Statistica Sinica **18**(2008), 749-768

ASYMPTOTICALLY EFFICIENT PRODUCT-LIMIT ESTIMATORS WITH CENSORING INDICATORS MISSING AT RANDOM

Qihua Wang[1,2] and Kai W. Ng[2]

[1]*Chinese Academy of Science and* [2]*The University of Hong Kong*

Abstract: In this paper, we develop methods for estimating a survival function with censoring indicators missing at random. The resulting methods lead to the use of imputation and inverse probability weighting. We give several asymptotically efficient PL estimators. All the estimators are proved to be strongly uniformly consistent and weakly convergent to a Gaussian process. Further, it is shown that these estimators are asymptotically efficient. A simulation study was carried out to evaluate the finite sample performances of the proposed estimators and compare the proposed estimators with van der Laan and McKeague's (1998) estimator under missing at random (MAR) and missing completely at random (MCAR) assumptions, respectively.

Key words and phrases: Missing at random, product-limit estimator, random censorship.

1. Introduction

Statistical analysis of lifetime or failure time data is frequently based on censored observations. Under random censorship, Kaplan and Meier (1958) suggested a product-limit (PL) estimator to estimate a survival function. The PL estimator is widely used in survival analysis and has been extensively investigated. It has many desirable properties such as asymptotic efficiency (Wellner (1982)).

To describe the product-limit estimator, let T denote a random variable representing lifetime with distribution function (*d.f.*) F, and let C denote a random variable describing right censoring time with *d.f.* G. It is assumed that T is independent of C. Under random censorship, one observes (X, δ), where $X = T \wedge C$ and $\delta = I[T \leq C]$, with $I(\cdot)$ the indicator function. Suppose that the data consist of independent and identically distributed observations $\{X_i, \delta_i\}$ for $i = 1, 2, \ldots, n$. Kaplan and Meier (1958) defined the PL estimator of the survival distribution $S(t) = 1 - F(t)$ as

$$\widehat{S}_{KM}(t) = \prod_{i:X_i \leq t} \Big(\frac{n - R_i}{n - R_i + 1}\Big)^{\delta_i}, \tag{1.1}$$

where R_i denotes the rank of X_i with the X-sample.

Clearly, the PL estimator requires that the censoring indicator is always observed. In some practical problems, however, the censoring indicator δ (or cause of failure information) is missing for a variety of reasons. For example, in a bioassay experiment some subjects might not be autopsied to save expense, or autopsy and hospital case notes can be inconclusive; in epidemiological studies relevant death certificate information can be missing due to emigration. When the censoring indicator is missing, a simple method is to ignore the missing data and invoke the Kaplan-Meier PL estimator. However, this complete case (CC) estimator is highly inefficient if there is a significant degree of missingness. Also, this estimator is consistent only for the special case when the censoring indicator is missing completely at random (MCAR), in the sense that the missing mechanism is independent of everything else.

Under MCAR, some authors have proposed improvements on the CC estimator. The first attempt was made by Dinse (1982), who used the EM algorithm to obtain a non-parametric maximum likelihood estimate (NPMLE). Lo (1991) showed that there are infinitely many NPMLEs and some of them are consistent; he constructed two alternative estimators, one of which is consistent and asymptotically normal. Gijbels, Lin and Ying (1993) and McKeague and Subramanian (1998) further improved this estimator. Increasing attention has been paid to survival analysis with missing censoring information in recent years, and not limited to the estimation of survival functions. Goetghebeur and Ryan (1990) derived a modified logrank test to compare survival in two groups, and Dewanji (1992) suggested a modification of that approach. Goetghebeur and Ryan (1995) extended the results of Goetghebeur and Ryan (1990) to the proportional hazard regression models. Tsiatis, Davidian and Mcneney (2002) used a form of multiple imputation methods for testing treatment differences in survival distributions.

When censoring indicators are missing under random censorship, the observed data are (X_i, δ_i, ξ_i), where the $X_i's$ are always observed and $\xi_i = 0$ if δ_i is missing, otherwise $\xi_i = 1$. Throughout this paper, we assume that δ is missing at random (MAR). The MAR assumption implies that ξ and δ are conditionally independent given X. That is, $P(\xi = 1|X, \delta) = P(\xi = 1|X)$. MCAR described above is a special case of MAR. MAR is a common assumption for statistical analysis with missing data and is reasonable in many practical situations, see Little and Rubin (1987, Chap. 1). Recently, van der Laan and McKeague (1998) found an estimator of $S(t)$ by an artificial reduced data approach. This estimator is shown to be asymptotically efficient for the reduced data under a slightly stronger missing assumption (See their formula (9)) than the usual MAR. Their estimating approach is to find the nonparametric maximum likelihood estimator (NPMLE) of $S(t)$ based on reduced data produced by a discretization of X that depends on a partition. The estimator can be unappealing in practice, especially for small samples, since it requires a special partition, some artificially chosen points and an artificial binning of the data. For small sample sizes, the estimator may have serious bias since the approach restricts the NPMLE of $S(t)$ to be discrete with point masses at all complete observations (X_i, δ_i) and on one (or more) artificially chosen point in each region that contains no complete observations.

In this paper, we develop methods that produce several asymptotically efficient PL estimators of $S(t)$. Our estimators make more efficient use of the available data, they do not require binning of the data, and are easier to calculate and implement in practice. Under the usual MAR assumption, all the estimators are proved to be strongly uniformly consistent and weakly convergent to a Gaussian process. Furthermore, we show that the proposed estimators are asymptotically efficient by establishing an asymptotic linear representation with the efficient influence curve of van der Laan and McKeague (1998).

Our estimating methods lead to the use of imputation and the inverse probability weighting. Imputation has become a popular method for handling missing data. See, e.g., Rubin (1987), Rao and Shao (1992), Cheng (1994), Lipsitz, Zhao and Molenberghs (1998), Barnard and Rubin (1999), Robins and Wang (2000), Wang and Rao (2001), Wang and Rao (2002a,b), Wang, Linton and Härdle (2004), to list just a few. This popularity largely stems from the fact that, once the missing values are filled in, standard complete-data methods can be readily applied to statistical analysis. The inverse probability weighting approach is another popular method for handling missing data and has been paid considerable attention in the literature. See, e.g., Robins and Rotnitzky (1992), Zhao, Lipsitz and Lew (1996), Wang, Wang, Gutierrez and Carroll (1998) and Robins, Rotnitzky and Zhao (1994). It is noted that imputation and weighting approaches are usually applied to regression problems with missing responses or missing covariates. They do not seem natural here. By Dikta (1998), however, $S(t)$ can be represented as a functional of $m(x) = E[\delta|X = x]$, a regression function of the indicator δ on X. This, together with product-limit approach, leads to the use of imputation and weighting.

This paper is organized as follows. In Section 2, we produce four asymptotically efficient estimators through imputation and weighting methods. In Section 3, we give their asymptotic properties, including consistency, asymptotic representations, weak convergence and asymptotic efficiency. In Section 4, some simulation results are reported to evaluate the finite-sample performances of the proposed estimators, and we compare our estimators with that of van der Laan and McKeague (1998). The proofs for the main results are delayed to Appendices A and B.

2. Estimation

Let H denote the distribution function of X, and take $H_1(t) = P(X \leq t, \delta = 1)$. The cumulative hazard function $\Lambda(t)$ corresponding to F is then given by

$$\Lambda(t)=\int_0^t \frac{1}{1-F(x)}dF(x)=\int_0^t \frac{1}{1-H(x)}dH_1(x). \tag{2.1}$$

By Dikta (1998), we have

$$H_1(t)=P(\delta=1,\ X\le t)=\int_0^t m(x)dH(x),$$

where $m(x)=P(\delta=1|X=x)=E[\delta|X=x]$. This, together with (2.1), gives

$$\Lambda(t)=\int_0^t \frac{m(x)}{1-H(x)}dH(x). \tag{2.2}$$

Let $H_n(t)=n^{-1}\sum_{i=1}^n I[X_i\le t]$, $H_n(t-)=\lim_{x\uparrow t}H_n(x)$, and $H_{n1}(t)=n^{-1}\sum_{i=1}^n I[X_i\le t,\delta_i=1]$. If one can define an estimator of $m(x)$, say $m_n(x)$, from the observed data (X_i,δ_i,ξ_i), $i=1,\ldots,n$, $\Lambda(t)$ can then be estimated by

$$\Lambda_n(t)=\int_0^t \frac{m_n(x)}{1-H_n(x-)}dH_n(x)=\sum_{i:X_i\le t}\frac{m_n(X_i)}{n-R_i+1}, \tag{2.3}$$

where R_i is as defined in the introduction. Then $S(t)=\exp\{-\Lambda(t)\}$ can be estimated by $exp(-\Lambda_n(t))$. By the approximation $\exp(-x)\approx 1-x$, we have

$$\exp(-\Lambda_n(t))=\prod_{i:X_i\le t}\Big(\exp\Big\{-\frac{1}{n-R_i+1}\Big\}\Big)^{m_n(X_i)}\approx\prod_{i:X_i\le t}\Big(\frac{n-R_i}{n-R_i+1}\Big)^{m_n(X_i)}.$$

This motivates us to consider the PL estimator

$$S_n(t)=\prod_{i:X_i\le t}\Big(\frac{n-R_i}{n-R_i+1}\Big)^{m_n(X_i)}. \tag{2.4}$$

We first use the inverse probability weighting approach to estimate $m(\cdot)$. Let

$$\pi_n(x)=\frac{\sum_{i=1}^n \xi_i W\Big(\frac{x-X_i}{b_n}\Big)}{\sum_{i=1}^n W\Big(\frac{x-X_i}{b_n}\Big)},$$

where $W(\cdot)$ is a kernel function and b_n is a bandwidth sequence. Note that $\pi_n(x)$ is the well-known Nadaraya-Watson kernel regression estimator of $\pi(x)=P(\xi=$

$1|X = x)$. Then, take

$$\widehat{m}_n(x) = \frac{\sum_{i=1}^{n} \Big(\frac{\xi_i \delta_i}{\pi_n(X_i)}\Big) K\Big(\frac{x - X_i}{h_n}\Big)}{\sum_{i=1}^{n} \Big(\frac{\xi_i}{\pi_n(X_i)}\Big) K\Big(\frac{x - X_i}{h_n}\Big)},$$

where $K(\cdot)$ is a kernel function and h_n is a bandwidth sequence. See, e.g., Robins et al. (1994) and Hirano, Imbens and Ridder (2003)

The first weighted estimator, say $\widehat{S}_{n,W}(t)$, is $S_n(t)$ with $m_n(\cdot)$ taken to be $\widehat{m}_n(\cdot)$.

Note that $\widehat{S}_{n,W}(t)$ is actually the Kaplan-Meier PL estimator (1.1) with δ_i replaced by $\widehat{m}_n(X_i)$ for $i = 1, \ldots, n$. Intuitively, this estimator can be modified by defining an estimator to be the Kaplan-Meier estimator with the missing δ_i replaced by $\widehat{m}_n(X_i)$ only. This leads to an imputation estimator given by

$$\widehat{S}_{n,I}(t) = \prod_{i:X_i \le t} \Big(\frac{n - R_i}{n - R_i + 1}\Big)^{\xi_i \delta_i + (1-\xi_i)\widehat{m}_n(X_i)}. \tag{2.5}$$

This estimator can also be motivated by the fact $E[\xi\delta + (1 - \xi)m(X)] = E[\delta]$ under MAR.

If one replaces $\widehat{m}_n(\cdot)$ in $\widehat{S}_{n,I}(t)$ with

$$\widetilde{m}_n(x) = \frac{\sum_{i=1}^{n} \xi_i \delta_i K\Big(\frac{x - X_i}{h_n}\Big)}{\sum_{i=1}^{n} \xi_i K\Big(\frac{x - X_i}{h_n}\Big)},$$

another imputation estimator, say $\widetilde{S}_{N,I}(t)$, is then obtained.

Let $\pi(x) = P(\xi = 1|X = x)$. Note that under MAR we have $E[\xi\delta/\pi(X) + (1 - \xi/\pi(X))m(X)] = E[\delta]$ and $\pi_n(x)$ is the kernel regression estimator of $\pi(x)$. This motivates us to define another inverse probability weighted estimator as

$$\widetilde{S}_{n,W}(t) = \prod_{i:X_i \le t} \Big(\frac{n - R_i}{n - R_i + 1}\Big)^{\frac{\xi_i \delta_i}{\pi_n(X_i)} + (1 - \frac{\xi_i}{\pi_n(X_i)})\widetilde{m}_n(X_i)}. \tag{2.6}$$

It should be pointed out that the inverse probability weighting approach was first introduced by Robins and Rotnitzky (1992). In the four estimators proposed here, the two imputation estimators and the second inverse probability weighted estimator reduce to the usual Kaplan-Meier estimator, and $\widehat{S}_{n,W}(t)$ reduces to a smoothed Kaplan-Meier estimator when the censoring indicators are observed completely. In this case, however, the influence curve, and hence the asymptotic distribution of $\widehat{S}_{n,W}(t)$, reduces to that of the Kaplan-Meier estimator.

The proposed estimators can be extended to some more realistic situations. Often, in survival analysis and biomedical studies, covariate information is collected when some censoring indicators are missing. A possible extension is to estimate the conditional hazard and conditional survival function by incorporating covariate information with a kernel smoothing method when censoring indicators are missing at random.

3. Asymptotic Properties

Let $\widehat{S}_n(t)$ denote one of $\widehat{S}_{n,W}(t)$, $\widehat{S}_{n,I}(t)$, $\widetilde{S}_{n,I}(t)$ and $\widetilde{S}_{n,W}(t)$, and $\tau_H = \inf\{t : H(t) = 1\}$.

Theorem 3.1. *Under the assumptions given in Appendix A, we have* $\sup_{0\le t\le\tau_0} |\widehat{S}_n(t) - S(t)| \xrightarrow{a.s.} 0$ *for* $0 < \tau_0 < \tau_H$.

To derive the weak convergence result, and prove the asymptotic efficiency of $\widehat{S}_n(t)$, we first establish an asymptotic representation of $\widehat{S}_n(t)$.

Theorem 3.2. *Under the assumptions listed in Appendix B, we have*

$$\widehat{S}_n(t) - S(t) = n^{-1}\sum_{i=1}^{n} \mathrm{IC}(X_i, \delta_i, \xi_i; t) + o_p(n^{-\frac{1}{2}}) \quad \text{for} \quad t < \tau_H,$$

where

$$\begin{aligned}\mathrm{IC}(X_i, \delta_i, \xi_i; t) = -S(t)\Big[&\frac{(\xi_i - \pi(X_i))(\delta_i - m(X_i))}{\pi(X_i)(1 - H(X_i))} I[X_i \le t] \\ &+ \int_0^{t\wedge X_i} \frac{d\widetilde{H}_1(s)}{(1 - H(s))^2} + \frac{I[X_i \le t, \delta_i = 1]}{1 - H(X_i)}\Big]\end{aligned}$$

with $\widetilde{H}_1(t) = P(X > t, \delta = 1)$.

The pointwise iid representation is used only for proving weak convergence. A separate work establishes that the remainder term in the iid representation is uniformly negligible. van der Laan and McKeague (1998) also established a pointwise iid representation with an efficient influence curve $IC_t^*(X, \delta, \xi)$ for the estimator (16) of their paper. It can be shown that the asymptotic representation is the same as that in Theorem 3.2 but it should be pointed out that their $IC_t^*(X, \delta, \xi)$ contains two typos.

The influence curve of the Kaplan-Meier PL estimator, say $\mathrm{IC}_{KM}(X,\delta)$, and due to Lo and Singh (1985), is

$$\mathrm{IC}_{KM}(X,\delta)=S(t)\Big(\int_0^{t\wedge X}\frac{dH_1(x)}{(1-H(X))^2}-\frac{I[X\le t,\delta=1]}{1-H(X)}\Big).$$

The correct efficient influence curve $\mathrm{IC}_t^*(X,\delta,\xi)$ in (16) of van der Laan and McKeague (1998) is

$$\begin{aligned}IC_t^*(X,\delta,\xi)=S(t)\Big[&-\Big(I(X\le t)-\frac{I(X\le t,\xi=1)}{\pi(X)}\Big)\frac{k(X)}{1-H(X)}\\&-\frac{I[X\le t,\delta=1,\xi=1]}{(1-H(X))\pi(X)}\\&+\int_0^t\frac{1}{(1-H(X))^2}\frac{I[X>x,\xi=1]}{\pi(X)}dH_1(x)\\&+\Big(1-\frac{\xi}{\pi(X)}\Big)\int_0^{t\wedge X}\frac{dH_1(x)}{(1-H(x))^2}\Big],\end{aligned}$$

if F is continuous, where $k(x)=dH_1(x)/dH(x)$. In the proof of the following Theorem 3.3, it is shown that

$$IC(X,\delta,\xi;t)=IC_t^*(X,\delta,\xi), \tag{3.1}$$

and shows that our estimators are asymptotically efficient.

Theorem 3.3. *Under assumptions of Theorem* 3.2*, we have*

$$\sqrt{n}(\widehat{S}_n(t)-S(t))\xrightarrow{\mathcal{D}}W(t),t\in[0,\tau_H)$$

and $\widehat{S}_n(t)$ is asymptotically efficient, where $W(t)$ is a Gaussian process with

$$\begin{aligned}&EW(t)=0,\\&\mathrm{Cov}\,(W(t_1),W(t_2))=S(t_1)S(t_2)\Big[-\int_0^{t_1\wedge t_2}\frac{d\widetilde{H}_1(s)}{(1-H(s))^2}\\&\qquad+\int_0^{t_1\wedge t_2}\frac{m(s)(1-m(s))}{(1-H(s))^2}\Big(\frac{1}{\pi(s)}-1\Big)dH(s)\Big].\end{aligned}$$

Clearly, the weak convergence result reduces to that of the Kaplan-Meier PL estimator, derived by Gill (1983) when $\pi(x)=1$. Theorem 3.3 implies

$$\sqrt{n}(\widehat{S}_n(t)-S(t))\xrightarrow{\mathcal{L}}N(0,\sigma^2(t)),$$

where

$$\sigma^2(t)=S^2(t)\Big[-\int_0^t\frac{d\widetilde{H}_1(s)}{(1-H(s))^2}+\int_0^t\frac{m(s)(1-m(s))}{(1-H(s))^2}\Big(\frac{1}{\pi(s)}-1\Big)dH(s)\Big].$$

This shows that all the proposed estimators have the same asymptotic distribution. Clearly, the asymptotic variance reduces to that of Kaplan-Meier PL estimator when $\pi(x) = 1$. A direct method to estimate the asymptotic variance is to use the "Plug in" technique by obtaining the estimators of $H(\cdot)$, $H_1(\cdot)$, $\pi(\cdot)$ and $m(\cdot)$. An alternative is to use the jackknife method to estimate the asymptotic variance.

All the suggested estimators $\widehat{S}_n(t)$ are global functionals of $\widehat{m}_n(\cdot)$ or $\widetilde{m}_n(\cdot)$, and hence $n^{1/2}$-rate asymptotic normality of the estimators indicates that a proper choice of both h_n and b_n depends only on the second order terms of the mean square error of the estimators. This implies that the selection of the bandwidths may not be critical for estimating $S(t)$. This is also suggested by our simulation results.

4. Simulation Results

4.1. A simulation comparison under MAR

We carried out a simulation study to evaluate the finite-sample properties of the proposed estimators and to compare the finite sample performances of the proposed estimators with the Kaplan-Meier estimator and with the van der Laan and McKeague (1998) estimator under MAR, in terms of the close fit of the curves of these estimators and their mean integrated squared errors. The Kaplan-Meier estimator can serve as a gold standard, even though it is practically unachievable because of the missingness of censoring indicators.

In the simulation, the life variable T and the censoring variable C were generated from exponential distributions $E(1)$ and $E(1/4)$ for a 20% censoring rate, from $E(1)$ and $E(2/3)$ for a 40% censoring rate, and from $E(1)$ and $E(7/3)$ for a 70% censoring rate, respectively. The sample size was taken to be $n = 30$, 60, 100 and 200. The missing mechanism followed $\text{logit}(\pi(x)) = \theta_1 + \theta_2 x$ with different $\theta = (\theta_1, \theta_2)$. For the censoring rate of 0.2, θ was $(1.25, 0.13)$ and $(0.5, -0.10)$, so that the average missing rates were about 0.2 and 0.4, respectively; for the censoring rate of 0.4, θ was taken to be $(1.25, 0.15)$ and $(0.70, -0.28)$, so that the average missing rates were about 0.2 and 0.4, respectively; for the censoring rate of 0.7, θ was taken to be $(1.40, -0.12)$ and $(0.45, -0.18)$ so that the average missing rates were about 0.2 and 0.4, respectively. To calculate the proposed estimators, the kernel functions $W(\cdot)$ and $K(\cdot)$ were taken to be $W(u) = 1/2$ for $|u| \le 1$, 0 otherwise, and $K(u) = (15/16)(1 - 2u^2 + u^4)$ for $|u| \le 1$, 0 otherwise. The bandwidths (h_n, b_n) were taken to be $(n^{-1/3}, n^{-1/3})$. To calculate $\widehat{S}_{VM}(t)$, we used the partition which consists of k points on a regular grid with $k = 50$.

We generated 5,000 Monte Carlo random samples of size $n = 30$, 60, 100 and 200 under every different combination of censoring rates and missing rates.

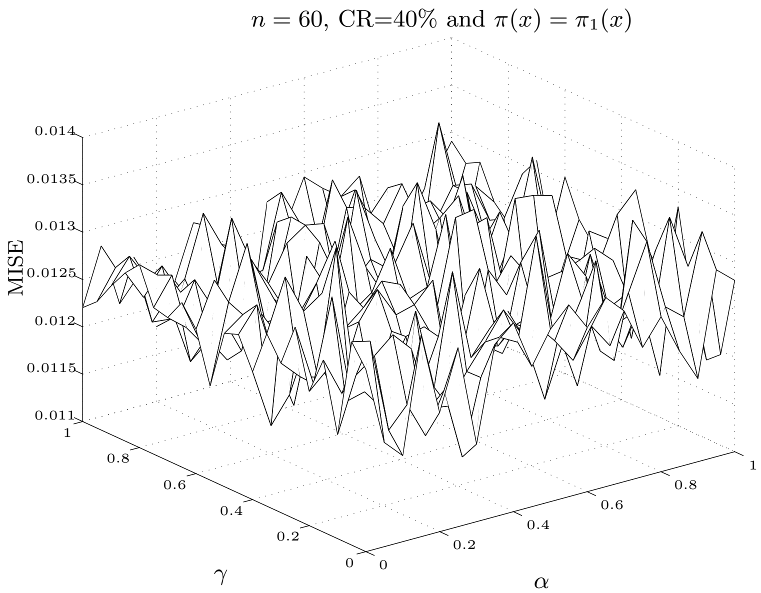

Figure 4.1. MISE Curves of $\widetilde{S}_{n,W}(t)$ on (α, γ), when $b_n = n^{-\alpha}$ and $h_n = n^{-\gamma}$.

From 5,000 simulated values of these estimators, we calculated the MISE. Also, to evaluate the effect of the bandwidths on the MISE, as an example, we plotted MISE curve of $\widetilde{S}_{n,W}(t)$ on the bandwidth (h_n, b_n). The MISE was calculated over the interval $[0, 2]$. The results are reported in Figure 4.1 and Table 4.1.

From Figure 4.1, the MISEs for different combinations of b_n and h_n are generally between 0.0115 and 0.0140. This suggests that the effect of the bandwidths h_n and b_n on the MISE is not critical, and hence the selection of the bandwidths may not be critical for the estimation of $S(t)$, as pointed out in the last paragraph of Section 3.

From Table 4.1, the proposed estimators all have similar MISE and close to that of the Kaplan-Meier estimator. This suggests that all the estimators perform well and have similar finite-sample performances. $\widehat{S}_{VM}(t)$ has about one or two times larger MISE than the proposed estimators and the Kaplan-Meier estimator when $n = 30$ or 60 and the censoring rate is 0.2 or 0.4. The MISE of $\widehat{S}_{VM}(t)$ is still larger than that of the proposed estimators when the censoring rate is 70%. This shows that $\widehat{S}_{VM}(t)$ does not work well for small or medium sample size, and that the proposed estimators outperform $\widehat{S}_{VM}(t)$ in terms of MISE. However, the difference between the proposed estimators and $\widehat{S}_{VM}(t)$ decreases in terms of MISE when sample size increases. $\widehat{S}_{VM}(t)$ has larger or slightly larger MISE than the proposed estimators when $n = 100$ or 200. As pointed out in the introduction, $\widehat{S}_{VM}(t)$ can be unappealing in practice, especially for small samples.

Table 4.1. Mean integrated square error under MAR.

n	Estimators	20% censoring		40% censoring		70% censoring	
		$\pi_1(x)$	$\pi_2(x)$	$\pi_1(x)$	$\pi_2(x)$	$\pi_1(x)$	$\pi_2(x)$
	$\widehat{S}_{KM}$	0.0148	0.0149	0.0205	0.0204	0.0753	0.0730
	$\widehat{S}_{n,W}$	0.0156	0.0174	0.0225	0.0274	0.0826	0.0975
	$\widehat{S}_{n,I}$	0.0159	0.0176	0.0225	0.0278	0.0848	0.1011
30	$\widetilde{S}_{n,I}$	0.0156	0.0173	0.0224	0.0273	0.0842	0.1006
	$\widetilde{S}_{n,W}$	0.0160	0.0179	0.0230	0.0287	0.0883	0.1110
	$\widehat{S}_{VM}$	0.0408	0.0429	0.0458	0.0463	0.1047	0.1212
	$\widehat{S}_{KM}$	0.0074	0.0075	0.0101	0.0099	0.0391	0.0409
	$\widehat{S}_{n,W}$	0.0078	0.0084	0.0109	0.0129	0.0465	0.0565
	$\widehat{S}_{n,I}$	0.0079	0.0085	0.0113	0.0131	0.0465	0.0584
60	$\widetilde{S}_{n,I}$	0.0078	0.0083	0.0109	0.0128	0.0451	0.0573
	$\widetilde{S}_{n,W}$	0.0080	0.0086	0.0115	0.0135	0.0474	0.0625
	$\widehat{S}_{VM}$	0.0132	0.0134	0.0194	0.0195	0.0559	0.0727
	$\widehat{S}_{KM}$	0.0044	0.0042	0.0059	0.0060	0.0255	0.0254
	$\widehat{S}_{n,W}$	0.0046	0.0057	0.0064	0.0071	0.0312	0.0375
	$\widehat{S}_{n,I}$	0.0046	0.0057	0.0066	0.0073	0.0316	0.0383
100	$\widetilde{S}_{n,I}$	0.0046	0.0057	0.0064	0.0071	0.0301	0.0376
	$\widetilde{S}_{n,W}$	0.0047	0.0058	0.0067	0.0075	0.0314	0.0409
	$\widehat{S}_{VM}$	0.0056	0.0059	0.0101	0.0103	0.0370	0.0447
	$\widehat{S}_{KM}$	0.0023	0.0021	0.0029	0.0029	0.0127	0.0136
	$\widehat{S}_{n,W}$	0.0025	0.0026	0.0031	0.0034	0.0163	0.0217
	$\widehat{S}_{n,I}$	0.0025	0.0026	0.0032	0.0035	0.0163	0.0220
200	$\widetilde{S}_{n,I}$	0.0025	0.0026	0.0031	0.0034	0.0149	0.0213
	$\widetilde{S}_{n,W}$	0.0025	0.0026	0.0032	0.0036	0.0156	0.0227
	$\widehat{S}_{VM}$	0.0025	0.0029	0.0037	0.0042	0.0162	0.0236

Note: $\widehat{S}_{VM}$ and $\widehat{S}_{KM}$ refer to the van der Laan and Mckeague (1998) estimator and the Kaplan-Meier PL estimator respentisy, $\pi_1(x)$ and $\pi_2(x)$ are the missing probability functions $\pi(x)$ with average missing rates about 0.2 and 0.4, respectively.

In Figures 4.2 and 4.3, we plotted the curve of the true survival distribution function and the curves of $\widehat{S}_{n,W}(t)$, $\widehat{S}_{n,I}(t)$, $\widetilde{S}_{n,I}(t)$, $\widetilde{S}_{n,W}(t)$, $\widehat{S}_{VM}(t)$ and $\widehat{S}_{KM}(t)$, where $\widehat{S}_{VM}(t)$ refers to the van der Laan and McKeague (1998) estimator and $\widehat{S}_{KM}(t)$ to the Kaplan-Meier estimator given in (1.1). Each estimated curve in Figure 4.2 is based on one samples. Each curve in Figure 4.3 is based on 5,000 samples. That is, each estimated curve in Figure 4.3 is the average of 5,000

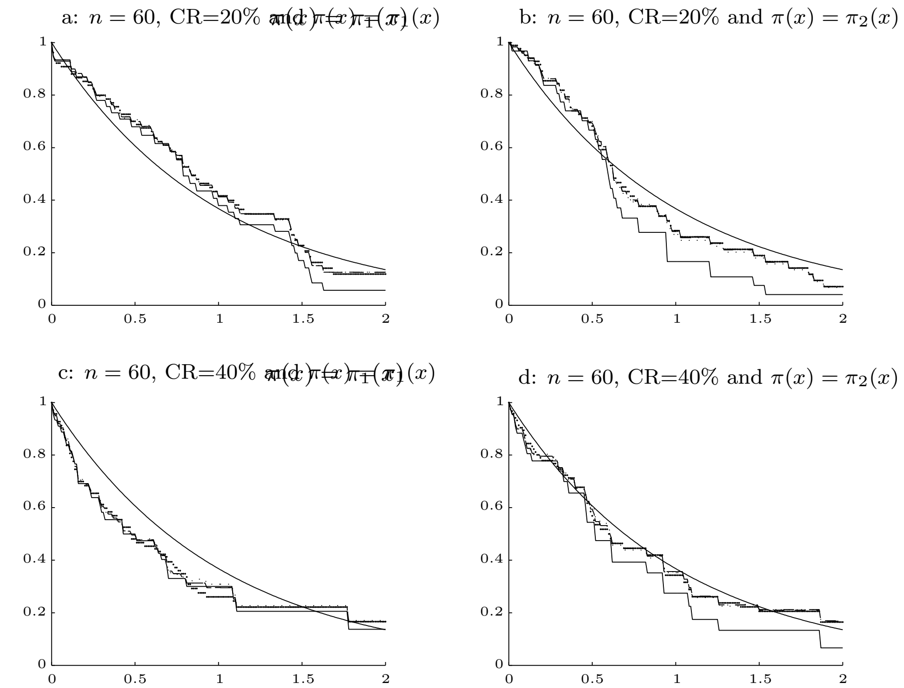

Figure 4.2. Curves for $\widehat{S}_{n,W}(t)$, $\widehat{S}_{n,I}(t)$, $\widetilde{S}_{n,I}(t)$, $\widetilde{S}_{n,W}(t)$, $\widehat{S}_{VM}(t)$, and the true survival function of $S(t)$, where $\widehat{S}_{VM}(t)$ refers to the van der Laan and Mckeague (1998) estimator. The dotted curve is $\widehat{S}_{n,W}(t)$; the dashed curve is $\widehat{S}_{n,I}(t)$; the dash-dotted curve is $\widetilde{S}_{n,I}(t)$; the solid stair curve is $\widetilde{S}_{n,W}(t)$; the step solid curve is $\widehat{S}_{VM}$ and the smooth solid curve is $S(t)$. Each estimated curve is based on one sample. CR refers to censoring rate.

survival function estimators.

From Figures 4.2 and 4.3, the curves of all the estimators including $\widehat{S}_{VM}(t)$ are close to the true survival curve, and the curves of the proposed estimators overlap each other. This suggests that the estimators perform well, and that the proposed estimators perform similarly in terms of bias.

4.2. A simulation comparison under MCAR

We carried out a simulation study to compare the finite sample performances of the proposed estimators with that of the Kaplan-Meier PL estimator (1.1) and the van der Laan and Mckeague (1998) estimator, $\widehat{S}_{VM}(t)$, under MCAR.

In the simulation, the life variable T and censoring variable C were generated from exponential distributions $E(1)$ and $E(1/4)$ for a 20% censoring rate, and from $E(1)$ and $E(2/3)$ for a 40% censoring rate, respectively. The sample size was taken to be $n = 30$, 60, 100 and 200. $\pi(x)$, the non-missing rate, was taken to be 0.4, 0.6 or 0.8.

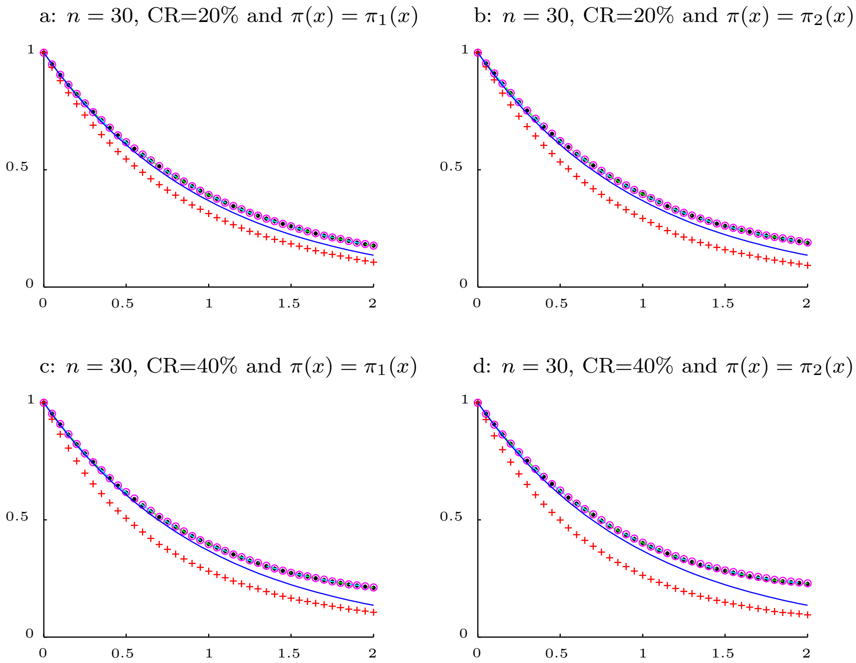

Figure 4.3. Curves for $\widehat{S}_{n,W}(t)$, $\widehat{S}_{n,I}(t)$, $\widetilde{S}_{n,I}(t)$, $\widetilde{S}_{n,W}(t)$, $\widehat{S}_{VM}$ and the true survival function of $S(t)$, where $\widehat{S}_{VM}$ refers to the van der Laan and Mckeague (1998) estimator. The star curve is $\widehat{S}_{n,W}(t)$; the dash-dotted curve is $\widehat{S}_{n,I}(t)$; the dashed curve is $\widetilde{S}_{n,I}(t)$; the circle curve is $\widetilde{S}_{n,W}(t)$; the plus curve is $\widehat{S}_{VM}(t)$ and solid curve is $S(t)$. Each estimated curve is based on 5,000 samples. CR refers to censoring rate.

To calculate the proposed estimators, the kernel functions $W(\cdot)$ and $K(\cdot)$ were taken to be $W(u) = 1/2$ for $|u| \leq 1$, 0 otherwise, and $K(u) = (15/16)(1 - 2u^2 + u^4)$ for $|u| \leq 1$, 0 otherwise. The bandwidths (h_n, b_n) were taken to be $(n^{-1/3}, n^{-1/3})$. To calculate $\widehat{S}_{VM}(t)$, the partition consists of k points on a regular grid with $k = 50$.

We generated 5,000 Monte Carlo random samples of size $n = 30$, 60, 100 and 200 under every different combination of censoring rate and missing rate. From 5,000 simulated values of the proposed estimators, the Laan-Mckeague estimator, and the Kaplan-Meier PL estimator, we calculated the MISE over the interval $[0, 2]$. The results are reported in Table 4.2.

From Table 4.2, the MISEs of the proposed estimators are similar and slightly larger than that of the Kaplan-Meier PL estimator when the non-missing rate is

Table 4.2. Mean integrated square error under MCAR when sample size n = 30, 60, 100 and 200.

	20% censoring			40% censoring		
Estimators	$\pi = 0.4$	$\pi = 0.6$	$\pi = 0.8$	$\pi = 0.4$	$\pi = 0.6$	$\pi = 0.8$
			$n = 30$			
$\widehat{S}_{KM}$	0.0144	0.0145	0.0148	0.0211	0.0205	0.0201
$\widehat{S}_{n,W}$	0.0201	0.0173	0.0156	0.0400	0.0278	0.0225
$\widehat{S}_{n,I}$	0.0202	0.0176	0.0159	0.0402	0.0279	0.0229
$\widetilde{S}_{n,I}$	0.0199	0.0172	0.0156	0.0397	0.0276	0.0224
$\widetilde{S}_{n,W}$	0.0208	0.0179	0.0161	0.0423	0.0290	0.0234
$\widehat{S}_{VM}$	0.0418	0.0398	0.0357	0.0460	0.0453	0.0451
			$n = 60$			
$\widehat{S}_{KM}$	0.0072	0.0073	0.0073	0.0100	0.0099	0.0100
$\widehat{S}_{n,W}$	0.0103	0.0084	0.0077	0.0184	0.0130	0.0111
$\widehat{S}_{n,I}$	0.0104	0.0085	0.0078	0.0186	0.0132	0.0114
$\widetilde{S}_{n,I}$	0.0101	0.0084	0.0077	0.0182	0.0129	0.0110
$\widetilde{S}_{n,W}$	0.0108	0.0087	0.0079	0.0194	0.0137	0.0116
$\widehat{S}_{VM}$	0.0164	0.0133	0.0134	0.0220	0.0198	0.0196
			$n = 100$			
$\widehat{S}_{KM}$	0.0045	0.0044	0.0044	0.0059	0.0060	0.0061
$\widehat{S}_{n,W}$	0.0066	0.0056	0.0045	0.0109	0.0080	0.0066
$\widehat{S}_{n,I}$	0.0066	0.0057	0.0046	0.0110	0.0082	0.0068
$\widetilde{S}_{n,I}$	0.0065	0.0056	0.0045	0.0107	0.0080	0.0066
$\widetilde{S}_{n,W}$	0.0065	0.0056	0.0046	0.0105	0.0084	0.0069
$\widehat{S}_{VM}$	0.0095	0.0061	0.0056	0.0152	0.0116	0.0102
			$n = 200$			
$\widehat{S}_{KM}$	0.0022	0.0022	0.0022	0.0029	0.0029	0.0029
$\widehat{S}_{n,W}$	0.0027	0.0024	0.0023	0.0050	0.0035	0.0030
$\widehat{S}_{n,I}$	0.0027	0.0024	0.0024	0.0050	0.0036	0.0031
$\widetilde{S}_{n,I}$	0.0026	0.0024	0.0023	0.0049	0.0035	0.0030
$\widetilde{S}_{n,W}$	0.0028	0.0025	0.0024	0.0052	0.0037	0.0031
$\widehat{S}_{VM}$	0.0036	0.0026	0.0024	0.0063	0.0043	0.0036

Note: $\widehat{S}_{VM}$ and $\widehat{S}_{KM}$ refer to the estimator of van der Laan and McKeague (1998) and to the Kaplan-Meier PL estimator, respectively.

0.6 or 0.8. This suggests that these estimators perform similarly and work well when non-missing rate is not too low. However, it was observed that $\widehat{S}_{VM}(t)$ had far larger (n = 30, 60), larger (n = 100), or slightly larger (n = 200) MISE than the proposed estimators. This suggests that our estimators outperform $\widehat{S}_{VM}(t)$ in terms of MISE.

Appendix A: Outline of Proof for Strongly Uniform Consistency

Here, we give only an outline of the proof of Theorem 2.1. The detailed proof is given in the on-line supplement (http://www.stat.sinica.edu.tw/statistica/submission). We first list some conditions needed for strongly uniform consistency.

(A.m): $m(x)$ is a continuous function.

(A.H): $H(\cdot)$ has uniformly continuous probability density function $h(\cdot)$.

(A.W): $W(\cdot)$ is a kernel function with bounded variation and bounded support satisfying $\int W(u)du = 1$ and $\int |W(u)|du < \infty$.

(A.K): $K(\cdot)$ is a bounded kernel function with bounded support satisfying $\int K(u)du = 1$.

(A.b_n): $b_n \to 0$, $(nb_n)^{-1}\log n \to 0$.

(A.h_n): $h_n \to 0$, $(nh_n)^{-1}\log n \to 0$.

Let

$$\widehat{\Lambda}_{n,W}(t) = \sum_{i:X_i\le t} \frac{\widehat{m}_n(X_i)}{n-R_i+1},$$

$$\widehat{\Lambda}_{n,I}(t) = \sum_{i:X_i\le t} \frac{\xi_i\delta_i + (1-\xi_i)\widehat{m}_n(X_i)}{n-R_i+1},$$

$$\widetilde{\Lambda}_{n,I}(t) = \sum_{i:X_i\le t} \frac{\xi_i\delta_i + (1-\xi_i)\widetilde{m}_n(X_i)}{n-R_i+1},$$

$$\widetilde{\Lambda}_{n,W}(t) = \sum_{i:X_i\le t} \frac{\frac{\xi_i\delta_i}{\pi_n(X_i)} + \Big(\frac{1-\xi_i}{\pi_n(X_i)}\Big)\widetilde{m}_n(X_i)}{n-R_i+1}.$$

Let $\widehat{\Lambda}_n(t)$ denote one of $\widehat{\Lambda}_{n,W}(t)$, $\widehat{\Lambda}_{n,I}(t)$, $\widetilde{\Lambda}_{n,I}(t)$ and $\widetilde{\Lambda}_{n,W}(t)$. To prove Theorem 3.1, we first prove the following lemma.

Lemma A.1. *Under assumptions of Theorem* 3.1*, we have* $\sup_{0\le t\le\tau_0}|\widehat{\Lambda}_n(t) - \Lambda(t)| \xrightarrow{a.s.} 0$*, where* $0<\tau_0<\tau_H$ *and* τ_H *is as defined in Section* 3.

Proof. We only prove that Lemma A.1 is true for $\widetilde{\Lambda}_{n,W}(t)$. The other three cases can be proved similarly.

For $\widetilde{\Lambda}_{n,W}(t)$, we have

$$\widetilde{\Lambda}_{n,W}(t) - \Lambda(t) = \Bigg(\sum_{i:X_i\le t} \frac{\frac{\xi_i\delta_i}{\pi_n(X_i)} + \Big(\frac{1-\xi_i}{\pi_n(X_i)}\Big)\widetilde{m}_n(X_i)}{n-R_i+1} - \sum_{i:X_i\le t} \frac{\frac{\xi_i\delta_i}{\pi_n(X_i)} + \Big(\frac{1-\xi_i}{\pi_n(X_i)}\Big)m(X_i)}{n-R_i+1}\Bigg)$$

$$
\begin{aligned}
&+\Big(\sum_{i:X_i\le t}\frac{\frac{\xi_i\delta_i}{\pi_n(X_i)}+\Big(\frac{1-\xi_i}{\pi_n(X_i)}\Big)m(X_i)}{n-R_i+1}\\
&-\sum_{i:X_i\le t}\frac{\frac{\xi_i\delta_i}{\pi(X_i)}+\Big(\frac{1-\xi_i}{\pi(X_i)}\Big)m(X_i)}{n-R_i+1}\Big)\\
&+\Big(\sum_{i:X_i\le t}\frac{\frac{\xi_i\delta_i}{\pi(X_i)}+\Big(\frac{1-\xi_i}{\pi(X_i)}\Big)m(X_i)}{n-R_i+1}-\Lambda(t)\Big)\\
&:=\zeta_{n1}(t)+\zeta_{n2}(t)+\zeta_{n3}(t). \qquad \text{(A.1)}
\end{aligned}
$$

It can be proved $\sup_{0\le t\le\tau_0}|\zeta_{n1}(t)| \xrightarrow{a.s.} 0$, $\sup_{0\le t\le\tau_0}|\zeta_{n2}(t)| \xrightarrow{a.s.} 0$ and $\sup_{0\le t\le\tau_0}|\zeta_{n3}(t)| \xrightarrow{a.s.} 0$. This proves that Lemma A.1 is true for $\widetilde{\Lambda}_{n,W}(t)$.

Proof of Theorem 3.1. By Taylor expansion, it is easy to obtain

$$
\widehat{S}_n(t)-S(t)=(-\widehat{\Lambda}_n(t)+\Lambda(t))\exp\{-\Lambda(t)\}+R_n(t), \qquad \text{(A.2)}
$$

where

$$
R_n(t)=(\log\widehat{S}_n(t)+\widehat{\Lambda}_n(t))\exp\{-\Lambda(t)\}+\frac{\exp\{c_n(t)\}}{2}(\log\widehat{S}_n(t)+\Lambda(t))^2,
$$

$$
\min\{\log\widehat{S}_n(t),-\Lambda(t)\}\le c_n(t)\le\max\{\log\widehat{S}_n(t),-\Lambda(t)\}.
$$

It can be proved that

$$
\sup_{0\le t\le\tau_0}|\log\widehat{S}_n(t)+\widehat{\Lambda}_n(t)|\xrightarrow{a.s.}0 \qquad \text{(A.3)}
$$

for any τ_0 such that $0<\tau_0<\tau_H$.

By Lemma A.1, (A.2) and (A.3), Theorem 3.1 is then proved.

Appendix B: Outline of proofs for asymptotic representation, weak convergence and asymptotic efficiency

Here, we give only an outline of the proofs of Theorem 3.2 and 3.3. The detailed proofs are given in the on-line supplement (http://www.stat.sinica.edu.tw/statistica/submission). The following are the conditions needed for asymptotic representation, weak convergence and asymptotic efficiency.

(C.m). $m(\cdot)$ has continuous derivatives up to order $k>1$.

(C.K). $K(\cdot)$ is a kernel function of order k with bounded support.

(C.W). i. $W(\cdot)$ is a probability density kernel function with bounded support.

ii. $\int W^2(s)ds<\infty$.

(C.H). $H(\cdot)$ has probability density $h(\cdot)$ and $h(\cdot)$ has derivatives up to order of $k > 1$.

(C.π). i. $\pi(\cdot)$ has derivatives up to order of $k > 1$.
ii. $\inf_t \pi(t) > 0$.

(C.h_n). $nh_n \to \infty$ and $nh_n^{2k} \to 0$ for k in (C.H).

(C.b_n). $b_n \to 0$ and $nb_n \to \infty$.

Lemma A.2. *Under the assumptions of Theorem* 3.2, *we have*

$$\widehat{\Lambda}_n(t) - \Lambda(t) = -\frac{1}{n}\sum_{i=1}^{n} \frac{\mathrm{IC}(X_i, \delta_i, \xi_i)}{S(t)} + o_p(n^{-\frac{1}{2}}),$$

where IC(X, δ, ξ) is as defined in Theorem 3.2.

Proof. (a) We first prove Lemma A.2 is true for $\widehat{\Lambda}_{n,W}(t)$. We have

$$\begin{aligned}\widehat{\Lambda}_{n,W}(t) - \Lambda(t) &= \Big(\sum_{i:X_i \le t} \frac{\widehat{m}_n(X_i)}{n - R_i + 1} - \sum_{i:X_i \le t} \frac{m(X_i)}{n - R_i + 1}\Big) \\ &\quad + \Big(\sum_{i:X_i \le t} \frac{m(X_i)}{n - R_i + 1} - \Lambda(t)\Big). \\ &:= I_{n1}(t) + I_{n2}(t). \end{aligned} \tag{A.4}$$

$I_{n1}(t)$ can be represented as

$$I_{n1}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{\xi_i(\delta_i - m(X_i))}{\pi(X_i)(1 - H(X_i))} I[X_i \le t] + o_p(n^{-\frac{1}{2}})). \tag{A.5}$$

By Dikta (1998), for $I_{n2}(t)$ in (A.4), we have

$$\begin{aligned} I_{n2}(t) &= \frac{1}{n}\sum_{i=1}^{n}\int_0^{t \wedge X_i} \frac{d\widetilde{H}_1(s)}{(1 - H(s))^2} + \frac{1}{n}\sum_{i=1}^{n} \frac{I[X_i \le t, \delta_i = 1]}{1 - H(X_i)} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \frac{m(X_i) - \delta_i}{1 - H(X_i)} I[X_i \le t] + o_p(n^{-\frac{1}{2}}). \end{aligned} \tag{A.6}$$

Lemma A.2 then holds for $\widehat{\Lambda}_{n,W}(t)$ from (A.4), (A.5) and (A.6).

(b) Secondly, we prove Lemma A.2 is true for $\widehat{\Lambda}_{n,I}(t)$. We have

$$\begin{aligned}\widehat{\Lambda}_{n,I}(t) - \Lambda(t) &= \Bigg(\sum_{i:X_i \le t} \frac{\xi_i\delta_i + (1 - \xi_i)m(X_i)}{n - R_i + 1} - \Lambda(t)\Bigg) \\ &\quad + \Bigg(\sum_{i:X_i \le t} \frac{\xi_i\delta_i + (1 - \xi_i)\widehat{m}_n(X_i)}{n - R_i + 1} - \sum_{i:X_i \le t} \frac{\xi_i\delta_i + (1 - \xi_i)m(X_i)}{n - R_i + 1}\Bigg)\end{aligned}$$

$$:= T_{n1}(t) + T_{n2}(t). \tag{A.7}$$

For $T_{n1}(t)$, we have

$$\begin{aligned} T_{n1}(t) &= \frac{1}{n}\sum_{i=1}^{n} \frac{(1-\xi_i)(m(X_i)-\delta_i)}{1-H(X_i)} I[X_i \le t] \\ &+ \frac{1}{n}\sum_{i=1}^{n}\left(\int_0^{t\wedge X_i} \frac{d\widetilde{H}_1(s)}{(1-H(s))^2} + \frac{I[X_i \le t, \delta_i = 1]}{1-H(X_i)}\right) + o_p(n^{-\frac{1}{2}}). \end{aligned} \tag{A.8}$$

It can be proved that

$$T_{n2}(t) = \frac{1}{n}\sum_{j=1}^{n} \frac{(1-\pi(X_j))\xi_j(\delta_j - m(X_j))}{\pi(X_j)(1-H(X_j))} I[X_j \le t] + o_p(n^{-\frac{1}{2}}). \tag{A.9}$$

From (A.7), (A.8) and (A.9), Lemma A.2 is proved for $\widehat{\Lambda}_{n,I}(t)$.

(c) Similar to (b), Lemma A.2 holds for $\widetilde{\Lambda}_{n,I}(t)$.

(d) Finally, we prove that Lemma A.2 holds for $\widetilde{\Lambda}_{n,W}(t)$.

For $\zeta_{n1}(t)$ in (A.1), we have

$$\begin{aligned} \zeta_{n1}(t) &= \frac{1}{n}\sum_{i=1}^{n} \frac{\left(1-\frac{\xi_i}{\pi(X_i)}\right) I[X_i \le t]\left[(nh_n)^{-1}\sum_{j=1}^{n}\xi_j(\delta_j - m(X_j))K\left(\frac{X_i - X_j}{h_n}\right)\right]}{(1-H(X_i))h(X_i)\pi(X_i)} \\ &+ \frac{1}{n}\sum_{i=1}^{n} \frac{\left(1-\frac{\xi_i}{\pi(X_i)}\right) I[X_i \le t]\left[(nh_n)^{-1}\sum_{j=1}^{n}\xi_j(m(X_j) - m(X_i))K\left(\frac{X_i - X_j}{h_n}\right)\right]}{(1-H(X_i))h(X_i)\pi(X_i)} \\ &+ o_p(n^{-\frac{1}{2}}) := \zeta_{n1,1}(t) + \zeta_{n1,2}(t) + o_p(n^{-\frac{1}{2}}). \end{aligned} \tag{A.10}$$

It can be proved that $E(\zeta_{n1,1}^2(t)) = O((n^2 h_n)^{-1})$. This proves $\zeta_{n1,1}(t) = o_p(n^{-1/2})$ as $nh_n \to \infty$. Similarly, it can be proved $E\zeta_{n1,2}^2(t) = O(h_n/n)$, which implies $\zeta_{n1,2}(t) = o_p(n^{-1/2})$ by the condition that $m(\cdot)$ has bounded derivative of order 1 and $K(\cdot)$ is a kernel function with bounded support. By (A.10), we then have $\zeta_{n1}(t) = o_p(n^{-1/2})$. Similarly, we have $\zeta_{n2}(t) = o_p(n^{-1/2})$. For $\zeta_{n3}(t)$ defined in (A.1), we have

$$\begin{aligned} \zeta_{n3}(t) &= \frac{1}{n}\sum_{i=1}^{n}\left(\int_0^{t\wedge X_i} \frac{d\widetilde{H}_1(s)}{(1-H(s))^2} + \frac{I[X_i \le t, \delta_i = 1]}{1-H(X_i)}\right) \\ &+ \frac{1}{n}\sum_{i=1}^{n} \frac{(\xi_i - \pi(X_i))(\delta_i - m(X_i))}{\pi(X_i)(1-H(X_i))} I[X_i \le t] + o_p(n^{-\frac{1}{2}}). \end{aligned} \tag{A.11}$$

This, together with (A.1), gives the desired result.

Proof of Theorem 3.2. It can be proved that

$$\log \widehat{S}_n(t) + \widehat{\Lambda}_n(t) = o_p(n^{-\frac{1}{2}}). \tag{A.12}$$

This together with (A.2) and Lemma A.2 proves Theorem 5.2.

Proof of Theorem 3.3. van der Laan and McKeague (1998) used the iid representation of their estimator to prove asymptotic efficiency and weak convergence. To prove Theorem 3.3, we need only prove that our estimators have the same asymptotic representation as that of van der Laan and McKeague (1998). This is done by showing that the influence curves $I(X, \delta, \xi; t)$ and $IC_t^*(X, \delta, \xi)$ are equal.

By the fact $k(x) = m(x)$, easily seen from $H_1(t) = \int_0^t m(x)dH(x)$, we have

$$\begin{aligned}
IC_t^*(X, \delta, \xi) &= S(t)\Big[-\frac{I[X \le t](\pi(X) - \xi)m(X)}{\pi(X)S(X)} \\
&\quad -\frac{I[X \le t](\pi(X) - \xi)\delta}{\pi(X)S(X)} + \int_0^{t\wedge} \frac{dH_1(x)}{(1 - H(x))^2}\Big] \\
&= -\frac{(\xi - \pi(X))(\delta - m(X))}{\pi(X)(1 - H(X))} I[X \le t] \\
&\quad - \int_0^{t\wedge X} \frac{d\widetilde{H}_1(s)}{(1 - H(s))^2} - \frac{I[X \le t, \delta = 1]}{1 - H(X)} \\
&= IC(X, \delta, \xi; t).
\end{aligned}$$

Acknowledgement

This research was supported by National Science Fund for Distinguished Young Scholars in China (10725106), National Natural Science Foundation of China (10671198), National Science Fund for Creative Research Groups in China and a grant from the Research Grants Council of the Hong Kong, China (HKU 7050/06P). The authors would like to thank I. W. McKeague for his assistance and for useful discussion.

References

Barnard, J. and Rubin, D. B. (1999). Small-sample degrees of freedom with multiple imputation. *Biometrika* **86**, 948-955.

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* **89**, 81-87.

Dewanji, A. (1992). A note on a test for competing risks with missing failure type. *Biometrika* **79**, 855-857.

Dikta, D. (1998). On semiparametric random censorship models. *J. Statist. Plann. Inference* **66**, 253-279.

Dinse, G. E. (1982). Nonparametric estimation for partially-complete time and type of failure data. *Biometrics* **38**, 417-431.

Gijbels, I., Lin, D. Y. and Ying, Z. (1993). Non- and semi-parametric analysis of failure time data with missing failure indicators. Preprint.

Gill, R. D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11**, 49-58.

Goetghebeur, E. J. and Ryan, L. (1990). A modified logrank test for competing risks with missing failure type. *Biometrika* **77**, 207-211.

Goetghebeur, E.J. and Ryan, L. (1995). Analysis of competing risks survival data when some failure types are missing. *Biometrika* **82**, 821-833.

Hirano, K., Imbens, G. and Ridder, G. (2003), Efficient estimation of average treatment effects using the propensity score. *Econometrica* **71**, 1161-1189.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

Lipsitz, S. R., Zhao, L. P. and Molenberghs, G. (1998). A semiparametric method of multiple imputation. *J. Roy. Statist. Soc. Ser. B* **60**, 127-144.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley & Sons, New York.

Lo, S.-H. (1991). Estimating a survival function with incomplete cause-of-death data. *J. Multivariate Anal.* **39**, 217-235.

Lo, S.-H. and Singh, K. (1985). The product-limit estimator and the bootstrap: some asymptotic representations. *Probab. Theory Related Fields* **71**, 455-465.

McKeague, I. W. and Subramanian, S. (1998). Product-limit estimators and Cox regression with missing censoring information. *Scand. J. Statist.* **25**, 589-601.

Rao, J. N. K. and Shao, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* **79**, 811-822.

Robins, J. and Rotnitzky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. *AIDS Epidermiology-Methodological Issues.* (Edited by Jewell N., Dietz K. and Farewell V), 297-331. Birlzhauser, Boston.

Robins, J., Rotnizky, A. and Zhao, L. P. (1994), Estimation of Regression Coefficients when Some Regressors Are Not Always Observed. *J. Amer. Statist. Assoc.* **89**, 846-866.

Robins, J. M. and Wang, N. (2000). Inference for imputation estimators. *Biometrika* **87**, 113-124.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. Wiley, New York.

Shorack, G. R. and Wellner, J. A. (1986). Empirical Processes with Applications to Statistics. John Wiley & Sons, New York.

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.

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.

Wang, C. Y., Wang, S., Gutierrez, R. G. and Carroll, R. J. (1998). Local linear regression for generalized linear models with missing data. *Ann. Statist.* **26**, 1028-1050.

Wang, Q. H. and Rao, J. N. K. (2001). Empirical likelihood for linear regression models for missing response. *Canad. J. Statist.* **29**, 597-608.

Wang, Q. H. and Rao, J. N. K. (2002a). Empirical likelihood-based inference in linear models with missing data. *Canad. J. Statist.* **29**, 563-576.

Wang, Q. H. and Rao, J. N. K. (2002b). Empirical likelihood-based inference under imputation for missing response data. *Ann. Statist.* **30**, 896-924.

Wang, Q. H., Linton, O. and Härdle, W. (2004). Semiparametric regression analysis with missing response at random. *J. Amer. Statist. Assoc.* **99**, 334-345.

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

Zhao, L. P., Lipsitz, S. R. and Lew, D. (1996). Regression analysis with missing covariate data using estimating equations. *Biometrics* **52**, 1165-1182

Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100080, China.

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: qhwang@amss.ac.cn

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: kaing@hku.hk

(Received June 2006; accepted October 2006)