Bootstrap Bandwidth for Estimation in the Missing Censoring Indicator Model

Sundar Subramanian[1]
Department of Mathematical Sciences
New Jersey Institute of Technology
Newark, New Jersey

Derek Bean
Department of Statistics
University of California
Berkeley

Abstract

In cancer survival studies, death certificate information can be missing, or incidental and fatal occurrences may be indistinguishable for some subjects, leading to missing censoring indicators (MCIs). For the framework of right censored data with MCIs, sub-density function kernel estimators play a significant role for estimating a survival function. Data-driven bandwidths for computing these kernel estimators are proposed. The bandwidths are obtained as minimizers of certain estimates of the mean integrated squared error (MISE). It is shown that the smoothed bootstrap offers a motivation for choosing the proposed MISE estimates for minimization. The efficacy of the proposed procedures is investigated through several simulation studies. Three illustrations are provided, using a mice data set, a data set extracted from the SEER database, and the well-known Primary Biliary Cirrhosis data.

KEY WORDS: Integrated squared bias, Inverse probability weighted, Least squares cross validation, Missing at random, Pilot bandwidth, Plug-in bandwidth selectors

[1]This research was supported by a National Institute of Health grant CA 103845.

1 INTRODUCTION

In survival studies, a number of data sets have the feature that the recorded failure type is dichotomous, namely death due to a particular specified disease as opposed to mortality arising from one of several other unspecified diseases. Typically, the focus is on estimating the probability of survival of patients suffering from the specified disease. For example, a number of registries of the Surveillance, Epidemiology, and End Results (SEER) Program database contain data specifying death as having occurred due to cancer or not due to cancer. In some cases, the cause of death (COD) is unknown. In the New Mexico registry, considering only deaths due to Nonhodgkins Lymphoma, out of 47 patients who were first diagnosed with the disease at age 43, 25 had unknown COD, while 9 died due to other causes. These and a plethora of other similar data can be extracted using the SEER*Stat 6.2.4. software. A second example concerns the survival times and disease status at death for 58 female RFM mice suffering from nonrenal vascular disease (NVD), analyzed by Dinse (1986). Of the 33 mice that died with NVD present, 8 died due to the disease, 19 from other known causes, and 6 had unknown COD.

Considering the time of death due to the specified disease as a failure time, the time of death due to the unspecified diseases as a censoring time, and unknown COD as a missing censoring indicator, the data can be subsumed under the framework of random right censorship with missing censoring indicators (MCIs). The focus of this article is to address comprehensively, for right censored survival data with MCIs, the important survivorship issue of estimating the true proportion of patients who will survive beyond any desired period

of time. Specifically, the issue is the investigation of (bootstrap-based) data-driven quantities, needed for the computation of certain conditional probabilities, leading to the provision of efficient survival function estimates when there are MCIs.

Clearly, the MCI model can be seen as arising from the classical random censorship model. Let T denote a failure time, C an independent censoring time, and let X denote the minimum of T and C. Suppose that ξ is a "missingness" indicator, assuming the value 1 or 0 according as whether the censoring indicator $\delta = I(T \leq C)$ is observed or not. We shall assume that δ is missing at random (MAR), which means that $P(\xi = 1|X, \delta) = P(\xi = 1|X)$. The data for the MCI model are n iid replicates of the triplet $Y = (X, \xi, \sigma)$, where $\sigma = \xi\delta$. Note that when $\sigma = 1$, the observation is uncensored. On the other hand, when $\sigma = 0$, we observe a censored observation ($\xi = 1$, $\delta = 0$) or a missing indicator ($\xi = 0$).

When there are MCIs, the Kaplan–Meier estimator of a survival function is unsuitable. Therefore, recent research has focused on finding a suitable alternative. Van der Laan and McKeague (1998) initiated asymptotically efficient estimation through their reduced data nonparametric maximum likelihood estimator. Their procedure, however, requires partitioning the interval of estimation, for which a moderate-sample optimal binning strategy was not investigated. In fact, unlike for standard right censoring, nonparametric estimation of a survival function from MCI data entails computation of estimates of conditional probabilities of binary variables. Therefore, smoothing, along with the related important issue of smoothing parameter (bandwidth) selection, is unavoidable. Two competing asymptotically efficient kernel-based estimators under the MAR framework have been investigated

(Subramanian 2004, 2006). These estimators were based on two different representations of the cumulative hazard function, in each case expressible as a smooth mapping of conditional probabilities of indicator variables and certain readily estimated functions of the distribution of X. Specifically, the choice of $p(t) = P(\delta = 1|X = t)$ leads to Dikta's (1998) representation, while that of $\pi(t) = P(\xi = 1|X = t)$ leads to the inverse probability weighted (IPW) representation. Indeed, the IPW technique has proven useful in several different settings, see Horvitz and Thompson (1952), Koul, Susarla, and van Ryzin (1981), Robins and Rotnitsky (1992), among others. Moving on to the problem at hand, the conditional probabilities can be represented in terms of certain subdensities, which can be estimated by kernel methods. For effective performance of kernel estimators, however, data driven smoothing parameters are essential, see Silverman (1986), Eubank (1988), Müller (1988) and Härdle (1989), among others. Toward this end, we investigate the efficacy of some bootstrap-based bandwidth (BW) selection procedures that target analysis of MCI data.

Note that ξ, σ, and δ are binary random variables whose conditional expectations are the objects of estimation, see Section 2.3 for details. To address the main issue, therefore, we shall focus on estimating the conditional expectation of a generic binary variable. Suppose that $(X_i, \eta_i), i = 1, \ldots, n$ are n iid pairs of (X, η), where η is the binary variable depending on X, and that interest centers on estimating the conditional probability $\upsilon(t) = P(\eta = 1|X = t) = h_1(t)/h(t)$. Here $h_1(t)$ is the subdensity corresponding to $P(X \leq t, \eta = 1)$ and $h(t)$ is the density of X. The estimator of $h_1(t)$ is given by $\hat{h}_1(t|a_n) = n^{-1} \sum_{i=1}^{n} \eta_i K_{a_n}(t - X_i)$, where $K_{a_n}(\cdot) = a_n^{-1} K(\cdot/a_n)$, K is a kernel function and a_n is a (generic) BW sequence tending to

0 as $n \to \infty$. For simplicity, we shall suppress the subscript in a_n and refer simply as a henceforth. The estimator of $h(t)$ is of course $\hat{h}(t|a) = n^{-1}\sum_{i=1}^{n} K_a(t - X_i)$ (cf. Silverman 1986), where, to simplify notation, we have used the same BW a.

The criterion often used for assessing the performance of kernel estimators is the mean integrated squared error (MISE), see, for example, Silverman (1986) or Wand and Jones (1995). In the present scenario this takes the form

$$M_{h_1}(a) \;=\; E\int \left(\hat{h}_1(x|a) - h_1(x)\right)^2 dx. \tag{1.1}$$

As in density estimation, $M_{h_1}(a)$ admits a representation which is the sum of the integrated variance and integrated squared bias. In particular, denoting $R(f) = \int f^2(x)dx$, $\rho = P(\eta = 1)$, and the convolution operator by *, it can be shown that

$$M_{h_1}(a) \;=\; (na)^{-1}\rho\, R(K) - n^{-1}R(K_a * h_1) + R\left\{(K_a * h_1)(x) - h_1(x)\right\}. \tag{1.2}$$

Analogous to density estimation, therefore, the MISE, given by Eq. (1.2), depends on the unknown subdensity h_1.

An alternative standard practice is to obtain a BW that minimizes an asymptotic representation of the MISE, denoted by $\tilde{M}_{h_1}(a)$, which represents the asymptotically leading terms of $M_{h_1}(a)$. We shall refer to this quantity henceforth as AMISE. For $l = 0, 1, 2$, let $\mu_l(K) = \int u^l K(u)du$. Note that $\mu_0(K) = 1$, that is, K is a probability density. After some calculations as in Cao (1993) or Wand and Jones (1995), assuming that $\mu_1(K) = 0$, it can be shown that $M_{h_1}(a) = \tilde{M}_{h_1}(a) + o\left\{(na)^{-1} + a^4\right\}$, where the AMISE is given by

$$\tilde{M}_{h_1}(a) \;=\; (na)^{-1}\,\rho\, R(K) + \frac{1}{4}a^4(\mu_2(K))^2 R(h_1''). \tag{1.3}$$

Here h_1'' is the second derivative of h_1. Its minimizer a_{AMISE} is then

$$a_{\text{AMISE}} = \left\{ \frac{\rho\, R(K)}{nR(h_1'')(\mu_2(K))^2} \right\}^{\frac{1}{5}}. \tag{1.4}$$

Therefore, the optimal rate of $O(n^{-1/5})$ for the subdensity BW is the same as that for density estimation. A plug-in BW selector based on Eq. (1.4) would require an estimate of $R(h_1'')$ (Hall and Marron 1987; Jones and Sheather 1991), the integrated squared subdensity second derivative. Estimation of a subdensity derivative, however, would entail additional BW estimation, which may be unattractive from the standpoint of estimating survival functions.

In this article we focus on obtaining suitable estimates of M_{h_1}, which we shall denote by $\hat{M}_{h_1}$, whose minimizations provide us with the desired data-driven BWs. Specifically, we investigate the approach that employs a pilot kernel estimate for h_1 and the usual sample proportion $\hat{\rho}$ for ρ in Eq. (1.2). It turns out that such an estimate $\hat{M}_{h_1}$ can be motivated through smoothed bootstrap considerations (Hall, Marron and Park 1992; Muller 1985; Staniswallis 1989).

Denote by $\hat{h}_1(x|g)$ a pilot kernel subdensity estimate of $h_1(x)$ with BW g and by $\hat{h}_1^*(x|a)$ the kernel subdensity estimate based on a smoothed bootstrap sample, that is, a sample drawn from the pilot subdensity $\hat{h}_1(x|g)$. We show that $\hat{M}_{h_1}$ is the theoretical expectation over bootstrap samples, E^*, of the bootstrap integrated squared error (ISE):

$$\hat{M}_{h_1}(a) = E^* \int \left(\hat{h}_1^*(x|a) - \hat{h}_1(x|g) \right) dx. \tag{1.5}$$

Grund and Polzehl (1997), among others, considered the above criterion in the context of density estimation, that is, with the density h in place of the subdensity h_1 in Eq. (1.5).

Before them, Taylor (1989), Faraway and Jhun (1990) considered an empirical criterion, that is, when E^* is taken as the average over a certain number of bootstrap replications, again for density estimation.

BW selection methods broadly come under one of two categories (Grund and Polzehl, 1997). The first, which includes cross validation (CV) and bootstrap BW selection methods, minimizes an empirical estimate of the MISE (Taylor, 1989; Faraway and Jhun, 1990). The second approach minimizes AMISE, leading to plug-in methods (Park and Marron, 1990; Hall, Sheather, Jones, and Marron, 1991; Jones, Marron, and Park, 1991). Efron (1982), and Silverman and Young (1987) investigated smoothed bootstrap to improve estimator performance. For the pilot BW, Hall, DiCiccio, and Romano (1989) recommended CV. Faraway and Jhun (1990) employed least squares CV for pilot BW selection in density estimation. See also Cao, Antonio, and Manteiga (1994), and Jones, Marron, and Sheather (1996) for surveys on BW selection for density estimation.

All the approaches of data-driven BW selection studied thus far have focused only on density estimation. In this article, we investigate data-driven BW selection for computing subdensities, since our prime motivation is in its utility for estimating survival functions. First we consider the efficacy of the empirical smoothed bootstrap of Faraway and Jhun (1990) in the context of subdensity estimation. Alternatively, instead of the empirical criterion, we shall compute the theoretical expectation E^* and obtain its minimizer. The latter approach avoids actual resampling for computing the BW based on the data at hand. Both approaches require an initial estimate of $h_1(x)$, however. We plug in the pilot kernel sub-

density estimate $\hat{h}_1(x|g)$, where the pilot BW g is selected through least squares CV. Based on simulation studies, this approach is reported to work well for density estimation (Grund and Polzehl 1997).

The article is organized as follows. In Section 2, we present smoothed bootstrap choice of BW estimation for a subdensity and provide three illustrations. In Section 3, we present the results of our simulation studies. Section 4 contains some concluding remarks. Relevant detailed calculations are shown in the Appendix.

2 BW for subdensity estimation

In this section, we first describe the smoothed bootstrap approach for subdensity BW estimation. Before we present our illustrations, however, we also provide a review of two survival function estimators whose moderate sample performance, using an "optimal" BW, is the object of comparison.

2.1 Smoothed bootstrap for a subdensity BW

Recall that the main issue is the estimation of the conditional expectation of a binary variable η given X, and that $(X_i, \eta_i), i = 1, \ldots, n$ are n iid pairs of (X, η). Denote the usual bootstrap sample by $\left\{(\tilde{X}_i, \tilde{\eta}_i)_{1 \leq i \leq n}\right\}$. The bootstrap estimate of h_1, denoted by $\tilde{h}_1$, is given by $\tilde{h}_1(x|a) = n^{-1} \sum_{i=1}^{n} K_a(x - \tilde{X}_i)\tilde{\eta}_i$. As in density estimation, we can show that the bootstrap bias of this estimate is zero. Therefore, from Eq. (1.2) the MISE of $\tilde{h}_1(x|a)$ is $O(n)$, which in turn gives a BW independent of n. This is undesirable, since the BW does not shrink to 0 with increasing n; see also Marron (1992) for a discussion of this aspect in

the context of density estimation.

The smoothed bootstrap eliminates the problem arising from the usual bootstrap. Since the focus is on a subdensity BW, we would need to perform smoothing for the resampled X_i's whose $\eta_i = 1$, using a pilot BW g (discussed in the next subsection) selected for the kernel subdensity estimate $\hat{h}_1(x|g)$. In particular, the CS1 resampling plan of Manteiga et al. (1996) can be employed to generate the smoothed resamples, but such an approach would require two pilot BWs, one each for $\eta = 1$ and $\eta = 0$, which needlessly increases the computational burden. In fact, it suffices to ensure that the resampled X_i's with $\eta_i = 1$ are properly smoothed, since the "improperly smoothed" X_i corresponding to $\eta_i = 0$ would be annihilated and do not figure in the analysis, see Eq. (2.6) below. In the following, therefore, we achieve our goal using only the pilot BW g corresponding to the subdensity estimate $\hat{h}_1(x|g)$. Let ϵ denote an independent random deviate drawn from the kernel K. For a given g, the *empirical iterated* smoothed bootstrap (EISB) works as follows:

Step 1: Resample from $(X_i, \eta_i), i = 1, \ldots, n$ to obtain $(\tilde{X}_i, \tilde{\eta}_i), i = 1, \ldots, n$.

Step 2: For each $i = 1, \ldots, n$, obtain $X_i^* = \tilde{X}_i + g\epsilon_i$. Also define $\eta_i^* = \tilde{\eta}_i$.

Step 3: Define the smoothed bootstrap kernel density estimate of h_1 as follows:

$$\hat{h}_1^*(x|a) \quad = \quad n^{-1} \sum_{i=1}^{n} K_a(x - X_i^*)\eta_i^* \tag{2.6}$$

Note that the "improper" X_i^* are annihilated by the η_i^* appearing in the sum above.

Step 4: Repeat the above procedure L times, for L bootstrap samples and L estimators.

Step 5: Form the empirical approximation to Eq. (1.5), also denoted by $\hat{M}_{h_1}(a)$ for conve-

nience, by taking E^* as the empirical mean. That is,

$$\hat{M}_{h_1}(a) \;=\; L^{-1}\sum_{j=1}^{L}\int\left(\hat{h}^*_{1,j}(x|a)-\hat{h}_1(x|g)\right)^2 dx. \tag{2.7}$$

Step 6: Obtain the one-step bootstrap choice of the BW by minimizing $\hat{M}_{h_1}(a)$ over a.

If a two-step bootstrap BW is desired, the value obtained in step 5 can be assigned as the new initial choice g, and steps 2 through 5 can be repeated.

Let $C(a) = (na)^{-1}\hat{\rho}\,R(K) + R\left(K_a * \hat{h}_1 - \hat{h}_1\right)$. From Eq. (A.3), the actual expression for the right hand side of Eq. (1.5) is given by

$$\hat{M}_{h_1}(a) \;=\; C(a) - n^{-1}R\left(K_a * \hat{h}_1\right). \tag{2.8}$$

Note that $C(a)$ is the sum of two quantities, the first of which approximates the integrated variance and the second is an estimate of the integrated squared bias. The minimizer of $\hat{M}_{h_1}(a)$ given by Eq. (2.8) defines the *theoretical* smoothed bootstrap (TSB) choice of BW. On the other hand, the minimizer of $C(a)$ defines the *smoothed cross validation* (SCV) choice of BW, see, for example, Hall, Marron, and Park (1992). Finally, after standard calculations, it can be shown that the integrated squared bias term in $C(a)$ takes the form $a^4\mu_2(K)^2R(\hat{h}''_1)/4$ plus a negligible $o(a^4)$ term. It follows that $\hat{M}_{h_1} = \hat{\tilde{M}}_{h_1} + o\left\{(na)^{-1} + a^4\right\}$, and the BW that minimizes $\hat{\tilde{M}}_{h_1}$ is given by Eq. (1.4), but ρ replaced with $\hat{\rho}$ and $R(h''_1)$ replaced with $R(\hat{h}''_1)$. We shall call this estimator *asymptotic* smoothed bootstrap (ASB) choice of BW. Note that $\hat{h}''_1$ is the second derivative of $\hat{h}_1$, and can be calculated in a straightforward way, but would require that K have a non-vanishing second derivative. As before, if a two-step BW is desired g can be updated and the process repeated.

2.2 Cross validation BW

We specify the pilot BW through least squares CV. Here we give a justification for the approach. Denote by $\hat{h}_{1,-i}$ the subdensity estimate based on all of $(X_j, \eta_j), j = 1, \ldots, n$, except the pair (X_i, η_i). We define the CV criterion by

$$\lambda(a) \;=\; \int \hat{h}_1^2(x|a)dx - 2n^{-1}\sum_{i=1}^{n} \eta_i \hat{h}_{1,-i}(X_i|a). \tag{2.9}$$

Indeed, for each i, we have that $E\left(\eta_i \hat{h}_{1,-i}(X_i|a)\middle| X_i, \eta_i\right) = \eta_i \int K_a(X_i - x)h_1(x)\,dx$; see, for example Silverman (1986). Then, by a conditioning argument, it follows that

$$\begin{aligned}
E\left(\lambda(a)\right) &= E\int \hat{h}_1^2(x|a)dx - 2n^{-1}\sum_{i=1}^{n} E\left(\eta_i \int K_a(X_i - x)h_1(x)dx\right) \\
&= E\int \hat{h}_1^2(x|a)dx - 2E\int \left(n^{-1}\sum_{i=1}^{n} \eta_i K_a(X_i - x)\right) h_1(x)dx \\
&= E\int \hat{h}_1^2(x|a)dx - 2\,E\int \hat{h}_1(x|a)h_1(x)dx \\
&= M_{h_1}(a) - R\left(h_1\right),
\end{aligned}$$

implying that the same value minimizes both $M_{h_1}(a)$ and $E\left(\lambda(a)\right)$. This suggests that it is reasonable to choose the minimizer of $\lambda(a)$, which gives the pilot BW g.

2.3 Survival function estimators

Let a, b, and c denote three BW sequences. The estimators of the survival function of T arise from two representations for $\Lambda(t)$, the cumulative hazard of T. Denote the distribution of X by $H(x)$, its survival function by $\bar{H}(x)$, and let $p(x) = P(\delta = 1|X = x)$. Under MAR, $p(x) = p_1(x)/\pi_1(x)$, where $p_1(x)$ and $\pi_1(x)$ are the derivatives with respect to x of

$P(X \leq x, \delta = 1)$ and $P(X \leq x, \xi = 1)$ respectively. The first representation is due to Dikta (1998) and takes the form

$$\Lambda(t) \quad = \quad \int_0^t p(x) \frac{dH(x)}{\bar{H}(x-)}. \tag{2.10}$$

Assuming MAR, the standard kernel-based estimator of $p(x)$ is given by

$$\hat{p}(x) \quad = \quad \frac{\sum_{i=1}^n K_a(x - X_i)\sigma_i}{\sum_{i=1}^n K_b(x - X_i)\xi_i} \quad \equiv \quad \frac{\hat{p}_1(x)}{\hat{\pi}_1(x)}. \tag{2.11}$$

The estimator $\hat{\Lambda}_D(t)$ is obtained by plugging into Eq. (2.10) the estimator $\hat{p}(x)$ and the empirical estimators of H and $\bar{H}$. The standard product integral of $\hat{\Lambda}_D(t)$, provides a Dikta-type estimator of the survival function $S(t)$, see Subramanian (2004) for details.

Let $H_{11}(x) = P(X \leq x, \xi = 1, \sigma = 1)$. The second representation results from an "inverse probability of nonmissingness weighted" (IPW) form, assuming MAR:

$$\Lambda(t) \quad = \quad \int_0^t \frac{1}{\pi(x)} \frac{dH_{11}(x)}{\bar{H}(x-)}, \tag{2.12}$$

where $\pi(x) = P(\xi = 1 | X = x)$. Under MAR, $\pi(x) = \pi_1(x)/h(x)$. The standard kernel-based estimator of $\pi(x)$ is given by

$$\hat{\pi}(x) \quad = \quad \frac{\sum_{i=1}^n K_b(x - X_i)\xi_i}{\sum_{i=1}^n K_c(x - X_i)} \quad \equiv \quad \frac{\hat{\pi}_1(x)}{\hat{h}(x)}. \tag{2.13}$$

Again, $\hat{\Lambda}_I(t)$ is the plug-in estimator obtained using $\hat{\pi}(x)$, and the empirical estimators of H_{11} and $\bar{H}$. The product integral yields $\hat{S}_I(t)$, the IPW estimator of $S(t)$, see Subramanian (2006) for details.

2.4 Illustrations using real data sets

Dinse (1986) reported the survival time and disease status at death for 58 female RFM mice. At necropsy each mouse was examined for nonrenal vascular disease (NVD). Of the

33 mice that died with NVD present, 8 died due to the disease ($\xi = 1, \sigma = 1$), 19 from other known causes ($\xi = 1, \sigma = 0$), and 6 had unknown cause of death ($\xi = 0$). For illustrating the MCI model, we shall assume that deaths due to NVD occur independently of deaths due to other causes. The Epanechnikov kernel was used. For computational convenience, the observed failure times were divided by 1100 and the minimization was performed over a fine grid of 500 BW values between 0.001 and 0.5. Numerical integrations were performed over a grid of 200 points between 0 and 1. The number of bootstrap samples was 500. The BWs in days are given in Table 1 below.

Table 1. BWs (in days) for the mice data.

Parameter	*CV*	*TSB*	*EISB*
h	55.0	72.6	84.7
π_1	49.5	63.8	79.2
p_1	22.0	35.2	53.9

The estimated survival curves using the two-step EISB BWs are shown in Figure 1.

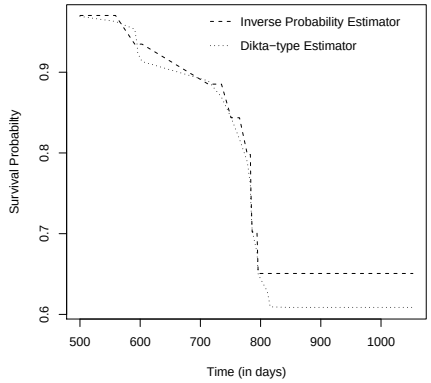

Figure 1: Comparison of estimators for the mice data set

We next extracted data from the New Mexico registry on Nonhodgkins Lymphoma pa-

tients. Out of 47 patients who were first diagnosed with the disease at age 43, 25 had unknown cause of death (MCI), while 9 died due to other causes and hence censored. Five different observations had ties, which were resolved by adding 0.001 to each of the repeats. For computational convenience, the observed failure times were divided by the maximum (294), and the minimization was performed over a fine grid of 500 BW values between 0.001 and 0.5. Numerical integrations were performed over a grid of 200 points between 0 and 1.1. The number of bootstrap samples was 500. The BWs in months are given in Table 2 below.

Table 2. BWs (in months) for Nonhodgkins Lymphoma data from Seer database.

Parameter	*CV*	*TSB*	*EISB*
h	17.35	18.52	22.64
π_1	12.64	14.70	17.05
p_1	11.47	13.23	17.35

The estimated survival curves using the two-step EISB BWs are shown in Figure 2.

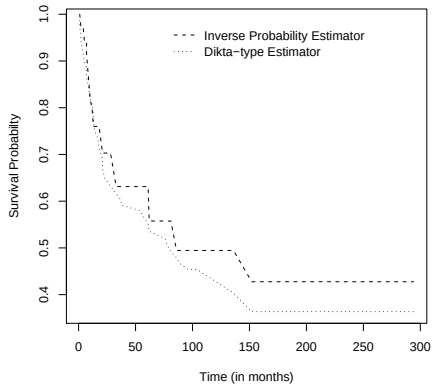

Figure 2: Comparison of estimators for the Nonhodgkins Lymphoma data set extracted from the Seer database.

For our final illustration, we employ the Primary Biliary Cirrhosis (PBC) right censored data set found in Fleming and Harrington (1991). We used the version available at the link

http://lib.stat.cmu.edu/datasets/pbc. The data are from the Mayo Clinic trial in PBC of the liver conducted between 1974 and 1984. Right censored survival data are available for 418 PBC patients. The observed time is the number of days between registration and the earliest of death, transplantation, or study analysis time in July, 1986, the latter two consituting censoring. We considered the 374 female patients for our analysis and imputed missingness by taking $\pi(t) = e^t/(1+e^t)$. The observed proportion of missingness was 44%. The observed times were divided by 4800, to restrict them between 0 and 1, and the minimization was performed over a fine grid of 500 BW values between 0.001 and 0.5. Numerical integrations were performed over a grid of 200 points between 0 and 1. The number of bootstrap samples was 500. The BWs in days are given in Table 3 below.

Table 3. BWs (in days) for the PBC data.

Parameter	*CV*	*TSB*	*EISB*
h	451.2	537.6	734.4
π_1	312.0	537.6	873.6
p_1	316.8	662.4	916.8

Due to the fact that missingness was artificially introduced, the benchmark in this case, namely the Kaplan–Meier estimator, can be computed. The IPW and Dikta-type estimated survival curves were calculated using the two-step EISB BWs. These curves are shown in Figure 3, along with the Kaplan–Meier curve.

3 Numerical results

We performed three different simulation studies to investigate the operating characteristics of the CV, EISB, and TSB choice of BWs. According to Faraway and Jhun, "a good

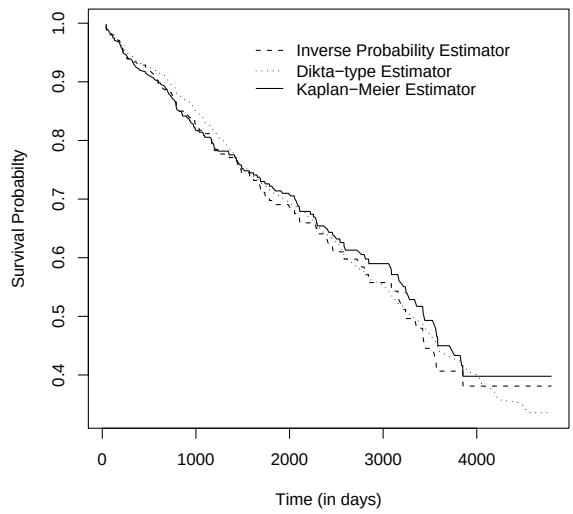

Figure 3: Comparison of estimators for the Primary Biliary Cirrhosis data set. Missingness imputed using $\pi(t) = e^t/(1+e^t)$.

benchmark with which to measure the performance" of the above three BW estimators is the ISE choice, which provides the best possible BW given complete knowledge of the underlying subdensity. We have not provided results for the SCV choice of BW. Recall that, unlike the TSB choice, the SCV choice of BW ignores an asymptotically negligible term during the minimization. For all the simulations, we employed the Epanechnikov kernel, the sample size was 100, the number of bootstrap samples was 200, and the number of Monte Carlo replications was 1000.

Simulation 1 The minimum X was taken to be uniformly distributed on $(0,1)$. 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 $1-r = P(\delta=1) = \int_0^1 p(t)dt = 11c/12$, so that c equals $12(1-r)/11$. We only consider values of r greater than $1/12$, namely $0.1, 0.2, 0.3$

and 0.4, giving censoring rates (CRs) of 10%, 20%, 30%, and 40% respectively. With these choices $S(t) = \exp(-\Lambda(t))$, where $\Lambda(t) = c(-0.75\log(1-t)+t+t^2/2)$. The conditional probability of $\xi = 1$ given X, denoted by $\pi(t)$, is taken to be $e^t/(1+e^t)$, giving a missingness rate (MR) of about 38%. The numerical integrations were performed over a grid of 100 points. Minimizations were over BWs taken in a fine grid of values of the unit interval. Table 4 gives the mean and standard deviation of the four different BW choices for computing the kernel estimator of the subdensity $p_1(x)$. The two-step EISB choice of BW performed better than its one-step counterpart and is closest to the ISE choice mean BW.

Table 4. Simulation 1. Mean and standard deviation (SD) of BW for subdensity p_1.

	ISE BW		*CV BW*		*TSB BW*		*EISB BW*	
CR	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
10%	0.335	0.149	0.128	0.062	0.182	0.071	0.296	0.106
20%	0.349	0.158	0.122	0.060	0.181	0.076	0.304	0.112
30%	0.368	0.165	0.116	0.059	0.184	0.082	0.308	0.121
40%	0.390	0.175	0.112	0.060	0.189	0.090	0.320	0.131

In Table 5, we provide the mean ratio of the estimated ISE of a method to the ISE minimized over all choices of BW. The efficiency of the one-step EISB choice was considerably worse than the two-step counterpart, which performs the best among the choices considered.

Table 5. Simulation 1. Relative efficiencies of CV, TSB and EISB BWs for subdensity p_1.

	CV		*TSB*		*EISB*	
CR	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
10%	3.45	4.01	2.25	2.23	1.68	1.47
20%	4.03	5.26	2.52	2.92	1.78	1.67
30%	4.84	6.11	2.86	3.17	1.93	1.71
40%	6.12	9.05	3.56	5.19	2.35	3.35

In Table 6, we give the MISE and standard deviation for the IPW and Dikta-type survival function estimators, computed using the two-step EISB choice of BW. The Dikta-type

estimator performs marginally better than the IPW estimator.

Table 6. Simulation 1. Comparison of IPW and Dikta-type survival function estimators using the two-step EISB choice of BW for various CRs.

CR	*IPW*		*Dikta-type*	
	MISE	*SD*	*MISE*	*SD*
10%	0.0234	0.0111	0.0227	0.0106
20%	0.0235	0.0118	0.0230	0.0115
30%	0.0232	0.0119	0.0229	0.0120
40%	0.0225	0.0126	0.0225	0.0128

Simulation 2 The failure time distribution was Weibull of the type $F(x) = 1-\exp(-(\alpha x)^2)$ (cf. Zhu, Yuen, and Tang 2002; Dikta, Kvesic, and Schmidt, 2006). The censoring distribution was exponential with mean 1. Then $p(x) = 2\alpha x/(1+2\alpha x)$. The conditional probability of nonmissingness was taken as the two parameter logit model: $\pi(x) = \exp(\theta_1 + \theta_2 x)/(1 + \exp(\theta_1+\theta_2 x))$. The parameters α and θ_1, θ_2 were chosen to get the following combinations of CR and MR: (10%, 44%), (20%, 35%), (25%, 19%), and (42%, 15%). From Tables 7 and 8, it may be noted that the two-step EISB choice of BW performs best in the sense of nearness to the ISE choice as well as in terms of relative efficiency.

Table 7. Simulation 2. Mean and standard deviation (SD) of BW for π_1 and p_1.

			ISE BW		*CV BW*		*TSB BW*		*EISB BW*	
Parameter	*CR*	*MR*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
π_1	10%	44%	0.063	0.016	0.035	0.012	0.052	0.012	0.064	0.016
	20%	35%	0.124	0.028	0.068	0.026	0.092	0.028	0.120	0.029
	25%	19%	0.152	0.033	0.100	0.039	0.118	0.036	0.150	0.033
	42%	15%	0.263	0.066	0.171	0.070	0.197	0.063	0.261	0.064
p_1	10%	44%	0.056	0.014	0.034	0.011	0.052	0.012	0.065	0.017
	20%	35%	0.114	0.026	0.065	0.026	0.094	0.029	0.123	0.032
	25%	19%	0.198	0.043	0.090	0.038	0.121	0.038	0.154	0.039
	42%	15%	0.303	0.074	0.162	0.075	0.226	0.070	0.299	0.083

Table 8. Simulation 2. Relative efficiencies of CV, TSB and EISB BWs for π_1 and p_1.

			CV		*TSB*		*EISB*	
Parameter	*CR*	*MR*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
π_1	10%	44%	3.42	4.25	2.13	3.38	1.99	3.23
	20%	35%	5.16	10.35	3.12	6.35	2.35	5.74
	25%	19%	4.17	8.55	2.80	4.82	2.03	3.73
	42%	15%	3.13	5.05	2.32	2.97	1.76	3.02
p_1	10%	44%	1.55	1.05	1.28	0.86	1.32	0.84
	20%	35%	2.73	3.42	1.72	1.73	1.54	1.50
	25%	19%	3.76	5.00	2.42	3.21	1.88	2.76
	42%	15%	4.24	7.18	2.33	3.37	1.87	3.45

In Table 9, as in Table 6, we provide the MISE and standard deviation for the IPW and Dikta-type estimators, computed using the two-step EISB choice of BW. The Dikta-type estimator performs marginally better than the IPW estimator for Simulation 2.

Table 9. Simulation 2. Comparison of IPW and Dikta-type survival function estimators using the two-step EISB choice of BW for various CRs.

CR	*MR*	*IPW*		*Dikta-type*	
		MISE	*SD*	*MISE*	*SD*
10%	44%	0.0005	0.0004	0.0004	0.0003
20%	35%	0.0010	0.0009	0.0008	0.0007
25%	19%	0.0012	0.0011	0.0011	0.0011
42%	15%	0.0029	0.0025	0.0027	0.0027

Simulation 3 The failure time distribution was taken as the Beta with parameters $\theta > 0$ and 1. The censoring distribution was again exponential with mean 1. Then $p(x) = \theta x^{\theta-1}/(1 - x^{\theta} + \theta x^{\theta-1})$. The conditional probability of nonmissingness was taken as $\pi(x) = \Phi(\gamma x)$, where $\Phi(x)$ is the standard normal cumulative distribution function. The parameters θ and γ were chosen to get the following combinations of CR and MR: (15%, 35%), (25%, 30%),

and (35%, 15%). From Tables 10 and 11, it may be noted that the two-step EISB choice of BW performs best overall, although in a couple of cases the TSB choice has better efficiency.

Table 10. Simulation 3. Mean and standard deviation (SD) of BW for π_1 and p_1.

			ISE BW		*CV BW*		*TSB BW*		*EISB BW*	
Parameter	*CR*	*MR*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
π_1	15%	35%	0.032	0.013	0.001	0.001	0.065	0.025	0.030	0.032
	25%	30%	0.067	0.021	0.022	0.018	0.043	0.025	0.060	0.030
	35%	15%	0.100	0.002	0.079	0.025	0.090	0.017	0.097	0.010
p_1	15%	35%	0.011	0.003	0.001	0.001	0.067	0.024	0.030	0.032
	25%	30%	0.024	0.011	0.018	0.016	0.042	0.026	0.055	0.031
	35%	15%	0.098	0.006	0.074	0.027	0.089	0.020	0.097	0.012

Table 11. Simulation 3. Relative efficiencies of CV, TSB and EISB BWs for π_1 and p_1.

			CV		*TSB*		*EISB*	
Parameter	*CR*	*MR*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
π_1	15%	35%	20.37	9.84	1.52	0.70	3.44	2.88
	25%	30%	4.12	5.33	1.71	1.46	1.40	1.08
	35%	15%	1.81	2.41	1.28	0.81	1.09	0.77
p_1	15%	35%	9.37	5.27	3.11	1.30	2.64	1.90
	25%	30%	1.75	1.38	1.13	0.18	1.16	0.24
	35%	15%	1.45	1.48	1.16	0.80	1.04	0.19

In Table 12, as in Tables 3 and 6, we provide the MISE and standard deviation for the IPW and Dikta-type estimators, computed using the two-step EISB choice of BW. The Dikta-type estimator performs better than the IPW estimator for Simulation 3.

Table 12. Simulation 3. Comparison of IPW and Dikta-type survival function estimators using the two-step EISB choice of BW for various CRs.

CR	*MR*	*IPW*		*Dikta-type*	
		MISE	*SD*	*MISE*	*SD*
15%	35%	0.0157	0.0290	0.0106	0.0236
25%	30%	0.0049	0.0097	0.0038	0.0086
35%	15%	0.0027	0.0032	0.0025	0.0028

4 Conclusion

Bootstrap-based BW selection procedures for computing kernel subdensity estimators were investigated in this paper. The methods mirror their counterpart for density estimation. The two-step empirical iterated smoothed bootstrap produces superior BW estimates as evidenced by several simulation studies. This "optimal" BW was plugged into the Dikta-type and IPW survival function estimators and a MISE analysis was performed. The Dikta-type estimator produces marginally better estimates compared to the IPW estimator. The methods were applied to two real data sets and one real data set with simulated MCIs.

Appendix

Bias and variance derivation For convenience we shall denote $\hat{h}_1(x|a), \hat{h}_1^*(x|a)$ etc as $\hat{h}_{1,a}(x), \hat{h}_{1,a}^*(x)$. Note that

$$\begin{aligned}\hat{h}_{1,a}^*(x) &= n^{-1}\sum_{i=1}^{n} K_a(x - X_i^*)\eta_i^* \\ &= \frac{1}{n}\sum_{i=1}^{n} K_a\left(x - \left(\tilde{X}_i + g\epsilon_i\right)\right)\tilde{\eta}_i.\end{aligned}$$

For each fixed $i = 1, \ldots, n$ we have

$$K_a\left(x - \left(\tilde{X}_i + g\epsilon_i\right)\right)\tilde{\eta}_i = \begin{cases} K_a\left(x - (X_1 + g\epsilon_i)\right)\eta_1 & \text{with probability } n^{-1} \\ K_a\left(x - (X_2 + g\epsilon_i)\right)\eta_2 & \text{with probability } n^{-1} \\ \ldots \\ K_a\left(x - (X_n + g\epsilon_i)\right)\eta_n & \text{with probability } n^{-1}. \end{cases}$$

Therefore,

$$
\begin{aligned}
E^*(\hat{h}^*_{1,a}(x)) &= \frac{1}{n}\sum_{i=1}^{n} E\left[K_a\left(x-\left(\tilde{X}_i+g\epsilon_i\right)\right)\tilde{\eta}_i\middle|(X_1,\eta_1),\ldots,(X_n,\eta_n)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} E\left[K_a\left(x-(X_j+g\epsilon_i)\right)\eta_j\middle|(X_1,\eta_1)\ldots,(X_n,\eta_n)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\eta_j\int K_a(x-y)K_g(y-X_j)dy \\
&= \frac{1}{n}\sum_{i=1}^{n}\int K_a(x-y)\left\{\frac{1}{n}\sum_{j=1}^{n}K_g(y-X_j)\eta_j\right\}dy \\
&= \frac{1}{n}\sum_{i=1}^{n}\int K_a(x-y)\hat{h}_{1,g}(y)dy \\
&= \int K_a(x-y)\hat{h}_{1,g}(y)dy \;:=\; \left(K_a*\hat{h}_{1,g}\right)(x).
\end{aligned}
$$

Therefore the bias is given by

$$
B^*(x) \;=\; \left(K_a*\hat{h}_{1,g}\right)(x)-\hat{h}_{1,g}(x). \tag{A.1}
$$

Next we have

$$
\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}K_a\left(x-\left(\tilde{X}_i+g\epsilon_i\right)\right)\tilde{\eta}_i\right)=\begin{cases}\mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}K_a\left(x-(X_1+g\epsilon_i)\right)\eta_1\right) & \text{w.p. } n^{-1}\\ \mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}K_a\left(x-(X_2+g\epsilon_i)\right)\eta_2\right) & \text{w.p. } n^{-1}\\ \ldots & \\ \mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}K_a\left(x-(X_n+g\epsilon_i)\right)\eta_n\right) & \text{w.p. } n^{-1}.\end{cases}
$$

Therefore

$$
\begin{aligned}
\mathrm{Var}^*(\hat{h}^*_{1,a}(x)) &= \frac{1}{n}\sum_{j=1}^{n}\mathrm{Var}\left(n^{-1}\sum_{i=1}^{n}K_a\left(x-(X_j+g\epsilon_i)\right)\eta_j\right) \\
&= \frac{1}{n^3}\sum_{j=1}^{n}\mathrm{Var}\left(\sum_{i=1}^{n}K_a\left(x-(X_j+g\epsilon_i)\right)\eta_j\right).
\end{aligned}
$$

Note that

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^{n} K_a\left(x-(X_j+g\epsilon_i)\right)\eta_j\right) &= E\left(\left(\sum_{i=1}^{n} K_a\left(x-(X_j+g\epsilon_i)\right)\eta_j\right)^2\right) \\
&\quad -\left(E\left(\sum_{i=1}^{n} K_a\left(x-(X_j+g\epsilon_i)\right)\eta_j\right)\right)^2 \\
&= n\left(K_a^2 * \hat{h}_{1,g}\right)(x) + n(n-1)\left(K_a * \hat{h}_{1,g}\right)^2(x) \\
&\quad - n^2\left(K_a * \hat{h}_{1,g}\right)^2(x) \\
&= n\left(\left(K_a^2 * \hat{h}_{1,g}\right)(x) - \left(K_a * \hat{h}_{1,g}\right)^2(x)\right).
\end{aligned}$$

Therefore

$$\mathrm{Var}^*(\hat{h}^*_{1,a}(x)) = \frac{1}{n}\left(\left(K_a^2 * \hat{h}_{1,g}\right)(x) - \left(K_a * \hat{h}_{1,g}\right)^2(x)\right). \tag{A.2}$$

We now obtain

$$\begin{aligned}
\hat{M}_{h_1}(a)) &= \frac{1}{n}\int\left\{\left(K_a^2 * \hat{h}_{1,g}\right)(x) - \left(K_a * \hat{h}_{1,g})^2(x)\right)\right\}dx \\
&\quad + \int\left\{(K_a * \hat{h}_{1,g})(x) - \hat{h}_{1,g}(x)\right\}^2 dx \\
&= \frac{1}{n}\int\left(K_a^2 * \hat{h}_{1,g}\right)(x)dx + \left(1-\frac{1}{n}\right)\int(K_a * \hat{h}_{1,g})^2(x)dx \\
&\quad - 2\int(K_a * \hat{h}_{1,g})(x)\hat{h}_{1,g}(x)dx + \int \hat{h}_{1,g}(x)^2 dx.
\end{aligned}$$

The first term on the right side of the above equation becomes

$$\begin{aligned}
\frac{1}{n}\int\left(K_a^2 * \hat{h}_{1,g}\right)(x)dx &= \frac{1}{n}\int_{-\infty}^{\infty}\left\{\int_{x-aM}^{x+aM} K_a^2(x-y)\hat{h}_{1,g}(y)dy\right\}dx \\
&= \frac{1}{na}\int_{-\infty}^{\infty}\left\{\int_{-M}^{M} K^2(u)\hat{h}_{1,g}(x-hu)du\right\}dx \\
&= \frac{1}{na}\int_{-M}^{M} K^2(u)\left\{\int_{-\infty}^{\infty}\hat{h}_{1,g}(x-hu)dx\right\}du \\
&= \frac{1}{na}\int_{-M}^{M} K^2(u)\left\{\int_{-\infty}^{\infty}\hat{h}_{1,g}(z)dz\right\}du \\
&= \frac{1}{na}\int_{-M}^{M} K^2(u)\left\{\frac{1}{ng}\sum_{i=1}^{n}\eta_i\int_{-\infty}^{\infty}K\left(\frac{z-X_i}{g}\right)dz\right\}du \\
&= \frac{1}{na}\left(\frac{1}{n}\sum_{i=1}^{n}\eta_i\right)\int_{-M}^{M}K^2(u)du.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\hat{M}_{h_1}(a) &= \frac{1}{na}\left(\frac{1}{n}\sum_{i=1}^{n}\eta_i\right)\int_{-M}^{M}K^2(u)du + \left(1-\frac{1}{n}\right)\int(K_a * \hat{h}_{1,g})^2(x)dx \\
&\quad - 2\int(K_a * \hat{h}_{1,g})(x)\hat{h}_{1,g}(x)dx + \int\hat{h}_{1,g}(x)^2dx. \qquad \text{(A.3)}
\end{aligned}$$

References

Bowman, A.W., 1984. An alternative method of cross-validation for the smoothing of density estimates. Biometrika 71, 353–360.

Cao, R., 1993. Bootstrapping the mean integrated squared error. Journal of Multivariate Analysis 45, 137–160.

Cao, R., Cuevas, A., Manteiga, W.G., 1994. A comparative study of several smoothing methods in density estimation. Computational Statistics and Data Analysis 17, 153–176.

Dikta, G., 1998. On semiparametric random censorship models. Journal of Statistical Planning and Inference 66, 253–279.

Dinse, G., 1986. Nonparametric prevalence and mortality estimators for animal experiments with incomplete cause-of-death data. Journal of the American Statistical Association 81, 328–336.

Efron, B., 1982. The Jackknife, the Bootstrap and Other Resampling Plans. SIAM, Philadelphia.

Eubank, R.L., 1988. Spline Smoothing and Nonparametric Regression. Dekker, New York.

Faraway, J.J., Jhun, M., 1990. Bootstrap choice of bandwidth for density estimation. Journal of the American Statistical Association 85, 1119–1122.

Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis, Wiley.

Grund, B., Polzehl, J., 1997. Bias corrected bootstrap bandwidth selection. Journal of Nonparametric Statistics 7, 797–817.

Hall, P., DiCiccio, T., Romano, J., 1989. On smoothing and the bootstrap. Annals of Statistics 17, 692–704.

Hall, P., Marron, J.S., 1987. Estimation of integrated squared density derivatives. Statistics and Probability Letters 6, 109–115.

Hall, P., Marron, J.S., 1991. Lower bounds for bandwidth selection in density estimation. Probability Theory and Related Fields 90, 149–173.

Hall, P., Marron, J.S., Park, B.U., 1992. Smoothed cross-validation. Probability Theory and Related Fields 90, 1–20.

Hall, P., Sheather, S.S., Jones, M.C., Marron, J.S., 1991. On optimal data-based bandwidth selection for kernel density estimation. Biometrika 78, 263–269.

Härdle, W., 1989. Applied Nonparametric Regression. Econometrics Society Monograph Series, No. 19, Cambridge University Press, Cambridge, 1989.

Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47, 663–685.

Jones, M.C., Sheather, S.J., 1991. Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. Statistics and Probability Letters 11, 511–514.

Jones, M.C., Marron, J.S., Sheather, S.J., 1996. A brief survey of bandwidth selection for density estimation. Journal of the American Statistical Association 91, 401–407.

Koul, H., Susarla, V., Van Ryzin, J., 1981. Regression analysis with randomly right censored data. Annals of Statistics 9, 1276–1288.

Müller, H.-G., 1985. Empirical bandwidth choice for nonparametric kernel regression by means of pilot estimators. Statistics and Decisions. Supplement no. 2, 193–206.

Müller, H.-G., 1988. Nonparametric Analysis of Longitudinal Data. Springer, New York.

Park, B.U., Marron, J.S., 1990. Comparison of data-driven bandwidth selectors. Journal of

the American Statistical Association 85, 66–72.

Robins, J.M., and Rotnitzky, A., 1992. Recovery of information and adjustment for dependent censoring using surrogate markers. in AIDS Epidemiology - Methodological Issues, eds. N. Jewell, K. Dietz, and V. Farewell, Boston: Birkhauser, pp. 297-331.

Rudemo, M., 1982. Empirical choice of histograms and kernel density estimators. Scandinavian Journal of Statistics 9, 65–78.

Silverman, B.W., 1986. Density Estimation. Chapman and Hall, London, New York.

Silverman, B.W., Young, G.A., 1987. The bootstrap: To smooth or not to smooth? Biometrika 74, 469–479.

Staniswallis, J.G., 1989. Local bandwidth selection for kernel estimates. Journal of the American Statistical Association 84, 284–288.

Subramanian, S., 2004. Asymptotically efficient estimation of a survival function in the missing censoring indicator model. Journal of Nonparametric Statistics 16, 797–817.

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

Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman and Hall/CRC, London, New York.