \frac{B_{i,m}^{(l+1)}}{\Phi_m^{(l+1)}}.$$

Outline

Monte Carlo simulations

Consider two mixture of MJPs with the following set of parameter values:

1. State space $\mathbb{S} = \{1, \ldots, 6\} \cup \{7\}$ (with $\Delta = \{7\}$).
2. Initial distribution $\boldsymbol{\pi}^\top = (1, 0, 0, 0, 0, 0, 0)$.
3. Phase-generator matrix $\mathbf{T}_1$ of $X^{(1)}$:

$$\mathbf{T}_1 = \begin{pmatrix} -2 & 2 & 0 & 0 & 0 & 0 \\ 0.05 & -2.05 & 2 & 0 & 0 & 0 \\ 0 & 0.05 & -2.05 & 2 & 0 & 0 \\ 0 & 0 & 0.05 & -2.05 & 2 & 0 \\ 0 & 0 & 0 & 0.05 & -2.05 & 2 \\ 0 & 0 & 0 & 0 & 0.05 & -2.05 \end{pmatrix},$$

Monte Carlo simulations: cont'd

1. Phase-generator matrix $\mathbf{T}_2$ of $X^{(2)}$:

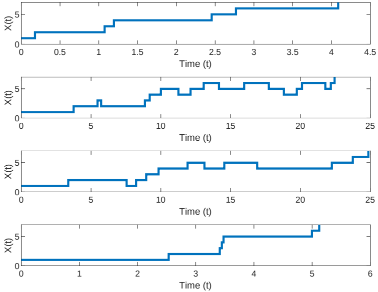
$$\mathbf{T}_2 = \begin{pmatrix} -0.3889 & 0.3889 & 0 & 0 & 0 & 0 \\ 0.1 & -0.4889 & 0.3889 & 0 & 0 & 0 \\ 0 & 0.1 & -0.4889 & 0.3889 & 0 & 0 \\ 0 & 0 & 0.1 & -0.4889 & 0.3889 & 0 \\ 0 & 0 & 0 & 0.1 & -0.4889 & 0.3889 \\ 0 & 0 & 0 & 0 & 0.1 & -0.4889 \end{pmatrix}.$$

2. The respective mixing probability matrices for $X^{(1)}$ and $X^{(2)}$:

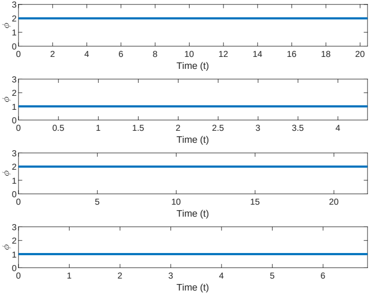
$$\mathbf{S}_1 = \mathrm{diag}(0.2, 0.3, 0.4, 0.5, 0.6, 0.7)$$
$$\mathbf{S}_2 = \mathrm{diag}(0.8, 0.7, 0.6, 0.5, 0.4, 0.3).$$

3. Generate $N = 20,000$ independent sample paths of X.

Sample paths $\{X_k\}$ with the respective regimes Φ_m



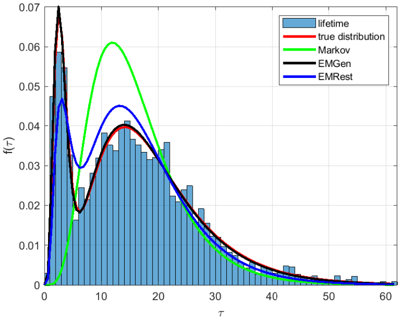
(a) sample paths X_k.



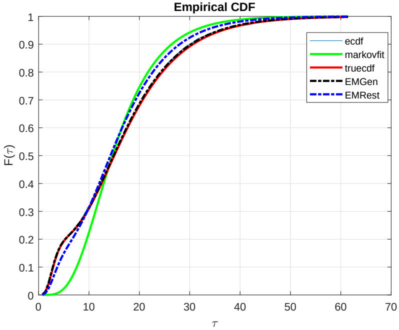
(b) regime Φ_m.

Figure: Sample paths of X with respective regimes Φ_m.

EM estimation based on the sample paths $\{X_k\}$



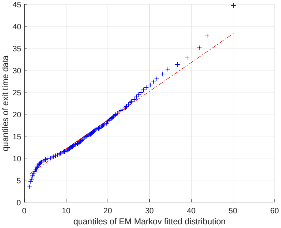
(a) exit time pdf $f(\tau)$.



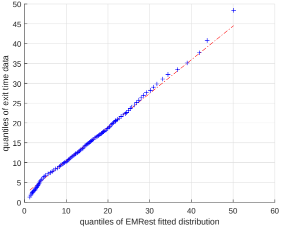
(b) exit time cdf $F(\tau)$.

Figure: Histogram of exit times and fitted model.

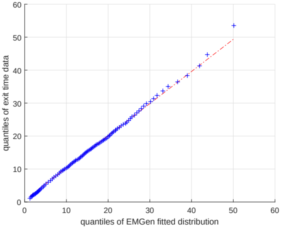
QQ plot of sampled exit times and the distribution



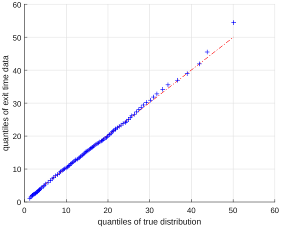
(a) Markov model.



(b) restricted mixture.

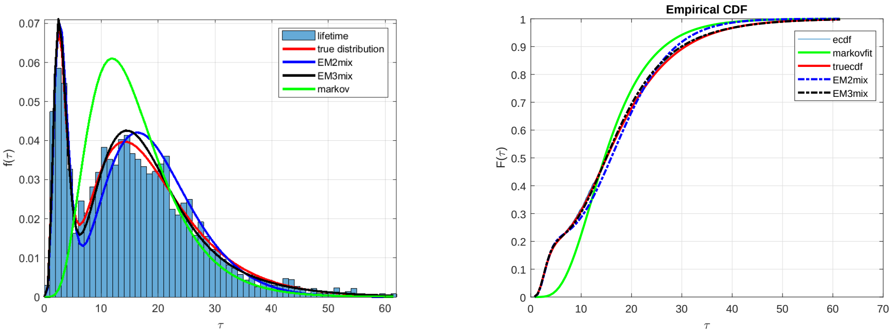


(c) general mixture.



(d) true distribution.

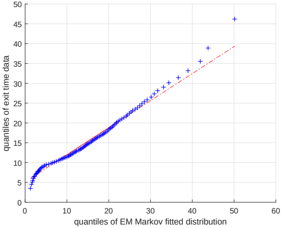
EM estimation based on the exit times $\{\tau_k\}$



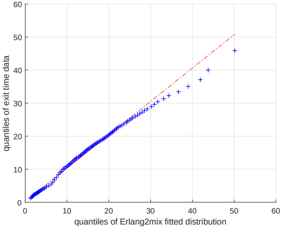
(a) exit time pdf $f(\tau)$. (b) exit time cdf $F(\tau)$.

Figure: Histogram and EM fitted exit time distributions.

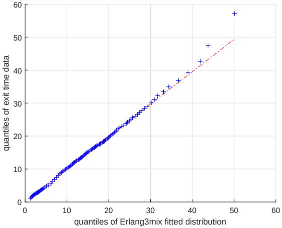
QQ plot of sampled exit times and the distribution



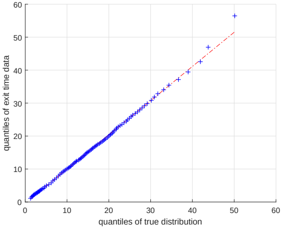
(a) Markov model.



(b) EM 2 mixture of distributions.



(c) EM 3 mixture of distributions.



(d) true distribution.

Outline

Ongoing/future works

1. Use BIC for model selection
2. Provide estimate of the variance for the estimators
3. Inclusion of covariates for the estimation
4. Multi absorbing states for competing risks analysis
5. Observation under censoring
6. Estimation under discrete observation of the sample paths (Summer/Winter work at the University of Copenhagen)

Acknowledgement: BS acknowledges the support from the school.

Thank You!

References

O. O. Aalen, O Borgan, & H. K. Gjessing. (2008). *Survival and Event History Analysis: A Process Point of View*, Springer.

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

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

H. Albrecher & S. Asmussen, S. (2010). *Ruin Probabilities*, 2nd Edition, World Scientific.

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

S. Asmussen. (2003). *Applied Probability and Queues*, 2nd Edition, Springer.

S. Asmussen, O. Nerman, & M. Olsson. (1996). Fitting phase-type distributions via the EM algorithm. *Scandinavian Journal of Statistics*, **23**: 419-441.

D. Assaf & B. Levikson. (1982). Closure of phase type distributions under operations arising in reliability theory. *Annals of Probability*, **10**: 265-269.

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

H. Balzter. (2000). Markov chain models for vegetation dynamics. Ecological Modelling, **126**: 139-154.

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

T. R. Bielecki & M. Rutkowski. (2002).*Credit Risk: Modeling, Valuation, and Hedging*. Springer.

References: cont'd

- I. Blumen, M. Kogan, & P. J. McCarthy. (1955). The industrial mobility of labor as a probability process. *Cornell Studies of Industrial Labor Relations.*, Vol. 6, Ithaca, N.Y., Cornel University Press.
- L. Breuer & D. Baum. (2005). *An Introduction to Queueing Theory and Matrix-Analytic Methods*, Springer.
- P. Buchholz, J. Kriege, & I. Felko. (2014). *Input Modeling with Phase-Type Distributions and Markov Models: Theory and Applications*, Springer.
- S. R. Chakravarthy & M. F. Neuts. (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.
- J. C. Duan, J. Sun, & T. Wang. (2012). Multiperiod corporate default prediction - a forward intensity approach. *Journal of Econometrics*, **170**: 191-209.
- D. Duffie, L. Saita, & K. Wang. (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics*, **83**: 635-665.
- D. Duffie & K. J. Singleton. (2003). *Credit Risk: Pricing, Measurement, and Management*, Princeton University Press.
- H. Frydman & B. A. Surya. (2019). The mixture of Markov jump processes: Monte Carlo and the EM estimation. *Working paper.*
- H. Frydman & T. Schuermann. (2008). Credit rating dynamics and Markov mixture models. *Journal of Banking and Finance*, **32**: 1062-1075.
- H. Frydman. (2005). Estimation in the mixture of Markov chains moving with different speeds. *Journal of the American Statistical Association*, **100**: 1046-1053.
- H. Frydman. (1984). Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association*, **79**: 632-638.

References: cont'd

- R. Jarrow & S. Turnbull. (1995). Pricing derivatives on financial securities subject to credit risk, *The Journal of Finance*, **50**: 53-86.
- D. Lando. (2004). *Credit Risk Modeling: Theory and Applications*. Princeton University Press.
- S. C. K. Lee & X. S. Lin. (2010). Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal*, **14**: 107-130.
- X. S. Lin & X. Liu. (2007). Markov aging process and phase-type law of mortality. *North American Actuarial Journal*, **11**: 92-109.
- M. F. Neuts. (1981). *Matrix-Geometric Solutions in Stochastic Models*, Johns Hopkins University Press, Baltimore.
- M. F. Neuts. (1975). Probability distributions of phase-type. In *Liber Amicorum Prof. Emiritus H. Florin*, 173-206, University of Louvain, Belgium.
- H. Okamura & T. Dohi. (2015). Phase-type software reliability model: parameter estimation algorithms with grouped data. *Annals of Operations Research*, 1-32.
- P. E. Pfeifer & R. L. Carraway. (2000). Modeling customer relationship as Markov chains. *Journal of Interactive Marketing*, **14**: 43-55.
- M. Pollak. (1985). Optimal detection of a change in distribution. *Annals of Statistics*, **13**:206-227.
- T. Rolski, H. Schmidli, V. Schmidt, & J. Teugels. (1998). *Stochastic Processes for Insurance and Finance*, Willey.
- B. A. Surya. (2018). Distributional properties of the mixture of continuous-time absorbing Markov chains moving at different speeds, *Stochastic Systems*, **8**: 29-44.