Multiple Imputations Based Estimation of Survival Functions

Sundarraman Subramanian*

Abstract

We derive the asymptotic distribution of the multiple imputations based Kaplan–Meier estimator from right censored data with missing censoring indicators and perform a theoretical and numerical comparison with a competing semiparametric survival function estimator, including a robustness study.

KEY WORDS: Bracketing integral, Dikta semiparametric estimator, Functional delta method, Gaussian process, Maximum likelihood, Stochastic equicontinuity.

1 Introduction

In this article, we focus on semiparametric estimation of a survival function from right censored data with missing censoring indicators (MCIs). For the familiar setting of standard right censorship, the observed random variables are X and δ, where $X = \min(T, C)$, $\delta = I(T \leq C)$ is the censoring indicator, T is the lifetime of interest, and C is an independent censoring variable. It is well known that the Kaplan–Meier (KM) estimator of $S(t)$, the survival function of T, is nonparametric and asymptotically efficient. The estimator is unsuitable when there are MCIs, however.

*Center for Applied Mathematics and Statistics, Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, New Jersey

The data for the MCI model are $\{(X_i, \xi_i, \sigma_i)_{1 \le i \le n}\}$, where $\xi_i = 1$ when δ_i is observed and is 0 otherwise, and $\sigma_i = \xi_i \delta_i$. The censoring indicators δ_i are assumed to be missing at random (MAR), which implies that $P(\xi = 1 | X = x, \delta = d) = P(\xi = 1 | X = x) = \pi(x)$. Letting $p(x) = P(\delta = 1 | X = x)$ this also means that, conditional on X, the missingness and censoring indicators are independent: $P(\sigma = 1 | X = x) = \pi(x)p(x)$. For this model, van der Laan and McKeague (1998), and Subramanian (2004a, 2006) addressed nonparametric estimation of $S(t)$. Unlike for standard right censored data, however, nonparametric estimation of $S(t)$ with MCIs involves smoothing, along with the related complicated issue of bandwidth specification. In addition, it is well known that smoothing produces less precise fits when the data are sparse, which is indeed the case when there are MCIs. For these reasons, a semiparametric approach to estimation is desirable.

Dikta (1998) introduced semiparametric random censorship models, which is an approach that produces estimators having significant efficiency gains over nonparametric estimators when the parametric model for $p(x) = p(x, \theta), \theta \in I\!\!R^p$, is specified correctly. Subramanian (2004b) and Subramanian and Bandyopadhyay (2008) extended Dikta's (1998) approach to the case of MCIs. In practice, however, parametric inference is susceptible to possible misspecification of the model, so much so that it is important to investigate the impact that departures from the true model would have on the efficiency of estimates, see also Satten, Datta, and Williamson (1998). In particular, the robustness of estimators based on semiparametric random censorship models are yet unknown. An alternative approach is imputation of MCIs using the assumed parametric model $p(x, \theta)$ and then computing the estimator based on several "completed" data sets. This multiple imputations approach has been implemented in a regression setting (Lu and Tsiatis 2001; Tsiatis, Davidian and McNeney 2001) and is reported to possess good robustness properties. It would be of interest to examine whether this robustness extends to the present scenario, and also whether or not the estimator outperforms the Dikta semiparametric estimator in the presence of misspecification. In addition, the large sample behavior of the estimator remains to be investigated. These are the main issues that we address in this paper.

The multiple imputations approach for the MCI set up works as follows. We stipulate a suitable parametric model $p(x, \theta)$ for $p(x)$. The parameter is estimated via maximum likelihood. Using the estimated conditional probability, missing δ are imputed to form $M \geq 1$ "synthetic" data sets, each of which is then regarded as a "complete" data set and the KM estimator is computed. The average of the M single imputation estimates of $S(t)$ provides the multiple imputations estimate. Tsiatis, Davidian and McNeney (2001), and Lu and Tsiatis (2001) implemented this method for competing risks with covariates and missing cause of failure information. Multiple imputations was introduced by Rubin (1987) and also investigated by Wang and Robins (1998).

We derive the asymptotic distribution of the multiple imputations estimator using the functional delta method (Gill and Johansen 1990) and the methods of Lu and Tsiatis (2001). When the parametric model $p(x, \theta)$ is correctly specified, it turns out that the Dikta semiparametric estimator is asymptotically more efficient than the multiple imputations based estimator. This is also exemplified through an example and several plots are displayed to illustrate efficiency, as measured by the difference in the asymptotic variances of the estimators at various time points. We also study the effect of misspecification through simulation studies. For many cases the performance of the multiple imputations estimator is superior when the assumed model departs significantly from the true model that generates the data.

In Section 2, we introduce the multiple imputations based estimator and then investigate its large sample properties, including theoretical comparison with the MCI-based Dikta semiparametric estimator. In Section 3, we present the various numerical results and plots. We conclude with a brief conclusion Section. The Appendix includes some technical material.

2 Survival function estimation

Denote the cumulative hazard function by $\Lambda(t)$, and define $H_1(t) = P(X \leq t, \delta = 1)$ and $y(t) = P(X \geq t)$. Survival function estimators are typically compactly differentiable mappings of estimators of these quantities (Gill and Johansen 1990). We first present the rationale for focusing on comparing estimators of $H_1(t)$ rather than estimators $\hat{S}(t)$ of $S(t)$. We next introduce the single imputation estimator of $S(t)$ and derive its large sample distribution theory, and then follow it up with the multiple imputations analysis. Finally we present the theoretical comparison with the MCI-based Dikta semiparametric estimator.

2.1 Reduction through functional delta method

Let τ be such that $H(\tau) < 1$, where H denotes the distribution of X. Let $\hat{H}_1(t)$ denote (any) estimator of $H_1(t)$, and $Y(t)$ denote the empirical estimator of $y(t)$. To perform our stated comparison of estimators of $S(t)$, it will suffice to analyze the large sample behavior of $n^{1/2}(\hat{H}_1(t) - H_1(t))$ for different choices of $\hat{H}_1(t)$. This reduction in analysis is facilitated through the functional delta method of Gill and Johansen (1990), who have shown that the weak convergence of $n^{1/2}\{\hat{S}(t) - S(t)\}$ follows from the weak convergence of the basic bivariate process $n^{1/2}\{\hat{H}_1(t) - H_1(t), Y(t) - y(t)\}$ and compact differentiability of a composition of these two basic processes. Formally, we let (Z_{H_1}, Z_y) denote a bivariate zero-mean Gaussian process, $\|\cdot\|_\infty$ the supremum norm, and $\|\cdot\|_\infty^\vee$ the max supremum norm. Suppose that $n^{1/2}\{\hat{H}_1(t) - H_1(t), Y(t) - y(t)\} \xrightarrow{\mathcal{D}} (Z_{H_1}, Z_y)$, in $(D[0,\tau] \times D_-[0,\tau], \|\cdot\|_\infty^\vee)$, as $n \to \infty$, where the covariance function of the limiting process (Z_{H_1}, Z_y) will (typically) be obtained from the asymptotic representation of $\{\hat{H}_1(t) - H_1(t)\}$ and $\{Y(t) - y(t)\}$. By the functional delta method, $n^{1/2}\{\hat{S}(t) - S(t)\}$ converges weakly in $(D[0,\tau], \|\cdot\|_\infty)$ to $-S(t)\int_0^t y^{-1}(s)dW(s)$ as $n \to \infty$, where $W(t) = Z_{H_1}(t) - \int_0^t Z_y(s)d\Lambda(s)$, see p. 1537 of Gill and Johansen (1990). Note that Z_{H_1} will differ for various choices of $\hat{H}_1$, but the second component of $W(t)$ will remain unchanged, as will be the covariance between $\hat{H}_1(t)$ and $Y(t)$ for the choices $\hat{H}_1(t)$ that we investigate. Therefore, we need only focus on comparing (the two) competing estimators of $H_1(t)$. For the sake of completeness, however, we also provide the asymptotic result for $\hat{S}(t)$ in Subsection 2.4, including the expression for the limiting covariance function.

2.2 The multiple imputations estimator

We introduce some more notation. Specify $p(x)$ parametrically through $p(x) = p(x, \theta)$ where p is known up to the k-dimensional parameter θ. Let θ_0 denote the true value of θ and $p_0(s) = p(s, \theta_0)$. For $r = 1, \ldots, k$, write $p_r(s, \theta_0) = \partial p(s, \theta)/\partial\theta_r|_{\theta=\theta_0}$ and $P_0(s) = [p_1(s, \theta_0), \ldots, p_k(s, \theta_0)]^T$. Let $I_0 = E[\pi(X)P_0(X)P_0^T(X)/(p_0(X)(1 - p_0(X)))]$, where $\pi(s) =$

$P(\xi = 1|X = s)$. Also, let $\alpha(s,t) = P_0^T(s) I_0^{-1} P_0(t)$. The MLE of θ can be obtained by maximizing the likelihood based on complete cases $\xi_i = 1$ (Lu and Tsiatis 2001). For the MLE $\hat{\theta}$, assuming all the standard regularity conditions (cf. Dikta 1998), we have

$$n^{1/2}(\hat{\theta} - \theta_0) = I_0^{-1} n^{-1/2} \sum_{i=1}^{n} \frac{\xi_i(\delta_i - p_0(X_i))}{p_0(X_i)(1 - p_0(X_i))} P_0(X_i) + o_p(1), \tag{1}$$

see Subramanian (2004b). Let $\hat{p}(s) = p(s, \hat{\theta})$. Also, let $\Delta_i(\xi_i, \theta)$ equal δ_i if $\xi_i = 1$ and $\Delta_i(\xi_i, \theta)$ equal, in distribution, to a Bernoulli random variable with success probability $p(X_i, \theta)$ if $\xi_i = 0$. Following Lu and Tsiatis (2001) or Tsiatis et al. (2001), we propose the single imputation estimator of $H_1(t)$ as $\hat{H}_1(t) = n^{-1} \sum_{i=1}^{n} I(X_i \le t) \Delta_i(\xi_i, \hat{\theta})$. For $s \le t$, define

$$V_S(s,t) = H_1(s)(1 - H_1(t)) + \int_0^s \int_0^t (1 - \pi(u)\pi(v))\alpha(u,v) dH(u) dH(v). \tag{2}$$

The proof of the following theorem uses some of the methods of Lu and Tsiatis (2001).

Theorem 1 *Assume that $\theta \in \Theta \subset I\!R^p$ and Θ is bounded. Assume also that $p(x,\theta)$ satisfies the Lipschitz-like condition $|p(x,\theta) - p(x,\theta^*)| \le M(x)\|\theta - \theta^*\|$, with $M(X)$ having finite expectation. Then, the single imputation estimator of $H_1(t)$ converges weakly in $D[0,\tau]$ to a zero mean Gaussian process Z_{H_1} with covariance function $V(s,t)$, for $s \le t$, given by Eq. (2).*

Proof Let $D_i(t,\theta) = I(X_i \le t)\Delta_i(\xi_i,\theta) - H_1(t)$, and $\mu_D(t,\theta)$ denote its expected value. Let $\dot{\mu}_D(t,\theta_0) = \partial \mu_D(t,\theta)/\partial\theta$ evaluated at $\theta = \theta_0$. Note that $\dot{\mu}_D(t,\theta_0)$ is a $k \times 1$ vector. Also,

$$\begin{aligned} \mu_D(t,\theta) &= E\{I(X \le t)\Delta(\xi,\theta)\xi\} + E\{I(X \le t)\Delta(\xi,\theta)(1-\xi)\} - H_1(t) \\ &= E\{\pi(X)p_0(X)I(X \le t)\} + E\{(1-\pi(X))p(X,\theta)I(X \le t)\} - H_1(t), \end{aligned}$$

where the last step is obtained by conditioning on X. It follows that

$$\dot{\mu}_D(t,\theta_0) = E\{P_0(X)(1-\pi(X))I(X \le t)\} = \int_0^t P_0(u)(1-\pi(u)dH(u).$$

We employ standard techniques from empirical process theory to obtain the following result:

$$n^{-1/2}\left[\sum_{i=1}^{n}\left\{D_i(t,\hat{\theta})-\mu_D(t,\hat{\theta})\right\}-\sum_{i=1}^{n}\left\{D_i(t,\theta_0)-\mu_D(t,\theta_0)\right\}\right] \xrightarrow{P} 0. \qquad (3)$$

To prove (3), let U denote a uniform random variable on $(0,1)$, generated independent of the data. Consider the sequence of processes $\{n^{-1/2}Q_n(\theta),\theta\in\Theta\}$, where $Q_n(\theta)=\sum_{i=1}^{n}\{f(\theta)-\mu_D(\theta)\}$, and $f(\theta)=\{\sigma+(1-\xi)I(U\le p(X,\theta))\}\,I(X\le t)-H_1(t)$. Denote by $\mathcal{F}$ the class of functions $\{f(\theta),\theta\in\Theta\}$. Note that for each $\theta, f(\theta)$ is a random variable. In the Appendix, we show that $\mathcal{F}$ is P-Donsker. This means that the sequence of processes $\{n^{-1/2}Q_n(\theta),\theta\in\Theta\}$ converges in distribution to a zero mean Gaussian process. In turn, we have stochastic equicontinuity, from which Eq. (3) follows via Lemma 14.3 of Tsiatis (2006). Then,

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n}D_i(t,\hat{\theta}) &= n^{-1/2}\sum_{i=1}^{n}D_i(t,\theta_0)+n^{1/2}\left(\mu_D(t,\hat{\theta})-\mu_D(t,\theta_0)\right)+o_p(1)\\ &= n^{-1/2}\sum_{i=1}^{n}D_i(t,\theta_0)+(\dot{\mu}_D(t,\theta_0))^T\left[n^{1/2}(\hat{\theta}-\theta_0)\right]+o_p(1). \end{aligned}$$

Using Eq. (1), it follows that $n^{-1/2}\sum_{i=1}^{n}D_i(t,\hat{\theta})=n^{-1/2}\sum_{i=1}^{n}[D_i(t,\theta_0)+C_i(t,\theta_0)]+o_p(1)$, where

$$C_i(t,\theta_0) = \frac{\xi_i(\delta_i-p_0(X_i))}{p_0(X_i)(1-p_0(X_i))}(\dot{\mu}_D(t,\theta_0))^T I_0^{-1}P_0(X_i). \qquad (4)$$

Note that $(X,\Delta(\xi,\theta_0))$ is equal in distribution to (X,δ). Dropping the subscript i, note that $E(D(t,\theta_0))=0$. It can be shown that, for $s\le t$, $E(D(s,\theta_0)D(t,\theta_0))=H_1(s)(1-H_1(t))$.

We can also show that, for $s \le t$,

$$
\begin{aligned}
E(C(s,\theta_0)C(t,\theta_0)) &= (\dot{\mu}_D(s,\theta_0))^T I_0^{-1} E\left[\frac{\pi(X)P_0(X)P_0^T(X)}{p_0(X)(1-p_0(X))}\right] I_0^{-1}\dot{\mu}_D(t,\theta_0) \\
&= (\dot{\mu}_D(s,\theta_0))^T I_0^{-1}\dot{\mu}_D(t,\theta_0) \\
&= \int_0^s \int_0^t (1-\pi(u))(1-\pi(v))\alpha(u,v)dH(v)dH(u).
\end{aligned}
$$

Next we can show that, for $s \le t$,

$$
\begin{aligned}
E(D(s,\theta_0)C(t,\theta_0)) &= E\left[\frac{\xi(\delta-p_0(X))^2}{p_0(X)(1-p_0(X))}P_0^T(X)I(X\le s)\right] I_0^{-1}\dot{\mu}_D(t,\theta_0) \\
&= E\left[\pi(X)P_0^T(X)I(X\le s)\right] I_0^{-1}\dot{\mu}_D(t,\theta_0) \\
&= \int_0^s \pi(u)\left\{\int_0^t (1-\pi(v))\alpha(u,v)dH(v)\right\}dH(u), \\
E(C(s,\theta_0)D(t,\theta_0)) &= (\dot{\mu}_D(s,\theta_0))^T I_0^{-1} E\left\{P_0(X)\pi(X)I(X\le t)\right\} \\
&= \int_0^t \pi(u)\left\{\int_0^s (1-\pi(v))\alpha(v,u)dH(v)\right\}dH(u).
\end{aligned}
$$

Since $\alpha(u,v)=\alpha(v,u)$, it follows that the process $n^{-1/2}\sum_{i=1}^n D_i(t,\hat{\theta})$ converges weakly to a Gaussian process with covariance function given by $V_S(s,t)$, for $s\le t$, see Eq. (2). $\square$

The single imputation estimator described above may be regarded as any one of j imputation estimators for $j=1,\ldots,M$. That is, we may write the above single imputation estimator as the j-th imputation estimator $\hat{H}_{1,j}(t) = n^{-1}\sum_{i=1}^n I(X_i \le t)\Delta_{ij}(\xi_i,\hat{\theta})$. The multiple imputations estimator takes the form $M^{-1}\sum_{j=1}^M \hat{H}_{1,j}(t)$. For $s\le t$, define

$$
V_M(s,t) \;=\; V_S(s,t) - \left(1-\frac{1}{M}\right)\int_0^s (1-p_0(u))(1-\pi(u))dH_1(u). \tag{5}
$$

The theorem below gives the asymptotic distribution of the multiple imputation estimator. Note that the limit (denoted by the same $Z_{H_1}(t)$) is different from the limit in Theorem 1.

Theorem 2 *Under the same conditions as in the statement of Theorem 1, the multiple imputation estimator of* $H_1(t)$ *converges weakly in* $D[0,\tau]$ *to a zero mean Gaussian process* Z_{H_1} *with covariance function* $V_M(s,t)$*, for* $s \le t$*, given by Eq. (5).*

Proof Define $D_{ij}(t,\theta) = I(X_i \le t)\Delta_{ij}(\xi_i,\theta) - H_1(t)$. Using the approximations developed in Theorem 1 it follows that

$$n^{-1/2}\frac{1}{M}\sum_{j=1}^{M}\sum_{i=1}^{n} D_{ij}(t,\hat{\theta}) = n^{-1/2}\sum_{i=1}^{n}\left\{\frac{1}{M}\sum_{j=1}^{M} D_{ij}(t,\theta_0) + C_i(t,\theta_0)\right\} + o_p(1).$$

Write the i-th summand of the above representation as $Q_{ij}(t,\theta_0) = M^{-1}\sum_{j=1}^{M} D_{ij}(t,\theta_0) + C_i(t,\theta_0)$, where $C_i(t,\theta_0)$ is given by Eq. (4). For $s \le t$, let $V_M(s,t) = E\left\{Q_{ij}(s,\theta_0)Q_{ij}(t,\theta_0)\right\}$ (same as Eq. (5), as we show below). Then we have, for $s \le t$, that

$$\begin{aligned}
V_M(s,t) &= \frac{1}{M^2}E\left\{\sum_{j=1}^{M} D_{ij}(s,\theta_0)\sum_{k=1}^{M} D_{ik}(t,\theta_0)\right\} + \frac{1}{M}E\left\{\sum_{j=1}^{M} D_{ij}(s,\theta_0)C_i(t,\theta_0)\right\} \\
&\quad + \frac{1}{M}E\left\{C_i(s,\theta_0)\sum_{j=1}^{M} D_{ij}(t,\theta_0)\right\} + E\left\{C_i(s,\theta_0)C_i(t,\theta_0)\right\} \\
&= \frac{1}{M}E\left\{D_{ij}(s,\theta_0)D_{ij}(t,\theta_0)\right\} + \frac{M-1}{M}E\left\{D_{ij}(s,\theta_0)D_{ik}(t,\theta_0)\right\} \\
&\quad + E\left\{D_{ij}(s,\theta_0)C_i(t,\theta_0)\right\} + E\left\{C_i(s,\theta_0)D_{ij}(t,\theta_0)\right\} + E\left\{C_i(s,\theta_0)C_i(t,\theta_0)\right\} \\
&= V_S(s,t) + \left(\frac{1}{M} - 1\right)E\left\{D_{ij}(s,\theta_0)D_{ij}(t,\theta_0)\right\} + \frac{M-1}{M}E\left\{D_{ij}(s,\theta_0)D_{ik}(t,\theta_0)\right\} \\
&= V_S(s,t) - \left(1 - \frac{1}{M}\right)E\left[D_{ij}(s,\theta_0)\left\{D_{ij}(t,\theta_0) - D_{ik}(t,\theta_0)\right\}\right] \\
&= V_S(s,t) - \left(1 - \frac{1}{M}\right)E\left[\Delta_{ij}(\xi_i,\theta_0)I(X_i \le s)\left\{\Delta_{ij}(\xi_i,\theta_0) - \Delta_{ik}(\xi_i,\theta_0)\right\}\right]
\end{aligned}$$

If $\xi_i = 1$, then $\Delta_{ij}(\xi_i, \theta_0) = \Delta_{ik}(\xi_i, \theta_0) = \delta_i$. If $\xi_i = 0$, then the expectation is non zero only if $\Delta_{ij}(\xi_i, \theta_0) = 1$ and $\Delta_{ik}(\xi_i, \theta_0) = 0$, which has probability $p_0(X)(1 - p_0(X))$. Therefore,

$$\begin{aligned} V_M(s,t) &= V_S(s,t) - \left(1 - \frac{1}{M}\right) E\left\{p_0(X)(1 - p_0(X))(1 - \pi(X)) I(X \le s)\right\} \\ &= V_S(s,t) - \left(1 - \frac{1}{M}\right) \int_0^s (1 - p_0(u))(1 - \pi(u)) dH_1(u). \end{aligned}$$ □

2.3 Theoretical comparison with Dikta semiparametric estimator

Define $V(t) = \lim_{M\to\infty} V_M(t,t)$. Note that this quantity is the smallest variance that the multiple imputation estimator of $H_1(t)$ can hope to achieve, and is given by the expression

$$V(t) = V_S(t,t) - \int_0^t (1 - p_0(u))(1 - \pi(u)) dH_1(u). \tag{6}$$

The MCI model-specific Dikta semiparametric estimator of the survival function of T is given by $\hat{S}_D(t) = \pi_{s\le t}\left(1 - d\hat{\Lambda}_D(s)\right) = \pi_{s\le t}\left(1 - \left\{\hat{p}(s) d\hat{H}(s)/(1 - \hat{H}(s-))\right\}\right)$, where π is the product integral and $\hat{\Lambda}_D(t)$ is the estimator of the cumulative hazard of T. The variance of the Dikta semiparametric estimator of $H_1(t)$ is given by (Subramanian 2004b)

$$V_D(t) = \int_0^t p_0(u) dH_1(u) - H_1^2(t) + \int_0^t \int_0^t \alpha(u,v) dH(u) dH(v). \tag{7}$$

To compare the expressions given by Eqs. (6) and (7), let $R(t) = V(t) - V_D(t)$. Then

$$\begin{aligned} R(t) &= \int_0^t (1 - p_0(u)) dH_1(u) - \int_0^t (1 - p_0(u))(1 - \pi(u)) dH_1(u) \\ &\quad - \int_0^t \int_0^t \pi(u)\pi(v)\alpha(u,v) dH(u) dH(v) \\ &= \int_0^t (1 - p_0(u))\pi(u) dH_1(u) - \int_0^t \int_0^t \pi(u)\pi(v)\alpha(u,v) dH(u) dH(v). \end{aligned}$$

We now show that $R(t) \ge 0$, which means that the Dikta semiparametric estimator is asymptotically more efficient than the multiple imputation estimator of $H_1(t)$. Defining

$b \in I\!R^k$ by $b = [E\{\pi(X)p_1(X,\theta_0)I(X \le t)\}, \ldots, E\{\pi(X)p_k(X,\theta_0)I(X \le t)\}]^T$, we note that the second term of $R(t)$ equals $-\langle b, I_0^{-1}b\rangle$. Write $dH^*(s) = p_0(s)(1-p_0(s))dH(s)$. For any $h \in I\!R^k\backslash\{0\}$, we then have that

$$\begin{aligned}
\langle h, b\rangle^2 &= \left[\int_0^t \pi(u)\left(\sum_{r=1}^k \frac{h_r p_r(u,\theta_0)}{p_0(u)(1-p_0(u))}\right) dH^*(u)\right]^2 \\
&= \left[\int_0^t \pi^{1/2}(u)\left(\sum_{r=1}^k \frac{\pi^{1/2}(u) h_r p_r(u,\theta_0)}{p_0(u)(1-p_0(u))}\right) dH^*(u)\right]^2
\end{aligned}$$

Applying Cauchy-Schwartz inequality, we obtain

$$\begin{aligned}
\langle h, b\rangle^2 &\le \int_0^t \pi(u)(1-p_0(u))dH_1(u) \int_0^t \frac{\pi(u)h^T P_0(u)P_0^T(u)h}{p_0(u)(1-p_0(u))} dH(u) \\
&= \left(\int_0^t \pi(u)(1-p_0(u))dH_1(u)\right)\langle h, I_0 h\rangle.
\end{aligned}$$

From Rao (1973, If1.1) or Dikta (1998), this implies that

$$\langle b, I_0^{-1}b\rangle \equiv \sup_{h\in I\!R^k\backslash\{0\}} \frac{\langle h, b\rangle^2}{\langle h, I_0 h\rangle} \le \int_0^t \pi(u)(1-p_0(u))dH_1(u).$$

Therefore $R(t) \ge 0$.

2.4 Completion of weak convergence result

Recall from Section 2.1 that, assuming continuity of $S(t)$, the process $n^{1/2}\{\hat{S}(t) - S(t)\}$ converges weakly in $(D[0,\tau], \|\cdot\|_\infty)$ to $-S(t)\int_0^t y^{-1}(s)dW(s)$ as $n \to \infty$, where $W(t) = Z_{H_1}(t) - \int_0^t Z_y(s)d\Lambda(s)$. We now compute the covariance function of $W(t)$. Interestingly, unlike in the case of standard right censored data, $W(t)$ does not have independent increments. Let $\eta(t) = \int_0^t I(X_i \ge u)d\Lambda(u) - H_1(t)$, the empirical counterpart of $\int_0^t Z_y(s)d\Lambda(s)$ for $n = 1$. For notational convenience, write $\alpha(t) = M^{-1}\sum_{j=1}^M D_{ij}(t,\theta_0)$. Following some routine calculations, we can show that

$$\begin{aligned}
E\left(\alpha(s)\eta(t)\right) &= H_1(s)\Lambda(s) - \int_0^s H_1(u)d\Lambda(u) - H_1(s)H_1(t), \\
E\left(\alpha(t)\eta(s)\right) &= H_1(t)\Lambda(s) - \int_0^s H_1(u)d\Lambda(u) - H_1(s)H_1(t), \\
E(\eta(s)\eta(t)) &= \int_0^s \Lambda(u)dH_1(u) + \int_0^s (H_1(t) - H_1(u))d\Lambda(u) - H_1(s)H_1(t).
\end{aligned}$$

Adding the above three terms, we obtain the quantity $-H_1(s)H_1(t)$, which is also included in the expression for $V_S(s,t)$ given by Eq. (2). It follows that, for $s \leq t$,

$$\begin{aligned}
\mathrm{Cov}(W(s), W(t)) &= H_1(s) + \int_0^s \int_0^t (1 - \pi(u)\pi(v))\alpha(u,v)dH(u)dH(v) \\
&\quad - \left(1 - \frac{1}{M}\right) \int_0^s (1 - p_0(u))(1 - \pi(u))dH_1(u).
\end{aligned}$$

It now follows by the functional delta method that the multiple imputation estimator $n^{1/2}(\hat{S}(t) - S(t))$ converges weakly to a Gaussian process with covariance function, for $s \leq t$, given by $S(s)S(t)B(s,t)$, where

$$\begin{aligned}
B(s,t) &= \int_0^s \frac{1}{y(u)} d\Lambda(u) + \int_0^s \int_0^t (1 - \pi(u)\pi(v)) \frac{\alpha(u,v)}{y(u)y(v)} dH(u)dH(v) \\
&\quad - \left(1 - \frac{1}{M}\right) \int_0^s \frac{(1 - p_0(u))(1 - \pi(u))}{y(u)} d\Lambda(u).
\end{aligned}$$

The asymptotic variance of the multiple imputation estimator, for large enough M, approaches a quantity which is the same as $S^2(t)$ times $B(t,t)$, but where the factor $(1 - M^{-1})$ in the third term of $B(t,t)$ is replaced with just 1. From Section 2.3, this quantity is at least as large as the asymptotic variance of the Dikta semiparametric estimator of $S(t)$.

3 Numerical results

We first report the results of some simulation studies designed to assess the robustness of the estimators discussed in this paper. Later, we present the results of an efficiency study.

3.1 Robustness study

Simulation 1 For our first simulation study, the minimum X was exponential with mean 1. The true model for the conditional probability was $p(x,\theta) = \exp(\theta_1+\theta_2 x)/(1+\exp(\theta_1+\theta_2 x))$, where the parameter θ_2 was fixed at 5.2, and θ_1 was assigned several values from -1 to 1, giving various censoring rates (CRs) between 5% and 20%. With this set up, we have

$$S(x) = \left\{ \frac{1+\exp(\theta_1+\theta_2 x)}{1+\exp(\theta_1)} \right\}^{-\frac{1}{\theta_2}}.$$

The conditional probability of nonmissingness was $\pi(x) = 1 - \exp(-\exp(x))$, which gives a 15% missingness rate (MR). Misspecification of $p(x)$ was enforced by always fitting the one-parameter model $p(x,\theta_2) = \exp(\theta_2 x)/(1+\exp(\theta_2 x))$ from the generated data. The estimate $\hat{\theta}_2$ was obtained by the Newton–Raphson procedure. The mean integrated squared errors (MISEs) of the multiple imputations based and Dikta semiparametric estimators were computed over the interval $[0, H^{-1}(0.9)]$. The results were based on 10,000 simulations, each of sample size 100, see Table 1 below. Note that the misspecification of $p(x)$ increases when θ_1 is farther away from 0. The multiple imputations based estimator has smaller average MISE and standard deviation than the Dikta semiparametric estimator when θ_2 falls outside the interval $[-0.5, 0.5]$, and hence has superior performance for this range of values of θ_2. For this simulation example, $M = 5$ gives optimal results for most of the cases investigated.

Table 1. Simulation 1. Mean and standard deviation (SD) of mean integrated squared error of the Dikta semiparametric and multiple imputations based estimators. The number of imputations is denoted by M, missingness rate is 15%, and censoring rate varies between 20% ($\theta_1 = -1$) and 5% ($\theta_1 = 1$).

		Dikta semiparametric		*Multiple Imputation*	
θ_1	M	*Mean*	*SD*	*Mean*	*SD*
-1.0	1	0.0109	0.0051	0.0100	0.0048
	5	0.0109	0.0051	0.0097	0.0047
	10	0.0110	0.0051	0.0098	0.0046
	25	0.0111	0.0052	0.0097	0.0047
-0.75	1	0.0096	0.0046	0.0096	0.0046
	5	0.0096	0.0046	0.0094	0.0045
	10	0.0097	0.0046	0.0095	0.0045
	25	0.0097	0.0046	0.0093	0.0045
-0.50	1	0.0088	0.0043	0.0093	0.0045
	5	0.0088	0.0042	0.0090	0.0044
	10	0.0090	0.0043	0.0091	0.0044
	25	0.0089	0.0043	0.0091	0.0044
-0.25	1	0.0085	0.0041	0.0091	0.0044
	5	0.0084	0.0041	0.0088	0.0042
	10	0.0086	0.0041	0.0089	0.0042
	25	0.0085	0.0041	0.0089	0.0043
0.0	1	0.0082	0.0040	0.0088	0.0042
	5	0.0083	0.0040	0.0087	0.0042
	10	0.0084	0.0040	0.0088	0.0041
	25	0.0083	0.0041	0.0087	0.0042
0.25	1	0.0084	0.0041	0.0088	0.0043
	5	0.0083	0.0040	0.0085	0.0041
	10	0.0084	0.0040	0.0086	0.0041
	25	0.0084	0.0041	0.0086	0.0042
0.50	1	0.0085	0.0041	0.0087	0.0042
	5	0.0083	0.0040	0.0083	0.0040
	10	0.0085	0.0041	0.0084	0.0040
	25	0.0085	0.0042	0.0085	0.0041
0.75	1	0.0086	0.0042	0.0086	0.0042
	5	0.0085	0.0041	0.0083	0.0040
	10	0.0087	0.0042	0.0085	0.0041
	25	0.0086	0.0042	0.0084	0.0040
1.0	1	0.0088	0.0043	0.0085	0.0041
	5	0.0088	0.0043	0.0084	0.0041
	10	0.0089	0.0043	0.0085	0.0041
	25	0.0090	0.0045	0.0085	0.0042

Simulation 2 The failure time distribution was Weibull of the type $F(x) = 1 - \exp(-(\alpha x)^2)$. The censoring distribution was exponential with mean 1. Then $p(x, \alpha) = 2\alpha x/(1 + 2\alpha x)$. The conditional probability of nonmissingness was the two parameter logit model: $\pi(x) = \exp(\theta_1 + \theta_2 x)/(1 + \exp(\theta_1 + \theta_2 x))$. The parameters α and θ_1, θ_2 were chosen to get the following combinations of CR and MR: (10%, 44%), (20%, 35%), (25%, 19%), and (42%, 15%). Misspecification of p was induced by always fitting the simple proportional hazards model $p(x, \theta) = \theta, \theta > 0$, from the generated data. The MISEs of the Dikta semiparametric and multiple imputations based estimators, computed over the interval $[0, H^{-1}(0.9)]$, were based on 10,000 simulations, each of sample size 100, see Table 2 below. The multiple imputations based estimator performs better than the Dikta semiparametric estimator.

Table 2. Simulation 2. Mean and standard deviation (SD) of mean integrated squared error of Dikta semiparametric and multiple imputations based estimators for M imputations.

			Dikta semiparametric		*Multiple Imputation*	
CR	*MR*	*M*	*Mean*	*SD*	*Mean*	*SD*
10%	44%	1	0.0006	0.0003	0.0006	0.0003
		5	0.0006	0.0003	0.0006	0.0003
		10	0.0006	0.0003	0.0006	0.0003
		25	0.0006	0.0003	0.0006	0.0003
20%	35%	1	0.0014	0.0007	0.0014	0.0007
		5	0.0015	0.0007	0.0013	0.0006
		10	0.0015	0.0007	0.0013	0.0006
		25	0.0015	0.0006	0.0013	0.0006
25%	19%	1	0.0020	0.0009	0.0018	0.0008
		5	0.0021	0.0009	0.0017	0.0008
		10	0.0020	0.0009	0.0017	0.0008
		25	0.0020	0.0009	0.0017	0.0008
42%	15%	1	0.0054	0.0021	0.0043	0.0021
		5	0.0054	0.0021	0.0041	0.0019
		10	0.0054	0.0021	0.0041	0.0019
		25	0.0054	0.0021	0.0041	0.0019

3.2 Efficiency comparison

The failure and censoring distributions were Weibull, with $F(x) = 1 - \exp(-(\alpha x)^\beta)$ and $G(x) = 1 - \exp(-(\gamma x)^\nu)$. Introducing new parameterizations, given by $\theta_1 = \beta\alpha^\beta/(\nu\gamma^\nu)$ and $\theta_2 = \nu - \beta$, we can write the true model for the conditional probability as $p(x,\theta) = \theta_1/(\theta_1 + x^{\theta_2})$, where $\theta_1 > 0$ and $\theta_2 \in I\!R$, see Dikta, Kvesic, and Schmidt (2006). We used $F(x) = 1 - \exp(-4x)$ and $G(x) = 1 - \exp(-(\nu x)^2)$. Then, $\theta_1 = 2/\nu^2$ and $\theta_2 = 1$, giving $p(x,\nu) = 2/(2+\nu^2 x)$. Note that the censoring rate is given by

$$1 - P(\delta = 1) = 1 - \frac{4}{\nu}\exp\left(\frac{4}{\nu^2}\right)\int_{\frac{2}{\nu}}^{\infty} e^{-t^2}dt.$$

From this, we obtained the values $\nu = 1.037, 1.705, 2.45$, and 3.37 for CRs 10%, 20%, 30% and 40%. We used $\pi(x) = e^x/(1+e^x)$, which gave a missingness rate between 44.5% and 46.3% for the values of ν above. Plots of $R^*(t) = V_M(t,t) - V_D(t)$ versus $t \in [0, H^{-1}(0.9)]$ for $M = 1, 10, 50, 100$ and ∞, and for CRs 10%–40% are shown in figures 1-4.

4 Conclusion

Based on simulation results, there is evidence that the multiple imputations based estimator is likely to be more robust than the Dikta semiparametric estimator. The improvement in its performance is proportional to the degree of misspecification. Although more number of imputations improves the asymptotic efficiency of the estimator in terms of reducing the asymptotic variance, in practice 5 to 10 imputations would seem sufficient. Since misspecification would be, most likely, the norm in practice, the procedure implemented in this article would be very useful to employ when there are MCIs.

Acknowledgements The author thanks Dr. Kaifeng Lu for sharing with him some variance calculations that were not included in the paper of Lu and Tsiatis (2001).

Appendix

We denote by α the maximum distance between any two members of $\Theta \subset I\!\!R^p$. We assume that Θ is bounded. Recall that U denote a uniform random variable on $(0,1)$, generated independent of the data. Also, recall that $Q_n(\theta) = \sum_{i=1}^n \{f(\theta) - \mu_D(\theta)\}$, and $f(\theta) = \{\sigma + (1-\xi)I(U \leq p(X,\theta))\}\, I(X \leq t) - H_1(t)$. Denote by $\mathcal{F}$ the class of functions $\{f(\theta), \theta \in \Theta\}$. Note that for each θ, $f(\theta)$ is a random variable. We show that $\mathcal{F}$ is P-Donsker.

Fix $\rho > 0$. Let C_θ denote a cube of width $2\rho/p^{1/2}$, centered at $\theta \in \Theta$. Then $C_\theta \subset B_\theta$, where $B_\theta = \{\theta' : \|\theta' - \theta\| \leq \rho\}$. Consider the class $\mathcal{F}_\theta = \{f(\theta') : \theta' \in C_\theta \cap \Theta\}$. Note that θ is the center of a (generic) cube with the aforementioned width. The Lipschitz condition, together with $\theta' \in C_\theta \cap \Theta \subset B_\theta$, implies that $p(x,\theta') \leq p(x,\theta) + \rho m(x)$ and $p(x,\theta') \geq p(x,\theta) - \rho m(x)$. It follows that $[f_*(\theta), f^*(\theta)]$ constitutes a bracket for $\mathcal{F}_\theta$, where

$$\begin{aligned} f_*(\theta) &= \{\sigma + (1-\xi)I(U \leq p(X,\theta) - \rho m(X))\}\, I(X \leq t) - H_1(t), \\ f^*(\theta) &= \{\sigma + (1-\xi)I(U \leq p(X,\theta) + \rho m(X))\}\, I(X \leq t) - H_1(t). \end{aligned}$$

Recall that $\gamma = E(m(X)) < \infty$. Since

$$\begin{aligned} E\left\{(f^*(\theta) - f_*(\theta))^2\right\} &= E\{(1-\xi)I(X \leq t)I(p(X,\theta) - \rho m(X) \leq U \leq p(X,\theta) - \rho m(X))\} \\ &= E\{(1-\pi(X))I(X \leq t)2\rho m(X)\} \;\leq\; 2\rho\gamma, \end{aligned}$$

we see that $[f_*(\theta), f^*(\theta)]$ actually constitutes an ϵ bracket for $\mathcal{F}_\theta$ when we take $\rho = \epsilon^2/(2\gamma)$.

We have shown that an ϵ-bracket covering $\mathcal{F}_\theta$ is associated with the cube C_θ whose width is $\epsilon^2/(\gamma p^{1/2})$. We can cover Θ with fewer than $(\alpha\gamma p^{1/2}/\epsilon^2)^p$ cubes. It is clear then that we can cover $\mathcal{F}$ by the same number of ϵ-brackets, that is, the number of such ϵ-brackets is $O(\epsilon^{-2p})$. It follows that the bracketing integral is proportional to $\int_0^1 \sqrt{-\log\epsilon}\, d\epsilon$. The latter integral is, by a change of variable, equal to $\int_0^\infty u^{1/2}e^{-u}du$, which is finite. Thus, by Theorem 19.5 of van der Vaart (1998), the class $\mathcal{F}$ is P-Donsker. This concludes our proof. $\square$

References

Dikta, G., 1998. On semiparametric random censorship models. *J. Statist. Plann. Inference* **66**, 253–279.

Dikta, G., Kvesic, M., and Schmidt, C. (2006). Bootstrap Approximations in Model Checks for Binary Data. *Journal of the American Statistical Association* **101**, 521–530.

Gill, R. D. and Johansen, S. (1990). A survey of product-integration with a view toward application in survival analysis. *Ann. Statist.* **18**, 1501–1555.

Lu K. and Tsiatis, A. A. (2001). Multiple imputation methods for estimating regression coefficients in proportional hazards models with missing cause of failure. *Biometrics* **57**, 1191–1197.

Rao, C. R., 1973. *Linear statistical inference and its applications.* 2nd ed. Wiley, New York.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* Wiley, New York.

Satten, G. A., Datta, S., and Williamson, J. M. (1998). Inference based on imputed failure times for the proportional hazards model with interval-censored data. *Journal of the*

American Statistical Association **93**, 318–327.

Subramanian, S., 2004a. Asymptotically efficient estimation of a survival function in the missing censoring indicator model. *J. Nonpara. Statist.* 16, 797–817.

Subramanian, S., 2004b. The missing censoring-indicator model of random censorship. Handbook of Statistics 23: Advances in Survival Analysis. Eds. N. Balakrishnan and C.R.Rao. 123–141.

Subramanian, S., 2006. Survival analysis for the missing censoring indicator model using kernel density estimation techniques. *Statist. Methodol.* 3, 125–136.

Subramanian, S., Bandyopadhyay, D., 2008. Semiparametric left truncation and right censorship models with missing censoring indicators. *Statist. Probab. Lett.*, **78**, 2572–2577.

Tsiatis, A.A., Davidian, M., and McNeney, B. (2002). Multiple imputation methods for testing treatment differences in survival distributions with missing cause of failure. *Biometrika* **89** 238–244.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data.* Springer, New York.

Van der Laan, M. J. and McKeague, I. W. (1998). Efficient estimation from right-censored data when failure indicators are missing at random. *Ann. Statist.* **26** 164–182.

Van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

Wang, N. and Robins, J. M. (1998). Large sample inference in parametric multiple imputation. *Biometrika* **85** 935–948.

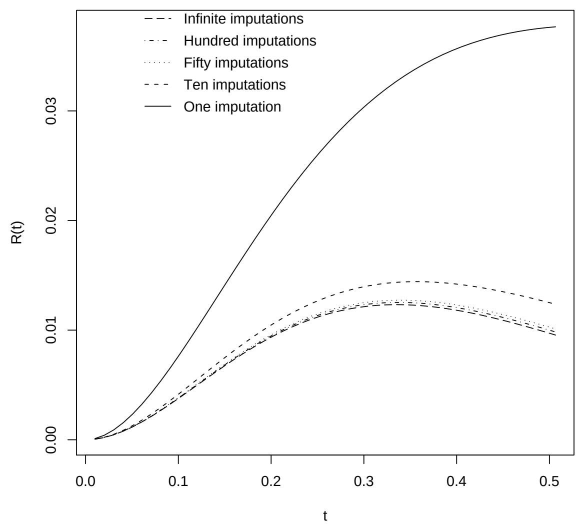

Figure 1: Efficiency comparison for 10% censoring.

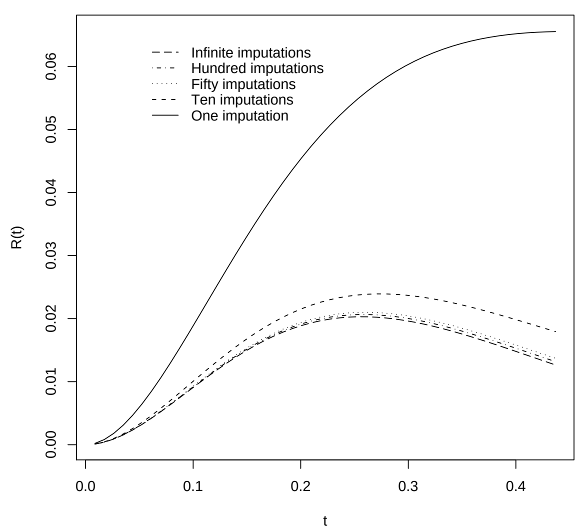

Figure 2: Efficiency comparison for 20% censoring.

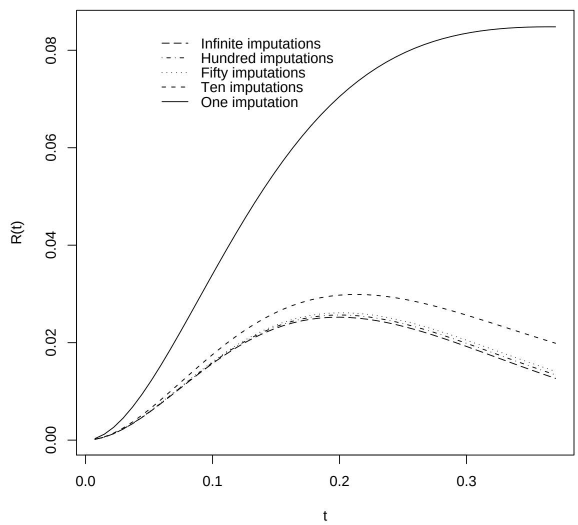

Figure 3: Efficiency comparison for 30% censoring.

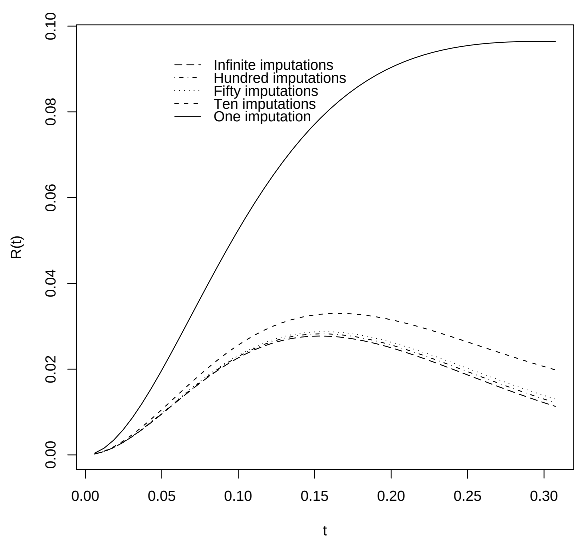

Figure 4: Efficiency comparison for 40% censoring.