Parametric Inference for Interval-Censored Data from Semi-Markov Multistate Models

Yang Yang
Department of Statistics
University of Michigan

2010 Joint Statistical Meetings

Joint work with Yves Atchade and Vijay Nair
Motivation

- The traditional failure time analysis can be viewed as a simple two-state model with ‘alive’ state and ‘failed’ state.
- In a general multistate modeling framework, there are a finite number of states and a subject spends a random amount of time in each state and then moves to another state.

When data on state occupancy times and transitions are available, multistate models lead to more informative inference than the regular failure time analysis.
Multistate Models: Simple examples

- **Usual two-state model**
  - alive \[\rightarrow\] fail

- **Repairable system – 2-state model**
  - alive \[\rightarrow\] fail 1 \[\rightarrow\] fail 2 \[\rightarrow\] ...

- **Multiples failure modes**
  - alive \[\rightarrow\] fail 1
  - alive \[\rightarrow\] fail 2
  - alive \[\rightarrow\] fail 3

- **Illness-death model**
  - well \[\rightarrow\] ill
  - ill \[\rightarrow\] death
Semi-Markov Multistate Models

Initial state distribution $\pi$, i.e. system starts in state $j$ with probability $\pi(j)$
$\rightarrow$ assumed that system $J(t)$ always starts in ‘good’ condition: $\pi(1) = 1$

$J(t)$ moving among states according to transition probability
$P_{ij}(s, t) \equiv \mathbb{P}(J(t) = j|J(s) = i)$.
$\rightarrow$ assumed here $J(t)$ is time-homogeneous: $P_{ij}(s, t) = P_{ij}(0, t - s)$

$P_{ij}(s, t)$ can be determined by:
$F_{ij}(t; \psi)$ – time spent in state $i$ before moving to state $j$
$p_{ij}$ – transition from state $i$ to state $j$

Denote $P(t; \theta) = \{P_{ij}(0, t; \theta)\}$, $\theta = (p_{ij}, \psi)$ the parameters of interest
Markov Multistate Models

Define transition intensity rates:

\[ q_{ij}(t) = \lim_{\Delta t \to 0} \frac{P_{ij}(t, t + \Delta t)}{\Delta t}, \quad j \neq i \]

Time-homegeneous case: \( q_{ij}(t) \equiv \lambda_{ij} \)

\( Q = \{\lambda_{ij}\} \) is the transition intensity matrix with \( \lambda_{ii} = -\sum_{j \neq i} \lambda_{ij} \)

For Markov models,

1. the parameters of interest: \( \theta = \{\lambda_{ij}\} \)

2. Especially, \( P(t; \theta) = \exp(Q(\theta)t) \)
Interval Censoring

In practice, data from the multi-state process $J(t)$ is available only periodically, leading to interval-censored data.

For instance, instead of exact transition times $\tau_s$ and visited states $i_s$, only the state information $x_k$ are observed at some discrete time points $t_k$.

$$Z = \{ (\tau_s, i_s); s = 1, \ldots, S \}$$

$$Y = \{ (t_k, x_k); k = 1, \ldots, K \}$$

Figure: $Z$ latent complete history, $Y$ observed censored history
Inference Based On Interval-censored Data

Likelihood for this example:

\[ L(\theta|Y) = \mathbb{P}(J(t_0) = x_0, J(t_1) = x_1, J(t_2) = x_2, J(t_3) = x_3) \]

If \( J(t) \) is Markovian, the likelihood can be simplified to

\[ L(\theta|Y) = \mathbb{P}(J(t_0) = x_0) \times \mathbb{P}(J(t_1) = x_1|J(t_0) = x_0) \times \mathbb{P}(J(t_2) = x_2|J(t_1) = x_1) \]
\[ \times \mathbb{P}(J(t_3) = x_3|J(t_2) = x_2) \]
\[ = \pi(x_0)P_{x_0,x_1}(t_1 - t_0; \theta) \times P_{x_1,x_2}(t_2 - t_1; \theta) \times P_{x_2,x_3}(t_3 - t_2; \theta), \]

which can be maximized numerically to estimate \( \theta \).

However, this simplification cannot be extend to general semi-Markov models. The likelihood function involves high dimensional integral.
Challenge with Interval-censored Data

Underlying $J(t)$

One observed sample path $Y = \{(t_k, x_k); k = 1, 2, 3\}$

Can we derive any information about $Z = \{(\tau_s, \iota_s); s = 1, \ldots, S\}$?

In reality, we don’t even know the number of transitions $S$. 
Multistate **Degradation** Models: Progressive or Acyclic

**One-step progressive**

**Multi-step progressive**

One observed sample path \( Y = \{(t_k, x_k); k = 1, 2, 3\} \)

Denote \( \delta = (i_s) \) the complete path.

For progressive models, there are only a **finite** number of possible paths.
Likelihood-based Inference

The incomplete data likelihood $L(\theta|Y)$ and the score function

$$h(\theta, Y) \equiv \nabla_\theta \log L(\theta|Y)$$

are difficult to compute directly.

Let $L(\theta|Z)$ be the complete data likelihood and $H(\theta, Z) \equiv \nabla_\theta \log L(\theta|Z)$. Note that

$$h(\theta, Y) = \int H(\theta, Z)p(Z|\theta, Y)dZ$$

where $p(Z|\theta, Y)$ is the conditional distribution of $Z$ given $\theta$ and $Y$. 
Likelihood-based Inference (contd.)

The complete data likelihood $L(\theta|Z)$ is easy to obtain even for general semi-Markov models. But $p(Z|\theta, Y)$ is difficult to work with.

Use Monte Carlo methods to sample $\{Z_1, Z_2, ..., Z_m\}$ from $p(\cdot|\theta, Y)$ and empirically estimate the score function

$$h(\theta, Y) = \int H(\theta, Z)p(Z|\theta, Y)dZ$$

by

$$\bar{H}(\theta, Z) = \frac{1}{m} \sum_{i=1}^{m} H(\theta, Z_i)$$

Here we suppress $\bar{H}$’s dependency on $Y$.

Details of sampling will be discussed later.
Stochastic Approximation

Gu and Kong (PNAS, 1998) – modified version of Robbins & Monro (1951) Recall (Louis, 1982) that $\mathbf{I}_Y(\theta) = \mathbf{I}_Z(\theta) - \mathbf{I}_{Z|Y}(\theta)$: estimate $\mathbf{I}_Y(\theta)$ by

$$\bar{I}(\theta, Z) = \frac{1}{m} \sum_i I(\theta, Z_i) - \frac{1}{m} \sum_i H(\theta, Z_i)H(\theta, Z_i)^T + \bar{H}(\theta, Z)\bar{H}(\theta, Z)^T.$$ 

Let $\gamma_n$ be a decreasing nonnegative sequence: $\sum \gamma_n = \infty$ and $\sum \gamma_n^2 < \infty$. Updating algorithm for $\theta_n$ and the information matrix $\Gamma_n$:

$$\begin{align*}
\Gamma_n &= \Gamma_{n-1} + \gamma_n (\bar{I}(\theta_{n-1}, Z^{(n)}) - \Gamma_{n-1}) \\
\theta_n &= \theta_{n-1} + \gamma_n \Gamma_{n-1}^{-1} \bar{H}(\theta_{n-1}, Z^{(n)})
\end{align*}$$

Here $Z^{(n)} = (Z^{(n)}_i, i = 1, \ldots, m)$, with $Z^{(n)}_i \sim p(\cdot | \theta_{n-1}, Y)$.

Convergence of $\theta_n$ and $\Gamma_n$ (Gu and Kong, 1998)
Bayesian Inference via Data Augmentation

Similarly, Bayesian inference can be made as long as sampling $Z$ is possible. Assume prior $p(\theta)$ for $\theta$. Update through

(i). For each observation $j$ in the sample, conditional on $Y_j$ and $\theta_{n-1}$, augment complete history $Z_j^{(n)}$ from $p(\cdot | \theta_{n-1}, Y_j)$ (possibly via MCMC sampling).

(ii). Sample $\theta_n$ from conditional distributions depends on $Z^{(n)}$.

$$p(\theta | Y, Z^{(n)}) = p(\theta | Z^{(n)})$$

Repeat the two steps many times, and the draws of $\theta$ are eventually from the posterior $p(\theta | Y)$ as desired.
Sampling $Z$ : One-step Progressive

Given $\{x_1, ..., x_k\}$, the sequence of transitions $\{i_1, ..., i_s\}$ is known.
Example : $x_1 = 1$, $x_2 = 2$, $x_3 = 4$ →
Only the sojourn times $\{\tau_1, ..., \tau_s\}$ are unknown.
Suppose $\tau_i \sim f_{i,i+1}(\cdot)$, independent.

The distribution of $Z$ given $\theta$ and $Y$ is

$$g(\tau_1, \tau_2, \tau_3 | \theta, Y) = f_{12}^\theta(\tau_1)f_{23}^\theta(\tau_2)f_{34}^\theta(\tau_3)$$

$$\times 1\{t_1 < \tau_1 \leq t_2, \ t_2 < \tau_1 + \tau_2 \leq t_3, \ t_2 < \tau_1 + \tau_2 + \tau_3 \leq t_3\}.$$  

Gibbs sampling can be used to sample from this conditional distribution.
Sampling $Z$ : Multi-step Progressive

Example: $x_1 = 1$, $x_2 = 2$, $x_3 = 4$ →

Now, both transitions $\{i_1, \ldots, i_s\}$ and sojourn times $\{\tau_1, \ldots, \tau_s\}$ unknown.

The distribution of the complete history $Z = (\delta, \tau)$ given $\theta$ and $Y$ is

$$p(\delta, \tau | \theta, Y) \propto \sum_k 1\{\delta = k\} g_k(\tau),$$

where each $g_k$ represents the unnormalized truncated distribution as the one-step progressive case. Note that the number of values $k$ can take is finite.

**Complication:** The dimension of the unknown $Z = (\delta, \tau)$ can vary.
Sampling Z : Multi-step Progressive (contd.)

Two approaches:

1. Calculate the conditional probabilities of each path $\delta$, given $\theta_n$ and $Y$, then use Gibbs sampling to sample as the one-step progressive case;

2. Use reversible jump MCMC to handle the varying dimension directly.

In our experience, reversible jump MCMC is computationally faster.
Sampling $Z$ : Multi-step Progressive (contd.)

1. Conditional Sampling: calculate

$$p_1 = \frac{\int g_1(\tau'_1, \tau'_2)d\tau'}{\int g_1(\tau'_1, \tau'_2)d\tau' + \int g_2(\tau_1, \tau_2, \tau_3)d\tau'}$$

sample $\delta$ from Multinomial(1, $p_1$, 1 − $p_1$), then sample $\tau \sim g_\delta$.

2. Reversible Jump MCMC Sampling: construct a dimension matching transformation as follows:

$T : (u, \tau'_2) \rightarrow (\tau_2, \tau_3)$ with $u \sim U(0, 1)$

$$\tau_2 = \tau'_2, \quad \tau_3 = u(t_3 - \tau'_2)$$

The Jacobian factor is $|T| = t_3 - \tau'_2$. 
Illustration of algorithm in Markov case:

Dash line – direct MLE; red – conditional; green – RJMCMC

$n = 100, \quad t_k - t_{k-1} = 0.4$
Simulation study: 3-state semi-Markov

\[ F_{12}(t) = F_{13}(t) = \text{Weib}(\gamma_1, \beta_1) \]

\[ F_{23}(t) = \text{Weib}(\gamma_2, \beta_2) \]

\[ \beta_1 = 0.8, \quad \gamma_1 = 1.2 \]

\[ \beta_2 = 1, \quad \gamma_2 = 2 \]

\[ p_{12} = 0.4 \]

\[ n = 100, \quad t_k - t_{k-1} = 0.3 \]

RJMCMC about 30 times faster.
Remarks

- Computationally intensive approach but general progressive models → for moderate number of states
- Extensions to covariates (time non-varying)
- Non-progressive models – possible in principle
- Nonparametric estimator