

Robust inference for event probabilities with non-Markov event data

BY DAVID V. GLIDDEN

Department of Epidemiology and Biostatistics, University of California,

San Francisco, California 94143-0560, U.S.A.

email: david@biostat.ucsf.edu

SUMMARY

Multistate event data, in which a single subject is at risk for multiple events, is common in biomedical applications. This paper considers nonparametric estimation of the vector of probabilities of state membership at time t . Estimators, derived under the Markov assumption, have been shown (Datta and Satten, 2001) to be consistent for data which is non-Markov. Inference, however, must take into account possibly non-Markov transitions when constructing confidence bands for event curves. We develop robust confidence bands for these curves, evaluate them via simulation and illustrate the method on two datasets.

Key words: Aalen-Johansen estimator; Confidence Band; Multivariate failure time data; Product integral.

1. INTRODUCTION

In many biomedical contexts subjects are followed over a period of time and may experience events of multiple types. This yields so-called multistate event data. Survival analysis, repeated events data, competing risks data and the illness-death model are all examples of this type of data. This data can be represented as a stochastic process, $\tilde{X}(t)$, with its value at time t denoting a subject's state.

Consider the following example of multistate event data. Figure 1 shows a diagram of a multistate representation of an AIDS clinical trial (Hardy et al., 1992). This trial compared two treatments for secondary prevention of a serious AIDS-related complication, *Pneumocystis carinii* pneumonia (PCP). Subjects in the trial were at risk of dying and/or contracting PCP. All patients began in state 1 'alive without PCP'. They moved either to 'alive with secondary PCP' (state 2) or to 'dead prior to secondary PCP' (state 3). From state 2 they can then progress to 'dead following PCP' (state 4). This setting was also considered and described in detail by Datta, Satten and Datta (2000) who termed it the 'three-stage irreversible illness-death model'. See Andersen (1988), Begg and Larson (1982) or Klein, Keiding and Copelan (1993) for other biomedical examples of multistate event data.

Figure 1 about here.

The analysis of multistate event data is complicated by multiple events within a subject. Simple summaries can be very useful for describing the experience of a group or comparing the impact of an intervention. A familiar summary for such data is a K -dimensional row-vector process

$$P(t) = \{P_1(t), \dots, P_K(t)\}, \quad t \in [0, \tau]$$

where $P_k(t) = \text{pr}\{\tilde{X}(t) = k\}$ is the probability that a subject is in state k at time t . This summary has been discussed by many authors (Datta et al., 2000; Hoover et al., 1996; Pepe,

Longton and Thornquist, 1991; Temkin, 1978). Datta et al. (2000) termed $P(t)$ as the 'stage-occupation probabilities' and both Datta et al. (2000) and Hoover et al. (1996) provide an excellent description of the interpretation and utility of $P(\cdot)$ in biomedical examples.

An estimator of $P(t)$ is available for the non-homogeneous Markov case using the estimator proposed by Aalen and Johansen (1978). However, the Markov assumption is often not realistic in practice. For example, in an AIDS study it was shown (Finkelstein et al., 1996) that patients who develop one kind of AIDS-associated infection develop other infections and/or die more quickly. This observation would make a Markov assumption for the aforementioned AIDS study unrealistic.

Recently, Datta and Satten (2001) studied the properties of the Aalen-Johansen estimator for independently censored non-Markov data. They proved that the estimator is consistent even when the state transitions are non-Markov. We review their result in Section 2 and extend it to argue that $\sqrt{n}(\hat{P} - P)$ converges weakly to a Gaussian process. We introduce a simulation technique to approximate the distribution of the process and to develop confidence bands for the estimator. Lastly, simulation studies and examples from an AIDS clinical trial and a leukemia study demonstrate the method in practice.

2. PRELIMINARIES

2.1. *Data and notation*

Let $\tilde{X}_i(t) (i = 1, \dots, n)$ be a stochastic process denoting the state of a subject i at time $t \in [0, \tau]$. Assume the process $\tilde{X}(\cdot)$ takes integer values from 1 to K , with K bounded. Each process begins at a common starting time in an index state, arbitrarily coded 1. Thus, $\text{pr}\{\tilde{X}(0) = 1\} = 1$. The value of the process corresponds to an arbitrary coding of possible states. Assume the process has right-continuous sample paths and makes a bounded number of transitions. However, we make no assumptions about the possible structure of transitions.

The data can be represented by a set of counting and at-risk processes. Define:

$$\tilde{N}_{ijk}(t) = \#\{s \leq t : \tilde{X}_i(s-) = j, \tilde{X}_i(s) = k\} \text{ and } \tilde{Y}_{ij}(t) = 1\{\tilde{X}_i(t-) = j\}$$

($j = 1, \dots, K; k \neq j; i = 1, \dots, n$) where $1(\cdot)$ is the indicator function. The counting process $\tilde{N}_{ijk}(t)$ counts the number of direct transitions between state j and state k for the i th subject observed by time t . The process $\tilde{Y}_{ij}(t)$ is a binary left-continuous function which indicates whether the i th subject was in state j at $t-$. The history of events to time t generates the filtration

$$F_t = \sigma\{\tilde{N}_{ijk}(s), \tilde{Y}_{ij}(s+), 0 \leq s \leq t, i = 1, \dots, n, j = 1, \dots, K, k \neq j\}.$$

We allow the process to be subject to random right-censoring. Let C_i represent a right-censoring time for subject i . Assume that $\tilde{X}_i(\cdot)$ and C_i are independent, with $\{\tilde{X}_i(\cdot), C_i\}$ independent and identically distributed ($i = 1, \dots, n$). The observed data are $\{X_i(t), H_i(t), t \in [0, \tau]\}$ ($i = 1, \dots, n$). The process $X_i(t) := \tilde{X}_i(t \wedge C_i)$, is the right-censored realization of $\tilde{X}_i(\cdot)$ at t and $H_i(t) := 1(C_i \geq t)$ indicates whether the process is observed at time t , where $a \wedge b = \min(a, b), t \in [0, \tau]$. The censored counting and at-risk processes are

$$N_{ijk}(t) = \#\{s \leq t : X_i(s-) = j, X_i(s) = k, H_i(s) = 1\} = \int_0^t 1(C_i \geq s) \tilde{N}_{ijk}(ds) \quad (j \neq k)$$

and

$$Y_{ij}(t) = 1\{X_i(t-) = j, H_i(t) = 1\} = \tilde{Y}_{ij}(t) 1(C_i \geq t).$$

Define n independent and identically distributed K by K matrices, $N_i(t)$ and $Y_{iD}(t)$. The (j, k) th element ($j \neq k$) of matrix $N_i(t)$ is $N_{ijk}(t)$. The j th diagonal element of $N_i(t)$ is

$$- \sum_{k \neq j} N_{ijk}(t), \quad (j = 1, \dots, K).$$

The matrix $Y_{iD}(t)$ is a diagonal matrix with j th diagonal element $Y_{ij}(t)$.

2. CONSISTENT ESTIMATION OF $P(t)$

Datta and Satten (2001) showed that the function $P(t)$ could be written as the product integral of a rate function. Specifically, the representation uses the 'partly conditional transition rate' proposed by Pepe and Cai (1993). The partly conditional transition rate $\lambda_{jk}(t)$ from state j to k at time t is defined as

$$E\{\tilde{N}_{ijk}(dt)|\tilde{Y}_{ij}(t) = 1\} = \lambda_{jk}(t)dt \quad j \neq k, \quad (i = 1, \dots, n) \quad (2.1)$$

The cumulative transition rate from j to k is $\Lambda_{jk}(t) = \int_0^t \lambda_{jk}(s)ds$. The rate function is easy to work with because it conditions only on the immediate past. Thus, even when the processes paths depend in a complex way on its history, (2.1) retains a simple form.

Define

$$\Lambda_{kk}(t) = - \sum_{j \neq k}^K \Lambda_{kj}(t)$$

and let $\Lambda(t)$ be a K by K matrix of the cumulative rate functions in (2.1) with elements $\Lambda(t) = \{\Lambda_{jk}(t)\}$. Then Datta and Satten (2001) showed that

$$P(t) = p \prod_{(0,t]} (I + d\Lambda), \quad (2.2)$$

where I is the K by K identity matrix and the function \prod is the matrix-valued product integral. The 1 by K row vector p has k th element $p_k = \text{pr}\{\tilde{X}(0) = k\}$. All processes start in state 1, so the first element of p is 1 and all others are zero. Thus, p times the product integral simply extracts the top row of the matrix. The product integral is defined for finite matrix-valued real measures on the Borel subsets of $(0, \tau]$. Let G be such a measure with $G(t)$ denoting $G\{(0, t]\}$. The product integral over $(0, t]$ is the limit of

$$\lim_{\max |t_l - t_{l-1}| \rightarrow 0} \prod_{l=1}^L \{I + G(t_l) - G(t_{l-1})\},$$

where $0 = t_0 < t_1, \dots, t_L = t$ is a partition of $(0, t]$, and Π denotes matrix multiplication. For more on product integration consult the definitive work of Gill and Johansen (1990).

Equation (2.2) establishes the relationship between $P(\cdot)$ and the rate matrix $\Lambda(\cdot)$. Datta and Satten consider the estimator for $\Lambda_{jk}(\cdot)$ from independently right-censored data of the

form

$$\hat{\Lambda}_{jk}(t) = \int_0^t \bar{Y}_j^{-1}(s) \bar{N}_{jk}(ds) \quad (j = 1, \dots, K; k = 1, \dots, K).$$

where $\bar{Y}_j(t) = n^{-1} \sum Y_{ij}(t)$ and $\bar{N}_{jk}(t) = n^{-1} \sum N_{ijk}(t)$. Let $\bar{N}(t) := n^{-1} \sum N_i(t)$ and $\bar{Y}_D(t) := n^{-1} \sum Y_{iD}(t)$. Using this notation, an estimator of $\Lambda(\cdot)$ is

$$\hat{\Lambda}(t) = \int_0^t \bar{Y}_D^{-1}(s) \bar{N}(ds).$$

The form of (2.2) suggests the estimator of $P(t)$. The estimator is

$$\hat{P}(t) = \{\hat{P}_1(t), \dots, \hat{P}_K(t)\} = p \prod_{(0,t]} (I + d\hat{\Lambda}). \quad (2.3)$$

The estimator (2.3) is just the Aalen-Johansen estimator. Datta and Satten provide the consistency of the estimators for both $\hat{\Lambda}$ and \hat{P} for non-Markov data. The consistency of P follows because of the representation (2.1) and the continuity of the product integral, which was shown by Gill and Johansen (1990).

3. ASYMPTOTICS AND INFERENCE

3.1. Asymptotic theory

The product integral is not only continuous but also compactly differentiable (Gill and Johansen, 1990; Andersen et al. 1993, p. 111) in an appropriate topology. Thus, if weak convergence is established for $\hat{\Lambda}(\cdot)$, then this theory can be transferred to $\hat{P}(\cdot)$. In Appendix A, empirical process theory gives large-sample results for $\bar{N}(\cdot)$ and $\bar{Y}_D(\cdot)$. By representing $(\hat{P} - P)(\cdot)$ as a series of well-behaved maps acting on $\bar{N}(\cdot)$ and $\bar{Y}_D(\cdot)$, we argue the desired result. The template for the proof is Gill and Johansen's proof of weak convergence for the Kaplan-Meier estimator. In Appendix A we show that $n^{\frac{1}{2}}(\hat{P} - P)(t)$ converges weakly to $G(t) = \{G_1(t), \dots, G_K(t)\}$, a vector of Gaussian processes, and that $n^{\frac{1}{2}}(\hat{P} - P)(t)$ has the same limiting distribution as the process $W(t) := \{W_1(t), \dots, W_K(t)\} := n^{-\frac{1}{2}} \sum \Phi_i(t)$ where

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

and

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

where $Y_D(s) = \lim_{n \rightarrow \infty} \bar{Y}_D(s)$. The elements of the vector $G(t)$ have limiting covariance:

$$\text{cov}\{G_j(s), G_k(t)\} = \xi_{jk}(s, t) = E\{\Phi_{1j}(s)\Phi_{1k}(t)\}, \quad 0 \leq s \leq t \leq \tau,$$

($j = 1, \dots, K$), ($k = 1, \dots, K$), where $\Phi_{1k}(t)$ is the k th element of the row-vector process $\Phi_1(t)$. The covariance function can be consistently estimated by $\hat{\xi}_{jk}(s, t) = n^{-1} \sum \hat{\Phi}_{ij}(s)\hat{\Phi}_{ik}(t)$ where

$$\hat{\Phi}_i(t) = p \int_0^t \prod_{(0,s)} (I + d\hat{\Lambda}) \bar{Y}_D^{-1}(s) \{N_i(ds) - Y_{iD}(s)\hat{\Lambda}(ds)\} \prod_{(s,t)} (I + d\hat{\Lambda}).$$

The consistency proof for the covariance function is found in Appendix B.

The complicated nature of the covariance function precludes analytical evaluation of the process. However, we exploit a flexible technique developed by Lin, Wei and Ying (1993) to approximate the distribution of $n^{\frac{1}{2}}(\hat{P} - P)(\cdot)$ through simulation.

To approximate $n^{\frac{1}{2}}(\hat{P} - P)(\cdot)$, we use the process $\tilde{W}(t) = \{\tilde{W}_1(t), \dots, \tilde{W}_K(t)\}$, with

$$\tilde{W}(t) = n^{-\frac{1}{2}} \sum_{i=1}^n \hat{\Phi}_i(t) Z_i,$$

where (Z_1, \dots, Z_n) are independent standard normal random variables that are also independent of the data. It can be shown that, conditional on the observed data, $\tilde{W}(t)$ is asymptotically tight with covariance function $\hat{\xi}(s, t)$. By consistency of $\hat{\xi}(s, t)$, shown in Appendix B, the limiting distributions of $n^{\frac{1}{2}}(\hat{P} - P)(\cdot)$ and $\tilde{W}(\cdot)$ are identical. To simulate from the distribution of $\tilde{W}(t)$, we generate independent standard normal random variables (z_{1m}, \dots, z_{nm}) ($m = 1, \dots, M$) where M is the desired number of realizations of $\tilde{W}(\cdot)$. Let

$$\tilde{W}^{(m)}(t) = \{\tilde{W}_1^{(m)}(t), \dots, \tilde{W}_K^{(m)}(t)\} = n^{-\frac{1}{2}} \sum_{i=1}^n \hat{\Phi}_i(t) z_{im}, \quad (m = 1, \dots, M).$$

The simulation method will form the basis for developing confidence bands for $n^{\frac{1}{2}}(\hat{P} - P)(\cdot)$.

3.2. Simultaneous confidence bands

We can develop a $(1 - \alpha)$ simultaneous confidence band for $\hat{P}_k(\cdot)$ over the interval $[t_1, t_2]$. Various types of confidence bands for $\hat{P}_k(\cdot)$ can be developed by considering the family of transformed processes

$$B_k(t) = n^{\frac{1}{2}} g_n^{(k)}(t) [\phi_k\{\hat{P}_k(t)\} - \phi_k\{P_k(t)\}], \quad (k = 1, \dots, K).$$

To improve coverage, we use the functions $\phi_k(\cdot)$ to transform $P_k(\cdot)$ from $[0, 1]$ to $(-\infty, \infty)$. The weight functions $g_n^{(k)}(t)$ determine the shape of the bands in time. Let $\phi_k(\cdot)$ have a non-zero derivative $\phi_k'(\cdot)$ and the $g_n^{(k)}(t)$ converge to non-negative bounded functions $g^{(k)}(t)$, $(k = 1, \dots, K)$. The process $B_k(t)$ is then asymptotically equivalent to $\hat{B}_k(t) = g_n^{(k)}(t) \phi_k'\{\hat{P}_k(t)\} W_k(t)$ which can be approximated by the process

$$\tilde{B}_k(t) = g_n^{(k)}(t) \phi_k'\{\hat{P}_k(t)\} \tilde{W}_k(t),$$

where $\tilde{W}_k(t)$ is the k th element of $\tilde{W}(t)$. Let constants $q_\alpha^{(k)}(t_1, t_2)$ $(k = 1, \dots, K)$ satisfy

$$\text{pr} \left\{ \sup_{t_1 \leq t \leq t_2} |\hat{B}_k(t)| > q_\alpha^{(k)}(t_1, t_2) \right\} := \alpha, \quad (k = 1, \dots, K).$$

We then form an approximate $(1 - \alpha)$ confidence band for $\phi_k(P_k)$ using

$$\phi_k\{\hat{P}_k(t)\} \pm n^{-\frac{1}{2}} q_\alpha^{(k)}(t_1, t_2) / g_n^{(k)}(t), \quad t_1 \leq t \leq t_2.$$

This band can be converted to a band for $\hat{P}_k(\cdot)$. The constants $q_\alpha^{(k)}(t_1, t_2)$ $(k = 1, \dots, K)$ are obtained by the following simulation method. We sample M realizations of $\tilde{B}_k^{(m)}(t)$ using

$$\tilde{B}_k^{(m)}(t) = g_n^{(k)}(t) \phi_k'\{\hat{P}_k(t)\} \tilde{W}_k^{(m)}(t),$$

which is equal to

$$\tilde{B}_k^{(m)}(t) = g_n^{(k)}(t) \phi_k'\{\hat{P}_k(t)\} \left\{ n^{-\frac{1}{2}} \sum_{i=1}^n \hat{\Phi}_{ik}(t) z_{im} \right\}, \quad (m = 1, \dots, M).$$

Define $\tilde{q}_m^{(k)}(t_1, t_2) = \sup_{t \in [t_1, t_2]} \left| \tilde{B}_k^{(m)}(t) \right|$. Then an estimate of $q_\alpha^{(k)}(t_1, t_2)$ is the $(1 - \alpha)$ percentile of $\tilde{q}_m^{(k)}(t_1, t_2)$, $(m = 1, \dots, M)$.

It is possible to obtain a continuity-corrected band (Lin et al., 1993) by using

$$\tilde{q}_m^{(k)}(t_1, t_2) = \sup_{t \in [t_1, t_2]} \left| \tilde{B}_k^{(m)}(t) \vee g_n^{(k)}(t-) \left| n^{\frac{1}{2}} [\phi_k\{\hat{P}_k(t-)\} - \phi_k\{\hat{P}_k(t)\}] + \frac{\tilde{B}_k^{(m)}(t)}{g_n^{(k)}(t)} \right| \right|.$$

Lin et al. reported that the continuity correction improves coverage, particularly in small samples.

We choose $\phi_1(x)$ to be $\log\{-\log(x)\}$ and $\phi_k(x) = \log\{-\log(1-x)\}$ for $k = 2, \dots, K$ because it is analogous to the familiar transformation used for the Kaplan-Meier estimator. Weight functions can be chosen which reduce to equal-precision (EP) bands (Nair, 1984) or Hall-Wellner (HW) bands (Hall and Wellner, 1980) when $K = 1$. The weight function $g_{1k}(t)$ for EP bands and $g_{2k}(t)$ for HW bands are given below:

$$g_{1k}(t) = \frac{\hat{P}_k(t)}{\hat{\sigma}_k(t)}, \quad g_{2k}(t) = \frac{\hat{P}_k(t)}{1 + \hat{\sigma}_k^2(t)}, \quad \hat{\sigma}_k^2(t) = \frac{\hat{\xi}_{kk}(t, t)}{\hat{P}_k(t)}, \quad (k = 1)$$

$$g_{1k}(t) = \frac{1 - \hat{P}_k(t)}{\hat{\sigma}_k(t)}, \quad g_{2k}(t) = \frac{1 - \hat{P}_k(t)}{1 + \hat{\sigma}_k^2(t)}, \quad \hat{\sigma}_k^2(t) = \frac{\hat{\xi}_{kk}(t, t)}{1 - \hat{P}_k(t)}, \quad (k = 2, \dots, K).$$

4. SIMULATION STUDIES

Simulation studies were undertaken to examine the empirical coverage of confidence bands constructed by the methods in the last section. Two settings are explored. The first is the illness-death setting represented in Figure 1. We also consider a second setting with two states. All patients begin in state 1 but may move freely and repeatedly between states 1 and 2. This two-state model has been used to model unemployment or psychiatric admissions (e.g., Keiding and Andersen, 1989).

The illness-death simulations borrow from the approach of Couper and Pepe (1997). To generate data, a constant hazard is used for transitions into state 2 and state 3. Specifically, $\lambda_{12}(t) = 0.054$ and $\lambda_{13}(t) = 0.048$, $t \in [0, 25]$. A non-Markov transition from state 2 to state 4 is achieved by setting the time spent in state 2 to be a constant multiple of the time of entry into state 2. If T_2 is the time of entry into state 2, then the time of entry into state 4 is $(1 + f)T_2$, where $f = 1.76$. Note that the process is neither fully Markov nor is it semi-Markov. The censoring times were uniformly distributed on $[0, 73]$ and $[0, 26]$, which censors 25% and 50% of subjects respectively.

The second set of simulations used a mixing distribution to induce non-Markov transitions. Each subject had a latent random effect ξ sampled from a gamma distribution with mean 1 and variance θ . The parameter θ governs the degree of dependence between transitions. Conditional on ξ , the transition intensities are $\alpha_{12}(t) = \xi\lambda_1$ and $\alpha_{21}(t) = \xi\lambda_2$, $t \in [0, 1]$. In the simulations, the dependence parameter was set to 0.0, 2.0 and 4.0. When θ equals 0.0, transitions are Markov. A value of $\theta = 4.0$ produces highly dependent transition times. The event probability for state 1 is

$$P_1(t) = \frac{\lambda_2}{\lambda_1 + \lambda_2} + \frac{\lambda_1}{\lambda_1 + \lambda_2} \{1 + \theta(\lambda_1 + \lambda_2)t\}^{-\theta^{-1}}, \quad \theta > 0,$$

and

$$P_1(t) = \frac{\lambda_2}{\lambda_1 + \lambda_2} + \frac{\lambda_1}{\lambda_1 + \lambda_2} \exp\{-(\lambda_1 + \lambda_2)t\}, \quad \theta = 0$$

with $P_2(t) = 1 - P_1(t)$. To ensure that event probabilities remained similar as θ varied, values of λ_1 and λ_2 are also shifted. We set $\lambda_1/\lambda_2 = 4$ and $\lambda_1 + \lambda_2 = \log(2)$, 6.25, and 10.5 when $\theta = 0, 2.0$ and 4.0, respectively.

Simulations were conducted to evaluate the empirical coverage probabilities of the 0.95 confidence bands. To improve coverage of the confidence bands, bands were restricted to [5,20] for the illness-death setting and [0.10,0.80] for the two-state setting. Censoring was varied at 0%, 25% and 50% and the sample sizes were 50, 100 and 200. The empirical coverage probabilities of the confidence bands were estimated from 5000 simulated datasets. For each simulated dataset, the boundary value q_α was estimated from 500 realizations of $\tilde{W}(\cdot)$.

The results for the illness-death setting are given in Table 1. They show that the EP and HW bands perform well for all states. Table 2 shows coverage for EP bands for P_1 both with and without the continuity correction (CC and NCC, respectively). We find accurate coverage for both the Markov ($\theta = 0$) and non-Markov settings. When the censoring is heavy and the sample size is large, the NCC bands give approximately correct coverage but the CC bands tend to cover more frequently than the nominal level.

Tables 1 and 2 about here.

5. DATA EXAMPLES

5.1. AIDS clinical trial

Hardy et al. (1992) conducted a randomized clinical trial for prevention of a serious AIDS-related complication. Three hundred ten AIDS patients who had recently recovered from a bout of *Pneumocystis carinii* pneumonia (PCP) were randomized to receive either trimethoprim sulfamethoxazole (TS) or aerosolized pentamidine (AP). One hundred fifty four patients received TS and 156 patients received AP. The study followed patients for PCP recurrence and/or death. The data analysis of Hardy et al. examined the effect of the

treatments on PCP recurrence and death separately. Figure 1 shows a possible graphical representation of the data as a multistate model. A multistate treatment provides a more unified analysis of the effect of treatment on both PCP and death.

We applied the Aalen-Johansen method, separately in the two treatment groups. It appears that there is little difference between treatment in risk of death. However, non-lethal PCP recurrence is more frequent in the AP group. This is demonstrated by estimating and graphing the probability a subject is in state 2. These curves, shown in Figure 2, graph the probability that a subject is alive with a PCP recurrence as a function of treatment and weeks of follow-up. The figure also includes EP bands from the curves from 9.5 to 20.5 weeks for the TS group and from 5.0 to 24.0 weeks for the AP group.

Figure 2 about here.

5.2. Bone marrow transplant for leukemia

We analyze data from a multicenter study of allogenic bone marrow transplantation for acute leukemia. Copelan et al. (1991) report the results of bone marrow transplants for 137 patients with either acute myelogenous leukemia or acute lymphoblastic leukemia. Following the transplant, patients are at risk of multiple events, including the development of acute graft-versus-host disease (GVHD) and platelet recovery, which is the return of platelets to a normal levels ($\geq 40 \times 10^9/l$). In addition, a transplant may fail because leukemia returns or because a patient dies prior to remission. One way to track the success of a transplant is to report the proportion of patients in the most advantageous state. These are the patients who are alive, have not relapsed, and have recovered platelets. We term this state as state 2. Figure 3 shows $\hat{P}_2(t)$ with both EP and HW confidence bands. We see the initial increase in the first 50 days in $\hat{P}_2(t)$ as patients recovery platelets. After day 100, $\hat{P}_2(t)$ decreases as some patients succumb to relapse or death.

Figure 3 about here.

6. DISCUSSION

It is simple to relax the index state condition that $\text{pr}\{X(0) = 1\} = 1$. First, group subjects by their state at time 0, then apply the Aalen-Johansen method to each group separately. Finally, pool the estimates by calculating a weighted linear combination using weights proportional to the number in each state at time 0.

Comparison of these results to results on estimation of marginal hazard ratios for cluster failure time data (Wei, Lin and Weissfeld, 1989) is natural. Like Wei, a 'working independence' estimator consistently estimates a population-level parameter. However, inference requires taking account about possible clustering in assessing variability. Here, we develop robust covariance estimator to correct for dependence.

The pointwise bands for $P(t)$ do not depend on the covariance structure of $n^{\frac{1}{2}}(\hat{P} - P)(\cdot)$. They depend only on $\text{var}\{(\hat{P} - P)(t)\}$, which is the same for Markov and non-Markov data. Thus pointwise bands based on a Markov assumption (Andersen, et al. 1993, p. 294) will give correct coverage even for non-Markov data. Only simultaneous confidence bands need to account for possible non-Markov transitions.

Our method is unable to accommodate censoring schemes which depend on the history of the process unless the data are, in fact, Markov. A useful extension of our approaches would be to allow for stage-dependent censoring, like the method of Datta et al. (2000).

ACKNOWLEDGMENT

The author thanks Ø. Borgan, C.F. Spiekerman, C.E. McCulloch, S.C. Shiboski, D. Hart and M.C. Wang for their advice and assistance. He thanks R.D. Gill for the product integral symbol. This research was supported by grant HL65411 from the U.S. National Heart, Lung and Blood Institute.

REFERENCES

- Aalen, O.O. and Johansen, S. (1978). An empirical transition matrix for non homogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* **5**, 141-150.
- Andersen, P.K. (1988). Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Statistics in Medicine* **7**, 661-670.
- Andersen, P.K., Borgan, Ø., Gill, R.D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer-Verlag.
- Begg, C.B. and Larson, M. (1982). A study of the use of the probability-of-being-in-response function as a summary of tumor response data. *Biometrics* **38**, 59-66.
- Copelan, E.A., Biggs, J.C., Thompson, J.M., Crilley, P., Szer, J., Klein, J.P., Kapoor, N., Avalos, B.R., Cunningham, I., Atkinson, K., Downs, K., Harmon, G.S., Daly, M.B., Brodsky, I., Bulova, S.I., and Tutschka, P.J. (1991). Treatment for acute myelocytic leukemia with allogenic bone marrow transplantation following preparation with Bu/Cy. *Blood* **78**, 838-843.
- Couper, D. and Pepe, M.S. (1997). Modelling prevalence of a condition: chronic graft-versus-host disease after bone marrow transplantation. *Statistics in Medicine* **16**, 1551-15571.
- Datta, S., Satten, G.A., and Datta, S. (2000). Nonparametric estimation for the three-stage irreversible illness-death model. *Biometrics* **56**, 841-847.
- Datta, S. and Satten, G.A. (2001). Validity of the Aalen Johansen estimators of stage occupation probabilities and Nelson Aalen integrated transition hazards for Non-Markov models. *Statistics and Probability Letters* **55**, 403-411.

- Finkelstein, D.M., Williams, P.L., Molenberghs, G., Feinberg, J., Powderly, W.G., Kahn, J., Dolin, R., and Cotton, D. (1996). Patterns of opportunistic infections in patients with HIV infection. *Journal of Acquired Immune Deficiency Syndromes* **12**, 38-45.
- Gill, R.D. and Johansen, S. (1990). A survey of product-integration with a view towards applications in survival analysis. *Annals of Statistics* **18**, 1501-1555.
- Hall, W.J. and Wellner, J.A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67**, 133-143.
- Hardy, W.D., Feinberg, J., Finkelstein, D.M., Power, M.E., He, W., Kaczka, C., Frame, P.T., Holmes, M., Waskin, H., Fass, R.J., Powderly, W.G., Steigbigel, R.T., Zuger, A., and Holzman, R.S. (1992). A controlled trial of trimethoprim-sulfamethoxazole or aerosolized pentamidine for secondary prophylaxis of *Pneumocystis carinii* pneumonia in patients with the acquired immunodeficiency syndrome. *New England Journal of Medicine* **327**, 1842-1848.
- Hoover, D.R., Peng, Y., Saah, A.J., Detels, R.R., Day, R.S., and Phair, J.P. (1996). Using multiple decrement models to estimate risk and morbidity from specific AIDS illnesses. *Statistics in Medicine* **15**, 2307-2321.
- Keiding, N. and Andersen, P.K. (1989). Nonparametric estimation of transition intensities and transition probabilities: A case study of a two-state Markov process. *Applied Statistics* **38**, 319-329.
- Klein, J.P., Keiding, N., and Copelan, E.A. (1993). Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplantation patients. *Statistics in Medicine* **12**, 2315-2332.
- Lin, D.Y., Wei, L.J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557-572.

- Nair, V.N. (1984). Confidence bands for survival functions with censored data: a comparative study. *Technometrics* **26**, 265-275.
- Pepe, M.S., Longton, G., and Thornquist, M. (1991). A qualifier Q for the survival function to describe the prevalence of a transient condition. *Statistics in Medicine* **10**, 413-421.
- Pepe, M.S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates. *Journal of the American Statistical Association* **88**, 811-820.
- Temkin, N.R. (1978). An analysis for transient states with applications to tumor shrinkage. *Biometrics* **34**, 571-580.
- van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer-Verlag.
- Wei, L.J., Lin, D.Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065-1073.

APPENDIX A

Weak convergence for $\sqrt{n}(\hat{P} - P)(\cdot)$

Denote $D^{K,K}[0, \tau]$ as the space of K by K matrices which are right-continuous processes with left-hand limits. Let $D_-^{K,K}[0, \tau]$ represent the space of K by K matrices with elements which are left-continuous processes with right-hand limits. We define the norm of a K by K matrix A as

$$\|A\| = \max_{1 \leq j \leq K} \sum_{k=1}^K |A_{jk}|$$

and the supremum norm of a K by K matrix-valued process on $[0, \tau]$ as

$$\|B\|_\infty = \sup_{t \in [0, \tau]} \|B(t)\|,$$

with the variation norm $\|\cdot\|_v$ analogously defined. The product space $D^{K,K}[0, \tau] \times D_-^{K,K}[0, \tau]$ employs the maximum supremum norm.

Define $N(t) = \lim_{n \rightarrow \infty} \bar{N}(t)$, $Y_D(t) = \lim_{n \rightarrow \infty} \bar{Y}_D(t)$. Under random censoring, $N(t) = \int_0^t \text{pr}(C \geq s) \tilde{N}(ds)$ and $Y_D(t) = \text{pr}(C \geq t) \tilde{Y}_D(t)$, where $\tilde{N}(\cdot)$ and $\tilde{Y}_D(\cdot)$ are $\lim_{n \rightarrow \infty} n^{-1} \sum \tilde{N}_i(\cdot)$ and $\lim_{n \rightarrow \infty} n^{-1} \sum \tilde{Y}_{Di}(\cdot)$ respectively.

Note $\Lambda(t) := \int_0^t \tilde{Y}_D(s)^{-1} \tilde{N}(ds) = \int_0^t Y_D(s)^{-1} N(ds)$ under random right censoring. Equation (2.2) established $P = p \mathcal{I}(I + d\Lambda)$.

Each of the functions $\bar{N}, \bar{Y}_D, N, Y_D$ are bounded functions. Because we assume a finite number of transitions in $[0, \tau]$, they may be written as the sum of monotone functions (even though they may not be monotone functions themselves) The function \hat{P} is the result of applying the mappings,

$$(\bar{N}, \bar{Y}_D) \rightarrow \hat{\Lambda} \rightarrow \hat{P},$$

which move through the spaces

$$D^{K,K}[0, \tau] \times D_-^{K,K}[0, \tau] \rightarrow D^{K,K}[0, \tau] \rightarrow D^{1,K}[0, \tau].$$

The proof of weak convergence proceeds by proving that the mappings are compactly differentiable where Y_D is a diagonal matrix and where the function $\|\Lambda\|_v < \infty$.

Weak convergence follows from a Donsker result for the empiricals combined with the compact differentiability of the mappings. The first mapping is a combination of diagonal matrix inversion with integration. This is compactly differentiable by adaptation of a proof by van der Vaart and Wellner (1996, p. 382). The final mapping, product integration, was shown to be compactly differentiable for processes of bounded variation by Gill and Johansen (1990). Recall that \bar{N} and \bar{Y}_D are bounded and note each can be written as the sum of monotone processes. It is easy to show that the finite dimensional distributions of $n^{\frac{1}{2}}(\bar{N} - N, \bar{Y}_D - Y_D)$ converge weakly to (Z_N, Z_Y) in $D^{K,K}[0, \tau] \times D_-^{K,K}[0, \tau]$. The representation of \bar{N} and \bar{Y} as the sum of monotone processes allows application of a result by van der Vaart and Wellner (1996, p. 215) which establishes that (Z_N, Z_Y) are tight Gaussian processes with right-continuous sample paths. The sample paths are continuous if the intensities of the process is absolutely continuous over $(0, \tau]$.

The mappings

$$f_1 : (x, y) \rightarrow v = \int_{[0, \cdot]} y^{-1} dx$$

$$f_2 : v \rightarrow w = p \prod_{[0, \cdot]} (I + dv)$$

are compactly differentiable at $x = N, y = Y_D$ and $v = \Lambda$ with derivatives

$$f_1'(x, y) \cdot (h_x, h_y) = \int_{[0, \cdot]} y^{-1} dh_x + \int_{[0, \cdot]} -h_y y^{-2} dx = (h_v)$$

$$f_2'(v) \cdot (h_v) = p \int_{[0, \cdot]} \prod_{(0, s)} (I + dv) dh_v \prod_{(s, \cdot]} (I + dv),$$

where h_y is a diagonal matrix. The derivatives act on $(h_x, h_y) = (Z_N, Z_Y)$. By the chain rule, $n^{\frac{1}{2}}(\hat{P} - P)(t)$ converges weakly to

$$p \int_0^t \prod_{(0, s)} (I + d\Lambda) \left\{ Y_D^{-1}(s) Z_N(ds) - Y_D^{-2}(s) Z_Y(s) N(ds) \right\} \prod_{(s, t]} (I + d\Lambda)$$

which is a 1 by K vector Gaussian process $G(t)$ on $D^{1,K}[0, \tau]$ with right-continuous sample paths. Note that whenever the j th element of Y_D is equal to zero, then the corresponding derivative is zero. That is because, for any such interval, it is easy to see that in such an interval both $\hat{\Lambda}_{jk}$ and \hat{P}_j are equal to their estimands, which are equal to zero.

The norm for $D^{1,K}[0, \tau]$ is equivalent to the more familiar maximum sup norm so the weak convergence is equivalent to weak convergence under the maximum sup norm. From the above, the covariance for $n^{\frac{1}{2}}(\hat{P} - P)(t)$ is not obvious. It is more transparent from the representation implied by theorem II.8.1 of Andersen et al. (1993, p. 111). By that theorem, $n^{\frac{1}{2}}(\hat{P} - P)(t)$ is asymptotically equivalent to

$$n^{\frac{1}{2}} p \int_0^t \prod_{(0,s)} (I + d\Lambda) \left[Y_D^{-1}(s) \{ \bar{N} - N \}(ds) - Y_D^{-2}(s) \{ \bar{Y}_D - Y_D \}(s) N(ds) \right] \prod_{(s,t]} (I + d\Lambda),$$

which is equal to

$$n^{\frac{1}{2}} p \int_0^t \prod_{(0,s)} (I + d\Lambda) Y_D^{-1}(s) \left\{ \bar{N}(ds) - \bar{Y}_D(s) Y_D^{-1}(s) N(ds) \right\} \prod_{(s,t]} (I + d\Lambda),$$

which is, by definition,

$$n^{-\frac{1}{2}} \sum_{i=1}^n p \int_0^t \prod_{(0,s)} (I + d\Lambda) \Psi_i(ds) \prod_{(s,t]} (I + d\Lambda) := n^{-\frac{1}{2}} \sum_{i=1}^n \Phi_i(t).$$

The covariance for $n^{\frac{1}{2}}(\hat{P} - P)(t)$ is then easily expressed by the mean of product of the components of Φ .

APPENDIX B

Consistency of covariance estimator $\hat{\xi}$

Define

$$\xi(s, t) = E\{\Phi_1(s)^T \Phi_1(t)\} \tag{B.1}$$

and

$$\hat{\xi}(s, t) = n^{-1} \sum_{i=1}^n \hat{\Phi}_i(s)^T \hat{\Phi}_i(t), \tag{B.2}$$

for any $0 \leq s \leq t \leq \tau$. Then $\xi_{jk}(s, t)$ and $\hat{\xi}_{jk}(s, t)$ are the (j, k) th elements of (B.1) and (B.2), respectively. Consistency of $\hat{\xi}_{jk}(s, t)$ is implied by

$$\left\| n^{-1} \sum_{i=1}^n \hat{\Phi}_i(s)^T \hat{\Phi}_i(t) - \Phi_i(s)^T \Phi_i(t) \right\| \rightarrow 0 \quad (\text{B.3})$$

in probability as $n \rightarrow \infty$. To show (B.3), it is sufficient to show

$$\max_{1 \leq i \leq n} \|\hat{\Phi}_i(t)\| = O_{p^*}(1), \quad (\text{B.4})$$

where p^* denotes outer probability, and

$$\max_{1 \leq i \leq n} \|\Phi_i(t)\| = O_{p^*}(1) \quad (\text{B.5})$$

and

$$\max_{1 \leq i \leq n} \|\hat{\Phi}_i(t) - \Phi_i(t)\| \rightarrow 0. \quad (\text{B.6})$$

The elements of $\hat{\Lambda}$ are monotone and bounded in probability, implying that $\|\hat{\Lambda}\|_\infty$ and hence $\|\hat{\Lambda}\|_v$ are also $O_{p^*}(1)$. This implies that $\max_{1 \leq i \leq n} \|\hat{\Psi}_i\|_v$ is also $O_{p^*}(1)$. For all t (B.4) is bounded above by

$$4 \left\| \prod_{(0, s]} (I + d\hat{\Lambda}) \right\|_\infty \|\hat{\Psi}_i\|_v \left\| \prod_{(s, t]} (I + d\hat{\Lambda}) \right\|_v. \quad (\text{B.7})$$

The third term of (B.7) is bounded above by $\|(I + \hat{\Lambda})\|_\infty \|\hat{\Lambda}\|_v$. The first term of (B.7) is bounded above by $\exp(\|\hat{\Lambda}\|_\infty)$. The bound for (B.5) is similar since $\|\Lambda\|_v$ is bounded. Finally (B.6) is bounded above by

$$\begin{aligned} & \max_{1 \leq i \leq n} \left\| \int_{(0, s]} \prod_{(0, s]} (I + d\hat{\Lambda}) \hat{\Psi}_i(ds) \left\{ \prod_{(s, \cdot]} (I + d\hat{\Lambda}) - \prod_{(s, \cdot]} (I + d\Lambda) \right\} \right\|_\infty + \\ & \left\| \int_{(0, s]} \prod_{(0, s]} (I + d\hat{\Lambda}) (\hat{\Psi}_i - \Psi_i)(ds) \prod_{(s, \cdot]} (I + d\Lambda) \right\|_\infty + \\ & \left\| \int \left\{ \prod_{(0, s]} (I + d\hat{\Lambda}) - \prod_{(0, s]} (I + d\Lambda) \right\} \Psi_i(ds) \prod_{(s, \cdot]} (I + d\Lambda) \right\|_\infty. \end{aligned}$$

The terms converge to 0 by the uniform consistency of $\hat{\Lambda}$, by the convergence of the empiricals, by the continuity of the product integral.

Fig. 1. Multistate representation of an AIDS clinical trial. PCP denotes *P. carinii* pneumonia.

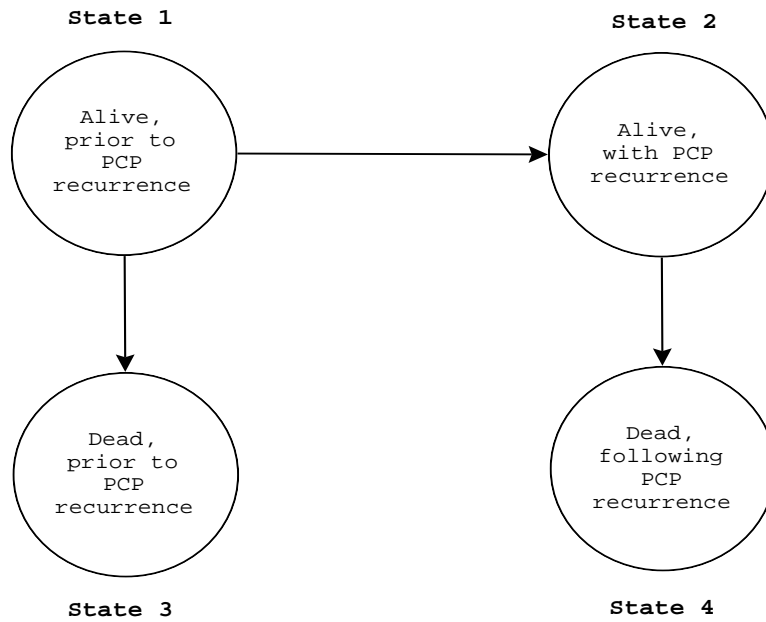


Fig. 2. The probability of being alive with a recurrence of pneumonia in AIDS trial with equal-precision confidence bands. Two treatments are displayed: aerosolized pentamidine (AP) and trimethoprim sulfamethoxazole (TS).

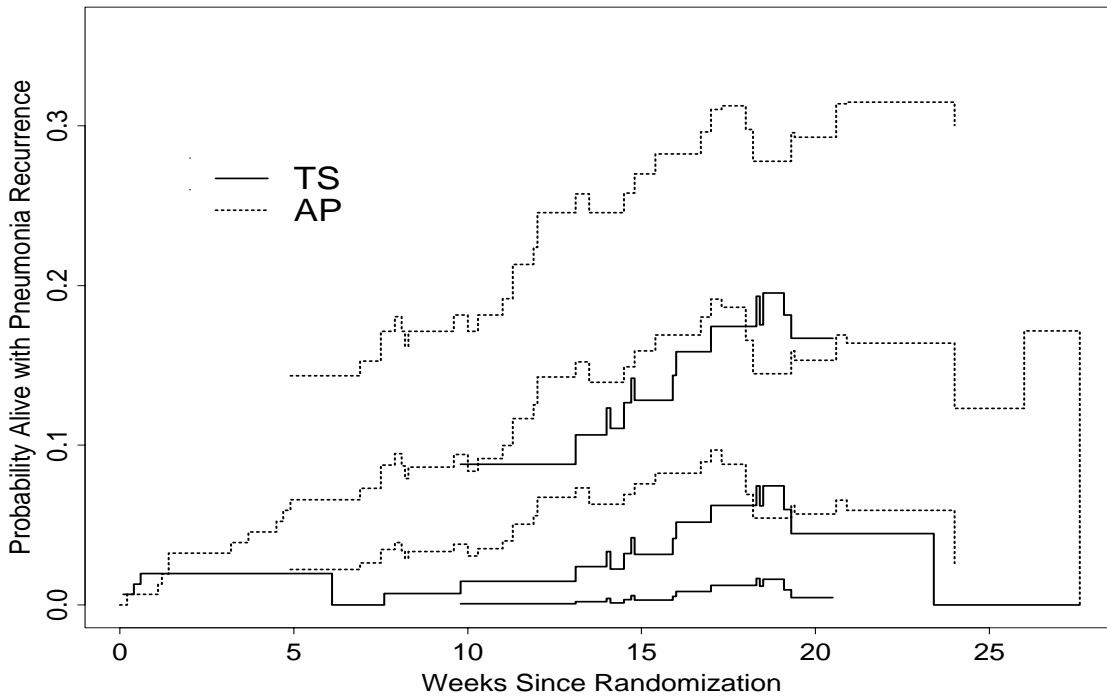


Fig. 3. The probability of being alive and recurrence-free with platelet recovery in the leukemia bone marrow transplant example, with Hall-Wellner type (HW) and equal precision type (EP) confidence bands.

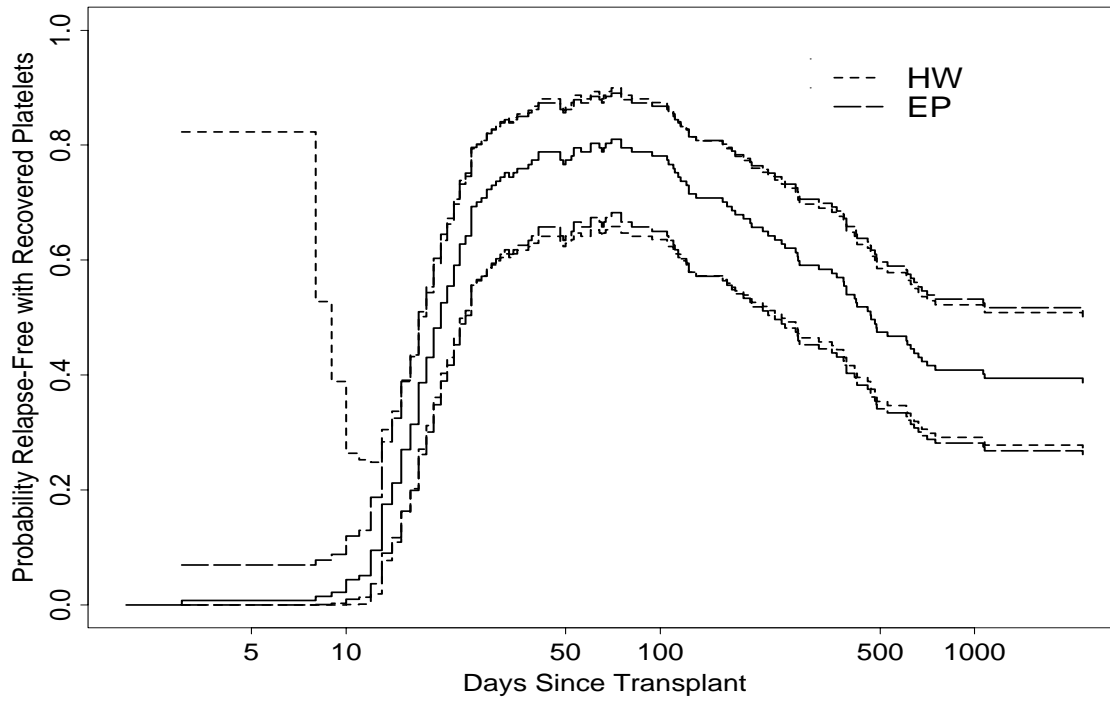


Table 1. *Empirical coverage of Equal-Precision (EP) and Hall-Wellner (HW) 0.95 confidence bands for the illness-death setting. Coverage is based on 5,000 simulations.*

Number of observations=nobs and censoring percentage=cens.

nobs	cens	K = 1		K = 2		K = 3		K = 4	
		EP	HW	EP	HW	EP	HW	EP	HW
50	0%	0.94	0.95	0.95	0.96	0.95	0.95	0.95	0.95
	25%	0.95	0.95	0.95	0.96	0.95	0.95	0.95	0.95
	50%	0.95	0.95	0.94	0.94	0.94	0.95	0.94	0.93
100	0%	0.94	0.95	0.95	0.96	0.94	0.94	0.95	0.94
	25%	0.94	0.95	0.94	0.96	0.95	0.94	0.95	0.94
	50%	0.94	0.95	0.95	0.95	0.94	0.94	0.95	0.94
200	0%	0.94	0.95	0.94	0.96	0.95	0.95	0.94	0.95
	25%	0.94	0.94	0.95	0.96	0.95	0.94	0.94	0.94
	50%	0.94	0.95	0.95	0.95	0.95	0.94	0.94	0.94

Table 2. Empirical coverage of Equal-Precision 0.95 continuity-corrected (CC) and uncorrected (NCC) confidence bands for $P_1(t)$. Coverage is based on 5000 simulations. Number of observations=*nobs*, censoring percentage=*cens*. Dependence is indexed by θ .

nobs	cens		$\theta=0$	$\theta=2$	$\theta=4$
50	0%	CC	0.94	0.94	0.96
		NCC	0.93	0.94	0.95
	25%	CC	0.94	0.95	0.96
		NCC	0.93	0.94	0.96
	50%	CC	0.95	0.95	0.97
		NCC	0.91	0.93	0.94
100	0%	CC	0.95	0.95	0.95
		NCC	0.94	0.94	0.95
	25%	CC	0.95	0.95	0.95
		NCC	0.94	0.94	0.95
	50%	CC	0.97	0.96	0.97
		NCC	0.94	0.94	0.94
200	0%	CC	0.95	0.95	0.96
		NCC	0.95	0.95	0.95
	25%	CC	0.95	0.95	0.95
		NCC	0.95	0.95	0.95
	50%	CC	0.97	0.97	0.96
		NCC	0.95	0.95	0.95