Multi-state models
Model diagnostics and model extensions

Andrew Titman
Lancaster University

November 28, 2008
Overview

- Introduction to multi-state modelling
- Interval censoring/panel observation
- Model diagnostics
- Non-parametric prevalence estimation
- Semi-Markov models
Multi-state models

- Used in *Event history analysis*.
  - Times of occurrence of events
  - Types of event that occurred.
- Consider an event as a transition from one state to another
  - E.g. healthy to diseased
  - E.g. diseased to death
- Process modelled in continuous time.
Example: Simple survival model

Alive \rightarrow Dead
Example: Competing risks model
Example: Illness-death model without recovery

[Diagram showing a flowchart with states Healthy, Ill, and Dead, with arrows indicating transitions.]
Example: Illness-death model with recovery
Areas of application

- Dementia
- Psoriatic arthritis
- Breast cancer
- Diabetic retinopathy
- HIV/AIDS
- Employment/Unemployment
- Smoking prevention
Mathematical framework

▶ Models defined by transition intensities between states

\[ \lambda_{rs}(t, \mathcal{F}_t, z(t)) = \lim_{\delta t \downarrow 0} \frac{\mathbb{P}(X(t + \delta t) = s | X(t) = r, \mathcal{F}_t, z(t))}{\delta t} \]

where \( \mathcal{F}_t \) is the history or filtration of the process and \( z(t) \) are covariates.

▶ Common assumptions
  ▶ Markov:
  \[ \lambda_{rs}(t, \mathcal{F}_t) = \lambda_{rs}(t) \]
  ▶ Semi-Markov:
  \[ \lambda_{rs}(t, \mathcal{F}_t) = \lambda_{rs}(t - t_r) \]
  where \( t_r \) is the time of entry into state \( r \).
  ▶ Cox-Markov model:
  \[ \lambda_{rs}(t, \mathcal{F}_t, z) = \lambda_{rs}(t) \exp(\beta^T z) \]
Types of observation

- Complete observation
- Right censored
- Left truncated and right censored
  - e.g. delayed entry into the study in a known state.
- Interval censored
- Panel observed
  - Observe patients’ disease status at discrete time points
- Current status data
  - Only observe patients’ disease status once
  - Also called Type I interval-censoring
- Panel observed with misclassification.
Right-censoring and left-truncation

- Methods for survival data can be generalised
- Aalen-Johansen estimator generalises the Nelson-Aalen/Kaplan-Meier to multi-state data
  - Makes a Markov assumption, but intensities otherwise non-parametric.
  - Marginal estimates of the intensity are robust for non-Markov processes.
- Generalisations of Cox models
  - Cox-Markov model
  - Cox-semi-Markov model
- Assumes non-informative censoring.
Interval censored and panel observed data
Interval censored and panel observed data

- Estimation more difficult
- Generally rely on parametric models
- Under a Markov assumption can write the likelihood as

\[
L(\theta) = \prod_{i,j} p_{x_{ij},x_{i+1,j}}(t_{ij}, t_{i+1j}; \theta)
\]

- Relying on observation process that generates \( t_{ij} \) being ignorable
  - Given observation time \( t_i, t_{i+1} \) does not depend on the trajectory of \( X(t) \) after \( t_i \).
Calculation of transition probabilities

- For Markov processes, the transition probabilities satisfy the Kolmogorov Forward equations

\[
\frac{dP(t_0, t)}{dt} = P(t_0, t)Q(t).
\]

where \( Q(t) \) is the matrix of transition intensities.

- For a time homogeneous Markov process, where \( Q(t) \equiv Q_0 \) we have

\[
P(t_0, t_1) = \exp \left( Q_0(t_1 - t_0) \right)
\]

- Other cases where transition probabilities are tractable
  - Time transformation models: \( Q(t) = Q_0f(t) \), for some non-negative scalar function \( f \).
  - Piecewise constant transition intensities.

- Otherwise need to solve KFE numerically.
Model diagnostics

- Use of parametric models mean strong assumptions made
  - Markov property
  - Time homogeneity or a particular form to time inhomogeneity
  - Patient heterogeneity, conditional on measured covariates
  - Proportional intensities assumption for covariates

- Model diagnostics needed to assess validity of these assumptions
  - Informal/graphical methods
  - Formal goodness-of-fit tests.
  - Tests against specific alternatives.
Plots against non-parametric alternatives

- Often for *illness-death* type models, the absorbing state is death and the time to death is known up to right censoring
- Plot Kaplan-Meier estimate of times to absorption against estimates from fitted parametric model
Plots against non-parametric alternatives

- Assessment of fit informal
  - Confidence limits are pointwise
  - Confidence limits don’t take into account that both parametric and non-parametric estimates are from the same data.
- Limited power to detect model misspecification.
  - Don’t take into account fit for intermediate states.
Estimates of prevalence

- Would like to compare non-parametric prevalence estimates with the fitted parametric model.
  - i.e. estimates of $P_{1s}(0, t)$, for $s = 1, \ldots, S$.
- However, good non-parametric estimates are not easy to obtain.
- Existing methods:
  - Simplistic: “Prevalence counts”
  - Complicated: Non-parametric maximum likelihood estimate assuming Markov.
Prevalence counts

- Want to estimate proportion of subjects in state $r$ at time $t$.
- But only know subject’s state at times $t_{ij} \neq t$.
- Apply some form of interpolation.
  - E.g. if $X(t_{ij}) = r$ and $X(t_{(i+1)j}) = s$ assume $r \to s$ occurred at midpoint.
  - Then use a simple moment estimate $\hat{P}_{1s}(0, t) = \hat{N}_s(t)/N_0(t)$.
- Patients who reach the absorbing state remain under observation until time they would have been censored had they survived, e.g. end of study date.
- Estimates may be reasonable if gaps between observations small.
- More generally, will have systematic bias.
  - Will not be able to tell if discrepancy with fitted parametric model is real or an effect of interpolation.
Non-parametric maximum likelihood estimate

- Assume time inhomogeneous Markov process
  - For purposes of prevalence estimation, estimates are robust for non-Markov processes.
- Use self-consistency estimation algorithm (Turnbull 1976)
  - Establish the points/regions of support for the transition survival functions.
  - Start with arbitrary estimates of survival functions
  - Then calculate $N_i^{(r)} = E[\text{Number in state } r \text{ at } t_i | D, \hat{F}]$ and $N_i^{(rs)} = E[\text{Number of state } r \text{ to } s \text{ transitions at } t_i | D, \hat{F}]$.
  - New estimate of transition probability at time $t_i$ is then $\frac{N_i^{(rs)}}{N_i^{(r)}}$.
  - Repeat until convergence.
- Convergence often slow.
- No asymptotic theory for variance estimates.
  - Can bootstrap but will be time consuming.
Comparison of methods

- Prevalence counts based on approximate interpolation
  - Easy to compute
  - Method current used in msm package in R.
  - May be too biased to be useful in many cases.

- NPMLE under Markov assumption
  - Gives good non-parametric estimate
  - Typically much harder to fit than original parametric model we wanted to assess.

- Is there some compromise method? e.g. unbiased but computationally simpler.
Compromise method

- Methods of estimation available for current status data (e.g. Datta and Sundaram, 2006).
  - For progressive models where only one path possible to a particular state.
  - Uses isotonic regression: Pooled-Adjacent-Violators algorithm.
  - Estimation procedure is quick and easy to implement e.g. using isoreg() in R.
- Treat observations for panel observed data as if they come from distinct individuals.
  - Under certain circumstances will get consistent estimates.
  - Some loss in efficiency.
Compromise method

- Consistent estimates maintained if
  1. Observation process entirely independent of the multi-state process
  2. Observations will continue if patient has reached the absorbing state

  - Condition 1. is (approximately) met in many cases
    - Though it is a stronger condition than non-informative sampling times.

- Condition 2. rarely met, especially for illness-death models.
Two-stage estimation procedure

- For illness-death type models where the time to absorption is known up to right censoring:
  - Survival estimate from Kaplan-Meier (not using intermediate observations) and NPMLE using all data coincide.
  - The conditional process, conditioning on not having entered the absorbing state, does usually meet condition 2.
    - If a patient survives, they will still have observations until some (independent) censoring time.
- Motivates estimating the survival process using Kaplan-Meier
- Motivates estimating the conditional process using the isotonic regression method.
- Unconditional prevalences estimates obtained by combining the two.
Example: Cardiac allograft vasculopathy in post-heart transplantation patients

- 596 post-heart transplant patients from Papworth hospital.
- Clinic visits for angiogram
  - Scheduled after 2 years and on a yearly basis thereafter.
  - Actual visit times varied considerably.
  - Presence of CAV determined by the angiogram.
- Time of death known to exact day.
CAV example: Comparison of estimates

State 1 prevalence

Isotonic
NPMLE
PP
Markov
CAV example: Comparison of estimates

State 2 prevalence

<table>
<thead>
<tr>
<th>Prevalence</th>
<th>Isotonic</th>
<th>NPMLE</th>
<th>PP</th>
<th>Markov</th>
</tr>
</thead>
</table>

Time (years)
CAV example: Comparison of estimates

State 3 prevalence

Prevalence

Isotonic NPMLE PP Markov

Time (years)
Limitations of isotonic regression method

- Need fully independent sampling process
  - ‘Doctor’s care’ (where next observation time determined as a result of disease status at current time) not acceptable.
- Range of models where approach permissible is limited.
  - Progressive models where there is only one path to any of the transient states.
  - Cannot incorporate covariates
  - Individual transition survival functions not identifiable
- In general, this approach to goodness-of-fit lacks power, whether use isotonic regression or NPMLE.
Semi-Markov models

- Usually Markov assumption made in analysis of panel observed multi-state data.
- Many medical applications where Markov assumption not reasonable
- Motivates use of semi-Markov models
  - Time homogeneous: $\lambda_{rs}(t, F_t) = \lambda_{rs}(s)$
  - Time inhomogeneous: $\lambda_{rs}(t, F_t) = \lambda_{rs}(t, s)$
Fitting semi-Markov models

- Likelihood for semi-Markov models not straightforward
  - Cannot factorise into distinct transitions.
  - All observations for a subject must be considered together.
- Progressive models: requires numerically evaluating multi-dimensional integral.
  - Feasible if number of states in model is small e.g. 3 or 4.
- Non-progressive models
  - Numerically solve integral equations:
    - Feasible if number of states is small and constrain one or more states to have exponential sojourn time
    - Constrain other states to have a ‘guarantee time.’
  - Simulation based estimation
    - Stopping time resampling.
Phase-type semi-Markov models

- Alternatively, constrain the semi-Markov model to have *phase-type* sojourn distributions.
  - Implies addition of latent states
  - Method common in analysis of stochastic processes (Erlang, Cox)
  - Limited use in statistical modelling (Crespi *et al* 2005)

- Advantages of this approach:
  - Likelihood is tractable: methods for hidden Markov models are applicable
  - Incorporation of classification error of observed states is straightforward
  - Existing software can be used to fit them in certain cases ( msm package in R)
Example: 3 state disease model with recovery

1: Healthy

2: BOS

3: Dead

$\lambda_{12}(s)$

$\lambda_{21}(s)$

$\lambda_{13}(s)$

$\lambda_{23}(s)$
Example: 3 state disease model with recovery
Phase-type semi-Markov model

- Transition intensities depend on $s$, time since entry into current state

$$\lambda(s) = \sum_{i} \mu_i \mathbb{P}\{\text{Phase} = i | \text{Time } s \text{ since entry into current state}\}$$

where $\mu_i$ are transition intensities out of the state conditional on being in phase $i$.

- Can reparametrise so that $\mu_2 = \mu_1 \tau$, then $\tau$ is a measure of departure from an exponential sojourn distribution.
  - $\tau > 1$ implies increasing transition intensities with respect to time since entry into the state
  - $\tau < 1$ implies decreasing intensities.
Likelihood for phase-type model

- Markov property does not apply

\[ L_j = P(X_{1j}, \ldots, X_{nj}) \]
\[ = P(X_{1j})P(X_{2j}|X_{1j})P(X_{3j}|X_{1j}, X_{2j}) \ldots P(X_{nj}|X_{1j}, \ldots, X_{(n-1)j}) \]
\[ \neq P(X_{1j})P(X_{2j}|X_{1j})P(X_{3j}|X_{2j}) \ldots P(X_{nj}|X_{(n-1)j}) \]

- Use the Forward algorithm for HMMs to calculate likelihood
- Can be expressed as a matrix product
  - Matrices composed of transition probabilities from the Markov chain and observation indicators

\[ e_{rs} = 1\{X(t) = s|X^*(t) = r\} \]

- For models with classification error, these become observation probabilities

\[ e_{rs} = P\{O(t) = s|X^*(t) = r\} \]
BOS dataset

- 364 post-lung transplantation patients at Papworth hospital
  - 242 heart-lung transplant patients, 122 double lung.
- Bronchiolitis obliterans syndrome
  - Deterioration in lung function after transplantation
  - Defined by decrease in FEV$_1$ compared to a baseline measure established in first 6 months after transplant.
  - Not defined before 6 months.
  - Cohort only includes 6 months survivors of transplant.
- Patients scheduled to visit clinic at 9 months, 12 months and then 6 monthly intervals.
  - Patients with infections had fortnightly visits for some periods
  - To limit effect of informative observation, only observations at the nearest times to those scheduled were considered.
- Observed BOS status subject to classification error.
BOS dataset: Results

- Large improvement in likelihood between the hidden Markov model and hidden semi-Markov models

<table>
<thead>
<tr>
<th>Model</th>
<th>-2 × LL</th>
<th>No. pars</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM</td>
<td>3005.06</td>
<td>9</td>
</tr>
<tr>
<td>HSMM</td>
<td>2976.51</td>
<td>13</td>
</tr>
</tbody>
</table>

- Modified likelihood ratio statistic also significant
  - 24.48 compared to approximate $\chi^2_2$.

- Semi-Markov model suggests decreasing transition intensities with respect to time since entry into a state

  $\hat{\tau}_1 = 0.34(0.18, 0.61), \hat{\lambda}_1 = 0.21(0.05, 0.91)$

  $\hat{\tau}_2 = 0.31(0.15, 0.67), \hat{\lambda}_2 = 1.37(0.29, 6.55)$. 
BOS dataset: Improvement in fit

- Improvement in the fit to the survival curve.

- General Pearson-type goodness-of-fit test suggests fit acceptable: $T = 37.1$ compared to approximate $\chi^2_{32}$. 

![Graphs showing survival curves for Heart-lung transplant and Double-lung transplant]
BOS dataset: Further issues

- Patient heterogeneity
  - Lung transplantation patients come from varied diagnostic groups e.g. cystic fibrosis, bronchiectasis, chronic obstructive pulmonary disease.
  - Trajectories of FEV\textsubscript{1} known to exhibit either acute or chronic decline (Jackson et al 2002).
  - Decreasing transition intensities could be due to frailty effects.

- Fluctuations in lung function
  - FEV\textsubscript{1} can fluctuate through time
  - If these fluctuations are on sufficiently long time scale, assumption of independent misclassification in our model may be inappropriate.
  - ‘True’ BOS process should be progressive, ie. recovery not possible.
Further areas of research

- Non-parametric estimates
  - Cases where independent sampling times can’t be assumed.
  - Semi-parametric estimates to incorporate covariates.

- Semi-Markov models
  - Incorporate time inhomogeneity.
  - Extension of phase-type models to other non-Markov models.
  - Overlap with random-effects models.
Acknowledgements and References

- Joint work with Linda Sharples at the MRC Biostatistics Unit, Cambridge.

References

- Turnbull B. The empirical distribution function with arbitrarily grouped, censored and truncated data. 1976, JRSS B.


Constructing formal tests of Markov property

- Given that $\tau = 1$ implies the process is time homogeneous Markov, can in principle construct a likelihood ratio test of the Markov property.
  - However, if $\tau = 1$, the parameter $\lambda$ governing the speed of progression from phase 1 to phase 2 becomes redundant and unidentifiable.
  - The likelihood ratio statistic therefore does not have a standard $\chi^2_{|\theta_1|−|\theta_0|}$ asymptotic null.
- Could test $\tau = 1$ by assuming $\lambda = \lambda_0$ as known.
  - But would lose power if $\tau \neq 1$ and $|\lambda − \lambda_0|$ is large.
Constructing tests of the Markov property

▶ Instead consider a penalised likelihood

\[ pl(\theta) = l(\theta) + \sum_r \{ C_r \log(\lambda_r) - C_r \lambda_r \alpha_r \} \]

▶ If \( C \) sufficiently large penalised LRT has approximate \( \chi^2_{|\tau|} \) null.
▶ Provided \( C \) not too large, will maintain power when \( |\lambda - 1/\alpha| \) is large.
▶ Avoids problems of identifiability if one or more state is close to being exponential.