The Kaplan-Meier Estimator from a Systems-Theoretic Perspective

Ronald W. Butler
Department of Statistical Science
Southern Methodist University
Dallas, TX 75275 U.S.A.
rbutler@smu.edu

Douglas A. Bronson
Neptune and Company
Lakewood, CO 80215 U.S.A.
dbronson@neptuneinc.org

October 10, 2007

Summary

The lifetime of a patient is considered from a systems-theoretic perspective that accommodates for the possibility of right censoring and left truncation. The systems-theoretic solution results in the Kaplan-Meier estimator as the appropriate estimate for survival that adjusts for censoring risk and truncation in both the single event and competing risks settings. This systems-theoretic viewpoint uses multi-state semi-Markov models to represent the lifetime of the patient and also provides a circuitry interpretation for the passage of the patient through censoring states and for delayed entry into the study. While the Kaplan-Meier estimator is a nonparametric maximum likelihood estimator, this alternative systems-theoretic and circuitry motivation provides another interpretation for the Kaplan-Meier from a quite different physical perspective. More fundamentally, these systems-theoretic results provide independent support for the use of likelihood procedures. When covariates are present, simple extensions lead to some commonly used estimates for the baseline survival in both single event and competing risks settings.

Some key words: Competing risks; Flowgraph; Kaplan-Meier; Left truncation; Right censoring; Self-consistency; Semi-Markov; Systems theory.

1 Introduction and overview

One aim of survival analysis is to use censored and truncated survival data to determine the lifetime distribution F^0 of a patient that is free from the risk of censoring and adjusted for truncation. As described below, this may be accomplished from a systems-theoretic perspective. When censoring and truncation are independent of survival time, the systems-theoretic solution to the problem is shown to result in the Kaplan-Meier estimator $\hat{F}^0$ for F^0. Without truncation, this estimator has been traditionally motivated as either a nonparametric maximum likelihood estimator (Kaplan-Meier, 1958), a redistribute-to-the-right estimator (Efron, 1967), or as an estimator that satisfies a self-consistency relationship (Efron, 1967). With truncation and censoring, it has been suggested as a nonparametric maximum likelihood estimator by Turnbull (1976). In the current paper, the Kaplan-Meier estimator is shown to be the systems-theoretic solution to generalizations of the self-consistency equations introduced in Efron (1967), however expressed in terms of their Laplace-Stieltjes transforms.

The systems point-of-view uses an analog circuit to represent the progressive stages in the life of the patient; see Figure 1. Suppose the certainty of the patient is represented by an electrical charge of 1. The arrival of the patient into the study corresponds to the arrival of a Dirac function to the input node of a circuit used to model the lifetime of the patient. The transient currents that pass through the various wires over time represents the rates at which "portions" of the virtual patient pass between nodes in the circuit. At the output node, which corresponds to the death state of the patient, a plot of current versus time is the impulse response function for the system driven by Dirac input; it is also the survival density $dF^0(t)$ of the patient whose accumulation of probability over time must be the total probability or the charge of 1.

An empirical version of this analog circuit based on right-censored and left-truncated data approximates the system that describes a patient's lifetime; see Figure 2. The corresponding empirical impulse response function for the empirical circuit is shown to be $d\hat{F}^0(t)$, the mass points for the Kaplan-Meier estimator.

The proposed analog circuit is represented mathematically as a multi-state semi-Markov process whose flowgraph provides a schematic diagram for the circuit; see Figure 1. This flowgraph is labelled with Laplace-Stieltjes transforms which are the "transmittances" through the various wires. The Laplace-Stieltjes transform for the impulse response at the output node is the system transfer function and for the empirical version of the circuit will be recognized as the Laplace-Stieltjes transform for the Kaplan-Meier mass probabilities. It is common to refer to this transfer function as the "solution" to the circuit or system.

This circuitry or systems-theoretic approach has been previously used in Butler & Bronson (2002) to estimate passage time distributions in multi-state survival models that lack censoring using bootstrap methods. A major aim and motivation for the present work has been to incorporate censoring into such models thus subsuming the treatment of censoring into the systems-theoretic approach. The results of this paper, as concerns the extensions to the competing risk setting, suggest that Laplace-Stieltjes transforms for the Kaplan-Meier mass probabilities can be used as transmittances in such models and can be justified as modules or chips within the circuitry to deal with the existence of right censoring from system states. For a more complete description of these systems-theoretic methods, see Butler (2001, 2007 Ch. 13) as it concerns multi-state semi-Markov processes.

A fundamentally important yet unintended consequence of this work is the support that the systems-theoretic approach provides for the use of maximum likelihood as a method of statistical inference. The systems-theoretic framework of this paper motivates the Kaplan-Meier estimator, a nonparametric maximum likelihood estimator, as an adjustment for right censoring and left truncation. By equating the passage of a virtual patient through the various states of a multi-state semi-Markov process with the flow of current through an electrical circuit, the fundamental concept of current additivity in parallel circuits leads to use of the Kaplan-Meier estimator to make the adjustment. Thus this framework provides a separate and fundamentally different motivation for an estimator that would otherwise be motivated in terms of mathematical likelihood. The physical basis for the systems-theoretic motivation and its independence from likelihood concepts give additional reassurance for the Kaplan-Meier estimator and more fundamentally for the likelihood principle and its maximization as a method for statistical inference.

Finally, covariate u may be entered into the single event setting by assuming that covariate dependence in $F^0(t)$ enters through the proportional hazards structure. In this context, survival time X^θ has survival $S^0(t)^\theta$ with $S^0(t)$ as baseline survival and $\theta = \exp(\beta^T u)$. Within the circuitry interpretation of this systems-theoretic approach, some commonly used estimates of baseline survival $S^0(t)$ may be determined. These estimates use the idea that the risk of X^θ is equivalent to the risk of θ independent hypothetical (virtual) baseline subjects that represent a charge of θ. The relative risk for all the X^θ subjects are assessed in terms of the corresponding risks for the collection of their virtual baseline counterparts. Extensions of these ideas to the competing risks setting are also given.

2 Censored and truncated lifetimes

Let random variable X^0 be the lifetime for a patient in the population with distribution function $F^0(x)$ and survival function $S^0(x)$. Suppose a random truncation time T^0 that is independent of X^0 which may also be interpreted as age upon entry into the study. Let the censoring time Z^0 be random, independent of X^0, and dependent on T^0 only through the fact that the event $\{T^0 < Z^0]$ is assumed to have probability

1. If $T^0 = t$, then also assume that the conditional distribution function of $Z^0|T^0 = t$ is the distribution function G^0 restricted to (t, ∞) or the residual life distribution for G^0 given $Z^0 > t$. For an untruncated patient who enters into the study, the truncation time T^0 is observed along with $\min(X^0, Z^0)$. This is the single event setting since one event time X^0 is considered.

Since only the smaller of X^0 and Z^0 is observed when $T^0 < \min(X^0, Z^0)$, it is convenient to define the following competitive variables and their conditional distributions:

$$\begin{aligned} T &\overset{d}{=} T^0 \,|\, \{T^0 < X^0\} \sim E\,(x) \\ X &\overset{d}{=} X^0 \,|\, \{T^0 < X^0 < Z^0\} \sim F\,(x) \\ Z &\overset{d}{=} Z^0 \,|\, \{T^0 < Z^0 < X^0\} \sim G\,(z) \end{aligned}$$

with $p_1 = \mathrm{pr}\{X^0 < Z^0 \,|\, T^0 < X^0\}$ and $p_0 = 1 - p_1$. The three random variables $T, X,$ and Z represent competitive values for truncation time, lifetime, and censoring time respectively, and the probability p_1 is also competitive. All three distributions and p_1 are estimable from the observed data. The support for all random variables is assumed to be $(0, \infty)$.

2.1 Semi-Markov systems

The lifetime of a random subject that may be right-censored and left-truncated is shown in the semi-Markov flowgraph of Figure 1. If times are expressed in terms of age, then a subject is born into node B at time 0. The transmittance input to node B takes the value 1 and is the Laplace-Stieltjes transform for a Dirac function input at time 0. An untruncated subject enters the study in the upper portion of the flowgraph at time $T \in [t, t + dt)$ through state 1_t where t indexes one among the continuum of truncation-time states $\{1_t : t > 0\}$. An observed lifetime occurs when the subject passes directly from $1_t \to$ D with D as the absorbing death state. A right-censored subject passes to state R_z among the continuum of states $\{\mathrm{R}_z : z > t\}$ where z is the absolute time of censoring. After censoring, the subject's subsequent unobserved lifetime depends on z as indicated by the transition from $\mathrm{R}_z \to$ D. The unobserved direct transition B $\to$ D indicated at the bottom of the flowgraph is the transmittance to death for a truncated subject. All transition times in the semi-Markov flowgraph are observed except for passage from B $\to$ D and $\mathrm{R}_z \to$ D.

Each pathway in the flowgraph is labelled with its *transmittance* defined as the product of the state transition probability times the moment generating function for the holding time in the originating state. For example, the transition B $\to 1_t$ occurs in time t hence the moment generating function is e^{st} with probability

$$dL(t) = \mathrm{pr}\{T^0 \in [t, t+dt), T^0 < X^0\} = \tau dE(t) \tag{1}$$

where $\tau = \mathrm{pr}\{T^0 < X^0\} = \int_0^\infty dL(t)$. With lifetime y, the transmittance $1_t \to$ D with incremental transition time $y - t$ is

$$M_t(s) = \int_t^\infty e^{s(y-t)} dB_t(y) \tag{2}$$

where $dB_t(y)$ is the probability the subject has observed lifetime y after entering the study at time t, or

$$dB_t(y) = \mathrm{pr}\{X^0 \in [y, y+dy), Z^0 > y \,|\, T^0 \in [t, t+dt), T^0 < X^0\}. \tag{3}$$

A subject who is censored at time $z > t$ makes transition $1_t \to \mathrm{R}_z$ in time $z - t$ with probability

$$dQ_t(z) = \mathrm{pr}\{Z^0 \in [z, z+dz), X^0 > z \,|\, T^0 \in [t, t+dt), T^0 < X^0\}, \tag{4}$$

hence the transmittance $e^{s(z-t)} dQ_t(z)$. The fact that this transmittance depends on the destination state R_z is the reason for referring to the flowgraph as semi-Markov.

Two of the transmittances $N(s,z)$ and $\Upsilon(s)$ are associated with transition times that are not directly observable. These transmittances are reexpressed in Lemma 1 in terms of quantities that are estimable as a result of the independent censoring and truncation assumptions.

Lemma 1 *Suppose that E^0, F^0, and G^0 have no common jump points so that all Riemann-Stieltjes integrals are defined. The transmittances that correspond to unobserved transition times are estimable though the following relationships.*

$$N(s,z) = \frac{e^{-sz}}{S^0(z)} \int_z^\infty e^{sy} dF^0(y) \tag{5}$$

$$\Upsilon(s) = \int_0^\infty e^{sy} \left\{ 1 - \tau \int_0^y \frac{dE(t)}{S^0(t)} \right\} dF^0(y). \tag{6}$$

Proof: For (5),

$$N(s,z) = \int_z^\infty e^{s(y-z)} dH_z(y) \tag{7}$$

where, for $y > z > t$,

$$\begin{aligned} dH_z(y) &= \mathrm{pr}\{X^0 \in [y, y+dy) \mid Z^0 \in [z, z+dz), Z^0 < X^0\} \\ &= dF^0(y)/S^0(z). \end{aligned} \tag{8}$$

Substitution of (8) into (7) leads to (5).

Derivation of (6), requires first determining the relationship of $dE(t)$ to $dE^0(t)$ as

$$\begin{aligned} dE(t) &= \mathrm{pr}\{T^0 \in [t, t+dt) \mid T^0 < X^0\} \\ &= \mathrm{pr}\{T^0 \in [t, t+dt), t < X^0\}/\tau \\ &= dE^0(t) S^0(t)/\tau. \end{aligned} \tag{9}$$

Thus

$$\begin{aligned}
\Upsilon(s) &= \int_0^\infty e^{sy}\,\mathrm{pr}\left\{X^0 \in (y, y+dy), X^0 < T^0\right\} \\
&= \int_0^\infty e^{sy}\{1 - E^0(y)\} dF^0(y) \\
&= \int_0^\infty e^{sy}\left\{1 - \int_0^y dE^0(t)\right\} dF^0(y)
\end{aligned}$$

which, upon using (9), gives (6). $\square$

Semi-Markov modelling to explain right censoring (but without truncation) was previously considered in Lagakos, Sommer, & Zelen (1978) and later Anderson *et al* (1993, Ex. III2.8). Their models record passage time up to $\min(X^0, Z^0)$ and represent censoring as a single absorbing node in a semi-Markov competing risks model. These models do not consider transitions after censoring as is required when explaining all state changes connected with lifetime X^0. In modelling the entire lifetime X^0, a single censoring state cannot express the semi-Markov structure after censoring; such states must be indexed as $\{\mathrm{R}_z\}$ in order to allow subsequent transition to state D in time $X^0 - z$. These other models are different because they address different concerns: the determination of a nonparametric maximum likelihood estimator for F^0. The current model rather explains the entire passage time X^0 and requires a semi-Markov model as in Figure 1 to take account of the unobserved residual lifetime after censoring from $\min(X^0, Z^0)$ to X^0.

A Markov model that allows for random left truncation has also been given in Anderson *et al* (1993, Ex. III3.3). This model shows the parallel connection of untruncated and truncated paths but otherwise has not been expanded to also allow for right censoring as in Figure 1.

2.2 Self-consistency of F^0

The assumptions of independent censoring and truncation provide a context in which it is possible to remove the censoring risk factor and adjust for truncation when estimating F^0. When considered in terms of the flowgraph, the density $dF^0(x)$ of X^0 is the output of the analog circuit at node D and the plot of $dF^0(x)$ versus x would be the output seen on an oscilloscope were it to be attached to D. The Laplace-Stieltjes transform of $F^0(x)$ is the transfer function of the system and is determined by summing all parallel transmittances from B $\rightarrow$ D. Thus

$$\int_0^\infty e^{sy} dF^0(y) = \Upsilon(s) + \Delta(s) + \Xi(s) \tag{10}$$

is the respective sum over a virtual patient who may have been truncated, may have had an observed lifetime, or may have been censored with

$$\Delta(s) = \int_{t=0}^{t=\infty} e^{st} dL(t) M_t(s) \tag{11}$$

$$\Xi(s) = \int_{t=0}^{t=\infty} e^{st} dL(t) \left\{ \int_{z=t}^{z=\infty} e^{s(z-t)} dQ_t(z) N(s,z) \right\}. \tag{12}$$

The Stieltjes integrals in (10), (11) and (12) exist so long as the distributions have no common jump points. Expressions (10), (11) and (12) may also be derived from first principles without the flowgraph presentation. However, the use of flowgraphs and the basic principle of summing over parallel conductances gives an intuitive physical interpretation for the preservation of probability flow through the multi-state process.

A general estimable solution $F^0(y)$ to (10) follows provided Υ, Δ, and Ξ can be expressed in terms of F^0 and the estimable distribution functions F, G, and H. As shown in the Appendix,

$$\Delta(s) = \tau p_1 \int_0^\infty e^{sy} dF(y) \tag{13}$$

$$\Xi(s) = \tau p_0 \int_0^\infty e^{sy} \left\{ \int_0^y \frac{dG(z)}{S^0(z)} \right\} dF^0(y) \tag{14}$$

and these expressions may be substituted into (10) along with (6) to yield

$$\begin{aligned}\int_0^\infty e^{sy} dF^0\left(y\right) &= \int_0^\infty e^{sy}\left\{1-\tau\int_0^y \frac{dE(t)}{S^0(t)}\right\} dF^0(y) \\ &\quad +\tau p_1\int_0^\infty e^{sy} dF\left(y\right)+\tau p_0\int_0^\infty e^{sy}\left\{\int_0^y \frac{dG(z)}{S^0(z)}\right\} dF^0(y).\end{aligned} \tag{15}$$

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

$$\begin{aligned}dF^0\left(y\right) &= \left\{1-\tau\int_0^y \frac{dE(t)}{S^0(t)}\right\} dF^0(y) \\ &\quad +\tau p_1 dF\left(x\right)+\tau p_0\left\{\int_0^y \frac{dG(z)}{S^0(z)}dz\right\} dF^0(y).\end{aligned} \tag{16}$$

Cancelling $dF^0(y)$ on both sides as well as the common factor τ leads to

$$dF^0\left(y\right)=\left\{\int_0^y \frac{dE(t)}{S^0(t)}-p_0\int_0^y \frac{dG(z)}{S^0(z)}\right\}^{-1} p_1 dF\left(y\right) \tag{17}$$

when the term in curly braces is not zero or negative. This is the population version of the self-consistent equation originally introduced by Efron (1967) for the estimation of $F^0\left(x\right)$ without truncation. If there is no truncation, then $dE(0)=1$ so $\int_0^y dE(t)/S^0(t)=1$ and (17) gives Efron's result for the population distribution.

Simple computations show that the right side of (17) is $dF^0(y)$. Expression (9) leads to

$$\int_0^y \frac{dE(t)}{S^0(t)}=\frac{E^0(y)}{\tau}.$$

The use of Bayes theorem on $dG(z)$ in the second term of (17) leads to

$$\begin{aligned}p_0\int_0^y \frac{dG(z)}{S^0(z)} &= p_0\int_0^y \frac{\mathrm{pr}\left\{Z^0\in[z,z+dz)\,|\,Z^0<X^0,T^0<X^0\right\}}{S^0(z)} \\ &= p_0\int_0^y \frac{dG^0(z)S^0(z)}{p_0\tau S^0(z)}=\frac{G^0(y)}{\tau}.\end{aligned}$$

Since $\{Z^0\leq y\}\subseteq\{T^0\leq y\}$, the term in curly braces in (17) is

$$\frac{E^0(y)}{\tau}-\frac{G^0(y)}{\tau}=\frac{\mathrm{pr}(T^0\leq y<Z^0)}{\mathrm{pr}(T^0\leq X^0)}.$$

The right side of (17) is now

$$\left\{\frac{\mathrm{pr}(T^0 \leq y < Z^0)}{\mathrm{pr}(T^0 \leq X^0)}\right\}^{-1} p_1\,\mathrm{pr}\{X^0 \in [y, y+dy)|T^0 < X^0 < Z^0\} = dF^0(y)$$

when Bayes theorem is used on the last probability for $dF(y)$.

3 A self-consistent estimator for $F^0(x)$

Untruncated data consist of n_1 lifetimes, observed as the pairs $\{(t_{1i}, x_i) : i = 1, \ldots, n_1\}$ where truncation time $t_{1i} < x_i$, and n_0 censored values $\{(t_{0j}, z_j) : j = 1, \ldots, n_0\}$ with $t_{0j} < z_j$. Distribution functions E, F, and G are estimated by their empirical counterparts $\hat{E}(t), \hat{F}(x)$, and $\hat{G}(z)$ based on $\{t_{1i}\} \cup \{t_{0j}\}$, $\{x_i\}$, and $\{z_j\}$ respectively while $\hat{p}_1 = n_1/n.$ with $n. = n_0 + n_1$.

Figure 2 shows a semi-Markov flowgraph that is an empirical version of the graph in Figure 1. Each subject contributes a separate path from B $\rightarrow$ D with weight $\tau/n.$. The unobserved branches $\mathrm{R}_{z_j} \rightarrow$ D and B $\rightarrow$ D direct have empirical transmittances

$$\begin{aligned}
\hat{N}(s, z_j) &= \frac{e^{-sz_j}}{\hat{S}^0(z_j)} \int_{z_j}^{\infty} e^{sy} d\hat{F}^0(y) \\
\hat{\Upsilon}(s) &= \int_0^{\infty} e^{sy} \left\{1 - \int_0^y \frac{\tau d\hat{E}(t)}{\hat{S}^0(t)}\right\} d\hat{F}^0(y)
\end{aligned}$$

where estimate $\hat{F}^0(y) = 1 - \hat{S}^0(t)$ is presumed to exist. Summing over all parallel connections gives the empirical counterpart to (15) as

$$\int_0^{\infty} e^{sy} d\hat{F}^0\,(y) = \hat{\Upsilon}(s) + \frac{\tau}{n.} \sum_{i=1}^{n_1} e^{st_{1i}} e^{s(x_i - t_{1i})} + \frac{\tau}{n.} \sum_{j=1}^{n_0} e^{st_{0j}} e^{s(z_j - t_{0j})} \hat{N}(s, z_j). \qquad (18)$$

The second term is $\tau n_1/n. \int_0^{\infty} e^{sy} d\hat{F}(y)$ while the last term is

$$\frac{\tau n_0}{n.} \int_0^{\infty} \frac{1}{\hat{S}^0(z)} \left\{\int_z^{\infty} e^{sy} d\hat{F}^0(y)\right\} d\hat{G}(z) = \frac{\tau n_0}{n.} \int_0^{\infty} e^{sy} \left\{\int_0^y \frac{d\hat{G}(z)}{\hat{S}^0(z)}\right\} d\hat{F}^0(y).$$

Inverting the Laplace-Stieltjes transforms in (18) leads to

$$\hat{C}(y)d\hat{F}^0(y) := \left\{ \int_0^y \frac{d\hat{E}(t)}{\hat{S}^0(t)} - \hat{p}_0 \int_0^y \frac{d\hat{G}(z)}{\hat{S}^0(z)} \right\} d\hat{F}^0(y) = \hat{p}_1 d\hat{F}(y) \tag{19}$$

as the defining equation for self-consistency. The solution to (19) is summarized in Theorem 2.

Theorem 2 *Suppose that $\hat{E}(t)$, $\hat{F}(t)$, and $\hat{G}(t)$ have no common jump points. Let $x_* = \min\{x_i\}$ and $x^* = \max(\{x_i\}, \{z_j\})$ and suppose, without any loss in generality, that censored values less than x_* have already been deleted as uninformative. If $\mathcal{N}_t$ is the number of subjects at risk at time t, the assumption that $\mathcal{N}_t > 0$ for all $t \in (x_*, x^*)$ assures that there is a unique self-consistent solution to (19) over (x_*, x^*) which is the Kaplan-Meier estimator.*

Proof. Since $d\hat{F}(y) = 0$ for $y \notin \{x_i\}$, the support for $d\hat{F}^0(y)$ can only be $\{x_i\}$ and also regions in (x_*, x^*) for which $\hat{C}(y) = 0$. The latter possibility will be eliminated with the assumption that $\mathcal{N}_t > 0$ for all $t \in (x_*, x^*)$. For the sake of argument, suppose that $x_1 < x_2 < \cdots < x_{n_1}$. For any t between $\min[\{t_{1i}\}, \{t_{0j}\}]$, and x_1, $\hat{C}(t) > 0$ and $d\hat{F}^0(t) = 0$ so that $\hat{S}^0(t) = 1$ over this range. At $x_1 = x_*$, (19) is

$$\hat{C}(x_1)d\hat{F}^0(x_1) = \left(\frac{\mathcal{N}_1}{n.} - 0 \right) d\hat{F}^0(x_1) = \frac{1}{n.} \tag{20}$$

where $\mathcal{N}_1$ denotes the number truncated before x_1 and therefore at risk, while 0 is the number censored before x_1. The solution to (20) is $d\hat{F}^0(x_1) = 1/\mathcal{N}_1$ so that $\hat{S}(x_1) = 1 - 1/\mathcal{N}_1$.

More generally, the following recursions

$$\frac{1}{d\hat{F}^0(x_l)} = \frac{\mathcal{N}_l}{\hat{S}^0(x_{l-1})} \qquad \text{and} \qquad \hat{S}^0(x_l) = \hat{S}^0(x_{l-1}) \left(1 - \frac{1}{\mathcal{N}_l} \right) \tag{21}$$

are shown to hold for $l = 1, \ldots, n_1$ which are those for the Kaplan-Meier estimator. The $l = 1$ case above has been shown to hold with $x_0 = x_1^-$.

The proof of (21) proceeds by using induction wherein recursion l in (21) is shown to imply recursion $l+1$. Suppose that $\Delta T_{l+1} = n.\{\hat{E}(x_{l+1}) - \hat{E}(x_l)\}$ and $\Delta R_{l+1} = n.\hat{p}_0\{\hat{G}(x_{l+1}) - \hat{G}(x_l)\}$ count the number of truncation and right censoring times within (x_l, x_{l+1}). From (19),

$$\begin{aligned}\frac{1}{d\hat{F}^0(x_{l+1})} &= \sum_{k=0}^{1} \sum_{\{j\,:\,t_{kj}\,<\,x_{l+1}\}} \frac{1}{\hat{S}^0(t_{kj})} - \sum_{\{j\,:\,z_j\,<\,x_{l+1}\}} \frac{1}{\hat{S}^0(z_j)} \\ &= \frac{1}{d\hat{F}^0(x_l)} + \frac{\Delta T_{l+1}}{\hat{S}^0(x_l)} - \frac{\Delta R_{l+1}}{\hat{S}^0(x_l)}\end{aligned}$$

where the constant value $\hat{S}^0(t) \equiv \hat{S}^0(x_l)$ over $t \in [x_l, x_{l+1})$ has been used. From (21), this is

$$\begin{aligned}\frac{1}{d\hat{F}^0(x_{l+1})} &= \frac{\mathcal{N}_l}{\hat{S}^0(x_{l-1})} + \frac{\Delta T_{l+1} - \Delta R_{l+1}}{\hat{S}^0(x_{l-1})\,(1 - 1/\mathcal{N}_l)} \\ &= \frac{\mathcal{N}_l}{\hat{S}^0(x_{l-1})\,(\mathcal{N}_l - 1)}(\mathcal{N}_l - 1 + \Delta T_{l+1} - \Delta R_{l+1}) \\ &= \frac{\mathcal{N}_{l+1}}{\hat{S}^0(x_l)} \end{aligned} \tag{22}$$

upon using the recursion for $\hat{S}^0(x_l)$ in (21). Using $\hat{S}^0(x_{l+1}) = \hat{S}^0(x_l) - d\hat{F}^0(x_{l+1})$ and (22) gives the remaining recursion. $\square$

4 Irregularities and examples

Uniqueness of the solution to the self-consistent equations in (19) may be lost if no subjects are at risk during a portion of the informative time span, i.e., there is a $t^0 \in (x_*, x^*)$ for which $\mathcal{N}_{t^0} = 0$. Furthermore, in this instance $\hat{S}^0$ does not have to place all mass on $\{x_i\}$ but rather can place non-zero mass onto an interval that contains t^0 since $\hat{C}(t^0) = 0$. Let t^* be the next time point above t^0, which is necessarily a truncation time, and let z_* be the next time point below t^0 which is necessarily a lifetime or censoring time. If z_* is a censoring time (lifetime), then interval (z_*, t^*)

can (cannot) hold non-zero mass in the solution to the self-consistency equations. Support intervals such as (z_*, t^*) were first noted by Frydman (1994) as additional sites capable of holding mass for the nonparametric maximum likelihood estimate when there is truncation. Such sites were not mentioned in Turnbull's (1976) original account dealing with general interval censoring and truncation. These points are illustrated using two simple examples.

Example 1. Consider the ordered data

$$t_{x_1} < t_{x_2} < x_1 < t_{z_1} < z_1 < x_2 < t_{x_3} < x_3$$

in which t_{x_i} is the truncation time for x_1, etc. At x_2, the two subjects entered into the study are no longer at risk but the third subject has not yet entered. The example violates the conditions of Theorem 2 since there are no subjects at risk during the interval (x_2, t_{x_3}). The self-consistent solution places mass 1/2 on x_1 and 1/2 on x_2. This leads to $\hat{S}^0(x_2) = 0$ and $\hat{C}(x_2) = 2$ from which the value for

$$\hat{C}(t_{x_3}) = \hat{C}(x_2) + \frac{1}{\hat{S}^0(t_{x_3})}$$

is undefined due to division by $\hat{S}^0(t_{x_3}) = 0$.

In computing the nonparametric maximum likelihood estimate, the support set from Turnbull (1976) is x_i with probability s_i for $i = 1, 2, 3$. The respective likelihood terms contributed by x_1, z_1, x_2, and x_3 are

$$L = s_1(s_2 + s_3)\frac{s_2}{s_2 + s_3}\frac{s_3}{s_3} = s_1 s_2$$

and the maximum likelihood estimate agrees with the self-consistent estimate with $\hat{s}_1 = \hat{s}_2 = 1/2$.

Example 2. Interchange z_1 and x_2 from Example 1 to get ordered data

$$t_{x_1} < t_{x_2} < x_1 < t_{z_1} < x_2 < z_1 < t_{x_3} < x_3.$$

Again there are no subjects at risk during (z_1, t_{x_3}) so Theorem 2 is violated. The self-consistent solution places mass $1/2$ on x_1, $1/4$ on x_2, mass p in (z_1, t_{x_3}), and mass $1/4 - p$ at x_3 for any $p \in [0, 1/4)$. In the self-consistent solution,

$$\hat{C}(z_1) = 4 - \frac{1}{\hat{S}^0(z_1)} = 4 - 4 = 0.$$

By allowing arbitrary mass p in (z_1, t_{x_3}), then

$$\hat{C}(t_{x_3}) = \hat{C}(z_1) + \frac{1}{\hat{S}^0(t_{x_3})} = \frac{1}{1/4 - p} > 0$$

for any $p \in [0, 1/4)$ which leaves $d\hat{F}^0(x_3) = 1/4 - p$.

The support set for the maximum likelihood estimate is determined from Frydman (1994) as x_1, x_2, (z_1, t_{x_3}), and x_3 with probabilities $s_1, \ldots, s_4$. The nonparametric likelihood is

$$L = s_1 s_2 \frac{s_3 + s_4}{s_2 + s_3 + s_4} = s_1 s_2 \frac{1 - s_1 - s_2}{1 - s_1}$$

which attains the same collection of maxima as the self-consistent solution.

5 Proportional hazards extensions in single event settings

Suppose data consist of $n.$ subjects with responses $(t_i, x_i, \delta_i, u_i)$ for $i = 1, \ldots, n.$ where the respective values are truncation time, lifetime/censoring time, indicator of lifetime response, and covariate vector. For notational convenience suppose that $x_1 < \cdots < x_{n.}$. In the context of the proportional hazards model, subject i with lifetime X^{θ_i} has survival function $\mathrm{pr}(X_{\theta_i} > t) = S^0(t)^{\theta_i}$ with $\theta_i = \exp(\beta^T u_i)$. This subject's survival function is the same as that for θ_i independent virtual baseline subjects. Equivalently, the hazard for X^{θ_i} is the sum of the hazards of these θ_i virtual baseline subjects.

By working with a total of $\theta. = \sum_{i=1}^{n.} \theta_i$ independent virtual subjects instead of $n.$ heterogeneous subjects, extensions to the estimation of baseline survival $S^0(t)$ in the proportional hazards setting can be easily determined if ties in responses are allowed. Allowance for ties in the Kaplan-Meier estimator is easily justified on the grounds that it is the limiting estimator when ties are separated slightly and then allowed to approach one another in the tied configuration. We shall work under such limiting assumptions and also use the total collection of $\theta.$ virtual subjects. Note that the difficult case that allows for tied values of x_i with different covariate values θ_i is not under consideration.

Under such arrangements, both the partial likelihood for β and two baseline estimates for $S^0(t)$ can be easily justified. For an assumed β, the partial likelihood under both right censoring and left truncation is

$$L_p(\beta) = \prod_{i=1}^{n.} \left(\frac{\theta_i}{\sum_{j \in R_i} \theta_j} \right)^{\delta_i} \tag{23}$$

where R_i is the risk set at time x_i so that $j \in R_i$ whenever $t_j < x_i < x_j$. The partial likelihood is simply the product of observed Bernoulli probabilities for the virtual baseline subjects at the set of lifetimes.

Estimation of baseline survival for an assumed value of β is simply

$$\hat{S}^0(t) = \prod_{\{i\,:\,x_i \leq t\}} \left(1 - \frac{1}{\sum_{j \in R_i} \theta_j} \right)^{\delta_i}, \tag{24}$$

which is the Kaplan-Meier form of the Breslow or Nelson-Aalen estimate. Without truncation, both (23) and (24) have been motivated as maximum profile semiparametric likelihood estimates by Johansen (1983). From the point-of-view of evaluating the lifetimes of $\theta.$ independent virtual subjects, the justification of (24) is simply that $1 - 1/\sum_{j \in R_i} \theta_j$ is the probability that a single baseline patient survives lifetime point x_i from within the collection of $\sum_{j \in R_i} \theta_j$ baseline subjects who are at risk at time x_i^-. Thus (24) adjusts the observed probabilities pertaining to θ_i-subjects so they are relevant to a baseline subject.

The Kalbfleisch & Prentice estimator (2002, Eq. 4.36)

$$\hat{S}^0_{KP}(t) = \prod_{\{i\,:\,x_i \leq t\}} \left(1 - \frac{\theta_i}{\sum_{j\in R_i}\theta_j}\right)^{\delta_i/\theta_i}$$

is equally simple to motivate. The probability that θ_i virtual patients survive time x_i is $1 - \theta_i / \sum_{j\in R_i}\theta_j$ when computed from among $\sum_{j\in R_i}\theta_j$ subjects at risk at time x_i^-. This is the survival probability $S^0(t)^{\theta_i}$ for X_{θ_i}, and so to have it apply to a baseline patient, it needs to be raised to the $1/\theta_i$ power.

By equating the risk of a θ_i-subject with θ_i independent baseline subjects, these three commonly used estimators from the proportional hazards setting can be motivated. Indeed it is the structure of the proportional hazards setting that allows for and gives validity to these simple arguments.

6 Competing Risks

In the classical competing risks setting, there are multiple event times $X_1^0, \ldots, X_K^0$ with distribution $F^0(x_1, \ldots, x_K)$ that compete with independent censoring time Z^0 and independent left truncation time T^0. The value and index for $M^0 = \min\{X_k^0\}$ are observed if the events are untruncated, $T^0 < M^0$, and uncensored, $M^0 < Z^0$. The aim is to estimate the collection of subdistributions

$$F_k^0(x) = \mathrm{pr}(X_k = M^0 \leq x) \qquad k = 1, \ldots, K$$

associated with F^0 from competitive data that are subject to right censoring and left truncation. The distributional structure of the censoring and truncation variables supposes $T^0 \sim E^0$ and $Z^0 | T^0 = t \sim G^0$ restricted to (t, ∞) with T^0 and Z^0

independent of $X_1^0, \ldots, X_K^0$. The survival function of M^0 is $S_+^0(t) = 1 - F_+^0(t)$ where $F_+^0(t) = \sum_{k=1}^K F_k^0(x)$.

The data are observed to come from the competitive distributions

$$
\begin{aligned}
&T \stackrel{d}{=} T^0 \,|\, \{T^0 < X^0\} \sim E\,(x) \\
&X_k \stackrel{d}{=} X_k^0 \,|\, \{T^0 < X_k^0 = M^0 < Z^0\} \sim F_k\,(x) \\
&Z \stackrel{d}{=} Z^0 \,|\, \{T^0 < Z^0 < M^0\} \sim G\,(z)\,.
\end{aligned}
$$

for $k = 1, \ldots, K$. The competitive probabilities $p_k = \mathrm{pr}(X_k^0 = M^0 < Z^0)$ for $k \geq 1$ and $p_0 = \mathrm{pr}(Z^0 < M^0)$ are estimable from the data and add to 1.

6.1 Population flowgraph

Figure 3 shows the competing risk flowgraph when there are $K = 2$ possible events that are subject to random right censoring and left truncation. The transmittances along with estimable expressions, determined from arguments similar to those in section 2, are $dL(t) = \tau dE(t)$ where $\tau = \mathrm{pr}(T^0 < M^0)$ and $M_{kt}(s) = \int_t^\infty e^{s(y-t)} dB_k(y)$ where

$$
\begin{aligned}
dB_k(y) &= \mathrm{pr}\left\{X_k^0 = M^0 \in [y, y+dy), M^0 < Z^0 \,|\, T^0 \in [t, t+dt), T^0 < M^0\right\} \\
dQ_t(z) &= \mathrm{pr}\left\{Z^0 \in [z, z+dz), Z^0 < M^0 \,|\, T^0 \in [t, t+dt), T^0 < M^0\right\} \\
N_k(s, z) &= \frac{e^{-sz}}{S_+^0(z)} \int_z^\infty e^{sy} dF_k^0(y) \\
\Upsilon_k(s) &= \int_0^\infty e^{sy} \left\{1 - \tau \int_0^y \frac{dE(t)}{S_+^0(t)}\right\} dF_k(y).
\end{aligned}
$$

Summing over all paths from B to D$_k$, within which lifetimes of type k are observed, gives transmittance

$$
\Delta_k(s) = \tau p_k \int_0^\infty e^{sy} dF_k(y).
$$

Summing from B to D_k but passing through censored states gives

$$\Xi_k(s) = \tau p_0 \int_0^\infty e^{sy} \left\{ \int_0^y \frac{dG(z)}{S_+^0(z)} \right\} dF_k^0(y)$$

The Laplace-Stieltjes transform for subdistribution F_k^0 is the sum over all transmittances from B to D_k or $\Upsilon_k(s) + \Delta_k(s) + \Xi_k(s)$. Transform inversion leads to

$$dF_k^0\,(y) = \left\{ \int_0^y \frac{dE(t)}{S_+^0(t)} - p_0 \int_0^y \frac{dG(z)}{S_+^0(z)} \right\}^{-1} p_k\, dF_k\,(y) \tag{25}$$

for $k = 1, \ldots, K$.

6.2 Empirical flowgraph

The data consist of n_k observed events of type k and n_0 censored with total sample size $n. = n_0 + \sum_{k=1}^K n_k$. Suppose x_{ki} is the ith observed event time of type k with associated truncation time $t_{ki} < x_{ki}$. The times $\{t_{0i} : i = 1, \ldots, n_0\}$ are the truncation times for the censored data. Truncation distribution E is estimated using $\hat{E}$, the empirical distribution of $\{t_{kj} : k = 0, \ldots, K; j = 1, \ldots, n_k\}$ while $\hat{F}_k$ and $\hat{G}$ are the empirical distributions of $\{x_{ki} : i = 1, \ldots, n_k\}$ and censoring times $\{z_j : j = 1, \ldots, n_0\}$ respectively. Estimate $\hat{p}_k = n_k/n.$.

Figure 4 shows the empirical competing risk flowgraph with $K = 2$. Summing over all paths from B to D_k leads to the self-consistency equations

$$\hat{C}(y) d\hat{F}_k^0(y) = \hat{p}_k d\hat{F}_k(y) = 1/n. \qquad k = 1, \ldots, K \tag{26}$$

where

$$\hat{C}(y) = \int_0^y \frac{d\hat{E}(t)}{\hat{S}_+^0(t)} - \hat{p}_0 \int_0^y \frac{d\hat{G}(z)}{\hat{S}_+^0(z)}.$$

The solution to (26) is now summarized in Theorem 3.

Theorem 3 *Suppose that $\hat{E}(t)$, $\{\hat{F}_k(t)\}$, and $\hat{G}(t)$ have no common jump points. Let $x_* = \min\{x_{ki}\}$ and $x^* = \max(\{x_{ki}\}, \{z_j\})$ and suppose, without any loss in generality,* *that censored values less than x_* have already been deleted as uninformative. The assumption that $\mathcal{N}_t > 0$ for all $t \in (x_*, x^*)$, where $\mathcal{N}_t$ is the number of subjects at risk at time t, assures that there is a unique self-consistent solution to the K equations in (26) over (x_*, x^*) and that the subdistribution solutions are those associated with the Kaplan-Meier estimator.*

Proof. The proof is the same inductive proof used for Theorem 2. Assumption $\mathcal{N}_t > 0$ assures that mass is only placed on event times. Order the $N = n. - n_0$ event times as $x_1 < \cdots < x_N$ and suppose the associated event types are $i(1), \ldots, i(N)$. The $l = 1$ case of induction holds and states that $d\hat{F}^0_{i(1)}(x_1) = 1/\mathcal{N}_1$ and $\hat{S}^0_+(x_1) = 1 - 1/\mathcal{N}_1$ where $\mathcal{N}_1$ is the number of subjects at risk at time $x_1 = x_*$. Assuming the lth case,

$$\frac{1}{d\hat{F}^0_{i(l)}(x_l)} = \frac{\mathcal{N}_l}{\hat{S}^0_+(x_{l-1})} \qquad \hat{S}^0_+(x_l) = \hat{S}^0_+(x_{l-1})\left(1 - 1/\mathcal{N}_l\right),$$

it can be shown through induction that the $(l+1)$st case holds. $\square$

Subdistribution estimate $\hat{F}^0_k$ accumulates Kaplan-Meier probabilities $d\hat{F}^0_k$ at event times of type k where the Kaplan-Meier estimate $\hat{S}^0_+(x) = 1 - \sum_{k=1}^K \hat{F}^0_k(x)$ for the survival of M^0 has been computed by using the pooled set of event times. If x^* is a censored value, then $\hat{S}^0_+(x)$ and $\{\hat{F}^0_k(x)\}$ are indeterminate for $x > x^*$.

These subdistribution estimates may also be expressed in terms of the individual Kaplan-Meier estimates for the marginal distributions of $\{X^0_k\}$ in the independent competing risk setting. This setting supposes $\{X^0_k\}$ are independent so that $F^0(x_1, \ldots, x_K) = \prod_{k=1}^K F^0(x_k)$. Let $\tilde{F}^0_k(x)$ denote the Kaplan-Meier estimator of $F^0_k(x)$ computed with non-k event times treated as censored along with data that are truly censored. It is a simple derivation to show the subdistribution estimate

$$d\hat{F}^0_k(x) = d\tilde{F}^0_k(x) \prod_{1=i\neq k}^{K} \left\{1 - \tilde{F}^0_i(x)\right\}.$$

That the independent competing risk model should happen to lead to the same estimate reflects the well-known fact that competing risks data contain no information about the dependence structure of the joint distribution of $X_1^0, \ldots, X_K^0$.

7 Proportional hazards extensions in competing risk settings

Prentice *et al* (1978) review the options proposed by Holt (1978) for including covariates in the Cox model and offer two models for this setting. In the first, regression coefficients are cause-specific so that a subject with covariate u would have the cause-specific hazard $\theta_k \lambda_{0k}(t)$ where $\theta_k = \exp(\beta_k^T u)$ and the baseline hazard is $\lambda_{0k}(t) = dF_k^0(t)/S_+^0(t)$. For this model, the K cause-specific parameter sets $\{\theta_1, \lambda_{01}(\cdot)\}, \ldots, \{\theta_K, \lambda_{0K}(\cdot)\}$ are L-independent (Barndorff-Nielsen, 1978, §3.3) in the sense that the likelihood is completely separable into such groups of parameters. Accordingly, cause specific baseline hazards, subdistributions, and regressions are estimated separately by using the methods of section 5.

Under the second model, the cause specific baseline hazards are proportional so the kth event type hazard is $\theta_k \lambda_0(t)$ with $\theta_k = \exp(\alpha_k + \beta_k^T u)$. In this model, the subject's hazard is the same as that of $\sum_{i=1}^K \theta_k$ independent virtual baseline subjects and can be treated as such for purposes of estimation. In regression estimation, this leads to the partial likelihood given in Prentice *et al* (1978, Pg. 547) as well as a logistic model for event type. Correspondingly, estimates of $\lambda_0(t)$ have support over the pooled collection of event times and for each event time are based on the subjects at risk at that time from among the $\sum_{i=1}^K \theta_k$ virtual baseline subjects.

8 Appendix

Proof of (13): Substituting (1), (2), and (3) into (11) gives

$$\begin{aligned}
\Delta(s) &= \tau \int_{t=0}^{t=\infty} e^{st} dE(t) \int_{y=t}^{y=\infty} e^{s(y-t)} dB_t(y) \\
&= \tau \int_{y=0}^{y=\infty} e^{sy} \int_{t=0}^{t=y} \mathrm{pr}\left\{X^0 \in [y, y+dy), Z^0 > y, T^0 \in [t, t+dt) \,|\, T^0 < X^0\right\} \\
&= \tau \int_0^\infty e^{sy}\, \mathrm{pr}\{X^0 \in [y, y+dy), Z^0 > X^0 \,|\, T^0 < X^0\} = \tau p_1 \int_0^\infty e^{sy} dF(y).
\end{aligned}$$

Proof of (14): Substituting (1) and (5) into (12) gives

$$\Xi(s) = \tau \int_{t=0}^{t=\infty} dE(t) \int_{z=t}^{z=\infty} \frac{dQ_t(z)}{S^0(z)} \int_z^\infty e^{sy} dF^0(y) = \int_0^\infty e^{sy} dB(y)$$

where

$$dB(y) = \tau dF^0(y) \int_{z=0}^{z=y} \frac{1}{S^0(z)} \int_{t=0}^{t=z} dQ_t(z) dE(t). \tag{27}$$

Substituting (4) into (27), the integral in t is

$$\begin{aligned}
\int_{t=0}^{t=z} \mathrm{pr}\{Z^0 &\in [z, z+dz), X^0 > Z^0, T^0 \in [t, t+dt) \,|\, T^0 < X^0\} \\
&= \mathrm{pr}\{Z^0 \in [z, z+dz), X^0 > Z^0 \,|\, T^0 < X^0\} \\
&= \mathrm{pr}\{Z^0 \in [z, z+dz) \,|\, T^0 < Z^0 < X^0\}\, \mathrm{pr}\{X^0 > Z^0 \,|\, T^0 < X^0\} \\
&= dG(z) p_0. \qquad (28)
\end{aligned}$$

Substitution of (28) into (27) gives (14).

References

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

[2] Barndorff-Nielsen, O.E. (1978). *Information and Exponential Families in Statistical Theory.* New York: Wiley.

[3] Butler, R.W. (2000). Reliabilities for feedback systems and their saddlepoint approximation. *Statist. Sci.* **15**, 279-298.

[4] Butler, R.W. (2007). *Saddlepoint Approximations with Applications.* Cambridge: Cambridge University Press.

[5] Butler, R.W. & Bronson, D.A. (2002). Bootstrapping survival times in stochastic systems by using saddlepoint approximations. *J. R. Statist. Soc. B* **64**, 31-49.

[6] Efron, B. (1967). The two sample problem with censored data. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* **4**, 831-53.

[7] Frydman, H. (1994). A note on nonparametric estimation of the distribution function from interval-censored and truncated observations. *J. R. Statist. Soc. B* **56**, 71-4.

[8] Holt, J.D. (1978). Competing risk analyses with special reference to matched pair experiments. *Biometrika* **65**, 159-66.

[9] Johansen, S. (1983). An extension of Cox's regression model. *Int. Statist. Rev.* **51**, 258-62.

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

[11] Kaplan, E.L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Statist. Assoc.* **53**, 457-81.

[12] Lagakos, S. W., Sommer, C. J. & Zelen, M. (1978). Semi-Markov models for partially censored data. *Biometrika* **65**, 311-7.

[13] Prentice, R.L., Kalbfleisch, J.D., Peterson Jr., A.V., Flournoy, N, Farewell, V.T. & Breslow, N.E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541-54.

[14] Turnbull, B. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. R. Statist. Soc. B* **38**, 290-5.

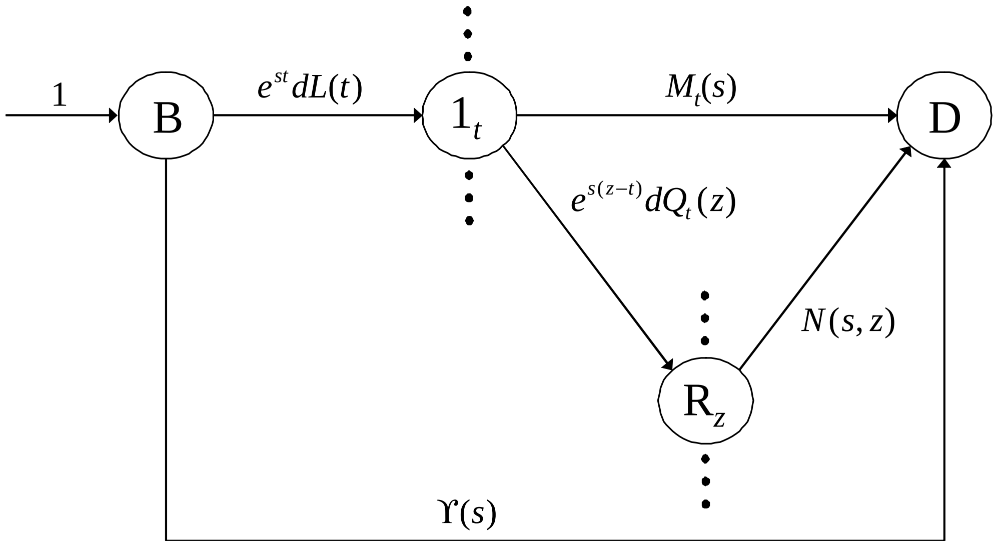

Fig. 1: Semi-Markov flowgraph for a virtual patient's lifetime subject to a single risk.

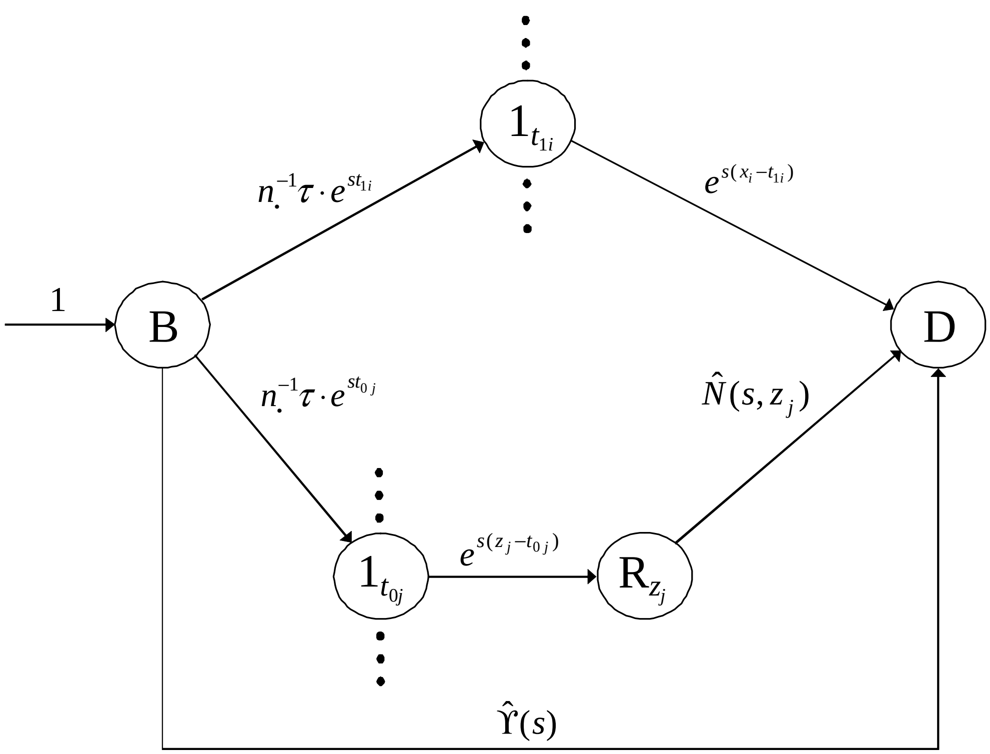

Fig. 2: Empirical semi-Markov flowgraph providing an approximation for the single-risk flowgraph of Figure 1.

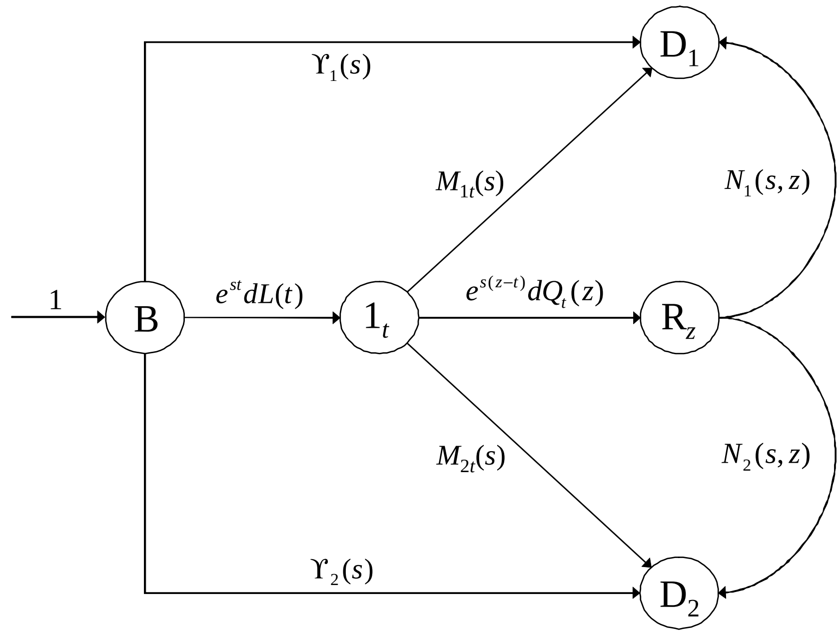

Fig. 3: Semi-Markov flowgraph for a virtual patient's lifetime subject to two competing risks.

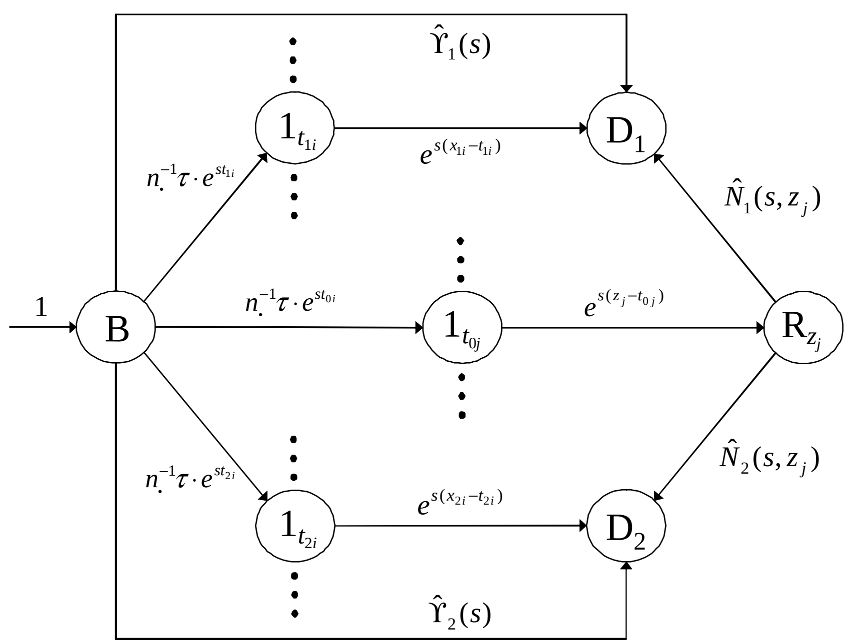

Fig. 4: Empirical flowgraph providing a semi-Markov approximation for the competing-risks flowgraph in Figure 3.