Bootstrapping Survival Times in Stochastic Systems using Saddlepoint Approximations

Ronald W. Butler* and Douglas A. Bronson
Department of Statistics
Colorado State University
Fort Collins, CO 80523 USA

September 4, 2001

Abstract

The single bootstrap is implemented using a saddlepoint approximation to determine estimates for the survival and hazard functions of first passage times in complicated semi-Markov processes. The double bootstrap is also implemented by resampling saddlepoint inversions and provides $BC_a$ confidence bands for these functions. Confidence intervals for the mean and variance of first passage times are easily computed. A new characterization of the asymptotic hazard rate for survival times is presented and leads to an indirect method for constructing its bootstrap confidence interval.

Keywords: Bootstrap; Cofactor rule; Saddlepoint approximation, Semi-Markov process; System

1 Introduction

This paper shows how saddlepoint methods may be used to implement bootstrap computations for prediction and estimation of first passage time characteristics in complicated semi-Markov systems. These computations, when used in conjunction with more complicated systems, would not be practically feasible without the use of saddlepoint methods. It is convenient and instructive to introduce the procedures as they relate to the estimation methods presented in Davison and Hinkley (1988) and then generalize them to the full class of finite state semi-Markov systems.

Suppose waiting time $T = \sum_{i=1}^{m} X_i$ has an unknown distribution $F$ which is the convolution of a fixed number $m$ of independent and identically distributed (iid) variables with unknown distribution $G$. Let the data be summarized in an empirical distribution $\hat{G}$ based upon $n$ iid observations from $G$. Bootstrap inference for $\mu$, the mean of $T$, using saddlepoint methods has been addressed in Davison and Hinkley (1988) who provide percentile confidence intervals for $\mu$. These percentiles are based

*Corresponding author
upon the distribution of a resampled \(T^*\) with moment generating function (mgf) \(\hat{M}^m\) where \(\hat{M}\) is the empirical mgf

\[
\hat{M}(s) = \int \exp(sx) \, d\hat{G}(x).
\]  

(1)

Percentiles are approximated by inverting \(\hat{M}^m(s)\) using the Lugannani and Rice (1980) approximation to determine the appropriate bootstrap percentiles of \(T^*\) for use in setting these confidence intervals to a prescribed level.

In a systems setting, the waiting time \(T\) for the context above may be considered the first passage time from \(1 \to m + 1\) in a series connection of states for the semi-Markov process specified in the flowgraph of Figure 1.

Fig. 1. Series connection for an iid sum.

Nodes 1, \ldots, \(m + 1\) specify system states and the \(M_G\) transmittance, labelling transitions among states, is the mgf of \(G\) and identifies the holding time in the state of origin before instantaneous passage into the destination state. The total time of passage \(T\) has mgf \(M_G^T\) and the resampled \(T^*\) has the mgf \(\hat{M}^m\) used in Devison and Hinkley (1988).

In this paper, the single and double bootstrap are used for predictive inference about \(T\). The single bootstrap provides an estimate for the survival function of \(T\), or \(\bar{F}(t) := 1 - F(t)\), as well as the hazard rate function \(z(t) = f(t)/\bar{F}(t)\), where \(f(\cdot)\) is the supposed density for \(T\). These estimates are the saddlepoint approximations to the survival function and hazard rate of \(T^*\) obtained by inverting its mgf \(\hat{M}^m\).

The double bootstrap is introduced as a means for providing \(BC_a\) confidence bands for these estimated functions. If \(\hat{G}^*\) is a resampled value of \(\hat{G}\) with mgf \(\hat{M}^*\) determined from \(\hat{G}^*\) as in (1), then a doubly resampled \(T^{**}\) has the mgf \((\hat{M}^*)^m\) for the system in Figure 1. Distributional characteristics for \(T^{**}\), determined through saddlepoint inversion of \(B = 999\) realizations of \((\hat{M}^*)^m\), provide the second layer of resampling for constructing these bands. In addition, this effort provides bootstrap estimates of the guarantee of coverage in guaranteed coverage tolerance intervals of \(T\) as described in Aitchison and Dunsmore (1975, chap. 6) or Guttmann (1970).

The important contribution of this paper concerns the extension of these ideas to bootstrap inferences for survival and failure times in very general semi-Markov systems. Figure 1 has illustrated
only the simplest possible \((m + 1)\)-state system whereas we shall consider arbitrary \(m\)-state systems
with many distinct branching pathways incorporating recurrent states that result from a hierarchy of feedback loops in the flowgraph of state transitions. Figure 2 shows a simple example of this sort which is analyzed in §3.1. It shows the survival time \(T\) of a patient as the first passage time from \(1 \rightarrow 6\).

![Flowgraph showing the degenerative states of dementia.](image)

The feedback loops \(2 \rightarrow 3 \rightarrow 2, 3 \rightarrow 4 \rightarrow 3, 2 \rightarrow 3 \rightarrow 4 \rightarrow 3 \rightarrow 2\), etc. result in a countably infinite number of distinct paths over which we must sum in order to determine the mgf of first passage. The renewal theory necessary for performing such sums and which also facilitates saddlepoint approximation is outlined in the next section. The single and double bootstrap, implemented through saddlepoint approximation, provide \(BC_a\) confidence bands for the survival function of \(T\) and its hazard rate function. They also determine the tolerance interval of form \((0, t_a)\) that provides \(g\)-guarantee for \(100(1 - \gamma)\%\) coverage of \(T\).

Parametric inference about \(T\) often concerns its mean and variance for which estimates can be computed directly from the resampled mgfs. The \(BC_a\) confidence intervals computed for these moments illustrate a much simpler application of the double bootstrap in the systems theory context.

Another parameter of interest is the asymptotic failure rate of \(T\), when it exists. A new characterization of this limit is presented in §5 as the right edge of the convergence strip for its mgf. This simple result provides a means for estimating the asymptotic failure rate. Its point estimate is the smallest real pole of the mgf of \(T^*\). The determination of such poles for \(B\) realizations of the mgf of \(T^{**}\) leads to a \(BC_a\) confidence band for the asymptotic failure rate.

Two examples are considered. The flowgraph of Figure 2 was given by Commenges (1986) and
represents the degenerative transitions occurring for a dementia patient. A completely nonparametric bootstrap analysis estimates the survival function and hazard rate of such a patient and finds confidence bands on these functions. The second example is a semi-parametric bootstrap analysis of a GI/M/1 queue. It provides predictive and estimative analyses concerning passage times to queue lengths 5 and 10 starting from 0.

2 Semi-Markov Processes

A semi-Markov process or system with state space \( S = \{1, \ldots, m\} \) is a generalization of a finite Markov chain in which holding times in states may be nonexponential and may also depend upon the destination states of the transitions. While the behavior of a discrete time Markov chain is characterized in terms of an \( m \times m \) matrix of transition probabilities \( P = (p_{ij}) \), that of a semi-Markov system is characterized by two \( m \times m \) matrices which may be collapsed into one: a transition probability matrix \( P = (p_{ij}) \) for the underlying Markov chain of state transitions, and \( M(s) = \{M_{ij}(s)\} \), a matrix of 1-step mgfs. These probabilities and distributions characterize the dynamic behavior of the system in the following manner. Upon entering state \( i \), the next state of the system is randomly determined by the distribution given in the \( i \)th row of \( P \). If this is state \( j \), then the holding time in state \( i \), before proceeding to \( j \), is given by the distribution with mgf \( M_{ij}(s) \). These two matrices are combined into a transmittance matrix

\[
Q(s) = \{Q_{ij}(s)\} = P \odot M(s) = \{p_{ij}M_{ij}(s)\}
\]

which characterizes the dynamic behavior of the semi-Markov system and provides the basis for our bootstrap and saddlepoint computations. Each entry of \( Q \) is a transmittance defined as a probability \( \times \) a moment generating function.

2.1 First passage transmittances

Suppose \( T \) is the first passage time from state 1 to \( m \). The first passage transmittance is accordingly

\[
f_{1m}F_{1m}(s) := \mathbb{E}\{e^{sT}1(T<\infty)\},
\]

where \( f_{1m} = \Pr(T < \infty) \) is the probability of passage and \( F_{1m}(s) \) is the conditional mgf of \( T \) given \( T < \infty \). When \( f_{1m} < 1 \) the distribution of \( T \) is defective with \( \Pr(T = \infty) = 1 - f_{1m} \). We have of course taken state 1 as the source and state \( m \) as the destination state without any loss in generality.
Theorem 1 The first passage transmittance from state 1 to $m \neq 1$ is
\begin{equation}
 f_{1m} \mathcal{F}_{1m} (s) = \frac{(m,1) \text{-cofactor of } I_m - Q(s)}{(m,m) \text{-cofactor of } I_m - Q(s)} := \frac{(-1)^{m+1} |\Psi_{m1}(s)|}{|\Psi_{mm}(s)|},
\end{equation}
where $\Psi_{ij}(s)$ is the $(i,j)$th minor of $I_m - Q(s)$. The ratio (2) is well-defined over a maximal convergence neighbourhood of 0 of the form $(-\infty, c)$ for some $c > 0$ under these conditions:

1. The system states $S = \{1, \ldots, m\}$ are exactly those relevant to passage from $1 \rightarrow m$ with all relevant states included and no irrelevant states. State $i$ is said to be relevant to first passage from state 1 to $m$ if during passage it is a possible intermediate state. States 1 and $m$ are designated as relevant if passage $1 \rightarrow m$ is possible.

2. The maximal common neighbourhood of convergence for the moment generating functions in the first $m - 1$ rows of $Q(s)$ is an open neighbourhood of 0.

Proof. See Butler (2000a,b). $lacksquare$

When the source and destination states are both state 1, then the first return transmittance has a different form.

Theorem 2 The first return transmittance for state 1 is
\begin{equation}
 f_{11} \mathcal{F}_{11} (s) = 1 - \frac{|I_m - Q(s)|}{(1,1) \text{-cofactor of } I_m - Q(s)} := 1 - \frac{|I_m - Q(s)|}{|\Psi_{11}(s)|}.
\end{equation}
The ratio (3) is well-defined over an maximal convergence neighbourhood of 0 of the form $(-\infty, c)$ for some $c > 0$ under these conditions:

1. The system states $S = \{1, \ldots, m\}$ are exactly those relevant to passage from $1 \rightarrow 1$.

2. The maximal common neighbourhood of convergence for the moment generating functions in $Q(s)$ is an open neighbourhood of 0.

Proof. See Butler (2000a,b). $lacksquare$

Mason (1953, 1956) gave a complicated expression for $f_{1m} \mathcal{F}_{1m} (s)$ that is still used in control theory as may be seen from Phillips and Harbor (1996). Pyke (1961, thm. 4.2) and Howard (1964, 1971, §§10.10, 11.11) gave a simpler expression but it too is not especially amenable to saddlepoint implementation. These two alternative expressions have been shown to be analytically equivalent to (2) and (3) in Butler (2000a).
2.2 Saddlepoint Approximations

The cofactor rules of Theorems 1 and 2 lead to explicit saddlepoint expressions that are convenient for use with the Lugannani and Rice (1980) and density approximations as discussed in Daniels (1987). Let $\mathcal{K}(s) = \ln F_{1,m}(s)$ be the cumulant generating function (cgf) of $T$. The survival function approximation is

$$
\bar{F}_1(t) = 1 - \Phi (\bar{\omega}) - \phi (\bar{\omega}) \left( \frac{1}{\bar{\omega}} - \frac{1}{\bar{u}} \right) \quad t \neq \mathbb{E}(T) = \mathcal{K}'(0)
$$

where $\Phi$ and $\phi$ are the standard normal CDF and density, $\bar{\omega} = \bar{\omega}(\bar{s})$ and $\bar{u} = \bar{u}(\bar{s})$ depend on $\bar{s}$ according to

$$
\bar{\omega} = \text{sgn}(\bar{s}) \sqrt{2 (\bar{s} \mathcal{K}'(\bar{s}))} \quad \text{and} \quad \bar{u} = \bar{s} \sqrt{\mathcal{K}''(\bar{s})},
$$

and saddlepoint $\bar{s}$ solves $\mathcal{K}'(\bar{s}) = t$ for $t > 0$. The standard density estimate

$$
f_1(t) = \frac{1}{\sqrt{2\pi \mathcal{K}''(\bar{s})}} \exp \left( -\frac{1}{2} \frac{\bar{u}^2}{\bar{s}} \right),
$$

given by Daniels (1954), combines with $\bar{F}_1(t)$ to provide two approximations for the hazard rate:

$$
z_1(t) = \frac{f_1(t)}{\bar{F}_1(t)} \quad \text{and} \quad z_2(t) = \frac{f_1(t)}{\bar{F}_1(t) \int_0^\infty f_1(u) du}.
$$

The saddlepoint equation $\mathcal{K}'(\bar{s}) = t$ has a simple form since $F_{1,m}$ is a ratio of the cofactor determinants in both (2) and (3). For passage from $1 \to m \neq 1$,

$$
\mathcal{K}' = \text{tr} \left( \Psi_{m1}^{-1} \Psi_{m1}' - \Psi_{mm}^{-1} \Psi_{mm}' \right)
$$

with $\Psi_{ij} = \partial \Psi_{ij} / \partial s$ and the dependence on $s$ suppressed. In computing $\bar{u}$, the second derivative is also explicit as

$$
\mathcal{K}'' = \text{tr} \left\{ \Psi_{m1}^{-1} \Psi_{m1}'' - (\Psi_{m1}^{-1} \Psi_{m1}')^2 - \Psi_{mm}^{-1} \Psi_{mm}' + (\Psi_{mm}^{-1} \Psi_{mm}')^2 \right\}.
$$

These simple expressions for $\mathcal{K}'$ and $\mathcal{K}''$ follow as a direct consequence of the simplicity of $\mathcal{K}$ as expressed through the cofactor rules (2) and (3). For any $m$ of reasonable size, the same derivative expressions are horrendous when determined using the expressions for $\mathcal{K}$ in Mason (1953, 1956) or Pyke (1961) and Howard (1964, 1971). This places severe limits on the computational tractability of these other expressions. By contrast, the cofactor rules have allowed the authors to successfully perform saddlepoint computations with general systems of size $m = 250$ with no numerical difficulties.
3 Nonparametric Bootstrap

Data for inference about a survival time, which is a first passage from $1 \rightarrow m$ in a semi-Markov system, can come in a variety of complete and incomplete forms. We shall only consider the simplest complete case in which our data consist of the full histories $\mathcal{H}_1, \ldots, \mathcal{H}_h$ of $h$ iid patients or systems. Here, $\mathcal{H}_i$ contains the sequence of states and holding times during the $i^{th}$ patient's lifetime. These $h$ patients are iid which means that their first passage is from $1 \rightarrow m$ and follows a common transmittance matrix $Q$.

The minimal requirement for bootstrapping is that an empirical estimate $\hat{Q}(s)$ of the transmittance matrix is available and that a sensible resampling mechanism can be implemented to determine $\hat{Q}^*$, a resampled value of $\hat{Q}$. With the full history data $\{\mathcal{H}_i\}$, this estimate is

$$\hat{Q}(s) = \hat{P} \circ \hat{M}(s) = \{ \hat{p}_{ij} \hat{M}_{ij}(s) \}.$$  

Matrix $\hat{P} = (\hat{p}_{ij}) = (n_{ij}/n_i)$ consists of empirical transition rates among the states pooled across all $h$ patients. Empirical mgf $\hat{M}_{ij}$ is computed from the pooled collection of holding times in state $i$ before passage to $j$. Transition $i \rightarrow j$ is not possible when $p_{ij} = 0$ and in such cases $\hat{p}_{ij} = 0$. When $p_{ij} > 0$, we shall also assume sufficient data that $\hat{p}_{ij} > 0$; should this not happen, then we might add a nominal amount in the appropriate cells. This is necessary to prevent the $\hat{Q}$ process from having irrelevant states that are relevant in the true $Q$ process. Indeed the theory of primitive matrices in Seneta (1981, p. 3) assures that the relevant states of $\hat{Q}$ are exactly those in $Q$ when

$$\text{sgn}(\hat{p}_{ij}) = \text{sgn}(p_{ij})$$  

for all $i,j$. The presence of irrelevant states in the $\hat{Q}$ process can cause removable singularities at $s = 0$ in the cofactor rules of (2) and (3) (see Butler, 1997a). This in turn leads to a destabilization of the numerical saddlepoint computation in a wide range about the mean of the distribution.

A resampled survival time $T^*$ is obtained by simulating the first passage time of the semi-Markov process characterized as having transmittance matrix $\hat{Q}$. It is important at this point to recognize that $\hat{Q}$ itself indexes a semi-Markov process that is an estimate of the unknown process $Q$. Thus, the estimation space for the processes is closed and the distribution of $T^*$ is approximated through saddlepoint inversion to provide an estimate of the distribution of $T$.

Theorem 3 A resampled $T^*$ has first passage transmittance $\hat{f}_{1m}\hat{F}_{1m}(s)$ obtained by substituting $\hat{Q}$ into (2) in place of $Q$. Estimates for the survival and hazard rate functions of $T$ are those for $T^*$.
and are computed from saddlepoint inversion of $\hat{F}_{1m}(s)$.

**Proof.** Each transition of $T^*$ is simulated using probabilities and holding times determined from empirical data in $\hat{Q}$. The proof is therefore immediate. ■

A doubly resampled $T^{**}$ is the passage time using a resampled $\hat{Q}^* = \{\hat{p}_{ij}^*, \hat{M}_{ij}^*\}$. Resampled values $\hat{M}_{ij}^*$ are based upon sampling the $i \to j$ transitions $n_{ij}$ times with replacement. Resampled probabilities are $\hat{p}_{ij}^* = n_{ij}^*/n_i$, where the vector

$$\{n_{1i}^*, \ldots, n_{im}^*\} | \{n_{ij}\} \sim \text{Multinomial} \ (n_i; \hat{p}_{i1}, \ldots, \hat{p}_{im}).$$

(11)

Again we see that zero values for $n_{ij}^*$, when $\hat{p}_{ij} > 0$, can create irrelevant states in the system transmittance $\hat{Q}^*$. We therefore substitute nominal value $\varepsilon > 0$ for such 0-values to avoid this difficulty and correspondingly diminish the other nonzero values proportionately to make $n_i^* = n_i$. Thus, in all resampled versions of $\hat{Q}^*$, we are able to maintain as relevant those states known to be relevant from $\hat{Q}$.

**Theorem 4** A resampled $T^{**}$ has first passage transmittance $\hat{f}_{1m}^* \hat{F}_{1m}(s)$ obtained by substituting $\hat{Q}^*$ into (2) in place of $\hat{Q}$. An ensemble of $B$ estimates for the survival and hazard rate functions of $T^{**}$ provide confidence bands for the point estimates based upon $\hat{F}_{1m}(s)$.

The resampling just described has not taken into account the random variation in $\{n_i\}$ and is therefore referred to as having **fixed row totals (frt)**. To compensate for this, a different sort of resampling might be considered. A resample of the patients, e.g. a random sample with replacement of size $h$ from the indices $\{1, \ldots, h\}$, followed by pooling of their transition rates, leads to the determination of $\{n_i^*\}$ as a replacement for the fixed values of $\{n_i\}$ in (11). Such a resampling of the patients can be used to initiate each of the $B$ resamples in (11) and accounts for the random variation in $\{n_i\}$. We refer to this method as having **resampled row totals (rrt)**.

The $B$ resampled determinations of $\hat{Q}^*$ are inverted using saddlepoint approximations over a grid of time points $\{t_i\}$ to form ensembles of estimates of survival and hazard rate functions. The saddlepoint inversions have effectively eliminated the need for the second layer of resampling within the $B$ outer layer resamples. This method for efficient implementation of the double bootstrap was introduced in Hinkley and Shi (1989) and further developed by DiCiccio et al. (1992a,b, 1994) in different contexts.

To determine $BC_{a}$ confidence bands, the accelerations $\{a_i\}$ must be estimated over all the grid points $\{t_i\}$. The underlying data structure is multisample with each nonzero component of $\hat{M}(s)$
representing an independent sample. The skewness estimates for \( \{ a_i \} \) with multisamples (Davison and Hinkley, 1997, (5.23)) are determined as estimates of the jackknife estimators. The true jackknife estimators (Efron and Tibshirani, 1993, p.186) over the grid \( \{ a_i \} \) would require summing over \( \prod_{i,j=1}^{m} n_{ij} \) ensembles of saddlepoint inversions created by leaving out all combinations of single observations from each of the samples. Since this amount of computation is prohibitive, a sum of 200 randomly sampled combinations has been used instead. A sample size increase to 1000 showed little change in the resulting \( BC_a \) intervals in the example considered below. We may infer from this that 200 further ensemble inversions are sufficient in estimation of \( \{ a_i \} \) for \( BC_a \) confidence bands. This sufficed in our example, but it may not in other examples. Overall 1200 saddlepoint inversions are used in the example below.

3.1 Dementia Example

In the flowgraph of Figure 2, state 1 represents good health, states 2-4 increasingly severe states of dementia, state 5 a terminal state, and state 6 is death. Since the data and specifics of the French study in Commenges (1986) are not available, we assume a specific system transmittance matrix so that bootstrap accuracy may be assessed. Let

\[
Q(s) = \begin{pmatrix}
0 & 0.6 g(2, 2) & 0.25 g(4, 8) & 0.15 g(9, 27) & 0 & 0 \\
0 & 0 & 0.85 r(\frac{1}{2} \sqrt{\pi}) & 0 & 0.15 \text{ig}(1, \frac{1}{2}) & 0 \\
0 & 0.7 r(\sqrt{\pi}) & 0 & 0.2 r(\sqrt{\pi}) & 0.1 \text{ig}(\frac{1}{2} \sqrt{2}, \frac{1}{2} \sqrt{2}) & 0 \\
0 & 0 & 0.7 r(\frac{3}{2} \sqrt{\pi}) & 0 & 0.3 \text{ig}(\sqrt{2}, \sqrt{2}) & 0 \\
0 & 0 & 0 & 0 & 0 & \text{ig}(1, 1)
\end{pmatrix}
\]

(12)

where \( g(a, b) \) and \( \text{ig}(a, b) \) are gamma and inverse Gaussian mgfs with mean \( a \) and variance \( b \), and \( r(a) \) is a Raleigh distribution mgf with mean \( a \). From this system, we simulate data consisting of the complete histories of 25 patients starting at state 1 and ending in absorbing state 6. The data of these 25 histories determine \( Q(s) \). No adjustment to \( \hat{P} \) for irrelevant states was necessary since our data satisfied (10). If \( T \) is the passage time from 1 \( \rightarrow \) 6, neither its density nor survival function can be exactly determined from (12), however saddlepoint versions may be computed so that "exact" comparison may be made with respect to these saddlepoint quantities.
3.1.1 Predictive Inference

Table 1 shows the accuracy achieved in using saddlepoint approximations to determine estimates and 90\% BC\alpha confidence intervals for five right-tail percentiles of \( T \). For example, the first column lists survival probability 0.25 followed by its associated "exact" percentile \( \bar{F}_{1}^{-1}(0.25) = 19.0 \) in the second column. This percentile has been determined from saddlepoint inversion based upon the cofactor rule applied to \( Q(s) \) in (12). The same saddlepoint inversion, but based upon \( \hat{Q}(s) \), provides the estimate 19.9 listed under SA. The accuracy of saddlepoint estimate 19.9 may be compared to the direct bootstrap resampling estimate 20.1 listed under Sim. This resampling estimate required simulation of \( 2 \times 10^5 \) values of \( T^* \) by repeatedly passing from 1 \( \rightarrow \) 6 through the system \( \hat{Q}(s) \); the resulting empirical survival function yielded estimate 20.1 as its 25th percentile. Both of the outer layer resampling schemes, frt and rrt, were used in conjunction with saddlepoint approximation at the inner level to determine 90\% BC\alpha confidence intervals for the 25th survival percentile 19.0; these intervals are (15.7, 26.9) and (15.9, 26.8) respectively. Direct double bootstrap frt resampling produced the BC\alpha interval (15.1, 26.8) listed under the Sim columns. Its computation consisted of resampling \( B = 999 \) replications of \( \hat{Q}^* \) and thereafter generating \( 2 \times 10^5 \) values of \( T^{**} \) from each system \( \hat{Q}^* \). For each of the 5 survival percentiles, the Sim bounds agree closely with the frt bounds which in turn also show a striking degree of similarity to the rrt bounds. This suggests that the simpler resampling scheme with fixed row totals may be adequate in this example even with the relatively small amount of data associated with 25 patient histories. Little difference was seen between frt and rrt resampling in all the remaining computations of this example so additional results are shown only for frt resampling.

Figures 3 and 4 show the "exact" density \( f_1(t) \) and survival function \( \bar{F}_1(t) \) of \( T \) as solid lines along with their estimates (short dashed lines) based upon saddlepoint inversion of \( \hat{Q}(s) \). Empirical estimates (dotted lines), based upon simulating \( 2 \times 10^5 \) values of \( T^* \) (and using kernel estimation for the density) are almost graphically indistinguishable from their saddlepoint counterparts. The long dashes in Figure 4 enclose 90\% BC\alpha confidence bands computed on a grid of 201 time points. "Exact" hazard rate \( z_2(t) \) for \( T \) is shown in Figure 5 along with its saddlepoint estimate from inverting \( \hat{Q}(s) \), 90\% BC\alpha confidence bands, and a simulation estimate using kernel density estimation. Relative errors in density and survival estimation, as compared with the "exact" versions, along with errors using simulation are shown in Figure 6. The proposed estimates using saddlepoint inversions are smoother and more consistently accurate than simulation. The plots using saddlepoint methods
required 0.38 hours as compares with 12.01 hours for the simulation.

3.1.2 Parametric Inference

The mean and standard deviation of $T$ are easily estimated using the cofactor rule in Theorem 1. The results appear in the first row of Table 2. The exact mean is $\hat{\mathcal{F}}_{16}(0) = 14.4$ which, along with $\hat{\mathcal{F}}_{16}'(0) = 322$, determines the standard deviation as 10.7. The same computations based upon $\hat{\mathcal{F}}_{16}$ determine the estimates 15.3 and 11.1. $BC_\alpha$ confidence intervals may be computed in a couple seconds by resampling $B$ values of $\hat{\mathcal{F}}_{16}^*$ and $\hat{\mathcal{F}}_{16}^{**}(0)$; the 90% $BC_\alpha$ confidence bands are shown as the cofactor rule method.

The second row in Table 2 treats the survival data as an iid model and ignores the information in the data about the semi-Markov structure. The 25 complete histories are collapsed into the 25 patient survival times with sample mean and standard deviation of 15.3 and 10.6 as listed in the table. The 90% $BC_\alpha$ intervals are also given.

The benefit of using the semi-Markov structure as opposed to the iid model when estimating the mean and standard deviation is revealed through the disparity in coverage accuracy for the two confidence methods. The original data of 15 complete histories were simulated 10,000 times and the 90% confidence intervals of Table 2 repeatedly computed. Table 3 displays the attained coverages of percentile and of $BC_\alpha$ intervals using two levels for the number of jackknife eliminations in acceleration estimation. Coverages using 200 and 1000 jackknife eliminations are comparable.

Bootstrap intervals using the cofactor rule method provide intervals with very accurate coverage whereas the iid method shows undercoverage for the mean and substantial undercoverage for the standard deviation. Such small sample undercoverage using the iid method has been recently noted in Polansky (1999). He explains that the undercoverage derives from the discrepancy between the bounded support of the bootstrap distribution as compares with the unbounded tail of the true passage distribution.

Undercoverage, when using the cofactor rule method with a feedback system such as in Figure 2, cannot occur as a result of the phenomenon described by Polansky (1999). This applies both to confidence bands for the survival function as well as for moments. The reason for this is that the presence of a feedback loop extends the support of $T^*$ and $T^{**}$ to $\infty$ so the right tail of support for $T^*$ and $T^{**}$ matches that of $T$. These situations differ however in finite cascading semi-Markov systems within which states cannot be repeated. Here there are no feedback loops and the number
of states bounds the number of transitions; thus $T^*$ and $T^{**}$ have bounded support in the right tail. If their right tail support differs from that of $T$, then the cofactor rule method is susceptible to undercoverage.

4 Semi-parametric Bootstrap

In some situations $Q = Q(G, \lambda)$ depends only upon an unknown parameter $\lambda$ and distribution function $G$. Such was the situation in Davison and Hinkley (1988) which involved $G$ but not $\lambda$. Finite state $M/G/1$ and $GI/M/1$ queues are examples of complex systems in which the resampling is part parametric and part nonparametric.

4.1 Example: GI/M/1 queue

Suppose states in $S = \{0, \ldots, m\}$ describe the length of a $GI/M/1$ queue of tasks (including one under service) that starts at 0 and "fails" upon arrival at length $m$ in time $T_m$. The transmittance matrix $Q$ has dimension $m + 1$ and depends only upon two quantities: the interarrival distribution $G$ of the renewal process of tasks, and $\lambda > 0$, the fixing rate of the Exponential ($\lambda$) fixing time for the single server. With $m = 5$, this transmittance matrix is

$$Q(s) = \begin{pmatrix}
0 & U_0(s) & 0 & 0 & 0 & 0 \\
0 & Q_{11}(s) & U_0(s - \lambda) & 0 & 0 & 0 \\
0 & Q_{21}(s) & U_1(s - \lambda) & U_0(s - \lambda) & 0 & 0 \\
0 & Q_{31}(s) & U_2(s - \lambda) & U_1(s - \lambda) & U_0(s - \lambda) & 0 \\
0 & Q_{41}(s) & U_3(s - \lambda) & U_2(s - \lambda) & U_1(s - \lambda) & U_0(s - \lambda) \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix},$$

where the entries are generally

$$U_i(s) = \int_0^\infty we^{sw}dG(w) \quad i = 0, \ldots, m - 2$$

$$Q_{41}(s) = U_0(s) - \sum_{k=0}^{i-1} U_k(s - \lambda) \quad i = 1, \ldots, m - 1.$$

To understand this matrix, consider for example entry $(5, 3)$ for the change $4 \rightarrow 2$. Upon leaving state 4 after time $W = w$, there must have been 3 tasks completed during this time for passage to state 2; therefore the entry is an expected product of $\{a Poisson (\lambda W) probability for 3\} \times e^{sw}$ or $U_3(s - \lambda)$.
There are several data schemes possible according to the way the process is observed. We suppose the data are iid \( \{(w_i, x_i) : i = 1, \ldots, n\} \) where \( w_i \) is the \( t \)th interarrival time and \( x_i | w_i \sim \text{Poisson} (\lambda w_i) \) counts the number of tasks completed in the interim. Two sufficient statistics summarize this data: \( \hat{G} \), the empirical distribution of \( \{w_i\} \), and \( \hat{\lambda} = x_i / w_i \), the MLE as a ratio of sums. The distribution \( T_m^* \) has the transmittance \( Q(\hat{G}, \hat{\lambda}) \).

**Theorem 5** The transmittance matrix \( Q(\hat{G}, \hat{\lambda}) \) characterizes a semi-Markov process whose first passage time to length \( m \) has the distribution \( T_m^* \).

The distribution of \( T_m^* \) has transmittance \( Q(\hat{G}^*, \hat{\lambda}^*) \) where the starred estimates are determined by resampling with replacement \( n \) pairs from \( \{(w_i, x_i) : i = 1, \ldots, n\} \).

### 4.1.1 Predictive Inference

As an example, take \( G \) as Gamma \((2, 2)\) with mean 1 and variance \( \frac{1}{2} \). Suppose the single server completes tasks at rate \( \lambda = 5/4 \) and \( n = 100 \). With \( m = 5 \), the true survival function \( \bar{F}(t) \) of \( T_5 \) is plotted in Figure 7 as a solid line along with its estimate \( \bar{F}_1(t) \) (short dashed line) based upon saddlepoint inversion of \( Q(\hat{G}, \hat{\lambda}) \). An empirical estimate (dotted line), based upon simulating \( 2 \times 10^5 \) values of \( T_5^* \), is graphically indistinguishable from \( \bar{F}_1(t) \). The long dashes enclose 90\% \( BC_\alpha \) confidence bands computed on a grid of 201 time points. The same features but with \( n = 500 \) are shown in Figure 8 where the greater informativeness of the data has narrowed the \( BC_\alpha \) bands.

Returning to the \( n = 100 \) setting, the true hazard rate of \( T_5 \) is shown in Figure 9 along with its saddlepoint estimate \( \bar{z}_2(t) \) from inverting \( Q(\hat{G}, \hat{\lambda}) \), 90\% \( BC_\alpha \) confidence bands, and a simulation estimate. Relative errors in density and survival estimation are shown in Figure 10. The saddlepoint estimates and confidence bands used 1/35th of the cpu time required by the simulations.

Computation of the exact density and survival function of \( T_5 \) is not generally possible with a GI/M/1 model. When \( G \) is Gamma \((2, 2)\), the computation becomes possible and is the rationale for our choice here. Butler and Bronson (2000) show that, since a Gamma \((2, 2)\) is the sum of two Exponential \((2)\) variables, the number of system states can be doubled to create a Markov process for which \( T_5 \) is a passage time. Its distribution is therefore computable exactly as a phase-type distribution (Aalen, 1995).

Table 4 shows the accuracy achieved in using saddlepoint approximations (SA) to determine estimates and 90\% \( BC_\alpha \) confidence intervals for various survival percentiles of \( T_5 \). For the various
survival probabilities, their associated exact percentiles are given as Exact. The SA entry just below is the saddlepoint approximation to Exact using $Q(G,\lambda)$, the true cdf of $T_5$. Point estimates and $BC_a$ confidence bands based upon saddlepoint approximation (SA) and simulation (Sim) are shown for sample sizes $n = 100$ and 500. There is a striking degree of similarity between the SA and Sim entries.

Tolerance intervals with 90% guarantee of coverage $1- (\text{survival prob.})$ are displayed in the final column again based upon saddlepoint and simulation methods. For example, with coverage $1- 0.25$ and $n = 500$, 46.0 is found by both methods as the smallest $c > 0$ for which

$$\Pr \{ \Pr (T_5 > c) \geq 0.75 \} \geq 0.9,$$

(14)
e.g. there is 90% guarantee of 75% coverage. Solution to (14) using saddlepoint approximations proceeds as follows: Take $B = 999$ saddlepoint inversions of $Q(\hat{G}^*, \hat{\lambda}^*)$ to determine $B$ survival approximations $\{ \hat{F}^* \}$ from which we extract their 25th survival percentiles as $\{ \hat{F}^*^{-1}(0.25) \}$. The 90th percentile of the cdf of this sequence is the value of $c$ that solves (14). Sim entries were determined by implementing the double bootstrap: For each of the $B$ resampled systems $Q(\hat{G}^*, \hat{\lambda}^*)$ with passage time $T_5^*$, $2 \times 10^5$ generations of $T_5^{**}$ were simulated and the 25th percentile of the empirical survival function of these $2 \times 10^5$ generations was determined. Value $c$ is the 90th percentile of the cdf of these empirical survival percentiles. The table entries show striking agreement between the saddlepoint methods and the double bootstrap methods they approximate.

Table 5 provides cpu times needed to determine estimates using the single bootstrap (Est.) and confidence bands using the double bootstrap (CIs) for the various sample sizes.

Further discussion and numerical work concerning first passage to queue length 10 or $T_{10}$ may be found in Butler and Bronson (2000).

4.1.2 Parametric Inference

Bootstrap estimates and confidence bounds of the mean and standard deviation for the distributions of passage time $T_5$ are given in Table 6. Estimates of the coverage probabilities of these intervals are shown in Table 7 and were computed in the same manner as those in Table 3 based upon 10,000 repetitions. The target coverage of 90% is accurately achieved in all instances. For this example, the simpler percentile intervals appear quite accurate in coverage.

The stationary distribution for a semi-Markov process has a mass function that can also be estimated. Our queue of length 5 is not stationary, but we can make it so by changing the queue
discipline. Queue length is only observed after interarrival epochs of time $G$. The holding time in state 5 is thus $G$ and leads to a return to state 5 should one or less tasks be completed in the interim; passage to states 1, ..., 4 results from the completion of two or more tasks. This makes the queue a semi-Markov process whose stationary distribution depends only upon $G$ and $\lambda$ (see Ross, 1983 §4.8). Table 8 displays 90% $BC_a$ confidence intervals on all of its stationary probabilities.

5 Asymptotic Hazard Rates

Theorem 6 presents a new characterization of the asymptotic average hazard rate. The result has importance apart from our use of it with passage time distributions for semi-Markov systems.

The cumulative hazard rate or hazard of survival time $T$ is defined as

$$\Lambda(t) = \int_0^t \frac{f(x)}{F(x)} dx = -\ln \{\overline{F}(t)\}.$$ 

**Theorem 6** Suppose waiting time $T$ has moment generating function $M(s)$. The convergence strip for $M(\cdot)$ is either $(-\infty, b)$ or $(-\infty, b]$ for $0 \leq b \leq \infty$ if and only if

$$\lim_{t \to \infty} \inf \frac{\Lambda(t)}{t} = b.$$ 

**Proof.** See Butler and Robinson (2001). □

When the asymptotic hazard rate exists, it may be determined using L'Hospital's rule as

$$\lim_{t \to \infty} \frac{\Lambda(t)}{t} = \lim_{t \to \infty} \frac{\Lambda'(t)}{1} = b$$

under conditions that assure passage time $T$ has a continuous density for sufficiently large $t$. For an extensive discussion of these results and their relationship with large deviation theory, see Butler and Robinson (2001).

5.1 Bootstrap Estimation

The asymptotic hazard rate for passage time $T$ is the smallest positive real pole of $\mathcal{F}_{1m}$ which is the smallest real root of $|\Psi_{mn}(s)| = 0$. It may be estimated by finding the comparable pole in $\hat{\mathcal{F}}_{1m}$ or the smallest real root of $|\hat{\Psi}_{mn}(s)| = 0$ instead. Such roots, when computed for $B$ repetitions of the resampled $\hat{\mathcal{F}}_{1m}$, allow for the computation of $BC_a$ confidence intervals for the asymptotic hazard rate.
In the dementia example, the true asymptotic rate is 0.0995 with estimate 0.0955 determined from $\hat{F}_{1m}$. The 90% $BC_{\alpha}$ confidence band is (0.0654, 0.128).

For the queue example, the true asymptotic rate is 0.0396. Estimates with sample sizes 100 and 500 are 0.0479 and 0.0437 respectively. The corresponding $BC_{\alpha}$ confidence bands are (0.0270, 0.0840) and (0.0320, 0.0575). The coverage accuracies of the $BC_{\alpha}$ intervals and percentile intervals (not provided) with sample size 100 were assessed by computing 2500 repetitions of these intervals. Respective empirical coverages of 0.892 and 0.885 were observed demonstrating quite accurate coverage for both.

6 Final Remarks

There are many theoretical and methodological issues connected with these saddlepoint/bootstrap methods that remain to be considered. On the theoretical side, analytical results are needed to compare the two approaches to bootstrap sampling, fixed row total versus random row total. Practically, there is also a need to allow the system dynamics to depend upon patient covariates and to allow for the possibility for censoring from the various system states.

7 Acknowledgments

The research of the authors has been supported by the National Science Foundation under grants DMS-9625396 and DMS-9970785.

References


Table 1. (Dementia survival distribution). Estimates and 90% $BC_a$ confidence intervals for various percentiles of $T$ at the survival probabilities listed in the first column. Saddlepoint intervals based upon frt and rrt outer layer resampling are compared with double bootstrap intervals (Sim) using $B \times 2 \times 10^5$ simulations of $T^{**}$.

<table>
<thead>
<tr>
<th>Survival &quot;Exact&quot;</th>
<th>Estimate</th>
<th>$BC_a$ Lower</th>
<th>$BC_a$ Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob. Perc.</td>
<td>SA Sim</td>
<td>frt rrt Sim</td>
<td>frt rrt Sim</td>
</tr>
<tr>
<td>0.50 11.5</td>
<td>12.3 12.6</td>
<td>9.7 9.8 9.9</td>
<td>15.9 15.9 15.1</td>
</tr>
<tr>
<td>0.25 19.0</td>
<td>19.9 20.1</td>
<td>15.7 15.9 15.1</td>
<td>26.9 26.8 26.8</td>
</tr>
<tr>
<td>0.10 28.4</td>
<td>29.6 29.7</td>
<td>23.2 23.5 23.5</td>
<td>40.6 41.6 41.0</td>
</tr>
<tr>
<td>0.05 35.4</td>
<td>36.9 37.0</td>
<td>28.7 29.0 29.0</td>
<td>51.0 52.6 50.5</td>
</tr>
<tr>
<td>0.01 51.7</td>
<td>53.8 53.7</td>
<td>41.4 41.8 41.7</td>
<td>75.4 78.2 74.5</td>
</tr>
</tbody>
</table>

Table 2. (Dementia parameter estimates). Confidence intervals for the mean and the standard deviation of $T$. The iid model assumes there are 25 iid survival times.

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$BC_a$</td>
<td>$BC_a$</td>
</tr>
<tr>
<td></td>
<td>Exact Est.</td>
<td>Lower Upper</td>
</tr>
<tr>
<td>Cofactor Rule</td>
<td>14.4 15.3 12.3 20.4</td>
<td>10.7 11.1 8.5 16.0</td>
</tr>
<tr>
<td>iid model</td>
<td>15.3 11.9 19.1</td>
<td>10.6 8.1 14.2</td>
</tr>
</tbody>
</table>

Table 3 (Dementia coverage probabilities). Estimated coverage probabilities for the 90% confidence intervals illustrated in Table 2.
<table>
<thead>
<tr>
<th>Survival Prob.</th>
<th>Exact SA</th>
<th>Sample Size</th>
<th>Estimate SA Sim</th>
<th>BCα Lower SA Sim</th>
<th>BCα Upper SA Sim</th>
<th>Guar. Tol. SA Sim</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>21.3</td>
<td>100</td>
<td>18.6 18.6</td>
<td>12.2 12.6</td>
<td>29.7 29.5</td>
<td>27.5 27.5</td>
</tr>
<tr>
<td></td>
<td>21.2</td>
<td>500</td>
<td>19.6 19.6</td>
<td>15.8 15.3</td>
<td>25.3 24.3</td>
<td>24.1 24.1</td>
</tr>
<tr>
<td>0.25</td>
<td>38.8</td>
<td>100</td>
<td>33.1 33.0</td>
<td>20.5 21.0</td>
<td>55.6 55.4</td>
<td>50.1 50.1</td>
</tr>
<tr>
<td></td>
<td>38.8</td>
<td>500</td>
<td>35.6 35.4</td>
<td>27.9 26.8</td>
<td>46.7 44.6</td>
<td>46.0 46.0</td>
</tr>
<tr>
<td>0.10</td>
<td>62.0</td>
<td>100</td>
<td>52.3 52.1</td>
<td>31.6 32.2</td>
<td>89.2 89.7</td>
<td>80.3 77.8</td>
</tr>
<tr>
<td></td>
<td>62.0</td>
<td>500</td>
<td>56.6 56.3</td>
<td>43.8 41.8</td>
<td>75.7 71.7</td>
<td>70.7 70.7</td>
</tr>
<tr>
<td>0.05</td>
<td>79.5</td>
<td>100</td>
<td>66.8 66.4</td>
<td>40.0 40.9</td>
<td>115.5 115.0</td>
<td>105.4 100.4</td>
</tr>
<tr>
<td></td>
<td>79.6</td>
<td>500</td>
<td>72.6 72.3</td>
<td>55.9 53.5</td>
<td>97.3 92.4</td>
<td>92.8 92.8</td>
</tr>
<tr>
<td>0.01</td>
<td>120.1</td>
<td>100</td>
<td>100.5 99.7</td>
<td>59.2 60.7</td>
<td>173.6 177.2</td>
<td>158.2 150.7</td>
</tr>
<tr>
<td></td>
<td>120.4</td>
<td>500</td>
<td>109.5 110.0</td>
<td>83.9 80.2</td>
<td>147.6 143.0</td>
<td>139.6 139.6</td>
</tr>
</tbody>
</table>

Table 4. (Queue). Estimates and 90% $BC_α$ confidence intervals for various percentiles of $T_5$ using the survival probabilities listed in the first column. The $BC_α$ intervals have been computed using saddlepoint methods (SA) and $B \times 2 \times 10^8$ simulations (Sim) of $T_5^\ast$. The last column has cutoffs for 90% guaranteed coverage tolerance intervals of coverage 1− (survival prob.).

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Est. (sec)</th>
<th>CIs (hrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SA Sim</td>
<td>SA Sim</td>
</tr>
<tr>
<td>100</td>
<td>0.9</td>
<td>29</td>
</tr>
<tr>
<td>500</td>
<td>2.2</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 5. (Queue). CPU times for estimating the first passage time density and survival function of $T_5$.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Exact SA</th>
<th>BCα</th>
<th>BCα Lower</th>
<th>BCα Upper</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>BCα</td>
<td></td>
<td></td>
<td>Std. Dev.</td>
</tr>
<tr>
<td></td>
<td>Exact</td>
<td>Est.</td>
<td>Lower</td>
<td>Upper</td>
<td>Exact</td>
</tr>
<tr>
<td>100</td>
<td>29.0</td>
<td>25.0</td>
<td>15.8</td>
<td>40.9</td>
<td>25.3</td>
</tr>
<tr>
<td>500</td>
<td>26.7</td>
<td>21.0</td>
<td>34.7</td>
<td></td>
<td>23.0</td>
</tr>
</tbody>
</table>

Table 6. (Queue). Confidence intervals for the mean and the standard deviation of $T_5$.
Table 7. (Queue). Coverage probabilities for 90% $BC_a$ and percentile confidence intervals.

<table>
<thead>
<tr>
<th>Bootstrap Method</th>
<th>Sample Size 100</th>
<th>Sample Size 500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>$BC_a$</td>
<td>0.8928</td>
<td>0.8917</td>
</tr>
<tr>
<td>Percentile</td>
<td>0.8904</td>
<td>0.8904</td>
</tr>
</tbody>
</table>

Table 8. (Queue). 90% confidence intervals for the stationary distribution of the altered queue.

<table>
<thead>
<tr>
<th>State</th>
<th>Exact</th>
<th>Sample Size</th>
<th>$BC_a$</th>
<th>$BC_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>Lower</td>
<td>Upper</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.318</td>
<td>0.722</td>
<td>0.156</td>
<td>0.393</td>
</tr>
<tr>
<td>1</td>
<td>0.318</td>
<td>0.722</td>
<td>0.156</td>
<td>0.393</td>
</tr>
<tr>
<td>0</td>
<td>0.235</td>
<td>0.222</td>
<td>0.165</td>
<td>0.251</td>
</tr>
<tr>
<td>1</td>
<td>0.235</td>
<td>0.222</td>
<td>0.165</td>
<td>0.251</td>
</tr>
<tr>
<td>2</td>
<td>0.174</td>
<td>0.180</td>
<td>0.171</td>
<td>0.183</td>
</tr>
<tr>
<td>3</td>
<td>0.174</td>
<td>0.180</td>
<td>0.171</td>
<td>0.183</td>
</tr>
<tr>
<td>4</td>
<td>0.127</td>
<td>0.145</td>
<td>0.101</td>
<td>0.182</td>
</tr>
<tr>
<td>5</td>
<td>0.146</td>
<td>0.181</td>
<td>0.093</td>
<td>0.320</td>
</tr>
</tbody>
</table>

21
Fig. 3. (Dementia). Density function estimates for $T$.

Fig. 4. (Dementia). Survival function estimates and confidence bands.
Fig. 5. (Dementia). Hazard rate estimates and confidence bands.

Fig. 6. (Dementia). Relative errors in estimation.
Fig. 7. (Queue). Survival function confidence bands for $T_5$ with $n = 100$.

Fig. 8. (Queue). Survival function confidence bands for $T_5$ with $n = 500$. 
Fig. 9. (Queue). Hazard rate function confidence bands for $T_3$ with $n = 100$.

Fig. 10. (Queue). Relative errors in estimation for $T_3$ with $n = 100$. 