T E C H N I C A L
R E P O R T

0345

PRESMOOTHED KAPLAN-MEIER AND NELSON-AALEN ESTIMATORS

Ricardo CAO, Ignacio LOPEZ-DE-ULLIBARRI,
Paul JANSSEN and Noël VERAVERBEKE

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Presmoothed Kaplan-Meier and Nelson-Aalen Estimators

Ricardo Cao and Ignacio López-de-Ullibarri
Departamento de Matemáticas,
Universidade da Coruña, A Coruña 15071, Spain

Paul Janssen and Noël Veraverbeke

Limburgs Universitair Centrum,
Universitaire Campus B-3590 Diepenbeek, Belgium

July 7, 2003

Abstract

In this paper a modification of the Kaplan-Meier and Nelson-Aalen estimators in the right random censorship model is studied. The new estimators are obtained by replacing the censoring indicator variables in the classical definitions by values of a nonparametric regression estimator. Asymptotic normality is obtained and it is shown that this presmoothing idea leads to a gain in asymptotic mean squared error. A local plug-in bandwidth selector is introduced and the problem of optimal pilot bandwidth selection for this estimator is studied. The gain of the presmoothed estimators with automatic plug-in bandwidth selector is demonstrated in a simulation study.

Abbreviated Title: Presmoothed Kaplan-Meier estimator.
AMS 1991 Subject classifications. Primary: 62G07; secondary: 60F05, 62G20.
Key words and phrases: Bandwidth selection, censored data, kernel smoothing, plug-in bandwidth.

1 Presmoothed estimators for censored data

Let $Y_1, \ldots, Y_n$ be independent, identically distributed (iid) positive random variables (survival times or failure times) with unknown continuous distribution function (df) $F(\cdot)$. In the right random censorship model these survival times are censored to the right by positive iid random variables $C_1, \ldots, C_n$ with unknown continuous df $G(\cdot)$. For each $i = 1, \ldots, n$ we observe (T_i, δ_i) where $T_i = Y_i \wedge C_i$

and $\delta_i = \mathbf{1}\{Y_i \le C_i\}$. We assume the Y_i's independent of the C_i's. Therefore the df $H(\cdot)$ of T_i satisfies

$$1 - H(t) = (1 - F(t))(1 - G(t)).$$

The variable δ_i indicates whether Y_i is censored ($\delta_i = 0$) or uncensored ($\delta_i = 1$). The success probability for this binary outcome is

$$\gamma = P(\delta_1 = 1) = E(\delta_1) = P(Y_1 \le C_1) = \int_0^\infty (1 - G(t))dF(t) = H^u(+\infty)$$

where $H^u(t) = P(T_1 \le t, \delta_1 = 1)$ is the subdistribution function of the uncensored observations. Note that

$$H^u(t) = \int_0^t p(s)dH(s)$$

where

$$p(t) = P(\delta_1 = 1 \mid T_1 = t) = E(\delta_1 \mid T_1 = t).$$

The function $p(\cdot)$ is the conditional probability that the observation is non-censored given that $T_1 = t$. The importance of the function $p(\cdot)$ is clear from the following relations (see also Dikta (1998)):

$$\Lambda_F(t) = \int_0^t \frac{1}{1 - H(s^-)} dH^u(s) = \int_0^t p(s)d\Lambda_H(s) \tag{1}$$

with $\Lambda_F(\cdot)$ and $\Lambda_H(\cdot)$ the cumulative hazard functions corresponding to $F(\cdot)$ and $H(\cdot)$.

From (1) we easily obtain

$$1 - F(t) = \exp(-\Lambda_F(t)) = \exp\left(-\int_0^t p(s)d\Lambda_H(s)\right) \tag{2}$$

and

$$\lambda_F(t) = p(t)\lambda_H(t) \tag{3}$$

with $\lambda_F(\cdot)$ and $\lambda_H(\cdot)$ the hazard functions.

Note that $p(t) \equiv 1$ in the case of no censoring. If δ_1 is independent of T_1, i.e.,

$$p(t) = E(\delta_1 = 1 \mid T_1 = t) = E(\delta_1) = \gamma$$

then (1) implies that we have the Koziol-Green proportional hazards model (see Koziol and Green (1976)):

$$1 - F(t) = \exp(-\Lambda_F(t)) = \exp(-\gamma\Lambda_H(t)) = (1 - H(t))^\gamma$$

or equivalently $1 - G(t) = (1 - F(t))^\beta$ with $\beta = (1-\gamma)/\gamma$.

The classical estimator for $\Lambda_F(t)$ is the Nelson-Aalen estimator (see Nelson (1972) and Aalen (1978))

$$\Lambda_n^{NA}(t) = \sum_{T_{(i)} \le t} \frac{\delta_{[i]}}{n - i + 1} \tag{4}$$

where $T_{(1)} \le \ldots \le T_{(n)}$ are the ordered T_i's and the $\delta_{[i]}$'s are the concomitants. The intuitive idea behind the Nelson-Aalen estimator is to consider the purely empirical version of $\Lambda_F(t) = \int_0^t \frac{p(s)}{1 - H(s^-)} dH(s)$, i.e., at the i-th ordered jump (of size n^{-1}) of $H_n(s) = n^{-1} \sum_{i=1}^n \mathbf{1}\{T_i \le s\}$ we estimate $\frac{p(s)}{1 - H(s^-)}$ by the purely empirical value $\frac{\delta_{[i]}}{1 - (i-1)/n} = \frac{n\delta_{[i]}}{n - i + 1}$. It is however clear that in this expression we can replace $\delta_{[i]}$ by a more smooth (parametric or nonparametric) estimator of $p(\cdot)$. An appealing intuitive idea is to estimate $p(T_{(i)})$ by $p_n(T_{(i)})$ where

$$p_n(t) = \frac{(nb)^{-1} \sum_{i=1}^n K\left(\frac{t - T_i}{b}\right) \delta_i}{(nb)^{-1} \sum_{i=1}^n K\left(\frac{t - T_i}{b}\right)} = \frac{n^{-1} \sum_{i=1}^n K_b(t - T_i)\delta_i}{n^{-1} \sum_{i=1}^n K_b(t - T_i)}$$

with $K(\cdot)$ a kernel, $K_b(u) = \frac{1}{b} K(u/b)$ and $b \equiv b_n$, $n = 1, 2, \ldots$, a bandwidth sequence ($p_n(\cdot)$ is the Nadaraya-Watson kernel estimator for $p(\cdot)$ based on the binary responses δ_i with covariates T_i, $i = 1, \ldots, n$). This yields the presmoothed estimator

$$\Lambda_n^P(t) = \sum_{T_{(i)} \le t} \frac{p_n(T_{(i)})}{n - i + 1}. \tag{5}$$

It should be noted that the parametric version of this idea appeared in Dikta (1998). He proposed working with a parametric estimator for $p(\cdot)$. In the unpublished Diploma Thesis by Ziegler (1995) the asymptotic distribution of some modified presmoothed Kaplan-Meier process has been established and used to construct confidence bands for the underlying distribution function.

It is immediate from (3) that $\lambda_F(t)$ can be estimated by means of a presmoothed hazard function estimator

$$\lambda_n^P(t) = p_n(t)\lambda_n(t) \tag{6}$$

where $\lambda_n(t)$ is an estimator of $\lambda_H(t)$ (e.g. the Watson and Leadbetter (1964a,b) kernel estimator). Note that the estimator like the one in (6) is simply the product of two estimators based on the iid observations at hand. Indeed $p_n(t)$

is based on (T_i, δ_i), $i = 1, \ldots, n$ and $\lambda_n(t)$ is based on T_i, $i = 1, \ldots, n$. Any possible nonparametric estimators for $p(t)$ and for $\lambda_H(t)$ can be used to make a product estimator like in (6). The asymptotic properties of such estimators are easily derived from those of the two factors, see the forthcoming Ph.D. Dissertation by López-de-Ullibarri for further details.

Some good features of the presmoothed Nelson-Aalen estimator, defined by (5), are:

(i) $\Lambda_n^P(\cdot)$ has a jump at any of the observations, so that from a graphical point of view $\Lambda_n^P(\cdot)$ provides more information on the local behaviour than the classical Nelson-Aalen estimator.

(ii) Using a binary regression smoother to estimate $p(\cdot)$ means that we can extrapolate the available information to better describe the tail behaviour. This feature should be clear from the graphical performance of this estimator.

(iii) The presmoothed Nelson-Aalen estimator has a smaller asymptotic variance than the Nelson-Aalen estimator. In this paper it will be shown that this results in a better mean squared error performance, showing that presmoothing is beneficial.

Using the first equation in (2), expressions (4) and (5) and the approximation $e^{-x} \simeq 1 - x$ for x close to 0, we easily obtain the following two estimators for $1 - F(t) = \exp(-\Lambda_F(t))$, the survival function at t:

$$1 - F_n^{KM}(t) = \prod_{T_{(i)} \leq t} \left(1 - \frac{\delta_{[i]}}{n - i + 1}\right) \tag{7}$$

and

$$1 - F_n^P(t) = \prod_{T_{(i)} \leq t} \left(1 - \frac{p_n(T_{(i)})}{n - i + 1}\right). \tag{8}$$

The estimator in (7) is the classical Kaplan-Meier estimator (see Kaplan and Meier (1958)) while (8) gives the new presmoothed estimator proposed in this paper. It is straightforward to show that, for t such that $H(t) < 1$,

$$1 - F_n^{KM}(t) = \exp(-\Lambda_n^{NA}(t)) + O_P(n^{-1})$$

and also that

$$1 - F_n^P(t) = \exp(-\Lambda_n^P(t)) + O_P(n^{-1}).$$

In this paper the following items on the presmoothed Nelson-Aalen estimator and the presmoothed Kaplan-Meier estimator will be covered. In Section 2 we give representations for $\Lambda_n^P(t) - \Lambda_F(t)$ and for $F_n^P(t) - F(t)$ and, as an application, we study their asymptotic distributional behaviour. The beneficial effect of presmoothing is demonstrated in Section 3. There we look at the asymptotic mean squared error of the dominant part of the presmoothed Nelson-Aalen estimator and the presmoothed Kaplan-Meier estimator, and we show that these $AMSE$'s are smaller than the corresponding expressions for the Nelson-Aalen and the Kaplan-Meier estimators. Section 4 includes some proposal of a plug-in bandwidth selector in this setup. The asymptotic results presented there for this selector show the rate of convergence of this data driven bandwidth to its populational counterpart. A small simulation study is included, in Section 5, to show the considerable gain that may be attained, in terms of efficiency, by the presmoothed estimator with plug-in bandwidth. Finally, Section 6 contains the auxiliary lemmas and some of the proofs.

2 Asymptotic representations

Our results will require the following conditions.

On the kernel function, K:

(K.1) K is a non negative, symmetric, twice differentiable function of bounded variation, with bounded second derivative. It also satisfies $\int_{-L}^{L} K(x)dx = 1$, K has support in the interval $[-L, L]$, for some $L > 0$ and $K(L) = K'(L) = K''(L) = 0$.

On the conditional probability of uncensoring, p:

(P.1) p is five times differentiable in $[0, \infty)$, with continuous fifth derivative.

(P.2) $p(0) = 1$ and $\varepsilon = \sup\{t : p(x) = 1, \forall x \in [0, t)\} > 0$.

On the distribution function H :

(H.1) There exists some t_0 such that $\varepsilon < t_0$ and $H(t_0) < 1$, H is five times differentiable in $[0, t_0]$, with fifth continuous derivative and there exists some $\delta > 0$ such that $H'(t) = h(t) > \delta$, $\forall t \in [\varepsilon/2, t_0]$.

Conditions (K.1), (P.1) and (H.1) are standard regularity conditions. The degree of differentiability in (P.1) and (H.1) could be relaxed for the asymptotic representations in this section. However, this is not the case for the rates of the plug-in bandwidth that will be presented in Section 4. Condition (P.2) is a technical one and may look rather surprising. It essentially states that a lifetime cannot be censored by an arbitrary small number. There should exist some positive lower bound for censoring times. This does not seem to be a restrictive condition for real data applications.

2.1 Presmoothed Nelson-Aalen estimator

With $H_n(s) = n^{-1} \sum_{i=1}^{n} \mathbf{1}(T_i \leq s)$, let $\widehat{\Lambda}_H(t)$ be the empirical estimator of $\Lambda_H(t)$:

$$\widehat{\Lambda}_H(t) = \int_0^t \frac{dH_n(s)}{1 - H_n(s-)}.$$

We then have that

$$\begin{aligned}\Lambda_n^P(t) - \Lambda_F(t) &= \int_0^t p(s) d\left(\widehat{\Lambda}_H(s) - \Lambda_H(s)\right) \\ &+ \int_0^t (p_n(s) - p(s)) d\Lambda_H(s) \\ &+ \int_0^t (p_n(s) - p(s)) d\left(\widehat{\Lambda}_H(s) - \Lambda_H(s)\right) \\ &:= (I) + (II) + (III).\end{aligned}$$

Theorem 1 Assume (K.1), (H.1) and $b = c_0 n^{-\alpha} + o(n^{-\alpha})$ for some $1/4 < \alpha < 1/2$ and some $c_0 > 0$. Then

$$\Lambda_n^P(t) - \Lambda_F(t) = \overline{\Lambda_n^P}(t) - \Lambda_F(t) + o_P(n^{-1/2}) \tag{9}$$

with

$$\overline{\Lambda_n^P}(t) = \Lambda_F(t) + \frac{1}{n}\sum_{i=1}^n (g_1(T_i) - g_2(T_i) + g_3(T_i, \delta_i)) \tag{10}$$

where

$$g_1(T_i) = \frac{p(t)}{1 - H(t)}(\mathbf{1}(T_i \le t) - H(t)) \tag{11}$$

$$g_2(T_i) = \int_0^t \frac{\mathbf{1}(T_i \le s) - H(s)}{1 - H(s)} p'(s) ds \tag{12}$$

$$g_3(T_i, \delta_i) = \int_0^t \frac{K_b(s - T_i)(\delta_i - p(s))}{1 - H(s)} ds. \tag{13}$$

The asymptotic normality of $\Lambda_n^P(t) - \Lambda_F(t)$ is an easy consequence of the asymptotic representation in Theorem 1. For an explicit result we need a closed formula for the asymptotic variance. Based on moment calculations, collected in Lemma 2, we obtain in Theorem 3, a nice expression for the asymptotic variance. This expression clearly shows the role played by presmoothing. We need the following notations:

$$d_K = \int_{-L}^{L} v^2 K(v) dv \qquad \mathsf{K}(v) = \int_{-L}^{v} K(s) ds \qquad e_K = \int_{-L}^{L} v \mathsf{K}(v) K(v) dv \tag{14}$$

$$\begin{aligned}\alpha(t) &= \int_0^t \frac{\frac{1}{2}p''(s)h(s) + p'(s)h'(s)}{1 - H(s)} ds \\ &= \frac{1}{2}\left[\frac{p'(t)h(t)}{1 - H(t)} + \int_0^t p'(s)\left\{\frac{h'(s)}{1 - H(s)} - \frac{h(s)^2}{(1 - H(s))^2}\right\}\right] ds\end{aligned} \tag{15}$$

$$q(t) = \frac{p(t)(1 - p(t))h(t)}{(1 - H(t))^2}. \tag{16}$$

Lemma 2 Assume (K.1) and (C.1). Then,

$$E[g_3(T_1,\delta_1)] = d_K\alpha(t)b^2 + o(b^2) \tag{17}$$

$$\mathrm{Var}[g_1(T_1)] = \frac{p(t)^2 H(t)}{1-H(t)} \tag{18}$$

$$\mathrm{Var}[g_2(T_1)] = 2\int_0^t (p(t)-p(s))\frac{H(s)}{1-H(s)}p'(s)ds \tag{19}$$

$$\mathrm{Var}[g_3(T_1,\delta_1)] = \int_0^t q(v)dv - 2bq(t)e_K + O(b^2) \tag{20}$$

$$E[g_1(T_1)g_2(T_1)] = p(t)\int_0^t \frac{H(s)}{1-H(s)}p'(s)ds \tag{21}$$

$$E[g_1(T_1)g_3(T_1,\delta_1)] = O(b^2) \tag{22}$$

$$E[g_2(T_1)g_3(T_1,\delta_1)] = O(b^2). \tag{23}$$

Theorem 3 Assume (K.1) and (C.1). Then

$$\mathrm{Var}\left(\overline{\Lambda_n^P}(t) - \Lambda_F(t)\right) = n^{-1}[\gamma(t) - 2e_K bq(t) + O(b^2)]$$

where

$$\gamma(t) = \int_0^t \frac{dH^u(s)}{(1-H(s))^2}. \tag{24}$$

2.2 Presmoothed Kaplan-Meier estimator

To obtain an iid representation for the presmoothed Kaplan-Meier estimator, we rely on the relation $1-F_n^P(t) = \exp(-\Lambda_n^P(t)) + O_P(n^{-1})$. Therefore a second order Taylor expansion yields

$$\begin{aligned} F_n^P(t) - F(t) &= \exp(-\Lambda_F(t)) - \exp(-\Lambda_n^P(t)) + O_p(n^{-1}) \\ &= (1-F(t))(\Lambda_n^P(t) - \Lambda_F(t)) \\ &\quad -\frac{1}{2}(\Lambda_n^P(t) - \Lambda_F(t))^2\exp(-\eta_n(t)) + O_P(n^{-1}) \end{aligned}$$

with $\eta_n(t)$ a stochastic intermediate value between $\Lambda_n^P(t)$ and $\Lambda_F(t)$. Moreover it follows from Theorems 2.1 and 2.3 that $\Lambda_n^P(t) - \Lambda_F(t) = O_P(n^{-1/2})$, so that

$$F_n^P(t) - F(t) = (1-F(t))(\Lambda_n^P - \Lambda_F(t)) + O_P(n^{-1}).$$

We therefore have the following result.

Theorem 4 Assume (K.1), (H.1) and $b = c_0 n^{-\alpha} + o(n^{-\alpha})$, for some $1/4 < \alpha < 1/2$ and some $c_0 > 0$. Then,

$$F_n^P(t) - F(t) = \overline{F_n^P}(t) - F(t) + o_P(n^{-1/2})$$

with

$$\overline{F_n^P}(t) = F(t) + (1 - F(t))\frac{1}{n}\sum_{i=1}^{n}(g_1(T_i) - g_2(T_i) + g_3(T_i, \delta_i))$$

and

$$\text{Var}\left(\overline{F_n^P}(t) - F(t)\right) = n^{-1}(1 - F(t))^2[\gamma(t) - 2e_K bq(t) + O(b^2)].$$

3 Mean squared error: beneficial effect of presmoothing

In this section we show the beneficial effect of presmoothing. In fact we show that the asymptotic mean squared errors ($AMSE$) of $\overline{\Lambda_n^P}(t)$ and $\overline{F_n^P}(t)$ are smaller than the mean squared errors of the Nelson-Aalen estimator and of the Kaplan-Meier estimator.

We illustrate this for the $AMSE$ of $\overline{\Lambda_n^P}(t)$. This $AMSE$ is defined as follows

$$AMSE\left(\overline{\Lambda_n^P}(t)\right) = AVar\left(\overline{\Lambda_n^P}(t)\right) + \left(ABias\left(\overline{\Lambda_n^P}(t)\right)\right)^2$$

where $AVar$ is the sum of the first two order terms of the variance and $ABias$ is the dominant term of the bias of the linear approximation in the asymptotic representation of Theorem 1. From Theorem 3 we have that

$$AVar\left(\overline{\Lambda_n^P}(t)\right) = n^{-1}\gamma(t) - 2e_K q(t)n^{-1}b$$

where $q(t)$, e_K and $\gamma(t)$ are given in (16), (14) and (24).

For the asymptotic bias we have from (17) in Lemma 2 that

$$ABias\left(\overline{\Lambda_n^P}(t)\right) = d_K\alpha(t)b^2$$

where $\alpha(t)$ is given in (15).

We therefore have that

$$AMSE\left(\overline{\Lambda_n^P}(t)\right) = n^{-1}\gamma(t) - 2e_K q(t)n^{-1}b + d_K^2\alpha(t)^2 b^4. \tag{25}$$

Thus, the asymptotic optimal bandwidth, $b_{OPT}(t) = \operatorname*{argmin}_{h>0} AMSE\left(\overline{\Lambda_n^P}(t)\right)$, is

$$b_{OPT}(t) = \left(\frac{e_K q(t)}{2d_K^2\alpha(t)^2 n}\right)^{1/3}. \tag{26}$$

For the asymptotically optimal bandwidth (26), $AMSE\left(\overline{\Lambda_n^P}(t)\right)$ becomes

$$n^{-1}\gamma(t) - \frac{3}{2^{4/3}}\left(\frac{e_K^4 q(t)^4}{d_K^2\alpha(t)^2}\right)^{1/3} n^{-4/3}.$$

This expression shows that the version $\overline{\Lambda_n^P}(t)$ of the presmoothed Nelson-Aalen estimator is more efficient than the classical Nelson-Aalen estimator. Indeed, for the latter we have that

$$Var(\Lambda_n^{NA}(t)) = n^{-1}\gamma(t) + O\left(n^{-3/2}\right).$$

This makes clear that the second order term of the variance of the Nelson-Aalen estimator is negligible with respect to the order $n^{-4/3}$. Hence the second order efficiency of $\overline{\Lambda_n^P}(t)$ with respect to $\Lambda_n^{NA}(t)$ is shown. The validity of the above expression for $Var(\Lambda_n^{NA}(t))$ follows from the iid representation for $\Lambda_n^{NA}(t) - \Lambda_F(t)$ together with the order of the moments of the remainder term in there (see e.g. Lo, Mack and Wang (1989) or Gijbels and Wang (1993)).

Similar properties for $\Lambda_n^P(t)$ are not easy to derive since the $o_P\left(n^{-1/2}\right)$ term in (9) is not negligible enough to obtain such a result from our previous discussion and Theorem 1. A deeper analysis of this term would be a rather complicated task. However, the simulation results in Section 5 point out that this second order efficiency is also present for $\Lambda_n^P(t)$.

4 Plug-in bandwidth selection

Using the asymptotic expression (25), it is easy to obtain the following expression for the weighted asymptotic mean integrated squared error:

$$\begin{aligned} AMISE_w\left(\overline{\Lambda_n^P}(t)\right) &= n^{-1}\int_0^\infty \gamma(t)w(t)dt - 2e_k n^{-1} b \int_0^\infty q(t)w(t)dt \\ &\quad + d_k^2 b^4 \int_0^\infty \alpha(t)^2 w(t)dt \end{aligned}$$

where w is a positive weight function. Therefore, the optimal bandwidth, in the sense of $AMISE_w\left(\overline{\Lambda_n^P}(t)\right)$, is

$$b_{OPT} = \arg\min_{b>0} AMISE_w\left(\overline{\Lambda_n^P}(t)\right) = \left(\frac{e_k Q}{2d_k^2 nA}\right)^{1/3}$$

where

$$\begin{aligned} Q &= \int_0^\infty q(t)w(t)dt \\ A &= \int_0^\infty \alpha(t)^2 w(t)dt. \end{aligned}$$

From now on, we will consider the following plug-in bandwidth selector of b

$$\widehat{b} = \left(\frac{e_k \widehat{Q}}{2d_k^2 n\widehat{A}}\right)^{1/3}$$

where

$$\widehat{Q}=\frac{1}{n}\sum_{i=1}^{n}\left(1-H_n(T_i)+\frac{1}{n}\right)^{-2}p_n(T_i)\left(1-p_n(T_i)\right)w(T_i) \tag{27}$$

$$\widehat{A}=\int_0^{\infty}\widehat{\alpha}(t)^2w(t)dt$$

$$\widehat{\alpha}(t)=\int_0^t\left(1-H_n(s)+\frac{1}{n}\right)^{-1}\left(p_n''(s)h_n(s)/2+p_n'(s)h_n'(s)\right)ds$$

and p_n, p_n' and p_n'' are the Nadaraya-Watson estimator of p and its first and second derivatives,

$$p_n(t)=\frac{\frac{1}{n}\sum K_g(t-T_i)\delta_i}{\frac{1}{n}\sum K_g(t-T_i)}=\frac{\psi_n(t)}{h_n(t)} \tag{28}$$

$$p_n'(t)=\frac{\psi_n'(t)h_n(t)-\psi_n(t)h_n'(t)}{h_n(t)^2} \tag{29}$$

$$p_n''(t)=\frac{\psi_n''(t)h_n(t)^2-\psi_n(t)h_n''(t)h_n(t)-2\psi_n'(t)h_n'(t)h_n(t)+2\psi_n(t)h_n'(t)^2}{h_n(t)^3} \tag{30}$$

where $f^{(k)}$ denotes the k-th derivative of f,

$$\begin{aligned}\psi_n^{(k)}(t) &= \frac{1}{n}\sum K_g^{(k)}(t-T_i)\delta_i\\ h_n^{(k)}(t) &= \frac{1}{n}\sum K_g^{(k)}(t-T_i)\\ K_g^{(k)}(t) &= \frac{1}{g^{k+1}}K^{(k)}\left(\frac{t}{g}\right)\end{aligned}$$

and the estimator $p_n^{(k)}(t), k=1,2,\ldots$, is the k-th derivative of $p_n(t)$. The functions h_n and $h_n'$are the Parzen-Rosenblatt kernel estimators of the density h and its first derivative and H_n is the empirical distribution function of the T_i.

Typically, the effective calculation of $\widehat{b}$ requires the election of some pilot bandwidths, g_1 and g_2, for the estimators $\widehat{A}$ and $\widehat{Q}$, respectively. As a preliminary step for the choice of the pilot bandwidths, some asymptotic expressions for the mean squared errors of the dominant parts of $\widehat{A}$ and $\widehat{Q}$ will be obtained in Subsection 6.2. The criterion to choose the pilot bandwidths g_1 and g_2 will consist in minimizing the dominant part of those expressions.

In the rest of this paper, we will make use of some further assumptions:

On the weight function w:

(W.1) w is non negative, with support within the interval $(\varepsilon/2, t_0)$ and twice differentiable in $[\varepsilon/2, t_0]$, with continuous second derivative.

On the pilot bandwidths, g_1 and g_2:

(V.1) $ng_1^3 \left(\log \frac{1}{g_1}\right)^{-3} \to \infty$ and $ng_1^6 \to 0$.

(V.2) $ng_2^{8/3} \to \infty$ and $ng_2^4 \to 0$.

Our next result gives the rate of convergence of the plug-in bandwidth selector when using the asymptotically optimal pilot bandwidths, as it will be detailed in Subsection 6.2.

Theorem 5 Under conditions (K.1), (P.1), (P.2), (H.1), (W.1), (V.1) and (V.2) and using the pilot bandwidths in (41) and (47) we have

$$\widehat{b} - b_{OPT} = O_P\left(n^{-11/15}\right)$$

and

$$\frac{\widehat{b} - b_{OPT}}{b_{OPT}} = O_P\left(n^{-2/5}\right).$$

Practical implementation of this plug-in bandwidth needs of selecting the pilot bandwidths g_1 and g_2. To do this, we used equations (41) and (47) and estimated the underlying functions H, h and p in (42), (43), (48) and (49) using a lognormal parametric fit for the first two and a logistic fit for the last one.

5 Simulations

Some simulations have been carried out in order to evaluate the practical performance of the presmoothed Nelson-Aalen estimator with plug-in bandwidth selector. To fulfill condition (P.2), some shifted version of a Weibull distribution has been considered for the censoring time. For some $\varepsilon > 0$, we define $C \overset{d}{=} W^{\varepsilon}(\alpha_G, \beta_G)$, which means $C - \varepsilon \overset{d}{=} W(\alpha_G, \beta_G)$, while $Y \overset{d}{=} W(\alpha_F, \beta_F)$, where $W(\alpha, \beta)$ denotes the Weibull distribution with shape parameter α and scale parameter β, with density $f(x) = \beta\alpha x^{\alpha-1}\exp(-\beta x^{\alpha}), \quad x > 0$.

Table I collects the parameters used for these two distributions in the four models considered here. The cumulative observable distribution function, H, and the conditional probability of uncensoring, p, pertaining to these models are plotted in Figure 1. The vertical dotted lines in this Figure indicate the left and right endpoints of the support of the weight function, w, which has been set to a constant within these limits. These endpoints have been selected to meet the condition (W.1). For comparison reasons the unconditional censoring probability for these models is very similar (between 0.32 and 0.34).

[Table I about here]

[Figure 1 about here]

Given some estimator of the hazard function, Λ_n, we define the weighted mean integrated squared error as

$$MISE_w(\Lambda_n) = E\left[\int (\Lambda_n(t) - \Lambda(t))^2 w(t)\, dt\right].$$

Using 500 samples of size $n = 30$, 200 and 500, the $MISE_w$ ratio:

$$\frac{MISE_w(\Lambda_n^P)}{MISE_w(\Lambda_n^{NA})}$$

has been approximated by simulation for a grid in a wide range of possible bandwidths. These functions of the smoothing parameter are plotted in Figures 2–5 for the four models considered here. Values of the $MISE_w$ ratio below 1 indicate that the presmoothed Nelson-Aalen estimator is better than the ordinary Nelson-Aalen estimator for these bandwidths.

[Figure 2 about here]

[Figure 3 about here]

[Figure 4 about here]

[Figure 5 about here]

Figures 2–5 show that there are quite wide ranges of presmoothing parameters for which the new estimator is better than the classical Nelson-Aalen estimator. The minimal $MISE_w$ ratio may be about 0.8 to 0.95, depending on the model and the sample size. Typically, those optimal $MISE_w$ ratios get closer to 1 as the sample size increases. Some special case is model 4, for which, for any possible presmoothing factor in an extremely large range, the $MISE_w$ ratio is smaller than 1. Figure 1 shows that, for model 3, the p function is almost constant in the interval $[0.6, 1.2]$ where almost all observed data fall, so the Koziol-Green model nearly holds. This means that the ACL estimator of Λ (see Abdushukurov (1987) and Cheng and Lin (1987)), that corresponds to and infinitely large presmoothing parameter, is more efficient than the classical Nelson-Aalen estimator.

In order to investigate the practical performance of the plug-in bandwidth proposed in Section 4 we first obtained this bandwidth selector, $\widehat{b}$, for 500 samples of size $n = 500$ drawn from models 1-4 and computed a Parzen-Rosenblatt kernel density estimation, using the Sheather-Jones bandwidth selector. These curves, together with the optimal $AMISE_w$ bandwidth, b_{OPT}, are plotted in Figure 6. Although for models 1 and 3 $\widehat{b}$ presents a clear bias, it is a reasonable selector for b_{OPT}. The simulation results indicate that for model 1 the plug-in selector is within 26% of deviation from the optimal bandwidth in 90% of the cases. For model 3 the same statement is only valid within 100% of deviation. This is not surprising since a large fluctuation of the bandwidth around its optimal value gives a small loss in terms of $MISE_w$ (see the flat shape of the functions in Figure 4 around their minima).

[Figure 6 about here]

Similar empirical studies to those performed above for a range of presmoothing parameters have been carried out for the presmoothing Nelson-Aalen estimator with plug-in presmoothing parameter. The $MISE_w$ ratio for the data-driven presmoothed estimator, $\Lambda^P_{n,\hat{b}}$, has been computed:

$$\frac{MISE_w\left(\Lambda^P_{n,\hat{b}}\right)}{MISE_w\left(\Lambda^{NA}_n\right)}.$$

Some Monte Carlo approximation of this quantity based on 500 samples has been computed for models 1–4 and different sample sizes. These results are collected in Table II. The figures in this table show that the presmoothed Nelson-Aalen estimator with automatic plug-in bandwidth is about 5% to 12% more efficient than the classical Nelson-Aalen estimator (in terms of $MISE_w$) for models 2–4. The results for model 1 are worse than for the other three. This is possibly caused by the oversmoothing effect of the plug-in bandwidth that can be observed in Figures 2 and 6.

[Table II about here]

6 Proofs

6.1 Proofs of the results of Section 2

Proof of Theorem 1

As to term (I) in (9), we have the well known iid representation

$$\widehat{\Lambda}_H(s) - \Lambda_H(s) = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}(T_i \leq s) - H(s)}{1 - H(s)} + r_n(s) \qquad (31)$$

where for $t > 0$ such that $H(t) < 1$:

$$\sup_{0 \le s \le t} | r_n(s) | = O(n^{-1} \log n) \qquad \text{a.s.}$$

as $n \to \infty$. This result follows from the more general representation in the censored data case, due to Lo and Singh (1986). Using integration by parts gives that

$$(I) = \frac{1}{n} \sum_{i=1}^{n} (g_1(T_i) - g_2(T_i)) + O(n^{-1} \log n) \qquad \text{a.s.}$$

with g_1 and g_2 as in (11) and (12).

For the term (II) in (9), we have, with $h_n(s) = n^{-1} \sum_{i=1}^{n} K_b(s - T_i)$:

$$p_n(s) - p(s) = \frac{n^{-1} \sum_{i=1}^{n} K_b(s - T_i)(\delta_i - p(s))}{h(s)} - (p_n(s) - p(s))(h_n(s) - h(s)) \frac{1}{h(s)}. \qquad (32)$$

This gives that

$$(II) = \frac{1}{n} \sum_{i=1}^{n} g_3(T_i, \delta_i) + o_p\left(n^{-1/2}\right) \qquad (33)$$

with g_3 given by (13). The $o_p\left(n^{-1/2}\right)$ remainder term in (33) is obtained as follows:

$$\left| \int_0^t (p_n(s) - p(s))(h_n(s) - h(s)) \frac{1}{h(s)} d\Lambda_H(s) \right|$$

$$\le \frac{t}{1 - H(t)} \left\{ \sup_{0 \le s \le t} | p_n(s) - p(s) | \right\} \left\{ \sup_{0 \le s \le t} | h_n(s) - h(s) | \right\}. \qquad (34)$$

Under the given conditions, it follows from Lemma 1 and Theorem B in Mack and Silverman (1982) that

$$\sup_{0 \le s \le t} | p_n(s) - p(s) | = O_p\left((nb)^{-1/2} (\log(1/b))^{1/2}\right)$$

and

$$\sup_{0 \le s \le t} | h_n(s) - E(h_n(s)) | = O_p\left((nb)^{-1/2} (\log(1/b))^{1/2}\right).$$

By a standard Taylor expansion argument, it also follows that $\sup_{0 \le s \le t} | E(h_n(s)) - h(s) | = O_p\left(b^2\right)$. Therefore, the right hand side in (34) is $o_p\left(n^{-1/2}\right)$.

For the term (III) in (9), we first plug in representation (31), which yields

$$(III) = \int_0^t (p_n(s) - p(s)) d\left(\frac{H_n(s) - H(s)}{1 - H(s)}\right) + o\left(n^{-1} \log n\right) \qquad \text{a.s.}$$

$$= \int_0^t \frac{p_n(s) - p(s)}{1 - H(s)} dH_n(s) - \int_0^t \frac{p_n(s) - p(s)}{1 - H(s)} dH(s)$$

$$+ \int_0^t \frac{(p_n(s) - p(s))(H_n(s) - H(s))}{(1 - H(s))^2} dH(s) + o(n^{-1} \log n). \qquad (35)$$

The absolute value of the third term in the right hand side of (35) is bounded above by

$$\frac{1}{(1 - H(t))^2} \left\{ \sup_{0 \le s \le t} \mid p_n(s) - p(s) \mid \right\} \left\{ \sup_{0 \le s \le t} \mid H_n(s) - H(s) \mid \right\} = o_p\left(n^{-1/2}\right)$$

since Theorem A in Mack and Silverman (1982) gives that $\sup_{0 \le s \le t} \mid p_n(s) - p(s) \mid = o_p(1)$ and a classical empirical process result gives that $\sup_{0 \le s \le t} \mid H_n(s) - H(s) \mid = O_p\left(n^{-1/2}\right)$.

For the second term in the right hand side of (35), we have by (32) that it equals

$$\frac{1}{n} \sum_{i=1}^n \int_0^t \frac{K_b(s - T_i)(\delta_i - p(s))}{1 - H(s)} ds + o_p\left(n^{-1/2}\right).$$

The first term in (35) is equal to

$$\frac{1}{n} \sum_{i=1}^n \frac{p_n(T_i) - p(T_i)}{1 - H(T_i)} \mathbf{1}(T_i \le t)$$

$$= \frac{1}{n^2} \sum_{i,j=1}^n \frac{K_b(T_i - T_j)(\delta_j - p(T_i))}{(1 - H(T_i)) h(T_i)} \mathbf{1}(T_i \le t) + o_p\left(n^{-1/2}\right)$$

by using representation (32) and the same bounds for $\sup_{0 \le s \le t} \mid p_n(s) - p(s) \mid$ and $\sup_{0 \le s \le t} \mid h_n(s) - h(s) \mid$ as before. Introducing

$$\varphi(T_i, T_j, \delta_j) = \frac{K_b(T_i - T_j)(\delta_j - p(T_i))}{(1 - H(T_i)) h(T_i)} \mathbf{1}(T_i \le t)$$

we have that

$$\begin{aligned} (III) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\varphi(T_i,T_j,\delta_j) \\ &- \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{K_b(s-T_i)(\delta_i-p(s))}{1-H(s)}ds + o_p\left(n^{-1/2}\right) \end{aligned}$$

since $\frac{1}{n^2}\sum_{i=1}^{n}\varphi(T_i,T_i,\delta_i) = O_p\left(\frac{1}{n^{3/2}b}\right) = o_p\left(n^{-1/2}\right)$, because it is $\frac{1}{n^2 b}$ times a sum of zero mean, iid random variables with finite variance. By symmetrization of the kernel φ, i.e. by putting

$$\psi(T_i,\delta_i,T_j,\delta_j) = \frac{1}{2}(\varphi(T_i,T_j,\delta_j)+\varphi(T_j,T_i,\delta_i)),$$

we obtain that

$$(III) = U_n - \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{K_b(s-T_i)(\delta_i-p(s))}{1-H(s)}ds + o_p\left(n^{-1/2}\right) \tag{36}$$

where U_n is the U-statistic with symmetric kernel ψ, i.e.

$$U_n = \left(\begin{array}{c} n \\ 2 \end{array}\right)^{-1}\sum_{1\leq i<j\leq n}\psi(T_i,\delta_i,T_j,\delta_j).$$

For the Hajek projection of U_n, we have

$$\begin{aligned} G(T_1,\delta_1) &= E[\psi(T_1,\delta_1,T_2,\delta_2)\mid T_1,\delta_1] \\ &= \frac{1}{2}\left\{E(\varphi(T_1,T_2,\delta_2)\mid T_1) + E(\varphi(T_2,T_1,\delta_1)\mid T_1,\delta_1)\right\} \\ &:= \frac{1}{2}\left\{G_1(T_1)+G_2(T_1,\delta_1)\right\}. \end{aligned}$$

We calculate

$$\begin{aligned} G_1(t_1) &= E(\varphi(T_1,T_2,\delta_2)\mid T_1=t_1) \\ &= E\left[\frac{K_b(t_1-T_2)[\delta_2-p(t_1)]}{(1-H(t_1))h(t_1)}\mathbb{1}\{t_1\leq t\}\right] \\ &= \frac{\mathbb{1}\{t_1\leq t\}}{(1-H(t_1))h(t_1)}\int K_b(t_1-t_2)\{E[\delta_2\mid T_2=t_2]-p(t_1)\}h(t_2)dt_2 \\ &= \frac{\mathbb{1}\{t_1\leq t\}}{(1-H(t_1))h(t_1)}\int K_b(t_1-t_2)[p(t_2)-p(t_1)]h(t_2)dt_2 \end{aligned}$$

and

$$\begin{aligned} G_2(t_1, d_1) &= E(\varphi(T_2, T_1, \delta_1) \mid T_1 = t_1, \delta_1 = d_1) \\ &= E\left[\frac{K_b(T_2 - t_1)}{(1 - H(T_2))h(T_2)}[d_1 - p(T_2)]1\{T_2 \le t\}\right] \\ &= \int K_b(t_2 - t_1)\frac{d_1 - p(t_2)}{1 - H(t_2)}I(t_2 \le t)dt_2. \end{aligned}$$

From U-statistic theory, it follows that

$$U_n = -\theta_n + \frac{1}{n}\sum_{i=1}^{n} G_1(T_i) + \frac{1}{n}\sum_{i=1}^{n} G_2(T_i, \delta_i) + o_p\left(n^{-1/2}\right)$$

where

$$\theta_n = E(\psi(T_1, \delta_1, T_2, \delta_2))$$

and, from (36),

$$(III) = -\theta_n + \frac{1}{n}\sum_{i=1}^{n} G_1(T_i) + o_p\left(n^{-1/2}\right).$$

To show that $\theta_n = O(b^2)$, note that by Taylor expansion it is easy to see that $E[G_1(T_1)] = O(b^2)$, so that we only have to show that $E[G_2(T_1, \delta_1)] = O(b^2)$.

Use $p(T_1) = E[\delta_1 \mid T_1]$ to obtain, for h small enough,

$$\begin{aligned} E[G_2(T_1, \delta_1)] &= -b\int_0^{t-bL} \frac{p'(t_1)}{1 - H(t_1)}\left[\int_{-L}^{L} vK(v)dv\right]dH(t_1) \\ &\quad -h\int_{t-bL}^{t+bL} \frac{p'(t_1)}{1 - H(t_1)}\left[\int_{-L}^{(t-t_1)/b} vK(v)dv\right]dH(t_1) + O(b^2) \\ &= O(b^2) \end{aligned}$$

Similarly, it follows that $E[G_1^2(T_1] = O(b^2)$, so that we can conclude that $(III) = o_p\left(n^{-1/2}\right)$ (since $nb^4 \to 0$).

Proof of Lemma 2

Using the fact that $p(u) = E(\delta_1 \mid T_1 = u)$, we have

$$\begin{aligned} E[g_3(T_1, \delta_1)] &= E\left[\int_0^t \frac{K_b(s - T_1)(\delta_1 - p(s))}{1 - H(s)}ds\right] \\ &= \int_0^t \int \frac{K_b(s - u)(p(u) - p(s))}{1 - H(s)}h(u)du\ ds \\ &= \int_0^t \int \frac{K(v)(p(s - bv) - p(s))}{1 - H(s)}h(s - bv)dv\ ds. \end{aligned}$$

Under the smoothness conditions of (C), performing a Taylor expansion of $p(s-bv)$ and $h(s-bv)$, and using that $\int_{-L}^{L} K(v)dv = 1$ and $\int_{-L}^{L} vK(v)dv = 0$, we obtain

$$E[g_3(T_1,\delta_1)][= d_K\alpha(t)b^2 + o(b^2)$$

with d_K and $\alpha(t)$ given in (14) and (15).

The derivation of (18) and (19) is straightforward. For $\text{Var}[g_3(T_1,\delta_1)]$ we have the following:

$$\begin{aligned}
& E[g_3^2(T_1,\delta_1)] \\
&= E\left[\iint \frac{K(u_1)[\delta_1 - p(T_1+bu_1)]K(u_2)[\delta_1 - p(T_1+bu_2)]}{(1-H(T_1+bu_1))(1-H(T_1+bu_2))} \right. \\
&\qquad\qquad \left. \times\ \mathbf{1}\{T_1 \le (t-bu_1)\wedge(t-bu_2)\}du_1du_2\right] \\
&= \iiint \frac{K(u_1)K(u_2)\{p(t_1) - p(t_1)[p(t_1+bu_1)+p(t_1+bu_2)] + p(t_1+bu_1)p(t_1+bu_2)\}}{(1-H(t_1+bu_1))(1-H(t_1+bu_2))} \\
&\qquad\qquad \times \mathbf{1}\{t_1 \le t - b(u_1\vee u_2)\}dH(t_1)du_1du_2 \\
&= \iiint \frac{p(t_1)(1-p(t_1))h(t_1)}{(1-H(t_1))^2}\mathbf{1}\{t_1 \le t-b(u_1\vee u_2)\}K(u_1)K(u_2)du_1du_2 + O(b^2) \\
&= \iint Q(t-b(u_1\vee u_2))K(u_1)K(u_2)du_1du_2 + O(b^2)
\end{aligned}$$

where $Q(s) = \int_0^s q(v)dv$. Now

$$\begin{aligned}
& \int\int Q(t-b(u_1\vee u_2))K(u_1)K(u_2)du_1du_2 \\
&= 2\int_{-L}^{L}\int_{-L}^{u_2} Q(t-bu_2)K(u_1)K(u_2)du_1du_2 \\
&= 2\int_{-L}^{L} Q(t-bu_2)\mathsf{K}(u_2)K(u_2)du_2 \\
&= 2Q(t)\int_{-L}^{L}\mathsf{K}(u_2)K(u_2)du_2 - 2bq(t)\int_{-L}^{L}u_2\mathsf{K}(u_2)K(u_2)du_2 + O(b^2) \\
&= Q(t)\mathsf{K}^2(u_2)\big|_{-L}^{L} - 2bq(t)e_K + O(b^2) \\
&= \int_0^t \frac{p(v)(1-p(v))}{(1-H(v))^2}dH(v) - 2bq(t)e_K + O(b^2).
\end{aligned}$$

The derivation of (21) is straightforward. For (22), we have

$$E[g_1(T_1)g_3(T_1,\delta_1)] = p(t)\int_0^t\int_0^t \frac{K_h(s-u)(p(u)-p(s))}{1-H(s)}h(u)duds$$

and this is $O(b^2)$ by Taylor expansion arguments. Similarly for (23).

Proof of Theorem 3. From expression (10) in Theorem 1 it follows that

$$\begin{aligned} n\text{Var}\left(\overline{\Lambda_n^P}(t) - \Lambda_F(t)\right) &= \text{Var}[g_1(T_1)] + \text{Var}[g_2(T_1)] \\ &\quad + \text{Var}[g_3(T_1, \delta_1)] - 2\text{Cov}(g_1(T_1), g_2(T_1)) \\ &\quad + 2\text{Cov}(g_1(T_1), g_3(T_1, \delta_1)) - 2\text{Cov}(g_2(T_1), g_3(T_1, \delta_1)). \end{aligned}$$

Now use the expressions in Lemma 2 and some extra algebra. ¤

6.2 Auxiliary results for Section 4

Let us mention that there exist similar expressions to (28), (29) and (30) but for populational quantities. Indeed,

$$p(t) = \frac{\psi(t)}{h(t)}$$

$$p'(t) = \frac{\psi'(t)h(t) - \psi(t)h'(t)}{h(t)^2} \tag{37}$$

$$p''(t) = \frac{\psi''(t)h(t)^2 - \psi(t)h''(t)h(t) - 2\psi'(t)h'(t)h(t) + 2\psi(t)h'(t)^2}{h(t)^3} \tag{38}$$

where h is the density function of the observed lifetime and $\psi = ph$.

Assumption (P.2), stated in Section 2, warranties that for any t in the interval $[0, \varepsilon/2]$ there exists some (common) value for the bandwidth used in the estimators $p_n''(s)$ and $p_n'(s)$ in $\widehat{\alpha}(t)$ such that for bandwidths smaller than that value, $\widehat{\alpha}(t) = 0$, with probability 1. For this reason when studying the asymptotic behaviour of $\widehat{A}$ we will consider its asymptotic equivalent term

$$\widehat{A}_{\varepsilon/2} = \int_{\varepsilon^\theta}^{\infty} \left(\int_{\varepsilon^\theta}^{t} \left(1 - H_n(s) + n^{-1}\right)^{-1} \left(p_n''(s)h_n(s)/2 + p_n'(s)h_n'(s)\right) \right)^2 dsw(t)dt$$

where $\varepsilon' = \frac{\varepsilon}{2}$.

In the rest of this subsection some lemmas and theorems useful to prove Theorem 5 will be stated. The proofs of these, as well as of some auxiliary results for them, are omitted here. They can be found in the Ph.D. Dissertation of López-de-Ullibarri and in the technical report by Cao, López-de-Ullibarri, Janssen and Veraverbeke (2003).

Lemma 6 Under conditions (K.1), (H.1), (W.1) and (V.1),

$$\widehat{A}_{\varepsilon/2} = \widehat{A}_1 + o_P\left(\widehat{A}_1\right)$$

where

$$\begin{aligned}\widehat{A}_1 &= \frac{1}{4}\int_{\varepsilon^\theta}^{\infty}\int_{\varepsilon^\theta}^{t}\int_{\varepsilon^\theta}^{t}(1-H(r))^{-1}(1-H(s))^{-1} \\ &\quad \times \left(\psi_n''(r) - p_n(r)h_n''(r)\right)\left(\psi_n''(s) - p_n(s)h_n''(s)\right)w(t)drdsdt. \quad (39)\end{aligned}$$

The term $\widehat{A}_1$ has still to be linearized, since the estimator p_n has a random denominator. To obtain such a linearization the expression $\psi_n'' - p_n h_n''$ is factorized as follows

$$\begin{aligned}\psi_n'' - p_n h_n'' &= \psi'' - ph'' + (\psi_n'' - \psi'') - p(h_n'' - h'') \\ &\quad -(\psi_n - \psi)h''h^{-1} + p(h_n - h)h''h^{-1} \\ &\quad +(p_n - p)((h_n - h)h''h^{-1} - (h_n'' - h'')) \quad (40)\end{aligned}$$

where we have used the following relation

$$p_n - p = (\psi_n - \psi)h^{-1} - p(h_n - h)h^{-1} - (p_n - p)(h_n - h)h^{-1}.$$

Substituting (40) in expression (39) we obtain

$$\begin{aligned}\widehat{A}_1 &= \frac{1}{4}\int_{\varepsilon^\theta}^{\infty}\int_{\varepsilon^\theta}^{t}\int_{\varepsilon^\theta}^{t}(1-H(r))^{-1}(1-H(s))^{-1} \\ &\quad \times \left(\psi''(r) - p(r)h''(r) + (\psi_n''(r) - \psi''(r)) - p(r)(h_n''(r) - h''(r))\right. \\ &\quad -(\psi_n(r) - \psi(r))h''(r)h(r)^{-1} + p(r)(h_n(r) - h(r))h''(r)h(r)^{-1} \\ &\quad \left.+(p_n(r) - p(r))((h_n(r) - h(r))h''(r)h(r)^{-1} - (h_n''(r) - h''(r)))\right) \\ &\quad \times \left(\psi''(s) - p(s)h''(s) + (\psi_n''(s) - \psi''(s)) - p(s)(h_n''(s) - h''(s))\right. \\ &\quad -(\psi_n(s) - \psi(s))h''(s)h(s)^{-1} + p(s)(h_n(s) - h(s))h''(s)h(s)^{-1} \\ &\quad \left.+(p_n(s) - p(s))((h_n(s) - h(s))h''(s)h(s)^{-1} - (h_n''(s) - h''(s)))\right)w(t)drdsdt.\end{aligned}$$

Now, using (37) and (38)

$$\begin{aligned}A &= \int_0^{\infty}\left(\int_0^t (1-H(s))^{-1}\left(p''(s)h(s)/2 + p'(s)h'(s)\right)ds\right)^2 w(t)dt \\ &= \frac{1}{4}\int_0^{\infty}\left(\int_0^t (1-H(s))^{-1}\left(\psi''(s) - p(s)h''(s)\right)ds\right)^2 w(t)dt \\ &= \frac{1}{4}\int_0^{\infty}\int_0^{t}\int_0^{t}(1-H(r))^{-1}(1-H(s))^{-1} \\ &\quad \times \left(\psi''(r) - p(r)h''(r)\right)\left(\psi''(s) - p(s)h''(s)\right)w(t)drdsdt\end{aligned}$$

which implies

$$
\begin{aligned}
\widehat{A}_1 \;=\;& A+\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}\big((\psi_n''(r)-\psi''(r))-p(r)(h_n''(r)-h''(r))\\
&-(\psi_n(r)-\psi(r))h''(r)h(r)^{-1}+p(r)(h_n(r)-h(r))h''(r)h(r)^{-1}\big)\,\alpha(t)w(t)drdt\\
&+\frac{1}{4}\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(1-H(s))^{-1}\\
&\times\big((\psi_n''(r)-\psi''(r))-p(r)(h_n''(r)-h''(r))\\
&-(\psi_n(r)-\psi(r))h''(r)h(r)^{-1}+p(r)(h_n(r)-h(r))h''(r)h(r)^{-1}\big)\\
&\times\big((\psi_n''(s)-\psi''(s))-p(s)(h_n''(s)-h''(s))\\
&-(\psi_n(s)-\psi(s))h''(s)h(s)^{-1}+p(s)(h_n(s)-h(s))h''(s)h(s)^{-1}\big)\,w(t)drdsdt\\
&+\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(p_n(r)-p(r))\\
&\times\big((h_n(r)-h(r))h''(r)h(r)^{-1}-(h_n''(r)-h''(r))\big)\,\alpha(t)w(t)drdt\\
&+\frac{1}{2}\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(1-H(s))^{-1}\\
&\times\big((\psi_n''(r)-\psi''(r))-p(r)(h_n''(r)-h''(r))\\
&-(\psi_n(r)-\psi(r))h''(r)h(r)^{-1}+p(r)(h_n(r)-h(r))h''(r)h(r)^{-1}\big)\\
&\times(p_n(s)-p(s))\big((h_n(s)-h(s))h''(s)h(s)^{-1}-(h_n''(s)-h''(s))\big)\,w(t)drdsdt\\
&+\frac{1}{4}\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(1-H(s))^{-1}\\
&\times(p_n(r)-p(r))\big((h_n(r)-h(r))h''(r)h(r)^{-1}-(h_n''(r)-h''(r))\big)\\
&\times(p_n(s)-p(s))\big((h_n(s)-h(s))h''(s)h(s)^{-1}-(h_n''(s)-h''(s))\big)\,w(t)drdsdt.
\end{aligned}
$$

Repeating the linearization of the second but last summand,

$$
\begin{aligned}
&\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(p_n(r)-p(r))\\
&\times\big((h_n(r)-h(r))h''(r)h(r)^{-1}-(h_n''(r)-h''(r))\big)\,\alpha(t)w(t)drdt\\
=\;&\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}\big((\psi_n(r)-\psi(r))h(r)^{-1}-p(r)(h_n(r)-h(r))h(r)^{-1}\big)\\
&\times\big((h_n(r)-h(r))h''(r)h(r)^{-1}-(h_n''(r)-h''(r))\big)\,\alpha(t)w(t)drdt\\
&-\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(p_n(r)-p(r))(h_n(r)-h(r))\\
&\times\big((h_n(r)-h(r))h''(r)h(r)^{-1}-(h_n''(r)-h''(r))\big)\,\alpha(t)w(t)drdt.
\end{aligned}
$$

Define

$$
\begin{aligned}
\widehat{A}_{11} \;=\;& \int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}\big((\psi_n''(r)-\psi''(r))-p(r)(h_n''(r)-h''(r))\\
&-h''(r)h(r)^{-1}\left((\psi_n(r)-\psi(r))-p(r)(h_n(r)-h(r))\right)\big)\,\alpha(t)w(t)drdt
\end{aligned}
$$

$$
\begin{aligned}
\widehat{A}_{12} &= \frac{1}{4}\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(1-H(s))^{-1} \\
&\quad \times \big((\psi_n''(r)-\psi''(r))-p(r)(h_n''(r)-h''(r)) \\
&\quad -h''(r)h(r)^{-1}\left((\psi_n(r)-\psi(r))-p(r)(h_n(r)-h(r))\right)\big) \\
&\quad \times \big((\psi_n''(s)-\psi''(s))-p(s)(h_n''(s)-h''(s)) \\
&\quad -h''(s)h(s)^{-1}\left((\psi_n(s)-\psi(s))-p(s)(h_n(s)-h(s))\right)\big)\, w(t)drdsdt
\end{aligned}
$$

$$
\begin{aligned}
\widehat{A}_{13} &= \int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}h(r)^{-1}\left((\psi_n(r)-\psi(r))-p(r)(h_n(r)-h(r))\right) \\
&\quad \times \big((h_n(r)-h(r))h''(r)h(r)^{-1}-(h_n''(r)-h''(r))\big)\,\alpha(t)w(t)drdt
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{A}_{14} &= \frac{1}{2}\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(1-H(s))^{-1} \\
&\quad \times \big((\psi_n''(r)-\psi''(r))-p(r)(h_n''(r)-h''(r)) \\
&\quad -h''(r)h(r)^{-1}\left((\psi_n(r)-\psi(r))-p(r)(h_n(r)-h(r))\right)\big) \\
&\quad \times(p_n(s)-p(s))\big((h_n(s)-h(s))h''(s)h(s)^{-1}-(h_n''(s)-h''(s))\big)\, w(t)drdsdt \\
&\quad -\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(p_n(r)-p(r))(h_n(r)-h(r)) \\
&\quad \times \big((h_n(r)-h(r))h''(r)h(r)^{-1}-(h_n''(r)-h''(r))\big)\,\alpha(t)w(t)drdt \\
&\quad +\frac{1}{4}\int_{\varepsilon^0}^{\infty}\int_{\varepsilon^0}^{t}\int_{\varepsilon^0}^{t}(1-H(r))^{-1}(1-H(s))^{-1} \\
&\quad \times(p_n(r)-p(r))((h_n(r)-h(r))h''(r)h(r)^{-1}-(h_n''(r)-h''(r))) \\
&\quad \times(p_n(s)-p(s))((h_n(s)-h(s))h''(s)h(s)^{-1}-(h_n''(s)-h''(s)))w(t)drdsdt
\end{aligned}
$$

to obtain the representation

$$\widehat{A}_1 - A = \widehat{A}_{11} + \widehat{A}_{12} + \widehat{A}_{13} + \widehat{A}_{14}.$$

Using a more compact notation, defining $\widetilde{A}_1 = \widehat{A}_1 - \widehat{A}_{14}$, it is straightforward to check that

$$\widehat{A}_1 - A = \widetilde{A}_1 - A + \widehat{A}_{14}.$$

Now, the main result for mean squared error of $\widetilde{A}_1$ can be stated.

Theorem 7 Under conditions (K.1), (P.1), (P.2), (H.1), (W.1) and (V.1),

$$MSE\left(\widetilde{A}_1\right) = AMSE\left(\widetilde{A}_1\right) + O\left(g_1^6\right) + o\left(n^{-1}g_1^{-1}\right) + o\left(n^{-2}g_1^{-6}\right)$$

where

$$AMSE\left(\widetilde{A}_1\right) = \left(C_1 g_1^2 + C_2 n^{-1} g_1^{-3}\right)^2$$

and, thus, the bandwidth that minimizes $AMSE\left(\widetilde{A}_1\right)$ is

$$g_{1,AMSE} = Cn^{-1/5} \tag{41}$$

where

$$C = \begin{cases} \left(-\frac{C_2}{C_1}\right)^{1/5}, & \text{if} \quad C_1 < 0 \\ \left(\frac{3C_2}{2C_1}\right)^{1/5}, & \text{if} \quad C_1 > 0 \end{cases}$$

$$C_1 = \frac{1}{2} d_K \int_\varepsilon^\infty \int_\varepsilon^t (1 - H(r))^{-1} \Big(p^{(4)}(r)h(r) + 4p^{(3)}(r)h'(r) + 5p''(r)h''(r) + 4p'(r)h^{(3)}(r) - 2p'(r)h(r)^{-1}h'(r)h''(r) \Big) \alpha_w(t) dr dt \tag{42}$$

and

$$C_2 = \frac{1}{4} c_{K^\theta} \int_\varepsilon^\infty (1 - H(x))^{-2} (1 - p(x)) p(x)h(x)w(x)dx. \tag{43}$$

The term $\widehat{A}_{14}$ can be proved to be negligible as stated in the following lemma.

Lemma 8 Under the conditions (K.1), (P.1), (P.2), (H.1), (W.1) and (V.1), we have

$$\widehat{A}_{14} = o_P\left(n^{-1}g_1^{-3}\right).$$

It remains to study the estimator of Q proposed in (27). First of all let us state some result that gives its dominant term.

Lemma 9 Under conditions (K.1), (H.1), (W.1).and (V.2), it holds

$$\widehat{Q} = \widehat{Q}_1 + o_P\left(\widehat{Q}_1\right)$$

where

$$\widehat{Q}_1 = \frac{1}{n}\sum_{i=1}^n (1 - H(T_i))^{-2} p_n(T_i) (1 - p_n(T_i)) w(T_i).$$

For the term $\widehat{Q}_1$, the representation

$$p_n (1 - p_n) = p (1 - p) + (p_n - p) (1 - 2p) - (p_n - p)^2$$

gives

$$\widehat{Q}_1 = \frac{1}{n}\sum_{i=1}^{n} p(T_i)\,(1-p(T_i))\,w(T_i)$$
$$+\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}\,(p_n(T_i)-p(T_i))\,(1-2p(T_i))\,w(T_i)$$
$$-\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}\,(p_n(T_i)-p(T_i))^2\,w(T_i) \tag{44}$$

The last two summands of (44) can be easily linearized using (32):

$$\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}\,(p_n(T_i)-p(T_i))\,(1-2p(T_i))\,w(T_i)$$
$$=\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}\,(\psi_n(T_i)-p(T_i)h_n(T_i))\,(1-2p(T_i))\,h(T_i)^{-1}w(T_i)$$
$$-\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}\,(p_n(T_i)-p(T_i))\,(h_n(T_i)-h(T_i))\,(1-2p(T_i))\,h(T_i)^{-1}w(T_i) \tag{45}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}\,(p_n(T_i)-p(T_i))^2\,w(T_i)$$
$$=\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}\,(\psi_n(T_i)-p(T_i)h_n(T_i))^2 h(T_i)^{-2}w(T_i)$$
$$-\frac{2}{n}\sum_{i=1}^{n}(\psi_n(T_i)-p(T_i)h_n(T_i))\,(p_n(T_i)-p(T_i))$$
$$\times\,(h_n(T_i)-h(T_i))\,(1-H(T_i))^{-2}\,h(T_i)^{-2}w(T_i)$$
$$+\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}\,(p_n(T_i)-p(T_i))^2\,(h_n(T_i)-h(T_i))^2 h(T_i)^{-2}w(T_i). \tag{46}$$

The same procedure is repeated for the second summand of (45), in order to obtain a linearized term with two factors of the type $(\psi_n - ph_n)$ or $(h_n - h)$, as in the first summand of (46),

$$\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}(p_n(T_i)-p(T_i))(h_n(T_i)-h(T_i))(1-2p(T_i))h(T_i)^{-1}w(T_i)$$
$$=\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}(\psi_n(T_i)-p(T_i)h_n(T_i))(h_n(T_i)-h(T_i))(1-2p(T_i))h(T_i)^{-2}w(T_i)$$
$$-\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}(p_n(T_i)-p(T_i))(h_n(T_i)-h(T_i))^2(1-2p(T_i))h(T_i)^{-2}w(T_i).$$

Define

$$Q_{11}=\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}p(T_i)(1-p(T_i))w(T_i)$$

$$\widehat{Q}_{12}=\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}(\psi_n(T_i)-p(T_i)h_n(T_i))(1-2p(T_i))h(T_i)^{-1}w(T_i)$$

$$\begin{aligned}\widehat{Q}_{13} &= -\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}(\psi_n(T_i)-p(T_i)h_n(T_i))^2h(T_i)^{-2}w(T_i)\\ &\quad-\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}(\psi_n(T_i)-p(T_i)h_n(T_i))(h_n(T_i)-h(T_i))\\ &\quad\times(1-2p(T_i))h(T_i)^{-2}w(T_i)\end{aligned}$$

and

$$\begin{aligned}\widehat{Q}_{14} &= \frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}(p_n(T_i)-p(T_i))(h_n(T_i)-h(T_i))^2(1-2p(T_i))h(T_i)^{-2}w(T_i)\\ &\quad+\frac{2}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}(\psi_n(T_i)-p(T_i)h_n(T_i))(p_n(T_i)-p(T_i))\\ &\quad\times(h_n(T_i)-h(T_i))h(T_i)^{-2}w(T_i)\\ &\quad-\frac{1}{n}\sum_{i=1}^{n}(1-H(T_i))^{-2}(p_n(T_i)-p(T_i))^2(h_n(T_i)-h(T_i))^2h(T_i)^{-2}w(T_i)\end{aligned}$$

to obtain the representation

$$\widehat{Q}_1-Q=Q_{11}-Q+\widehat{Q}_{12}+\widehat{Q}_{13}+\widehat{Q}_{14},$$

or, equivalently, defining $\widetilde{Q}_1=\widehat{Q}_1-\widehat{Q}_{14}$,

$$\widehat{Q}_1-Q=\widetilde{Q}_1-Q+\widehat{Q}_{14}.$$

Now, the mean squared error of $\widetilde{Q}_1$ is given in the next result.

Theorem 10 Under conditions (K.1), (P.1), (H.1), (W.1) and (V.2),

$$MSE\left(\widetilde{Q}_1\right) = AMSE\left(\widetilde{Q}_1\right) + O\left(g_2^6\right) + o\left(n^{-1}g_2^{-1}\right) + o\left(n^{-2}g_2^{-6}\right)$$

where

$$AMSE\left(\widetilde{Q}_1\right) = \left(D_1 g_2^2 + D_2 n^{-1} g_2^{-1}\right)^2$$

and the smoothing parameter that minimizes $AMSE\left(\widetilde{Q}_1\right)$ is

$$g_{2,AMSE} = Dn^{-1/3} \tag{47}$$

where

$$D = \begin{cases} \left(\frac{D_2}{2D_1}\right)^{1/3}, & \text{if} \quad D_1 < 0 \\ \left(-\frac{D_2}{D_1}\right)^{1/3}, & \text{if} \quad D_1 > 0 \end{cases}$$

$$D_1 = d_K \int_{\varepsilon^\theta}^{\infty} (1 - H(x))^{-2} (1 - 2p(x)) \left(p'(x)h'(x) + \frac{1}{2}p''(x)h(x)\right) w(x) dx \tag{48}$$

and

$$D_2 = -c_K \int_{\varepsilon^\theta}^{\infty} (1 - H(x))^{-2} p(x) (1 - p(x)) w(x) dx. \tag{49}$$

The term $\widehat{Q}_{14}$ is proved to be negligible in the following result.

Lemma 11 Under conditions (K.1), (P.1), (H.1), (W.1).and (V.2),

$$\widehat{Q}_{14} = O_P\left(n^{-3/2} g_2^{-3/2} \log\left(1/g_2\right)^{3/2}\right).$$

7 Acknowledgments

Research partly supported by the MCyT Grant BFM2002-00265 (European FEDER support included) and XUGA Grant PGIDT00PXI10501PN for the first two authors and by the International Scientific and Technological Cooperation Grant (BIL 00/28, Flemish Community, Belgium) and the Interuniversity Attraction Poles Research Network P5/24 of the Belgian State for the last two authors.

References

[1] Aalen, O. O.(1978). Nonparametric inference for a family of counting processes. Ann. Statist., **6**, 701-726.

[2] Abdushukurov, A. A. (1987). Estimation of the probability density and intensity function of the Koziol-Green model of random censoring. *Sankhya Ser. A*, **48**, 150-168.

[3] Cao, R., López-de-Ullibarri, I., Janssen, P. and Veraverbeke, N. (2003). Presmoothed Kaplan-Meier and Nelson-Aalen estimators. Technical Report #03-01, Department of Statistics and Operations Research, University of Santiago de Compostela. [www http://eio.usc.es/pub/reports.html]

[4] Cheng, P. E. and Lin, G. D. (1987). Maximum likelihood estimation of a survival function under the Koziol-Green proportional hazards model. *Statist. Probab. Lett.*, **5**, 75-80.

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

[6] Gijbels, I. and Wang, J. L. (1993). Strong representations of the survival function for truncated and censored data with applications. *J. Multivariate Anal.*, **47**, 210-229.

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

[8] Koziol, J. A. and Green, S. B. (1976). A Cramér-von Mises statistic for randomly censored data. *Biometrika*, **63**, 465–474.

[9] Lo, S. H., and Singh, K. (1986). The product-limit estimator and the bootstrap: some asymptotic representations. *Probab. Theory Rel. Fields*, **71**, 455-465.

[10] Lo, S. H., Mack, Y. P. and Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probab. Theory Rel. Fields*, **80**, 461-473.

[11] Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **61**, 405-415.

[12] Nelson, W. (1972). Theory and applications for hazard plotting for censored failure data. *Technometrics*, **14**, 945-965.

[13] Watson, G. S. and Leadbetter, M. R. (1964a). Hazard analysis I. *Biometrika*, **51**, 175-184.

[14] Watson, G. S. and Leadbetter, M. R. (1964b). Hazard analysis II. *Sankhya Ser. A*, **26**, 101-116.

[15] Ziegler, S. (1995). *Ein modifizierter Kaplan-Meier Schätzer*. Diplomarbeit, University of Gießen.

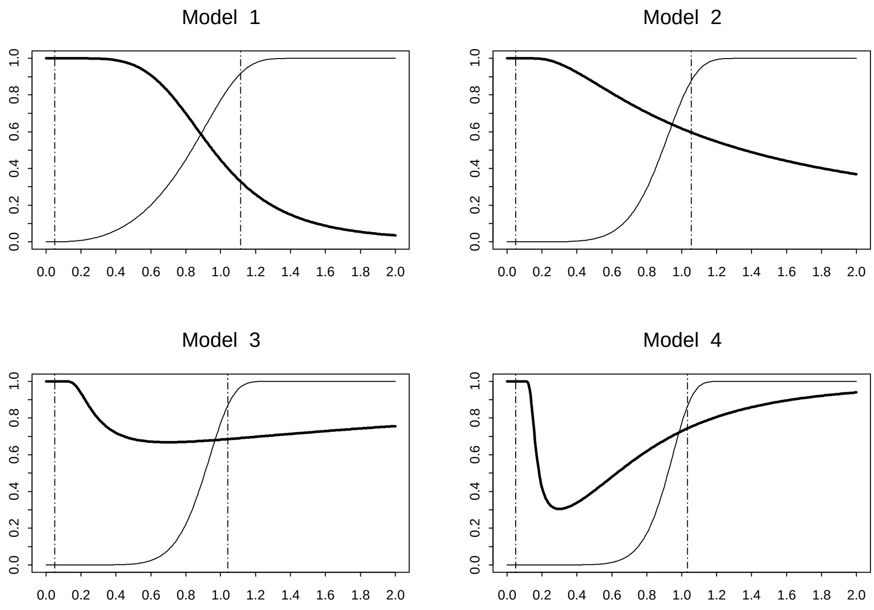

Figure 1. Functions H (thin line) and p (thick line) for models 1–4.

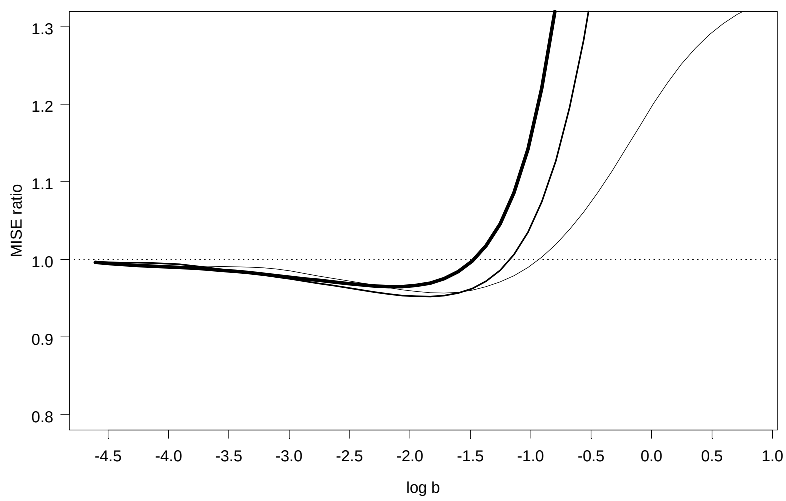

Figure 2. $MISE_w$ ratio of Λ_n^P with respect to Λ_n^{NA} for $n = 30$ (thin line), $n = 200$ (medium line) and $n = 500$ (thick line) for model 1.

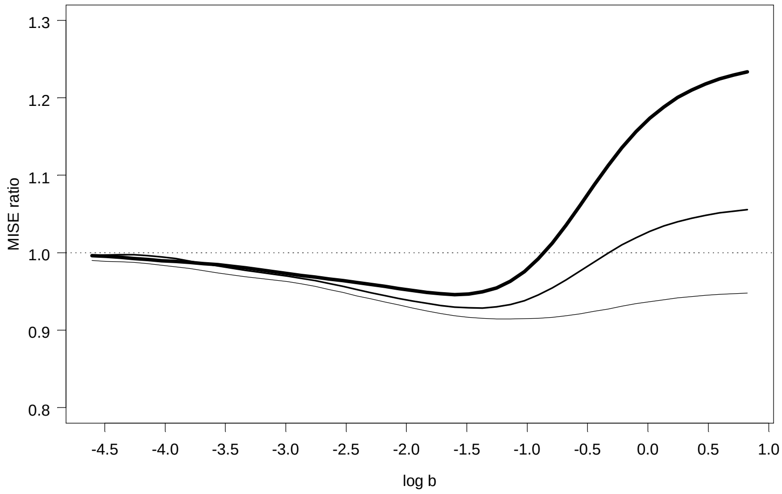

Figure 3. $MISE_w$ ratio of Λ_n^P with respect to Λ_n^{NA} for $n = 30$ (thin line), $n = 200$ (medium line) and $n = 500$ (thick line) for model 2.

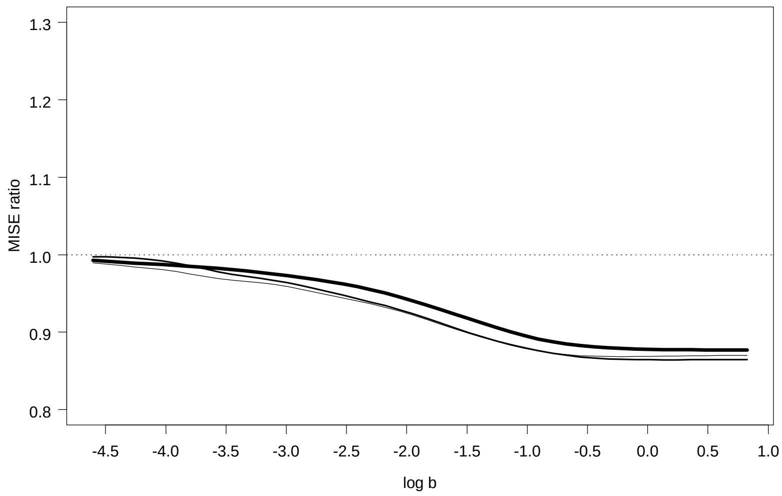

Figure 4. $MISE_w$ ratio of Λ_n^P with respect to Λ_n^{NA} for $n = 30$ (thin line), $n = 200$ (medium line) and $n = 500$ (thick line) for model 3.

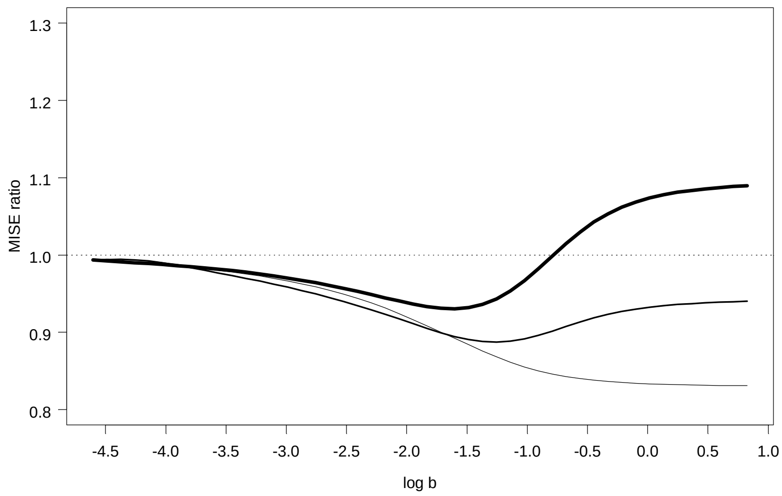

Figure 5. $MISE_w$ ratio of Λ_n^P with respect to Λ_n^{NA} for $n = 30$ (thin line), $n = 200$ (medium line) and $n = 500$ (thick line) for model 4.

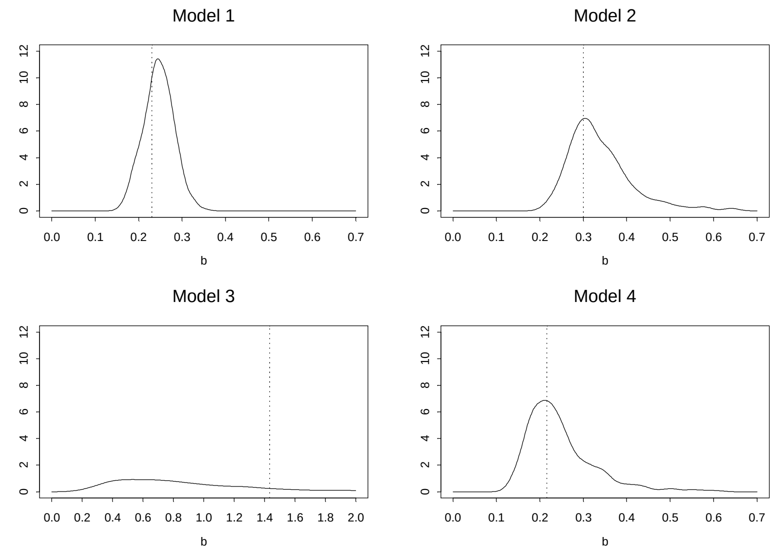

Figure 6. Kernel estimation of the density of the plug-in bandwidth selector, $\hat{b}$, for models 1-4 (solid line) and b_{OPT} bandwidth (dotted vertical line).

Table I. Parameters for the distribution of the lifetime ($W(\alpha_F, \beta_F)$) and the censoring variable ($W^{\varepsilon}(\alpha_G, \beta_G)$).

Model	α_F	β_F	α_G	β_G	ε
1	3	1	7	1	0.1
2	6	1	7	1	0.1
3	8	1	7	1	0.1
4	10	1	7	1	0.1

Table II. $MISE_w$ ratio of the presmoothed Nelson-Aalen estimator with plug-in bandwidth and the classical Nelson-Aalen estimator.

Model	$n = 30$	$n = 200$	$n = 500$
1	1.086	1.032	1.015
2	0.931	0.951	0.967
3	0.877	0.871	0.887
4	0.854	0.906	0.948