Bootstrap confidence bands for sojourn distributions in multistate semi-Markov models with right censoring

Ronald W. Butler
Department of Statistical Sciences, Southern Methodist University
Dallas, Texas 75275 USA
rbutler@smu.edu

Douglas A. Bronson
National Surgery Office, Veterans Health Administration
Denver, Colorado 80246 USA
douglas.bronson@va.gov

April 23, 2012

Summary

Transient semi-Markov processes have traditionally been used to describe the transitions of a patient through the various states of a multistate survival model. A survival distribution in this context is a sojourn through the states until passage to a fatal absorbing state or certain endpoint states. Using complete sojourn data, this paper shows how such survival distributions and associated hazard functions can be estimated nonparametrically and also how nonparametric bootstrap pointwise confidence bands can be constructed for them when patients are subject to independent right censoring from each state during the sojourn. Limitations to the estimability of such survival distributions that result from random censoring with bounded support are clarified. The methods are applicable to any sort of sojourn through any finite state process of arbitrary complexity involving feedback into previously occupied states.

Some key words: bootstrap; censoring; cofactor rule; proportional hazards; saddlepoint approximation; semi-Markov process; sojourn; survival analysis.

1. Introduction

Multistate survival models are usually transient semi-Markov processes whose states are considered clinically meaningful in the study of some disease. Subjects move among the states, sometimes repeating states, until reaching either a fatal absorbing state or some non-fatal endpoint state. Survival distributions that are of interest in such models are sojourn time distributions from an initial state to a destination state or a subset of such states. When complete histories of the sojourns of subjects through such systems are available, nonparametric statistical inference about such survival distributions is possible using the methodology developed in Butler & Bronson (2002) and further discussed in Butler (2007, Ch. 14). These references provide nonparametric function estimates and nonparametric bootstrap pointwise confidence bands for such survival functions and associated hazard and density functions. The basic idea is to use saddlepoint methods to determine function estimates and to compute resampled function estimates to which the BC_a or percentile bootstrap methods, as described in Davison & Hinkley (1997), can be applied to determine confidence bands.

This paper develops similar nonparametric methods to allow statistical inference when subjects are at risk of independent right censoring from each state of the semi-Markov process. Using complete histories about sojourns and censorings of subjects, we show how nonparametric function estimates and BC_a bootstrap pointwise confidence bands can be computed for the survival, hazard and density functions of arbitrary sojourns through general finite-state semi-Markov processes that may include transient states as well as irreducible subchains. Like the previous methods in Butler & Bronson (2002), this paper integrates transform methods, saddlepoint inversion and resampling methods. Transform methods determine the moment generating function of the sojourn distribution and an empirical moment generating function as its estimator. Saddlepoint inversion is required to obtain a survival estimate from this estimator. Finally, resampling methods are needed to obtain resampled survival estimates, from which confidence bands can be computed.

To accommodate random censoring, two key assumptions are made throughout most of the paper. First, right censoring times for subjects must be independent of all aspects of the subjects' sojourns through the semi-Markov process. This standard assumption is needed in order to use Kaplan–Meier estimators for exit-time distributions of the semi-Markov process. Secondly, the censoring mechanism must result in consistently estimable tails for exit-time distributions. Such estimability occurs when the distributional support of random censoring has least upper bound τ and $(0, \tau)$ encompasses the support of all exit-time distributions. Under such circumstances, we say the censoring is tail-estimable. Without tail-estimability, sojourn distributions are only estimable for $t < \tau$; see §4·3 for further discussion.

According to Chiang & Hsu (1976), the seminal paper to address the problem under consideration was Fix & Neyman (1951). The left panel of Fig. 1 is a flowgraph that is slightly more general than their multistate process and which will serve to illustrate our methods. The goal of Fix & Neyman was to model the survival time, or first passage from $1 \rightarrow 3$, of a patient under treatment for a disease such as cancer. Upon diagnosis, the patient enters state 1, an active state, and from there is subject to four competing risks, namely feedback into state 1, passage to recessive state 2, passage to death state 3, and right censoring, i.e. passage to state R_1. States 1 and 2 are transient states from which the patient may be censored, while state 3 is absorbing. To make the three-state process as general as possible, transitions $1 \rightarrow 1$, $2 \rightarrow 2$ and $1 \rightarrow \mathrm{R}_1$ have been added to the Fix & Neyman (1951) model and censoring may occur from all transient states. Apart from this, the main difference between the two models is that ours is a semi-Markov process whereas the original one is a time-homogeneous Markov process, which is parametric and has possibly unrealistic exponential holding times.

A less restrictive approach than Fix & Neyman noted by a referee maintains the Markov assumption for the multistate model but relaxes time-homogeneity. Nonparametric estimation of passage time distributions in this context uses the Aalen–Johansen (1978) estimator; see also Anderson, Hansen & Keiding (1991). Our approach is related by the fact that Aalen–Johansen estimators are used for estimating individual steps of the semi-Markov process; it differs in that these individual steps are assembled according to semi-Markov dynamics that violate Markov assumptions over multiple steps. See

§4·3 for further details.

Parametric approaches with semi-Markov assumptions have been considered by Butler & Huzurbazar (1997), when there is no censoring, and by Lô, Heritier & Hudson (2008), under independent right censoring. Such parametric semi-Markov models can be more realistic and useful than time-homogeneous Markov models, but inference is based on parametric assumptions which our nonparametric methods are designed to avoid. Apart from these references, no general methodology has been proposed for estimating sojourn times in semi-Markov processes, principally because the emphasis has been on addressing such problems in the time domain. Feedback among states as in Fig. 1 is difficult to handle in the time domain but not in the transform domain.

The nature and difficulty of our inference problem can be further clarified by considering the censoring-free semi-Markov process for the Fix & Neyman (1951) model shown in the right panel of Fig. 1. The goal is to determine confidence bands for the survival and hazard functions associated with first passage from 1 → 3 in this uncensored flowgraph. However, data are not observed from this flowgraph but rather from the censored flowgraph in the left panel. It will be seen that the removal of censoring states from the left panel corresponds to the removal of censoring risk by using Kaplan–Meier estimates, which in turn gives estimates for the dynamical parameters of the semi-Markov process shown in the right panel of Fig. 1.

2. Simulation solution using nested resampling

An estimate of the survival function can be simulated by using a scheme which we refer to as walking through the network. Suppose complete data are available to describe all sojourn details for previous patients. Independence assumptions associated with the semi-Markov process, along with independence among patients, allow such sojourn data to be partitioned in terms of the state exited. For each exit state the data are summarized in terms of independent pairs consisting of holding times and associated destination states.

If there were no censoring, then walking through the network would consist of entering state 1, randomly sampling an exit pair associated with state 1, passing to the chosen destination state, repeating the same random sampling for the next destination state, and so on. The accumulated holding times up to absorption in state 3 are X_1^*, a simulated sojourn time. Repeating this $M = 10^6$ times leads to the empirical distribution of $\{X_k^* : k = 1, \ldots, M\}$ which suffices as a survival estimate.

If censored holding times may be sampled upon exit from a state, then it is necessary to modify the simulation by using the redistribute-to-the-right method introduced in Efron (1967) and described in Miller (1981). With this algorithm, a random exit time is selected and used if it is a non-censored time; otherwise, the selected censored time is not used and those holding times exceeding it are randomly sampled in order to draw a larger non-censored time. This process is repeated until either a larger non-censored exit time is selected or the very largest exit time is selected. In the latter case, if the largest exit time is censored, the exit simulation from the current state is rejected and repeated in its entirety. Such a scheme can again be used to compute an empirical distribution of M simulated sojourn times as a survival estimate.

Bootstrap resampling of empirical survival estimates can be implemented by using the following nested resampling scheme. The partition of the data set, determined according to exit state, allows resampled data sets to be determined by independently resampling the (destination, exit time) pairs from each state. If there is a total of $n_{i\cdot}$ non-censored and censored exits from state i, then, subject to some restrictions described in §6, a random sample with replacement of $n_{i\cdot}$ such data pairs creates the resampled data for exiting from each state i in resampled network data set $\mathcal{D}_1^*$. This resampling is repeated independently B times to generate network data sets $\mathcal{D}_1^*, \ldots, \mathcal{D}_B^*$ and $M = 10^6$ walks are resampled through each of these B networks. Walks in network k give sojourn times $\{X_{kl}^{**} : l = 1, \ldots, M\}$ that determine survival estimate $\hat{S}_k^*(t)$ and the ensemble $\{\hat{S}_k^*(t) : k = 1, \ldots, B\}$ leads to bootstrap pointwise confidence bands.

The demands for implementation of these nested resampling methods are the same for the double bootstrap. Indeed the authors have previously described this method as a double bootstrap, but it may also be thought of as a single bootstrap that requires simulation to determine the point estimate $\hat{S}_k^*(t)$ in the bootstrap resampling. As noted in Booth & Presnell (1998), the resampling demands for the double bootstrap, which extend to our proposed nested resampling scheme, generally place its implementation beyond the range of practical computing. While it is possible to perform such computations with small systems having sojourns that require few state transitions, multistate models with a large number of highly interconnected states, complex feedback patterns, or prolonged transient behaviour make such simulation not practically feasible.

Saddlepoint methods will be used in lieu of the inner resampling of $\{X_{kl}^{**} : l = 1, \ldots, M\}$ to determine $\hat{S}_k^*(t)$. The methods were first introduced in the double bootstrap context by Hinkley & Shi (1989). Unlike simulation, the saddlepoint procedures can deal with large systems, complex feedback patterns and prolonged transient behaviour. The resampling at the inner layer to determine MB values for $\{X_{kl}^{**}\}$ is replaced by B analytical saddlepoint inversions.

3. Semi-Markov process models

3·1. *Cumulative incidence functions and transmittances*

Semi-Markov processes have traditionally been used to describe the transitions of a subject through the various states $\mathcal{S} = \{1, \ldots, m\}$ of a multistate survival model. The process can be characterized through a sequence of independent competing risks that describe the exit times and destination-state choices in $\mathcal{S}$ for the process. In the simplest setting, a subject enters state $j_0 = 1$ at time 0, and moves from state to state, remaining in each state for a random period. Upon entering state $j_0 = 1$, exit from state 1 is a competing risk situation for which the m-dimensional exit distribution $\mathcal{H}_1$ determines the holding time and the next state. If random vector $\{H_{11}, \ldots, H_{1m}\}$ has distribution $\mathcal{H}_1$, then $j_1 = \arg\min_{j \in \mathcal{S}} H_{1j}$ is the next state and the holding time in state 1 has the conditional distribution of H_{1j_1} given the value of $j_1 = \arg\min_{j \in \mathcal{S}} H_{1j}$. Upon entering state j_1 at time $H_{1j_1} = t_1$, exit from state j_1 becomes another competing risk with multivariate exit distribution $\mathcal{H}_{j_1}$ that depends on the state j_1 and which is otherwise independent of the past. The sojourn of the subject through $\mathcal{S}$ continues in this way, with the sojourn itself interpretable as a sequence of independent competing risk exits from the visited states starting from a known entry state. The observed data for the sojourn would consist of competing risk type data of the form $\{1, (j_1, t_1), (j_2, t_2), \ldots\}$.

The process dynamics are completely specified by the collection of m-dimensional exit distributions $\{\mathcal{H}_i : i \in \mathcal{S}\}$. Actually, each multivariate exit distribution $\mathcal{H}_i$ overspecifies the process of exiting from state i, since only the so-called cumulative incidence functions, or destination-specific subdistributions, associated with $\mathcal{H}_i$ are needed to specify movement in $\mathcal{S}$; see Klein & Moeschberger (2003, p. 52). If $\{H_{i1}, \ldots, H_{im}\}$ has distribution $\mathcal{H}_i$, define the cumulative incident function for destination j when leaving state i as

$$G_{ij}(t) = \mathrm{pr}\left(H_{ij} = \min\nolimits_{k \in \mathcal{S}} H_{ik} \le t\right) = p_{ij} F_{ij}(t),$$

where $p_{ij} = \mathrm{pr}\left(j = \arg\min_k H_{ik}\right)$ and $F_{ij}(t)$ is the conditional distribution of H_{ij} given $j = \arg\min_k H_{ik}$. Here $\{p_{ij} : j \in \mathcal{S}\}$ are exit probabilities out of state i and $F_{ij}(t)$ is the holding time distribution in state i if destination j is assured. The $m \times m$ matrix of cumulative incidence functions $G(t) = \{G_{ij}(t) : i, j \in \mathcal{S}\}$ characterizes the semi-Markov process and is called the semi-Markov kernel; see Medhi (1994, §7·2).

The Laplace–Stieltjes transforms for components of $G(t)$, i.e.,

$$T_{ij}(s) = \int_0^\infty e^{st} dG_{ij}(t) = p_{ij} \int_0^\infty e^{st} dF_{ij}(t) = p_{ij} M_{ij}(s),$$

also characterize the process dynamics and prove to be most useful for computing properties of the process through saddlepoint approximation. The function $T_{ij}(s)$ is called a transmittance and the $m \times m$ matrix

$$T\left(s\right) = \left\{T_{ij}\left(s\right)\right\} = \left\{p_{ij} M_{ij}\left(s\right)\right\}$$

is the transmittance matrix of the semi-Markov process. It is the Hadamard product $P \odot M\left(s\right)$, where $P = (p_{ij}) = T(0)$ is the transition probability matrix for state changes, and $M(s) = \{M_{ij}(s)\}$ is a matrix of 1-step moment generating functions. The transmittance matrix $T(s)$ characterizes the dynamic behaviour of the system in the following manner: upon entering state i, the next state of the system is randomly determined by the multinomial distribution given in the ith row of P. If this is state j, then the holding time in state i, before proceeding to j, has distribution with moment generating function $M_{ij}(s)$.

3·2. First-passage transmittances

Suppose that X is the first-passage time in the semi-Markov process of a sojourn that enters state 1 at time 0 and ends upon entering state m. State m may be an absorbing state, a transient state or one member of a irreducible subchain of the process. The first-passage transmittance is accordingly

$$f_{1m}\mathcal{F}_{1m}(s) = E\left\{e^{sX}1_{(X<\infty)}\right\},$$

where $f_{1m} = \text{pr}(X < \infty)$ is the probability of passage and $\mathcal{F}_{1m}(s)$ is the conditional moment generating function of X given $X < \infty$. If there is at least one absorbing state in $\mathcal{S}$ other than state m, then $f_{1m} < 1$ and the distribution of X is defective with $\text{pr}(X = \infty) = 1 - f_{1m}$.

The following cofactor rule for computing $f_{1m}\mathcal{F}_{1m}(s)$ was proven in Butler (2000) and discussed further in Butler (2007, §13·2·1). Let $\mathcal{R} \subseteq \mathcal{S}$ with $\text{card}\,\mathcal{R} = m_{\mathcal{R}}$ be the relevant states to the sojourn, defined as the set of all possible intermediate states during it. Let $\mathcal{R} = \{1, i_2, \ldots i_{m_{\mathcal{R}}-1}, m\}$ be ordered so that states 1 and m are the first and $m_{\mathcal{R}}$th elements and let $\mathcal{I} = \mathcal{S}\backslash\mathcal{R}$ be all irrelevant states, with $\text{card}\,\mathcal{I} = m - m_{\mathcal{R}}$. Let $T_{\mathcal{R}\mathcal{R}}(s)$ and $T_{\mathcal{I}\mathcal{I}}(s)$ denote the $m_{\mathcal{R}} \times m_{\mathcal{R}}$ and $(m - m_{\mathcal{R}}) \times (m - m_{\mathcal{R}})$ principal submatrices of $T(s)$ corresponding to the ordered states in $\mathcal{R}$ and $\mathcal{I}$.

Theorem 1. *The first-passage transmittance from state* 1 *to* $m \neq 1$ *is*

$$f_{1m}\mathcal{F}_{1m}(s) = \frac{(m_{\mathcal{R}}, 1)\text{-cofactor of } I_{m_{\mathcal{R}}} - T_{\mathcal{R}\mathcal{R}}(s)}{(m_{\mathcal{R}}, m_{\mathcal{R}})\text{-cofactor of } I_{m_{\mathcal{R}}} - T_{\mathcal{R}\mathcal{R}}(s)} = \frac{(-1)^{m_{\mathcal{R}}+1}\left|\Psi_{m_{\mathcal{R}}1}(s)\right|}{\left|\Psi_{m_{\mathcal{R}}m_{\mathcal{R}}}(s)\right|}, \qquad (1)$$

where $\Psi_{ij}(s)$ is the (i,j)th minor of $I_{m_{\mathcal{R}}} - T_{\mathcal{R}\mathcal{R}}(s)$. If all components of $T_{\mathcal{R}\mathcal{R}}(s)$ are analytic over $(-\infty, \varepsilon)$ for some $\varepsilon > 0$, then the ratio (1) is analytic over a maximal convergence neighbourhood of 0 of the form $(-\infty, c)$, for some $c > 0$. Denote the cofactor ratio in (1) by $f_{1m}\mathcal{F}_{1m}(s)$. If the cofactor rule were instead used with $I_m - T(s)$, so all irrelevant states are included, then

$$\frac{(m,1)\text{-cofactor of } I_m - T(s)}{(m,m)\text{-cofactor of } I_m - T(s)} = f_{1m}\mathcal{F}_{1m}(s) \times \frac{|I_{m-m_{\mathcal{R}}} - T_{\mathcal{I}\mathcal{I}}(s)|}{|I_{m-m_{\mathcal{R}}} - T_{\mathcal{I}\mathcal{I}}(s)|} \qquad (s \neq 0),$$

and a removable discontinuity can occur at $s = 0$ when $|I_{m-m_{\mathcal{R}}} - T_{\mathcal{I}\mathcal{I}}(0)| = 0$.

When the source and destination states are both state 1, or if the destination is an arbitrary subset of states $\mathcal{D} \subset \mathcal{S}$, then equally simple cofactor rules $f_{11}\mathcal{F}_{11}(s)$ and $f_{1\mathcal{D}}\mathcal{F}_{1\mathcal{D}}(s)$ are given in §13·2·6 and §13·3 of Butler (2007). The $T(s)$ matrix in those expressions is the $T_{\mathcal{R}\mathcal{R}}(s)$ matrix used here, since Butler (2007, Ch. 13) assumes that $\mathcal{S}$ has already been restricted to states relevant to the sojourn.

With all three destination types $l = m, 1$, or $\mathcal{D}$, if $f_{1l} < 1$, then absorbing states other than $m, 1$, or $\mathcal{D}$ exist in $\mathcal{S}$. These absorbing states, and perhaps some other transient states, are irrelevant states for the first passage and not part of the computation in (1). In such instances, the defective survival function for first-passage is $f_{1l}S_{1l}(t) + 1 - f_{1l}$, where $S_{1l}(t)$ is the non-defective survival for $\mathcal{F}_{1l}(s)$, and mass $1 - f_{1l}$ is placed at ∞.

Computation of survival time distributions in multistate survival models entails saddlepoint inversion of a first-passage moment generating function such as $\mathcal{F}_{1m}(s)$ in (1). Such approximate inversions are indicated throughout using tilded overscores to distinguish them from exact expressions. Expressions for computing saddlepoint density $\tilde{d}_{1m}(t)$ and survival $\tilde{S}_{1m}(t)$ approximations that can be used in conjunction with a cumulant generating function $\log \mathcal{F}_{1m}(s)$ are provided in Butler (2007, §§1·1·2, 1·2·1). A saddlepoint hazard rate approximation is $\tilde{z}_{1m}(t) = \tilde{d}_{1m}(t)/\tilde{S}_{1m}(t)$. These saddlepoint computations require the first two derivatives of $\log \mathcal{F}_{1m}(s)$, where $\mathcal{F}_{1m}$ is a ratio of cofactors as in (1). Such derivatives take on especially simple forms and are given in Butler (2007, ch. 13).

4. Point estimation of a survival function

4·1. Saddlepoint point estimation

The survival function for first-passage time X from $1 \to m$ is denoted through the unknown functional $S(t;\theta) = \text{pr}(X > t;\theta)$, where the system parameter $\theta = T(s)$ is the $m \times m$ transmittance matrix that characterizes the semi-Markov process. Even if an estimator $\hat{\theta} = \hat{T}(s)$ is available, $S(t;\hat{\theta})$ is intractable since S is unknown. However, an estimator can be determined by plugging $\hat{\theta}$ into the saddlepoint approximation $\tilde{S}(t;\theta)$ for $S(t;\theta)$ so that $\tilde{S}(t;\hat{\theta})$ is an estimator for $S(t;\theta)$; its use presumes there is a practically negligible difference between $\tilde{S}(t;\theta)$ and $S(t;\theta)$. This is supported by the many examples in Butler (2000, 2007, Ch. 13).

4·2. Estimation without censoring

In the absence of censoring, the computation of $\hat{\theta}$ from sojourn data through the network has been discussed in Butler & Bronson (2002). Pooling sojourn data for all N subjects, suppose there are n_{ij} transitions from $i \to j$ with holding times $x_{ij1}, ..., x_{ijn_{ij}}$. Then

$$\hat{T}_{ij}(s) = \frac{n_{ij}}{n_{i\cdot}} \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} \exp(s x_{ijk}) = \hat{p}_{ij} \hat{M}_{ij}(s),$$

with $n_{i\cdot} = \sum_{j \in \mathcal{S}} n_{ij}$, so that $\hat{\theta} = \hat{T}(s) = \{\hat{T}_{ij}(s)\}$. From the nonparametric likelihood of the sojourn data, it is clear that $\hat{\theta} = \hat{T}(s)$ is a sufficient statistic for θ but also the nonparametric maximum likelihood estimator for θ when data are uncensored.

The estimator $\hat{\theta} = \hat{T}(s)$ itself indexes a semi-Markov process whose exit-time distributions are discrete and whose behaviour provides estimates for the behaviour of the semi-Markov process indexed by θ. Since saddlepoint inversions are smooth functionals of $\theta = T(s)$, the suggested point estimator $\tilde{S}(t;\hat{\theta})$ is a nonparametric maximum likelihood estimator for $\tilde{S}(t;\theta)$.

4·3. Estimation with censoring

The estimation procedure just described can be modified to accommodate right censoring if each patient's censoring time satisfies the two conditions mentioned in §1: First, the censoring time is independent of all aspects of the patient's sojourn through the semi-Markov process. Secondly, the censoring mechanism is tail-estimable for exit-time distributions. The latter condition holds if all exit-time distributions $\{F_{ij}(t)\}$ have intervals of support that are subsets of $(0,\tau)$, where $\tau \leq \infty$ is the least upper bound for the support of the censoring time distribution $C(t)$.

Consider, for example, a subject who enters state 1 of the semi-Markov process in Fig. 1 at time 0, proceeds to state 2 in time t_1, returns to state 1 after holding for t_2, and is finally censored while holding in state 1 for time t_3^+. The data are $\{1,(2,t_1),(1,t_2),(\mathrm{R}_1,t_3^+)\}$ and the nonparametric likelihood for this sojourn is

$$p_{12}dF_{12}(t_1)p_{21}dF_{21}(t_2)\left\{1-\textstyle\sum_{k=1}^{3}p_{1k}F_{1k}(t_3^+)\right\}\times dC(t_1+t_2+t_3^+). \tag{2}$$

The portion of the likelihood for $T(s)$ separates from that for $C(t)$. When these data are used to label transmittances in the left panel of Fig. 1, about which they are informative as shown in Fig. 2, then t_1 is assigned to transmittance $1\to 2$, t_2 to $2\to 1$, and t_3^+ to $1\to R_1$. Once all holding times for all patients have been assigned, then exit data from each of the transient states are those of a classical competing risk data set with independent right censoring. Fig. 2 is an example of exit data from state 1.

Estimating transmittances $\{T_{ij}(s)\}$ for the semi-Markov process in the right panel of Fig. 1, that are free from censoring risk, begins by first estimating the associated cumulative incident functions $\{G_{ij}(t)\}$ using standard nonparametric maximum likelihood estimators as in Klein & Moeschberger (2003, eqn. 4.7.1). This is followed by the computation of their Laplace–Stieltjes transforms and a rescaling. In Fig. 2, for example, cumulative incident function estimators $\{\hat{G}_{1j}(t):j=1,2,3\}$ are computed by first estimating the overall exit distribution function $G_{1\cdot}(t)=\sum_{j=1}^{3}G_{1j}(t)$ as the ordinary Kaplan–Meier estimator $\hat{G}_{1\cdot}(t)$ that treats the uncensored exit times $\{x_1,\ldots,x_{13},y_1,\ldots,y_4,z_1,\ldots,z_4\}$ as event times and $\{w_1^+,\ldots,w_4^+\}$ as censored times. In the special case with no ties, the mass points assigned to each destination-specific cumulative incidence function are the Kaplan–Meier masses at the holding times for that destination; i.e., $d\hat{G}_{12}(x_i) = d\hat{G}_{1\cdot}(x_i)$ for $i = 1, \ldots, 13$, $d\hat{G}_{13}(y_j) = d\hat{G}_{1\cdot}(y_j)$, and $d\hat{G}_{11}(z_k) = d\hat{G}_{1\cdot}(z_k)$.

Such estimators are also Aalen–Johansen estimators (Anderson et al., 1993, Example IV·4·1). They are applicable for estimating the semi-Markov kernel $G(t)$ because isolated single-step exits from the states of such a process represent time-heterogeneous Markov processes; the Markov property breaks down when multiple steps are allowed.

Let $\hat{\tau}_1$ be the largest holding time in state 1. The Kaplan–Meier total at $\hat{\tau}_1$ is $\hat{G}_{1\cdot}(\hat{\tau}_1) = \sum_{k=1}^{3} \hat{G}_{1k}(\hat{\tau}_1) = 1$ if $\hat{\tau}_1$ is not a censored value. If, however, $\hat{\tau}_1$ is a censoring time, $\hat{G}_{1\cdot}(\hat{\tau}_1) < 1$ and the unallocated probability above $\hat{\tau}_1$ must be reallocated to the three mixture components when estimating the kernel $G(t)$ and transmittance matrix $T(s)$ so as to correctly reflect the known transience of state 1. The data provide no guidance about such reallocation, as the nonparametric likelihood is uninformative. However, when censoring is tail-estimable, any reallocation can be used and will be asymptotically correct as indicated in Theorem 3. Our discussion only considers the approach of Dinse & Larson (1986, p. 381) that reallocates this probability proportionately. This leads to transmittance estimators

$$\hat{T}_{1j}(s) = \{\hat{G}_{1\cdot}(\hat{\tau}_1)\}^{-1} \int_0^\infty \exp(st) d\hat{G}_{1j}(t) \qquad (j = 1, 2, 3), \tag{3}$$

and ensures that state 1 is transient in $\hat{T}(s)$. This rescaling is also equivalent to using the redistribute-to-the-right algorithm described in §2 wherein the exit simulation is redone in its entirety if the algorithm stops at $\hat{\tau}_1$ and $\hat{\tau}_1$ is a censoring time.

Justification for the rescaling in (3) is motivated as an attempt to accurately estimate transition probabilities $\hat{P} = \hat{T}(0)$. This occurs with rescaling if destination probabilities achieved before time $\hat{\tau}_1$ are roughly proportional to those after time $\hat{\tau}_1$; i.e., the unknown vectors $\{G_{1j}(\hat{\tau}_1)\}$ and $\{p_{1j} - G_{1j}(\hat{\tau}_1)\}$ are roughly proportional. Any such presumption, however, is untestable because the data lack information about exit times and destination probabilities above $\hat{\tau}_1$. Misjudgment in this reallocation will bias the estimators $\{\hat{T}_{1j}(s)\}$ with the size of the bias proportional to the unapportioned mass

$1-\hat{G}_{1\cdot}(\hat{\tau}_1)$. Fortunately, when censoring is tail-estimable, this unapportioned mass is asymptotically negligible, as noted in Theorem 3, so any such bias is also asymptotically negligible. In practical applications, however, these unapportioned mass sizes should be examined to determine their potential for bias when using $\hat{T}(s)$ as an estimator of $T(s)$, as this affects the accuracy of the overall method.

Such rescaling is unnecessary when a parametric model is used for $G(t)$, as in Lô et al. (2008). Such an approach, however, may introduce serious bias when models are misspecified.

The rescaled transmittance estimator $\hat{T}(s)$ indexes a semi-Markov process on $\mathcal{S}$. For this process, let $\hat{\mathcal{R}}$, perhaps different from $\mathcal{R}$, denote those states that are relevant to sojourn $1 \to m$. If X^* is a sojourn time for a walk through the network using the redistribute-to-the-right algorithm of §2, then its transmittance is the cofactor rule of Theorem 1 applied to $\hat{T}_{\hat{\mathcal{R}}\hat{\mathcal{R}}}(s)$, the relevant principal submatrix of $\hat{T}(s)$.

Theorem 2. *If X^* is a simulated sojourn time from $1 \to m$ as described in §2, then*

$$E\left\{\exp(sX^*)1_{(X^*<\infty)}\right\} = \hat{f}_{1m}\,\hat{\mathcal{F}}_{1m}(s) = \frac{(m_{\hat{\mathcal{R}}},1)\text{-cofactor of } I_{m_{\hat{\mathcal{R}}}} - \hat{T}_{\hat{\mathcal{R}}\hat{\mathcal{R}}}(s)}{(m_{\hat{\mathcal{R}}},m_{\hat{\mathcal{R}}})\text{-cofactor of } I_{m_{\hat{\mathcal{R}}}} - \hat{T}_{\hat{\mathcal{R}}\hat{\mathcal{R}}}(s)},$$

where $m_{\hat{\mathcal{R}}} = \operatorname{card}\hat{R}$ and the expectation is conditional upon the data. If $\hat{S}_M(t)$ is an empirical survival function for M independently simulated copies of X^*, then $\hat{S}_M(t)$ converges weakly to $S(t;\hat{\theta}) = \hat{f}_{1m}S_{1m}(t;\hat{\theta}) + 1 - \hat{f}_{1m}$ as $M \to \infty$, where $S_{1m}(t;\hat{\theta}) =$ $\mathrm{pr}(X^* > t \mid X^* < \infty)$. The saddlepoint approximation for $S(t;\hat{\theta})$ is

$$\tilde{S}(t;\hat{\theta}) = \hat{f}_{1m}\tilde{S}_{1m}(t;\hat{\theta}) + 1 - \hat{f}_{1m}, \tag{4}$$

where $\tilde{S}_{1m}(t;\hat{\theta})$ is the survival function obtained through saddlepoint inversion of $\hat{\mathcal{F}}_{1m}(s)$ using the continuous Lugannani–Rice approximation (Butler, 2007, eqn. 1.21).

For settings where $\hat{f}_{1m} < 1$, there is at least one absorbing state in $\mathcal{S}\backslash\hat{\mathcal{R}}$ that is irrelevant to first passage and makes $\mathrm{pr}(X^* = \infty) = 1 - \hat{f}_{1m} > 0$.

Proof of Theorem 2. In the simulation of a sojourn, each leg exiting a state is a random generation from the appropriate row of the rescaled transmittance matrix $\hat{T}(s)$.

Thus saddlepoint inversion of $\hat{\mathcal{F}}_{1m}(s)$, determined from $\hat{T}(s)$, leads to a saddlepoint approximation for $S_{1m}(t;\hat{\theta})$, the survival function for $\hat{\mathcal{F}}_{1m}(s)$. The weak convergence of $\hat{S}_M(t)$ to $S(t;\hat{\theta})$ follows from the law of large numbers. □

The theorem indicates that analytical computation of the saddlepoint estimator $\tilde{S}(t;\hat{\theta})$ can replace the simulation estimator $\hat{S}_M(t)$ as a substitute for the intractable plug-in estimator $S(t;\hat{\theta})$. A justification for using $\tilde{S}(t;\hat{\theta})$ as an estimator of $S(t;\theta)$ follows from the next result, which shows $\tilde{S}(t;\hat{\theta})$ is a uniformly consistent estimator of $\tilde{S}(t;\theta)$ as the number of patients $N \to \infty$. The proof is given in the Supplementary Material and makes use of Theorem 1 of Suzukawa (2002) with the following assumptions:

Assumptions. (i) Right censoring is random and independent of the semi-Markov dynamics. (ii) The censoring distribution $C(t)$ and all exit-time distributions $\{F_{ij}(t)\}$ are continuous. (iii) All exit-time distributions $\{F_{ij}(t)\}$ have moment generating functions that are convergent in a neighbourhood of zero. (iv) The censoring is tail-estimable, i.e. all exit-time distributions $\{F_{ij}(t)\}$ have intervals of support that are subsets of $(0,\tau)$, where $\tau \leq \infty$ is the least upper bound to the support of $C(t)$.

Theorem 3. *Subject to assumptions (i)–(iv) above, the saddlepoint survival in (4), density estimator $\tilde{d}_{1m}(t;\hat{\theta})$, and hazard estimator $\tilde{z}_{1m}(t;\hat{\theta}) = \tilde{d}_{1m}(t;\hat{\theta})/\tilde{S}_{1m}(t;\hat{\theta})$ converge uniformly in probability to population counterparts $\tilde{S}(t;\theta)$, $\tilde{d}_{1m}(t;\theta)$, and $\tilde{z}_{1m}(t;\theta)$ as $N \to \infty$ over compact subsets of t. Furthermore, if $\hat{\tau}_i$ is the largest holding time in state i, censored or otherwise, then*

$$\hat{G}_{i\cdot}(\hat{\tau}_i) = \int_0^{\hat{\tau}_i} d\hat{G}_{i\cdot}(t) \to \int_0^{\infty} dG_{i\cdot}(t) = 1, \qquad (N \to \infty).$$

This makes the rescaling used in $\hat{T}(s)$ asymptotically correct and $\hat{T}(s)$ converges uniformly in probability to $T(s)$ over compact subsets in the convergence region of $T(s)$.

Without tail-estimable censoring, the consistency of saddlepoint estimators is restricted to $0 < t < \tau$.

Theorem 4. *Under assumptions (i)–(iii), saddlepoint survival, density, and hazard estimators converge uniformly in probability to their population counterparts as $N \to \infty$*

over compact subsets of t in $(0,\tau)$. The kernel $G(t)$ is consistently estimable for $t \leq \tau$, but not for $t > \tau$, and $T(s)$ is not estimable for any s.

The proof and details about the saddlepoint estimators are given in the Supplementary Material. Implications of Type I censoring for consistency are also given.

5. Simulated example

A numerical example is considered for our generalization of the Fix–Neyman model. The model has been specified to satisfy the four assumptions preceding Theorem 3 so that tail-estimability is ensured. The data consist of 100 simulated sojourns through the system in the left panel of Fig. 1 starting in state 1 at time 0. These sojourns were generated by simulating sojourns through the censor-free system in the right panel of Fig. 1 and independently generating a censoring time from $C(t)$ that competes with the sojourn to determine the exit state in the left panel. The censor-free system had transmittance matrix

$$T(s) = \begin{pmatrix} 0{\cdot}3\,\mathrm{ig}(10{\cdot}5,11{\cdot}7) & 0{\cdot}3\,\mathrm{r}(17{\cdot}7) & 0{\cdot}4\,\mathrm{r}(22{\cdot}2) \\ 0{\cdot}5\,\mathrm{r}(13{\cdot}3) & 0{\cdot}5\,\mathrm{ig}(11{\cdot}0,8{\cdot}8) & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad (5)$$

where $\mathrm{ig}(a,b)$ is the inverse Gaussian moment generating function with mean a and standard deviation b. The moment generating function $\mathrm{r}(a)$ is for a Rayleigh distribution with mean a and can be easily derived in terms of the $\mathrm{erf}(\cdot)$ error function. Starting in state 1, the transmittance (5) admits a mean sojourn time through the uncensored system of Fig. 1 as $\mathcal{K}'(0) = 61{\cdot}575$ where $\mathcal{K}(s) = \log \mathcal{F}_{13}(s)$. The standard deviation of the sojourn time is $\mathcal{K}''(0)^{1/2} = 55{\cdot}3$ and the standardized third cumulant $\mathcal{K}'''(0)/\mathcal{K}''(0)^{3/2} = 2{\cdot}59$ suggests that the sojourn distribution is highly skewed.

The independent censoring distribution $C(t)$ is gamma with mean $62{\cdot}5$ and standard deviation $28{\cdot}0$, so the mean is roughly the same as $\mathcal{K}'(0)$ and the standard deviation is about half that of the sojourn. The standardized third cumulant is $0{\cdot}169$, suggesting an approximate normal shape. With the means roughly equal, there should be substantial censoring due to the severe skewness of the sojourn distribution. For this example, 40 of the 100 simulated sojourns were censored.

Kaplan–Meier estimates were computed for cumulative incidence functions in (5) that are associated with the censor-free system in Fig. 1. Rescaling in the estimation of $\hat{T}(s)$ was not necessary for these estimates since the largest holding times $\hat{\tau}_1 = 51{\cdot}4$ and $\hat{\tau}_2 = 28{\cdot}3$ were not terminated by censoring. Laplace–Stieltjes transforms of these cumulative incidence function estimates yield an empirical transmittance $\hat{\theta} = \hat{T}(s)$ with $\hat{\mathcal{R}} = \mathcal{R} = \{1, 2, 3\}$ that can be used in conjunction with the cofactor rule of Theorem 1 to determine the empirical first-passage moment generating function

$$\hat{\mathcal{F}}_{13}\,(s) = \frac{\{1 - \hat{T}_{22}(s)\}\hat{T}_{13}(s)}{\{1 - \hat{T}_{11}(s)\}\{1 - \hat{T}_{22}(s)\} - \hat{T}_{12}(s)\hat{T}_{21}(s)}$$

and passage probability $\hat{f}_{13} = 1$. Saddlepoint inversion of $\hat{\mathcal{F}}_{13}\,(s)$ leads to $\tilde{S}(t; \hat{\theta}) = \tilde{S}_{13}(t; \hat{\theta})$, as plotted in Fig. 3, which estimates $\tilde{S}(t; \theta)$, whose plot is also shown by inverting $\mathcal{F}_{13}\,(s)$. Bootstrap confidence bands using the BC$_a$ method are also shown and will be discussed in §6.

An estimate of the survival density $\tilde{d}_{13}(t; \hat{\theta})$, using saddlepoint inversion of $\hat{\mathcal{F}}_{13}\,(s)$, and the true saddlepoint density $\tilde{d}_{13}\,(t; \theta)$ are plotted in Fig. 4.

The hazard function estimate $\tilde{z}_{13}(t; \hat{\theta})$ is plotted in Fig. 3 along with $\tilde{z}_{13}(t; \theta)$, the true saddlepoint hazard function. BC$_a$ confidence bands are also shown. The stabilization of the hazard rate plot as $t \to \infty$ was also noted in Butler & Bronson (2002). In unpublished research, this limit has been characterized as the right-edge of the convergence strip of its associated moment generating function. This occurs at $c = 0{\cdot}0167$, the smallest real positive singularity of $\mathcal{F}_{13}\,(s)$. The comparable singularity for $\hat{\mathcal{F}}_{13}\,(s)$ is $\hat{c} = 0{\cdot}0163$. Such accuracy in estimating a tail characteristic results from using the entire structure of the semi-Markov process when computing $\hat{\mathcal{F}}_{13}\,(s)$ from $\hat{T}(s)$. This stabilization of hazard provides an important new observation about sojourns in the general theory of semi-Markov processes. This observation, confirmed more rigorously in unpublished research, is that passage time distributions in semi-Markov processes have exponential tails with rate parameter c when $c > 0$. Thus, they behave like passage times of Markov processes.

6. Bootstrapping in the transform domain

The BC$_a$ confidence bands for survival and hazard functions shown in Fig. 3 are the result of implementing the bootstrap but using saddlepoint approximations to compute each of the resampled estimates. Bootstrap resampling provides an ensemble of survival and hazard function estimates which lead to pointwise confidence bands to accompany the point estimates. The idea is to implement resampling to construct B resampled values of the empirical system parameter denoted by $\{\hat{\theta}_k^* = \hat{T}_k^*(s) : k = 1, \ldots, B\}$. These B transmittances determine B first-passage transmittances $\{\hat{f}_k^* \hat{\mathcal{F}}_k^*(s) : k = 1, \ldots, B\}$ whose saddlepoint inversions form an ensemble of survival function estimates $\{\tilde{S}(t; \hat{\theta}_k^*) : k = 1, \ldots, B\}$. Pointwise confidence bands using the bootstrap percentile or BC$_a$ methods can be determined by using a sufficiently fine grid of time points that assures smooth-looking curves.

Each $\hat{\theta}_k^* = \hat{T}_k^*(s)$ is obtained by extending Efron's (1981) scheme to the various competing risks. For each k, a resample of transitions out of each state i proceeds by randomly sampling $n_{i\cdot}$ (destination, holding-time) pairs with replacement from the $n_{i\cdot}$ exiting data pairs from state i. One caveat in using these resampled exits is that all possible uncensored state transitions out of state i must be represented in the resample, otherwise the entire resample from state i should be drawn again. If such resamples were not discarded, they would alter the system structure by creating irrelevant states in the resampled system $\hat{\theta}_k^* = \hat{T}_k^*(s)$ that are relevant states in the original data system $\hat{\theta} = \hat{T}(s)$ and therefore relevant to the true underlying system $\theta = T(s)$. This rejection scheme ensures that $\hat{\mathcal{R}}_k^* \equiv \hat{\mathcal{R}} \subseteq \mathcal{R}$ for all k, where $\hat{\mathcal{R}}_k^*$ are the relevant states of $\hat{T}_k^*(s)$.

Theorem 5. *If X^{**} is a sojourn time through the resampled system $\hat{\theta}^* = \hat{T}^*(s)$, then its first-passage transmittance is*

$$E\left\{\exp(sX^{**}) 1_{(X^{**}<\infty)}\right\} = \hat{f}_{1m}^* \hat{\mathcal{F}}_{1m}^*(s) = \frac{(m_{\hat{\mathcal{R}}}, 1)\text{-cofactor of } I_{m_{\hat{\mathcal{R}}}} - \hat{T}_{\hat{\mathcal{R}}\hat{\mathcal{R}}}^*(s)}{(m_{\hat{\mathcal{R}}}, m_{\hat{\mathcal{R}}})\text{-cofactor of } I_{m_{\hat{\mathcal{R}}}} - \hat{T}_{\hat{\mathcal{R}}\hat{\mathcal{R}}}^*(s)}.$$

If $\hat{S}_M^{**}(t)$ is an empirical survival function for M independently simulated copies of X^{**}, then $\hat{S}_M^{**}(t)$ converges weakly to $S(t;\hat{\theta}^*) = \hat{f}_{1m}^* S_{1m}^*(t;\hat{\theta}) + 1 - \hat{f}_{1m}^*$ as $M \to \infty$, where $S_{1m}^*(t;\hat{\theta}) = \mathrm{pr}(X^{**} > t \mid X^{**} < \infty)$. A saddlepoint approximation for the intractable resampled survival function $S(t;\hat{\theta}^*)$ is $\tilde{S}(t;\hat{\theta}^*) = \hat{f}_{1m}^* \tilde{S}_{1m}^*(t;\hat{\theta}) + 1 - \hat{f}_{1m}^*$, where $\tilde{S}_{1m}^*(t;\hat{\theta})$ is the Lugannani–Rice saddlepoint inversion of $\hat{\mathcal{F}}_{1m}^*(s)$. If the data are such that $\hat{f}_{1m} = 1$, then $\hat{f}_{1m}^* = 1$.

Proof. The rejection scheme for resampling ensures that transition probabilities in $\hat{T}(0)$ for system $\hat{\theta}$ that are non-zero remain non-zero in $\hat{T}^*(0)$ for $\hat{\theta}^*$. This ensures the same pattern of communicating states so that if $\hat{f}_{1m} = 1$, then $\hat{f}_{1m}^* = 1$. □

The values of X^{**} are the MB resampled sojourns on the inner layer of resampling described in §2. Theorem 5 emphasizes that the inner layer of resampling is unnecessary for determining an ensemble of survival functions to compute bootstrap confidence bands. Moreover, each saddlepoint survival function $\tilde{S}(t;\hat{\theta}_k^*)$ is equivalent to an empirical survival function $\hat{S}_M^{**}(t)$ computed with a very large simulation size M.

In some sojourn settings that have rare transitions without much data, the requirement that $\hat{\mathcal{R}}_k^* \equiv \hat{\mathcal{R}}$ for each bootstrap sample k may need to be relaxed in order to obtain reasonable bootstrap coverage accuracy.

7. Simulated example revisited

Figure 3 shows 90% BC$_a$ confidence bands for the survival and hazard functions of sojourn time X based on an ensemble of $B = 1000$ resampled estimates of the appropriate functions.

To assess coverage accuracy of the method, 90% bootstrap confidence intervals were constructed for selected percentiles of the sojourn distribution displayed in Table 1. For example, 82·7 is the exact 75th saddlepoint percentile that solves $0{\cdot}25 = \tilde{S}(82{\cdot}7;\theta)$ and percentile estimate 89·2 solves $0{\cdot}25 = \tilde{S}(89{\cdot}2;\hat{\theta})$. A 90% BC$_a$ confidence interval for this percentile is (74·6, 110·5). This interval was constructed from resampled percentiles determined by solving $\{0{\cdot}25 = \tilde{S}(q_i^*;\hat{\theta}_i^*) : i = 1\ldots, B\}$ so that $\{q_i^* : i = 1\ldots, B\}$ are resampled 75th percentiles that determine the 90% interval $(74{\cdot}6, 110{\cdot}5)$.

Also listed are the right boundaries for guaranteed coverage tolerance intervals of X that provide coverage $1-$ Right Perc. with a 90% guarantee; see Aitchison & Dunsmore (1975). For example, the tolerance interval $(0, 107{\cdot}4)$ gives coverage for 75% of the smallest values of X with a 90% guarantee.

Coverage percentages for the 90% BC_a confidence intervals of the right-tail percentiles were estimated by computing BC_a intervals for the percentiles using 1000 additional simulated data sets. From the median up to the 99th percentile, all coverages are reported in Table 1 as 90% when rounded to the nearest percentage.

The same 1000 data sets were used to assess coverage versus time for 90% BC_a confidence bands of $\tilde{S}(t;\theta)$. The right panel of Fig. 4 plots the resulting percentage relative error versus time or $100\% \times \{\text{empirical coverage}(t) - 0{\cdot}9\}/0{\cdot}1$ versus t from the 50th to the 99·5 percentiles of the sojourn distribution. Note the very high degree of accuracy for all values of t shown. Perhaps even more striking is the relative accuracy that is maintained at and above $t = 300$ which, for the distribution of X, is 4·3 standard deviations above the mean sojourn time of 62·5. We conjecture that such high coverage accuracy is maintained well into the tail of the sojourn distribution because the bootstrap is working at the simple poles $\hat{c}^*$, $\hat{c}$, and c that define the edges of convergence strips for $\hat{\mathcal{F}}^*_{1m}(s)$, $\hat{\mathcal{F}}_{1m}(s)$, and $\mathcal{F}_{1m}(s)$. The structure of these poles is similar since they are defined by the simple zeros of the denominators $|I_{m_{\hat{\mathcal{R}}}} - \hat{T}^*_{\hat{\mathcal{R}}\hat{\mathcal{R}}}(s)|$, $|I_{m_{\hat{\mathcal{R}}}} - \hat{T}_{\hat{\mathcal{R}}\hat{\mathcal{R}}}(s)|$, and $|I_{m_{\mathcal{R}}} - T_{\mathcal{R}\mathcal{R}}(s)|$.

The rescaling in (3), while not needed in computing $\hat{T}(s)$, inevitably occurred when computing $\{\hat{T}^*_k(s) : k = 1, ..., B\}$. It also occurred in computing $\hat{T}(s)$ and $\{\hat{T}^*_k(s) : k = 1, ..., B\}$ when generating the 1000 data sets used to determine the coverage percentages in Table 1 and the error plot in Fig. 4. The accuracy of coverages displayed in the table and plot reflect the success of the rescaling method in (3).

8. Proportional hazard extensions

For data possessing time-dependent covariates $x(t)$, there are no conceptual difficulties to extending these methods if the additional structure of the Cox-model is assumed for the competing risk settings when exiting each state i. The associated destination-specific hazards are

$$\lambda_{ij}(t;x) = \lambda_{ij0}(t)e^{\beta_{ij}^T x(t)} \qquad (j = 1, \ldots, m),$$

and the methods for semiparametric estimation described in Kalbfleisch and Prentice (2002, §8.2.3) and Lawless (2003, §9.4) are applicable for plug-in estimation of destination-specific cumulative incidence functions. For confidence intervals however, nontrivial computational issues are likely to arise in bootstrap resampling of (destination-state, holding-time, time-indexed covariate) triples from individual states, due to sampling unbalanced designs.

Acknowledgments

We thank James Booth for clarifying distinctions between double and single bootstrap interpretations of this work. We also thank two referees, the editor, and an associate editor whose comments substantially improved the presentation. The work was supported by the U.S. National Science Foundation.

Supplementary material

Supplementary material available at *Biometrika* online includes proofs of Theorems 3 and 4, discussion of consistency with censoring that is not tail-estimable, and consideration of Type I censoring.

References

Aalen, O. O. & Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand. J. Statist.* **5**, 141–50.

Aitchison, J. & Dunsmore, I. R. (1975). *Statistical Prediction Analysis.* Cambridge: Cambridge University Press.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). *Statistical Models Based on Counting Processes.* New York: Springer.

ANDERSEN, P. K., HANSEN, L. S. & KEIDING, N. (1991). Non- and semi-parametric estimation of transition probabilities from censored observations of a non-homogeneous Markov process. *Scand. J. Statist.* **18**, 153–67.

BOOTH, J. G. & PRESNELL, B. (1998). Allocation of Monte Carlo resources for the iterated bootstrap. *J. Comp. Graph. Stat.* **7**, 92–112.

BUTLER, R. W. (2000). Reliabilities for feedback systems and their saddlepoint approximation. *Statist. Sci.* **15**, 279–98.

BUTLER, R. W. (2007). *Saddlepoint Approximations with Applications.* Cambridge: Cambridge University Press.

BUTLER, R. W. & BRONSON, D. A. (2002). Bootstrapping survival times in stochastic systems by using saddlepoint approximations. *J. R. Statist. Soc. B* **64**, 31–49.

BUTLER, R. W. & HUZURBAZAR, A. V. (1997). Stochastic network models for survival analysis. *J. Am. Statist. Assoc.* **92**, 246–57.

CHIANG, C. & HSU, J. (1976). On multiple transition time in a simple illness death process - a Fix–Neyman model. *Math. Biosci.* **30**, 55–71.

DAVISON, A. C. AND HINKLEY, D. V. (1997) *Bootstrap Methods and their Application.* Cambridge: Cambridge University Press.

DINSE, G. & LARSON, M. (1986). A note on semi-Markov models for partially censored data. *Biometrika* **73**, 379–86.

EFRON, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.* **4**, 831–53.

EFRON, B. (1981). Censored data and the bootstrap. *J. Am. Statist. Assoc.* **76**, 312–9.

FIX, E. & NEYMAN, J. (1951). A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology* **23**, 205–41.

HINKLEY, D. V. & SHI, S. (1989). Importance sampling and the nested bootstrap. *Biometrika* **76**, 435–46.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data.* 2nd ed. New York: Wiley.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data.* 2nd ed. New York: Wiley.

Lô, S. N., Heritier, S., & Hudson, M. (2008). Saddlepoint approximation for semi-Markov processes with application to a cardiovascular randomized study. *Comp. Statist. Data Anal.* **53**, 683–98.

Medhi, J. (1994). *Stochastic Processes.* 2nd ed. New York: Wiley.

Miller, R. (1981). *Survival Analysis.* New York: Wiley.

Suzukawa, A. (2002). Asymptotic properties of Aalen–Johansen integrals for competing risk data. *J. Japan Statist. Soc.* **1** 77–93.

Table 1. Estimates for the distribution of first-passage time X from $1 \rightarrow 3$ as shown in the right panel of Fig. 1. Percentiles of X in column 2, associated with the right-tail probabilities in column 1, have 90% BC$_a$ confidence intervals given in columns 4 and 5 when computed from a single data set. The right boundaries of 90% guaranteed coverage tolerance intervals are in column 6. Coverage percentages for the BC$_a$ confidence intervals resulting from 1000 simulated data sets and rounded to the nearest integer are in column 7.

Right Probs.	Exact	Estimate	BC$_a$ Lower	BC$_a$ Upper	Guar. Tol. I.	Coverage %age
0·50	42·0	47·7	41·0	58·4	54·9	90·
0·25	82·7	89·2	74·6	110·5	107·4	90·
0·10	137·7	145·4	119·9	179·7	173·6	90·
0·05	179·4	188·1	153·5	233·7	225·3	90·
0·01	276·3	287·4	231·9	358·6	346·1	90·

Probs., Tail Probabilities; Guar. Tol. I., Guaranteed Tolerance Interval; %age, percentage.

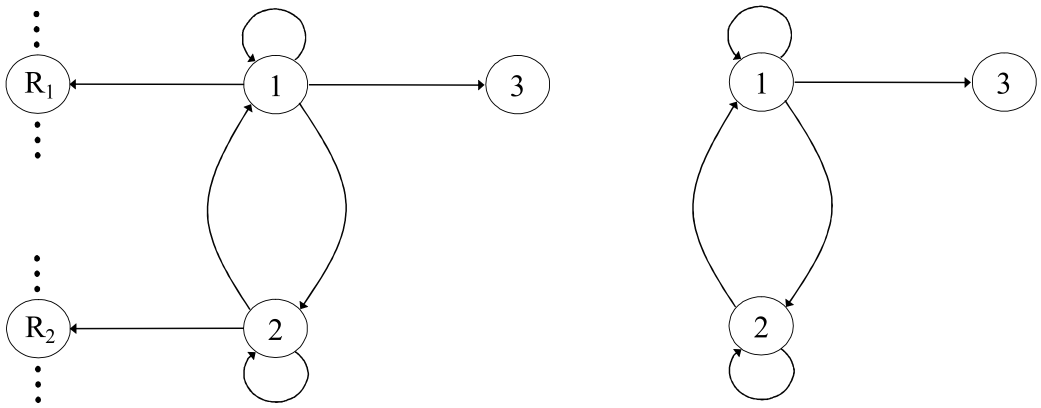

Fig. 1. (Left) The observed flowgraph for the generalized Fix & Neyman model allowing censoring from each transient state. (Right) The unobserved generalized Fix & Neyman flowgraph with censoring risk factors removed.

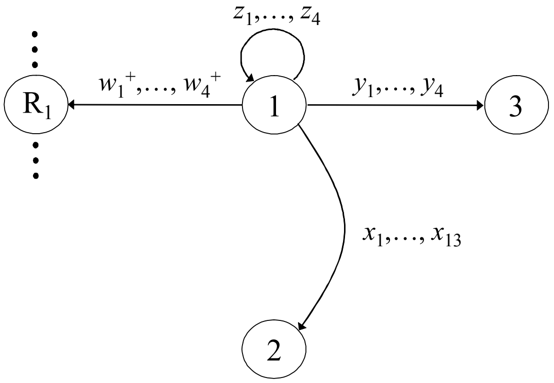

Fig. 2. Example of exit data from state 1 with censoring

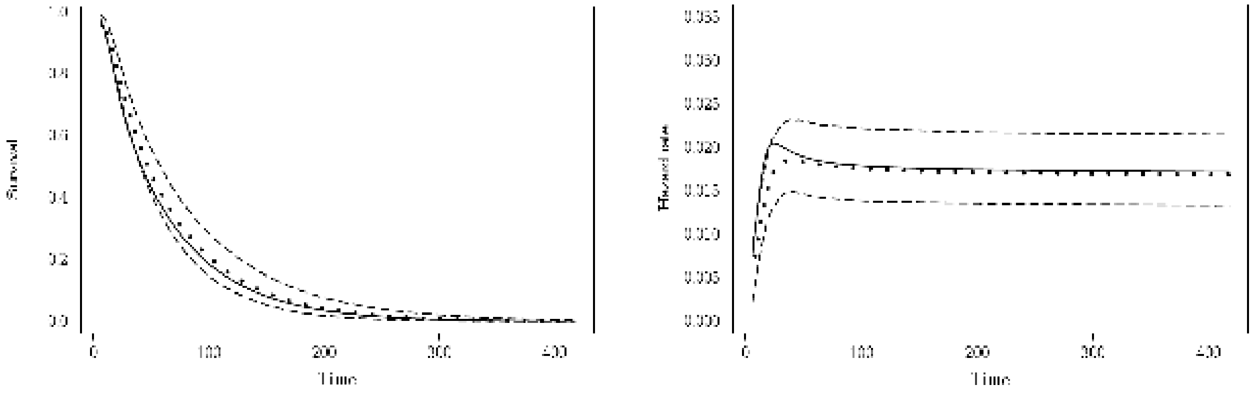

Fig. 3. (Left) Saddlepoint survival function $\tilde{S}(t;\theta) = \tilde{S}_{13}(t;\theta)$ (solid), its saddlepoint estimate $\tilde{S}(t;\hat{\theta})$ (dotted), and 90% BC$_a$ pointwise confidence bands (dashed). (Right) Saddlepoint hazard function $\tilde{z}_{13}(t;\theta)$ (solid), its saddlepoint estimate $\tilde{z}_{13}(t;\hat{\theta})$ (dotted), and 90% BC$_a$ pointwise confidence bands (dashed).

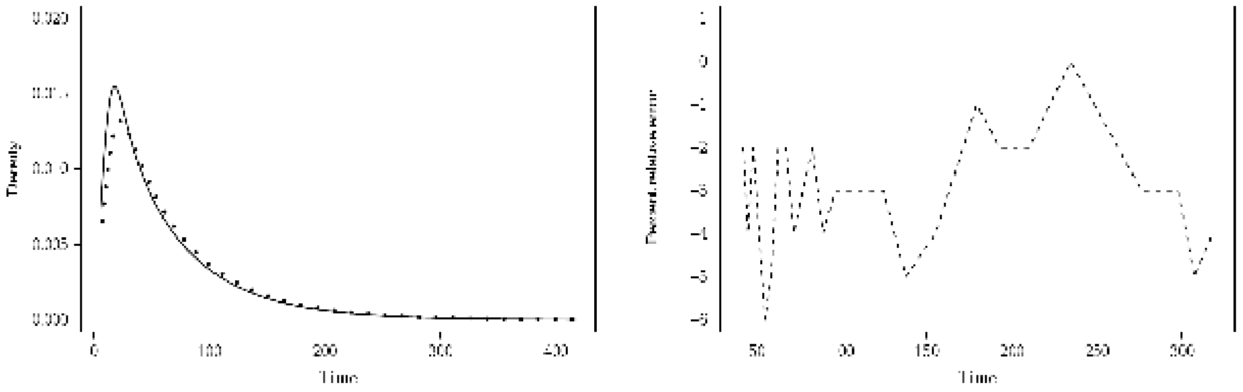

Fig. 4. (Left) Saddlepoint density function $\tilde{d}_{13}(t;\theta)$ (solid) and its saddlepoint estimate $\tilde{d}_{13}(t;\hat{\theta})$ (dotted). (Right) The percentage relative error (dashed) of coverage from nominal 90% coverage in survival function estimation. The relative error computation compares coverage error to 0·1.

Supplementary Material for Bootstrap confidence bands for sojourn distributions in multistate semi-Markov models with right censoring

R. W. Butler and D. A. Bronson (2012)

Proof of Theorem 3

Standard renewal theory modifications to the proof in Suzukawa (2002) suffice to show consistency. Suzukawa (2002) shows consistency for integral estimates, such as the first row of $\hat{T}(s)$, but in an isolated competing risk setting consisting of a single step. As such, his value of $n_{1\cdot}$ is fixed and not random as with our data. He lets $n_{1\cdot} \to \infty$ in the asymptotics but in our context $n_{1\cdot}$ is random and driven to ∞ with probability one as $N \to \infty$.

Denote random variable $n_{1\cdot}$ as N_1 and let event $A_{N_1} = \{|\hat{T}_{11}(s) - T_{11}(s)| > \eta\}$ for some $\eta > 0$. For any $\varepsilon > 0$, we must show that N_0 exists such that $\mathrm{pr}(A_{N_1}) < \varepsilon$ for $N > N_0$. The conditions of Theorem 3 allow Theorem 1 of Suzukawa (2002) to be applied conditional on N_1 so there exists $M(\varepsilon)$ such that the conditional probability $\mathrm{pr}(A_{N_1} \,|\, N_1) < \varepsilon/2$ for $N_1 > M(\varepsilon)$. Define event $B_\varepsilon = \{N_1 < M(\varepsilon)\}$ and let $N_0 = N_0(\varepsilon)$ be such that $\mathrm{pr}(B_\varepsilon) < \varepsilon/2$ when $N > N_0$. Then

$$\mathrm{pr}(A_{N_1}) = \mathrm{pr}(A_{N_1} \cap B_\varepsilon) + \mathrm{pr}(A_{N_1} \cap B_\varepsilon^C) \leq \varepsilon/2 + \sum_{k=M(\varepsilon)}^{\infty} \mathrm{pr}(A_{N_1} \mid N_1 = k)\, \mathrm{pr}(N_1 = k) < \varepsilon.$$

Thus all entries of $\hat{T}(s)$ are pointwise consistent for $T(s)$. Since each component is strictly increasing in s, pointwise consistency can be easily shown to extend to uniform consistency over compact sets. Uniform consistency for components of $\hat{T}'(s)$ and $\hat{T}''(s)$, the first two derivatives of $\hat{T}(s)$, is shown by using the same arguments.

In Theorem 2, since $\mathrm{pr}(\hat{\mathcal{R}} = \mathcal{R}) \to 1$ as $N \to \infty$, its $(m_{\hat{\mathcal{R}}}, m_{\hat{\mathcal{R}}})$-cofactor of $I_{m_{\hat{\mathcal{R}}}} - \hat{T}_{\hat{\mathcal{R}}\hat{\mathcal{R}}}(s)$ converges to $|\Psi_{m_{\mathcal{R}} m_{\mathcal{R}}}(s)|$ in probability as $N \to \infty$ uniformly in s. This ensures that its root $\hat{c}$ is consistent for c, the convergence strip boundary for $\mathcal{F}_{1m}(s)$. On compact subsets of $(-\infty, c)$, uniform consistency of $\hat{T}(s)$, $\hat{T}'(s)$, and $\hat{T}''(s)$ ensure the same for $\partial^k \hat{\mathcal{F}}_{1m}(s)/\partial s^k$, i.e. $\partial^k \hat{\mathcal{F}}_{1m}(s)/\partial s^k \to \partial^k \mathcal{F}_{1m}(s)/\partial s^k$ in probability uniformly on compact sets with $k = 0, 1$, and 2. Such uniformity for $k = 0, 1$ is needed to ensure consistency of the saddlepoint sequence $\hat{s}_t \to s_t$ as the roots of $\{\log \hat{\mathcal{F}}_{1m}(\hat{s}_t)\}' = t$. This, together with the uniform consistency of $\{\partial^k \hat{\mathcal{F}}_{1m}(s)/\partial s^k : k = 0, 1, 2\}$, ensure that saddlepoint estimates are uniformly consistent and that, for example, $\tilde{S}(t;\hat{\theta}) \to \tilde{S}(t;\theta)$ in probability uniformly over a corresponding range of t in the time domain.

Censoring is not tail-estimable

Suppose condition (iv) of Theorem 3 does not hold, and let $\mathcal{E}_i = \{j \in \mathcal{S} : F_{ij}(\tau) < 1\} \neq \emptyset$ for at least one i. The consistency properties of Theorem 3 continue to hold for some modified saddlepoint estimators, but only for $t < \tau$.

Proof of Theorem 4. There are three steps for determining the saddlepoint estimators. Consistency of these estimators then follows the same approach as used in Theorem 3.

The first step proposes estimators for the rows of $G(t)$. If $\mathcal{E}_i = \emptyset$, then the rescaled estimators in transmittance estimator (3), suffice for row i and are denoted as $\{\hat{G}^{\dagger}_{ij}(t) : j \in \mathcal{S}\}$. If $\mathcal{E}_i \neq \emptyset$, then use the following estimator

$$\hat{G}^{\dagger}_{ij}(t) = \hat{G}_{ij}(t)1_{\{t \leq \hat{\tau}_i\}} + \left[\hat{G}_{ij}(\hat{\tau}_i) + \frac{\hat{G}_{ij}(\hat{\tau}_i)}{\hat{G}_{i\cdot}(\hat{\tau}_i)}\{1 - \hat{G}_{i\cdot}(\hat{\tau}_i)\}H(t - \hat{\tau}_i)\right]1_{\{t > \hat{\tau}_i\}} \qquad (j \in \mathcal{S}),$$

that proportionately reallocates the unallocated mass $\{1 - \hat{G}_{i\cdot}(\hat{\tau}_i)\}$ to all destinations in $\mathcal{S}$ but placing the mass over $(\hat{\tau}_i, \infty)$. Here, H is any distribution function, such as an Exponential (1), that has a convergent Laplace-Stieltjes transform in an open neighbourhood of 0.

Step 2 is concerned with showing the resulting semi-Markov kernel $\hat{G}^{\dagger}(t) = \{\hat{G}^{\dagger}_{ij}(t)\}$ and its Laplace-Stieltjes transform $\hat{\theta}^{\dagger} = \hat{T}^{\dagger}(s)$ are consistent estimators of some population kernel $G^{\dagger}(t)$ and its transform $\theta^{\dagger} = T^{\dagger}(s)$, which differ from $G(t)$ and $T(s)$, but have the property that $G^{\dagger}(t) \equiv G(t)$ for $t < \tau$. If $\mathcal{E}_i = \emptyset$, then, from Theorem 3, rescaled estimator

$$\hat{G}^{\dagger}_{ij}(t) \to G_{ij}(t)$$

as $N \to \infty$ for all $t > 0$ and $j \in \mathcal{S}$. If $\mathcal{E}_i \neq \emptyset$, then

$$\begin{aligned} \hat{G}^\dagger_{ij}(t) &\to G_{ij}(t) 1_{\{t \leq \tau\}} + \left[G_{ij}(\tau) + \frac{G_{ij}(\tau)}{G_{i\cdot}(\tau)} \{1 - G_{i\cdot}(\tau)\} H(t-\tau) \right] 1_{\{t > \tau\}} \\ &= G^\dagger_{ij}(t). \end{aligned}$$

Note that $\{G^\dagger_{ij}(t)\}$ is a semi-Markov kernel, since each $G^\dagger_{ij}(t)$ is a cumulative incidence function, and

$$\begin{aligned} \sum_{j \in \mathcal{S}} G^\dagger_{ij}(\infty) &= \sum_{j \in \mathcal{S}} \left[G_{ij}(\tau) + \frac{G_{ij}(\tau)}{G_{i\cdot}(\tau)} \{1 - G_{i\cdot}(\tau)\} \right] \\ &= G_{i\cdot}(\tau) + \{1 - G_{i\cdot}(\tau)\} = 1. \end{aligned}$$

Note that

$$G^\dagger_{ij}(t) = G_{ij}(t) \qquad (t < \tau;\, i, j \in \mathcal{S}).$$

Thus the initial transitional dynamics of the daggered and undaggered processes are the same during time interval $(0, \tau)$.

For the final step, if $G^\dagger(t) = G(t)$ for $t < \tau$, then their respective first-passage distributions are identical during the same period but not afterwards. Thus, saddlepoint estimator $\tilde{S}(t; \hat{\theta}^\dagger)$, based on the daggered process, consistently estimates $\tilde{S}(t; \theta^\dagger) \simeq S(t; \theta^\dagger) = S(t; \theta)$ for $t < \tau$. $\square$

Type I censoring

Whether or not the consistency Theorems 3 and 4 are applicable with Type I censoring depends largely on how censoring distribution $C(t)$ is viewed, as either a continuous or discontinuous distribution. In the latter case, if $C(t) = H_0(t - \tau)$ where H_0 is the Heaviside function, then condition (ii) does not hold and both theorems are not applicable.

If, however, $C(t)$ is viewed as continuous with support on $[\tau - \varepsilon, \tau + \varepsilon]$ for sufficiently small $\varepsilon > 0$, then condition (ii), as it applies to $C(t)$, holds. Thus, with such an interpretation of Type I censoring, both theorems are applicable if the remaining conditions hold.

References

Suzukawa, A. (2002). Asymptotic properties of Aalen–Johansen integrals for competing risk data. *J. Japan Statist. Soc.* **1** 77–93.