Robust Inference for Event Probabilities with non-Markov Event Data

Dave Glidden, UCSF

03/07/02

Outline

- Multistate Event Data (MED)
  correlated, structured events in 1 subject
- A summary for MED
  'event probabilities'
- Nonparametric estimation
  product integration of transition rates
- Asymptotic theory
- Numerical studies
- Reflections

Slide 1

Multistate Event Data

- Subjects at risk for multiple events
  e.g., illness and death
- Data is stochastic process
  process value = subject's state
- Events have a 'longitudinal' structure
- Analysis should account for clustering

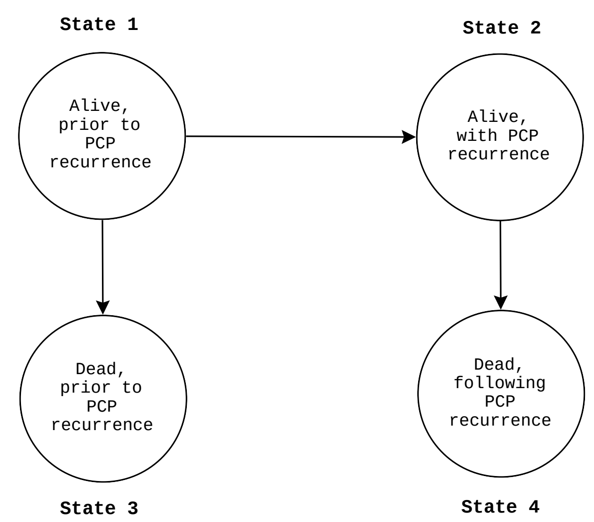
Slide 2

An Example

- AIDS Clinical Trial (ACTG021)
- Two treatments for recurrent pneumonia (PCP)
- Subjects at risk for death and/or pneumonia
- Can be represented a stochastic process, $X(t)$
  value of $X(\cdot)$ 'codes' the states
- Transitions likely non-Markov

Slide 3

Multistate Representation



Slide 4

Summaries for MED

- **Aalen-Johansen:**
  $P_{jk}(s,t) := \mathrm{pr}[X(t) = k | X(s) = j]$
  prob of moving from j at s to k at t
  $\{P(s,t)\}_{jk} = P_{jk}(s,t)$
- **My interest:**
  $P_k(t) := P_{1k}(0,t) = \mathrm{pr}[X(t) = k]$
  prob of being in state k at t
  $P(t)$ top row of $P(0,t)$
- **My objective:**
  Estimate $P(t)$, non-Markov rt-censored $X(\cdot)$

Slide 5

The Function $P(t)$

- $P(t) = \{P_1(t), \ldots, P_K(t)\}, t \in [0, \tau]$
- $P_k(t) = \mathrm{pr}\{X(t) = k\}$
  prob subject in state k at time t
- Examples: survivor fn, cumulative incidence fn
  Temkin's 'prob of being in response' fn
- Simple, interpretable, fully marginal

Slide 6

My Work

- Estimate $P(t)$ for rt. censored, non-Markov $X(\cdot)$
- Motivate Markov estimator for non-Markov data
  Aalen/Johansen (AJ) Estimator
- Show AJ estimator is consistent, Gaussian
- Develop robust confidence bands
- Evaluate performance in simulations
- Illustrate the method in practice

Slide 7

Data and Notation

- $X_i(t)$: state of subject i at time t
  $X(0) = 1;\ i = 1, \ldots, n;\ t \in [0, \tau]$
- p: K dim row vector, $p_k = \text{pr}\{X(0) = k\}$
- $X(\cdot)$ takes values in $\{1, \ldots, K\}$
- $N_{ijk}(t) = \#\{s \leq t, X_i(s-) = j, X_i(s) = k\}, \quad (j \neq k)$
  number of moves from j to k for subject i by t
- $Y_{ij}(t) = I\{X_i(t-) = j\}$
  indicates $X_i(\cdot)$ was in state j prior to t

Slide 8

More Notation

- Let $N_i(t)$ be K by K matrices, $\quad (i = 1, \ldots, n)$
- $\{N_i(t)\}_{jk} = N_{ijk}(t), \quad (j \neq k)$
- $N_{ijj}(t) := -\sum_{k \neq j} N_{ijk}(t), \quad (j = 1, \ldots, K)$
- Let $Y_{iI}(t)$ be K by K diagonal matrices
- $\{Y_{iI}(t)\}_{jj} = Y_{ij}(t), \quad (j = 1, \ldots, K)$
- $\bar{N}(t) := n^{-1} \sum_{i=1}^{n} N_i(t), \quad \bar{Y}_I(t) := n^{-1} \sum_{i=1}^{n} Y_{iI}(t)$
- $F_t = \sigma\{X_i(s), 0 \leq s \leq t, i = 1, \ldots, n\}$
  increasing σ-algebras generated by history of process

Slide 9

Aalen & Johansen (1978)

- $A_{jk}(dt) = E\{N_{jk}(dt) | F_{t-}\}, \quad (j \neq k)$
- $A_{jj}(t) = -\sum_{k \neq j} A_{jk}(t), \quad (j = 1, \ldots, K)$
- $A(t)$, K by K matrix
- $\hat{A}_{jk}(t) = \int_0^t \bar{Y}_j^{-1}(s) \bar{N}_{jk}(ds)$, $\hat{A}(t) = \int_0^t \bar{Y}_I^{-1}(s) \bar{N}(ds)$
- $P(s,t) = \prod_{(s,t]} (I + dA)$
  $\prod$ *product integration, I identity matrix*
- $\hat{P}(s,t) = \prod_{(s,t]} (I + d\hat{A})$

Slide 10

The (Matrix) Product Integral

- Denoted $\prod$
- G: a Borel measure on $(0,\tau]$, $G(t) := G\{(0,t]\}$
- The product integral over $(0,t]$ is

$$\lim_{\max|t_i - t_{i-1}| \to 0} \Pi_{i=1}^{n} \{I + G(t_i) - G(t_{i-1})\},$$

- $0 = t_0 < t_1, \ldots, t_n = t$ is a partition of $(0,t]$
- Π, matrix multiplication
- Gill & Johansen (1990) *Ann. Statist.*

Slide 11

Aalen & Johansen's Results

- $\hat{A}_{jk}(t) - A_{jk}(t)$ is a (F_t)-martingale
- $\hat{A}_{jk}(t)$ consistent, Gaussian
- Gill & Johansen (1990)
  - continuity of $\prod$ implies $\hat{P}(s,t)$ consistent
  - compact diff of $\prod$ implies $\hat{P}(s,t)$ Gaussian

Slide 12

Non-Markov Data

- Pepe and Cai: 'partly conditional transition rate'
  $\Lambda_{jk}(dt) = E\{N_{jk}(dt)|Y_j(t)=1\}, \quad (j \neq k)$
- $\Lambda_{jj}(t) := -\sum_{k\neq j} \Lambda_{jk}(t)$
- Markov data: $\Lambda_{jk}(dt) = A_{jk}(dt)$
- Non-Markov data: $\Lambda_{jk}(dt) = E\{A_{jk}(dt)|Y_j(t)=1\}$
- $\Lambda_{jk}(dt)$, transition rate from j at $t-$ to k at t
- $\Lambda(t)$: a K by K matrix

Slide 13

$\Lambda(t)$ and $P(t)$

- $E\{N_{jk}(dt)\} = P_j(t-)\Lambda_{jk}(dt)$
- $1(X(t)=k) = 1(X(0)=k) + \sum_{j\neq k}^{K}\{N_{jk}(t) - N_{kj}(t)\}$
- Taking expectations
  $P_k(t) = p_k + \sum_{j\neq k}^{K} \int_0^t \{P_j(s-)\Lambda_{jk}(ds) - P_k(s-)\Lambda_{kj}(ds)\}$
  $= P(t) = p + \int_0^t P(s-)\Lambda(ds)$
- By Gill and Johansen
  $P(t) = p \prod_{(0,t]} \left(I + d\Lambda\right)$

Slide 14

Nonparametric Estimation

- Pepe & Cai (1993) proposed $\hat{\Lambda}_{jk}(t) = \int_0^t \overline{Y}_j^{-1}(s)\overline{N}_{jk}(dt)$,
- $\hat{\Lambda}(t) = \int_0^t \overline{Y}_I^{-1}(s)\overline{N}(ds)$
- Let $\hat{P}(t) = p\prod_{(0,t]}\left(I + d\hat{\Lambda}\right)$
- $\hat{P}(t)$, matrix product over obs. event times
- Censoring absent: $\hat{P}_k(t) = n^{-1}\sum_{i=1}^{n} 1\{X_i(t) = k\}$

Slide 15

$\hat{P}(t)$ is the Aalen-Johansen Estimator

- $\hat{\Lambda}(t) = \int_0^t \overline{Y}_I^{-1}(s)\overline{N}(ds) = \hat{A}(t)$
- $A(t) := \Lambda(t)$ only when data Markov
- Thus, $\hat{\Lambda}(t) = \hat{A}(t)$ but $\Lambda(t) \neq A(t)$
- Note, $\hat{P}(t) =$ top row of $\hat{P}(0,t)$, implies
- $\hat{P}(\cdot)$ is the Aalen-Johansen estimator of $P(\cdot)$
- Motivated the Aalen-Johansen w/o Markov assumption

Slide 16

Consistency of $\hat{P}(t)$

- $N(t) := E\{\bar{N}(t)\}$, $Y_I(t) := E\{\bar{Y}_I(t)\}$
- By Glivenko Cantelli
  $\sup|\bar{N} - N| \to 0$, $\sup|\bar{Y}_I - Y_I| \to 0$
- That implies
  $\sup|\hat{\Lambda} - \Lambda| \to 0$
- Continuity of $\prod$ implies $\sup\|\hat{P}_k - P_k\| \to 0$,
  $(k = 1, \ldots, K)$

Slide 17

Weak Convergence

$\sqrt{n}(\hat{P} - P)(t)$ asymp equiv to $W(t) := n^{-\frac{1}{2}} \sum_{i=1}^{n} \Phi_i(t)$

$$\Phi_i(t) = p \int_0^t \prod_{(0,s)} (I + d\Lambda) \Psi_i(ds) \prod_{(s,t]} (I + d\Lambda)$$

$W(t) \Rightarrow G(t)$, a K-dim Gaussian process on $(0, \tau]$

$$\Psi_i(t) = \int_0^t Y_I^{-1}(s) \{N_i(ds) - Y_{iI}(s)\Lambda(ds)\},$$

$\mathrm{cov}\{G_j(s), G_k(t)\} = \xi_{jk}(s,t) = E\{\Phi_{1j}(s)\Phi_{1k}(t)\}$,
$G_j(t)$ jth element of $G(t)$, $\Phi_{1j}(t)$ jth element of $\Phi(t)$

Slide 18

A Useful Process, $\tilde{W}(t)$

- For arbitrary processes $\xi_{jk}(s,t)$ is not a simple function
- Not analytically tractable
- Define $\tilde{W}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \hat{\Phi}_i(t) Z_i$
- $\hat{\Phi}(t)$ estimate of $\Phi(t)$
- $Z_1, \ldots, Z_n$ are iid N(0,1), indep of data
- $\tilde{W}(t)$ has same limiting distn as $\sqrt{n}(\hat{P} - P)(t)$
- $\tilde{W}(t)$ can be obtained by simulation

Slide 19

Approximating $\sqrt{n}(\hat{P} - P)$

- $\tilde{W}(t)$ can be obtained by simulation
- Generate N(0,1) $(z_{1m}, \ldots, z_{nm})$ $(m = 1, \ldots, M)$
- $\tilde{W}^{(m)}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \hat{\Phi}(t) z_{im}$
- Idea due to Lin et al. (1993)
- Can use to develop confidence bands for $\sqrt{n}(\hat{P} - P)$
  Hall-Wellner (HW) or equal precision (EP) bands

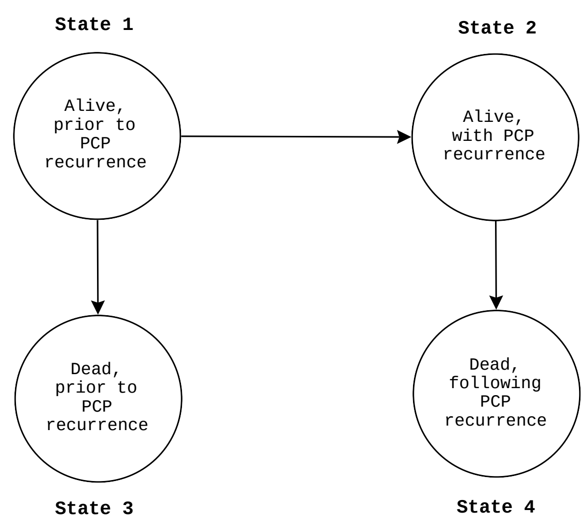
Slide 20

Simulation Studies

- Studied two multistate models
  illness death, 2 state with return
- Examined point estimation
  Markov data and Non-Markov data
  compared with ’naive’ estimator
- Looked at confidence band coverage
- Variety of sample sizes, censoring %

Slide 21

AIDS Example Set-Up



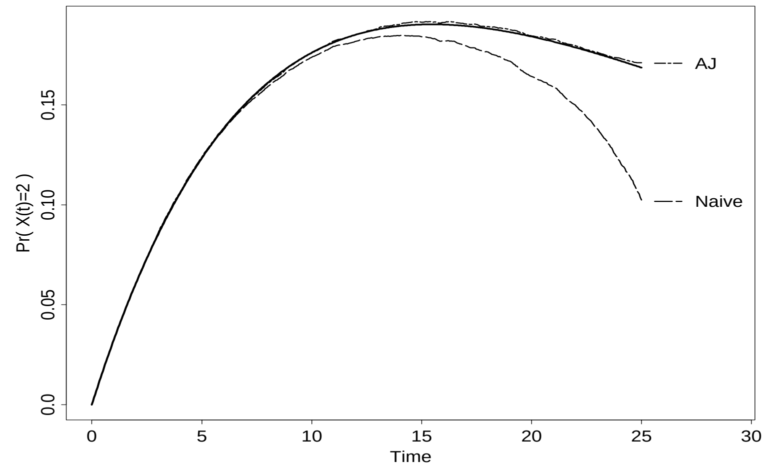
Slide 22

Point Estimation

- Considered PCP example, state 2
- Generated non-Markov data
- $\hat{P}_2(t)$, proportion alive with pneumonia
- Naive: prop in state 2 among those uncensored at t
- Aalen-Johansen approximately unbiased
- AJ does well w/small n, heavy censoring
- 5,000 simulations

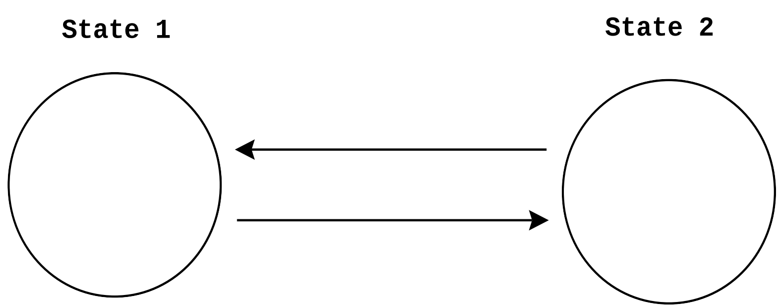
Slide 23

Bias: $n = 50$, 50% censoring



Slide 24

Two state model with return



Slide 25

Two-State Simulations

- Mixing dist to generate Non-Markov
- $\xi_i, (i = 1, \ldots, n)$ latent random effect
- ξ Gamma rv, mean 1 var θ
- $E\{N_{i12}(dt)|(F_{t-}), \xi_i\} = \alpha_1 \xi_i, \quad t \in [0, 1]$
- $E\{N_{i21}(dt)|(F_{t-}), \xi_i\} = \alpha_2 \xi_i, \quad t \in [0, 1]$
- θ=0, Markov
- Used $\theta = 0, 2, 4$

Slide 26

Confidence Band Coverage

- Empirical coverage of equal precision bands
- M=500, simulations from $\tilde{W}(\cdot)$
- Two-state setting
- With & without continuity correction (cc vs ncc)
- Effect of sample size n
- Effect of dependence between transitions
- 25% censoring, 5,000 simulations

Slide 27

Empirical Coverage (5000 Simulations)

nobs		θ=0	θ=2	θ=4
50				
	cc	0.94	0.95	0.96
	ncc	0.93	0.94	0.96
100				
	cc	0.95	0.95	0.95
	ncc	0.94	0.94	0.95
200				
	cc	0.95	0.95	0.95
	ncc	0.95	0.95	0.95

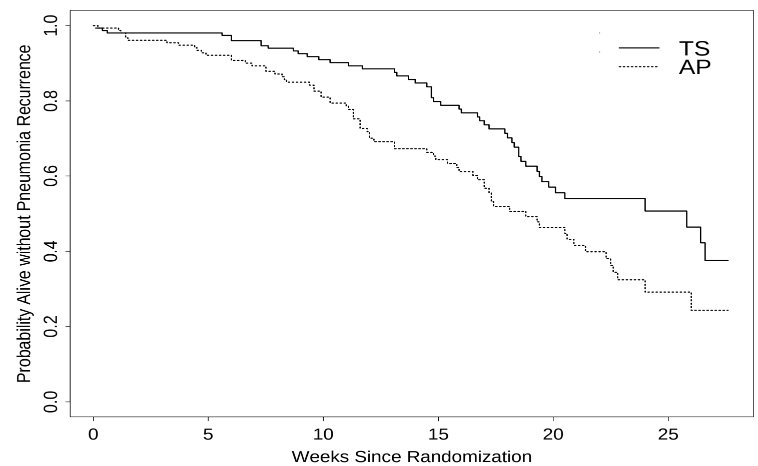
Slide 28

AIDS Example

- AIDS Clinical Trial (ACTG021)
- Two treatments for recurrent pneumonia (PCP)
- TS: Trimethoprim Sulfmethoxazole
  daily oral antibiotic
- AP: Aerosolized Pentamidine
  monthly inhaled rx
- 310 subjects: 154 TS, 156 AP

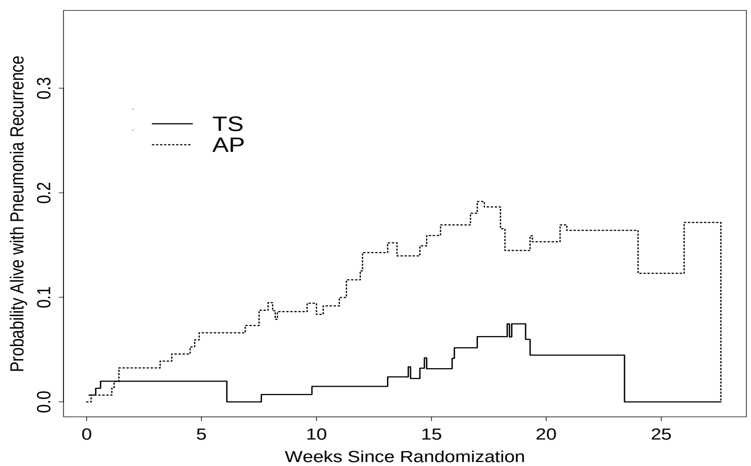
Slide 29

ACTG 021 – State 1



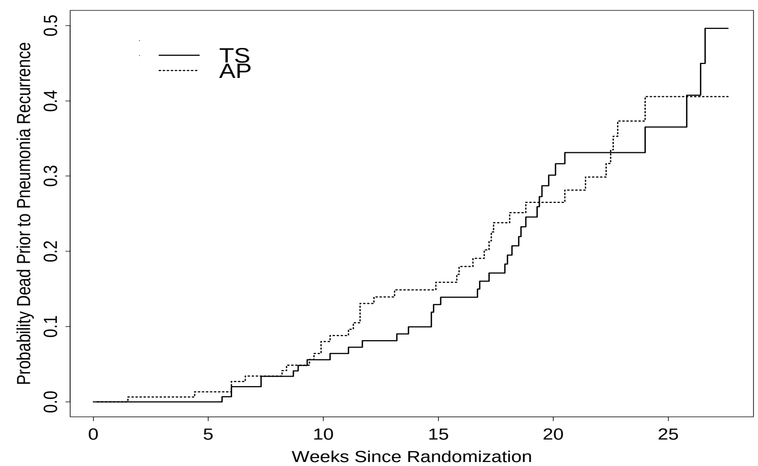
Slide 30

ACTG 021 – State 2



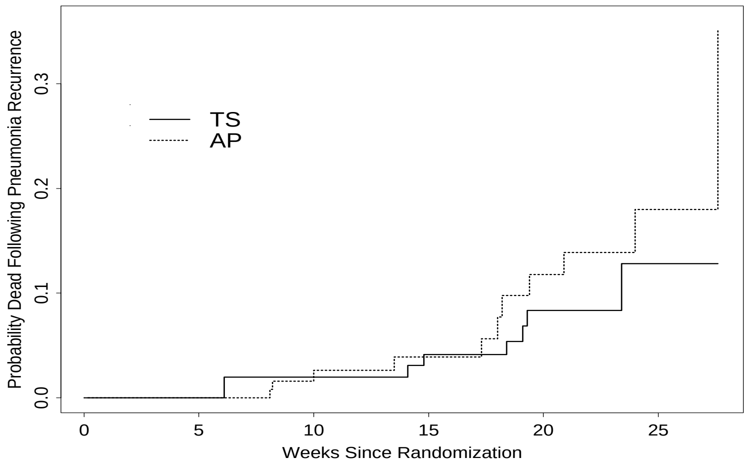
Slide 31

ACTG 021 – State 3



Slide 32

ACTG 021 – State 4



Slide 33

Discussion

- Aalen Johansen estimator is 'robust'
- Comparison with Wei (1989) natural
  'working independence' consistent for popn parameters
- Naive 'pointwise' bands have correct coverage
- Can relax assumptions
  $X(0) = 0$ *and* $s = 0$
- $P(\cdot)$: population prevalence
  has advantages, disadvantages

Slide 34

Many Thanks to:

- R. A. Olshen
- C.E. McCulloch
- Ø. Borgan
- S.C. Shiboski

Slide 35