Analysis of Progressive Multi-State Models with Misclassified States

Feng He and Grace Y. Yi

Department of Statistics and Actuarial Science
University of Waterloo

February 14, 2014

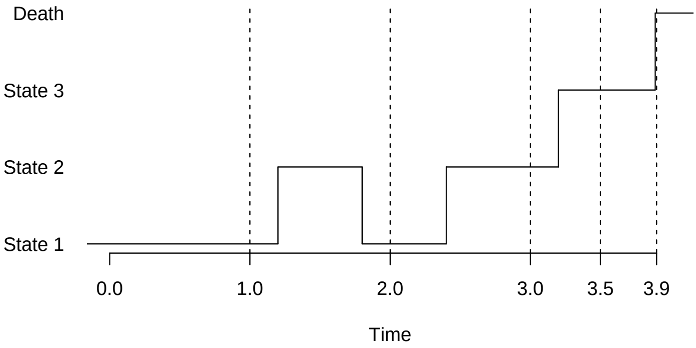
Overview

CAV data (Sharples *et al.*, 2003)

- Coronary allograft vasculopathy is a chronic disease and one of the most common causes of death among long-term survivors of heart transplantation
- The aim of this study: estimate the diagnostic accuracy of coronary angiography and identify patient risk factors for CAV onset and progression

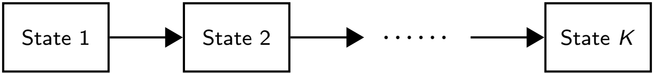
Multi-state models

- Longitudinal data collected from disease progression studies are often under panel/intermittent observation
  - Observation times are irregularly spaced
  - Exact times of transitions are interval censored



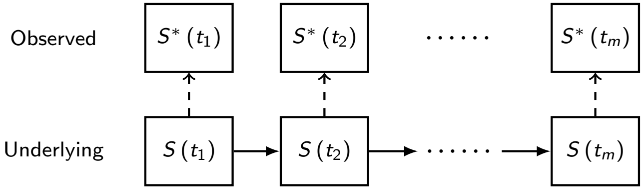
Multi-state models

- Continuous-time multi-state modelare commonly used for characterizing disease processes due to irregularly spaced observation times in the study.
- Unidirectional progressive Markov models
  - model the process of successive events reflecting an accumulation of damage or deterioration
  - Satten (1999); Cook *et al.* (2004); Sutradhar and Cook (2008); Chen *et al.* (2010)
- Statistical challenge: misclassification of the disease state
  - The poor quality of a diagnostic test
  - The impossibility of the accurate assessment
  - The reading error



Hidden Markov models

- Objective: understand the influence of covariates for the disease onset and progression and account for potential misclassification
- Hidden Markov models (HMMs)
  - The observation is governed by some probability distribution conditional on the unobserved state of an underlying Markov process
  - The observed process with misclassification in the progressive multi-state model no longer has the Markov property



Notation

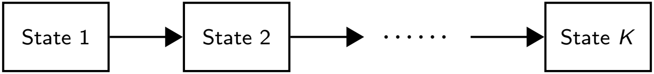
- Suppose an individual moves among K states
- $S(t)$: the underlying unobservable state at time t
- $S^*(t)$: the observed state subject to misclassification at time t
- $P_{ij}(s, s+t)$: the transition probability from state i at time s to j at time $(s+t)$

$$P_{ij}(s, s+t) = \Pr\left[S(s+t) = j \mid S(s) = i\right], \qquad s \geq 0, t > 0$$

- $q_{ij}(t)$: the transition intensity from state i to j at time t

$$q_{ij}(t) = \lim_{\Delta t \downarrow 0} \frac{P_{ij}(t, t+\Delta t)}{\Delta t}, \quad i \neq j; \qquad q_{ii}(t) = -\sum_{j \neq i} q_{ij}(t)$$

K-state progressive Markov model



- The process is irreversible and the transition only happens from one state to its consecutive state
  - for state $i = 1, \ldots, K-1$,

$$q_{i,i+1}(t) > 0, \quad q_{ii}(t) = -q_{i,i+1}(t) \quad \text{and} \quad q_{ij}(t) = 0, \quad j \neq i, i+1$$

  - for state K, $q_{Kj} = 0$, $j = 1, \ldots, K$
  - $q_i(t) = q_{i,i+1}(t)$, $i = 1, \ldots, K-1$; $q_K = 0$
- Time homogeneous:
  - $q_{i,i+1}(t) = q_i(t) = q_i > 0, \quad i = 1, \ldots, K-1$
  - $P_{ij}(s, s+t) = P_{ij}(0, t) = P_{ij}(t)$

K-state progressive Markov model

An explicit analytic expression for the transition probability from state i to state j in terms of transition intensities is given by Satten (1999) in the form

$$P_{ij}(t) = \begin{cases} \sum_{k=i}^{j} C_{ijk} \exp(-q_k t) & i \le j \\ 0 & i > j \end{cases}$$

where the coefficients are given by

$$C_{ijk} = \frac{\prod_{l=i}^{j-1} q_l}{\prod_{l=i,\, l\neq k}^{j} (q_l - q_k)} \qquad i \le k \le j$$

and $C_{kkk} = 1,\ i, j, k = 1, \ldots, K$.

Modelling scheme

- Regression model for transition intensities

$$q_i(\mathbf{x}) = q_{i0} \exp\left(\mathbf{x}^{\mathsf{T}} \boldsymbol{\beta}_i\right) \qquad i = 1, \ldots, K-1,$$

  - $\mathbf{x}$: a $p \times 1$ vector of prognostic variables
  - $q_{i0} = \exp(\beta_{i0})$: the baseline transition intensity out of state i
  - $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{ip})^{\mathsf{T}}$: of primary interest

Modelling scheme

- Multinomial logistic regression model for misclassification probabilities
  - Misclassification probabilities:

$$\begin{aligned} \pi_{ij}\left[C\left(t\right)\right] &= \Pr\left[S^{*}\left(t\right)=j \mid S\left(t\right)=i, C\left(t\right)\right], \qquad i \neq j \\ \pi_{ii}\left[C\left(t\right)\right] &= 1-\sum_{j\neq i}^{K} \pi_{ij}\left[C\left(t\right)\right] \end{aligned}$$

  - $C\left(t\right)$: the time-dependent misclassification predictor
  - Multinomial logistic regression model

$$\log\left\{\frac{\pi_{ij}\left[C\left(t\right)\right]}{\pi_{ii}\left[C\left(t\right)\right]}\right\} = \alpha_{ij0}+\alpha_{ijc}C\left(t\right), \qquad i \neq j$$

  - α_{ij0} and α_{ijc} : time-independent regression coefficients

Method 1:

Full likelihood based method

Likelihood function

- N individuals under study
- m_ℓ observation times for individual ℓ : $t_{\ell 1}, \ldots, t_{\ell m_\ell}$

The contribution $\mathcal{L}_\ell(\boldsymbol{\theta})$ can be written as a product of matrices,

$$\begin{aligned} \mathcal{L}_\ell(\boldsymbol{\theta}) &= \Pr\left(S^*_{\ell 1}, \ldots, S^*_{\ell m_\ell} \,\middle|\, H^C_{\ell, m_\ell+1}, \mathbf{x}_\ell; \boldsymbol{\theta}\right) \\ &= \mathbf{f}_\ell^\mathsf{T} \boldsymbol{\Pi}_{\ell 1} \prod_{r=2}^{m_\ell} \left(\mathbf{P}_{\ell r} \boldsymbol{\Pi}_{\ell r}\right) \mathbf{1} \end{aligned}$$

- $\mathbf{f}_\ell = \left[\Pr\left(S_{\ell 1} = i\right), i = 1, \ldots, K\right]^\mathsf{T}$; $\mathbf{1}$: a $K \times 1$ unit vector
- $\boldsymbol{\Pi}_{\ell r} = \mathrm{diag}\left[\Pr\left(S^*_{\ell r} \mid S_{\ell r} = j, C_{\ell r}; \boldsymbol{\alpha}_j\right), j = 1, \ldots, K\right]$
- $\mathbf{P}_{\ell r}$: the $K \times K$ transition probabilities with (i, j) entry

$$\Pr\left(S_{\ell r} = j \mid S_{\ell, r-1} = i, \mathbf{x}_\ell; \boldsymbol{\beta}\right),$$

Maximum likelihood estimation via the EM algorithm

- Motivation: the underlying states are not observed and can be considered as latent variables
- The log-likelihood of the complete data with the true states observed for individual ℓ is

$$\log \mathcal{L}_\ell^{c}(\boldsymbol{\theta}) = \sum_{r=1}^{m_\ell} \log \Pr\left(S_{\ell r}^{*} \mid S_{\ell r}, C_{\ell r}; \boldsymbol{\alpha}_{S_{\ell r}}\right) + \sum_{s=1}^{m_\ell - 1} \log \Pr\left(S_{\ell, s+1} \mid S_{\ell s}, \mathbf{x}_\ell; \boldsymbol{\beta}\right)$$

- E-step

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_k) = \sum_{\ell=1}^{N} E\left[\log \mathcal{L}_\ell^{c}(\boldsymbol{\theta}) \,\middle|\, H_{\ell, m_\ell+1}^{S^*}, H_{\ell, m_\ell+1}^{C}, \mathbf{x}_\ell; \boldsymbol{\theta}_k\right]$$

- M-step: The parameters related to the transition intensity model for the progression through underlying states are distinct from those related to the misclassification model for the observation process of underlying states.

Variance estimation

- Asymptotic distribution

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\mathrm{FL}}-\boldsymbol{\theta}\right) \xrightarrow{d} \mathbf{N}\left[\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\theta})\right]$$

  - $\mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$: the multivariate normal distribution
  - $\mathcal{I}(\boldsymbol{\theta})$: the Fisher information matrix
- Sandwich-type robust variance estimation

$$\mathbf{C}(\boldsymbol{\theta}) = \mathbf{A}^{-1}(\boldsymbol{\theta})\,\mathbf{B}(\boldsymbol{\theta})\,\mathbf{A}^{-1}(\boldsymbol{\theta})$$

with

$$\mathbf{A}(\boldsymbol{\theta}) = E\left[\frac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta};\mathbf{S}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right] \quad \text{and} \quad \mathbf{B}(\boldsymbol{\theta}) = E\left\{\left[\frac{\partial \log \mathcal{L}(\boldsymbol{\theta};\mathbf{S}^*)}{\partial\boldsymbol{\theta}}\right]^{\otimes 2}\right\}.$$

- Estimation of $\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\theta})$

$$\widehat{\mathbf{B}}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{\ell=1}^{N}\left[\frac{\partial \log \mathcal{L}_\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]^{\otimes 2} = \frac{1}{N}\sum_{\ell=1}^{N}\left\{\left.\frac{\partial Q_\ell(\boldsymbol{\theta},\boldsymbol{\theta}')}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}'=\boldsymbol{\theta}}\right\}^{\otimes 2};$$

$$\widehat{\mathbf{A}}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{\ell=1}^{N}\frac{\partial^2 \log \mathcal{L}_\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}.$$

Method 2:

Pairwise likelihood based method

Pairwise likelihood formulation

- The pairwise likelihood contributed from individual ℓ takes the matrix product form of

$$\begin{aligned}\mathcal{L}_\ell^p(\boldsymbol{\theta}) &= \prod_{r=1}^{m_\ell-1}\prod_{s=r+1}^{m_\ell} \Pr\left(S_{\ell r}^*, S_{\ell s}^* \mid C_{\ell r}, C_{\ell s}, \mathbf{x}_\ell; \boldsymbol{\theta}\right) \\ &= \prod_{s=2}^{m_\ell}\left(\mathbf{f}_\ell^{\mathsf{T}}\boldsymbol{\Pi}_{\ell 1}\mathbf{P}_{\ell 1 s}\boldsymbol{\Pi}_{\ell s}\mathbf{1}\right) \times \prod_{r=2}^{m_\ell-1}\prod_{s=r+1}^{m_\ell}\left(\mathbf{f}_\ell^{\mathsf{T}}\mathbf{P}_{\ell 1 r}\boldsymbol{\Pi}_{\ell r}\mathbf{P}_{\ell r s}\boldsymbol{\Pi}_{\ell s}\mathbf{1}\right)\end{aligned}$$

- The expected complete data pairwise log-likelihood for individual ℓ is

$$\begin{aligned}Q_\ell^p(\boldsymbol{\theta}, \boldsymbol{\theta}_k) = \sum_{s=2}^{m_\ell} E\Big[&\log \Pr\left(S_{\ell 1}^* \,\middle|\, S_{\ell 1}, C_{\ell 1}; \boldsymbol{\alpha}_{S_{\ell 1}}\right) + \log \Pr\left(S_{\ell s}^* \,\middle|\, S_{\ell s}, C_{\ell s}; \boldsymbol{\alpha}_{S_{\ell s}}\right) \\ &+ \log \Pr\left(S_{\ell s} \mid S_{\ell 1}, \mathbf{x}_\ell; \boldsymbol{\beta}\right) \,\Big|\, S_{\ell 1}^*, S_{\ell s}^*, C_{\ell 1}, C_{\ell s}; \boldsymbol{\theta}_k\Big] \\ &+ \sum_{r=2}^{m_\ell-1}\sum_{s=r+1}^{m_\ell} E\Bigg\{\log \Pr\left(S_{\ell r}^* \,\middle|\, S_{\ell r}, C_{\ell r}; \boldsymbol{\alpha}_{S_{\ell r}}\right) + \log \Pr\left(S_{\ell s}^* \,\middle|\, S_{\ell s}, C_{\ell s}; \boldsymbol{\alpha}_{S_{\ell s}}\right) \\ &+ \log \Pr\left(S_{\ell s} \mid S_{\ell r}, \mathbf{x}_\ell; \boldsymbol{\beta}\right) + \log\left\{\sum_{S_{\ell 1}}\left[\Pr\left(S_{\ell r} \mid S_{\ell 1}, \mathbf{x}_\ell; \boldsymbol{\beta}\right)\Pr\left(S_{\ell 1}\right)\right]\right\} \,\Bigg|\, S_{\ell r}^*, S_{\ell s}^*, C_{\ell r}, C_{\ell s}; \boldsymbol{\theta}_k\Bigg\}\end{aligned}$$

Variance estimation

- Asymptotic distribution

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\text{PL}} - \boldsymbol{\theta}\right) \xrightarrow{d} \mathbf{N}\left[\mathbf{0}, \mathbf{G}^{-1}(\boldsymbol{\theta})\right]$$

  - $\mathbf{G}(\boldsymbol{\theta})$: the Godambe information matrix (Godambe, 1960)

$$\mathbf{G}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\,\mathbf{J}(\boldsymbol{\theta})^{-1}\,\mathbf{H}(\boldsymbol{\theta})$$

  with

$$\mathbf{H}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\frac{\partial^2 \log \mathcal{L}^p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathsf{T}}}\right] \quad \text{and} \quad \mathbf{J}(\boldsymbol{\theta}) = \text{Var}_{\boldsymbol{\theta}}\left[\frac{\partial \log \mathcal{L}^p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]$$

    - $\mathcal{L}^p(\boldsymbol{\theta}) = \mathcal{L}^p(\boldsymbol{\theta}; \mathbf{S}^*)$

Analysis of CAV data

			Likelihood			Pairwise Likelihood		
State	Covariates		MLE	SE	*p*-value	MPLE	SE	*p*-value
$1 \to 2$	Intercept	β_{10}	−3.076	0.252	0.0000	−2.642	0.227	0.0000
	IHD	β_{11}	0.561	0.255	0.0279	0.525	0.275	0.0561
	Donor age	β_{12}	0.507	0.126	0.0001	0.319	0.152	0.0363
$2 \to 3$	Intercept	β_{20}	−2.646	0.498	0.0000	−3.273	0.706	0.0000
	IHD	β_{21}	0.488	0.527	0.3539	0.752	0.699	0.2818
	Donor age	β_{22}	−0.095	0.242	0.6941	0.337	0.296	0.2538

- Constant misclassification probability
- Forward selection of risk factors
    - IHD: preoperative ischemic heart disease
    - both IHD and dage have significant effects on the CAV onset
    - the effect of recipient age is not significant

Additional Feature:
Non-homogeneous population

Mover-stayer feature

- Two types of individuals: the stayer stays in the initial state, whereas the mover evolves according to a continuous-time Markov process
- ω_k : the probability of being a stayer
- Logistic model for the mover-stayer feature

$$\log\left\{\frac{\omega_k}{1-\omega_k}\right\} = \mathbf{x}_{1\ell}^{\mathsf{T}}\phi_k$$

  - $\mathbf{x}_{1\ell} = \left(1, \mathbf{x}_\ell^{\mathsf{T}}\right)^{\mathsf{T}}$
  - $\phi_k = (\phi_{k0}, \phi_{k1}, \ldots, \phi_{kp})^{\mathsf{T}}$
- Simulation results

Concluding Remarks

- The HMM is employed to simultaneously estimate the transition rates and account for potential misclassification
- The mover-stayer feature is incorporated
- Likelihood approach vs pairwise likelihood approach:
  - Robustness
  - The bivariate cases for the output independence assumption
- Direct maximization vs EM algorithm
  - Convenience
  - Numerical underflow (Leroux and Puterman, 1992)

The End