Published in final edited form as:
Stat Methodol. 2006 ; 3(2): 125–136. doi:10.1016/j.stamet.2005.09.014.

Survival analysis for the missing censoring indicator model using kernel density estimation techniques

Sundarraman Subramanian*
Department of Mathematics and Statistics, University of Maine, United States

Abstract

This article concerns asymptotic theory for a new estimator of a survival function in the missing censoring indicator model of random censorship. Specifically, the large sample results for an inverse probability-of-non-missingness weighted estimator of the cumulative hazard function, so far not available, are derived, including an almost sure representation with rate for a remainder term, and uniform strong consistency with rate of convergence. The estimator is based on a kernel estimate for the conditional probability of non-missingness of the censoring indicator. Expressions for its bias and variance, in turn leading to an expression for the mean squared error as a function of the bandwidth, are also obtained. The corresponding estimator of the survival function, whose weak convergence is derived, is asymptotically efficient. A numerical study, comparing the performances of the proposed and two other currently existing efficient estimators, is presented.

Keywords

Bandwidth sequence; Functional delta method; Independent increments; Kernel density estimator; Lyapounov central limit theorem; Standard Wiener process

1. Introduction

The missing censoring indicator (MCI) model arises from the random censorship model when the so-called failure indicators are missing for a subset of the sampled items. Let T be a failure time and C a censoring time independent of T. In the classical random censorship model one observes n i.i.d. observations of $X = \min(T, C) = T \wedge C$, and $\delta = I(X = T)$. The survival function of T, denoted by $S(t)$, is often of interest in survival studies and is estimated by the Kaplan–Meier [16] estimator, also known as the product-limit estimator. The Kaplan–Meier estimator has been well studied in survival analysis; its asymptotic normality was first derived by Breslow and Crowley [2], and its asymptotic efficiency was proved by Wellner [26]. Indeed, the estimator has been shown to possess excellent rates of convergence, see [11,3,4,7,19,15,6], among others. However, the Kaplan–Meier estimator is not suitable for the MCI model.

The data for the MCI model differ from standard right censored data in the following manner. Let ξ be an indicator variable that may depend on $X = \min(T, C)$, assuming the value 1 when the censoring indicator δ is observed and taking the value 0 otherwise. The observed data in the MCI model are, thus, n i.i.d. observations of (X,ξ,σ) where $\sigma = \xi\delta$. Furthermore, the censoring indicators are assumed to be missing at random (MAR), in which framework one has that $P(\xi = 1 \mid X = x, \sigma = d) = P(\xi = 1 \mid X = x) = \pi(x)$; see [21]. Indeed, under MAR ξ and δ

* Corresponding address: University of Maine, Mathematics and Statistics, 5752 Neville Hall University of Maine, 04469−5752 Orono, ME, United States. Tel.: +1 207 581 3934; fax: +1 207 581 3902. *E-mail address:* subraman@math.umaine.edu..

are conditionally independent given X. The unsuitability of the Kaplan–Meier estimator necessitates a search for new estimators of $S(t)$ in the MCI model.

Van der Laan and McKeague [17] first addressed efficient estimation of $S(t)$ in the MCI model. They introduced two estimators and provided analysis for one of them, viz. the “reduced-data” nonparametric maximum likelihood estimator (NPMLE), and derived the efficient influence curve (EIC); the second moment of the EIC provides the efficiency bound for estimating $S(t)$. More recently, Subramanian [24] derived the asymptotic normality of an estimator arising out of a representation specified by Dikta [9], and proved that it achieved the efficiency bound. On the other hand, when ξ and δ are completely independent, which is a stronger assumption than MAR, [18,12,20], among others, have also studied estimation of $S(t)$.

In the MCI model, however, several important issues need to be addressed. First, all approaches, including ours here, require estimates of conditional quantities, and hence smoothing. In the case of van der Laan and McKeague's [17] two estimators, expressions for bias, variance, and mean squared error (MSE) were not derived. They also required a “slightly stronger” assumption than MAR. Furthermore, they did not provide for their estimator (i) an almost sure representation with a rate for the remainder term, and (ii) uniform strong consistency with rate of convergence. Subramanian [24], on the other hand, derived approximate expressions for the bias and variance of his kernel-based estimator of $S(t)$, as a function of the bandwidth, but did not study (i) and (ii) listed above or weak convergence.

We propose a new and readily comprehensible estimator that is almost Kaplan–Meier-estimator-like and that is easy to compute as well. The estimator is based on a representation for $\Lambda(t)$, the cumulative hazard of T. Denote the distribution of X by $H(t)$, its survival function by $\bar{H}(t)$, and the subdistribution function of the “observed uncensored time” by $H_{11}(t) = P(X \le t, \xi = 1, \sigma = 1)$. Under MAR, it follows that $\Lambda(t)$ has the form

$$\Lambda(t) = \int_0^t \frac{1}{\pi(s)} \frac{\mathrm{d}H_{11}(s)}{\bar{H}(s-)}. \tag{1.1}$$

Let $\widehat{H}_{11}(t) = n^{-1}\sum_{i=1}^{n} I(X_i \le t, \xi_i = 1, \sigma_i = 1)$ and $\widehat{\bar{H}}(t) = n^{-1}\sum_{i=1}^{n} I(X_i > t)$ denote the empirical estimators of $H_{11}(t)$ and $\bar{H}(t)$. Let $\widehat{\pi}(t)$ denote a kernel estimator of $\pi(t)$ (cf. Section 2). Plugging these estimators into (1.1) we get our estimator $\widehat{\Lambda}(t)$. The product integral [1] then provides the estimator $\hat{S}(t)$. The main results in this paper are an almost sure representation for $\widehat{\Lambda}(t) - \Lambda(t)$, including a rate for the remainder term, asymptotic normality and uniform strong consistency of $\widehat{\Lambda}(t)$, an expression for the MSE of $\widehat{\Lambda}(t)$ as a function of the bandwidth, and weak convergence.

The paper is organized as follows. In Section 2 we formally introduce our proposed estimator, derive an almost sure representation for $\widehat{\Lambda}(t)$, state its asymptotic normality and uniform strong consistency, and derive its approximate mean squared error. We then derive the weak convergence of $n^{1/2}(\hat{S}(t - S(t))$ to a standard Wiener process. In Section 3 we present a numerical comparison study. In Section 4 we give a concluding discussion. Proof of Lemma 1 and some variance calculations are provided in Appendix A.

2. The estimator and large sample results

In this section we first state results for certain kernel estimators used in our approach. We then introduce our estimator and derive (i) an almost sure representation for $\left(\widehat{\Lambda}(t) - \widehat{\Lambda}(t)\right)$, (ii) mean squared error of $n^{1/2}\left(\widehat{\Lambda}(t) - \widehat{\Lambda}(t)\right)$, and (iii) weak convergence of $n^{1/2}(\hat{S}(t) - S(t))$.

2.1. Kernel estimators

Let τ be such that $\bar{H}(\tau) > 0$, and $K_v(u) = v^{-1} K(u/v)$, where K is a kernel function. The kernel density estimator of $h(t)$, the density of X, takes the form (e.g. [22])

$$h_n(t) = \frac{1}{n}\sum_{i=1}^{n} K_{a_n}(t - X_i), \qquad t \in \mathbb{R}, \tag{2.1}$$

where a_n is a bandwidth sequence. Let $\tilde{h}(t) = E(h_n(t))$. Denote the kernel estimator of $h_1(t) = P(X = t, \xi = 1)$ by $h_{n,1}(t) = n^{-1}\sum_{i=1}^{n} K_{a_n}(t - X_i)\xi_i$ and let $\tilde{h}_1(t) = E(h_{n,1}(t))$. Furthermore, let $\pi_n(t) = h_{n,1}(t)/h_n(t)$ denote the estimator of $\pi(t) = h_1(t)/h(t)$. Suppose that the following conditions hold, see also [23] or[8]:

B The sequence $a_n \downarrow 0$ satisfies $na_n^2/(\log(a_n))^2 \to \infty$, $na_n^4 \to 0$ and $|\log a_n|/\log\log n \to \infty$.

K (i) The kernel K is a probability density function with support on $[-M, M]$ for some $M > 0$. (ii) K is of bounded variation. (iii) The first moment of K is 0: $\int_{-M}^{M} uK(u)\,\mathrm{d}u = 0$.

D (i) The function h is bounded, bounded away from 0, and uniformly continuous on $\mathbb{R}$. (ii) The same holds for the function π. (iii) The functions h and π are each twice differentiable, and the derivatives are bounded and bounded away from 0 uniformly in $[0, \tau]$. (iv) The density of T, namely $f(x)$, is twice differentiable in $[0, \tau]$.

The condition $na_n^2/|\log(a_n)|^2 \to \infty$ is required for making $n^{1/2}$ times a remainder term (cf. Lemma 1) negligible. The same condition is required for weak convergence also. The condition $na_n^4 \to 0$ ensures that the limiting normal distribution has zero mean (That is, the bias is 0.). Specifically, the standard optimal one, namely $a_n = o(n^{-1/5})$ will not suffice.

Under conditions **B**, **K** and **D**, $\sup_{a_n M \le t \le \tau} |h_n(t) - h(t)| = O\left(\left(\frac{na_n}{\log a_n^{-1}}\right)^{-1/2} + a_n^2\right)$ a.s. (e.g. [8]), with analogous rates holding for h_n, $1(t)$ and $\pi_n(t)$.

2.2. Proposed estimator and an almost sure representation

We define our estimator of $\pi(t)$ by $\hat{\pi}(t) = \hat{h}_1(t)/\hat{h}(t)$, where $\hat{h}(t) = h_n(a_n M)I(t < a_n M) + h_n(t)I(t \ge a_n M)$, and $\hat{h}_1(t)$ is analogous to $\hat{h}(t)$ but with $h_n(\cdot)$ replaced with $h_{n,1}(\cdot)$. We define the estimator $\hat{S}(t)$ through the product integral: $\hat{S}(t) = \pi_{0 \le s \le t}\{1 - d\hat{\Lambda}(s)\}$, where

$$\hat{\Lambda}(t) = \int_0^t \frac{1}{\hat{\pi}(s)\hat{\bar{H}}(s-)}\mathrm{d}\hat{H}_{11}(s), \qquad t \le \tau. \tag{2.2}$$

Let $p(t) = P(\delta = 1 \mid X = t)$ and $P(t) = p(t)/(\pi(t)\bar{H}(t))$. Define $\mu_l(u) = \int_{-M}^{u} v^l K(v)\,\mathrm{d}v, l = 0, 1, 2$. Henceforth, for example, $f(x)^2$ means $(f(x))^2$. Define the following:

$$\zeta_{1,i}(t) = \int_0^t \frac{1}{\pi(x)\bar{H}(x)}\mathrm{d}\left(I(X_i \le x, \xi_i = 1, \sigma_i = 1) - H_{11}(x)\right), \tag{2.3}$$

$$\zeta_{2,i}(t) = \int_0^t \frac{I(X_i \le x) - H(x)}{\pi(x)\bar{H}(x)^2}\mathrm{d}H_{11}(x), \tag{2.4}$$

$$\zeta_{3,i}(t) = -(\xi_i - \pi(X_i))\,P(X_i)\,I(X_i \le t), \tag{2.5}$$

$$V(t)=\int_0^t \frac{\mathrm{d}\Lambda(x)}{\bar{H}(x)}+\int_0^t \frac{1-\pi(x)}{\pi(x)}(1-p(x))\frac{\mathrm{d}\Lambda(x)}{\bar{H}(x)}. \tag{2.6}$$

Here $V(t)$ is the information bound (the second moment of EIC) for estimating $\Lambda(t)$. Define

$$A_n(t)=(\xi-\pi(X))\ P(X)I(X\le 2a_nM)\mu_0\left(\frac{a_nM-X}{a_n}\right)$$

$$C_n(t)=(\xi-\pi(X))\ P(X)I(t-a_nM\le X\le t+a_nM)\left(1-\mu_0\left(\frac{t-X}{a_n}\right)\right)$$

$$D_n(t)=-(\xi-\pi(X))\ P(X)I(t\le X\le t+a_nM).$$

Let $\widehat{A}(t)=n^{-1}\sum_{i=1}^{n}\left[A_{n,i}(t)+B_{n,i}(t)+C_{n,i}(t)\right]$. Note that, although $A_n(t)$, $C_n(t)$, and $D_n(t)$ have mean 0, they have second as well as cross product moments of the order of $O(a_n)$, which are necessary for a proper analysis of the mean squared estimator (MSE) of $\widehat{\Lambda}(t)$, one goal of this article. The following lemma gives the almost sure representation for $\widehat{\Lambda}(t)-\widehat{\Lambda}(t)$.

Lemma 1—*Assume conditions* **B**, **K**, & **D**. *Then, for* $t\le\tau$,

$$\widehat{\Lambda}(t)-\Lambda(t)=\frac{1}{n}\sum_{i=1}^{n}\{\zeta_{1,i}(t)+\zeta_{2,i}(t)+\zeta_{3,i}(t)\}+\widehat{A}(t)+R(t), \tag{2.7}$$

where $n^{1/2}\hat{A}(t)$ converges to $N(0, O(a_n))$ *and is, therefore, negligible, and R(t) satisfies*

$$\sup_{0<t\le\tau}|R(t)|=O\left(\left(\frac{n}{a_n\log a_n^{-1}}\right)^{-1/2}+\left(\frac{na_n}{\log a_n^{-1}}\right)^{-1}+a_n^2\right)\text{a.s.}\quad as\quad n\to\infty. \tag{2.8}$$

The proof of Lemma 1 is quite technical and involved, and a concise version is given in Appendix A. For a detailed proof we refer the reader to our technical report entitled "Survival Analysis for the Missing Censoring Indicator Model" [25].

2.3. Approximate mean squared error of $\widehat{\Lambda}(t)$

By condition **D**(iv), the hazard function $\lambda(t)$ is twice differentiable. Let $B_n(t) = -\beta_n + \alpha_n(t)$, see Eqs. (A.2) and (A.3) for β_n and $\alpha_n(t)$. UsingTaylor expansions it can be shown that

$$\beta_n=\int_0^{a_nM}\lambda(s)\,\mathrm{d}s-\frac{1}{\pi(a_nM)}\int_0^{a_nM}\pi(s)\lambda(s)\,\mathrm{d}s=\frac{1}{2}a_n^2M^2\frac{\pi'(0)}{\pi(0)}\lambda(0)+o\left(a_n^2\right).$$

Furthermore, it can be shown that (cf. author's technical report)

$$\alpha_n(t)=-\frac{1}{2}a_n^2\mu_2(M)\int_0^t\left(2\frac{\pi'(y)}{\pi(y)}\frac{h'(y)}{h(y)}+\frac{\pi''(y)}{\pi(y)}\right)\lambda(y)\,\mathrm{d}y+o\left(a_n^2\right).$$

Therefore, up to $O\left(a_n^2\right)$, the bias is given by

$$B_n(t)=-\frac{1}{2}a_n^2M^2\frac{\pi'(0)}{\pi(0)}\lambda(0)-\frac{1}{2}a_n^2\mu_2(M)\int_0^t\left(2\frac{\pi'(y)}{\pi(y)}\frac{h'(y)}{h(y)}+\frac{\pi''(y)}{\pi(y)}\right)\lambda(y)\,\mathrm{d}y. \tag{2.9}$$

From Eq. (A.7), the variance of $n^{1/2}\left(\widehat{\Lambda}(t)-\widehat{\Lambda}(t)\right)$ is $V(t)+\sigma_n(t)^2+o(a_n)$. The approximate mean squared error of $\widehat{\Lambda}(t)$ is, thus, given by

$\mathrm{MSE}=\left(V(t)+\sigma_n(t)^2+nB_n(t)^2\right)/n=(V(t)+C_1a_n)/n+C_2a_n^4$ for appropriate constants C_1 and C_2. Minimization then gives the optimal bandwidth choice as $a_n = O(n^{-1/3})$.

2.4. Asymptotic normality of $\widehat{\Lambda}(t)$

Lemma 1 leads immediately to the following asymptotic normality result.

Proposition 1—*Assume conditions* **B**, **K**, & **D**. *Then, for* $t \le \tau$, *as* $n \to \infty$, *we have that* $n^{1/2}\left(\widehat{\Lambda}(t)-\Lambda(t)\right) \to N(0,V(t))$. *That is,* $\widehat{\Lambda}(t)$ *is asymptotically efficient for* $\Lambda(t)$.

2.5. Uniform strong consistency rate for $\widehat{\Lambda}(t)$

For proof of the following result see the technical report.

Proposition 2—*Assume conditions* **B**, **K**,& **D**. *Then, as* $n \to \infty$,

$$\sup_{0<t\le\tau}|\widehat{\Lambda}(t)-\Lambda(t)|=O\left(\left(\frac{n}{\log n}\right)^{-1/2}+a_n^2\right)\text{a.s.}$$

2.6. Weak convergence of $n^{1/2}(\hat{S}(t) - S(t))$

Let $D[0, \tau]$ and $D_{-}[0, \tau]$ denote the cadlag and "caglad" functions on $[0, \tau]$, W a standard Wiener process, and $\|\cdot\|_\infty$ the supremum norm. Note that Eq. (2.2) can be written as $\widehat{\Lambda}(t)=\int_0^t\widehat{\bar{H}}(s)^{-1}\mathrm{d}\widehat{H}_1(s)$, where $\hat{H}_1(t)$ takes the form $\widehat{H}_1(t)=\int_0^t\widehat{\pi}(s)^{-1}\mathrm{d}\widehat{H}_{11}(s)$. Define

$$\breve{\zeta}_1(t)=\int_0^t\frac{1}{\pi(x)}\mathrm{d}\left(I(X\le x,\xi=1,\sigma=1)-H_{11}(x)\right) \tag{2.10}$$

$$\breve{\zeta}_2(t)=-\int_0^t\left(I(X\ge s)-\bar{H}(s)\right)\mathrm{d}\Lambda(s) \tag{2.11}$$

$$\breve{\zeta}_3(t)=-(\xi-\pi(X))\frac{p(X)}{\pi(X)}I(X\le t) \tag{2.12}$$

$$V_1(t)=H_1(t)+\int_0^t\left(\frac{1-\pi(x)}{\pi(x)}\right)(1-p(x))\,\mathrm{d}H_1(x). \tag{2.13}$$

Here $V_1(t)$ is the information bound (the second moment of EIC) for estimating $H_1(t)$. We now present the main theorem of this article, the weak convergence of $n^{1/2}(\hat{S}(t) - S(t))$.

Theorem 1—*Assume* **B**, **K**, & **D**, *and that* $S(t)$ *is continuous. Then the process* $n^{1/2}(\hat{S}(t) - S(t))$ *converges weakly to* $S \cdot W(V_1(t))$. *In particular,* $\hat{S}(t)$ *is asymptotically efficient for* $S(t)$.

Proof—It can be shown as in the proof of Lemma 1 that

$$\widehat{H}_1(t)-H_1(t)=\frac{1}{n}\sum_{i=1}^{n}\left\{\breve{\zeta}_{1,i}(t)+\breve{\zeta}_{3,i}(t)\right\}+o_p\left(n^{-1/2}\right).$$

Note also that $n^{-1}\sum_{i=1}^{n}\breve{\zeta}_{2,i}(t)=-\int_0^t\left(\widehat{\bar{H}}(s-)-\bar{H}(s)\right)\mathrm{d}\Lambda(s)$. Now, Gill and Johansen [13] have shown that weak convergence of $n^{1/2}(\hat{S}(t) - S(t))$ follows directly from weak convergence of the basic bivariate process $n^{1/2}\left(\widehat{H}_1(t)-H_1(t),\widehat{\bar{H}}(t-)-\bar{H}(t)\right)$ and compact differentiability of a composition of these two basic processes. Specifically, let (Z_{H_1}, Z_H) denote a bivariate zero-mean Gaussian process, and $\|\cdot\|_\infty^\vee$ the max supremum norm. Then, it can be shown that the process $n^{1/2}\left(\widehat{H}_1(t) - H_1(t), \widehat{\bar{H}}(t-) - \bar{H}(t)\right) \xrightarrow{\mathcal{D}} \left(Z_{H_1}, Z_H\right)$, in $(D[0,\tau] \times D_-[0,\tau], \|\cdot\|_\infty^\vee)$, as $n \to$. The covariance structure of the limiting process (Z_{H_1}, Z_H) can be obtained from that of $\left(\widehat{H}_1(t) - H_1(t), \widehat{\bar{H}}(t-) - \bar{H}(t)\right)$. Thus, by the functional delta method, $n^{1/2}(\hat{S}(t) - S(t))$ converges weakly in $(D[0,\tau], \|\cdot\|_\infty)$ to $-S(t)\int_0^t \bar{H}(s)^{-1} \mathrm{d}U(s)$ as $n \to \infty$, where $U(t) = Z_{H_1}(t) - \int_0^t Z_H(s)\,\mathrm{d}\Lambda(s)$ (cf. [13], p. 1537).

To show that $U(x)$ has independent increments, it suffices to deal with $\gamma(x) = \gamma_1(x) + \gamma_2(x)$, where $\gamma_1(x) = \check{\zeta}_1(x) + \check{\zeta}_3(x)$. It is straightforward to show that, for $x < y$, $E(\gamma(x)\gamma(y)) = V_1(x)$. Note, in particular, that the variance function of $U(x)$ is $V_1(x)$. It follows that $E(\gamma(x)(\gamma(y) - \gamma(x))) = 0$, and this implies that $U(x)$ has independent increments. We then have exactly as in [13, p. 1538] that $n^{1/2}(\hat{S}(x) - S(x))$ converges weakly to $S \cdot W(V_1(x))$. The proof is complete. □

3. Numerical results

We now report the results of a simulation study comparing the performances of the three asymptotically efficient estimators, namely (1) the Dikta-type estimator proposed by Subramanian [24], (2) the reduced-data NPMLE, and (3) our proposed estimator. The comparisons are based on the mean integrated squared error (MISE).

For the kernel estimators we employed a uniform kernel on (−1, 1), and investigated with several different bandwidths (BWs). The BW that gave the lowest MISE and standard deviation (STDEV) among those used (0.025−0.30) was taken as the “best” BW. To compute the reduced-data NPMLE, we investigated with several subdivisions k of the interval of estimation $(0, \tau)$, where $\tau = 0.75$. The grid that provided the smallest MISE and STDEV among those used (5, 10, 25, 50, 75, 100, 150) was taken as the “best” one. For cases where it was difficult to choose the “best” grid, results are reported for more than one value of k.

The minimum X was taken to be uniformly distributed on (0, 1) and the conditional probability of non-censoring was taken to be $p(t) = c(1 - 0.25t^2)$. Denoting the censoring rate (CR) by r, we have that $1 - r = p(\delta=1) = \int_0^1 p(t)\,\mathrm{d}t = 11c/12$, so that c equals $12(1 - r)/11$. We only consider those values of r greater than 1/12, namely 0.1, 0.2, ..., 0.5(CRs 10%−50%). Note that $p(t)$ can exceed 1 when $r < 1/12$. With these choices of X and $p(t)$, it follows that $S(t) = \exp(-\Lambda(t))$, where $\Lambda(t) = c(-0.75 \log(1 - t) + t + t^2/2)$.

Two different models were employed for the missing mechanism, the *logit* and the *probit*. For the logit model we took $\pi(t) = e^t/(1 + e^t)$. Then, it follows that $P(\xi=1) = \int_0^1 \pi(t)\,\mathrm{d}t = \log((1+e)/2) = 0.620115$, which gives a missingness percentage of about 38. For the probit model we took $\pi(t) = \Phi(t/2)$, where $\Phi(t)$ is the standard normal cumulative distribution function, and this gives a missingness percentage of about 40.

In the case of the reduced-data NPMLE, for a bin with no complete observations we chose the smallest X_i in that bin to redistribute the mass. In each such bin, we assigned mass according to the rate $1 - r$. The results are based on 10,000 samples, each of size 100.

For the logit model, using Table 1a, we calculated 95% confidence intervals for the differences in the MISE of the three estimators. Note that for the 10% and 50% CRs the "best" k has been reported. When $k = 5$, the MISEs are not significantly different for 10%–30% CRs. For 40% CR, only the proposed estimator is significantly better than the reduced-data NPMLE. For 50% CR, however, the Dikta-type and proposed estimators are both significantly better than the reduced-data NPMLE. Indeed, the higher STDEV of the reduced-data NPMLE in some cases actually works in its favor (e.g. 40% CR), dominating the standard error of the difference. For finer partitions (k = 100 or 150), the reduced-data NPMLE performs significantly worse for moderate as well as higher CRs.

For the probit model, the reduced-data NPMLE is significantly better than *only* the Dikta-type estimator for 10% CR, see Table 1b below. When the partition is coarse, the MISEs of the reduced-data NPMLE and Dikta-type estimators are not significantly different for 20%–40% CRs, while the reduced-data NPMLE performs significantly worse than the proposed estimator for 40% CR. For 50% CR, however, the Dikta-type and proposed estimators are both significantly better. As for the logit model, the reduced-data NPMLE performs significantly worse for finer partitions and when the CR is moderate or high.

4. Concluding discussion

Although all the three estimators, namely the Dikta-type, the reduced-data NPMLE, and the proposed, are asymptotically optimal, our finite sample comparison study indicates that the reduced-data NPMLE may perform poorly for moderate to higher CRs. Even for low CRs, where the reduced-data NPMLE outperforms the Dikta-type estimator, the proposed estimator fares well. Contrary to the expectation that finer grids improve the reduced-data NPMLE performance, coarser grids provided best performance in some cases. It was also observed in our simulation study that coarser grids provided lower MISE but higher STDEV, while finer grids reduced the STDEV but not the MISE. Indeed in some cases it was hard to choose the "best" partition. We believe that the reduced-data NPMLE approach lacks definite practical guidelines for optimal partitioning of the interval of estimation.

The proposed and the Dikta-type estimators, on the other hand, would require the user to supply the bandwidth, a key parameter, for computing the estimator. Fortunately, bandwidth selection is a well investigated area and there is more than one solution. For example, one may provide an asymptotic representation for the MISE and choose the bandwidth that minimizes it [5] or its bootstrap estimate [10,14]. This will be addressed in future work.

Subramanian [24] has noted that the Dikta-type estimator is relatively less complicated to program on the computer. This is true of the proposed estimator also. The FORTRAN program segment that calculates the estimator runs only about 20 lines. For the reduced-data NPMLE, however, the user has to keep track of bins with no complete observations and provide appropriate mass at an arbitrarily chosen point, introducing additional complexity. Indeed, the kernel estimators facilitate coding. Thus computational ease is a definite plus for the proposed estimator. Furthermore, our approach does not rely on "reducing" the data and employs the readily understood substitution principle—unknown quantities are replaced with their empirical counterparts or kernel estimators. Our estimator, unlike the reduced-data NPMLE, would require only MAR to hold. Finally, we have provided an explicit expression for the approximate MSE of $\widehat{\Lambda}(t)$, not available for the reduced-data NPMLE.

Acknowledgements

I am grateful to referees and an Associate Editor for their useful comments, and to Prof. Jogesh Babu for allowing a revision. Thanks are also due to Prof. Peter Major for explaining a technicality in his published paper with Prof. L.

Rejtö. This research was initiated during the author's sabbatical leave at the University of Maine in Fall 2003. Research partly supported by National Institutes of Health grant CA 103845.

Appendix A

A.1. Proof of Lemma 1

For a detailed proof, see [25]. Let $\bar{h}(t) = E(\hat{h}(t))$ and $\bar{h}_1(t) = E(\hat{h}_1(t))$. Denote the empirical distribution function of X by $\hat{H}(t)$. The following basic representation for $\widehat{\Lambda}(t) - \Lambda(t)$ can be checked through direct calculations:

$$\begin{aligned}\widehat{\Lambda}(t) - \Lambda(t) = \; & \int_0^t \frac{1}{\pi(s)\bar{H}(s)} \mathrm{d}\left(\widehat{H}_{11}(s) - H_{11}(s)\right) + \int_0^t \frac{1}{\pi(s)} \frac{\widehat{H}(s-) - H(s)}{\bar{H}(s)^2} \mathrm{d}\widehat{H}_{11}(s) \\ & - \int_0^t \frac{\widehat{\pi}(s) - \pi(s)}{\widehat{\pi}(s)\pi(s)\bar{H}(s)} \mathrm{d}\widehat{H}_{11}(s) + \int_0^t \frac{\left(\widehat{H}(s-) - H(s)\right)^2}{\widehat{\pi}(s)\bar{\widehat{H}}(s-)\bar{H}(s)^2} \mathrm{d}\widehat{H}_{11}(s) \\ & - \int_0^t \frac{\widehat{\pi}(s) - \pi(s)}{\widehat{\pi}(s)\pi(s)} \frac{\widehat{H}(s-) - H(s)}{\bar{H}(s)^2} \mathrm{d}\widehat{H}_{11}(s).\end{aligned} \tag{A.1}$$

Write $\widehat{\Lambda}(t) - \Lambda(t) = \sum_{k=1}^5 I_k(t)$. It is straightforward to show that $I_1(t) = n^{-1} \sum_{i=1}^n \zeta_{1,i}(t)$. Next, it can be shown after some work that, almost surely, $I_2(t) = n^{-1} \sum_{i=1}^n \zeta_{2,i}(t) + O\left((n/\log n)^{-1}\right)$ and $\sup_{t \in [0,\tau]} |I_4(t)| = O\left((n/\log n)^{-1}\right)$. The negligibility of $I_5(t)$ will be clear after we address $I_3(t)$ below. First, after considerable calculations in can be shown that

$$I_{32}(t) \equiv \int_0^{a_n M} \frac{\widehat{\pi}(s) - \pi(s)}{\widehat{\pi}(s)\pi(s)\bar{H}(s)} \mathrm{d}\widehat{H}_{11}(s) = \beta_n + O\left(\left(\frac{n}{a_n \log a_n^{-1}}\right)^{-1/2} + a_n^3\right) \text{ a.s.,}$$

where

$$\beta_n \equiv \frac{1}{\pi(a_n M)} \int_0^{a_n M} \left(\pi(a_n M) - \pi(s)\right) \mathrm{d}\Lambda(s) = O\left(a_n^2\right). \tag{A.2}$$

Therefore, $I_{32}(t) = O\left(\left(\frac{n}{a_n \log a_n^{-1}}\right)^{-1/2} + a_n^2\right)$ a.s. Recall that $P(y) = p(y)/(\pi(y)\bar{H}(y))$. Next,

$$\begin{aligned} I_{31}(t) & \equiv - \int_{a_n M}^t \frac{\widehat{\pi}(s) - \pi(s)}{\widehat{\pi}(s)\pi(s)\bar{H}(s)} \mathrm{d}\widehat{H}_{11}(s) \\ & = - \int_{a_n M}^t \frac{\widehat{\pi}(s) - \pi(s)}{\pi(s)^2 \bar{H}(s)} \mathrm{d}\widehat{H}_{11}(s) + R_{31}(t) + R_{32}(t), \end{aligned}$$

where, using a representation for $\widehat{\pi}(t) - \pi(t)$, it can be shown that almost surely and uniformly for $t \in [a_n M, \tau]$ the remainder terms $R_{31}(t)$ and $R_{32}(t)$ are asymptotically negligible:

$$|R_{31}(t)| = \left| \int_{a_n M}^t \frac{\left(\widehat{\pi}(s) - \pi(s)\right)^2}{\widehat{\pi}(s)\pi(s)^2 \bar{H}(s)} \mathrm{d}\widehat{H}_{11}(s) \right| = O\left(\left(\left(\frac{n a_n}{\log a_n^{-1}}\right)^{-1/2} + a_n^2\right)^2\right) \text{ a.s.}$$

$$\begin{aligned} |R_{32}(t)| & = \left| \int_{a_n M}^t \frac{\widehat{\pi}(s) - \pi(s)}{\pi(s)^2 \bar{H}(s)} \mathrm{d}\left(\widehat{H}_{11}(s) - H_{11}(s)\right) \right| \\ & = O\left(\left(\frac{n a_n}{\log a_n^{-1}}\right)^{-1} + a_n^2 \left(\frac{n a_n}{\log a_n^{-1}}\right)^{-1/2}\right). \end{aligned}$$

Using the representation for $\widehat{\pi}(t) - \pi(t)$ it follows that

$$I_{31}(t) = n^{-1}\sum_{i=1}^{n}(\rho_{1,i}(t)+\rho_{2,i}(t))+\alpha_n(t) +O\left(\left(\frac{na_n}{\log a_n^{-1}}\right)^{-1}+a_n^2\left(\frac{na_n}{\log a_n^{-1}}\right)^{-1/2}+a_n^4\right)\text{ a.s.},$$

uniformly for $t \in [a_nM, \tau]$, where

$$\rho_{1,i}(t) = -\int_{a_nM}^{t}\left(K_{a_n}(y-X_i)\xi_i - \bar{h}_1(y)\right)P(y)\,dy, \rho_{2,i}(t)=\int_{a_nM}^{t}\left(K_{a_n}(y-X_i)-\bar{h}(y)\right)\pi(y)P(y)\,dy$$

and

$$\alpha_n(t)=\int_{a_nM}^{t}\left(\bar{h}(y)-h(y)\right)\pi(y)P(y)\,dy-\int_{a_nM}^{t}\left(\bar{h}_1(y)-h_1(y)\right)P(y)\,dy=O\left(a_n^2\right). \tag{A.3}$$

Let $t - a_nM=t_n^-, t+a_nM=t_n^+$. We can show that the 'non-a_n' terms of $\rho_1(t)$ are given by

$$-[\xi P(X)I(X\le t)-\Lambda(t)]-\left[\xi P(X)I(t\le X\le t_n^+)-\int_t^{t_n^+}d\Lambda(x)\right]$$

$$+\left[\xi P(X)\mu_0\left(\frac{a_nM-X}{a_n}\right)I(X\le 2a_nM)-\int_0^{2a_nM}\mu_0\left(\frac{a_nM-X}{a_n}\right)d\Lambda(x)\right]$$

$$+\left[\xi P(X)\left(1-\mu_0\left(\frac{t-X}{a_n}\right)\right)I(t_n^-\le X\le t_n^+)-\int_{t_n^-}^{t_n^+}\left(1-\mu_0\left(\frac{t-X}{a_n}\right)\right)d\Lambda(x)\right].$$

Likewise, we can show that the 'non-a_n' terms of $\rho_2(t)$ are given by replacing $\xi P(X)$ everywhere above with $\pi(X)P(X)$, and interchanging the signs outside the four square brackets. Therefore, it follows that the 'non-a_n' terms of $\rho_1(t) + \rho_2(t)$ are given by $\zeta_3(t) + A_n(t)\, C_n(t) + D_n(t)$. We next consider the '$a_n$' terms of $\rho_1(t) + \rho_2(t)$. The 'a_n' terms of $\rho_1(t)$ are

$$a_n\left[\xi P'(X)\mu_1\left(\frac{a_nM-X}{a_n}\right)I(X\le 2a_nM)-\int_0^{2a_nM}\mu_1\left(\frac{a_nM-X}{a_n}\right)P'(x)\pi(x)\,dH(x)\right]$$

$$-a_n\left[\xi P'(X)\mu_1\left(\frac{t-X}{a_n}\right)I(t_n^-\le X\le t_n^+)-\int_{t_n^-}^{t_n^+}\mu_1\left(\frac{t-X}{a_n}\right)P'(x)\pi(x)\,dH(x)\right].$$

Note that each term 'a_n' has mean 0 variance $o\left(a_n^2\right)$. Thus, by the Lyapounov central limit theorem, the 'a_n' terms of $n^{-1/2}\sum_{i=1}^{n}(\rho_{1,i}(t)+\rho_{2,i}(t))$ are each asymptotically negligible. This analysis can be extended to the 'a_n^2'terms and so on. Collecting all the remainder terms into $R(t)$, we have Eqs. (2.7) and (2.8). The lemma is proved. □

A.2. Variance calculations

The following expressions can be verified after straightforward calculations:

$$E\left(\zeta_1(t)^2\right)+E\left(\zeta_2(t)^2\right)+2E(\zeta_1(t)\zeta_2(t))=\int_0^t\frac{d\Lambda(x)}{\pi(x)\bar{H}(x)} \tag{A.4}$$

$$E\left(\zeta_3(t)^2\right)=\int_0^t p(x)\frac{1-\pi(x)}{\pi(x)}\frac{d\Lambda(x)}{\bar{H}(x)} \tag{A.5}$$

$$2E(\zeta_1(t)\zeta_3(t))+2E(\zeta_2(t)\zeta_3(t))=-2\int_0^t p(x)\frac{1-\pi(x)}{\pi(x)}\frac{d\Lambda(x)}{\bar{H}(x)}. \tag{A.6}$$

Note that adding Eqs. (A.4)–(A.6), gives $V(t)$, see Eq. (2.6). The following expressions may also be checked in a relatively straightforward manner:

$$E\left(A_n(t)^2\right)=a_n\frac{1-\pi(0)}{\pi(0)}p(0)^2h(0)\int_{-M}^{M}\mu_0(u)^2\mathrm{d}u+O\left(a_n^2\right)$$

$$E\left(C_n(t)^2\right)=a_n\frac{1-\pi(t)}{\pi(t)}p(t)^2\frac{h(t)}{\bar{H}(t)^2}\int_{-M}^{M}(1-\mu_0(u))^2\mathrm{d}u+O\left(a_n^2\right)$$

$$E\left(D_n(t)^2\right)=a_nM\frac{1-\pi(t)}{\pi(t)}p(t)^2\frac{h(t)}{\bar{H}(t)^2}+O\left(a_n^2\right)$$

$$2E(C_n(t)D_n(t))=-2a_n\frac{1-\pi(t)}{\pi(t)}p(t)^2\frac{h(t)}{\bar{H}(t)^2}\int_{-M}^{0}(1-\mu_0(u))\,\mathrm{d}u+O\left(a_n^2\right)$$

$$2E(\zeta_1(t)A_n(t))=2a_n\frac{1-\pi(0)}{\pi(0)}p(0)^2h(0)\int_{-M}^{M}\mu_0(u)\,\mathrm{d}u+O\left(a_n^2\right)$$

$$2E(\zeta_1(t)C_n(t))=2a_n\frac{1-\pi(t)}{\pi(t)}p(t)^2\frac{h(t)}{\bar{H}(t)^2}\int_{0}^{M}(1-\mu_0(u))\,\mathrm{d}u+O\left(a_n^2\right)$$

$$2E(\zeta_3(t)A_n(t))=-2a_n\frac{1-\pi(0)}{\pi(0)}p(0)^2h(0)\int_{-M}^{M}\mu_0(u)\,\mathrm{d}u+O\left(a_n^2\right)$$

$$2E(\zeta_3(t)C_n(t))=-2a_n\frac{1-\pi(t)}{\pi(t)}p(t)^2\frac{h(t)}{\bar{H}(t)^2}\int_{0}^{M}(1-\mu_0(u))\,\mathrm{d}u+O\left(a_n^2\right).$$

Let $\eta(t)=\sum_{k=1}^{3}\zeta_k(t)+A_n(t)+C_n(t)+D_n(t)$. The cross product moments of $\zeta_2(t)$ with $A_n(t)$, $C_n(t)$ or $D_n(t)$ is 0. Adding the equations, $E\left(\eta(t)^2\right)=V(t)+\sigma_n(t)^2+O\left(a_n^2\right)$, where

$$\begin{aligned}\sigma_n(t)^2= \quad & a_n\tfrac{1-\pi(0)}{\pi(0)}h(0)\,p(0)^2\left(\int_{-M}^{M}\mu_0(u)^2\mathrm{d}u\right)+a_n\tfrac{1-\pi(t)}{\pi(t)}\tfrac{h(t)}{\bar{H}(t)^2}p(t)^2\\ &\times\left(\int_{-M}^{M}\mu_0(u)^2\mathrm{d}u-2\int_{0}^{M}\mu_0(u)\,\mathrm{d}u+M\right).\end{aligned} \tag{A.7}$$

References

1. Andersen, PK.; Borgan, O.; Gill, RD.; Keiding, N. Statistical Models Based on Counting Processes. Springer-Verlag; New York: 1993.
2. Breslow N, Crowley J. A large sample study of the life table and product-limit estimates under random censorship. Ann. Statist 1974;2:437–453.
3. Burke MD, Csörgö S, Horváth L. Strong approximations of some biometric estimates under random censorship. Z. Wahrsch. verw. Gebiete 1981;56:87–112.
4. Burke MD, Csörgö S, Horváth L. A correction to and improvement of "Strong approximation of some biometric estimates under random censorship". Probab. Theory Related Fields 1988;79:51–57.
5. Cao R, Jácome MA. Presmoothed kernel density estimator for censored data. J. Nonparametr. Statist 2004;16:289–309.
6. Chen K, Lo S-H. On the rate of uniform convergence of the product-limit estimator: strong and weak Laws. Ann. Statist 1997;25:1050–1087.
7. Csörgö S, Horváth L. The rate of strong uniform consistency for the product-limit estimator. Z. Wahrsch. verw. Gebiete 1983;62:411–426.

8. Deheuvels, P. Uniform limit laws for kernel density estimators on possibly unbounded intervals. In: Limnios, N.; Nikulin, M., editors. Recent Advances in Reliability Theory: Methodology. Practice and Inference; Birkhauser, Boston: 2000. p. 477-492.

9. Dikta G. On semiparametric random censorship models. J. Statist. Plann. Inference 1998;66:253–279.

10. Faraway JJ, Jhun M. Bootstrap choice of bandwidth for density estimation. J. Amer. Statist. Assoc 1990;85:1119–1122.

11. Földes A, Rejtö L. Strong uniform consistency for nonparametric survival estimates from randomly censored data. Ann. Statist 1981;9:122–129.

12. Gijbels, I.; Lin, DY.; Ying, Z. Non- and semi-parametric analysis of failure time data with missing failure indicators, Technical Report 039−93. Mathematical Sciences Research Institute; Berkeley: 1993.

13. Gill RD, Johansen S. A survey of product-integration with a view toward application in survival analysis. Ann. Statist 1990;18:1501–1555.

14. González-Manteiga W, Cao R, Marron JS. Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. J. Amer. Statist. Assoc 1996;91:1130–1140.

15. Gu MG, Lai TL. Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation. Ann. Probab 1990;18:160–189.

16. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc 1958;53:457–481.

17. Van der Laan MJ, McKeague IW. Efficient estimation from right-censored data when failure indicators are missing at random. Ann. Statist 1998;26:164–182.

18. Lo S-H. Estimating a survival function with incomplete cause-of-death data. J. Multivariate. Anal 1991;39:217–235.

19. Major P, Rejtö L. Strong embedding of the estimator of the distribution function under random censorship. Ann. Statist 1988;16:1113–1132.

20. McKeague IW, Subramanian S. Product-limit estimators and Cox regression with missing censoring information. Scand. J. Statist 1998;25:589–601.

21. Rubin DB. Inference and missing data. Biometrika 1976;63:581–590.

22. Silverman, BW. Density Estimation for Statistics and Data Analysis. Chapman and Hall; 1986.

23. Stute W. A law of the logarithm for kernel density estimators. Ann. Probab 1982;10:414–422.

24. Subramanian S. Asymptotically efficient estimation of a survival function in the missing censoring indicator model. J. Nonparametr. Statist 2004;16:797–817.

25. Subramanian, S. Survival analysis for the missing censoring indicator model, Technical Report. Department of Mathematics and Statistics, University of Maine; Orono: 2005.

26. Wellner JA. Asymptotic optimality of the product limit estimator. Ann. Statist 1982;10:595–602.

Table 1

Mean integrated squared error of the estimators

(a) The logit model. Missingness percentage is about 38%.

CR (%)	Dikta-type			Reduced-data NPMLE			Proposed		
	MISE	**STDEV**	**BW**	**MISE**	**STDEV**	***k***	**MISE**	**STDEV**	**BW**
10	0.0210	0.0091	0.100	0.0208	0.0090	5	0.0208	0.0092	0.085
20	0.0204	0.0095	0.100	0.0202	0.0098	5	0.0202	0.0096	0.085
				0.0211	0.0091	100			
30	0.0194	0.0097	0.100	0.0194	0.0107	5	0.0192	0.0098	0.085
				0.0200	0.0089	150			
40	0.0179	0.0095	0.150	0.0181	0.0120	5	0.0177	0.0096	0.085
				0.0185	0.0089	150			
50	0.0157	0.0089	0.175	0.0166	0.0091	150	0.0156	0.0091	0.090

(b) The probit model. Missingness percentage is about 40%.

CR (%)	Dikta-type			Reduced-data NPMLE			Proposed		
	MISE	**STDEV**	**BW**	**MISE**	**STDEV**	***k***	**MISE**	**STDEV**	**BW**
10	0.0210	0.0091	0.10	0.0207	0.0092	10	0.0208	0.0093	0.080
20	0.0204	0.0095	0.10	0.0203	0.0100	5	0.0202	0.0097	0.080
				0.0211	0.0091	100			
30	0.0194	0.0097	0.10	0.0194	0.0109	5	0.0192	0.0099	0.085
				0.0200	0.0091	100			
40	0.0179	0.0095	0.15	0.0181	0.0122	5	0.0177	0.0097	0.085
				0.0185	0.0089	150			
50	0.0158	0.0091	0.15	0.0166	0.0091	150	0.0157	0.0092	0.085

Sample size is 100 and each MISE is based on 10,000 samples. Also given are grid size *k* and bandwidth BW.