DISSERTATION

BOOTSTRAPPING STOCHASTIC SYSTEMS IN SURVIVAL ANALYSIS

Submitted by

Douglas A. Bronson

Department of Statistics

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2001

UMI Number: 3013826

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

March 29, 2001

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY DOUGLAS A. BRONSON ENTITLED BOOTSTRAPPING STOCHASTIC SYSTEMS IN SURVIVAL ANALYSIS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Committee Member

Committee Member

Committee Member

Adviser

Department Head

ABSTRACT OF DISSERTATION

BOOTSTRAPPING STOCHASTIC SYSTEMS IN SURVIVAL ANALYSIS

The main focus of this paper is the first passage time distribution in a semi-Markov processes from an initial state to an absorbing state. The single bootstrap is implemented using a saddlepoint approximation to determine estimates for the survival and hazard functions of first passage. 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 characterization of the asymptotic hazard rate for survival times is presented and leads to an indirect method for constructing its bootstrap confidence intervals.

The lifetime of a patient in the presence of independent left- and right-censoring is considered from a systems theoretic point-of-view. The systems development introduces semi-Markov flowgraphs to represent the transitions of the patient into and out of censoring states up until the time of death. The empirical flowgraph, determined from the lifetime data subject to censoring, provides an estimate of the population version and is itself a semi-Markov flowgraph. Solving the empirical flowgraph has the effect of removing the censoring risk with both Kaplan-Meier and Turnbull survival estimators resulting. The fact that these estimators are produced highlights the strength of this particular point of view.

Now consider passage through a semi-Markov system in the presence of independent right-censoring. Interest lies in estimating the first passage time distribution with the censoring risk removed. Data is observed from a censored-system, producing an empirical flowgraph. This is an estimate of the population flowgraph and is also a semi-Markov system. The main idea is to amend the empirical flowgraph so that the censoring risk is removed. Since the system is free of censoring, previous saddlepoint and bootstrap methods can be used to compute confidence intervals for the survival and hazard functions of first passage. The methods are illustrated by providing a bootstrap/saddlepoint solution to the classical Fix and Neyman (1951) survival model.

Douglas A. Bronson
Department of Statistics
Colorado State University
Fort Collins, Colorado 80523
Spring 2001

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Professor Ron Butler. His endless energy and inspiration kept me going. I would also like to thank Amy. Without her love and support, this would not have been possible. Finally, I would like to thank my parents. Without them, truely nothing would have been possible.

Contents

1 Bootstrapping Survival Times in Stochastic Systems using Saddlepoint Approximations

1.1 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 generalized 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 μ, the mean of T, using saddlepoint methods has been addressed by Davison and Hinkley (1988) who provide percentile confidence intervals for μ. These percentiles are based upon the distribution of a resampled T^* with moment generating function (mgf) $\hat{M}^m$ where $\hat{M}$ is the empirical mgf

$$\hat{M}(s) = \int \exp(sx)\, d\hat{G}(x). \tag{1.1}$$

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 1.1. 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^m and the resampled T^* has the mgf $\hat{M}^m$ used in Davison 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

$$z(t) = \frac{f(t)}{\bar{F}(t)},$$

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 (1.1), then a doubly resampled T^{**} has the mgf $(\hat{M}^*)^m$ for the system in Figure 1.1. 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 Guttman (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 1.1 has illustrated only the simplest possible $(m+1)$-state system

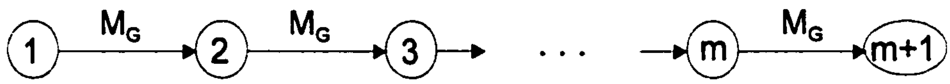

Figure 1.1 Series connection for an iid sum.

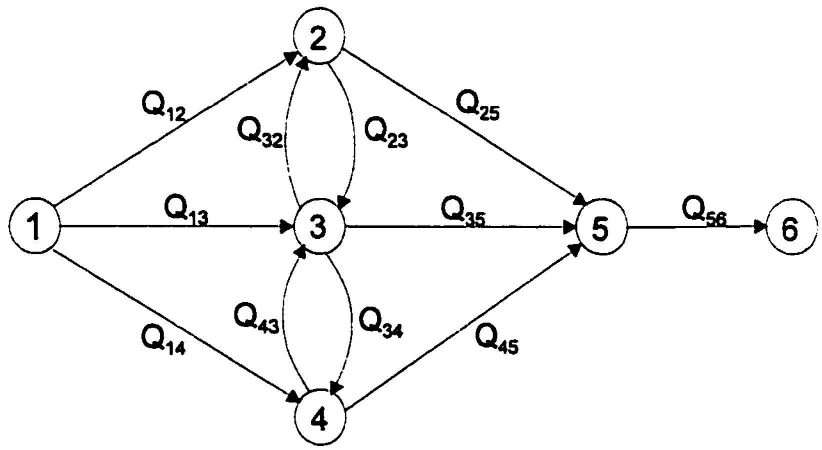

Figure 1.2 Flowgraph showing the degenrative states of dementia.

whereas we 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 1.2 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 $1 \to 6$. The feedback loops $2 \to 3 \to 2$, $3 \to 4 \to 3$, $2 \to 3 \to 4 \to 3 \to 2$, 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 $(0, t_u)$ 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 1.2 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 analyses that concerns passage times to queue lengths 5 and 10 starting from 0.

1.2 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 $\mathcal{P} = (p_{ij})$, 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 $\mathcal{M}(s) = \{M_{ij}(s)\}$, 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 ith 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

$$\mathcal{Q}(s) = \{Q_{ij}(s)\} = \mathcal{P} \odot \mathcal{M}(s) = \{p_{ij} M_{ij}(s)\}$$

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

1.2.1 First Passage Transmittances

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

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

where $f_{1m} = \Pr(T<\infty)$ is the probability of passage and $\mathcal{F}_{1m}(s)$ 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 1.1 *The first passage transmittance from state* 1 *to* $m \neq 1$ *is*

$$f_{1m}\mathcal{F}_{1m}(s) = \frac{(m,1)\text{-cofactor of } I_m - \mathcal{Q}(s)}{(m,m)\text{-cofactor of } I_m - \mathcal{Q}(s)} := \frac{(-1)^{m+1}|\Psi_{m1}(s)|}{|\Psi_{mm}(s)|}. \tag{1.2}$$

where $\Psi_{ij}(s)$ *is the* $(m-1)\times(m-1)$ *submatrix of* $I_m - \mathcal{Q}(s)$ *with its* i^{th} *row and* j^{th} *column removed. The ratio (1.2) is well-defined over a maximal convergence neighborhood 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 $1 \to m$ 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 it is a possible intermediate state during such passage. States 1 and m are designated as relevant if passage $1 \to m$ is possible.

2. The maximal common neighborhood of convergence for the moment generating functions in the first $n-1$ rows of $\mathcal{Q}(s)$ is an open neighborhood of 0.

Proof: See Butler (1997*a*, *b*). ■

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

Theorem 1.2 *The first return transmittance for state* 1 *is*

$$f_{11}\mathcal{F}_{11}(s) = 1 - \frac{|I_m - \mathcal{Q}(s)|}{(1,1)\text{-cofactor of } I_m - \mathcal{Q}(s)} := 1 - \frac{|I_m - \mathcal{Q}(s)|}{|\Psi_{11}(s)|}. \tag{1.3}$$

The ratio (1.3) is well-defined over an maximal convergence neighborhood 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 $1 \to 1$.

2. The maximal common neighborhood of convergence for the moment generating functions in $\mathcal{Q}(s)$ is an open neighborhood of 0.

Proof: See Butler (1997a, b). ■

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 equivalence to (1.2) and (1.3) in Butler (1997a).

When heavy-tailed distributions such as the Pareto are used as one step holding times, then the convergence regions for $\mathcal{F}_{1m}(s)$ and $\mathcal{F}_{11}(s)$ are no larger than that of the Pareto and therefore have form $(-\infty, 0]$. Another way of treating these issues is to redefine a transmittance as a probability × a characteristic function. This leads to the following result whose proof is the same as the theorems above.

Corollary 1.1 *If characteristic functions are used instead of moment generating functions when defining transmittances, then Theorems 1.1 and 1.2 provide the first passage transmittances based upon characteristic functions for* T *and defined over* *all real s. Now the conditions for this require only that the state space S consists of exactly those states relevant to passage.*

The corollary really offers no more generality than the theorems stated using mgfs. We shall therefore continue to work only with mgfs.

1.2.2 Saddlepoint Approximations

The cofactor rules of Theorems 1.1 and 1.2 lead to explicit saddlepoint expressions that are convenient for use with the Lugannani and Rice (1980) and density approximations as discussed by Daniels (1987). Let $\mathcal{K}(s) = \ln \mathcal{F}_{1m}(s)$ be the cumulant generating function (cgf) of T. The survival function approximation is

$$\bar{F}_1(t) = 1 - \Phi(\hat{w}) - \phi(\hat{w})\left(\frac{1}{\hat{w}} - \frac{1}{\hat{u}}\right) \qquad t \neq E(T) = \mathcal{K}'(0) \tag{1.4}$$

where Φ and ϕ are the standard normal CDF and density, $\hat{w} = \hat{w}(\hat{s})$ and $\hat{u} = \hat{u}(\hat{s})$ depend on $\hat{s}$ according to

$$\hat{w} = \operatorname{sgn}(\hat{s})\sqrt{2\{\hat{s}t - \mathcal{K}(\hat{s})\}} \qquad \text{and} \qquad \hat{u} = \hat{s}\sqrt{\mathcal{K}''(\hat{s})}, \tag{1.5}$$

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

$$f_1(t) = \frac{1}{\sqrt{2\pi\mathcal{K}''(\hat{s})}} \exp\left(-\frac{1}{2}\hat{w}^2\right), \tag{1.6}$$

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

$$z_1(t) = \frac{f_1(t)}{\bar{F}_1(t)} \qquad \text{and} \qquad z_2(t) = \frac{f_1(t)}{\bar{F}_1(t)\int_0^\infty f_1(u)\,du}. \tag{1.7}$$

The saddlepoint equation $\mathcal{K}'(\hat{s}) = t$ has a simple form since $\mathcal{F}_{1m}$ is a ratio of the cofactor determinants in (1.2) or (1.3), . For passage from $1 \to m \neq 1$, then

$$\mathcal{K}' = \operatorname{tr}\left(\Psi_{m1}^{-1}\Psi_{m1}' - \Psi_{mm}^{-1}\Psi_{mm}'\right) \tag{1.8}$$

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

$$\mathcal{K}'' = \operatorname{tr}\left\{\Psi_{m1}^{-1}\Psi''_{m1} - \left(\Psi_{m1}^{-1}\Psi'_{m1}\right)^2 - \Psi_{mm}^{-1}\Psi''_{mm} + \left(\Psi_{mm}^{-1}\Psi'_{mm}\right)^2\right\}. \tag{1.9}$$

The simplicity of these two expressions is only possible because of the simplicity for $\mathcal{K}$ as expressed through the cofactor rules (1.2) and (1.3). This structure has allowed the authors to successfully perform saddlepoint computations with general systems of size $m = 250$ without any difficulties. Such computations are not feasible using the $\mathcal{F}_{1m}$ expressions of Mason (1953, 1956) or Pyke (1961) and Howard (1964. 1971).

1.3 Nonparametric Bootstrap

Data for inference about a survival time, which is a first passage from $1 \to m$ 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 $1 \to m$ and follows a common transmittance matrix $\mathcal{Q}$.

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

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

Matrix $\hat{\mathcal{P}} = (\hat{p}_{ij}) = (n_{ij}/n_{i\cdot})$ 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 $i \to j$ is not possible when $p_{ij} = 0$ and in such cases $\hat{p}_{ij} = 0$. We shall also assume sufficient data that $\hat{p}_{ij} > 0$ whenever $p_{ij} > 0$; 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

$$\operatorname{sgn}(\hat{p}_{ij}) = \operatorname{sgn}(p_{ij}) \tag{1.10}$$

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 (1.2) and (1.3) (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 1.3 *A resampled T^* has first passage transmittance $\hat{f}_{1m}\hat{\mathcal{F}}_{1m}(s)$ obtained by substituting $\hat{Q}$ into (1.2) 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 $\hat{\mathcal{F}}_{1m}(s)$.*

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 $\hat{p}^{*}_{ij} = n^{*}_{ij}/n_{i\cdot}$ where the vector

$$(n^{*}_{i1}, \ldots, n^{*}_{im}) \mid \{n_{ij}\} \sim \text{Multinomial } (n_{i\cdot}; \hat{p}_{i1}, \ldots, \hat{p}_{im}). \qquad (1.11)$$

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 $n^{*}_{i\cdot} = n_{i\cdot}$. 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 1.4 *A resampled T^{**} has first passage transmittance $\hat{f}^{*}_{1m}\hat{\mathcal{F}}^{*}_{1m}(s)$ obtained by substituting $\hat{Q}^{*}$ into (1.2) 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 $\hat{\mathcal{F}}_{1m}(s)$.*

The resampling just described has not taken into account the random variation in $\{n_{i\cdot}\}$ 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 $\{n^{*}_{i\cdot}\}$ as a replacement for the fixed values of $\{n_{i\cdot}\}$ in (1.11). Such a resampling of the patients can be used to initiate each of the $B = 999$ resamples in (1.11) and accounts for the random variation in $\{n_{i\cdot}\}$. We refer to this method as having resampled row totals (rrt).

The $B = 999$ 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 $\hat{\mathcal{M}}(s)$ representing an independent sample. The skewness estimates for $\{a_i\}$ with multisamples (Davison and Hinkley, 1997, (5.28)) 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.

1.3.1 Dementia Example

In the flowgraph of Figure 1.2, 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

$$\mathcal{Q}(s) = \begin{pmatrix} 0 & 0.6\,\mathrm{g}(2,2) & 0.25\,\mathrm{g}(4,8) & 0.15\,\mathrm{g}(9,27) & 0 & 0 \\ 0 & 0 & 0.85\,\mathrm{r}(\frac{1}{2}\sqrt{\pi}) & 0 & 0.15\,\mathrm{ig}(1,\frac{1}{2}) & 0 \\ 0 & 0.7\,\mathrm{r}(\sqrt{\pi}) & 0 & 0.2\,\mathrm{r}(\sqrt{\pi}) & 0.1\,\mathrm{ig}(\frac{1}{2}\sqrt{2},\frac{1}{4}\sqrt{2}) & 0 \\ 0 & 0 & 0.7\,\mathrm{r}(\frac{3}{2}\sqrt{\pi}) & 0 & 0.3\,\mathrm{ig}(\sqrt{2},\sqrt{2}) & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathrm{ig}(1,1) \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{1.12}$$

where $\mathrm{g}(a,b)$ and $\mathrm{ig}(a,b)$ are gamma and inverse Gaussian mgfs with mean a and variance b, and $\mathrm{r}(a)$ 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 $\hat{\mathcal{Q}}(s)$. No adjustment to $\hat{P}$ for irrelevant states was necessary since our data satisfied (1.10). If T is the passage time from $1 \to 6$, neither its density nor survival function can be exactly determined from (1.12), however saddlepoint versions may be computed so that "exact" comparison may be made with respect to these saddlepoint quantities.

Predictive Inference Table 1.1 shows the accuracy achieved in using saddlepoint approximations (frt and rrt) to determine estimates and 90% BC_a confidence intervals for various right-tail percentiles of T. Right-tail α levels are listed in the first column and their associated "exact" percentiles, as determined from saddlepoint inversion based upon (1.12), are listed as "Exact". Saddlepoint inversion of $\hat{\mathcal{Q}}(s)$ provides the percentile point estimates listed as SA. The simulation estimators, denoted as Sim, are determined by making 2×10^5 passes through system $\hat{\mathcal{Q}}(s)$ and finding the appropriate empirical quantiles; their values agree closely with the SA estimates. Both of the outer layer resampling schemes frt and rrt are used along with saddlepoint approximation at the inner level to determine 90% BC_a confidence intervals for the various percentiles; they show little difference. This suggests that the resampling scheme with fixed row totals may be adequate and that the more elaborate rrt resampling is unnecessary even with the relatively small amount of data associated with 25 patient histories. In addition, little difference between frt and rrt resampling was seen in all the remaining computations for this example; we therefore show only those using frt resampling. Direct double bootstrap frt resampling leads to the BC_a intervals listed as Sim. They show a striking degree of similarity to the frt and rrt bounds.

Table 1.1 Estimates and 90% BC_a confidence intervals for various percentiles of T at the right-tail α-levels listed in the first column.

Right Perc.	"Exact"	Estimate		BC_a Lower			BC_a Upper		
		SA	Sim	frt	rrt	Sim	frt	rrt	Sim
0.50	11.5	12.3	12.6	9.7	9.8	9.9	15.9	15.9	15.1
0.25	19.0	19.9	20.1	15.7	15.9	15.1	26.9	26.8	26.8
0.10	28.4	29.6	29.7	23.2	23.5	23.5	40.6	41.6	41.0
0.05	35.4	36.9	37.0	28.7	29.0	29.0	51.0	52.6	50.5
0.01	51.7	53.8	53.7	41.4	41.8	41.7	75.4	78.2	74.5

Figures 1.3 and 1.4 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 2×10^5 values of T^* (and using kernel estimation for the density) are almost graphically indistinguishable from their saddlepoint counterparts. The long dashes in Figure 1.4 enclose 90% BC_a confidence bands computed on a grid of 201 time points. "Exact" hazard rate $z_2(t)$ for T is shown in Figure 1.5 along with its saddlepoint estimate from inverting $\hat{Q}(s)$, 90% BC_a confidence bands, and a simulation estimate using kernel density estimation. Relative errors in density and survival estimation, as compares with the "exact' versions, along with errors using simulation are shown in Figure 1.6. The proposed estimates using saddlepoint inversions are smoother and more consistently accurate than simulation.

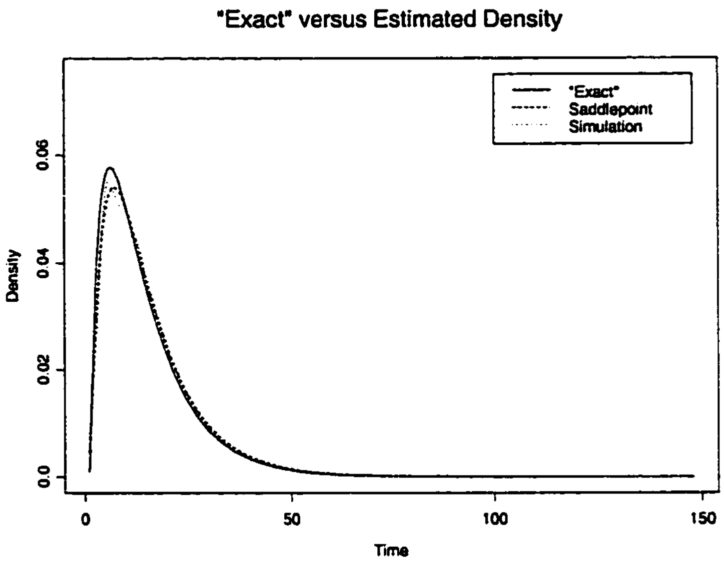

Figure 1.3 Density function estimates for T.

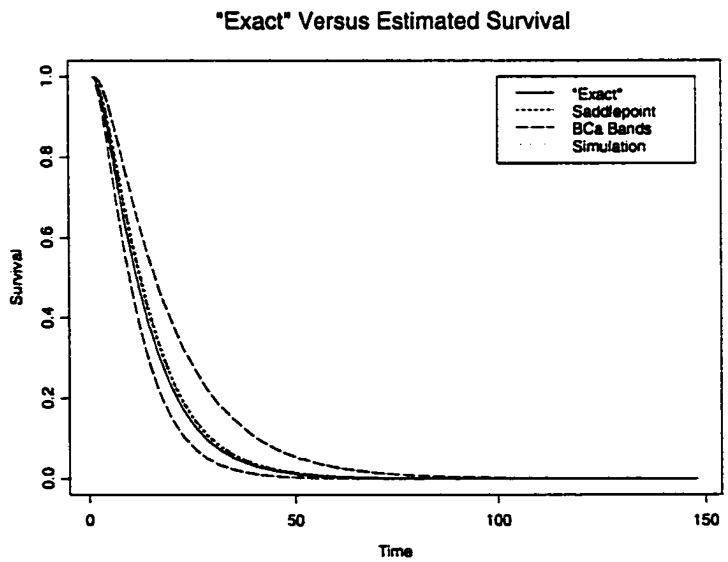

Figure 1.4 Survival function estimates and confidence bands.

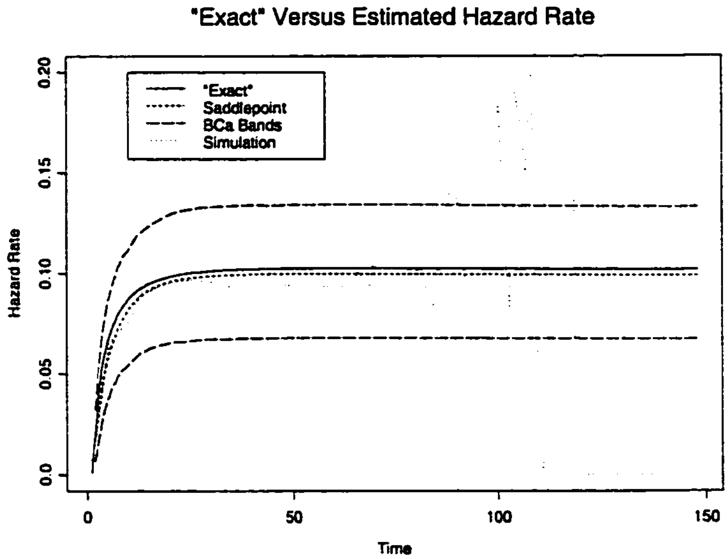

Figure 1.5 Hazard rate estimates and confidence bands.

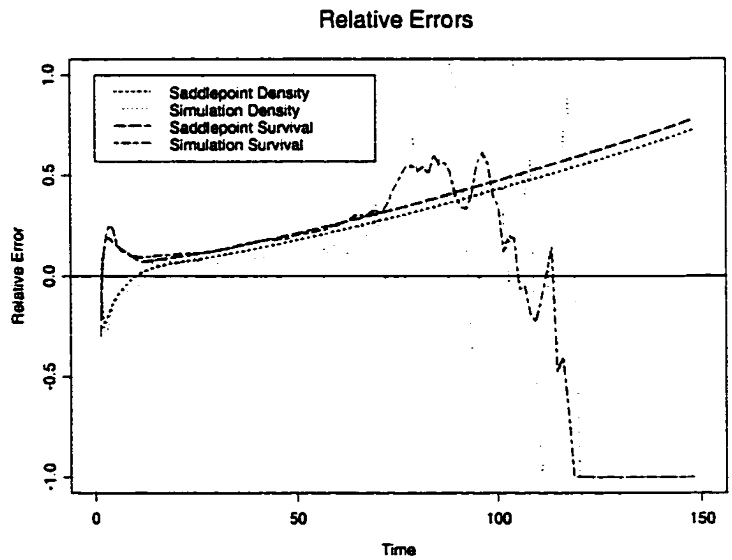

Figure 1.6 Relative errors in estimation.

Parametric Inference The mean and standard deviation of T are easily estimated using the cofactor rule in Theorem 1.1. This involves computing $\hat{\mathcal{F}}'_{1m}(0)$ and $\hat{\mathcal{F}}''_{1m}(0)$, the first two derivatives at zero. No root-finding is necessary in such calculation and the BC_a confidence intervals may be computed in a couple seconds using the first two derivatives of $\hat{\mathcal{F}}^{\bullet}_{1m}(s)$ at $s = 0$. Table 1.2 provides point estimates and 90% BC_a confidence bands listed as the Cofactor Rule method.

The iid method in Table 1.2 estimates these quantities using only the 25 patient survival times and assumes an iid nonparametric model. Unlike the Cofactor Rule method, this method does not assume or make use of the information in the data about the semi-Markov structure.

Table 1.2 Confidence intervals for the mean and the standard deviation of T.

	Mean				Standard Deviation			
Method	Exact	Est.	BC_a Lower	BC_a Upper	Exact	Est.	BC_a Lower	BC_a Upper
Cofactor Rule	14.4	15.3	12.3	20.4	10.7	11.1	8.5	16.0
iid model		15.3	11.9	19.1		10.6	8.1	14.2

The benefit of using the semi-Markov structure is revealed from the coverage accuracy of the two confidence methods shown in Table 1.3. The top group provides empirical coverages of 90% BC_a and percentile intervals with both methods based upon 10,000 repetitions and using 200 randomly chosen jackknife eliminations in the acceleration estimation. The bottom group provides the same information but using 1000 randomly chosen jackknife eliminations. A comparison of top versus bottom groups shows that the 200 eliminations in acceleration estimation provide very accurate coverage.

Table 1.3 Coverage Probabilities for the 90% confidence intervals of mean and standard deviation of T.

Number of Jackknife Eliminations	Bootstrap Method	Cofactor Rule		iid Model	
		Mean	Std. Dev.	Mean	Std. Dev.
200	BC_a	0.9058	0.9085	0.8626	0.7489
	Percentile	0.9024	0.9009	0.8584	0.7039
1000	BC_a	0.9076	0.9107	0.8727	0.7519
	Percentile	0.9049	0.9041	0.8683	0.7045

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. This small sample phenomenon for the latter method has also been recently noted in Polansky (1999). He explains that the undercoverage derives from the discrepancy between the bounded support of the bootstrap distribution as compared to the unbounded tail of the true passage distribution.

Undercoverage, when using the Cofactor Rule method with a feedback system such as in Figure 1.2, 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 the moments. The reason for this is that the presence of a feedback loop extends the support of T^* and T^{**} to ∞ so the right-tail of support for T^* and T^{**} matches that of T. The situation differs 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 susceptable to undercoverage.

1.4 Semi-parametric Bootstrap

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

1.4.1 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 $\mathcal{Q}$ 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 (λ) fixing time for the single server. With $m = 5$, this transmittance matrix is

$$\mathcal{Q}(s) = \begin{pmatrix} 0 & \mathcal{U}_0(s) & 0 & 0 & 0 & 0 \\ 0 & \mathcal{Q}_{11}(s) & \mathcal{U}_0(s-\lambda) & 0 & 0 & 0 \\ 0 & \mathcal{Q}_{21}(s) & \mathcal{U}_1(s-\lambda) & \mathcal{U}_0(s-\lambda) & 0 & 0 \\ 0 & \mathcal{Q}_{31}(s) & \mathcal{U}_2(s-\lambda) & \mathcal{U}_1(s-\lambda) & \mathcal{U}_0(s-\lambda) & 0 \\ 0 & \mathcal{Q}_{41}(s) & \mathcal{U}_3(s-\lambda) & \mathcal{U}_2(s-\lambda) & \mathcal{U}_1(s-\lambda) & \mathcal{U}_0(s-\lambda) \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where the entries are generally

$$\begin{aligned} \mathcal{U}_i(s) &= \tfrac{\lambda^i}{i!} \int_0^\infty w^i e^{sw} dG(w) & i &= 0, \ldots, m-2 \\ \mathcal{Q}_{i1}(s) &= \mathcal{U}_0(s) - \textstyle\sum_{k=0}^{i-1} \mathcal{U}_k(s-\lambda) & i &= 1, \ldots, m-1\,. \end{aligned} \tag{1.13}$$

To understand these entries consider, for example, entry $(4, 2)$. 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 (λW) probability for 3 times e^{sW} or $\mathcal{U}_3(s - \lambda)$.

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 ith interarrival time and $x_i|w_i \sim$ Poisson (λw_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 $\hat{\lambda} = x./w.$, the MLE as a ratio of sums. The distribution T_m^* has the transmittance $\mathcal{Q}(\hat{G}, \hat{\lambda})$.

Theorem 1.5 *The transmittance matrix $\mathcal{Q}(\hat{G}, \hat{\lambda})$ characterizes a semi-Markov process whose first passage time to length m has the distribution T_m^*.*

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

Predictive Inference As an example, take G as Gamma $(2, 2)$ with mean 1 and variance $\frac{1}{2}$. Suppose a 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 1.7 as a solid line along with its estimate $\bar{F}_1(t)$ (short dashed line) based upon saddlepoint inversion of $\mathcal{Q}(\hat{G}, \hat{\lambda})$. An empirical estimate (dotted line), based upon simulating 2×10^5 values of T_5^*, is graphically indistinguishable from $\bar{F}_1(t)$. The long dashes enclose 90% BC_a confidence bands computed on a grid of 201 time points. The same features but with $n = 500$ are shown in Figure 1.8 where the greater informativeness of the data has narrowed the BC_a bands. Returning to the $n = 100$ setting, the true hazard rate of T_5 is shown in Figure 1.9 along with its saddlepoint estimate $z_2(t)$ from inverting $\mathcal{Q}(\hat{G}, \hat{\lambda})$, 90% BC_a confidence bands, and a simulation estimate. Relative errors in density and survival estimation are shown in Figure 1.10. The saddlepoint estimates and confidence bands used 1/35 of the c.p.u. 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. In the Appendix, it is shown that

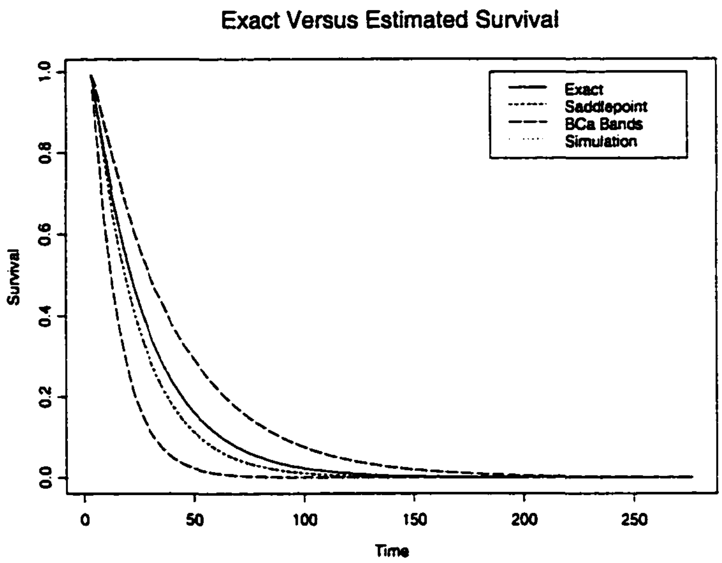

Figure 1.7 Survival function confidence bands for T_5 with $n = 100$.

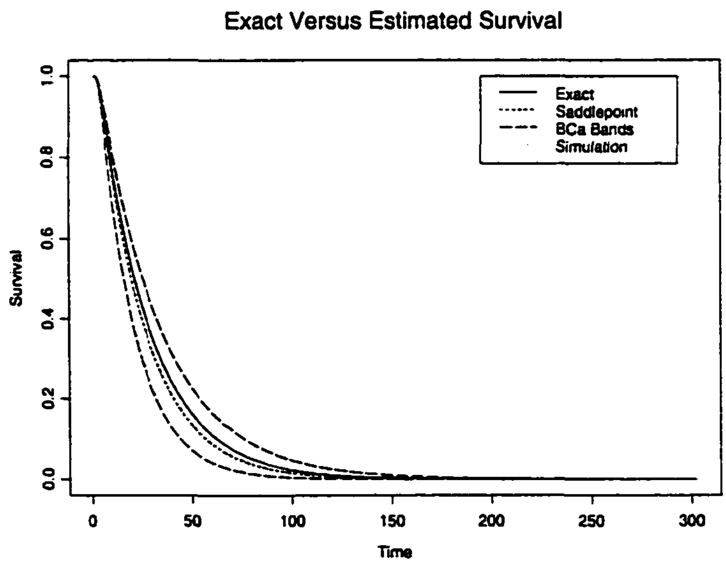

Figure 1.8 Survival function confidence bands for T_5 with $n = 500$.

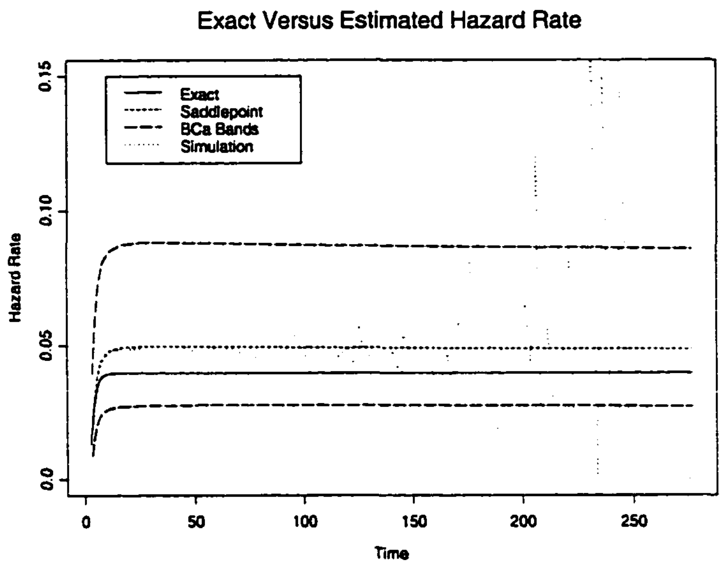

Figure 1.9 Hazard rate confidence bands for T_5 with $n = 100$.

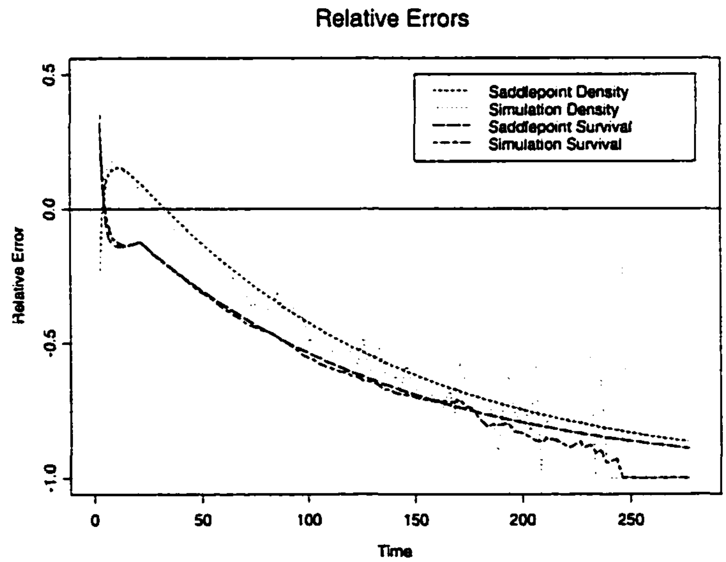

Figure 1.10 Relative errors in estimation for T_5 with $n = 100$.

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 1.4 shows the accuracy achieved in using saddlepoint approximations (SA) to determine estimates and 90% BC_a confidence intervals for various right-tail percentiles of T_5. For the various right-tail α levels, their associated exact percentiles are given as Exact. The SA entry just below is the saddlepoint approximation to Exact using $\mathcal{Q}(G, \lambda)$, the true cgf 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.

Table 1.4 Estimates and 90% BC_a confidence intervals for various percentiles of T_5 using the right-tail α-levels listed in the first column.

Right	Exact	Sample	Estimate		BC_a Lwr		BC_a Upr		Guar. Tol.	
Perc.	SA	Size	SA	Sim	SA	Sim	SA	Sim	SA	Sim
0.50	21.3	100	18.6	18.6	12.2	12.6	29.7	29.5	27.5	27.5
	21.2	500	19.6	19.6	15.8	15.3	25.3	24.3	24.1	24.1
0.25	38.8	100	33.1	33.0	20.5	21.0	55.6	55.4	50.1	50.1
	38.8	500	35.6	35.4	27.9	26.8	46.7	44.6	46.0	46.0
0.10	62.0	100	52.3	52.1	31.6	32.2	89.2	89.7	80.3	77.8
	62.0	500	56.6	56.3	43.8	41.8	75.7	71.7	70.7	70.7
0.05	79.5	100	66.8	66.4	40.0	40.9	115.5	115.0	105.4	100.4
	79.6	500	72.6	72.3	55.9	53.5	97.3	92.4	92.8	92.8
0.01	120.1	100	100.5	99.7	59.2	60.7	173.6	177.2	158.2	150.7
	120.4	500	109.5	110.0	83.9	80.2	147.6	143.0	139.6	139.6

Tolerance intervals of coverage $1-\alpha$ with 90% guarantee are displayed in the final column again based upon saddlepoint and simulation methods. For example with

$1 - \alpha = 0.75$ and $n = 500$, 46.0 is found by both methods as the smallest $c > 0$ for which

$$\Pr\{\Pr(T_5^* < c) \leq 0.75\} \geq 0.9,$$

e.g. there is 90% guarantee of 75% coverage. For fixed c, the outer probability of guarantee is saddlepoint approximated (SA) by (i) first using $B = 999$ saddlepoint inversions of $\mathcal{Q}(\hat{G}^*, \hat{\lambda}^*)$ to determine B CDF approximations evaluated at c; (ii) then counting the proportion of these smaller than 0.75. Sim entries were determined by implementing the double bootstrap: (i) for each of the B resampled systems $\mathcal{Q}(\hat{G}^*, \hat{\lambda}^*)$ with passage time T_5^*, 2×10^5 generations of T_5^{**} were simulated to estimate $\Pr(T_5^* < c)$; (ii) the proportion of $\Pr(T_5^* < c)$ estimates smaller than 0.75 approximates the outer guarantee probability and has been used to determine the Sim entries. Again the two methods show striking agreement.

Table 1.5 is a repetition of Table 4 but requiring passage to queue length 10 so that percentiles of T_{10} are estimated. Considerably more state transitions are required to reach queue length 10 and any inaccuracies in the estimation of G and λ are further accentuated leading to substantially wider confidence bands for the percentiles. For this reason sample sizes of $n = 500$ and 1000 were used. Only saddlepoint approximations are displayed since the c.p.u. time required to use the double bootstrap was prohibitive.

Table 1.6 provides c.p.u. times needed to determine estimates using the single bootstrap (Est.) and confidence bands using the double bootstrap (CIs) for the various queue lengths and sample sizes. The starred simulation times for length 10 were estimated using partial runs to avoid excessive run times.

Table 1.5 Estimates and 90% BC_a confidence intervals for various percentiles of T_{10} along with tolerance intervals.

Right Perc.	Exact SA	Sample Size	Estimate	BC_a Lower	BC_a Upper	Guar. Tol.
0.50	163.0	500	131.9	78.4	250.8	226.0
	162.3	1000	149.0	103.0	235.4	208.0
0.25	312.7	500	251.2	143.2	489.4	431.7
	312.6	1000	286.5	193.6	460.1	423.3
0.10	510.6	500	408.7	229.1	804.3	699.0
	511.4	1000	467.8	314.6	756.9	683.8
0.05	660.3	500	527.9	293.8	1042.6	904.7
	661.7	1000	605.2	405.3	981.6	873.3
0.01	1007.9	500	804.2	444.8	1596.0	1357.2
	1010.2	1000	923.5	615.9	1502.7	1323.4

Table 1.6 CPU times for estimating the density and survival function of T.

Queue Length	Sample Size	Est. (*secs*) SA	Sim	CIs (*hrs*) SA	Sim
5	100	0.9	29	0.29	9.0
5	500	2.2	32	0.94	11.4
10	500	3.3	159	1.4	57*
10	1000	5.1	167	3.1	74*

Parametric Inference Bootstrap confidence bounds of the mean and standard deviation for the distributions of passage times T_5 and T_{10} are given in Table 1.7.

Table 1.7 Confidence intervals for the mean and the standard deviation of the first passage time distributions.

		Mean				Std Dev			
Queue Length	Sample Size	Exact	Est.	BC_a Lwr	BC_a Upr	Exact	Est.	BC_a Lwr	BC_a Upr
5	100	29.0	25.0	15.8	40.9	25.3	21.0	12.0	36.1
	500		26.7	21.0	34.7		23.0	17.4	30.8
10	500	229.2	184.9	107.1	357.1	216.1	171.4	93.2	342.5
	1000		210.1	143.5	333.0		197.4	130.7	320.7

To assess coverage accuracy for BC_a and percentile intervals determined by saddlepoint inversion, the true coverage probabilities have been estimated for the confidence intervals described in Table 1.7. Ten thousand repetitions of these BC_a intervals along with percentile intervals were computed and the coverage frequencies are listed in Table 1.8 for $m = 5, 10$ and with sample sizes $n = 100, 500$, and 1000. The target coverage of 90% is accurately achieved in all instances. For this example, the simpler percentile intervals appear quite accurate in coverage.

Table 1.8 Coverage Probabilities for 90% BC_a and percentile confidence intervals.

Passage Length	Bootstrap Method	$n = 100$ Mean	$n = 100$ Std Dev	$n = 500$ Mean	$n = 500$ Std Dev	$n = 1000$ Mean	$n = 1000$ Std Dev
5	BC_a	0.8928	0.8917	0.8981	0.8987		
	Percentile	0.8904	0.8904	0.9013	0.9018		
10	BC_a			0.9005	.9004	0.9017	0.9016
	Percentile			0.9024	0.9028	0.9031	0.9026

To assess the difficulty in reaching queue lengths 5 and 10, we consider the distributions that count the number of state transitions needed to achieve such passages.

The means of these distributions are 29.0 and 229.0 respectively with standard deviations 23.6 and 212.2. These computations are determined for system $\mathcal{Q}(s)$ by using the system transition probabilities $\mathcal{Q}(0)$ in conjunction with fixed holding times of 1. This underlying Markov chain has transmittance $\mathcal{Q}(0)e^s$ which may be used with Theorem 1.1 to determine the first passage transmittance for the number of state transitions needed for passage of the original system $\mathcal{Q}(s)$. The first two derivatives at $s = 0$ give the mean and variance.

The stationary distribution for a semi-Markov process has a mass function that can also be estimated. Since our queue of length 5 is not stationary, we can make it so by changing the queue discipline so that arrivals are turned away while in state 5. This makes the queue a stationary semi-Markov process whose holding time in state 5, before passing to state 4, is Exponential (λ). The stationary distribution depends only upon G and λ (see Ross, 1983 §4.8) and 90% BC_a confidence intervals on each of its probabilities are displayed in Table 1.9.

Table 1.9 Ninety-percent confidence intervals for the stationary distribution of the altered queue.

State	Exact	Sample Size	Estimate	BC_a Lower	BC_a Upper
0	0	100	0	0	0
		500	0	0	0
1	0.318	100	0.272	0.156	0.393
		500	0.296	0.237	0.356
2	0.235	100	0.222	0.165	0.251
		500	0.229	0.207	0.245
3	0.174	100	0.180	0.171	0.183
		500	0.177	0.168	0.180
4	0.127	100	0.145	0.101	0.182
		500	0.135	0.114	0.155
5	0.146	100	0.181	0.093	0.320
		500	0.163	0.117	0.220

1.5 Asymptotic Hazard Rates

Theorem 1.6 presents a new characterization of the asymptotic average failure rate that has importance apart from its application here to passage time distribution for semi-Markov systems. For proofs of the theorems in this section, see Butler and Bronson (2000).

1.5.1 Characterization

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

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

The limiting average hazard is now characterized in terms of the associated mgf.

Theorem 1.6 *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*

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

The limiting average hazard exists in many instances so that the liminf in Theorem 1.6 is really a limit. Special cases include those distributions with monotone average hazard rates. Such monotonicity is an aging or antiaging property of the distribution according to whether it is increasing or decreasing. Such monotonicity settings commonly occur in the reliability modelling of coherent systems whose individual components have exponential failure times as discussed in Barlow and Proschan (1975, §4.2). Apart from these considerations, such as in considering passage times in semi-Markov systems, simple conditions have been given in Butler and Robinson (2000) to assure that both the hazard rate and the average hazard for T have limits. These conditions hold quite generally in semi-Markov applications, and are satisfied in both of our examples. We shall therefore simply suppose the conditions under which these limits exist.

Corollary 1.2 *Suppose passage time T has a continuous density for sufficiently large t and* $\lim_{t\to\infty} z(t)$ *exists. Then*

$$\lim_{t\to\infty} \frac{\Lambda(t)}{t} = \lim_{t\to\infty} z(t). \tag{1.14}$$

Example. The inverse Gaussian distribution with mean 1 and variance 1 has a moment generating function convergent on $\left(-\infty, \frac{1}{2}\right]$. Chhikara and Folks (1977, p.155) show that the limiting hazard rate is $\frac{1}{2}$. ■

1.5.2 Uniformity of Saddlepoint Approximations

The saddlepoint estimate of asymptotic failure rate based upon the unnormalized z_1 converges quite generally under very weak conditions.

Theorem 1.7 *Assume that the interior of the convergence strip for the moment generating function of T is $(-\infty, b)$ for $b > 0$. In saddlepoint approximation, suppose also that $\hat{w} \to \infty$ and $\liminf_{t\to\infty} \hat{u} > 0$ as $t \to \infty$. Then*

$$\lim_{t\to\infty} z_1(t) = b \neq \lim_{t\to\infty} z_2(t). \tag{1.15}$$

If furthermore $\lim_{t\to\infty} z(t) = b$, *then*

$$\lim_{t\to\infty} \frac{f_1(t)}{f(t)} = \lim_{t\to\infty} \frac{\bar{F}_1(t)}{\bar{F}(t)} \tag{1.16}$$

providing that one of the limits is finite; thus the saddlepoint density and survival distribution approximations have the same limiting relative error in the extreme tail.

According to this theorem, equality for the limiting error of the saddlepoint density and survival distributions is characteristic of the more regular settings in which saddlepoint methods are used. Such settings satisfy the conditions of the theorem and it is often simple to show that the saddlepoint density has a finite limiting relative error; therefore the limiting relative error for the Lugannani and Rice approximation is the same.

Example. Consider the inverse Gaussian example. Then $\hat{f}_1(t) \equiv f(t)$ (Daniels, 1980) so its relative error is 0; the limiting relative error of $\hat{\bar{F}}_1(t)$ must therefore be 0. This fact has been show in Jensen (1995, §2.4) through direct expansion. ■

In practice when considering moderate values of t, $z(t)$ is usually more accurately estimated using $z_2(t)$, the normalized approximation. For large t however, it is biased and $z_1(t)$ is consistent.

1.5.3 Bootstrap Estimation

The asymptotic hazard rate for passage time T is the smallest positive real pole of $\mathcal{F}_{1m}$ and may be estimated by finding the comparable root of $\hat{\mathcal{F}}_{1m}$ instead. Such poles, 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 $\hat{\mathcal{F}}_{1m}$. The 90% BC_a confidence band is $(0.0654, \ 0.128)$.

For the queue example, Table 1.10 provides the true asymptotic failure rates, their estimates, and 90% BC_a confidence intervals using the various queue lengths and sample sizes. The coverage accuracies of the BC_a and percentile (not shown) confidence intervals for queue length 5 with sample size 100 were assessed by computing 2500 such intervals. Empirical coverages of 0.892 and 0.885 respectively were observed demonstrating quite accurate coverage for both.

Table 1.10 Estimated and exact asymptotic hazard rates for T_5 and T_{10}.

Queue Length	Exact	Sample Size	Estimate	BC_a Lower	BC_a Upper
5	0.03957	100	0.04790	0.02702	0.08397
		500	0.04370	0.03198	0.05754
10	0.00463	500	0.00584	0.00292	0.01076
		1000	0.00507	0.00309	0.00766

2 Self-Consistent Estimates in Survival Analysis via a Systems Theoretic Approach

2.1 Introduction

The aim in survival analysis is to use censored survival data to determine the lifetime distribution F^0 of a patient without the risk of censoring. We introduce a systems theoretic point of view which accomplishes this and accommodates for the possibility of left- and right-censoring. Semi-Markov flowgraphs are constructed to represent the transitions of a random patient into and out of censoring states up until the time of death. For each semi-Markov population flowgraph, we also present its empirical flowgraph counterpart which is determined from the lifetime data subject to censoring. Each empirical flowgraph provides an estimate of its population version and is itself a semi-Markov flowgraph. Within the context of the constructed semi-Markov systems. the distribution F^0, or more precisely its Laplace-Stiltjes transform, is the "solution" to the population flowgraph. The same solution to its empirical counterpart forms its empirical estimate $\hat{F}^0$.

In the case of independent right-censoring, solutions to empirical flowgraphs yield a self-consistency relationship that was originally introduced by Efron (1967) from a different perspective. Its unique solution is the Kaplan-Meier (1958) estimator and is also the solution to the empirical flowgraph model. Thus the systems theoretic perspective provides further motivation for the validity of the Kaplan-Meier estimator in dealing with independent right-censoring. The Berliner and Hill (1988) estimator is also shown to be a solution to a slightly different empirical flowgraph.

A comparable development is presented for double-censoring where flowgraphs are introduced to accommodate both left and right censoring. Solutions to these empirical flowgraphs turn out to be the solutions to some self-consistency equations introduced by Tsai and Crowley (1985) which were used to represent the Turnbull (1974) estimators. These self-consistency relationships do not admit unique solutions and a complete characterization and listing of the possible solutions is presented in Butler and Bronson (2001); see also Mykland and Ren (1996). When the solutions are restricted to placing mass only at death times, then there is a unique solution and it turns out to be a particular form of a Turnbull-type estimator.

The main purpose of this paper is to provide a new unified framework within which one of the most basic problems of survival analysis, the removal of independent censoring risk factors in determining survival distributions, may be explained. Unity is provided through the use of a systems theoretic framework to model the semi-Markov processes that underlie the censoring mechanisms. This framework has been the dominate tool used to analyze linear systems in engineering and an extensive discussion of this point of view in stochastic modeling is given in Butler (2000). Systems theory has also been the dominant paradigm in other scientific disciplines and this paper also sets the stage for providing the same perspective in survival analysis. The strength of the approach can be measured in terms of the empirical estimates it provides: two well established estimators, the Kaplan-Meier and Turnbull estimators, and prospects for a considerable amount of further development.

Section 2.2 considers the independent right-censoring model proceeding from development of the population flowgraph and its solution to its empirical counterpart and solution. Section 2.3 extends this to the independent double-censoring model. Some overlooked properties of self-consistent Turnbull-type estimators are also discussed. In particular we consider its equivariance properties to the reflective group in which the problem is considered in both positive and negative time (so the roles of left and right censoring are reversed), and the indeterminacy in its tails.

2.2 Right-Censored Lifetimes

Let random variable X^0 be the death time for a patient with cumulative distribution function (c.d.f.) $F^0(x)$ and associated moment generating function (m.g.f.) $M^0(s)$. Independent of X^0, suppose the patient may be censored at random time Z^0 with c.d.f. $H^0(z)$. Since only the smaller of the two variables is observed, it is convenient to define the competitive variables

$$X \stackrel{d}{=} X^0 \mid \{X^0 < Z^0\} \sim F(x), \quad \text{with m.g.f. } M(s),$$
$$Z \stackrel{d}{=} Z^0 \mid \{Z^0 < X^0\} \sim H(z).$$

The support of all random variables is assumed to be $(0, \infty)$.

The lifetime of a random patient subject to right-censoring is shown in the semi-Markov flowgraph of Figure 2.1. Nodes 1 and D represent the alive and dead states respectively. States $\{\mathrm{R}_z : z > 0\}$ represent a continuum of censored states for which z is the time of censoring. Upon censoring at time z, the patient's subsequent time of death depends on z as indicated in the transition from $\mathrm{R}_z \to \mathrm{D}$. The flowgraph shows the collection of mutually exclusive pathways that the patient can follow: either direct transition $1 \to \mathrm{D}$ without censoring, or along the continuum of pathways through censoring nodes $\{\mathrm{R}_z\}$ as shown. Each pathway is labelled with its *transmittance* defined as the state transition probability times the m.g.f. for holding time in the originating state. For example,

$$p = \Pr(1 \to \mathrm{D}) = \Pr\left(X^0 < Z^0\right)$$

and $M(s)$ is the m.g.f. of the holding time in state 1 before one-step transition to D.

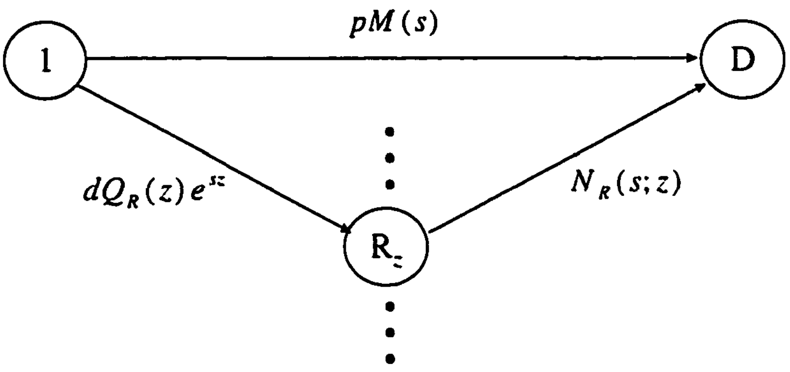

Figure 2.1 Semi-Markov flowgraph for patient lifetimes.

The probability of passing to censored state R_z is

$$dQ_R(z) = \Pr\left\{Z^0 \in (z - dz,\, z],\, Z^0 < X^0\right\} \tag{2.1}$$

and the holding time in state 1 has m.g.f. $\exp(sz)$. This holding time varies according the state of destination which is a characteristic of semi-Markov processes and the reason for referring to the flowgraph as semi-Markov.

The transmittance from $R_z \to D$ is the Laplace-Stieltjes transform for the holding time distribution in state R_z. Using the independence assumption for noninformative censoring, this is specified as

$$\begin{aligned} H(x;z) &= \Pr\left(X^0 \le x + z \mid Z^0 \in (z - dz,\, z],\, X^0 > z\right) \\ &= \frac{F^0(x+z) - F^0(z)}{1 - F^0(z)}. \end{aligned}$$

Its Laplace-Stieltjes transform is therefore

$$\begin{aligned} N_R(s;z) &= \int_0^\infty e^{sx} dH(x;z) = \int_0^\infty e^{sx} \frac{1}{1 - F^0(z)} dF^0(x+z) \\ &= \frac{e^{-sz}}{1 - F^0(z)} \int_0^\infty e^{s(x+z)} dF^0(x+z) \\ &= \frac{e^{-sz}}{1 - F^0(z)} \int_z^\infty e^{sx} dF^0(x). \end{aligned}$$

One of the primary aims of the random censoring model is to provide a context in which it is possible to remove the censoring risk factor and estimate the distribution of X^0 apart from its competition with censoring. In terms of the flowgraph, this is equivalent to "pruning" the continuum of censoring states $\{R_z\}$ so that only one direct transition from $1 \to D$ remains. The transmittance for this single transition is the sum over all the mutually exclusive parallel connections from $1 \to D$, and is

$M^0(s)$, the Laplace-Stieltjes transform for the c.d.f. of X^0. Thus,

$$\begin{aligned} M^0(s) &= pM(s) + \int_0^\infty e^{sz} N_R(s;z)\, dQ_R(z) \\ &= pM(s) + \int_0^\infty \frac{1}{1-F^0(z)} \left\{ \int_z^\infty e^{sx} dF^0(x) \right\} dQ_R(z) \\ &= pM(s) + \int_0^\infty e^{sx} \left\{ \int_0^x \frac{dQ_R(z)}{1-F^0(z)} \right\} dF^0(x). \end{aligned} \tag{2.2}$$

The Stieltjes integrals in (2.2) exist so long as F^0 and Q_R have no common jump points. Expression (2.2) may also be derived from first principles without the flowgraph presentation (see the Appendix). The use of flowgraphs however, provides further understanding of (2.2) by determining M^0 as the transfer function for a linear system that accommodates censoring.

Semi-Markov modeling for a right censoring mechanism was previously considered in Lagakos, Sommer, and Zelen (1978). Their model records passage time up to $\min(X^0, Z^0)$ and represents censoring as a single absorbing node. It does not however consider transitions after censoring as required when representing all possible state changes connected with a flowgraph expression for passage time X^0. In this latter setting, a single censoring state cannot express semi-Markov structure after censoring; such states must be indexed as $\{R_z\}$ to allow subsequent transition to state D in time $X^0 - z$. The model of Lagakos et al. (1978) is different because it addresses different concerns, the determination of a nonparametric m.l.e. for F^0; our model seeks to determine self-consistent relationships for the population distributions and their estimates. This requires the semi-Markov modelling of Figure 2.1 which can account for the additional time after censoring from $\min(X^0, Z^0)$ to X^0.

2.2.1 A Self-Consistent Expression for $F^0(x)$

The Laplace-Stieltjes expression in (2.2) provides a solution for $F^0(x)$ in terms of the competitive distributions $F(x)$ and $Q_R(z)$. In terms of these quantities, expression

(2.2) is

$$\int_0^{\infty} e^{sx} dF^0(x) = p \int_0^{\infty} e^{sx} dF(x) + \int_0^{\infty} e^{sx} \left\{ \int_0^x \frac{dQ_R(z)}{1 - F^0(z)} \right\} dF^0(x). \tag{2.3}$$

Inverting the transforms leads to the unique solution for $dF^0(x)$ as

$$dF^0(x) = p\, dF(x) + \left\{ \int_0^x \frac{dQ_R(z)}{1 - F^0(z)} \right\} dF^0(x) \tag{2.4}$$

or

$$dF^0(x) = \left\{ 1 - \int_0^x \frac{dQ_R(z)}{1 - F^0(z)} \right\}^{-1} p\, dF(x). \tag{2.5}$$

This is the population version of the self-consistent equation originally introduced by Efron (1967) for the estimation of $F^0(x)$.

Some elementary computations provide important intuition for expression (2.5) which concerns the relationship of the noncompetitive and unobservable probability $dF^0(x)$ with the competitive and observable probability $p\, dF(x)$.

Lemma 2.1 *If $X^0 \sim F^0$ and $Z^0 \sim H^0$ are independent then*

$$\frac{p\, dF(x)}{1 - H^0(x)} = dF^0(x) \qquad \textit{and} \qquad \frac{(1-p)\, dH(z)}{1 - F^0(z)} = dH^0(z). \tag{2.6}$$

Proof: The results follow by recalling that

$$\begin{aligned} p\, dF(x) &= \Pr\left\{ X^0 \in (x - dx, x], X^0 < Z^0 \right\} \\ &= \Pr\left\{ X^0 \in (x - dx, x] \right\} \Pr\left(Z^0 > x \right) \\ &= dF^0(x) \Pr\left(Z^0 > x \right), \end{aligned}$$

upon using the independence of censoring. ■

Lemma 2.1 states that $dF^0(x)$ is found by dividing the probability $p\, dF(x)$ by the factor $\Pr(Z^0 > x)$. In (2.5), this is accomplished with the division in curly braces. To see this, note that $dQ_R(z) = (1-p)\, dH(z)$ so that, by Lemma 2.1,

$$1 - \int_0^x \frac{dQ_R(z)}{1 - F^0(z)} = 1 - \int_0^x dH^0(z) = \Pr\left(Z^0 > x \right).$$

Thus identity (2.5) is equivalent to the first identity in (2.6); the division in curly braces is division by $\Pr(Z^0 > x)$ which effectively removes consideration of the event $X^0 < Z^0$ from the competitive probability computation $p\,dF(x)$.

We shall continue to discuss the principle involved in Lemma 2.1 when discussing double and interval censoring; it provides an intuitive understanding of the various self-consistent equations that one might otherwise find rather mysterious.

2.2.2 A Self-Consistent Estimator for $F^0(x)$

Data are observed according to the typical random censoring model and are expressed in a form consistent with the flowgraph model. Let $x_1 \leq \ldots \leq x_n$ be the observed deaths and $z_1 \leq \ldots \leq z_r$ the random right censoring times. An estimate of p is $\hat{p} = n/(n+r)$ and the empirical c.d.f.s of $\{x_i\}$ and $\{z_k\}$, denoted as $\hat{F}$ and $\hat{H}$, estimate F and H using realizations of the competitive responses. The empirical m.g.f. $\hat{M}(s)$ is the Laplace-Stieltjes transform of $\hat{F}(x)$. The empirical estimate for $Q_R(z)$ is $\hat{Q}_R(z) = (1-\hat{p})\hat{H}(z)$. These empirical versions of the flowgraph transmittances for Figure 2.1 allow us to now introduce Figure 2.2 as an empirical version of Figure 2.1.

Each censored value z_k must now index its own censoring node as well as its own remaining transmittance from $\mathrm{R}_{z_k} \to \mathrm{D}$ given by

$$\hat{N}_R(s; z_k) = \frac{e^{-sz_k}}{1 - \hat{F}^0(z_k)} \int_{z_k}^{\infty} e^{sx} d\hat{F}^0(x).$$

Note that we have presumed the existence of an estimate $\hat{F}^0(x)$ in order to define $\hat{N}_R(s; z_k)$. This presumption is the first step required to determine an equation for which $\hat{F}^0(x)$ is a self-consistent solution. Note also that the empirical flowgraph in Figure 2.2 is itself a semi-Markov system that estimates the true semi-Markov system of Figure 2.1.

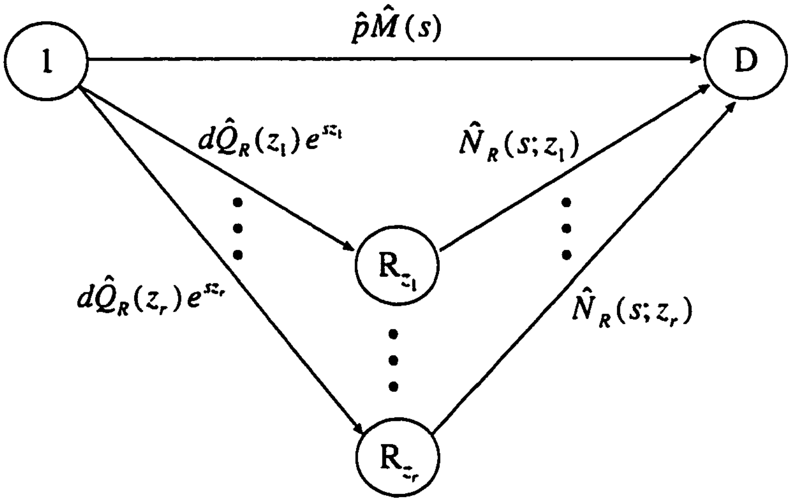

Figure 2.2 Empirical flowgraph for right-censoring representing a semi-Markov estimate of the flowgraph in Figure 1.

Our discussion for "pruning" the censored nodes from Figure 2.1 applies equally well to its empirical counterpart in Figure 2.2. In order for the various Stieltjes integrals to exist, we require that $\hat{F}$ and $\hat{Q}_R$ have no common jump points, i.e. no ties between deaths and censoring.

Theorem 2.1 *Suppose that $\hat{F}$ and $\hat{Q}_R$ have no common jump points and let $z^* :=$ $\max(x_n, z_r)$. For any $x < z^*$, the Kaplan-Meier estimator $\hat{F}^0(x)$ is the unique self-consistent estimator of $F^0(x)$, that solves the empirical flowgraph equation*

$$\int_0^\infty e^{sx} d\hat{F}^0(x) = \hat{p} \int_0^\infty e^{sx} d\hat{F}(x) + \int_0^\infty e^{sx} \left\{ \int_0^x \frac{d\hat{Q}_R(z)}{1 - \hat{F}^0(z)} \right\} d\hat{F}^0(x). \qquad (2.7)$$

This estimator is right-continuous and places mass

$$d\hat{F}^0(x_i) = \left\{ 1 - \int_0^{x_i} \frac{d\hat{Q}_R(z)}{1 - \hat{F}^0(z)} \right\}^{-1} \hat{p}\, d\hat{F}(x_i). \qquad (2.8)$$

$$= \left\{ (n+r) - \sum_{\{k : z_k < x_i\}} \frac{1}{1 - \hat{F}^0(z_k)} \right\}^{-1} \qquad (2.9)$$

at support point x_i for $i = 1, \ldots, n$. Values of $\hat{F}^0(x)$ for $x \geq z^*$ require the consideration of two distinct settings:

If $x_n = z^*$, the mass points in (2.8) add to 1 and $\hat{F}^0(x) \equiv 1$ for all $x \geq z^*$. Thus $\hat{F}^0(x)$ is uniquely defined for all x.

If $z_r = z^*$, the mass points in (2.8) sum to < 1. The self-consistent estimate is left continuous at z_r with

$$\hat{F}^0(z_r) = \hat{F}^0(z_r^-) = 1 - \left\{ (n+r) - \sum_{k=1}^{r-1} \frac{1}{1 - \hat{F}^0(z_k)} \right\}^{-1}.$$

For values $x > z^*$, the self-consistent estimate $\hat{F}^0(x)$ is flat, indeterminate, and assumes an arbitrary value in $[\hat{F}^0(z_k), 1]$.

This characterization of the self-consistent estimate differs slightly from that given in Efron (1967, theorem 7.1). Efron mistakenly has the unique self-consistent estimate set to $\hat{F}^0(x) = 1$ for any $x > z_r = z^*$ so that it must place a nonzero mass at z_r. This is actually only one of many solutions since $\hat{F}^0(x)$ is indeterminate and $d\hat{F}^0(x) = 0$ for any value $x > z_r = z^*$.

In order to have the self-consistent and Kaplan-Meier estimators identical in all settings, we adopt the convention throughout that both estimators assume indeterminate values when the largest observation is censored.

Proof: The solution has been derived above and its uniqueness follows from the uniqueness property of the Laplace-Stieltjes transforms. The mass in (2.9), is the mass assumed by the Kaplan-Meier as shown in Efron (1967). The left-continuity of $\hat{F}^0(x)$ at $x = z_r = z^*$ and its indeterminacy above z_r are shown in the Appendix. ∎

Efron (1967) has shown that the Kaplan-Meier estimate $\hat{F}^0(x)$ is the unique self-consistent solution to the equation

$$1 - \hat{F}^0(x) = \hat{p}\{1 - \hat{F}(x)\} + (1 - \hat{p})\{1 - \hat{H}(x)\} + \{1 - \hat{F}^0(x)\} \int_0^x \frac{d\hat{Q}_R(z)}{1 - \hat{F}^0(z)} \tag{2.10}$$

which is easily shown to be equivalent to (2.7) when $x < z^*$. Take the differential of each side and, upon recognition that $d\hat{H}(x_i) = 0 = d\hat{Q}_R(x_i)$, then

$$d\hat{F}^0(x_i) = \hat{p}\, d\hat{F}(x_i) + d\hat{F}^0(x_i)\} \int_0^{x_i} \frac{d\hat{Q}_R(z)}{1 - \hat{F}^0(z)}.$$

Solving this expression leads to (2.8) and its Laplace-Stiltjes transformation is (2.7).

Further understanding of relation (2.8) results by developing the empirical versions of the competitive/noncompetitive relationships expressed in (2.6).

Lemma 2.2 *The self-consistent (Kaplan-Meier) estimators $\hat{F}^0(x)$ and $\hat{H}^0(z)$, defined in Theorem 2.1, are related to their competitive empirical counterparts $\hat{F}(x)$*

and $\hat{H}(z)$ *according to*

$$\frac{\hat{p}\, d\hat{F}(x)}{1-\hat{H}^0(x)} = d\hat{F}^0(x) \tag{2.11}$$

and

$$\frac{(1-\hat{p})\, d\hat{H}(z)}{1-\hat{F}^0(z)} = d\hat{H}^0(z) \tag{2.12}$$

for all $x, z \in [0, z^*)$. *If, for example,* $z_r = z^*$, *then (2.11) and (2.12) together require that* $\hat{H}^0(y) = 1$ *for* $z \geq z_r$, $\hat{F}^0(z_r) = \hat{F}^0(z_r^-)$, *and are indeterminate about* $\hat{F}^0(x)$ *with* $x > z_r$.

Before proving this result, we consider the intuitive meaning it provides for understanding the masses in (2.8) and (2.9). From (2.11), write

$$d\hat{F}^0(x_i) = \frac{\hat{p}\, d\hat{F}(x_i)}{1-\hat{H}^0(x_i)}. \tag{2.13}$$

According to (2.12), we may now replace the denominator in (2.13) with

$$1-\hat{H}^0(x_i) = 1 - \int_0^{x_i} \frac{(1-\hat{p})\, d\hat{H}(z)}{1-\hat{F}^0(z)} = 1 - \int_0^{x_i} \frac{d\hat{Q}_R(z)}{1-\hat{F}^0(z)},$$

after using $d\hat{Q}_R(z) = (1-\hat{p})\, d\hat{H}(z)$. This leads directly to expression (2.8). The gist of this understanding may be summarized by making two points:

1. Division of the competitive probability $\hat{p}\, d\hat{F}(x_i)$ by the term in curly braces, a disguised version of the Kaplan-Meier estimator $1-\hat{H}^0(x_i)$, produces the non-competitive Kaplan-Meier probability $d\hat{F}^0(x_i)$. This becomes a very general mechanism for removing probabilistic competition.

2. The self-consistent equations are a result of the circularity arising from (a) specifying $d\hat{F}^0$ in terms of $d\hat{F}$ and $\hat{H}^0$ as in (2.11), followed by (b) specifying $\hat{H}^0$ in terms of $d\hat{Q}_R$ and $\hat{F}^0$ as in (2.12), so that $d\hat{F}^0$ is determined from $d\hat{F}$, $d\hat{Q}_R$ and $\hat{F}^0$ as in (2.8) and (2.9).

Proof: We first consider the situation on $[0, z^*)$ by introducing the notion of an independent risks *proxy* model as discussed in Crowder (2000). Suppose that $\hat{P}(x, z)$ is a joint empirical c.d.f. and let random variables $\hat{X}^0$ and $\hat{Z}^0$ be such that $(\hat{X}^0, \hat{Z}^0) \sim \hat{P}$. Then $\hat{P}$ is defined to be an *empirical proxy* distribution if, when viewed competitively, $\hat{X}^0 \,|\, \{\hat{X}^0 < \hat{Z}^0\} \sim \hat{F}$ and $\hat{Z}^0 \,|\, \{\hat{Z}^0 < \hat{X}^0\} \sim \hat{H}$. The independent empirical proxy distribution necessarily has $\hat{P}(x, z) = \hat{F}^0(x)\hat{H}^0(z)$ where $\hat{F}^0$ and $\hat{H}^0$ are the respective Kaplan-Meier estimates for F^0 and H^0. The indeterminacy of $\hat{F}^0$ or $\hat{H}^0$ above z^* assures indeterminacy for $\hat{P}$ with one or more component values above z^*; it is therefore not a probability distribution since it accounts only for a finite probability totalling $\hat{P}(x_n, z_r) < 1$. The independent empirical proxy distribution is only one of many empirical proxy models that yield the empirical probability distributions $\hat{F}$ and $\hat{H}$ when modified to account for competition.

Expression (2.8) is now governed by the independent proxy measure $\hat{P}$ in which $\hat{X}^0$ and $\hat{Z}^0$ are independent variables. It is the empirical version of the population expression (2.5) that is governed by the distribution $P(x, z) = F^0(x)H^0(y)$ and for which independence has been assumed rather than constructed through the notion of proxy models. Since independence is found in both settings and is the crux of the arguments in the population results of (2.6), then these results must extend to the empirical counterparts as stated in (2.11) and (2.12).

The situation for values exceeding z^* is shown through tedious consideration of the solutions possible in (2.11) and (2.12). ■

When ties exist between deaths and right censored times, the usual assumption that the censored values are slightly larger allows the presentation to be extended to all settings.

Example. Suppose the data are ordered according to

$$\begin{array}{ccccccccccccccccc} x_1 & < & z_1 & < & x_2 & < & z_2 & < & z_3 & < & z_4 & < & x_3 & < & x_4 & < & z_5 \\ D & & R & & D & & R & & R & & R & & D & & D & & R \end{array}. \tag{2.14}$$

Self-consistent survival estimates and their placement of point masses, as in (2.8), are given in the table below.

Table 2.1 Kaplan-Meier estimates for the data in (2.14).

i	$1-\hat{F}^0(x_i)$	$d\hat{F}^0(x_i)$	k	$1-\hat{H}^0(z_k)$	$d\hat{H}^0(z_k)$
1	$8/9$	$1/9$	1	$7/8$	$1/8$
2	$8/9 \times 6/7 = 16/21$	$8/63$	2	$7/8 \times 5/6 = 35/48$	$7/48$
3	$16/21 \times 2/3 = 32/63$	$16/63$	3	$35/48 \times 4/5 = 7/12$	$7/48$
4	$32/63 \times 1/2 = 16/63$	$16/63$	4	$7/12 \times 3/4 = 7/16$	$7/48$
			5	0	$7/16$

Table 2.2 demonstrates that the identities (2.11) and (2.12) provide the same masses as in Table 1. From (2.14), we have $\hat{p}\,d\hat{F}(x_i) = 4/9 \times 1/4 = 1/9$ and $(1 - \hat{p})\,d\hat{H}(y_j) = 5/9 \times 1/5 = 1/9$ so that

Table 2.2 Illustrating identities (2.11) and (2.12).

i	$\dfrac{\hat{p}\,d\hat{F}(x_i)}{1-\hat{H}^0(x_i)}$	$1-\hat{F}^0(x_i)$	k	$\dfrac{(1-\hat{p})\,d\hat{H}(z_k)}{1-\hat{F}^0(z_k)}$	$1-\hat{H}^0(z_k)$
1	$1/9 \div 1 = 1/9$	$1/9$	1	$1/9 \div 8/9 = 1/8$	$7/8$
2	$1/9 \div 7/8 = 8/63$	$5/21$	2	$1/9 \div 6/21 = 7/48$	$35/48$
3	$1/9 \div 7/16 = 16/63$	$31/63$	3	$1/9 \div 16/21 = 7/48$	$7/12$
4	$1/9 \div 7/16 = 16/63$	$16/63$	4	$1/9 \div 16/21 = 7/48$	$7/16$
			5	$1/9 \div 16/63 = 7/16$	0

Note that $\hat{p}\,d\hat{F}(z_5)/\{1-\hat{H}^0(z_5)\}$ is indeterminate of the form 0/0 and therefore undefined. The survival value $1-\hat{F}^0(z_5)=16/63$ has been used in the calculation of

$$(1-\hat{p})\,d\hat{H}(z_5)/\left\{1-\hat{F}^0(z_5)\right\}=\frac{1/9}{16/63}=\frac{7}{16}=d\hat{H}^0(z_5).$$

■

2.2.3 Self-Consistency of the Berliner-Hill Survival Estimate

The Kaplan-Meier estimate is the self-consistent solution to (2.8) that results from substituting $d\hat{F}(x)$ and $d\hat{Q}_R(z)$ for their population counterparts in (2.5). This use of empirical distributions amounts to placing point masses of $1/n$ on the death times and $1/r$ on the censoring times. Alternatively, consider the approach of Berliner and Hill (1988) in which $d\hat{F}(x)$ is replaced by a posterior measure that places mass $1/(n+1)$ just to the left of each of the observed death times and also to the left of the largest death time at $x_{n+1}:=\infty$. Each of the $1/(n+1)$ posterior masses may be either continuous or discrete and its entirety must be placed to the right of the next smallest observation. This leads to the competitive estimates

$$\tilde{p}=\frac{n+1}{n+r+1},\ d\tilde{Q}_R(z_k)=\frac{1}{n+r+1}\ \text{and}\ \tilde{F}(x_i)=\frac{i}{n+1} \tag{2.15}$$

for $i=1,\dots,n+1$. Now, the Berliner and Hill (1988) estimator becomes the unique self-consistent estimator when these tilded estimates are used in (2.8) to replace of their hatted counterparts.

Theorem 2.2 *When considered at death times $\{x_i : i=1,\dots,n\}$, the Berliner and Hill (1988) estimator of F^0 is the unique self-consistent estimate satisfying (2.8) using the tilded estimates in (2.15) in place of their hatted counterparts.*

Proof: Denote this estimate as $\tilde{F}^0$. Its mass points $d\tilde{F}^0(x_i^-)$ are

$$d\tilde{F}^0(x_i^-) = \left\{1 - \int_0^{x_i} \frac{d\tilde{Q}_R(z)}{1 - \tilde{F}^0(z)}\right\}^{-1} \tilde{p}\, d\tilde{F}(x_i^-)$$

$$= \left\{(n + r + 1) - \sum_{z_k < x_i} \frac{1}{1 - \hat{F}^0(z_k)}\right\}^{-1} \tag{2.16}$$

and correspond to those for the Berliner and Hill estimate as shown in the Appendix.

■

Since $x_{n+1}^- > z_r$ in all settings, the indeterminacy found in $\hat{F}^0(x)$ when $x > z_r > x_n$ cannot occur with this estimate.

2.3 Doubly-Censored Lifetimes

A patient is left censored when the event of interest, such as death, occurs before the patient is sampled or enters into the observational study. In this case, the recorded censored lifetime is the patient's age at the start of observation. A flowgraph accommodating such doubly-censored lifetimes is shown below. The top two paths, when considered alone, constitute the flowgraph for right-censored patients from §2 and the transmittances are much the same. The additional parallel path on the bottom has been inserted to accommodate the possibility of left-censoring as we explain below.

We continue with the notation of §2 using X^0 and Z^0 as the patient's noncompetitive death and right-censoring times respectively. In addition, we must now add the patient's age Y^0 upon entering observation. Assume that $Y^0 \sim G^0(y)$, (Y^0, Z^0) is independent of X^0, and $Y^0 < Z^0$ with probability 1. The standard model assumes that a patient is left-censored if $X^0 < Y^0$, an observed death when $Y^0 < X^0 < Z^0$, or right-censored if $Z^0 < X^0$. The support of all random variables is assumed to be $(0, \infty)$.

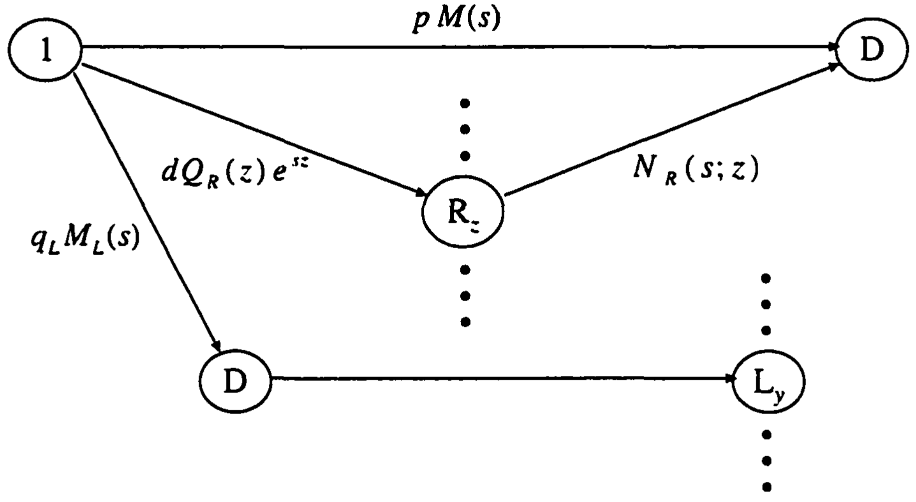

Figure 2.3 Semi-Markov flowgraph for patient lifetimes in the presence of left- and right-censoring.

The lower branch transmittance $q_L M_L(s)$ represents left-censoring in which $q_L = \Pr(X^0 < Y^0)$ is the probability of left-censoring, and $M_L(s)$ is the m.g.f. of the random variable

$$X_L \stackrel{d}{=} X^0 \mid \{X^0 < Y^0\} \sim F_L(x).$$

After passing to state D, a left-censored patient passes to a state L_y in total time y, where y is the left-censoring time. The states $\{ \mathrm{L}_y : y > 0\}$ comprise a continuum of possible left-censoring states just like the right-censoring states $\{\mathrm{R}_z : z > 0\}$.

2.3.1 A Self-Consistent Expression for $F^0(x)$

In Figure 2.3, we are only interested in passage from $1 \to D$, so the states $\{\mathrm{L}_y\}$ are actually unnecessary for describing this passage. Our reason for including these states is to deal with the problem arising from the fact that transform $M_L(s)$ does not have a directly observable empirical counterpart. We can determine an estimable form of M_L in terms of G^0 and F^0 by first conditioning on passage to state L_y and then summing over all distinct pathways through the states $\{\mathrm{L}_y\}$.

Lemma 2.3 *Suppose that $G^0(\cdot)$ is continuous (a.e.) with respect to dF^0. Then*

$$M_L(s) = \frac{1}{q_L}\int_0^\infty e^{sx}\left\{\int_x^\infty dG^0(y)\right\} dF^0(x)$$

Proof: Take the differential of

$$F_L(x) = \Pr\left\{X^0 \le x \mid X^0 < Y^0\right\} = \frac{1}{q_L}\int_0^x \left\{\int_u^\infty dG^0(y)\right\} dF^0(u),$$

and determine the Laplace-Stiltjes transform as

$$M_L(s) = \int_0^\infty e^{sx}\, dF_L(x) = \frac{1}{q_L}\int_0^\infty e^{sx}\left\{\int_x^\infty dG^0(y)\right\} dF^0(x).$$

■

To express $M_L(s)$ in terms of dF^0 and other estimable quantities, let

$$dQ_L(y) = \Pr\left\{Y^0 \in (y-dy, y],\ X^0 < Y^0\right\} = dG^0(y)\ F^0(y)$$

and rewrite M_L as

$$M_L(s) = \frac{1}{q_L}\int_0^\infty e^{sx}\left\{\int_x^\infty \frac{dQ_L(y)}{F^0(y)}\right\} dF^0(x), \tag{2.17}$$

which is now estimable from left-censored observations.

Our goal is to remove the risk factors due to left- and right-censoring and determine $F^0(x)$ with Laplace-Stieltjes transform $M^0(s)$. The removal of risk factors is equivalent to the pruning of all censored states from the flowgraph. This is accomplished by summing over the two continuums of mutually exclusive censored states $\{\mathrm{L}_y\}$ and $\{\mathrm{R}_z\}$ and reducing the flowgraph to the single transition $1 \to D$ whose transmittance must be $M^0(s)$. Thus we may equate

$$\begin{aligned} M^0(s) &= q_L M_L(s) + pM(s) + \int_0^\infty e^{sz} N_R(s; z)\, dQ_R(z) \\ &= \int_0^\infty e^{sx}\left\{\int_x^\infty \frac{dQ_L(y)}{F^0(y)}\right\} dF^0(x) + p\int_0^\infty e^{sx}\, dF(x) \\ &+ \int_0^\infty e^{sx}\left\{\int_0^x \frac{dQ_R(z)}{1-F^0(z)}\right\} dF^0(x) \end{aligned}$$

where $N_R(s; z)$ and $dQ_R(z)$ are defined as in the first section. A proof of this relationship from first principles is in the Appendix.

Inverting the Laplace-Stieltjes transform leads to a unique representation for $dF^0(x)$ as

$$dF^0(x) = \left\{\int_x^\infty \frac{dQ_L(y)}{F^0(y)}\right\} dF^0(x) + p\,dF(x) + \left\{\int_0^x \frac{dQ_R(z)}{1-F^0(z)}\right\} dF^0(x)$$

or

$$dF^0(x) = \left\{1 - \int_x^\infty \frac{dQ_L(y)}{F^0(y)} - \int_0^x \frac{dQ_R(z)}{1-F^0(z)}\right\}^{-1} p\,dF(x),$$

which is a generalization of (2.5) to include left-censoring.

2.3.2 A Self-Consistent Estimator for $F^0(x)$

Data are observed according to the random doubly-censored model described above. Let $x_1 \leq \cdots \leq x_n$ be the observed deaths, $y_1 \leq \cdots \leq y_l$ the left-censored times, and $z_1 \leq \cdots \leq z_r$ the right-censored times. Without loss in generality, assume that there are no ties in any of the observations. An estimate of p is $\hat{p} = n/(n+l+r)$ and the estimator $\hat{F}$, its empirical m.g.f. $\hat{M}$, and $d\hat{Q}_R(z_k)$ and $\hat{N}_R(s; z_k)$, for $k = 1, \ldots r$, are as given in §2 with the caveat that $r/(n+l+r)$ replaces $r/(n+r)$ as the total probability for $d\hat{Q}_R(z_k)$. These empirical versions of the flowgraph transmittances from Figure 2.3 allow for the introduction of the empirical flowgraph in Figure 2.4 which is also a semi-Markov system.

Transmittance $q_L M_L(s)$ is unobservable and its estimate $\hat{q}_L \hat{M}_L(s)$ must be found by marginalising over the distribution of observed left-censored values $y_1, \ldots, y_l$. This amounts to estimating $M_L(s)$ using (2.17) so that the bottom empirical transmittance from $1 \to D$ is

$$\hat{q}_L \hat{M}_L(s) = \int_0^\infty e^{sx} \left\{ \int_x^\infty \frac{d\hat{Q}_L(y)}{\hat{F}^0(y)} \right\} d\hat{F}^0(x),$$

with $d\hat{Q}_L$ placing mass $1/(n+l+r)$ on each y_i-value. Solving this empirical flowgraph in the same manner leads to the expression

$$\begin{aligned} \int_0^\infty e^{sx}\, d\hat{F}^0(x) &= \int_0^\infty e^{sx} \left\{ \int_x^\infty \frac{d\hat{Q}_L(y)}{\hat{F}^0(y)} \right\} d\hat{F}^0(x) + \hat{p} \int_0^\infty e^{sx}\, d\hat{F}(x) \\ &+ \int_0^\infty e^{sx} \left\{ \int_0^x \frac{d\hat{Q}_R(z)}{1 - \hat{F}^0(z)} \right\} d\hat{F}^0(x), \end{aligned} \tag{2.18}$$

where $\hat{F}^0$ is a presumed self-consistent solution. This is solved for $d\hat{F}^0(x)$ by following the same approach used with the population versions and gives

$$\hat{C}(x; \hat{F}^0) d\hat{F}^0(x) = d\hat{F}(x) \tag{2.19}$$

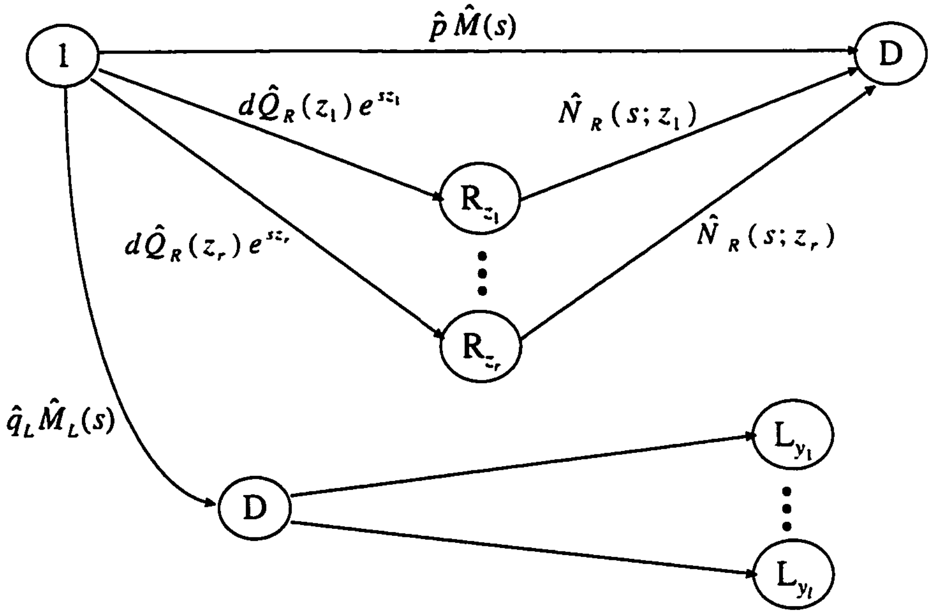

Figure 2.4 Empirical flowgraph for double censoring representing a semi-Markov estimate of the population flowgraph in Figure 2.3.

where

$$\hat{C}(x;\hat{F}^0) := 1 - \int_x^\infty \frac{d\hat{Q}_L(y)}{\hat{F}^0(y)} - \int_0^x \frac{d\hat{Q}_R(z)}{1-\hat{F}^0(z)}. \tag{2.20}$$

In order to characterize a solution to (2.19), we must first assure that the Stieltjes integrals in (2.20) are meaningful for all $\hat{F}^0$ including an $\hat{F}^0$ with potential step-discontinuities on $\{y_i\} \cup \{z_i\}$; this is done by first writing (2.20) as summations over $\{y_i\} \cup \{z_i\}$ before attempting to evaluate $\hat{C}(x)$. The support for self-consistent solutions is now characterized to assure that equality is maintained in (2.19).

Lemma 2.4 *Any self-consistent solution $\hat{F}^0(x)$ places point masses at the death times given by*

$$0 < d\hat{F}^0(x_i) = \hat{C}(x_i;\hat{F}^0)^{-1}\,\hat{p}\,d\hat{F}(x_i) \qquad i = 1,\ldots,n. \tag{2.21}$$

In addition, it may also distribute mass arbitrarily on the set

$$\left\{x : \hat{C}(x;\hat{F}^0) = 0\right\} \subset (0,\infty)\setminus\{x_i\}. \tag{2.22}$$

A complete answer to the determination of self-consistent solutions to (2.18) has been provided in Butler and Bronson (2001). If the range of support for $\hat{F}^0$ includes both the points in (2.21) and the region (2.22), then they have characterized the collection of self-consistent solutions to (2.19) and (2.20); the uniquely defined non-parametric m.l.e for this problem has been shown to be a member of this collection. Furthermore, a simple EM algorithm has been provided making it easy to determine all members in this collection including the global nonparametric m.l.e. For the present setting however, we restrict our search to $\hat{F}^0 \in \mathcal{F}_n :=$ {solutions placing all mass at the death points}.

Theorem 2.3 *Suppose, without loss in generality, that $\hat{F}$, $\hat{Q}_L$, and $\hat{Q}_R$ share no common points of discontinuity and let $y^* = \min(y_1, x_1)$ and $z^* = \max(x_n, z_r)$. Within*

$\mathcal{F}_n$, there exists a unique self-consistent estimator $\hat{F}^0(x)$ of $F^0(x)$ that solves the empirical flowgraph equation and is well-defined for any $x \in (y^*, z^*)$. Estimator $\hat{F}^0(x)$ agrees with a particular version of the Turnbull (1974) estimator, is right-continuous with support on $\{y^*, x_1, \ldots, x_n, z^*\}$, and puts mass (2.21) or

$$d\hat{F}^0(x_i) = \hat{C}(x_i; \hat{F}^0)^{-1}(n+l+r)^{-1} \tag{2.23}$$
$$= \left\{(n+l+r) - \sum_{\{j:y_j > x_i\}} \frac{1}{\hat{F}^0(y_j)} - \sum_{\{k:z_k < x_i\}} \frac{1}{1-\hat{F}^0(z_k)}\right\}^{-1}$$

at x_i for $i = 1, \ldots, n$. Corollary 2.2 deals with $\hat{F}^0$ at y^* and z^* and values outside of (y^*, z^*).

Proof: Tsai and Crowley (1985, equation 5.1) have stated that Turnbull's estimate is a self-consistent solution to the equation

$$1 - \hat{F}^0(x) = \hat{p}\left\{1 - \hat{F}(x)\right\} + \hat{p}_L\left\{1 - \hat{G}(x)\right\} + \hat{p}_R\left\{1 - \hat{H}(x)\right\}$$
$$- \int_x^{\infty} \frac{\hat{F}^0(x)}{\hat{F}^0(y)} d\hat{Q}_L(y) + \int_0^x \frac{1 - \hat{F}^0(x)}{1 - \hat{F}^0(z)} d\hat{Q}_R(z), \tag{2.24}$$

where $\hat{p}_L = \int d\hat{Q}_L(y)$ and $\hat{p}_R = \int d\hat{Q}_R(z)$ are the estimated probabilities of left- and right-censoring, respectively, and $\hat{G}(y) = \hat{Q}_L(y)/\hat{p}_L$ and $\hat{H}(z) = \hat{Q}_R(z)/\hat{p}_R$ are the respective empirical c.d.f.s of $y_1 \le \cdots \le y_l$ and $z_1 \le \cdots \le z_r$. Restricting attention to $\hat{F}^0 \in \mathcal{F}_n$, then differentials of (2.24) give

$$d\hat{F}^0(x_i) = \hat{p}\, d\hat{F}(x_i) + d\hat{F}^0(x_i) \int_{x_i}^{\infty} \frac{d\hat{Q}_L(y)}{\hat{F}^0(y)} + d\hat{F}^0(x_i) \int_0^{x_i} \frac{d\hat{Q}_R(z)}{1 - \hat{F}^0(z)},$$

which is the inversion of the Laplace-Stiltjes transforms in (2.18); thus the solutions are equivalent when searching in $\mathcal{F}_n$. The details of a more convincing proof have been given in Butler and Bronson (2001). There, the solution (2.23) has been shown to be the unique critical value of a nonparametric likelihood similar to that introduced by

Turnbull (1974); therefore it becomes the unique stationary point of an EM algorithm that is a modification to that used by Turnbull (1974). ■

The modifications to the Turnbull algorithm needed in computing $\hat{F}^0$ in (2.23) are now detailed. The algorithm is not uniquely defined because it depends on a choice of partition within which the data is categorized. The input to the algorithm consists of bin counts for the deaths, right- and left-censoring, as determined by the prespecified partition; thus the algorithm converges to a unique stationary value that is partition dependent. The limit point $\hat{F}^0$ in (2.23) results from taking the partition determined by y^*, z^*, and $\{x_i : i = 1, \ldots, m\}$ and ranging from $(0, y^*], (y^*, t_1], \ldots, (t_m, z^*]$ where t_1 (t_m) is either x_1 or x_2 (either x_{n-1} or x_n). See Butler and Bronson (2001) for the details.

Further understanding of the mass placement in (2.23) results from considering the adjustment term $\hat{C}(x_i; \hat{F}^0)$ in (2.23) used to modify the uniform distribution $1/(n+l+r)$. Consider the noncompetitive probability of the event $C_i = \{T$ is not censored in $(x_i^-, x_i]\}$. The product in (2.23) is $1/(n+l+r)$, the competitive probability of death in $(x_i^-, x_i]$, discounted or divided by $\widehat{\Pr}(C_i)$. To see this note that

$$\widehat{\Pr}(C_i) = 1 - \widehat{\Pr}(C_i^c) = 1 - \int_{x_i}^{\infty} d\hat{G}^0(y) - \int_0^{x_i} d\hat{H}^0(z). \tag{2.25}$$

Noncompetitive left- and right-censoring probabilities are the result of discounting the competitive censoring probabilities by the noncompetitive probabilities of death; thus

$$d\hat{G}^0(y) = \frac{d\hat{Q}_L(y)}{\hat{F}^0(y)} \qquad d\hat{H}^0(z) = \frac{d\hat{Q}_R(z)}{1 - \hat{F}^0(z)} \tag{2.26}$$

so that

$$d\hat{F}^0(x_i) = \widehat{\Pr}(C_i)^{-1} \frac{1}{n+l+r},$$

with $\widehat{\Pr}(C_i)^{-1}$ specified by substituting (2.26) into (2.25). This is exactly the self-consistency relationship expressed in (2.21) and (2.23).

The flowgraph estimate in (2.23) demonstrates that left-censored values and right-censored values play a symmetric role in determining $\hat{F}^0$. To emphasize this, suppose we multiply all observed times by -1 and refer to the ensuing double censoring analysis as occurring in negative time. For $x < 0$, let $\hat{F}^0_-(x)$ be the self-consistent estimate in negative time for $F^0_-(x) := 1 - F^0(-x)$.

Corollary 2.1 *The self-consistent estimator $\hat{F}^0$ is equivariant under the reflective group $\{+1, -1\}$. More precisely,*

$$\hat{F}^0_-(-x_i) = 1 - \hat{F}^0(x_i^-) \qquad i = 0, \ldots, n+1$$

where, for sake of notation, we take

$$x_0 = \begin{cases} 0 & \text{if } y^* = x_1 \\ y^* & \text{if } y^* = y_1 \end{cases} \qquad x_{n+1} = \begin{cases} \infty & \text{if } z^* = x_n \\ z^* & \text{if } z^* = z_r \end{cases} .$$

Estimator $\hat{F}^0$ may be indeterminate outside of $[y^*, z^*]$ when the extreme values are censored observations.

Corollary 2.2 *For $x < y^*$,*

$$\hat{F}^0(x) = \begin{cases} 0 & \text{if } y^* = x_1 \\ \text{indeterminate} & \text{if } y^* = y_1 \end{cases} .$$

For $x > z^*$,

$$\hat{F}^0(x) = \begin{cases} 1 & \text{if } z^* = x_n \\ \text{indeterminate} & \text{if } z^* = z_r \end{cases} .$$

In the indeterminate cases, $\hat{F}^0$ is right and left continuous on boundary points y^* and z^* respectively with

$$0 < \hat{F}^0(y^*) = \hat{F}^0(x_1) - d\hat{F}^0(x_1)$$

and

$$\hat{F}^0(z^*) = \hat{F}^0(x_n) < 1.$$

Proof: The last step of the Turnbull algorithm is a Kaplan-Meier computation that ignores the left-censored times. The last step guarantees this form for the right tail. The reflective symmetry property of Corollary 2.1 guarantees the mirror image property in the left tail. A formal demonstration of the indeterminacy is given in the Appendix. ■

Example. We demonstrate how Turnbull's algorithm adjusts the initial Kaplan-Meier estimate to account for the left-censored values. The data given in the table below were simulated with F^0 as exponential with mean 30.

time	2.8	3.1	7.8	12.8	17.2	23.0	30.7	32.3	33.0	39.5	46.0	47.1
type	D	L	R	D	D	L	D	D	R	R	L	D

The table 2.3 gives the death time x_i along with its initial Kaplan-Meier survival estimate, the Turnbull survival estimate $1 - \hat{F}^0(x_i)$, and the exact survival $1 - F^0(x_i)$.

Table 2.3 Example of the self-consistent estimate.

Death x_i	Kaplan-Meier	$1 - \hat{F}^0(x_i)$	$1 - F^0(x_i)$
2.8	0.8889	0.7682	0.9109
12.8	0.7619	0.6318	0.6523
17.2	0.6349	0.4953	0.5629
30.7	0.5079	0.3878	0.3595
32.3	0.3810	0.2804	0.3407
47.1	0.0000	0.0000	0.2079

Figure 2.5 shows how the Turnbull estimate moves the mass of the initial Kaplan-Meier estimate to the left thus deflating the Kaplan-Meier estimate at any fixed value x_i.

■

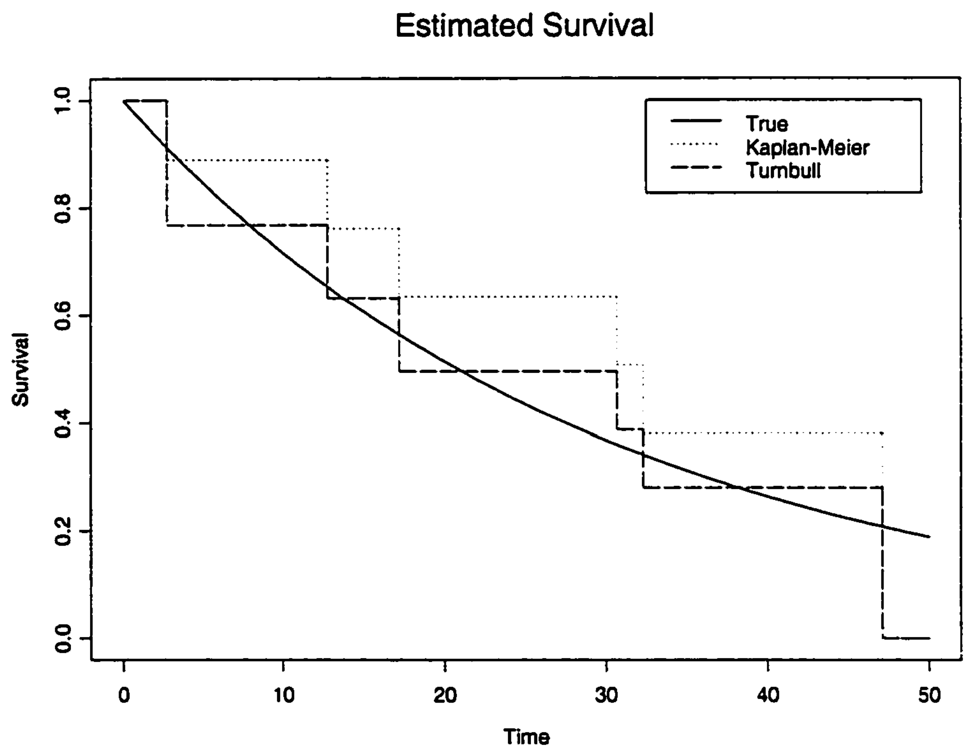

Figure 2.5 Graph demonstrating the effect Turnball's algorithm has on the initial Kaplan-Meier estimate.

3 Censoring in Semi-Markov Models

3.1 Introduction

We now consider the bootstrap inference methods of Chapter 1 for passage or survival times in semi-Markov systems with the additional complication: there is the possibility of independent right censoring from the various states of the system. Censored data are observed for patients transcending the states of a semi-Markov system that allows for censoring. The goal is to use this censored data to infer the survival time of a patient passing through the same states but free from the possibility of censoring. This removal of censoring risk amounts to creating a new censor-free semi-Markov system from the observed system in which the possibility for censoring from each system state has been removed. We show how a systems theoretic approach to the elimination of right censoring in the observed system leads to the use Kaplan-Meier estimates for subsurvivals in the new censor-free system. In the latter system, Laplace-Stiltjes transforms of these subsurvival estimates characterize the system dynamics. These transforms can be used, in conjunction with the cofactor rules of Butler (2000, thm. 1) and saddlepoint inversion, to determine bootstrap estimates for the survival distribution and hazard rate of a patient free from censoring risk. In this determination however, we use a single saddlepoint approximation to determine the bootstrap survival and hazard estimates in lieu of the resampling. The double bootstrap may be implemented with the more intensive use of 1000 saddlepoint approximations and provides confidence envelopes for the survival and hazard functions as well as other characteristics related to first passage time through the censor-free system.

The removal of censoring risk from a particular state or node of the observed system is discussed in Section 3.2. In the observed system, transitions out of a state compete with censoring to create the classical competing risks setting discussed in Chapter 9 of Cox and Oakes (1984), for example. We provide a new semi-Markov model to account for censoring in these competing risks settings. Our model describes all state transitions including those into and out of censored states. The model differs from those previously considered by providing for the unobservable transitions of censored patients into subsequent non-censored states. Within the context of our systems theoretic approach, we show that the removal of censored states from the system leads to the specification of certain self-consistency equations that determine censor-free subsurvival function estimates. These self-consistent estimates turn out to be the Kaplan-Meier subsurvival estimators introduced by Lagakos, Summer, and Zelen (1978) as nonparametric maximum likelihood estimates (m.l.e.s) for this competing risk setting. While our self-consistent subsurvival estimates may agree with those of Lagakos, Summer, and Zelen (1978), our model is different. As previously mentioned, Lagakos et al. have chosen to model only the observable portion of patient state transitions and not the transitions occurring after censoring since only this observable portion is required when determining nonparametric m.l.e.s.

The removal of censored states from the observed system leads to a new censor-free semi-Markov system with subsurvival distributions given by Kaplan-Meier estimates. This new system may be characterized through the specification of its branch transmittances which are now the Laplace-Stiltjes transforms for these Kaplan-Meier sub-survival or sub-c.d.f. estimators. Now, employing only minor modifications to the methods developed in Chapter 1, we can compute bootstrap estimators and confidence envelopes for the survival distribution and hazard function of a patient as shown in Section 3.3.

We illustrate these methods by considering a generalization of the classical Fix and Neyman (1951) survival model. According to Chiang and Hsu (1976), this is the seminal paper to address general consideration of our current objective: the determination of a patient's survival time with data observed on patient transitions through a multistate feedback system in which censoring may occur from the system states. Our generalization differs from their original formulation in several respects. Their system made Markov assumptions and therefore was essentially parametric. We weaken this rather strong assumption by considering semi-Markov models which allow us to work in a nonparametric framework. Our nonparametric analysis becomes possible with the use of three important new technologies: (i) The determination of a nonparametric estimate for the moment generating function (m.g.f.) of the patient's survival density in the feedback system as given by the cofactor rules in Butler (2000). These rules generalize and simplify the original work of Mason (1953, 1956); (ii) The inversion of this m.g.f. estimate using the saddlepoint approximations in Daniels (1954, 1987); (iii) The use of the single and double bootstraps to estimate and form confidence envelopes for the survival and hazard functions.

3.2 Competing Risks Model

Suppose the current state of a patient is subject to several different state transitions, or death types, that compete with independent censoring. For ease of presentation, we consider only two types of death since an understanding of the full generality is essentially the same. Denote the two non-competitive death times as X_1^0 and X_2^0 with bivariate c.d.f. $F^0(x_1, x_2)$ and assume the right-censoring time Z^0 is independent of (X_1^0, X_2^0). Only $W = \min(X_1^0, X_2^0, Z^0)$ is observed, so we may define the competitive random variables

$$X_k \stackrel{d}{=} X_k^0 \,|\, \{X_k^0 = W\} \;\sim F_k(x_k), \quad \text{with m.g.f. } M_k(s) \quad k = 1, 2$$
$$Z \stackrel{d}{=} Z^0 \,|\, \{Z^0 = W\} \;\sim H(z).$$

Figure 3.1 is a semi-Markov flowgraph for a patient subject to these 3 competing risks. Node 1 is the current alive state and nodes D_1 and D_2 are different death states. The continuum of nodes $\{\mathrm{R}_z : z > 0\}$ denotes the censored states which are indexed by the censoring time z. Once censored, the patient's death time is only dependent on the censoring time z and the state of destination which is consistent with the semi-Markov nature of the flowgraph as indicated by the transmittances leaving node R_z. The transmittances out of state 1 also reflect this semi-Markov nature and have

$$p_k = \Pr(1 \to \mathrm{D}_k) = \Pr\left\{X_k^0 = W\right\} \quad k = 1, 2$$
$$dQ_R(z) = \Pr(1 \to \mathrm{R}_z) = \Pr\{Z^0 \in (z - dz, z], \, Z^0 = W\}.$$

Note that $\int_0^\infty dQ_R(z) = 1 - p_1 - p_2$ so that the transition probabilities out of state 1 add to 1.

The transmittance from $\mathrm{R}_z \to \mathrm{D}_1$ is the Laplace-Stieltjes transform for the destination specific holding time distribution defined by

$$H_1(x_1; z) = \Pr\left\{X_1^0 \leq x_1 + z, \, X_1^0 < X_2^0 \,|\, Z^0 = W \in (z - dz, z]\right\}.$$

Using the independence of the noninformative censoring assumption, this may be expressed as

$$H_1(x_1; z) = \frac{G_1(x_1 + z) - G_1(z)}{1 - G_1(z) - G_2(z)}$$

where G_1 and G_2 are the sub-distribution (sub-c.d.f.) functions calculated from the

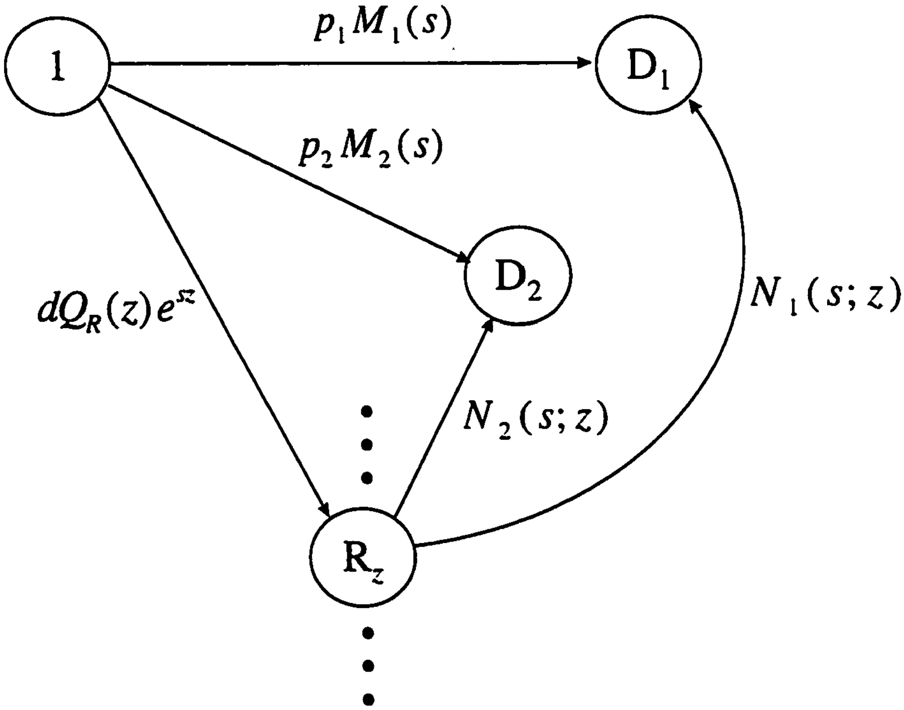

Figure 3.1 Semi-Markov flowgraph for the lifetimes of patients subject to two competing risks and right-censoring.

bivariate c.d.f. F^0 as

$$G_1(x_1) := \Pr\left(X_1^0 \le x_1,\, X_1^0 < X_2^0\right) \tag{3.1}$$

$$G_2(x_2) := \Pr\left(X_2^0 \le x_2,\, X_2^0 < X_1^0\right).$$

Similarly,

$$H_2(x_2; z) = \frac{G_2(x_2 + z) - G_2(z)}{1 - G_1(z) - G_2(z)}.$$

The Laplace-Stieltjes transform of $H_k(x_k; z)$ is therefore

$$\begin{aligned} N_k(s; z) &= \int_0^\infty e^{sx_k}\, dH_k(x_k; z) = \int_0^\infty e^{sx_k} \frac{dG_k(x_k + z)}{1 - G_1(z) - G_2(z)} \\ &= \frac{e^{-sz}}{1 - G_1(z) - G_2(z)} \int_z^\infty e^{sx_k}\, dG_k(x_k). \end{aligned}$$

The competing risks model in Figure 3.1 provides a framework within which we may remove the risk of right-censoring and estimate the sub-c.d.f. distributions in (3.1). The removal of right-censoring is achieved by summing over the mutually exclusive lifetime pathways leading to node D_k; the result of this is the expression given on the right side of (3.2). Such summing amounts to pruning the censoring states from the flowgraph so that only the transitions $1 \to \mathrm{D}_k$ for $k = 1, 2$ remain; the consequent flowgraph has a direct transmittance to state D_k which is the Laplace-Stieltjes transform for G_k given on the left side of (3.2). Thus,

$$\begin{aligned} \int_0^\infty e^{sx_k}\, dG_k(x_k) &= p_k M_k(s) + \int_0^\infty e^{sz}\, N_k(s; z)\, dQ_R(z) \qquad (3.2) \\ &= p_k M_k(s) + \int_0^\infty \left\{ \int_z^\infty e^{sx_k}\, dG_k(x_k) \right\} \frac{dQ_R(z)}{1 - G_1(z) - G_2(z)} \\ &= p_k M_k(s) + \int_0^\infty e^{sx_k} \left\{ \int_0^{x_k} \frac{dQ_R(z)}{1 - G_1(z) - G_2(z)} \right\} dG_k(x_k) \end{aligned}$$

for $k = 1, 2$. Equality (3.2) can also be developed without flowgraphs (see the Appendix), but the flowgraph approach provides a new understanding of competing risks from a systems theoretic perspective.

3.2.1 Self-Consistent Expressions for the Sub-cdf Functions

The Laplace-Stieltjes expressions in (3.2) provide solutions for G_1 and G_2 in terms of the competitive distributions F_1, F_2 and Q_R. Inversion of the Laplace-Stieltjes transforms in (3.2) gives

$$dG_k(x_k) = p_k\, dF_k(x_k) + \left\{ \int_0^{x_k} \frac{dQ_R(z)}{1 - G_1(z) - G_2(z)} \right\} dG_k(x_k)$$

or

$$dG_k(x_k) = \left\{ 1 - \int_0^{x_k} \frac{dQ_R(z)}{1 - G_1(z) - G_2(z)} \right\}^{-1} p_k\, dF_k(x_k), \tag{3.3}$$

for $k = 1, 2$.

Theorem 3.1 *The sub-cdf functions G_1 and G_2 specified in (3.3) are the unique self-consistent solution to the two equations in (3.2).*

Proof: First define $G(x) = G_1(x) + G_2(x)$. Then the two equations in (3.3) may be added together for common $x_1 = x_2 = x$ to get an equation with a unique solution for $dG(x)$ of the Kaplan-Meier form. Now define

$$G_{2:1}(z) = \frac{G_2(z)}{1 - G_1(z)} \quad \text{and} \quad F_{2:1}(x) = \frac{p_2 F_2(x)}{1 - G_1(x)},$$

etc. In (3.3), divide both sides of equation $k = 2$ by $1 - G_1(x)$ and, inside the curly braces, divide top and bottom of the integrand by $1 - G_1(z)$. Then (3.3) becomes

$$dG_{2:1}(x) = \left\{ 1 - \int_0^{x} \frac{dQ_{R:1}(z)}{1 - G_{2:1}(z)} \right\}^{-1} dF_{2:1}(x)$$

which provides a unique Kaplan-Meier solution for $G_{2:1}(x)$. Values for $G_{2:1}(x)$ and $G(x)$ uniquely determine the individual values of G_1 and G_2 at x. ∎

3.2.2 Mutually Independent Competing Risks

The assumption of independence between Z^0 and (X_1^0, X_2^0) allows for the determination of unique subsurvival functions G_1 and G_2 as the self-consistent solutions to (3.3). The stronger assumption of mutual independence among all three competing risks X_1^0, X_2^0, and Z^0 allows for the derivation of self-consistent expressions for F_1^0, F_2^0, and H^0 as the noncompetitive marginal distributions of X_1^0, X_2^0, and Z^0. In this case, sub-c.d.f. G_1 may be expressed as

$$dG_1(x_1) = dF_1^0(x_1)\left\{1 - F_2^0(x_1)\right\}$$

and similarly, $dG_2(x_2) = dF_2^0(x_2)\{1 - F_1^0(x_2)\}$, so that

$$1 - G_1(z) - G_2(z) = \Pr\left(X_1^0 > z, X_2^0 > z\right) = \left\{1 - F_1^0(z)\right\}\left\{1 - F_2^0(z)\right\}.$$

In terms of F_1^0 and F_2^0, the self-consistent expressions in (3.3) are

$$dF_1^0(x_1) = \left[1 - \int_0^{x_1} \frac{dQ_R(z)}{\{1 - F_1^0(z)\}\{1 - F_2^0(z)\}}\right]^{-1} \frac{p_1\, dF_1(x_1)}{1 - F_2^0(x_1)} \tag{3.4}$$

$$dF_2^0(x_2) = \left[1 - \int_0^{x_2} \frac{dQ_R(z)}{\{1 - F_1^0(z)\}\{1 - F_2^0(z)\}}\right]^{-1} \frac{p_2\, dF_2(x_2)}{1 - F_1^0(x_2)}. \tag{3.5}$$

These expressions provide an understanding for the successive removal of risk factors under mutual independence. First remove factor 2 from factor 1 and censoring. Write

$$\begin{aligned} \frac{p_1\, dF_1(x_1)}{1 - F_2^0(x_1)} &= \frac{\Pr\{X_1^0 \in (x_1 - dx_1, x_1], X_1^0 = W\}}{\Pr\{X_2^0 > x_1\}} \\ &= \Pr\left\{X_1^0 \in (x_1 - dx_1, x_1], X_1^0 < Z^0\right\} \\ &= p_{1:2}\, dF_{1:2}(x_1), \end{aligned} \tag{3.6}$$

which is the sub-c.d.f. distribution for risk factor 1 with factor 2 removed as indicated through the notation 1:2. Likewise

$$dQ_{R:2}(z) := \frac{dQ_R(z)}{1 - F_2^0(z)} \tag{3.7}$$

is the censoring sub-c.d.f. with risk factor 2 removed. With risk factor 2 removed from factor 1 and censoring, we now remove the censoring risk factor from factor 1. Substituting these results into expression (3.4) leads to the following result.

Lemma 3.1 *In the independent competing risks model in which X_1^0, X_2^0, and Z^0 are mutually independent, F_1^0 is the unique self-consistent solution to*

$$dF_1^0(x_1) = \left\{1 - \int_0^{x_1} \frac{dQ_{R:2}(z)}{1 - F_1^0(z)}\right\}^{-1} p_{1:2}\, dF_{1:2}(x_1). \tag{3.8}$$

This expression relates F_1^0 to $Q_{R:2}$ and $F_{1:2}$ in the same manner as F_1^0 is related to Q_R and F_1 when risk factor 2 has been excluded. Thus we have a sequential removal of risk factors: factor 2 is removed from factor 1 and censoring in the computation of $p_{1:2}\, dF_{1:2}(x_1)$ and $dQ_{R:2}(z)$; then censoring is removed from $p_{1:2}\, dF_{1:2}(x_1)$ using $dQ_{R:2}(z)$ as in (3.8). With the independence of all risk factors, these operations are completely analogous to way in which partial correlations are determined through successive removal of independent variable effects using linear regression.

3.2.3 Self-Consistent Estimators with Independent Competing Risks

Data are observed according to the independent competing risks model and are expressed in a form that reflects the flowgraph model. Let $x_{11} \le \cdots \le x_{1n_1}$ and $x_{21} \le \cdots \le x_{2n_2}$ be the observed death times due to independent risk factors 1 and 2, respectively. An estimate of p_k is $\hat{p}_k = n_k/(n_1 + n_2 + r)$, for $k = 1, 2$. The empirical c.d.f.s of $\{x_{1i}\}$ and $\{x_{2i}\}$, denoted by $\hat{F}_1$ and $\hat{F}_2$, estimate the competitive death time distributions F_1 and F_2. The empirical m.g.f.s $\hat{M}_1(s)$ and $\hat{M}_2(t)$ are the Laplace-Stieltjes transforms of $\hat{F}_1(x_1)$ and $\hat{F}_2(x_2)$.

The observed censoring times are denoted as $z_1 \leq \cdots \leq z_r$ with empirical c.d.f. $\hat{H}$, an estimate of the competitive distribution H. The empirical estimate of $Q_R(z)$ is $\hat{Q}_R(z) = (1 - \hat{p}_1 - \hat{p}_2)\, \hat{H}(z)$.

Empirical versions of the self-consistency equations in (3.4) and (3.5) are obtained by placing a hat on each of the c.d.f.s involved. These may be solved directly but a more elegant approach is to sequentially remove risk factors as previously described for the population in (3.8). Let $\hat{F}_k^0(x_k)$ be the Kaplan-Meier estimate of the c.d.f. for risk factor $k = 1, 2$ computed by treating as censored data the pooled observations from the other risk factor and actual censoring. Then

$$\hat{p}_{1:2}\, d\hat{F}_{1:2}(x_1) := \frac{\hat{p}_1\, d\hat{F}_1(x_1)}{1 - \hat{F}_2^0(x_1)} \qquad \text{and} \qquad d\hat{Q}_{R:2}(y) := \frac{d\hat{Q}_R(z)}{1 - \hat{F}_2^0(z)} \tag{3.9}$$

respectively define partially censored Kaplan-Meier estimates that are the sub-c.d.f. functions for the competition of factor 1 and censoring but with factor 2 removed. The expressions in (3.9) are substituted into the empirical versions of (3.4) and (3.5) to give $\hat{F}_1^0$ and $\hat{F}_2^0$ as the self-consistent solutions to

$$d\hat{F}_1^0(x_1) = \left[1 - \int_0^{x_1} \frac{d\hat{Q}_{R:2}(z)}{1 - \hat{F}_1^0(z)}\right]^{-1} \hat{p}_{1:2}\, d\hat{F}_{1:2}(x_1) \tag{3.10}$$

$$d\hat{F}_2^0(x_2) = \left[1 - \int_0^{x_2} \frac{d\hat{Q}_{R:1}(z)}{1 - \hat{F}_2^0(z)}\right]^{-1} \hat{p}_{2:1}\, d\hat{F}_{2:1}(x_2)\,. \tag{3.11}$$

This is summarized as follows.

Theorem 3.2 *Suppose that $\hat{F}_1$, $\hat{F}_2$ and $\hat{Q}_R$ share no common jump points and let $z^* = \max(x_{1n_1}, x_{2n_2}, z_r)$. Then the Kaplan-Meier c.d.f. estimators $\hat{F}_1^0(x)$ and $\hat{F}_2^0(x)$ are the unique self-consistent solutions over $x \in [0, z^*)$ to empirical flowgraph equations that are the empirical versions of (3.4) and (3.5).*

Estimators $\hat{F}_1^0(x)$ and $\hat{F}_2^0(x)$ were first suggested by Hoel (1972) and were derived as nonparametric m.l.e.s by Lagakos et al. (1978). Dinse and Larson (1986) clarified their relationship to Kaplan-Meier estimators. The novelty of this theorem is that it provides the framework within which these same estimators may be interpreted as the unique self-consistent estimators.

Computations for $\hat{F}_1^0(x)$ and $\hat{F}_2^0(x)$ in the presence of ties such as $x_1 = x_2 = y$ require the use a convention such as $x_2 = x_1^+$ and $y = x_2^+$, as used in Lagakos et al. (1978). The values of $\hat{F}_1^0(x)$ and $\hat{F}_2^0(x)$ for $x \geq z^*$ conform to the conventions specified in Chapter 2. For example, if $z^* = x_{1n_1}$, then $\hat{F}_1^0(x) = 0$ and $\hat{F}_2^0(x)$ and $\hat{H}^0(x)$ are indeterminate over $x \geq z^*$. The support of $\hat{F}_1^0(x)$ is over $x \in \{x_{1i}\}$ with masses

$$d\hat{F}_1^0(x_{1i}) = \left[1 - \int_0^{x_{1i}} \frac{d\hat{Q}_R(z)}{\{1-\hat{F}_1^0(z)\}\{1-\hat{F}_2^0(z)\}}\right]^{-1} \frac{\hat{p}_1\, d\hat{F}_1^0(x_{1i_1})}{1-\hat{F}_2^0(x_{1i_1})} \tag{3.12}$$

$$= \left[1 - \int_0^{x_{1i}} \frac{d\hat{Q}_{R:2}(z)}{1-\hat{F}_1^0(z)}\right]^{-1} \hat{p}_{1:2}\, d\hat{F}_{1:2}^0(x_{1i}) \tag{3.13}$$

that add to one. The support of $\hat{F}_2^0(x)$ and $\hat{H}^0(z)$ are $\{x_{2i}\}$ and $\{z_j\}$ respectively and the masses in each case add to < 1. Estimators $\hat{F}_2^0$ and $\hat{H}^0$ are left-continuous at z^* and indeterminate above z^*. The three risk factors involved are treated in an exchangable manner so the self-consistent estimators are analogous in the other settings in which $z^* = x_{2n_2}$ or z_r.

3.2.4 Self-Consistent Estimators for the Sub-c.d.f. Functions $G_1(x)$ and $G_2(x)$

Suppose that X_1^0 and X_2^0 are not independent as previously assumed but suppose that censoring time Z^0 remains independent of (X_1^0, X_2^0). Then self-consistent estimators of the sub-c.d.f. functions solving (3.3) can be determined which provide the unique solution to the empirical versions of (3.3).

The empirical semi-Markov version of the flowgraph in Figure 3.1 is shown in Figure 3.2. Each censored value z_j indexes its own censoring node in Figure 3.2 as

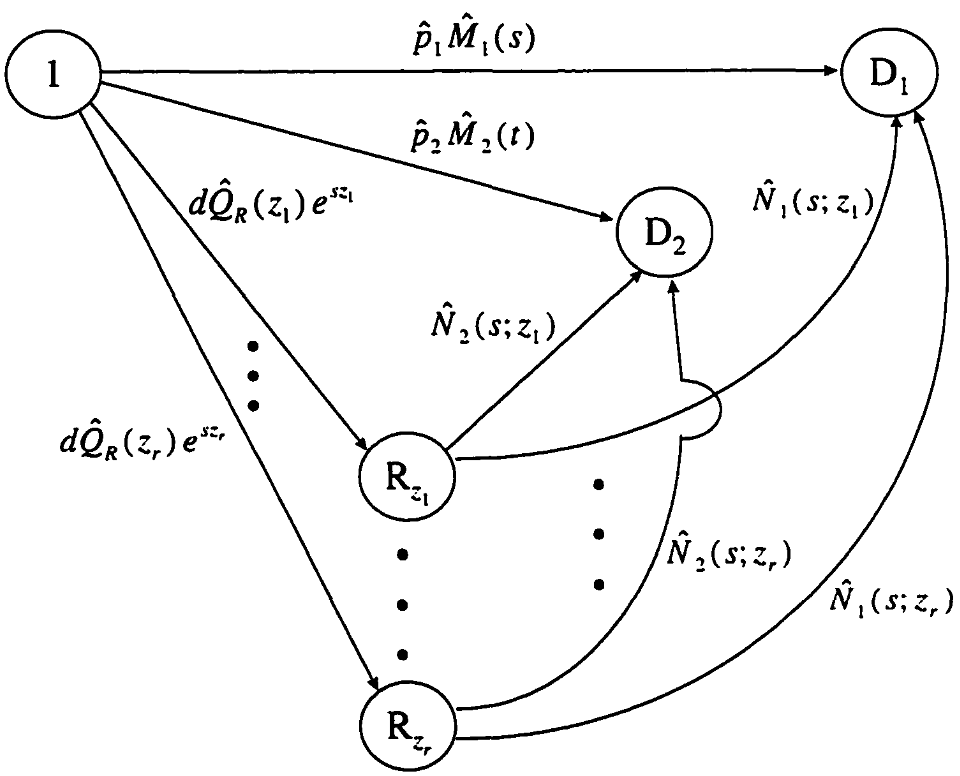

Figure 3.2 Empirical flowgraph for the competing risks model.

well as theremaining transmittances $R_{z_j} \to D_1$ and $R_{z_j} \to D_2$ given by

$$\hat{N}_k(s; z_j) = \frac{e^{-sz_j}}{1 - \hat{G}_1(z) - \hat{G}_2(y)} \int_{z_j}^{\infty} e^{sx_k} d\hat{G}_k(x_k), \quad \text{for} \quad k = 1, 2.$$

The empirical flowgraph in Figure 3.2 is a semi-Markov system that estimates the semi-Markov system in Figure 3.1 and is solved in exactly the same manner.

Theorem 3.3 *Suppose the distributions $\hat{F}_1$, $\hat{F}_2$ and $\hat{Q}_R$ have no common jump points. Then unique sub-c.d.f.s $\hat{G}_1(x_1)$ and $\hat{G}_2(x_2)$ exist over $x_k \in [0, z^*)$ as the self-consistent solutions to the empirical flowgraph equations*

$$\int_0^{\infty} e^{sx_k} d\hat{G}_k(x_k) = \hat{p}_k \int_0^{\infty} e^{sx_k} d\hat{F}_k(x_k)$$
$$+ \int_0^{\infty} e^{sx_k} \left\{ \int_0^{x_k} \frac{d\hat{Q}_R(z)}{1 - \hat{G}_1(z) - \hat{G}_2(z)} \right\} d\hat{G}_k(x_k)$$

for $k = 1, 2$. The support of $\hat{G}_k$ is $\{x_{ki} : i = 1, \ldots n_k\}$ with mass

$$d\hat{G}_k(x_{ki}) = \left\{ 1 - \int_0^{x_{ki}} \frac{d\hat{Q}_R(z)}{1 - \hat{G}_1(z) - \hat{G}_2(z)} \right\}^{-1} \hat{p}_k d\hat{F}_k(x_{ki}) \qquad (3.14)$$

for $k = 1, 2$.

Proof: The proof is the same as given for the population analogues. ∎

The Laplace-Stieltjes transforms of $\hat{G}_1$ and $\hat{G}_2$ are the empirical transmittances to states D_1 and D_2 in the censor-free system that has the censored states $\{R_{z_j} : j = 1, \ldots, r\}$ pruned. The next result simplifies their calculation by relating $\hat{G}_1$, $\hat{G}_2$ to the Kaplan-Meier estimates $\hat{F}_1^0$ and $\hat{F}_2^0$. A proof by induction is given in the Appendix.

Theorem 3.4 *Suppose data are observed under the competing risks model in which X_1^0 and X_2^0 may be dependent but (X_1^0, X_2^0) is independent of Z^0. Then unique self-consistent estimates of the sub-c.d.f.s exist and are*

$$d\hat{G}_1(x_1) = \{1 - \hat{F}_2^0(x_1)\} d\hat{F}_1^0(x_1) \qquad (3.15)$$
$$d\hat{G}_2(x_2) = \{1 - \hat{F}_1^0(x_2)\} d\hat{F}_2^0(x_2),$$

for all x_1 and x_2.

These relationships furthermore demonstrate that data from such a competing risks model contain no information about the dependence of the competing random variables X_1^0 and X_2^0. This fact has been pointed out by many authors including Tsiatis (1975). Reconsider the idea of *empirical proxy* models and suppose that random variables $(\hat{X}_1^0, \hat{X}_2^0, \hat{Z}^0)$ have c.d.f. $\hat{P}(x_1, x_2, z)$. Then $\hat{P}$ is an *empirical proxy* model if it leads to the empirical distributions $\hat{F}_1(x_1)$, $\hat{F}_2(x_2)$ and $\hat{H}(z)$ of the data when the components are viewed competitively. One of many such empirical proxies is the independent proxy model having

$$\hat{P}(x_1, x_2, y) = \hat{F}_1^0(x_1)\, \hat{F}_2^0(x_2)\, \hat{H}^0(z) \tag{3.16}$$

where $\hat{F}_1^0$, $\hat{F}_2^0$ and $\hat{H}^0$ are the Kaplan-Meier c.d.f. estimates. Estimates $\hat{G}_1$ and $\hat{G}_2$ in (3.15) are calculated as if (3.16) were its underlying proxy model. Thus these estimates are entirely compatible with the presumption that the underlying joint empirical distribution has independent components.

Now suppose that state 1 in Figure 3.2 is an arbitrary state of a semi-Markov system subject to censoring. For the corresponding censor-free system, the empirical transmittances out of state 1 are the Laplace-Stieltjes transforms of $\hat{G}_1$ and $\hat{G}_2$. These tranmittances are easily computed from their Kaplan-Meier estimates using the relationship in (3.15). If the largest observation is a death time, then the sum of the transition probabilities in the censor-free system is equal to one, e.g.,

$$\widehat{\Pr}(1 \to \mathrm{D}_1) + \widehat{\Pr}(1 \to \mathrm{D}_2) := \hat{G}_1(\infty) + \hat{G}_2(\infty) = 1. \tag{3.17}$$

If, however, the largest time is a censored time, then $\hat{G}_1(x_1)$ and $\hat{G}_2(x_2)$ are indeterminate for $x_1, x_2 > z^*$ so that the transition probabilities are also indeterminate.

The practical solution to such indeterminacy is to take

$$\hat{p}_{1:R} := \widehat{\Pr}(1 \to D_1) = \frac{\hat{G}_1(z^*)}{\hat{G}_1(z^*) + \hat{G}_2(z^*)} = 1 - \widehat{\Pr}(1 \to D_2)$$

and use Laplace-Stieltjes transforms

$$\hat{p}_{k:R}\hat{M}_{k:R}(s) := \frac{1}{\hat{G}_1(z^*) + \hat{G}_2(z^*)} \int_0^\infty e^{sx_k} d\hat{G}_k(x_k) \qquad k = 1, 2 \tag{3.18}$$

as the 1-step transmittances in the censor-free system. Here, the $: R$ notation indicates that the transmittance has the censoring factor removed and is associated with the censor-free system. This practical solution to the indeterminacy problem is used in the next section with the obvious extensions required when the competing risk consists of more than two types of death or transition.

Expression (3.18) summarizes computation for the transmittance to state D_k in the censor-free system using data derived from the competing risks setting with censoring. The expression is nonparametric and makes no distributional assumptions about the censoring other than its independence from (X_1^0, X_2^0). The additional assumption that censoring times are Exponential (λ) leads to simple alternative semi-parametric expressions for determining these transmittances.

Theorem 3.5 *Suppose that $p_k M_k(s)$ is the transmittance to state D_k in competition with censoring. Then*

$$p_{k:R} M_{k:R}(s) = p_k M_k(s + \lambda)$$

is the transmittance to state D_k in the censor-free system. Here, $p_{k:R} = p_k M_k(\lambda)$ and $M_{k:R}(s) = M_k(s + \lambda)/M_k(\lambda)$ is the m.g.f. of the λ-tilted distribution for the competitive holding time with censoring risk removed.

Proof: The transmittance from $1 \to D_k$ in the censor-free system is

$$\begin{aligned}
p_{k:R}M_{k:R}(s) &= \int_0^\infty e^{sx} dG_k(x) = \int_0^\infty e^{(s+\lambda)x} \left\{ \int_x^\infty \lambda e^{-\lambda z} dz \right\} dG_k(x) \\
&= \int_0^\infty \lambda e^{-\lambda z} \left\{ \int_0^z e^{(s+\lambda)x} dG_k(x) \right\} dz \\
&= E^{Z^0} \left[\int_0^{Z^0} e^{(s+\lambda)x} dG_k(x) \right] = E\left[e^{(s+\lambda)X_k^0} 1\left\{ X_k^0 = W \right\} \right] \\
&= p_k M_k(s+\lambda).
\end{aligned}$$

■

With data observed according to the competing risk model, estimation of transmittance $p_{k:R}M_{k:R}(s)$ proceeds with the parametric estimation of λ and the nonparametric estimation of $p_k M_k$ to give

$$\hat{p}_{k:R}\hat{M}_{k:R}(s) = \hat{p}_k \hat{M}_k(s+\hat{\lambda}). \tag{3.19}$$

This semi-parametric approach to estimation could be used in conjunction with a semi-parametric bootstrap to determine semi-parametric confidence envelopes for the system survival function. However, we choose to focus on the totally nonparametric approach that uses the nonparametric bootstrap.

3.3 Bootstrapping System Survival Times

Suppose T is the survival (first passage) time for a patient passing from state $1 \to D$, the death state, in a complicated semi-Markov system. The data are censored however T refers to a passage time in the censor-free system. We now demonstrate the computation of bootstrap confidence envelopes for the survival and hazard functions of T using censor-free transmittances estimated from censored data.

Assume that a patient can be censored from any state of the initial semi-Markov system. To ensure that the system retains its semi-Markov property, we must also assume that after each state change a new censoring risk begins from time 0 to compete with the other possible transitions from the new state. This implies that each transient state of the system must have it's own continuum of censoring nodes if we wish to model passage times through the system subject to censoring. Under such circumstances, the departure out of each transient state is the competing risk setting discussed in the previous section.

The removal of censoring risk from the system is effected through its separate removal from each of the transient states of the system. The estimated transmittance matrix $\hat{Q}$ of the censor-free system now has $(i, j)^{th}$ entry $\hat{p}_{ij:R}\hat{M}_{ij:R}(s)$ computed as described in (3.18), i.e.

$$\hat{p}_{ij:R}\hat{M}_{ij:R}(s) := \frac{1}{\sum_k \hat{G}_{ik}(z_i^*)} \int_0^\infty e^{sx} d\hat{G}_{ij}(x). \tag{3.20}$$

where z_i^* is the largest holding time in state i. Note that the summation in the denominator of (3.20) is 1 except when z_i^* is a censoring time, in which case it is < 1. Here, $\hat{G}_{ij}$ is the censor-free sub-c.d.f. given in Theorem 3.4 as

$$d\hat{G}_{ij}(x) = d\hat{F}_{ij}^0(x) \prod_{k \neq j} \left\{1 - \hat{F}_{ik}^0(x)\right\},$$

where $\hat{F}_{ij}^0$ is the Kaplan-Meier estimate of the noncompetitive transmittance from $i \to j$ considering all other transitions out of state i as censored values. Recall that we are not assuming the complete independence of the transmittances out of a transient state; Theorem 3.4 merely implies we should calculate $\hat{G}_{ij}$ as if they were. The collection of sub-c.d.f. estimates gives empirical transmittance matrix $\hat{Q} = \{\hat{p}_{ij:R}\hat{M}_{ij:R}(s)\}$.

3.3.1 Nonparametric Bootstrap

Consider how the data might be use to resample T^* as a passage time through the system subject to censoring: Pass from initial state 1 to terminal state D and sum the holding times that are resampled while transcending transient states in the pathway in the following way. From transient state i, suppose there are $n_{i\cdot}$ observed transitions out of i including censoring. Pick one of these at random. Follow it to its associated destination if it is not a censored value. If censored however, randomly pick one of the $n_{i\cdot}$ values that is larger; continue doing this until a non-censored value is resampled. Should this scheme ultimately fail to select a noncensored value, because z_i^* is censored, then restart the random selection from scratch using all $n_{i\cdot}$ transitions. This mechanism describes the redistribute to the right algorithm introduced by Efron (1967) and further discussed in Miller (1981, §3.2.1). The situation in which the selection ends with z_i^* and must be restarted, is equivalent to the normalization in (3.20) to assure that $\sum_j \hat{p}_{ij:R} = 1$. Thus this resampling scheme which accounts for censoring is equivalent to direct resampling from the estimated sub-c.d.f.s $\{\hat{G}_{ij} : i \to j\}$ computed as Kaplan-Meier estimates and whose Laplace-Stieltjes transforms comprise the ith row of $\hat{Q}(s)$. This leads to the following equivalence.

Theorem 3.6 *Suppose T^* is a passage time resampled as just described. Then, the m.g.f. of T^* is given by the first passage cofactor rule in Theorem 1 of Butler (2000) when applied to the system transmittance* $\hat{Q}(s) = \left\{ \hat{p}_{ij:R} \hat{M}_{ij:R}(s) \right\}$.

Proof: The proof that each competing risk resample has a distribution determined by the estimated sub-c.d.f.s $\{\hat{G}_{ij} : i \to j\}$ follows as a simple adaptation of the proof in Miller (1981, pp. 191-193) to account for several transitions out of state i rather than a single death transition. ∎

The main consequence of Theorem 3.6 is that we need not resample at all in order to use the single bootstrap. The analytic computation of a saddlepoint inversion, using the aforementioned cofactor rule applied to $\hat{Q}$, yields a saddlepoint approximation to the bootstrap distribution. Such inversion provides a point estimate of the survival and hazard function of T.

The double bootstrap provides an ensemble of survival and hazard estimates which lead to confidence envelopes to accompany the point estimates. The idea is to implement some form of outer sampling to construct 999 resampled values of the empirical system transmittance $\{\hat{Q}_i^*(s) : i = 1, \ldots, 999\}$. Then, 999 inversions of these transmittances along with the point estimate provide 1000 replicates of the survival distribution from which we determine confidence envelopes with the percentile or BC_a method.

Each $\hat{Q}^*$ is obtained by using the logical resampling extension of Efron's (1981) scheme to competing risks. For transitions out of state i, resample $n_{i\cdot}$ times with replacement from all $n_{i\cdot}$ holding times. If any of the resulting resample counts to noncensored states are zero, then discard the resample. The difficulty with trying to use such resamples is that they may alter the system structure by creating irrelevant states in system $\hat{Q}^*$ that are relevant in the system $\hat{Q}$ and therefore relevant to the true underlying system. Thus the use of systems $\hat{Q}^*$ with resampled zeros is not representative of what we know about the true underlying system and should not be allowed in the outer resampling. See Butler (2000, §4) for further discussion on the irrelevance of system states.

Using this outer resampling scheme, the resampled sub-c.d.f. estimates are computed as

$$d\hat{G}_{ij}^*(x) = d\hat{F}_{ij}^{0*}(x) \prod_{k \neq j} \left\{1 - \hat{F}_{ik}^{0*}(x)\right\}.$$

The Laplace-Stieltjes transforms of these sub-c.d.f.s as in (3.20) gives the resampled transmittance matrix

$$\hat{Q}^* = \left\{\hat{p}_{ij:R}^* \hat{M}_{ij:R}^*(s)\right\}.$$

Suppose that T^{**} is a system survival time simulated from system $\hat{Q}^{*}$ using the same algorithm as was used in determining T^{*} from system $\hat{Q}$. Then the values of T^{**} represent the second inner layer of resampling nested within the outer resample of the double bootstrap. This inner layer of resampling is unnecessary, however, and can be replaced with saddlepoint approximation as we now indicate.

Theorem 3.7 *Suppose T^{**} is a passage time simulated from the algorithm above as applied to the raw resampled data underlying the system $\hat{Q}^{*}(s)$. Then, the m.g.f. of T^{**} is given by the first passage cofactor rule in Theorem 1 of Butler (2000) when applied to the system transmittance $\hat{Q}^{*}(s)$.*

3.3.2 Numerical Example

Fix and Neyman (1951) introduced the system shown in the flowgraph of Figure 3.3. Their goal was to model the survival time of a patient under treatment for a disease such as cancer. Once diagnosed, the patient enters the system at state 1. Upon leaving diagnosis state 1, a patient may feedback into state 1, pass to state 2 implying apparent recovery from cancer, pass to state D implying death from cancer, or pass to state R_1 implying censoring and therefore either death from a cause other than cancer or loss from observation. States 1 and 2 are transient states and the system contains 3 feedback loops to : $1 \rightarrow 1$, $1 \rightarrow 2 \rightarrow 1$, and $2 \rightarrow 2$. State R_2 denotes censoring from state 2. Only state R_1 is not a part of the original Fix-Neyman system, and we have added it to consider censoring from all transient states.

Figure 3.3 is the population version of the semi-Markov system subject to censoring risk. The data are the complete patient histories describing the sequence of states and holding times during their passage from state 1 to D, R_1, or R_2.

Figure 3.4 shows the population flowgraph for the corresponding censor-free system. The goal is to provide confidence envelopes for the survival and hazard rate

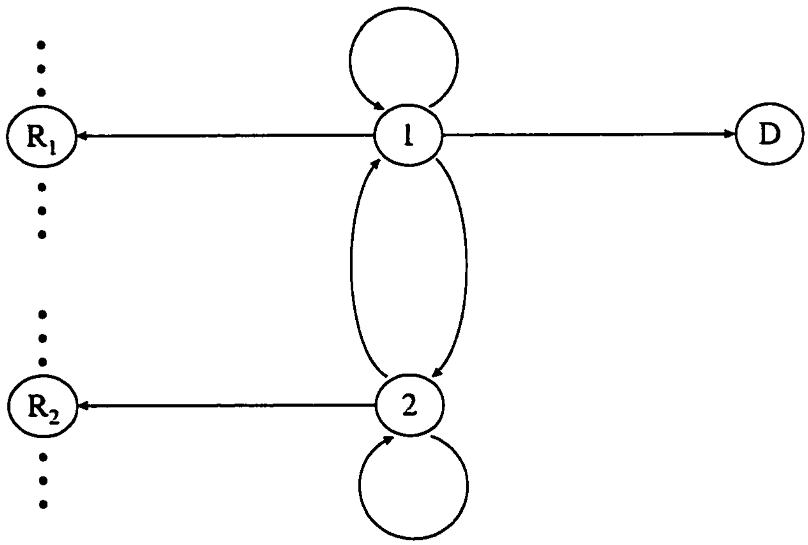

Figure 3.3 The observed flowgraph for the Fix-Neyman model.

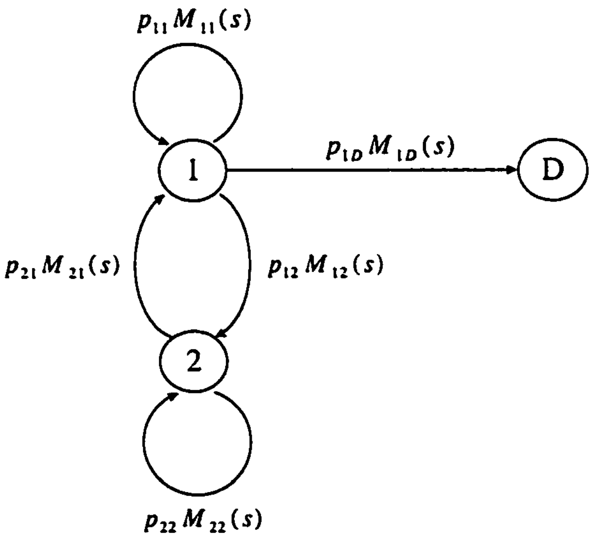

Figure 3.4 The Fix-Neyman flowgraph with the censoring risk factors removed.

functions describing T, the first passage time from $1 \to D$ in this censor-free system.

As a specific example, we simulated 100 histories for passage through Figure 3.3 by performing simulations related to the censor-free system in Figure 3.4 as described below. For the latter system assume the following transmittance matrix with states ordered as 1, 2, and D:

$$\mathcal{Q}(s) = \begin{pmatrix} 0.3\,\mathrm{ig}\,(10.5, 11.7) & 0.3\,\mathrm{r}\,(17.7) & 0.4\,\mathrm{r}\,(22.2) \\ 0.5\,\mathrm{r}\,(13.3) & 0.5\,\mathrm{ig}\,(11.0, 8.8) & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.21}$$

Here, $\mathrm{ig}(a, b)$ is the inverse Gaussian m.g.f. with mean a and standard deviation b, and $\mathrm{r}(a)$ is the Raleigh m.g.f. with mean a. The independent censoring distributions out of states 1 and 2 are Exponential(180) and Exponential(72), respectively. To obtain histories describing passage through the system in Figure 3.3, we consider passage through Figure 3.4 but, at each state transition, require that the state transition compete with independent censoring. For example, in leaving state 1, we determine the censor-free destination state as the outcome of a Bernoulli trial with probabilities (0.3, 0.3, 0.4), simulate its corresponding holding time, and then compare this time with an independently generated Exponential (180). If the censoring time is smaller, the destination in Figure 3.3 is R_1; otherwise it is 1, 2, or D as determined by the Bernoulli trial.

For the system in Figure 3.4, exact computation of the survival distribution of T is very difficult. However, the cofactor rule in Theorem 1 of Butler (2000) may be used with $\mathcal{Q}$ in (3.21) to determine a very accurate saddlepoint approximation. Table 3.1 provides right-tail percentiles of the saddlepoint approximation in the "Exact" column for the various tail probabilities listed to its left. Estimates of these percentiles are provided and computed from the saddlepoint inversion of $\hat{\mathcal{Q}}$ based on the 100 histories. The double bootstrap envelopes lead to the 90% BC_a confidence intervals listed for the various percentiles. Also given are the right boundaries of tolerance intervals of T that have 90% guarantee that the coverage is $1-$ Right Perc. For example, the interval $(0, 48.1)$ has been chosen to assure a 90% guarantee of 50% coverage for the values of T.

Table 3.1 Estimates, 90% BC_a confidence intervals and 90% Guaranteed Coverage Tolerance Intervals for the right-tail percentiles of T.

Right Perc.	"Exact"	Estimate	BC_a Lower	BC_a Upper	Guar. Tol.
0.50	42.0	41.8	36.4	50.7	48.1
0.25	82.8	78.3	67.1	97.3	97.1
0.10	137.8	127.8	107.8	159.4	153.4
0.05	179.4	165.4	137.9	207.1	197.2
0.01	276.4	252.7	208.0	317.6	303.6

Table 3.2 considers estimation of the exact mean and standard deviation for T as determined from its m.g.f. based on $\mathcal{Q}$. The estimates in the table have been based on $\hat{\mathcal{Q}}$ and confidence intervals have been determined from the resampled values $\{\hat{\mathcal{Q}}_i^*\}$. We also consider an i.i.d. model in order to see the benefit derived from using the semi-Markov structure of the survival time. This i.i.d. model treats the 100 censored survival times as derived from a single distribution subjected to independent censoring. Accordingly, the Kaplan-Meier survival function estimate $\hat{S}(t)$ has been calculated and the mean and standard deviation have been computed directly from it as

$$\hat{\mu} = -\int_0^\infty \hat{S}(t)dt \qquad \text{and} \qquad \hat{\sigma}^2 = 2\int_0^\infty t\hat{S}(t)dt - \hat{\mu}^2. \tag{3.22}$$

When the largest survival time is censored, the estimates in (3.22) are undefined and so we are forced to change this largest survival time to a death time (see Miller, 1981,

§3.4). This convention has the effect of biasing the mean and standard deviation estimates downward as seen in the table. Double bootstrap confidence intervals for the i.i.d. model are also biased downward.

Table 3.2 Estimates and 90% BC_a confidence intervals for the mean and standard deviation of T.

	Mean				Standard Deviation			
Method	Exact	Est.	BC_a Lower	BC_a Upper	Exact	Est.	BC_a Lower	BC_a Upper
Cofactor Rule	61.5	59.5	50.7	72.1	58.2	52.2	41.7	66.1
i.i.d. model		54.4	48.6	67.7		39.8	27.9	62.7

With only a single implementations of the double bootstrap in Table 3.2, we are not able to assess the coverage accuracy of the BC_a methods. To do this, we simulated 10,000 data sets and repeated the computations of Table 3.2, recording the observed coverage levels of the BC_a and the percentile confidence intervals. Table 3.3 presents these coverages with a target coverage of 0.9. The Cofactor Rule procedure using the semi-Markov structure achieves quite accurate coverage while the i.i.d. model exhibits severe undercoverage. The BC_a and Percentile coverages are similar for the Cofactor Rule procedure, but differ significantly for the i.i.d. model. The implication is that there are some severe deficiencies with the Percentile method when used with the i.i.d. model. The assumption that the largest survival time must be a death could be part of this deficiency. A potential solution to this problem would be to estimate the right-tail of the survival function using the Hill (1975) tail estimate rather than use our implicit truncation.

Recall from chapter one that the asymptotic hazard rate for survival time T is the right-edge of the convergence strip of the equivalent transmittance from $1 \to D$

Table 3.3 Coverage probabilities for the 90% confidence intervals illustrated in Table 3.2.

Bootstrap Method	Cofactor Rule		i.i.d. Model	
	Mean	Std. Dev.	Mean	Std. Dev.
BC_a	0.8942	0.9001	0.8105	0.5794
Percentile	0.8937	0.8944	0.7702	0.4759

where the equivalent transmittance is found by applying the Cofactor Rule to $\mathcal{Q}$. The true asymptotic hazard rate for this Fix-Neyman system is 0.0167. The estimate from the cofactor rule applied to $\hat{\mathcal{Q}}$ is 0.0185. A 90% BC_a confidence interval is (0.0146, 0.0227).

Figure 3.5 compares the "Exact" density, as determined by saddlepoint inversion based on $\mathcal{Q}$, to the bootstrap/saddlepoint estimate as determined by saddlepoint inversion of $\hat{\mathcal{Q}}$ using the 100 patient histories. Figures 3.6 and 3.7 compare the "Exact" survival and hazard rate functions to their bootstrap/saddlepoint estimates and 90% BC_a confidence bands. Figure 3.8 shows the relative error in estimation for the density, survival function and hazard rate function.

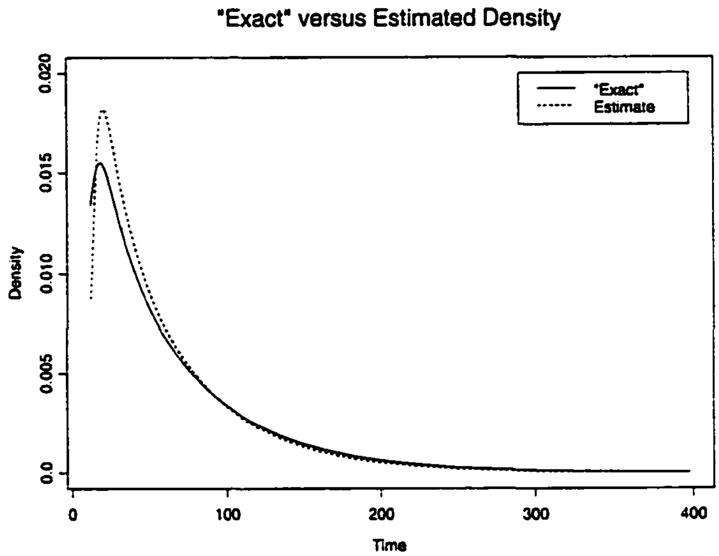

Figure 3.5 Density function for T.

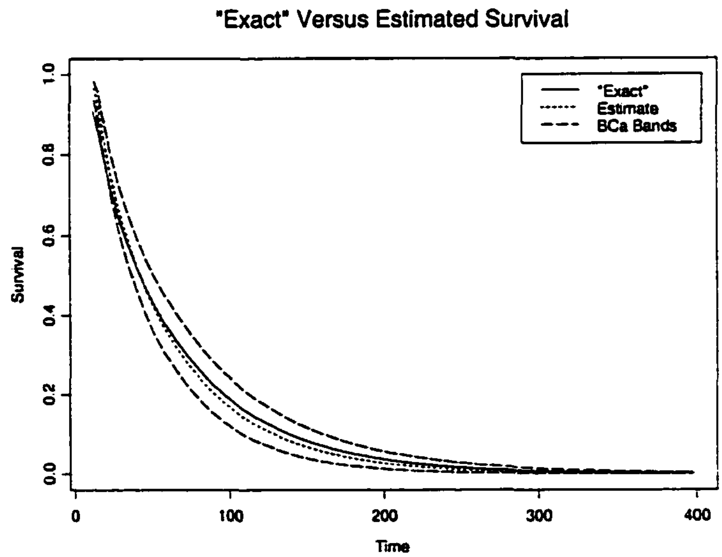

Figure 3.6 Survival function estimates and confidence bands.

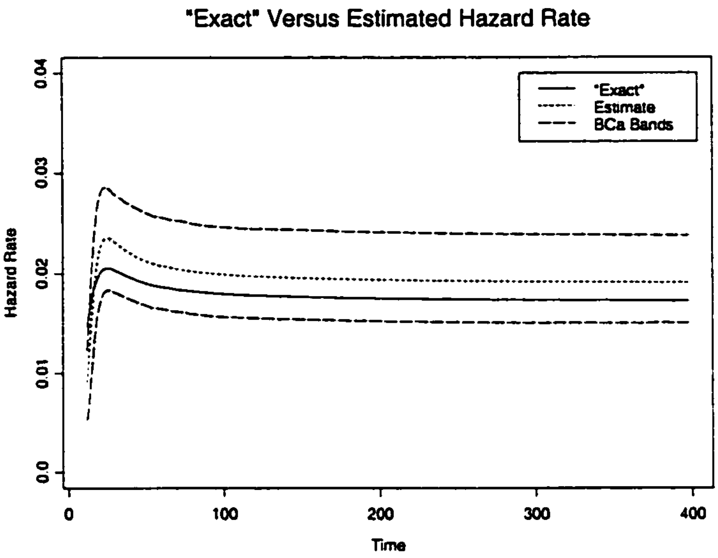

Figure 3.7 Hazard rate function estimates and confidence bands.

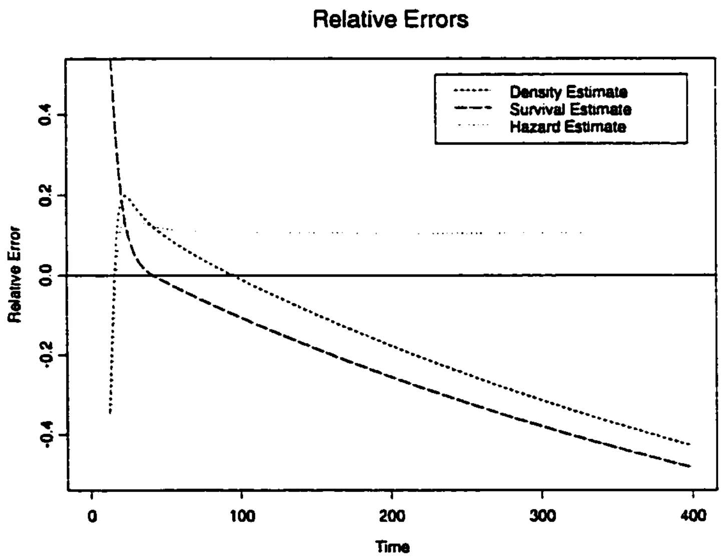

Figure 3.8 Relative errors in estimation.

A Appendix to Chapter 1

A.1 Calculation of the Exact First Passage Time Density and Distribution for T_5

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$(2, \beta)$ however, the computation becomes possible since a Gamma $(2, \beta)$ interarrival distribution can be expressed as the sum of two independent Exponential(β) variables. A new system, with about twice as many states, needs to be created which records the queue length at the end of each of these Exponential(β) variables. Such a system turns out to be Markov and the passage time T_5 for the $GI/M/1$ system is also a passage time in the new Markov system. As such, its distribution is computed exactly as a phase-type distribution (Aalen, 1995).

Consider passage time T_5 with G as a Gamma$(2, \beta)$ distribution and semi-Markov transmittance matrix

$$Q(s) = \begin{pmatrix} 0 & U_0(s) & 0 & 0 & 0 & 0 \\ 0 & T_{11}(s) & U_0(s-\lambda) & 0 & 0 & 0 \\ 0 & T_{21}(s) & U_1(s-\lambda) & U_0(s-\lambda) & 0 & 0 \\ 0 & T_{31}(s) & U_2(s-\lambda) & U_1(s-\lambda) & U_0(s-\lambda) & 0 \\ 0 & T_{41}(s) & U_3(s-\lambda) & U_2(s-\lambda) & U_1(s-\lambda) & U_0(s-\lambda) \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \text{(A.1)}$$

with states $0-5$.

We create the Markov process with ordered states $0, 0', 1, 1', \ldots, 4, 4', 5$ and transmittance matrix

$$Q_M(s) = \begin{pmatrix} 0 & M_1 & 0 & & & & & \\ 0 & 0 & M_1 & 0 & & & & \\ qM_2 & 0 & 0 & pM_2 & \ddots & & & \\ 0 & qM_2 & 0 & 0 & \ddots & 0 & & \\ & 0 & qM_2 & 0 & \ddots & pM_2 & 0 & \\ & & 0 & \ddots & \ddots & 0 & pM_2 & 0 \\ & & & \ddots & qM_2 & 0 & 0 & pM_2 \\ & & & & 0 & 0 & 0 & 0 \end{pmatrix} \tag{A.2}$$

where

$$M_1 = \frac{\beta}{\beta - s}, \quad M_2 = \frac{(\lambda + \beta)}{(\lambda + \beta) - s}$$

and

$$p = \frac{\beta}{\beta + \lambda}, \quad q = 1 - p.$$

Lemma A.1 *T_5, the first passage time from state $0 \to 5$ in the semi-Markov process (A.1) is also the first passage time from $0 \to 5$ in the Markov process (A.2).*

Proof: The semi-Markov system in (A.1) holds in state 0 for time Gamma $(2, \beta) = \text{Exp}_1(\beta) + \text{Exp}_2(\beta)$ before passing to state 1. Likewise the Markov system (A.2) holds in state 0 for time $\text{Exp}_1(\beta)$ before passing to state $0'$. Then, the Markov system holds in state $0'$ for time $\text{Exp}_2(\beta)$ before passing to state 1. Passage from $0 \to 1$ is therefore the same in both systems.

In state 1, the semi-Markov system waits time Gamma $(2, \beta) = \text{Exp}_1(\beta) + \text{Exp}_2(\beta)$. Should the job in queue have been fixed before time Gamma$(2, \beta)$, then the destination state is 0; otherwise it is state 1.

The Markov system, by comparison, records the same state changes, but does so in at least twice the number of steps by also recording the state of the queue after the first part of the interarrival time $\mathrm{Exp}_1(\beta)$. Starting in state 1, if the job in queue is fixed before time $\mathrm{Exp}_1(\beta)$, so that $\mathrm{Exp}(\lambda) < \mathrm{Exp}_1(\beta)$ with probability q, then the Markov system proceeds from $1 \to 0$ in time given by mgf M_2. From here, the remainder of the $\mathrm{Exp}_1(\beta)$ interarrival must be waited out. However, by the memoryless property of the exponential (by MPE), this waiting time is once again $\mathrm{Exp}_1(\beta)$ so it as if the system is starting anew at state 0 with another Gamma$(2, \beta)$ interarrival; such dynamics are those shown from states 0 and 1 in (A.2). Returning to the first part of the interarrival, consider the alternative event that $\mathrm{Exp}(\lambda) > \mathrm{Exp}_1(\beta)$. Then the Markov system proceeds from $1 \to 1'$ with probability p in time M_2. State $1'$ indicates that the queue length is 1 and unchanged midway through the Gamma$(2, \beta)$ interarrival time. From state $1'$, the second part of the interarrival $\mathrm{Exp}_2(\beta)$ competes with fixing time $\mathrm{Exp}(\lambda)$ by MPE. Should fixing not occur first, the Markov system goes to state 2 with probability p in time M_2 and the next Gamma$(2, \beta)$ interarrival is entertained to compete with $\mathrm{Exp}(\lambda)$ by MPE. Should fixing occur first, the system proceeds from $1' \to 0'$ in time M_2 and then must continue to wait out the remainder of the interarrival $\mathrm{Exp}_2(\beta)$ by MPE before passing from $0' \to 1$. The dynamics of these passages are captured in the transmittance in (A.2).

A description of the dynamics of passage from states $i = 2, 3, 4$ or $i = 2', 3', 4'$ in the Markov system is similar and is also captured in the transmittance (A.2). Passage from $i \to i'$ occurs when $\mathrm{Exp}_1(\beta) < \mathrm{Exp}(\lambda)$; by MPE, the fixing time starts anew again as $\mathrm{Exp}(\lambda)$ which now competes with $\mathrm{Exp}_2(\beta)$ to determine the next system state. Should $\mathrm{Exp}(\lambda) > \mathrm{Exp}_2(\beta)$, then $i' \to i+1$ with probability p in time M_2 but otherwise $i' \to (i-1)'$, with probability q in time M_2, wherein the remainder of the $\mathrm{Exp}_2(\lambda)$, which is again $\mathrm{Exp}_2(\lambda)$ by MPE, is waited out in competition with $\mathrm{Exp}(\lambda)$ once again. These dynamics appear in the transmittance matrix (A.2) which shows that the system is Markov. ∎

The infinitesimal generator for this Markov system is

$$
\mathcal{I} = \begin{pmatrix}
-\beta & \beta & 0 & & & & & \\
0 & -\beta & \beta & \ddots & & & & \\
\lambda & 0 & -(\lambda+\beta) & \ddots & & & & \\
0 & \lambda & 0 & \ddots & & 0 & & \\
 & 0 & \lambda & \ddots & & \beta & 0 & \\
 & & 0 & \ddots & 0 & -(\lambda+\beta) & \beta & 0 \\
 & & & \ddots & \lambda & 0 & -(\lambda+\beta) & \beta \\
 & & & & 0 & 0 & 0 & -(\lambda+\beta)
\end{pmatrix}
$$

The survival function of T_5 in this Markov system is the probability that the system, starting in state 0 at time 0, has not arrived at absorbing state 5 by time t. Suppose $p_{ij}(t)$ is the probability of being in state j at time t having started in state i at time 0. Then

$$\mathcal{P}(t) := (p_{ij}(t)) = \exp(\mathcal{I}t).$$

The survival function of T_5 is now

$$\bar{F}(t) = \sum_{k \in S} p_{0k}(t)$$

where $S = \{0, 0', 1, 1', 2, 2', 3, 3', 4, 4'\}$.

B Appendix to Chapter 2

B.1 Indeterminacy for the Self-Consistent Estimator

For the setting $z_r = z^*$ we first show that $\hat{F}^0(z_r) = \hat{F}^0(z_r^-)$. In this case, expression (2.10) at $x = z_r$ is

$$1 - \hat{F}^0(z_r) = \frac{1}{n+r}\sum_{k=1}^{r}\frac{1-\hat{F}^0(z_r)}{1-\hat{F}^0(z_k)}$$
$$= \frac{1}{n+r}\left(\sum_{k=1}^{r-1}\frac{1-\hat{F}^0(z_r)}{1-\hat{F}^0(z_k)} + 1\right). \qquad \text{(B.1)}$$

Thus, solving (B.1),

$$1 - \hat{F}^0(z_r) = \left\{(n+r) - \sum_{k=1}^{r-1}\frac{1}{1-\hat{F}^0(z_k)}\right\}^{-1}. \qquad \text{(B.2)}$$

This is also the value of $1 - \hat{F}^0(z_r^-)$ as may also be derived from (2.10).

The indeterminacy of $\hat{F}^0(x)$ for any $x > z_r = z^*$ also follows directly from (2.10) as

$$1 - \hat{F}^0(x) = \left\{1 - \hat{F}^0(x)\right\}\frac{1}{n+r}\sum_{k=1}^{r}\frac{1}{1-\hat{F}^0(z_k)}$$
$$= \left\{1 - \hat{F}^0(x)\right\}\frac{1}{n+r}\left\{\sum_{k=1}^{r-1}\frac{1}{1-\hat{F}^0(z_k)} + \frac{1}{1-\hat{F}^0(z_r)}\right\}. \qquad \text{(B.3)}$$

Substituting expression (B.2) into the latter term of (B.3) reduces the right hand expression to $\{1-\hat{F}^0(x)\}$. This value is therefore completely arbitrary and the reduction demonstrates its indeterminacy from the self-consistency equation.

B.2 Proof that the Berliner-Hill estimate is self-consistent

Let $c(k)$ be the number of right-censored observations in $[x_k, x_{k+1})$ for $k = 0, \ldots, n$, where $x_0 = 0$ and $x_{n+1} = \infty$. Define $C(k) = \sum_{i=0}^{k} c(i)$ and

$$\lambda_k = \frac{1}{(m+n+1) - k - C(k)}, \quad k = 0, 1, \ldots, n.$$

Then assuming no ties among death times, the Berliner-Hill estimate of survival at x_i is

$$1 - \tilde{F}^0_{BH}(x_i) = \prod_{k=0}^{i-1} (1 - \lambda_k).$$

with a mass at x_i of

$$d\tilde{F}^0_{BH}(x_i) = \lambda_{i-1} \prod_{k=0}^{i-2} (1 - \lambda_k).$$

The proof proceeds by induction. The basis for induction using (2.16) is

$$\begin{aligned} d\tilde{F}^0(x_1) &= \left\{ (n+r+1) - \sum_{z_k < x_1} \frac{1}{1 - \tilde{F}^0(z_k)} \right\}^{-1} \\ &= \{(n+r+1) - c(0)\}^{-1} \\ &= d\tilde{F}^0_{BH}(x_1). \end{aligned}$$

Now assuming the result is true for $x_1, \ldots, x_{i-1}$,

$$\begin{aligned} d\tilde{F}^0(x_i) &= \left\{ (n+r+1) - \sum_{z_k < x_i} \frac{1}{1 - \tilde{F}^0(z_k)} \right\}^{-1} \\ &= \left\{ (n+r+1) - c(0) - \sum_{k=1}^{i-1} \frac{c(k)}{1 - \tilde{F}^0(x_k)} \right\}^{-1} \\ &= \left\{ \frac{1}{\lambda_0} - \sum_{k=1}^{i-1} \frac{c(k)}{\prod_{j=0}^{k-1} (1 - \lambda_j)} \right\}^{-1}. \end{aligned}$$

After some algebra, we have

$$d\tilde{F}^0(x_i) = \frac{\prod_{j=0}^{i-2}(1-\lambda_j)}{\left(\frac{1}{\lambda_0}-1\right)\prod_{j=1}^{i-2}(1-\lambda_j) - \sum_{k=1}^{i-1}\left\{c(k)\prod_{j=k}^{i-2}(1-\lambda_j)\right\}}.$$

We can collapse the denominator by first collecting the two $\prod_{j=1}^{i-2}(1-\lambda_j)$ terms leaving

$$\left\{\frac{1}{\lambda_0} - 1 - c(1)\right\}\prod_{j=1}^{i-2}(1-\lambda_j) - \sum_{k=2}^{i-1}\left\{c(k)\prod_{j=k}^{i-2}(1-\lambda_j)\right\}.$$

Then note that $1/\lambda_0 - 1 - c(1) = 1/\lambda_1$ and remove the $1-\lambda_1$ term from the product giving

$$\left\{\frac{1}{\lambda_1} - 1\right\}\prod_{j=2}^{i-2}(1-\lambda_j) - \sum_{k=2}^{i-1}\left\{c(k)\prod_{j=k}^{i-2}(1-\lambda_j)\right\}.$$

Finally, repeating the steps above for the terms containing $\prod_{j=2}^{i-2}(1-\lambda_j), \ldots, (1-\lambda_{i-2})$ reduces the denominator to $1/\lambda_{i-1}$. Therefore,

$$\begin{aligned} d\tilde{F}^0(x_i) &= \lambda_{i-1}\prod_{j=0}^{i-2}(1-\lambda_j) \\ &= d\tilde{F}^0_{BH}(x_i). \end{aligned}$$

Hence the self-consistent masses in (2.16) are those corresponding to the Berliner-Hill survival estimate.

B.3 First principles proof of the self-consistent expression for the right-censoring model

The proof of the theorem requires that $M(\cdot)$ be stated in terms of $dF^0(\cdot)$. This is demonstrated in the following lemma.

Lemma B.1 *Suppose that $H^0(\cdot)$ is continuous almost everywhere (a.e.) with respect to the measure dF^0. Then*

$$M(s) = \frac{1}{p}\int_0^\infty e^{sx}\left\{1 - H^0(x)\right\} dF^0(x)$$

Proof: First,

$$F(x) = \Pr\left\{X^0 \le x \mid X^0 < Z^0\right\} = \frac{1}{p}\int_0^x dF^0(u)\int_u^\infty dH^0(y)$$
$$= \frac{1}{p}\int_0^x \left\{1 - H^0(u)\right\} dF^0(u). \tag{B.4}$$

The differential of (B.4) provides the Laplace-Stieltjes transform as

$$M(s) = \int_0^\infty e^{sx} dF(x) = \frac{1}{p}\int_0^\infty e^{sx}\left\{1 - H^0(x)\right\} dF^0(x).$$

The Stieltjes integral is undefined when H^0 and F^0 share a common point of step discontinuity. The assumption assures that this does not happen. ■

The next step is to write $dQ_R(z)$ in terms of $F^0(\cdot)$ and $H^0(\cdot)$.

$$dQ_R(z) = \Pr\left\{Z^0 \in (z - dz, z], Z^0 < X^0\right\}$$
$$= \Pr\left\{Y^0 \in (z - dz, z]\right\} \Pr\left(X^0 > z\right)$$
$$= dH^0(z)\left\{1 - F^0(z)\right\}. \tag{B.5}$$

Now,

$$M^0(s) = \int_0^\infty e^{sx} dF^0(x)$$
$$= \int_0^\infty e^{sx}\left\{1 - H^0(x)\right\} dF^0(x) + \int_0^\infty e^{sx} H^0(x) dF^0(x)$$
$$= pM(s) + \int_0^\infty e^{sx}\left\{\int_0^x dH^0(z)\right\} dF^0(x)$$

where the first term has been simplified using the mgf of X. The latter term $dG^0(y)$ is replaced by (B.5) to give

$$M^0(s) = pM(s) + \int_0^\infty e^{sx}\left\{\int_0^x \frac{dQ_R(z)}{1 - F^0(z)}\right\} dF^0(x).$$

This completes the proof.

B.4 First principles proof of the self-consistent expression for doubly censored lifetimes

The proof requires that $M(\cdot)$, the m.g.f. of $X \stackrel{d}{=} X^0 \mid Y^0 < X^0 < Z^0$, be stated in terms of $dF^0(\cdot)$. This is demonstrated in the following lemma.

Lemma B.2 *Suppose that $G^0(\cdot)$ and $H^0(\cdot)$ are continuous almost everywhere (a.e.) with respect to measure dF^0. Then*

$$M(s) = \frac{1}{p}\int_0^\infty e^{sx}\, G^0(x)\left\{1 - H^0(x)\right\} dF^0(x)$$

Proof: First,

$$\begin{aligned} F(x) &= \Pr\left\{X^0 \le x \mid Y^0 < X^0 < Z^0\right\} \\ &= \frac{1}{p}\int_0^x dF^0(u)\int_0^x dG^0(y)\int_x^\infty dH^0(z) \\ &= \frac{1}{p}\int_0^x G^0(u)\left\{1 - H^0(u)\right\} dF^0(u) \end{aligned} \tag{B.6}$$

The differential of (B.6) provides the Laplace-Stieltjes transform as

$$M(s) = \int_0^\infty e^{sx}\, dF(x) = \frac{1}{p}\int_0^\infty e^{sx}\, G^0(x)\left\{1 - H^0(x)\right\} dF^0(x)$$

The Stieltjes integral is undefined when G^0 or H^0 share a common point of discontinuity with F^0. The assumption assures that this doesn't occur. ∎

Next, recall

$$\begin{aligned} dQ_L(y) &= \Pr\left\{Y^0 \in (y - dy, y]\,, X^0 < Y^0\right\} = dG^0(y)\; F^0(y) \\ dQ_R(z) &= \Pr\left\{Z^0 \in (z - dz, zy]\,, Z^0 < X^0\right\} = dH^0(z)\left\{1 - F^0(z)\right\}. \end{aligned}$$

Now,

$$
\begin{aligned}
M^0(s) &= \int_0^\infty e^{sx}\, dF^0(x) \\
&= \int_0^\infty e^{sx} \Pr\{X^0 < Y^0\}\, dF^0(x) + \int_0^\infty e^{sx} \Pr\{Y^0 < X^0 < Z^0\}\, dF^0(x) \\
&+ \int_0^\infty e^{sx} \Pr\{X^0 < Z^0\}\, dF^0(x) \\
&= \int_0^\infty e^{sx} \left\{ \int_x^\infty dG^0(y) \right\} dF^0(x) + \int_0^\infty e^{sx} G^0(x)\{1 - H^0(x)\}\, dF^0(x) \\
&+ \int_0^\infty e^{sx} \left\{ \int_0^x dH^0(y) \right\} dF^0(x).
\end{aligned}
$$

Using the lemma and the above representations of $dQ_L(y)$ and $dQ_R(z)$ gives

$$
\begin{aligned}
M^0(s) = \int_0^\infty e^{sx} \left\{ \int_x^\infty \frac{dQ_L(y)}{F^0(y)} \right\} dF^0(x) + p\, M(s) \\
+ \int_0^\infty e^{sx} \left\{ \int_0^x \frac{dQ_R(z)}{1 - F^0(z)} \right\} dF^0(x).
\end{aligned}
$$

Which is the desired result.

B.5 Indeterminacy Results

To determine the c.d.f. at $y^* = y_1$ and $z^* = z_r$, we need to peel the right-censored observations from the left-tail and peel the left-censored observations from the right-tail as described above. So, using (2.24), the survival for smallest observation $y^* = y_1$ is

$$
\begin{aligned}
1 - \hat{F}^0(y_1) &= 1 - \frac{1}{n + l + r} \sum_{j=1}^{l} \frac{\hat{F}^0(y_1)}{\hat{F}^0(y_j)} \\
&= 1 - \frac{1}{n + l + r} \left\{ \sum_{j=2}^{l} \frac{\hat{F}^0(y_1)}{\hat{F}^0(y_j)} + 1 \right\}.
\end{aligned}
$$

Which implies

$$1 - \hat{F}^0(y_1) = 1 - \left\{ n + l + r - \sum_{j=2}^{l} \frac{1}{\hat{F}^0(y_j)} \right\}^{-1}.$$

The indeterminacy of $\hat{F}^0(x)$ for $x < y^* = y_1$ follows from (2.24) and using the value of $1 - \hat{F}^0(y_1)$ by

$$\begin{aligned} 1 - \hat{F}^0(x) &= 1 - \frac{\hat{F}^0(x)}{n+l+r} \sum_{j=1}^{l} \frac{1}{\hat{F}^0(y_j)} \\ &= 1 - \frac{\hat{F}^0(x)}{n+l+r} \left[\sum_{j=2}^{l} \frac{1}{\hat{F}^0(y_j)} + \left\{ n + l + r - \sum_{j=2}^{l} \frac{1}{\hat{F}^0(y_j)} \right\} \right] \\ &= 1 - \hat{F}^0(x) \end{aligned}$$

Therefore $\hat{F}^0(x)$ may take any value in $\left[0,\ \hat{F}^0(y_1)\right]$.

Again using (2.24), we may calculate the survival at the largest time $z^* = z_r$ as

$$\begin{aligned} 1 - \hat{F}^0(z_r) &= \frac{1}{n+l+r} \sum_{k=1}^{r} \frac{1 - \hat{F}^0(z_r)}{1 - \hat{F}^0(z_k)} \\ &= \frac{1}{n+l+r} \left\{ \sum_{k=1}^{r-1} \frac{1 - \hat{F}^0(z_r)}{1 - \hat{F}^0(z_k)} + 1 \right\} \end{aligned}$$

Thus, solving for $1 - \hat{F}^0(z_r)$ gives

$$1 - \hat{F}^0(z_r) = \left\{ n + l + r - \sum_{k=1}^{r-1} \frac{1}{1 - \hat{F}^0(z_k)} \right\}^{-1}.$$

The indeterminacy of $\hat{F}^0(x)$ for $x > z^* = z_r$ follows from (2.24) and using the value of $1 - \hat{F}^0(z_r)$ by

$$\begin{aligned} 1 - \hat{F}^0(x) &= \frac{1 - \hat{F}^0(x)}{n+l+r} \sum_{k=1}^{r} \frac{1}{1 - \hat{F}^0(z_k)} \\ &= \frac{1 - \hat{F}^0(x)}{n+l+r} \left[\sum_{k=1}^{r-1} \frac{1}{1 - \hat{F}^0(z_k)} + \left\{ n + l + r - \sum_{k=1}^{r-1} \frac{1}{1 - \hat{F}^0(z_k)} \right\} \right] \\ &= 1 - \hat{F}^0(x). \end{aligned}$$

Therefore $\hat{F}^0(x)$ may take any value in $\left[\hat{F}^0(z_r),\ 1\right]$.

C Appendix to Chapter 3

C.1 Proof of Theorem 3.4

Comparing the two systems of equations (3.14) and the empirical versions of (3.4) and (3.5), it suffices to show

$$1 - \hat{G}_1(z) - \hat{G}_2(z) = \left\{1 - \hat{F}_1^0(z)\right\}\left\{1 - \hat{F}_2^0(z)\right\} \tag{C.1}$$

for all censored values z.

The proof proceeds by induction. To the left of the first death, $\hat{G}_1$, $\hat{G}_2$, $\hat{F}_1^0$ and $\hat{F}_2^0$ are all equal to zero. Which implies (C.1) for all z_1 to the left of the first death. Suppose the first death is type one, say x_{11}. Then since (C.1) for all $z_1 < x_{11}$, we have

$$d\hat{G}_1(x_{11}) = \left\{1 - \hat{F}_2^0(x_{11})\right\} d\hat{F}_1^0(x_{11}). \tag{C.2}$$

Note that $1 - \hat{F}_2^0(x_{11}) = 1 - \hat{F}_2^0(z_1)$ for some $z_1 < x_{11}$. Now for censored values z_2 between x_{11} and the second death

$$\begin{aligned} 1 - \hat{G}_1(z_2) - \hat{G}_2(z_2) &= 1 - \left\{\hat{G}_1(z_1) + d\hat{G}_1(x_{11})\right\} - \hat{G}_2(z_2) \\ &= \left\{1 - \hat{F}_1^0(z_1)\right\}\left\{1 - \hat{F}_2^0(z_1)\right\} - \left\{1 - \hat{F}_2^0(x_{11})\right\} d\hat{F}_1^0(x_{11}) \\ &= \left\{1 - \hat{F}_1^0(z_1) - d\hat{F}_1^0(x_{11})\right\}\left\{1 - \hat{F}_2^0(z_1)\right\} \\ &= \left\{1 - \hat{F}_1^0(z_2)\right\}\left\{1 - \hat{F}_2^0(z_2)\right\}. \end{aligned}$$

If the first death is type two, then a symmetric argument to the one above produces the required result.

For the induction step, suppose (C.1) is true for all censored values less than the l^{th} death. Let z_l be a censored value between the $(l-1)^{th}$ and l^{th} deaths. Suppose the l^{th} death is type one and it is the k^{th} type one death. Then the induction hypothesis implies

$$d\hat{G}_1(x_{1k}) = \left\{1 - \hat{F}_2^0(x_{1k})\right\} d\hat{F}_1^0(x_{1k}).$$

Note that $1 - \hat{F}_2^0(x_{1k}) = 1 - \hat{F}_2^0(z_l)$. Now for censored values z_{l+1} between the l^{th} and $(l+1)^{th}$ deaths

$$\begin{aligned}
1 - \hat{G}_1(z_{l+1}) - \hat{G}_2(z_{l+1}) &= 1 - \left\{\hat{G}_1(z_l) + d\hat{G}_1(x_{1k})\right\} - \hat{G}_2(z_l) \\
&= \left\{1 - \hat{F}_1^0(z_l)\right\}\left\{1 - \hat{F}_2^0(z_l)\right\} - \left\{1 - \hat{F}_2^0(x_{1k})\right\} d\hat{F}_1^0(x_{1k}) \\
&= \left\{1 - \hat{F}_1^0(z_l) - d\hat{F}_1^0(x_{1k})\right\}\left\{1 - \hat{F}_2^0(z_l)\right\} \\
&= \left\{1 - \hat{F}_1^0(z_{l+1})\right\}\left\{1 - \hat{F}_2^0(z_{l+1})\right\}.
\end{aligned}$$

If the l^{th} death is type two, then a symmetric argument produces the desired result.

C.2 Proof that the transmittance $1 \to \mathrm{D}_k$ is $p_{k:R} M_{k:R}(s)$

The result will be demonstrated for the first risk factor with the result for the second following the lines of proof identically. Recall that $p_{1:R} = \Pr(X_1^0 < X_2^0)$ and $F_{1:R}(x_1) = \Pr(X_1^0 \le x_1 \mid X_1^0 < X_2^0)$ and

$$\begin{aligned}
p_1 &= \Pr\left\{X_1^0 = W\right\} \\
F_1(x_1) &= \Pr\left\{X_1^0 \le x_1 \mid X_1^0 = W\right\} \\
G_1(x_1) &= p_{1:R} F_{1:R}(x_1).
\end{aligned}$$

Then,

$$\begin{aligned}
M_{1:R}(s) &= \int_0^\infty e^{sx_1}\, dF_{1:R}(x_1) \\
&= \frac{1}{p_{1:R}} \int_0^\infty e^{sx_1}\, dG_1(x_1).
\end{aligned}$$

Now split the event $\{X_1^0 < X_2^0\}$ into the mutually exclusive events

$$\{X_1^0 < X_2^0,\, X_1^0 < Z^0\} \text{ and } \{X_1^0 < X_2^0,\, X_1^0 > Z^0\}$$

and define

$$\bar{F}_1(x_1) = \Pr\left(X_1^0 \le x_1,\, X_1^0 < X_2^0,\, X_1^0 > Z^0\right).$$

Note that $F_{1:R}(x_1) = p_1 F_1(x_1) + \bar{F}_1(x_1)$ so,

$$p_{1:R} M_{1:R}(s) = p_1 \int_0^\infty e^{sx_1} dF_1(x_1) + \int_0^\infty e^{sx_1} d\bar{F}_1(x_1).$$

To complete the proof, we need to calculate $d\bar{F}_1(x_1)$. Therefore, condition on a specific small censored region to get

$$\begin{aligned}\bar{F}_1(x_1) &= \int_0^{x_1} \Pr\left\{X_1^0 \le x_1,\, X_1^0 < X_2^0 \mid Z^0 \in (z - dz, z],\, Z^0 = W\right\} dQ_R(z) \\ &= \int_0^{x_1} \frac{G_1(x_1) - G_1(z)}{1 - G_1(z) - G_2(z)} dQ_R(z)\end{aligned}$$

$$\Rightarrow d\bar{F}_1(x_1) = \left\{\int_0^{x_1} \frac{dQ_R(z)}{1 - G_1(z) - G_2(z)}\right\} dG_1(x_1).$$

Hence,

$$\begin{aligned}p_{1:R} M_{1:R}(s) &= p_1 \int_0^\infty e^{sx_1} dF_1(x_1) + \int_0^\infty e^{sx_1} d\bar{F}_1(x_1) \\ &= p_1 M_1(s) + \int_0^\infty e^{sx_1} \left\{\int_0^{x_1} \frac{dQ_R(z)}{1 - G_1(z) - G_2(z)}\right\} dG_1(x_1),\end{aligned}$$

the desired expression.

References

[1] Aalen, O.O. (1995). Phase type distributions in survival analysis. *Scand. J. Statist.*, **22**, 447-463.

[2] Aitchison, J. and Dunsmore, I.R. (1975). *Statistical Prediction Analysis*. Cambridge: Cambridge University Press.

[3] Akritas, M.G. (1986). Bootstrapping the Kaplan-Meier Estimator. *Journal of the American Statistical Society*, **81**, 1032-1038.

[4] Barlow, R.E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. New York: Holt, Reinhart and Winston.

[5] Berliner, L.M. and Hill, B.M. (1988). Bayesian Nonparametric Survival Analysis. *Journal of the American Statistical Association*, **83**, 772-784.

[6] Butler, R.W. (1997a). First passage distributions in semi-Markov processes and their saddlepoint approximation. *Preprint*, Statistics Department, Colorado State University, Fort Collins.

[7] Butler, R.W. (1997b). Reliabilities for feedback systems and their saddlepoint approximation. *Preprint*, Statistics Department, Colorado State University, Fort Collins.

[8] Butler, R.W. (2000). Reliabilities for Feedback Systems and their Saddlepoint Approximation. *Statistical Science*, **15**, 279-298.

[9] Butler, R.W. and Bronson, D.A. (2000). Bootstrapping Survival Times in Stochastic Systems Using Saddlepoint Approximations. *Preprint*, Statistics Department, Colorado State University, Fort Collins.

[10] Butler, R.W. and Robinson, J. (2000). Uniformity of saddlepoint approximations for passage time distributions in semi-Markov processes. *In preparation*.

[11] Chhikara, R.S. and Folks, J.L. (1989). *The Inverse Gaussian Distribution*. New York: Marcel Dekker.

[12] Chiang, C. and Hsu, J. (1976). On Multiple Transition Time in a Simple Illness Death Process - a Fix-Neyman Model. *Mathematical Biosciences*, **30**, 55-71.

[13] Commenges, D. (1986). Semi-Markov and non-homogeneous Markov models in medicine. In *Semi-Markov Models: Theory and Applications* (ed. J. Janssen) pp. 423-436. New York: Plenum Press.

[14] Crowder, M. (2000). Characterizations of Competing Risks in Terms of Independent-risks Proxy Models. *Scand. J. Statist.*, **27**, 57-64.

[15] Daniels, H. (1954). Saddlepoint approximations in statistics. *Ann. Math. Statist.*, **25**, 631-650.

[16] Daniels, H. (1980). Exact saddlepoint approximations. *Biometrika*, **67**, 59-63.

[17] Daniels, H. (1987). Tail probability approximations. *Int. Statist. Rev.*, **55**, 37-48.

[18] Davison, A.C. and Hinkley, D.V. (1988). Saddlepoint approximations in resampling methods. *Biometrika*, **75**, 417-431.

[19] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and their Application.* Cambridge: Cambridge University Press.

[20] DiCiccio, T.J., Martin, M.A. and Young, G.A. (1992a). Analytic approximations for iterated bootstrap confidence intervals. *Statistics and Computing*, **2**, 161-171.

[21] DiCiccio, T.J., Martin, M.A. and Young, G.A. (1992b). Fast and accurate approximate double bootstrap confidence intervals. *Biometrika*, **79**, 285-295.

[22] DiCiccio, T.J., Martin, M.A. and Young, G.A. (1994). Analytic approximations to bootstrap distribution functions using saddlepoint approximations. *Statistica Sinica*, **4**, 281-295.

[23] Dinse, G. and Larson, M. (1986). A Note of Semi-Markov Models for Partially Censored Data. *Biometrika*, **73**, 379-386.

[24] Efron, B. (1967). The Two Sample Problem with Censored Data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 831-853.

[25] Efron, B. (1981). Censored Data and the Bootstrap. *Journal of the American Statistical Society*, **76**, 312-319.

[26] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

[27] Elandt-Johnson, R. and Johnson, N. (1980). *Survival Models and Data Analysis.* New York: John Wiley and Sons.

[28] Fix, E. and Neyman, J. (1951). A Simple Stochastic Model of Recovery, Relapse, Death and Loss of Patients. *Human Biology*, **23**, 205-241.

[29] Gu, M.G. and Zhang, C.H. (1993). Asymptotic Properties of Self- Consistent Estimators Based on Doubly Censored Data. *Annals of Statistics*, **21**, 611-624.

[30] Guttman, I. (1970). *Statistical Tolerance Regions: Classical and Bayesian.* London: Griffin.

[31] Hill, B.M. (1975). A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, **3**, 1163-1174.

[32] Hill, B. (1991). Bayesian Nonparametric Survival Analysis: A Comparison of the Kaplan-Meier and Berliner-Hill Estimators. *Nato Advanced Studies Workshop on Survival Analysis and Related Topics*, Invited paper.

[33] Hinkley, D.V. and Shi, S. (1989). Importance sampling and the nested bootstrap. *Biometrika*, **76**, 435-446.

[34] Hoel, D. (1972). A Representation of Mortality Data by Competing Risks. *Biometrics*, **28**, 475-488.

[35] Howard, R.A. (1964). System analysis of semi-Markov processes. *IEEE Trans. on Military Electronics*, **8**, 114-124.

[36] Howard, R.A. (1971). *Dynamic Probabilistic Systems, Volume II: Semi-Markov and Decision Processes.* New York: Wiley.

[37] Jensen, J. (1995). *Saddlepoint Approximations.* Oxford: Clarendon Press.

[38] Kaplan, E.L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Society*, **53**, 457-481.

[39] Lagakos, S., Sommer, C. and Zelen, M. (1978). Semi-Markov models for partially censored data. *Biometrika*, **65**, 311-317.

[40] Lugannani, R. and Rice, S. (1980). Saddlepoint approximations for the distribution of the sum of independent random variables. *Adv. Appl. Prob.*, **12**, 475-490.

[41] Mason, S.J. (1953). Feedback theory—Some properties of signal flow graphs. *Proceedings of Institute of Radio Engineers*, **41**, 1144-1156.

[42] Mason, S.J. (1956). Feedback theory—Further properties of signal flow graphs. *Proceedings of Institute of Radio Engineers*, **44**, 920-926.

[43] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models.* New York: Chapman and Hall.

[44] Miller, R. (1981). *Survival Analysis.* New York: John Wiley and Sons.

[45] Mykland, P. and Ren, J.-J. (1996). Algorithms for Computing Self-Consistent and Maximum Likelihood Estimators with Doubly Censored Data. *The Annals of Statistics*, **24**, 1740-1764.

[46] Phillips, C.L. and Harbor, R.D. (1996). *Feedback Control Systems.* New Jersey: Prentice Hall.

[47] Polansky, A.M. (1999). Upper bounds on the true coverage of bootstrap percentile type confidence intervals. *American Statistician*, **53**. 362-369.

[48] Pyke, R. (1961). Markov renewal processes with finitely many states. *Ann. Math. Stat.*, **32**, 1243-1259.

[49] Ross, S.M. (1983). *Stochastic Processes.* New York: Wiley.

[50] Seneta, E. (1981). *Non-negative Matrices and Markov Chains.* New York: Springer-Verlag.

[51] Tsai, W. and Crowley, J. (1985). A Large Sample Study of Generalized Maximum Likelihood Estimators from Incomplete Data Via Self-Consistency. *The Annals of Statistics*, **4**, 1317-1334.

[52] Tsiatis, A. (1975). A Nonidentifiability Aspect of the Problem of Competing Risks. *Proceedings of the National Academy of Sciences*, **72**, 20-22.

[53] Turnbull, B. (1974). Nonparametric Estimation of a Survivorship Function with Doubly Censored Data. *Journal of the American Statistical Association*, **69**. 169-173.

[54] Turnbull, B. (1976). The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society Series B*, **38**, 290-295.