Multiple imputations and the missing censoring indicator model

Accepted for publication in **Journal of Multivariate Analysis**

Sundarraman Subramanian
Center for Applied Mathematics and Statistics
Department of Mathematical Sciences
New Jersey Institute of Technology
USA

Abstract

Semiparametric random censorship (SRC) models (Dikta, 1998) provide an attractive framework for estimating survival functions when censoring indicators are fully or partially available. When there are missing censoring indicators (MCIs), the SRC approach employs a model-based estimate of the conditional expectation of the censoring indicator given the observed time, where the model parameters are estimated using only the complete cases. The multiple imputations approach, on the other hand, utilizes this model-based estimate to impute the missing censoring indicators and form several completed data sets. The Kaplan–Meier and SRC estimators based on the several completed data sets are averaged to arrive at the multiple imputations Kaplan–Meier (MIKM) and the multiple imputations SRC (MISRC) estimators. While the MIKM estimator is asymptotically as or less efficient than the standard SRC-based estimator that involves no imputations, here we investigate the performance of the MISRC estimator and prove that it attains the benchmark variance set by the SRC-based estimator. We also present numerical results comparing the performances of the estimators under several misspecified models for the above mentioned conditional expectation.

KEY WORDS: Asymptotic normality, Functional delta method, Lindeberg's condition, Maximum likelihood, Missing at random, Model-based resampling.

[1]Email: sundars@njit.edu

1 Introduction

There are two important approaches of estimating survival functions from right censored data. The nonparametric and most popular approach leads to the Kaplan–Meier (KM) or product limit estimator, which has several appealing properties such as asymptotic efficiency (Wellner, 1982). An alternative approach is based on semiparametric random censorship (SRC) models (Dikta, 1998) and leads to an estimator of the survival function with asymptotic variance not greater than that of the KM estimator, and potentially even smaller. The efficacy of the SRC approach, however, is rooted in the basic premise that the correct model be specified for the conditional expectation of the censoring indicator given the observed, possibly censored, event time; since, otherwise, the estimator would be inconsistent. When the censoring indicators are always available, therefore, the choice between the two approaches may present an intriguing dilemma as it represents a fundamental tradeoff between semiparametric efficiency and nonparametric "robustness" – the KM estimator is consistent, if less efficient than the possibly inconsistent SRC estimator. When there are MCIs, however, the KM estimator is inapplicable, and the "robustness" advantage of nonparametric approaches is perhaps neutralized by the need for smoothing, requiring the specification of data-based optimal bandwidths for computing the estimator (van der Laan and McKeague, 1998; Subramanian, 2004b; 2006; Subramanian and Bean, 2008). Apart from the effort needed to choose a suitable model, the SRC approach has no such frailties, which may well be a significant advantage when there are MCIs (Subramanian 2004a).

The approach of multiple imputations is useful when there are missing data (Rubin, 1987; Satten, Datta, and Williamson, 1998; Wang and Robins, 1998; Lu and Tsiatis, 2001; Tsiatis, Davidian, and McNeney, 2002; Srivastava and Dolatabadi, 2009; Subramanian, 2009). In this approach, the missing components are filled in with imputed values and parameter estimates are obtained from the completed data set, treating the imputed values as though they were actually observed. Estimates from multiple completed data sets are combined in some natural way, such as averaging, to further improve their precision. Kim (2006) investigated the finite sample properties of multiple imputations estimators while Schenker and Welsh (1988) derived asymptotic results.

In this article, we focus on multiple imputations based estimation of a survival function from right censored data with MCIs. For right censorship without MCIs, 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. Dikta

(1998) introduced SRC models, by proposing model-based estimation of the conditional expectation $E(\delta|X = t) = p(\delta = 1|X = t) = p(t)$ and proved that, when the model for $p(t)$ was correctly specified, the SRC estimator of $S(t)$, the survival function of T, was as or more efficient than the KM estimator. The data for the MCI model of random censorship are $\{(X_i, \xi_i, \sigma_i)_{1\leq i\leq n}\}$, where $\xi_i = 1$ when δ_i is observed and is 0 otherwise, and $\sigma_i = \xi_i\delta_i$. Subramanian (2004a) proved that the SRC estimator for the MCI model, denoted by $\hat{S}_D(t)$, was as or more efficient than nonparametric estimators. Subramanian (2009) investigated a multiple imputations based KM estimator (referred henceforth as the MIKM estimator), defined as the average of many single imputation KM estimators, and proved that the MIKM estimator was asymptotically *less efficient* than $\hat{S}_D(t)$. Naturally, the question arises as to whether there are alternative multiple imputations estimators which are better than the MIKM estimator, and whether they would attain the existing benchmark variance set by $\hat{S}_D(t)$. We address this issue by proposing the multiple imputations based SRC estimator, called the MISRC estimator, and derive its asymptotic distribution.

Note that $\hat{S}_D(t)$ is computed without recourse to any imputations. To obtain the model-based estimate of $p(t)$ used for computing $\hat{S}_D(t)$, we choose a suitable good-fitting model $p(t, \theta)$ [from candidates such as logit, probit, generalized proportional hazards, among others, see Dikta (1998)] and estimate the model parameter $\theta \in I\!R^k$ by maximum likelihood based on only the *complete* cases. We denote the maximum likelihood estimator (MLE) by $\hat{\theta}_D$. Estimating θ in this way still produces a consistent estimate under the assumption that the MCIs are *missing at random* (MAR), see Lu and Tsiatis (2001), Tsiatis, Davidian, and McNeney (2002), or Subramanian (2004a). Note that MAR implies that $P(\xi = 1|X = t, \delta = d) = P(\xi = 1|X = t) = \pi(t)$ (Rubin, 1976); and also means that, conditional on X, the missingness and censoring indicators are independent: $P(\sigma = 1|X = t) = \pi(t)p(t)$. The multiple imputations approach involves using the estimated conditional probability $p(t, \hat{\theta}_D)$ to impute missing δ, to form $M \geq 1$ completed data sets, and computing the SRC estimator, denoted by $\hat{S}^{(m)}(t)$. The average of the M single imputation SRC estimates $\hat{S}^{(m)}(t), m = 1, \ldots, M$ provides the MISRC estimator, to be denoted henceforth by $\hat{S}(t)$. Lu and Tsiatis (2001), and Tsiatis, Davidian, and McNeney (2002) implemented this method for competing risks with covariates and missing cause of failure information. We prove that when the model for $p(t)$ is specified correctly, the MISRC estimator $\hat{S}(t)$ is asymptotically equivalent to the SRC estimator $\hat{S}_D(t)$ and hence asymptotically as or more efficient than the MIKM estimator. We also carried out several numerical studies to compare the performance of the estimators when $p(t)$ was misspecified. The MIKM was more robust to misspecification.

Significantly, the multiple imputations procedure has connections with the *model-based resampling*, introduced by Dikta, Kvesic, and Schmidt (2006) for model checking in the context of binary data. Dikta et al. (2006) prescribe the following recipe for standard right censored data (that is, when there are no MCIs): Resample *all* the censoring indicators, based on the estimated model $\hat{p}_D(t) = p(t, \hat{\theta}_D)$. Dikta and Winkler (2009) implemented the extension to MCI data, resampling *only* the *non-missing* censoring indicators. *In contrast, the model-based resampling implicit in our multiple imputations procedure entails resampling (imputing) only the MCIs.* We do *not* resample (impute) the non-missing censoring indicators.

The rest of the article is organized as follows. In Section 2, we derive the asymptotic distribution of the MISRC estimator. In Section 3, we present several numerical results comparing the SRC, MIKM, and MISRC estimators. Section 4 focuses on some discussion and conclusions. Technical complements are included in an Appendix.

2 Multiple imputations estimation

Some of the notation below are from Dikta (1998). Specify a parametric model for $p(t)$ through $p(t) = p(t, \theta)$, where $p(\cdot)$ is known up to the k-dimensional parameter θ. Define

$$q(t, \theta) = \log p(t, \theta), \qquad \bar{q}(t, \theta) = \log(1 - p(t, \theta)).$$

Let θ_0 denote the true value of θ and define $p_0(t) = p(t, \theta_0)$, $q_0(t) = q(t, \theta_0)$, and $\bar{q}_0(t) = \log(1 - p_0(t))$. Note that $q_0(t) = \log p_0(t)$. Write $D_r(p(t, \theta))$ for the partial derivative of $p(t, \theta)$ with respect to θ_r; when it is evaluated at $\theta = \theta^*$, denote it by $D_r(p(t, \theta^*))$. Write $\mathrm{Grad}(p(t, \theta)) = [D_1(p(t, \theta)), \ldots, D_k(p(t, \theta))]^T$ and $C_\theta(t) = \mathrm{Grad}(p(t, \theta))\,(\mathrm{Grad}(p(t, \theta)))^T$. When $\theta = \theta_0$, we denote the matrix $C_{\theta_0}(t)$ by $C_0(t)$. Define the information matrices

$$I(\theta_0) \doteq I_0 = E\left(\frac{C_0(X)}{p_0(X)(1 - p_0(X))}\right), \qquad J(\theta_0) \doteq J_0 = E\left(\frac{\pi(X) C_0(X)}{p_0(X)(1 - p_0(X))}\right).$$

Note that the (r, s) elements of I_0 (case of no MCIs) and J_0 (case with MCIs) are given by

$$i_{r,s} = E\left(\frac{D_r(p_0(X)) D_s(p_0(X))}{p_0(X)(1 - p_0(X))}\right), \qquad j_{r,s} = E\left(\frac{\pi(X) D_r(p_0(X)) D_s(p_0(X))}{p_0(X)(1 - p_0(X))}\right). \quad (1)$$

Also, write $\alpha(u, v) = (\mathrm{Grad}(p_0(u)))^T J_0^{-1} \mathrm{Grad}(p_0(v))$. We denote the second order partial derivatives by $D_{r,s}(\cdot)$. We will need the following assumptions (cf. Dikta et al., 2006):

(**A1**) The functions $D_{r,s}(q(t, \theta))$ and $D_{r,s}(\bar{q}(t, \theta))$ are continuous with respect to θ at each $\theta \in$

$\mathcal{D} \subset I\!R^k$ and $t \in I\!R$. Also, the functions $D_r(q(\cdot,\theta))$, $D_r(\bar{q}(\cdot,\theta))$, $D_{r,s}(q(\cdot,\theta))$ and $D_{r,s}(\bar{q}(\cdot,\theta))$ are measurable for each $\theta \in \mathcal{D}$, and there exists a neighborhood of θ_0, $\mathcal{N}(\theta_0) \subset \mathcal{D}$, and a square integrable measurable function K $[E(K^2(X)) < \infty]$ such that for all $\theta \in \mathcal{N}(\theta_0)$, $t \in I\!R$, and $1 \le r, s \le k$, $|D_{r,s}(q(t,\theta))| + |D_{r,s}(\bar{q}(t,\theta))| + |D_r(q(t,\theta))| + |D_r(\bar{q}(t,\theta))| \le K(t)$.
(**A2**) The matrices I_0 and J_0 are positive definite.
(**A3**) The set $\mathcal{D}$ is bounded and convex and includes the true value θ_0 of θ in its interior. Also, as in **A1**, there exists an integrable envelope function $\tilde{K}(t)$ so that for $1 \le r \le k$, $D_r\left(\log\left(p(t,\theta)/(1-p(t,\theta))\right)\right) \le \tilde{K}(t)$ uniformly for $(t,\theta) \in I\!R \times \mathcal{D}$.

Recall that the MLE of θ, denoted by $\hat{\theta}_D$, is obtained by maximizing a likelihood formed using only the complete cases $\xi_i = 1$: $\tilde{l}_n(\theta) = \sum_{i=1}^n \{\sigma_i q(X_i,\theta) + (\xi_i - \sigma_i)\bar{q}(X_i,\theta)\}$. The MLE is asymptotically linear with influence function $J_0^{-1}\tilde{D}(\theta_0)$ (Subramanian, 2004a), where

$$\tilde{D}(\theta) \;=\; U(\theta)\xi(\delta - p(X,\theta)); \qquad U(\theta) \;=\; \frac{\mathrm{Grad}(p(X,\theta))}{p(X,\theta)(1-p(X,\theta))}, \tag{2}$$

and we deduce by the central limit theorem that $n^{1/2}(\hat{\theta}_D - \theta_0) \xrightarrow{\mathcal{D}} N(0, J_0^{-1})$. Let $\Delta(\xi,\theta,X) \equiv \Delta(\xi,\theta)$ denote a random variable that is defined as follows: When $\xi = 1$, $\Delta(\xi,\theta)$ equals δ; when $\xi = 0$, $\Delta(\xi,\theta)$ equals a Bernoulli random variable having $p(x,\theta)$ as the conditional success probability given $X = x$. That is, conditional on $X = x$ and $\xi = 0$, $\Delta(\xi,\theta)$ induces a Bernoulli distribution with success probability $p(x,\theta)$. The normalized log-likelihood function based on a completed data set is given by

$$l_n^*(\theta) \;=\; \frac{1}{n}\sum_{i=1}^n \left\{\Delta_i(\xi_i,\hat{\theta}_D)q(X_i,\theta) + (1-\Delta_i(\xi_i,\hat{\theta}_D))\bar{q}(X_i,\theta)\right\} \;\doteq\; \frac{1}{n}\sum_{i=1}^n \psi_i(\theta).$$

The m-th imputations-based MLE of θ, which we shall denote by $\hat{\theta}^{(m)}$, is a root of $S_n^*(\theta) = \mathrm{Grad}(l_n^*(\theta)) = \sum_{i=1}^n \mathrm{Grad}(\psi_i(\theta))/n$, where, using Eq. (2), we note that

$$\mathrm{Grad}(\psi_i(\theta)) \;=\; U_i(\theta)\left\{\Delta_i(\xi_i,\hat{\theta}_D) - p(X_i,\theta)\right\}. \tag{3}$$

Proof of consistency of $\hat{\theta}^{(m)}$, the estimator that solves $S_n^*(\theta) = 0$, is given in the Appendix.

Note that there are k partial derivatives of each element of the k-vector $S_n^*(\theta)$. We denote the resulting $k \times k$ matrix by $A_n^*(\theta)$, whose (r,s) element, $a_{r,s}^{n*}(\theta) \equiv D_{r,s}(l_n^*(\theta))$, is given by

$$a_{r,s}^{n*}(\theta) \;=\; \frac{1}{n}\sum_{i=1}^n \left\{\Delta_i(\xi_i,\hat{\theta}_D)D_{r,s}(q(X_i,\theta)) + (1-\Delta_i(\xi_i,\hat{\theta}_D))D_{r,s}(\bar{q}(X_i,\theta))\right\}. \tag{4}$$

Denote the distribution function of X by $H(t)$. As detailed in Subramanian (2009), thanks to Gill and Johansen's (1990) functional version of the delta method, it is enough to focus

on estimators of the subdistribution function $Q(t) = P(X \leq t, \delta = 1)$. More specifically, since each survival function estimator is defined through a series of compactly differentiable mappings beginning with the estimator of $Q(t)$ and $1 - H(t)$, asymptotic equivalence of any two estimators of $Q(t)$ leads to that of the two corresponding survival function estimators. Writing $1 - \hat{H}(t-) = Y(t)$, where $\hat{H}(t)$ is the empirical estimator of $H(t)$, we have

$$(\hat{Q}, Y) \;\longrightarrow\; \left(\hat{Q}, \frac{1}{Y}\right) \;\longrightarrow\; \int_{[0,\cdot]} \frac{1}{Y}\, d\hat{Q} \;\longrightarrow\; \prod_{[0,\cdot]} \left(1 - \frac{1}{Y}\, d\hat{Q}\right) \;\equiv\; \hat{S}. \tag{5}$$

Also, write $\hat{p}_D(t)$ for $p(t, \hat{\theta}_D)$ and define $\hat{Q}_D(t) = \int_0^t \hat{p}_D(s) d\hat{H}(s)$, which is the SRC estimator of $Q(t)$. Note that (Subramanian, 2004a) $n^{1/2}(\hat{Q}_D(t) - Q(t))$ is asymptotically linear with influence function $W(t) + V(t)$, where

$$W(t) \;=\; p_0(X) I(X \leq t) - Q(t), \qquad V(t) \;=\; \frac{\xi\,(\delta - p_0(X))}{p_0(X)(1 - p_0(X))} \int_0^t \alpha(s, X) dH(s). \tag{6}$$

For $m = 1, \ldots, M$, the m-th single imputation estimator of $Q(t)$ is defined as

$$\hat{Q}^{(m)}(t) \;=\; \int_0^t \hat{p}^{(m)}(s) d\hat{H}(s), \tag{7}$$

where $\hat{p}^{(m)}(t) = p(t, \hat{\theta}^{(m)})$. The multiple imputation estimator of $Q(t)$ is defined as $\hat{Q}(t) = \frac{1}{M} \sum_{m=1}^{M} \hat{Q}^{(m)}(t)$. Let $\mathcal{N}_k$ denote the k-variate normal distribution. The following lemma provides an important link to our main result.

Lemma 1 *Suppose that the distribution of X is continuous. Under assumptions* (**A1**) – (**A3**)*, and when the model-based resampling scheme is employed to impute MCIs,*

$$n^{1/2}(\hat{\theta}^{(m)} - \hat{\theta}_D) \;=\; I_0^{-1}\, n^{-1/2} \sum_{i=1}^{n} U_i(\theta_0)(1 - \xi_i)(\Delta_i(\xi_i, \theta_0) - p(X_i, \theta_0)) + o_{\mathbb{P}}(1). \tag{8}$$

Also, $n^{1/2}(\hat{\theta}^{(m)} - \hat{\theta}_D) \xrightarrow{\mathcal{D}} \mathcal{N}_k(0, I_0^{-1} - I_0^{-1} J_0 I_0^{-1})$.

Since the i-th summand on the right hand side of Eq. (8) is just $D_i(\theta_0) - \tilde{D}_i(\theta_0)$, where $D_i(\theta)$ is defined by Eq. (2) and

$$D_i(\theta) \;=\; U_i(\theta)(\Delta_i(\xi_i, \theta) - p(X_i, \theta)), \tag{9}$$

it suffices to show that the left hand side of Eq. (8) is asymptotically equivalent to the centered quantity $I_0^{-1} n^{-1/2} \sum_{i=1}^{n} \left\{ D_i(\theta_0) - \tilde{D}_i(\theta_0) \right\}$. This is proved in the Appendix.

We now state and prove our main result in the following theorem:

Theorem 1 *Under the assumptions stated in Lemma 1, and assuming that the survival function $S(t)$ is continuous, the MISRC estimator $\hat{S}(t)$ is asymptotically equivalent to the standard SRC estimator $\hat{S}_D(t)$.*

Proof Define $\beta(u,v) = [\mathrm{Grad}(p_0(u))]^T I_0^{-1}(I_0 - J_0)I_0^{-1}[\mathrm{Grad}(p_0(v))]$. Where convenient, we write $1-\hat{p}_D(t)$ as $\hat{\bar{p}}_D(t)$ in equation displays below. We first prove the asymptotic equivalence between $\hat{Q}(t)$ and $\hat{Q}_D(t)$ and then extend the result to $\hat{S}(t)$ and $\hat{S}_D(t)$ by invoking the functional delta method. Let $\hat{\theta}^*$ denote a value on the line segment joining $\hat{\theta}^{(m)}$ and $\hat{\theta}_D$. We employ a Taylor expansion of $p(s,\hat{\theta}^{(m)})$ around $\hat{\theta}_D$ to obtain

$$n^{1/2}\left(\hat{Q}^{(m)}(t) - \hat{Q}_D(t)\right) = \left[\int_0^t (\mathrm{Grad}(p(s,\theta_0)))^T dH(s)\right] n^{1/2}(\hat{\theta}^{(m)} - \hat{\theta}_D) + o_{I\!P}(1).$$

By Lemma 1, $n^{1/2}\left(\hat{Q}^{(m)}(t) - \hat{Q}_D(t)\right)$ is asymptotically linear with influence function

$$\left[\int_0^t (\mathrm{Grad}(p_0(s)))^T dH(s)\right] I_0^{-1}\, U(\theta_0)(1-\xi)(\Delta(\xi,\theta_0) - p(X,\theta_0)),$$

and we deduce the asymptotic normality of $n^{1/2}(\hat{Q}^{(m)}(t) - \hat{Q}_D(t))$ with asymptotic variance

$$\int_0^t (\mathrm{Grad}(p_0(s)))^T dH(s)\; I_0^{-1}(I_0 - J_0)I_0^{-1} \int_0^t \mathrm{Grad}(p_0(s))dH(s) = \int_0^t\int_0^t \beta(u,v)dH(v)dH(u).$$

Following the single imputation approximations derived thus, it readily follows that $Z_n(t) = n^{1/2}\left(\hat{Q}(t) - \hat{Q}_D(t)\right)$ is asymptotically linear with influence function

$$\left[\int_0^t (\mathrm{Grad}(p_0(s)))^T dH(s)\right] I_0^{-1}\, U(\theta_0)(1-\xi)\left(\frac{1}{M}\sum_{m=1}^M \Delta^{(m)}(\xi,\theta_0) - p(X,\theta_0)\right),$$

where $\Delta^{(m)}(\xi,\theta_0)$ denotes the m-th imputation-specific random variable which equals $\Delta(\xi,\theta_0)$ in distribution. The asymptotic variance of $Z_n(t)$ is given by

$$\frac{1}{M}\int_0^t\int_0^t \beta(u,v)dH(v)dH(u). \tag{10}$$

To derive the asymptotic distribution of $n^{1/2}(\hat{S}(t) - \hat{S}_D(t))$, denote the limit of $Y(t)$ by $y(t) = 1 - H(t)$ and define $W_n(t) = n^{1/2}(Y(t) - y(t))$. From the representation of (Z_n, W_n) as the normalized sum of i.i.d processes, the finite dimensional distributions of (Z_n, W_n) are multivariate normal with covariance structure for $s \le t$ given by

$$\begin{aligned}
\mathrm{Cov}(Z(s),Z(t)) &= \frac{1}{M}\int_0^s\int_0^t \beta(u,v)dH(v)dH(u) \\
\mathrm{Cov}(W(s),W(t)) &= y(s)(1-y(t)); \qquad \mathrm{Cov}(Z(s),W(t)) = \mathrm{Cov}(W(s),Z(t)) = 0.
\end{aligned}$$

It is well known that the sequence of distributions induced by W_n is tight, see Billinglsey (1968). Tightness of the sequence of distributions induced by Z_n follows as in the proof of Lemma 3.13 of Dikta (1998) and the continuous mapping theorem. Therefore, (Z_n, W_n) induces a tight sequence of distributions on the product space $D[0,\tau] \times D[0,\tau]$, where τ is such that $H(\tau) > 0$. The bivariate process (Z_n, W_n) thus converges weakly to the zero-mean bivariate Gaussian process (Z, W) with covariance structure in the preceding display. From this weak convergence, followed by an application of the functional delta method (cf. Gill and Johansen, 1990, p.1537), we obtain that $n^{1/2}(\hat{S}(t) - \hat{S}_D(t))$ is asymptotically equivalent to $-S(t)\int_0^t \left[d\left\{n^{1/2}\left(\hat{Q}(t) - \hat{Q}_D(t)\right)\right\}/(1-H(s))\right]$. It follows that the asymptotic distribution of $n^{1/2}(\hat{S}(t) - \hat{S}_D(t))$ is normal with mean 0 and variance given by

$$\sigma^2(t) \;=\; \frac{1}{M}\, S^2(t) \int_0^t \int_0^t \frac{\beta(u,v)}{(1-H(u))(1-H(v))} dH(u)dH(v). \tag{11}$$

By Theorem 25.4 of Billingsley (1986), as $M \to \infty$, $\hat{S}(t)$ and $\hat{S}_D(t)$ are asymptotically equivalent. □

3 Numerical results

Since the MIKM estimator is asymptotically as or less efficient than the SRC estimator (Subramanian, 2009), in this section, we first present a numerical study comparing the asymptotic variances of only $n^{1/2}(\hat{S}(t) - S(t))$ and $n^{1/2}(\hat{S}_D(t) - S(t))$ as a function of M, the number of imputations. Then we present the results of several misspecfication studies.

3.1 Comparison of asymptotic variances

The failure time was Weibull with distribution function $F(x) = 1 - \exp(-4x)$. The censoring was independent Weibull with distribution function $G(x) = 1 - \exp(-2x^2/\theta)$, where the censoring parameter $\theta > 0$ was obtained for censoring rates (CRs) 10%–40% from the expression $P(\delta = 0) = 1 - 2\sqrt{2\theta}\, e^{2\theta} \int_{\sqrt{2\theta}}^{\infty} e^{-t^2} dt$. The model for the conditional probability is then $p(x,\theta) = \theta/(\theta + x)$, which is a generalized proportional hazards model (GPHM), see Dikta (1998). We used $\pi(x) = e^x/(1+e^x)$, which gave a missingness rate (MR) of about 45% for the chosen values of θ. Plots of $\sigma^2(t)$ [see Eq. (11)] versus $t \in (0, H^{-1}(0.9)]$ for 10%–40% CRs are shown in Figure 1. The difference in the asymptotic variances of $n^{1/2}(\hat{S}(t) - S(t))$ and $n^{1/2}(\hat{S}_D(t) - S(t))$ declines with increasing M, and is negligible when $M = 25$ or more.

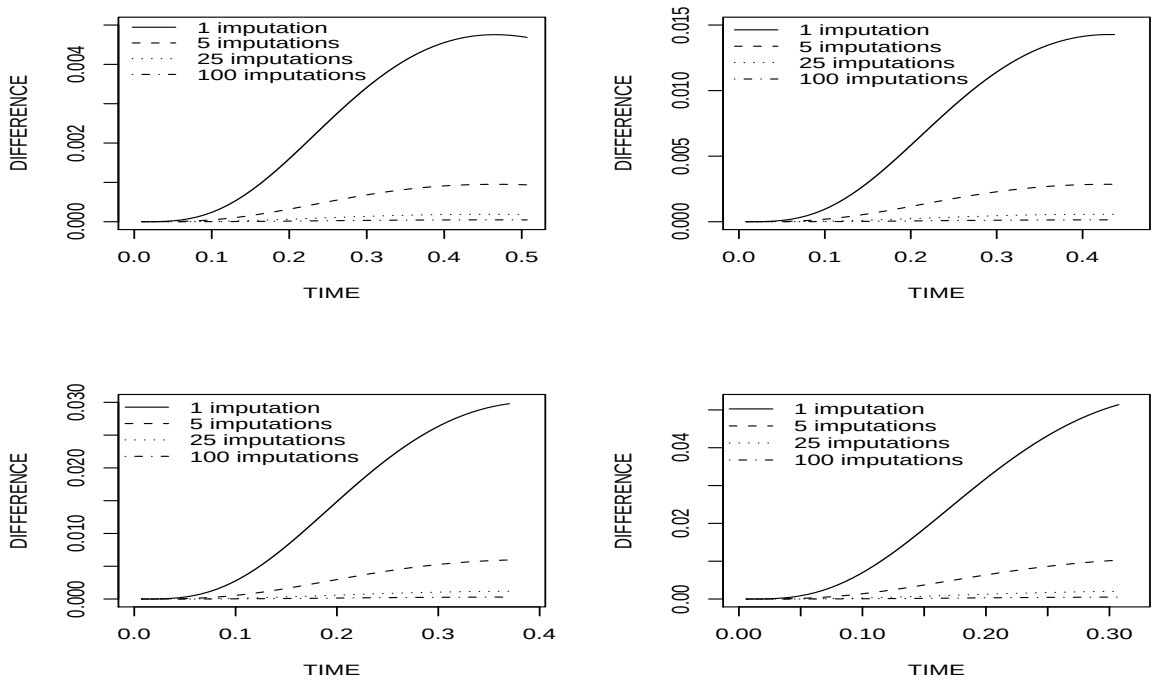

Figure 1: Efficiency comparison between the MISRC and SRC estimators of the survival function $S(t)$ for $t \in (0, H^{-1}(0.9)]$ and different censoring rates: Starting from the top left and in row major order are 10%, 20%, 30% and 40% censoring rates.

3.2 Misspecification performance studies

All our studies were based on 10,000 replications each of sample size 100.

3.2.1 First study

The minimum X was exponential with mean 1 and $p(x, \boldsymbol{\theta}) = \exp(\theta_1 + \theta_2 x)/(1 + \exp(\theta_1 + \theta_2 x))$, where θ_2 was fixed at 5.2 or 0.7. When $\theta_2 = 5.2$, we assigned several values for θ_1 from -2 to 0.5, giving CRs between 5% and 20%. When $\theta_2 = 0.7$, we assigned values for θ_1 from -1 to 1 giving CRs between 30% and 50%. We introduced misspecification of $p(x)$ by fitting $p(x, \theta_2) = \exp(\theta_2 x)/(1 + \exp(\theta_2 x))$ from the generated data. *Note that the misspecification of $p(x, \boldsymbol{\theta})$ increases when θ_1 is farther away from 0.* The MLE $\hat{\theta}_2$ was obtained by the Newton–Raphson procedure. The ranges for θ_1 given above were determined so that the induced

misspecification was not too extreme as to render calculation of the MLE of θ_2 infeasible. The survival function of T takes the form $S(x) = \{1 + \exp(\theta_1 + \theta_2 x)/(1 + \exp(\theta_1))\}^{-\frac{1}{\theta_2}}$. We considered $\pi(x) = 1 - \exp(-\exp(x))$, which gave a 15% MR. The average integrated squared errors (ISEs) of the MIKM, MISRC, and SRC estimators were computed over the interval $[0, H^{-1}(0.9)]$. The MIKM estimator was more robust to misspecification, having best performance when θ_1 was farther away from 0. The SRC and MISRC estimators performed equally well over the entire range of θ_1 that we considered, and better than the MIKM estimator when θ_1 was in the vicinity of 0; see Figure 2.

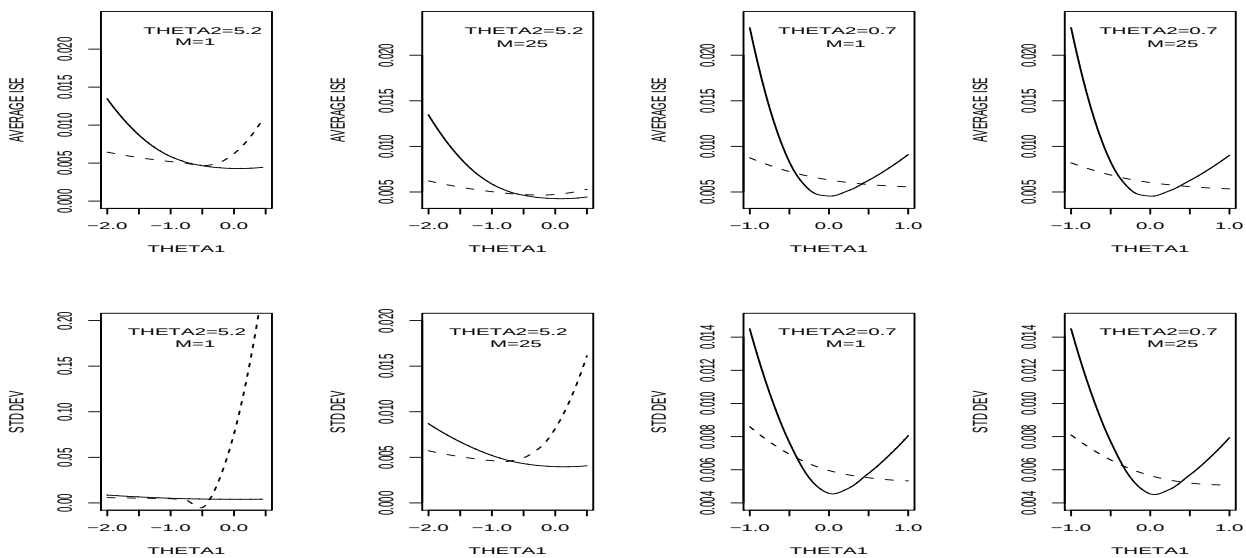

Figure 2: Results for Simulation 1. Estimated mean integrated squared errors (top row) and corresponding standard deviations (bottom row) for various censoring rates and 15% missingness rate. Solid line denotes the SRC, dotted line the MISRC (coincides with the solid line), and dashed line the MIKM estimators. M denotes the number of imputations.

3.2.2 Second study

The minimum X was uniform on $(0, 1)$ and we generated $p(x)$ according to the two-parameter complementary log-log model $p(x, \boldsymbol{\theta}) = 1 - \exp(-\exp(\theta_1 + \theta_2 x))$, where θ_2 was fixed at -4.92 or -5.92 and θ_1 was assigned several values from 3 to 6. When $\theta_2 = -4.92$, the

CR varied between 0% and 30% and when $\theta_2 = -5.92$, the CR varied between 3% and 40%. We introduced misspecification of $p(x)$ by fitting the model $p(x, \theta_2) = 1 - \exp(-\exp(4 + \theta_2 x))$ from the generated data. *Note that the misspecification of* $p(x, \boldsymbol{\theta})$ *increases when* θ_1 *is farther away from 4.* The MLE $\hat{\theta}_2$ was obtained by the Newton–Raphson procedure. The survival function takes the form $S(x) = \exp\left\{-\int_0^x (p(y, \boldsymbol{\theta})/(1-y))\, dy\right\}$. We considered $\pi(x) = \Phi(\rho x)$, where Φ is the cumulative distribution function of the standard normal distribution, with $\rho = 3.98$ giving a 10% MR. For values of θ_1 between 4 and 6, the MISRC estimator performs marginally better; for $\theta_1 \in (3, 4)$ the roles are reversed. The SRC and MISRC estimators coincide; see Figure 3.

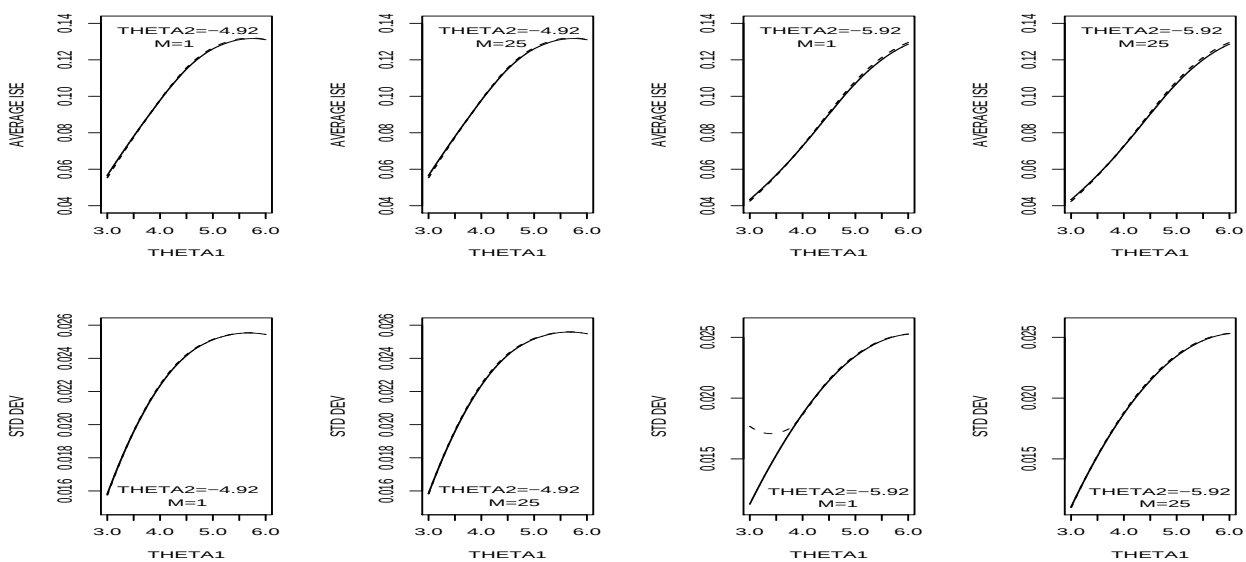

Figure 3: Results for simulation 2. Estimated mean integrated squared errors (top row) and standard deviations (bottom row) for various censoring rates and 10% missingness rate. Solid line denotes the SRC, dotted line the MISRC (coincides with the solid line), and dashed line the MIKM estimators. M denotes the number of imputations.

3.2.3 Third study

We considered the GPHM $p(x, \boldsymbol{\theta}) = \theta_1/(\theta_1 + x^{\theta_2})$, where $\theta_1 > 0$ and $\theta_2 \in \mathbb{R}$, which arises when the failure and censoring distributions are Weibull: $F(x) = 1 - \exp(-(\alpha x)^\beta)$ and

$G(x) = 1 - \exp(-(\gamma x)^{\nu})$, with $\theta_1 = \beta\alpha^{\beta}/(\nu\gamma^{\nu})$ and $\theta_2 = \nu - \beta$, see Dikta (1998). We used $\pi(x, \boldsymbol{\eta}) = \exp(\eta_1 + \eta_2 x)/(1 + \exp(\eta_1 + \eta_2 x))$. For Case 1A, we fixed $(\alpha, \beta, \gamma) = (1.5, 0.8.0.1)$ and for Case 1B we fixed $(1.7, 0.8, 0.3)$. When $\theta_2 = \nu - \beta = 0$, the GPHM reduces to the simple proportional hazards model $p(x, \kappa) = \kappa = \theta_1/(\theta_1 + 1)$, for which the complete cases MLE was $\hat{p}_D(x) = \sum_{i=1}^{n} \sigma_i / \sum_{i=1}^{n} \xi_i$ and the completed set MLE, $\hat{p}^{(m)}(x)$, was the proportion of $\delta = 1$ in the completed data set. We varied ν in a fine grid of values between 0.2 and 1.7 and also set $\eta_1 = 0.2$ and $\eta_2 = 0.5$. With these parameter choices, the CR varied between 8% and 46% for Case 1A, and between 2% and 40% for Case 1B. The MR was between 38% and 41%. We induced misspecification of $p(x)$ by fitting the simple proportional hazards model using the generated data. As ν varied in the selected range, θ_2 varied between -0.6 and 0.9, with $\theta_2 = 0$ representing no misspecification. Except for ν in the interval $(0.5, 1.0)$ (less misspecification), the MIKM performed best and was more robust, see Figure 4.

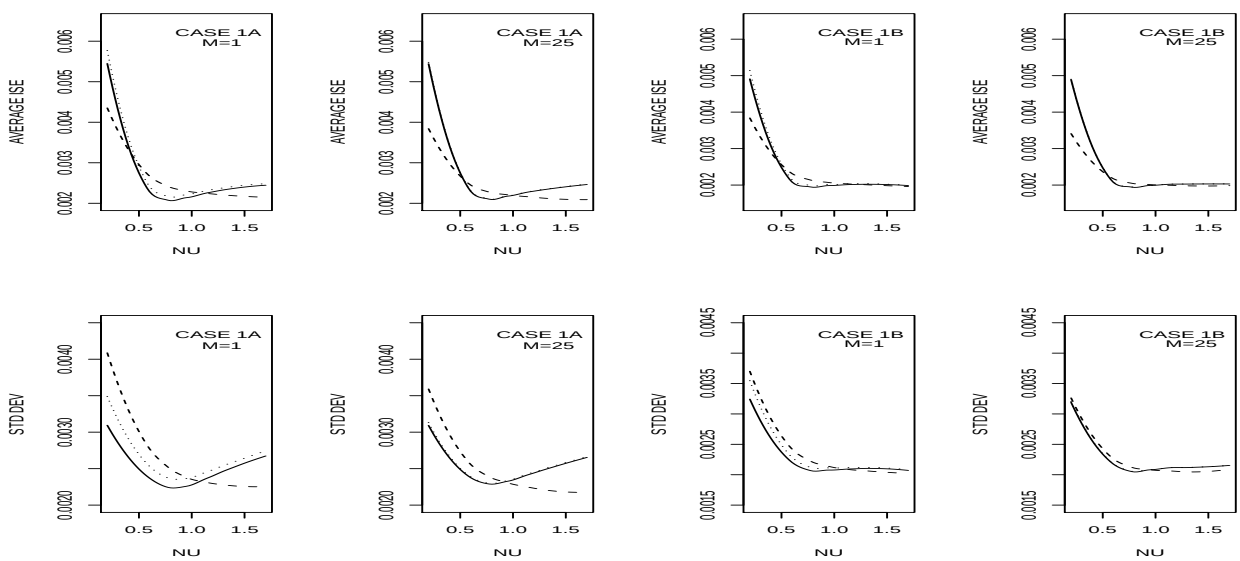

Figure 4: Results for simulation 3A: Estimated mean integrated squared errors (top row) and standard deviations (bottom row). For Case 1A the censoring rate varied from 8 to 46 percent. For Case 1B, it varied between 2 and 40 percent. The missingness rate varied between 38 and 41 percent for both cases. Solid line denotes the SRC, dotted line the MISRC, and dashed line the MIKM estimators. M denotes the number of imputations.

For the GPHM $p(x, \boldsymbol{\theta}) = \theta_1/(\theta_1 + x^{\theta_2})$, we also considered an alternative misspecification by using $\hat{p}_D(x) = \sum_{i=1}^n \sigma_i/(\sum_{i=1}^n(\xi_1(1-\delta_i)))$ and $\hat{p}^{(m)}(x)$ as the same ratio based on the completed data set. Note that the expression for $\hat{p}_D(x)$ is the MLE of θ_1 when the simple proportional hazards model is fitted, and represents high misspecification. For Case 2A we fixed $(\alpha, \beta, \gamma) = (1.5, 0.7, 0.2)$, and for Case 2B we fixed $(\alpha, \beta, \gamma) = (2.0, 0.7, 0.9)$. We varied ν between 0.2 and 1.4, which gave CR between 9% and 44% and MR between 37% and 40% for Case 2A; and CR between 28% and 52% and MR between 40% and 42% for Case 2B. The range of values for ν was chosen to ensure that the denominator of $\hat{p}_D(x)$ would not vanish. The MIKM estimator performed the best, indicating its robustness to misspecification.

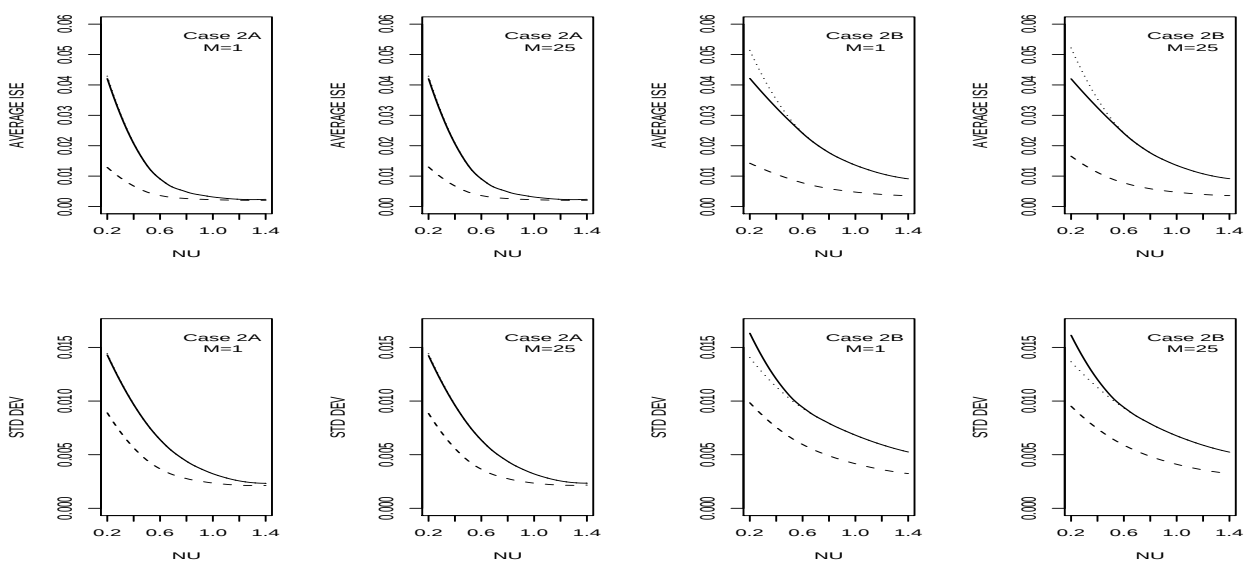

Figure 5: Estimated mean integrated squared errors (top row) and standard deviations (bottom row). For Case 2A the censoring rate varied from 9 to 44 percent and the missingness rate varied between 37 and 40 percent. For Case 2B, the censoring rate varied between 28 and 52 percent and the missingness rate varied between 40 and 42 percent. Solid line denotes the SRC estimator, dotted line denotes the MISRC estimator and dashed line denotes the MIKM estimator. M denotes the number of imputations.

4 Concluding discussion

In this article we have investigated multiple imputations aided SRC estimation of a survival function when the censoring indicators are partially missing. An appealing alternative viewpoint of multiple imputations of MCIs is from the perspective of model-based resampling. Analogous to estimation from standard right censored data involving no MCI's, where the SRC estimator is as or more efficient than the KM estimator when the correct model is specified for $p(t)$, the multiple imputations based SRC estimator is asymptotically as or more efficient than its KM counterpart, the MIKM estimator. In particular we have shown that, when $p(t)$ is correctly specified and as the number of imputations tends to infinity, the MISRC estimator is asymptotically equivalent to the standard SRC estimator of $S(t)$ involving no imputations.

SRC-based estimation of $S(t)$ can suffer from poor estimator performance when there is misspecification, however. Even when a prescribed model for $p(t)$ is considered appropriate for a certain situation, it still may not offer an adequate fit because the observations can have recording or measurement errors. In this context, the MIKM estimator offers a measure of insulation against misspecification, compensating for its asymptotic inefficiency, as is well evidenced by our numerical studies. To provide a rationale for the MIKM estimator's superior performance in the face of uncertainty pertaining to the model for $p(t)$, it may be noted that the standard SRC and MISRC estimators of the subdistribution $Q(t)$ utilize the censoring information only through the model-based estimate of $p(t)$. Misspecification of $p(t)$, therefore, manifests in an unreliable SRC estimate of $Q(t)$. Furthermore, multiple imputations of censoring indicators could snowball an unreliable SRC estimate into an even worse MISRC estimate of $Q(t)$, as seen in Case 2B, see Figure 5. The MIKM estimator, on the other hand, incorporates the censoring indicators directly into estimation, with the result

that, when there is misspecification, the non-missing and the imputed censoring indicators are both utilized to obtain a compromise final estimate having reduced unreliability.

Our finite sample results indicate that, when there is no misspecification, the MISRC estimator generally performs as well as the basic SRC estimator even when M, the number of imputations, is 1. Therefore, its divergence from the latter could be utilized as a sign of possible misspecification. Alternatively, since such divergence implies a potentially greater discrepancy between the MIKM and MISRC estimators, a formal model-check procedure may be implemented using the two estimators by introducing the process $\hat{\mathcal{R}}(t), t \geq 0$:

$$\hat{\mathcal{R}}(t) = n^{1/2}\left\{\hat{Q}_{KM}(t) - \hat{Q}(t)\right\} = n^{-1/2}\sum_{i=1}^{n}\left\{\frac{1}{M}\sum_{m=1}^{M}\left(\Delta^{(m)}(\xi_i, \hat{\theta}_D) - \hat{p}^{(m)}(X_i)\right)\right\} I(X_i \leq t),$$

and employing the well-known Kolmogorov–Smirnov (KS) or Cramér-von Mises (CvM) statistics based on $\hat{\mathcal{R}}(t)$. The process $\hat{\mathcal{R}}(t)$ is the multiple imputations version of the well-studied marked empirical process (Stute, 1997; Stute, González Manteiga, and Presedo Quindimil, 1998; Zhu, Yuen, and Tang, 2002; Dikta et al., 2006; Dikta and Winkler, 2009), whose functional convergence, under the null hypothesis of no misspecification, to a centered Gaussian process can be shown using the techniques developed in this paper. The continuous mapping theorem allows one to deduce distributional convergence of the KS and CvM statistics, whose asymptotic critical values can then be calibrated from their sample counterparts based on $\hat{\mathcal{R}}(t)$. This research direction is the subject of our ongoing investigations and the results will be reported after completion of the specific tasks outlined above.

Acknowledgements The author is grateful to the Editor-in-Chief and an Associate Editor for the opportunity to submit a revised version. The author extends his sincere thanks to a referee for remarks that resulted in a significantly improved version.

Appendix

Consistency of $\hat{\theta}^{(m)}$: Some of the ideas can also be found in Subramanian (2001) or Subramanian (2004c). Let V_γ denote a γ-neighborhood of θ_0. Defining

$$S_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} U_i(\theta)(p(X_i,\theta_0) - p(X_i,\theta)), \tag{A.1}$$

we can show after some calculations that

$$B_n(\theta_0) \doteq \partial S_n(\theta)/\partial\theta|_{\theta=\theta_0} = -\frac{1}{n}\sum_{i=1}^{n}\frac{C_0(X_i)}{p(X_i,\theta_0)(1-p(X_i,\theta_0))},$$

so that by the strong law of large numbers $B_n(\theta_0) \xrightarrow{\text{a.s.}} -\mathbb{E}(C_0(X)/(p_0(X)(1-p_0(X)))) \doteq -I_0$, which is negative definite, by Condition **A2**. Since $S_n(\theta_0) = 0$, this implies that the sequence $S_n(\theta)$ is bounded away from 0 for any $\theta \neq \theta_0$. Hence it suffices to show that $S_n(\hat{\theta}^{(m)}) \xrightarrow{P} 0$ as $n \to \infty$, which would follow if we can prove that (cf. Condition **A3** for reference to $\mathcal{D}$)

$$\sup_{\theta\in\mathcal{D}} \|S_n^*(\theta) - S_n(\theta)\| \quad \xrightarrow{P} \quad 0. \tag{A.2}$$

To prove Eq. (A.2), we utilize the fact that the joint distribution of (X,δ) and $(X,\Delta(\xi,\theta_0))$ are the same (Lu and Tsiatis, 2001). We have that

$$S_n^*(\theta) - S_n(\theta) \;=\; \frac{1}{n}\sum_{i=1}^{n} U_i(\theta)\left(\Delta_i(\xi_i,\hat{\theta}_D) - p(X_i,\theta_0)\right) \;\doteq\; T_{n1}(\theta) + T_{n2}(\theta) + T_{n3}(\theta),$$

where, with $\mathcal{U}_1,\ldots,\mathcal{U}_n$ denoting independent random numbers on $(0,1)$,

$$\begin{aligned}
T_{n1}(\theta) &= \frac{1}{n}\sum_{i=1}^{n} U_i(\theta)\xi_i\left(\delta_i - p(X_i,\theta_0)\right),\\
T_{n2}(\theta) &= \frac{1}{n}\sum_{i=1}^{n} U_i(\theta)(1-\xi_i)\left(I(\mathcal{U}_i \le p(X_i,\theta_0)) - p(X_i,\theta_0)\right),\\
T_{n3}(\theta) &= \frac{1}{n}\sum_{i=1}^{n} U_i(\theta)(1-\xi_i)\left(I(\mathcal{U}_i \le p(X_i,\hat{\theta}_D)) - I(\mathcal{U}_i \le p(X_i,\theta_0))\right).
\end{aligned}$$

Assume for simplicity that the dimension $k = 1$. Write $I_i(\hat{\theta}_D,\theta_0) = I(\mathcal{U}_i \le p(X_i,\hat{\theta}_D)) - I(\mathcal{U}_i \le p(X_i,\theta_0))$. Note that $|I_i(\hat{\theta}_D,\theta_0)| = 1$ with conditional probability $|\hat{p}_D(X_i) - p_0(X_i)|$

when $\hat{p}_D(X_i) \geq p_0(X_i)$ or $\hat{p}_D(X_i) < p_0(X_i)$. Applying Markov's inequality we have

$$\begin{aligned} P\left(\sup_{\theta\in\mathcal{D}} |T_{n3}(\theta)| \geq \epsilon\right) &\leq \frac{1}{\epsilon}\mathbb{E}\left(\sup_{\theta\in\mathcal{D}} |U(\theta)||I_i(\hat{\theta}_D,\theta_0)|\right) \\ &= \frac{1}{\epsilon}\mathbb{E}\left(\sup_{\theta\in\mathcal{D}} |U(\theta)|\cdot|\hat{p}_D(X_i)-p_0(X_i)|\right) \\ &\leq \frac{1}{\epsilon}\mathbb{E}\left(\sup_{\theta\in\mathcal{D}} |U(\theta)| \sup_{\tilde{\theta}\in V_\gamma} |p(X,\tilde{\theta})-p(X,\theta_0)|\right), \end{aligned}$$

which, by Condition **(A3)** and Lebesgue's theorem, tends to 0 as $\gamma \to 0$. To show that $T_{n1}(\theta)$ and $T_{n2}(\theta)$ are each $o_p(1)$ uniformly for $\theta \in \mathcal{D}$, it suffices to focus on the generic version $T_n(\theta) = n^{-1}\sum_{i=1}^n U_i(\theta)(\delta_i - p(X_i,\theta_0))$. Write $\hat{H}_1(t) = n^{-1}\sum_{i=1}^n I(X_i \leq t, \delta = 1)$. Then we can write $n^{1/2}T_n(\theta)$ as the difference of two empirical processes

$$\mathbb{G}_{n1}h = n^{1/2}\int hd(\hat{H}_1 - H_1), \qquad \mathbb{G}_{n2}g = n^{1/2}\int gd(\hat{H} - H),$$

where $H_1(t) = P(X \leq t, \delta = 1)$, and

$$h(t,\theta) = D_1\left(\log\frac{p(t,\theta)}{1-p(t,\theta)}\right), \qquad g(t,\theta) = h(t,\theta)p(t,\theta_0),$$

are indexed by $\theta \in \mathcal{D}$; see sections 19.2 – 19.7 of van der Vaart (1998). By Example 19.7 of van der Vaart (1998) the class $\mathcal{H} = \{h(\cdot,\theta) : \theta \in \mathcal{D}\}$ is Donsker when $h(\cdot,\theta)$ satisfies a mild Lipshitz-like condition. In an analogous way, the class $\mathcal{G} = \{g(\cdot,\theta) : \theta \in \mathcal{D}\}$ is also Donsker. Therefore we deduce by the continuous mapping theorem that $\sup_{\theta\in\mathcal{D}} \mathbb{G}_{n1}(\cdot,\theta)$ and $\sup_{\theta\in\mathcal{D}} \mathbb{G}_{n2}(\cdot,\theta)$ converge weakly to Gaussian limits. This implies that $T_n(\theta) = o_p(1)$ uniformly for $\theta \in \mathcal{D}$, completing the proof of the consistency of $\hat{\theta}^{(m)}$. □

Proof of Lemma 1 We employ some ideas from Dikta et al. (2006). Taylor expansion of $S_n^*(\hat{\theta}^{(m)})$ about $\hat{\theta}_D$ yields $0 = S_n^*(\hat{\theta}_D) + [A_n^*(\hat{\theta}^*)](\hat{\theta}^{(m)} - \hat{\theta}_D)$, where $\hat{\theta}^*$ is a value on the line segment joining $\hat{\theta}^{(m)}$ and $\hat{\theta}_D$, from which we deduce that $n^{1/2}[A_n^*(\hat{\theta}^*)](\hat{\theta}^{(m)} - \hat{\theta}_D) = -n^{1/2}S_n^*(\hat{\theta}_D)$. We find the limit of $A_n^*(\hat{\theta}^*)$ and then prove the asymptotic normality of

$n^{1/2}S_n^*(\hat{\theta}_D)$. Recall that $A_n^*(\theta) = \left[a_{r,s}^{n*}(\theta)\right]_{1\le r,s\le k}$, where

$$a_{r,s}^{n*}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\Delta_i(\xi_i,\hat{\theta}_D)D_{r,s}(q(X_i,\theta)) + (1-\Delta_i(\xi_i,\hat{\theta}_D))D_{r,s}(\bar{q}(X_i,\theta))\right\}.$$

We will also need the quantity

$$a_{r,s}^{n}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\Delta_i(\xi_i,\theta_0)D_{r,s}(q(X_i,\theta)) + (1-\Delta_i(\xi_i,\theta_0))D_{r,s}(\bar{q}(X_i,\theta))\right\}.$$

Let V_γ denote a gamma-neighborhood of θ_0. Since

$$\begin{aligned} a_{r,s}^{n*}(\theta) - a_{r,s}^{n*}(\theta_0) &= \frac{1}{n}\sum_{i=1}^{n}\Delta_i(\xi_i,\hat{\theta}_D)\left(D_{r,s}(q(X_i,\theta)) - D_{r,s}(q(X_i,\theta_0))\right) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(1-\Delta_i(\xi_i,\hat{\theta}_D))(D_{r,s}(\bar{q}(X_i,\theta)) - D_{r,s}(\bar{q}(X_i,\theta_0))), \end{aligned}$$

we note that

$$\begin{aligned} |a_{r,s}^{n*}(\theta) - a_{r,s}^{n*}(\theta_0)| &\le \frac{1}{n}\sum_{i=1}^{n}|D_{r,s}(q(X_i,\theta)) - D_{r,s}(q(X_i,\theta_0))| \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}|D_{r,s}(\bar{q}(X_i,\theta)) - D_{r,s}(\bar{q}(X_i,\theta_0))|, \end{aligned}$$

so that

$$\begin{aligned} \sup_{\theta\in V_\gamma}|a_{r,s}^{n*}(\theta) - a_{r,s}^{n*}(\theta_0)| &\le \frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in V_\gamma}|D_{r,s}(q(X_i,\theta)) - D_{r,s}(q(X_i,\theta_0))| \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in V_\gamma}|D_{r,s}(\bar{q}(X_i,\theta)) - D_{r,s}(\bar{q}(X_i,\theta_0))|. \end{aligned}$$

We employ Markov's inequality to get

$$\begin{aligned} P(|a_{r,s}^{n*}(\hat{\theta}^*) - a_{r,s}^{n*}(\theta_0)| > \epsilon) &\le P(\sup_{\theta\in V_\gamma}|a_{r,s}^{n*}(\theta) - a_{r,s}^{n*}(\theta_0)| > \epsilon) \\ &\le \frac{1}{\epsilon}\mathbb{E}(\sup_{\theta\in V_\gamma}|a_{r,s}^{n*}(\theta) - a_{r,s}^{n*}(\theta_0)|). \end{aligned}$$

The expectation on the right hand side is bounded above by

$$\mathbb{E}\left(\sup_{\theta\in V_\gamma}|D_{r,s}(q(X,\theta)) - D_{r,s}(q(X,\theta_0))| + \sup_{\theta\in V_\gamma}|D_{r,s}(\bar{q}(X,\theta)) - D_{r,s}(\bar{q}(X,\theta_0))|\right),$$

which, by Condition **(A1)** and Lebesgue's theorem, tends to 0 as $\gamma \to 0$. Therefore, we have

$$A_n^*(\hat{\theta}^*) = A_n^*(\theta_0) + o_{I\!P}(1).$$

Furthermore, we can write $a_{r,s}^{n*}(\theta_0) = a_{r,s}^n(\theta_0) + T_{n4}(\theta_0) + o_{I\!P}(1)$, where

$$\begin{aligned} T_{n4}(\theta_0) &= \frac{1}{n}\sum_{i=1}^n \Big\{ \Big(\Delta_i(\xi_i, \hat{\theta}_D) - \Delta_i(\xi_i, \theta_0)\Big) D_{r,s}(q(X_i, \theta_0)) \\ &\quad - \Big(\Delta_i(\xi_i, \hat{\theta}_D) - \Delta_i(\xi_i, \theta_0)\Big) D_{r,s}(\bar{q}(X_i, \theta_0)) \Big\}. \end{aligned}$$

Now, $T_{n4}(\theta_0) = o_{I\!P}(1)$, see the proof of consistency above dealing with a reminder term T_{n3}. Alternatively, we can prove this using empirical process theory, see Eq. (A.3) and the arguments accompanying it below. Using iterated expectation with conditioning on X, it follows from the conditional independence between ξ and δ given X that

$$\begin{aligned} I\!E(a_{r,s}^n(\theta_0)) &= I\!E\left(\Delta(\xi, \theta_0) D_{r,s}(q(X, \theta_0)) + (1 - \Delta(\xi, \theta_0)) D_{r,s}(\bar{q}(X, \theta_0))\right) \\ &= I\!E\Big[\pi(X)\left\{p_0(X) D_{r,s}(q(X, \theta_0)) + (1 - p_0(X)) D_{r,s}(\bar{q}(X, \theta_0))\right\} \\ &\quad + (1 - \pi(X))\left\{p_0(X) D_{r,s}(q(X, \theta_0)) + (1 - p_0(X)) D_{r,s}(\bar{q}(X, \theta_0))\right\}\Big] \\ &= I\!E\left[p_0(X) D_{r,s}(q_0(X)) + (1 - p_0(X)) D_{r,s}(\bar{q}_0(X))\right] \\ &= -I\!E\left[\frac{D_r(p_0(X)) D_s(p_0(X))}{p_0(X)(1 - p_0(X))}\right] \equiv -i_{r,s}. \end{aligned}$$

We conclude that $A_n^*(\hat{\theta}^*) = -I_0 + o_{I\!P}(1)$.

We next obtain an asymptotic representation for $n^{1/2} S_n^*(\hat{\theta}_D)$. Recall that

$$n^{1/2} S_n^*(\hat{\theta}_D) = n^{-1/2} \sum_{i=1}^n \mathrm{Grad}(\psi_i(\hat{\theta}_D)) \;=\; n^{-1/2} \sum_{i=1}^n D_i(\hat{\theta}_D),$$

where $D_i(\theta) = U_i(\theta)(\Delta_i(\xi_i, \theta) - p(X_i, \theta))$, see Eq. (3) and Eq. (9). Let $\mu_D(\theta)$ denote its expected value. Let $\dot{\mu}_D(\theta_0) = \partial \mu_D(\theta)/\partial\theta|_{\theta=\theta_0}$. Since $E(\Delta_i(\xi_i, \theta)|X_i) = \pi(X_i) p_0(X_i) + (1 - \pi(X_i) p(X_i, \theta)$, using iterated expectation with conditioning on X, we have

$$\begin{aligned} \mu_D(\theta) &= E\left[U(\theta)\left\{\pi(X) p_0(X) + (1 - \pi(X)) p(X, \theta) - p(X, \theta)\right\}\right] \\ &= E\left[U(\theta) \pi(X)(p(X, \theta_0) - p(X, \theta))\right], \end{aligned}$$

from which we get $\dot{\mu}_D(\theta) = E\left[\dot{U}(\theta)\pi(X)(p(X,\theta_0) - p(X,\theta)) - U(\theta)\pi(X)(\mathrm{Grad}(p(X,\theta)))^T\right]$, so that $\dot{\mu}_D(\theta_0) = -J_0$. Working exactly as in the Appendix of Subramanian (2009) it can be seen that

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

Recall that $\tilde{D}_i(\theta_0) = U_i(\theta_0)\xi_i(\delta_i - p(X_i,\theta_0))$, see Eq. (2). From Eq. (A.3) we now have that

$$\begin{aligned} n^{1/2}S_n^*(\hat{\theta}_D) \equiv n^{-1/2}\sum_{i=1}^{n} D_i(\hat{\theta}_D) &= n^{-1/2}\sum_{i=1}^{n} D_i(\theta_0) + n^{1/2}\left(\mu_D(\hat{\theta}_D) - \mu_D(\theta_0)\right) + o_{I\!P}(1) \\ &= n^{-1/2}\sum_{i=1}^{n} D_i(\theta_0) + \dot{\mu}_D(\theta_0)\, n^{1/2}(\hat{\theta}_D - \theta_0) + o_{I\!P}(1) \\ &= n^{-1/2}\sum_{i=1}^{n} D_i(\theta_0) - J_0 J_0^{-1} n^{-1/2}\sum_{i=1}^{n}\tilde{D}_i(\theta_0) + o_{I\!P}(1), \end{aligned}$$

where we used the asymptotic representation for $n^{1/2}(\hat{\theta}_D - \theta_0)$, see Eq. (2). It follows that

$$n^{1/2}(\hat{\theta}^{(m)} - \hat{\theta}_D) = I_0^{-1} n^{-1/2}\sum_{i=1}^{n}\left(D_i(\theta_0) - \tilde{D}_i(\theta_0)\right) + o_{I\!P}(1).$$

Clearly $D_i(\theta_0) - \tilde{D}_i(\theta_0) = U_i(\theta_0)(1-\xi_i)(\Delta_i(\xi_i,\theta_0) - p(X_i,\theta_0))$, which is a centered quantity since the conditional expectation of $\Delta_i(\xi_i,\theta_0)$ given X_i is $p(X_i,\theta_0)$. It is easy to show that

$$E\left[\left(D(\theta_0) - \tilde{D}(\theta_0)\right)\left(D(\theta_0) - \tilde{D}(\theta_0)\right)^T\right] = I_0 - J_0.$$

By the multivariate central limit theorem, we infer that $n^{1/2}(\hat{\theta}^{(m)} - \hat{\theta}_D)$ is asymptotically distributed as $\mathcal{N}_k\left(0, I_0^{-1}(I_0 - J_0)I_0^{-1}\right)$ distribution. Proof of Lemma 1 is completed. □

References

P. Billingsley, Convergence of probability measures, Wiley, New York, 1968.

P. Billingsley, Probability and measure, Wiley, New York, 1986.

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

G. Dikta, M. Kvesic, C. Schmidt, Bootstrap approximations in model checks for binary data, J.Ȧmer. Statist. Assoc. 101 (2006) 521–530.

G. Dikta, T, Winkler, Zur Diagnostik binärer Datenmodelle unter unvollständigen Beobachtungen, Masters thesis, Fachhochschule Aachen, Standort Jülich, 2009.

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

J.K. Kim, Finite sample properties of multiple imputation estimators, Ann. Statist. 32 (2006) 766–783.

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

D.B. Rubin, Inference and missing data, Biometrika 63 (1976) 581–590.

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

G.A. Satten, S. Datta, J.M. Williamson, Inference based on imputed failure times for the proportional hazards model with interval-censored data, J. Amer. Statist. Assoc. 93 (1998) 318–327.

N. Schenker, A.H. Welsh, Asymptotic results for multiple imputation, Ann. Statist. 16 (1988) 1550–1566.

M.S. Srivastava, M. Dolatabadi, Multiple imputation and other resampling schemes for imputing missing observations, J. Multivar. Anal. (In press).

W. Stute, Nonparametric model checks for regression, Ann. Statist. 25 (1997) 613–641.

W. Stute, W. González Manteiga, M. Presedo Quindimil, Bootstrap approximations in model checks for regression, J. Amer. Statist. Assoc. 93 (1998) 141–149.

S. Subramanian, Parameter estimation in regression for long-term survival rate from censored data, J. Statist. Plann. Inference 99 (2001) 211–222.

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

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

S. Subramanian, Survival-rate regression using kernel conditional KaplanMeier estimators, J. Statist. Plann. Inference 123 (2004c) 187–205.

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

S. Subramanian, The multiple imputations based Kaplan–Meier estimator, Statist. Probab. Lett. 79 (2009), 1906–1914.

S. Subramanian, D. Bean, The missing censoring indicator model and the smoothed bootstrap, Computat. Statist. Data Anal. 53 (2008) 471–476.

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

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

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

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

J.A. Wellner, Asymptotic optimality of the product limit estimator, Ann. Statist. 10 (1982) 595–602.

L.X. Zhu, K.C. Yuen, N.Y. Tang, Resampling methods for testing random censorship models, Scand. J. Statist. 29 (2002) 111–123.