Estimation of discretely observed Markov Jump Processes with applications in survival analysis

Salim Serhan

Technical University of Denmark (DTU)

DTU Compute
Department of Applied Mathematics and Computer Science

Outline

- Problem formulation
- Complete-data problem
- EM-algorithm
- Extensions
- Conclusion

Problem formulation

- Consider a Markov Jump Process, $\{X(t)\}_{t\geq 0}$, of dimension k, initial probability vector $\boldsymbol{\pi}$ and generator $\mathbf{Q} = \mathbf{C} + \mathbf{D}$.
- $X(t)$ generates a Markovian arrival process (MAP).
- We examine following estimation problem: We observe state of $X(t)$ at certain discrete time points, as well as at the time of all arrivals in the MAP.
- It follows that the states have a physical interpretation.
- We wish to estimate $\boldsymbol{\theta} = (\boldsymbol{\pi}, \mathbf{C}, \mathbf{D})$.

Illustration

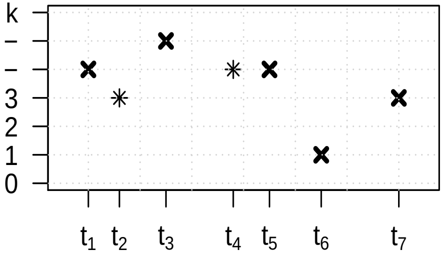
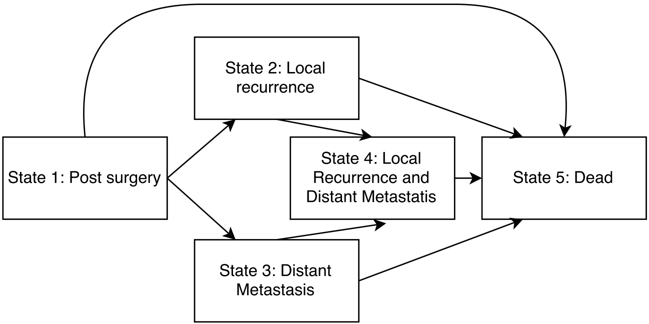


Figure: An illustration of the discrete observation sampling scheme. The stars are arrivals while the crosses are discrete observations.

- Observations are labeled as discrete observations or arrivals.

An example from survival analysis



$$\boldsymbol{\pi} = (1,0,0,0,0), \mathbf{C} = \begin{bmatrix} . & c_{12} & c_{13} & 0 & 0 \\ 0 & . & 0 & c_{24} & 0 \\ 0 & 0 & . & c_{34} & 0 \\ 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} \\ 0 & 0 & 0 & 0 & d_{25} \\ 0 & 0 & 0 & 0 & d_{35} \\ 0 & 0 & 0 & 0 & d_{45} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Illustration

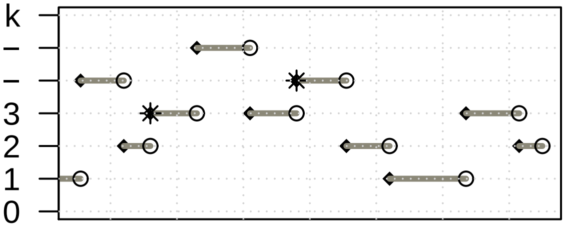


Figure: A complete sample path of the Markov jump process generating the MAP

Likelihood function

- The Complete-data likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{k} \pi_i^{b_i} \cdot \prod_{i=1}^{k} \prod_{j \neq i} c_{ij}^{n_{ij}} \exp(-c_{ij} z_i) \cdot \prod_{i=1}^{k} \prod_{j=1}^{k} d_{ij}^{\overline{n}_{ij}} \exp(-d_{ij} z_i).$$

- Where
  - b_i, the number of processes that start in state i,
  - z_i, the total time spent in state i,
  - n_{ij}, the total number of transitions from state i to state j not associated with an arrival,
  - $\overline{n}_{ij}$, the total number of transitions from state i to state j associated with an arrival,
- The maximum likelihood estimators are

$$\hat{\pi}_i = b_i, \quad \hat{c}_{ij} = \frac{n_{ij}}{z_i}, \quad \hat{d}_{ij} = \frac{\overline{n}_{ij}}{z_i}. \tag{1}$$

EM-algorithm

- Now consider the case of incomplete-data.
- We observe a vector of states $\mathbf{X} = (x_{t_1}, x_{t_2}, \ldots, x_{t_n})$, where $t_1 < t_2 < \ldots < t_n$.
- We also observe a vector of indicators $\mathbf{I} = (i_{t_1}, i_{t_2}, \ldots, i_{t_n})$. i_{t_h} equals 1 if the h'th observation is an arrival, 0 otherwise.
- The pair $(\mathbf{X}, \mathbf{I})$ is the incomplete data.
- For the E-step, we need expressions for $E(Z_k|\mathbf{X}, \mathbf{I})$, $E(N_{ij}|\mathbf{X}, \mathbf{I})$, $E(\overline{N}_{ij}|\mathbf{X}, \mathbf{I})$ and $E(B_i|\mathbf{X}, \mathbf{I})$

Some notation

- First, some notation. Put $\Delta_h = t_h - t_{h-1},\ h = 2, \ldots, (n-1)$, with $\Delta_h = t_1$.
- $M_{ij}^{k}(h) = E(Z_k | X(0) = i, X(\Delta_h) = j) =$ the expected sojourn time in state k, given that the process was initialised in state i and is in state j at time t.
- $f_{ij}^{kl}(h) = E(N_{kl} | X(0) = i, X(\Delta_h) = j) =$ the expected number of jumps not caused by an event from k to l, given that X was initialised in state i and is in state j after time t.
- $\overline{f}_{ij}^{kl}(h) = E(\overline{N}_{kl} | X(0) = i, X(\Delta_h) = j) =$ same as for $f_{ij}^{kl}(t)$, but for the number of jumps *caused* by an event.

Some notation

- Assuming homogeneity, we may then write
  - $E(Z_k|\mathbf{X}) = M^k_{\boldsymbol{\pi} x_{t_1}}(1) + \sum_{h=2}^{n} M^k_{x_{t_{h-1}} x_{t_h}}(h)$.
  - $E(N_{ij}|\mathbf{X}) = f^{ij}_{\boldsymbol{\pi} x_{t_1}}(1) + \sum_{h=2}^{n} f^{ij}_{x_{t_{h-1}} x_{t_h}}(h)$.
  - $E(\overline{N}_{ij}|\mathbf{X}) = \overline{f}^{ij}_{\boldsymbol{\pi} x_{t_1}}(1) + \sum_{h=2}^{n} \overline{f}^{ij}_{x_{t_{h-1}} x_{t_h}}(h)$.
  - $E(B_i|\mathbf{X}) = E(B_i|X(t_1) = x_{t_1}, I(t_1) = i_{t_1})$.
- Thus, the problem is reduced to finding expressions for $M, f, \overline{f}$ and $E(B_i|X_{t_1}, I_{t_1})$.

Integral calculation

- Define the matrices
  - $\mathbf{M}^{kk'}(h) = \int_0^{\Delta_h} \exp(\mathbf{C}u)\mathbf{e}_k\mathbf{e}_k' \exp(\mathbf{C}(\Delta_h - u))\mathrm{d}u.$
  - $\mathbf{M}^{kl'}(h) = \int_0^{\Delta_h} \exp(\mathbf{C}u)\mathbf{e}_k\mathbf{e}_l' \exp(\mathbf{C}(\Delta_h - u))\mathrm{d}u.$
- Where $\mathbf{e}_i$ is the i'th unit vector of appropriate dimension.
- A way to calculate the integrals is

$$\mathbf{M}^{kl'}(t) = \begin{pmatrix} I & \mathbf{0} \end{pmatrix} \exp\left(\begin{bmatrix} \mathbf{C} & \mathbf{e}_k\mathbf{e}_l' \\ \mathbf{0} & \mathbf{C} \end{bmatrix} t \right) \begin{pmatrix} \mathbf{0} \\ I \end{pmatrix},$$

- where I is the identity matrix of dimension $k \times k$ and $\mathbf{0}$ is a matrix of zeroes of dimension $k \times k$.

E-step formulas

- The E-step formulas are as follows, when $h \geq 2$

$$M_{ij}^{k}(h) = \frac{\mathbf{e}_i \mathbf{M}^{kk'}(h) \mathbf{D}^{i_{t_h}} \mathbf{e}_j}{\mathbf{e}_i \exp(\mathbf{C}\Delta_h) \mathbf{D}^{i_{t_h}} \mathbf{e}_j}, \quad f_{ij}^{kl}(h) = c_{kl} \frac{\mathbf{e}_i \mathbf{M}^{kl'}(h) \mathbf{D}^{i_{t_h}} \mathbf{e}_j}{\mathbf{e}_i \exp(\mathbf{C}\Delta_h) \mathbf{D}^{i_{t_h}} \mathbf{e}_j},$$

$$\overline{f}_{ij}^{kl}(h) = 0 \text{ for } l \neq j, \quad \overline{f}_{ij}^{kl}(h) = d_{kj} \frac{\mathbf{e}_i \exp(\mathbf{C}\Delta_h) \mathbf{D}^{i_{t_h}} \mathbf{e}_k}{\mathbf{e}_i \exp(\mathbf{C}\Delta_h) \mathbf{D}^{i_{t_h}} \mathbf{e}_j} \text{ for } l = j.$$

- When $h = 1$, replace all the $\mathbf{e}_i$ vectors by $\boldsymbol{\pi}$. Also,

$$E(B_i | X(t_1), I_{t_1}) = \frac{\boldsymbol{\pi}_i \mathbf{e}_i' \exp(\mathbf{C}t_1) \mathbf{D}^{i_1} \mathbf{e}_{x_{t_1}}}{\boldsymbol{\pi} \exp(\mathbf{C}t_1) \mathbf{D}^{i_1} \mathbf{e}_{x_{t_1}}}.$$

Covariates

- We can parameterize the transition intensities using covariates.
- Let $\mathbf{Z}$ denote the covariaties.
- A popular model in survival analysis is the Cox proportional hazards model

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)\exp\left(\boldsymbol{\beta}\mathbf{Z}\right).$$

- This gives an inhomogeneous model, unless we put $\lambda_0(t) = \lambda$.

Phase-type sojourn times

- Exponential sojourn times may be unrealistic.
- Consider the Markov jump process $Y(t)$ with an expanded state space

$$\{1_1, \ldots, 1_{m_1}\} \cup \{2_1, \ldots, 2_{m_2}\} \cup \ldots \cup \{k_1, \ldots, k_{m_k}\}$$

- Where $m_i, i = 1, 2, \ldots, k$ is the number of sub-states for the i'th batch state. Let $m = m_1 + m_2 + \ldots + m_k$ denote the dimension of the expanded state space.
- Canonical representations should be used. That is, Coxian structures with increasing mean sojourn times.
- The sub-states do not have a physical interpretation, i.e. we cannot observe them.
- $Y(t)$ is a semi-Markov jump process with the following relation to $X(t)$.

$$P(X(t) = r|Y(t) = r_i) = 1$$

- This is a hidden Markov model with deterministic state-dependent distributions.

Estimation with Phase-type sojourn times

- The likelihood function is

$$L(\boldsymbol{\theta}) = \boldsymbol{\pi} \left(\prod_{h=1}^{n} \boldsymbol{\Gamma}(h) \boldsymbol{P}(x_{t_h}) \right)$$

- Where $\mathbf{\Gamma}(h)$ is an $m \times m$ matrix, where the (i,j)-th element is $P(X(\Delta_h) = j | X(0) = i, I_{t_h} = i_{t_h})$. We find these by

$$\frac{\mathbf{e}_i \exp(\mathbf{C}\Delta_h) \mathbf{D}^{i_{t_h}} \mathbf{e_j}}{\mathbf{e}_i \exp(\mathbf{C}\Delta_h) \mathbf{D}^{i_{t_h}} \mathbf{1}}.$$

- Where $\mathbf{1}$ is a vector of ones of appropriate dimension.
- $\mathbf{P}(x_{t_h})$ is an $m \times m$ diagonal matrix, where the i'th diagonal elements is $P(X(t_h) = x_{t_h} | Y(t_h) = i)$

Misclassification models

- With a Hidden Markov Model defined, we can easily include the possibility of misclassification.
- This can be the case when there is uncertainty on the state observations.
- In survival analysis, this is known as a censored state.
- Let e_{rs} denote the probability of wrongly classifying $X(t)$ in batch-state s, when the true batch-state is r. We can write this as

$$P(X(t_h) = r | Y(t_h) = s) = e_{rs}.$$

- This gives categorical state-dependent distributions, and we may use the previous likelihood function.

Conclusion

- We have extended some EM-algorithms from the literature to account for different observation types.
- We have shown how these models may be applied to a certain model from survival analysis.
- Covariates can be included, with certain limitations.
- We can have phase-type sojourn times at the cost of a harder estimation problem.
- And finally, we can allow uncertainty on the state observations.

Further Work

- Derive formulas for the Fisher information matrix.
- Study the large sample properties of the algorithm.
- Develop estimators for non-homogeneous Markov processes.