Bernoulli **8**(5), 2002, 607–625

Density and hazard estimation in censored regression models

INGRID VAN KEILEGOM[1] and NOËL VERAVERBEKE[2]

[1]*Institut de Statistique, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium. E-mail: vankeilegom@stat.ucl.ac.be*
[2]*Department of Mathematics, Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek, Belgium. E-mail: noel.veraverbeke@luc.ac.be*

Let (X, Y) be a random vector, where Y denotes the variable of interest, possibly subject to random right censoring, and X is a covariate. Consider a heteroscedastic model $Y = m(X) + \sigma(X)\varepsilon$, where the error term ε is independent of X and $m(X)$ and $\sigma(X)$ are smooth but unknown functions. Under this model, we construct a nonparametric estimator for the density and hazard function of Y given X, which has a faster rate of convergence than the completely nonparametric estimator that is constructed without making any model assumption. Moreover, the proposed estimator for the density and hazard function performs better than the classical nonparametric estimator, especially in the right tail of the distribution.

We prove the weak convergence of both the density and the hazard function estimator. The results are obtained by constructing asymptotic representations for the two estimators and by making use of work by Van Keilegom and Akritas in which an estimator of the conditional distribution of Y given X is studied under the same model assumption.

Keywords: asymptotic representation; density function; hazard rate; heteroscedastic regression; right censoring; weak convergence

1. Introduction

Let Y denote a possible transformation of the variable of interest and let X be a covariate. The response Y is allowed to be subject to random right censoring, while X is completely observed. It is assumed that the vector (X, Y) follows the heteroscedastic regression model

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1.1}$$

where ε is independent of X, the function $m(\cdot)$ is the unknown regression curve and $\sigma(\cdot)$ is a conditional scale function, representing possible heteroscedasticity. Under this model we estimate the density and hazard function of Y given X based on kernel methods.

Given that (1.1) is the correct model, the estimators for our proposed conditional density and hazard have two important advantages over the completely nonparametric estimator (which does not make use of (1.1)). First, there is a difference in the practical performance of the two types of estimators, especially in the right tail of the distribution. The completely nonparametric kernel estimator for the conditional density or hazard at a given value x of X uses only data in a neighbourhood of x and hence behaves badly whenever the censoring is

1350–7265

heavy in a neighbourhood of x. The proposed estimator is based on all available data points in the full range of X, not only those in a neighbourhood of x. Because of this, the behaviour will be satisfactory provided there is a region of the support of X where the censoring of Y is light.

The second advantage of these new estimators is their faster rate of convergence. The completely nonparametric kernel estimator for the conditional density or hazard requires smoothing both over the covariate space and over the time axis. As a consequence, the rate of convergence of such estimators is $O_P((na_n^2)^{-1/2})$ (for one-dimensional covariates), where a_n is the bandwidth used in the smoothing process (see McKeague and Utikal 1990; Li and Doss 1995). Although the proposed estimators also require smoothing in two directions, their rate of convergence is $O_p((na_n)^{-1/2})$.

Much research has been devoted to the estimation of the density and hazard function. We focus on recent papers based on kernel estimators. In the case of independently and identically distributed (i.i.d.) censored data (without covariates), Lo *et al.* (1989) estimated the density and hazard function based on the Kaplan–Meier estimator, and obtained an i.i.d. representation and the asymptotic normality of these estimators. Müller and Wang (1994) introduced boundary-corrected kernels for a better estimation of the hazard function near the boundary of the covariate space. Gefeller and Dette (1992) and Dette and Gefeller (1995) used nearest-neighbour weights for estimating the hazard function. In the context of censored response data considered in a regression model with fully observable covariates, the nonparametric estimation of the conditional hazard function was studied by McKeague and Utikal (1990), Li and Doss (1995) and Li (1997), who used counting processes to obtain the asymptotic normality of their estimator. Gray (1996) used binning methods to estimate the conditional hazard function.

The paper is organized as follows. In the next section, we give the definition of the estimator of the density and hazard function and state the main assumptions under which the results will be derived. Section 3 describes the main results, and the proofs are given in the Appendix. In Section 4 we illustrate the performance of the proposed estimators in a simulation study.

2. Definitions and assumptions

Assume the vector (X, Y) satisfies the regression model (1.1), where the functions $m(\cdot)$ and $\sigma(\cdot)$ are respectively a location and scale functional. This means that there exist functionals S and T such that $m(x) = T(F_Y(\cdot|x))$ and $\sigma(x) = S(F_Y(\cdot|x))$, with

$$T(F_{aY+b}(\cdot|x)) = aT(F_Y(\cdot|x)) + b \quad \text{and} \quad S(F_{aY+b}(\cdot|x)) = aS(F_Y(\cdot|x)),$$

for all $a \geqslant 0$, and $b \in \mathbb{R}$, where $F_Y(\cdot|x)$ denotes the distribution of Y given $X = x$. The response Y is allowed to be subject to random right censoring. Let C be the censoring random variable which is conditionally independent of Y given X, and suppose that the observable random vector is (X, Z, Δ), where $Z = \min(Y, C)$ and $\Delta = I(Y \leqslant C)$. Finally, let (X_i, Z_i, Δ_i), $i = 1, \ldots, n$, denote independent replications of (X, Z, Δ). We use the notation $F(y|x) = P(Y \leqslant y|x)$, $G(y|x) = P(C \leqslant y|x)$, $H(y|x) = P(Z \leqslant y|x)$ and $H_1(y|x) = P(Z \leqslant y,$

$\Delta = 1|x)$. Further, denote $F_e(y) = P(\varepsilon \leqslant y)$ and $G_e(y) = P((C - m(X))/\sigma(X) \leqslant y)$, and for $E = (Z - m(X))/\sigma(X)$ we use the notation $H_e(y) = P(E \leqslant y)$, $H_{e1}(y) = P(E \leqslant y, \Delta = 1)$, $H_e(y|x) = P(E \leqslant y|x)$ and $H_{e1}(y|x) = P(E \leqslant y, \Delta = 1|x)$. The density functions of the above distribution functions will be denoted with lower-case letters.

In order to estimate the conditional density function, we first need to estimate the conditional distribution function $F(y|x)$. Note that under model (1.1),

$$F(y|x) = F_e\left(\frac{y - m(x)}{\sigma(x)}\right). \tag{2.1}$$

To estimate $m(x)$ and $\sigma(x)$ we will work with the particular definitions

$$m(x) = \int_0^1 F^{-1}(s|x)J(s)\,\mathrm{d}s, \qquad \sigma^2(x) = \int_0^1 F^{-1}(s|x)^2 J(s)\,\mathrm{d}s - m^2(x), \tag{2.2}$$

where $F^{-1}(s|x) = \inf\{t;\ F(t|x) \geqslant s\}$ is the quantile function of Y given x and $J(s)$ is a given score function satisfying $\int_0^1 J(s)\,\mathrm{d}s = 1$. Note that if the assumed independence of ε and X holds for particular location and scale functionals then it holds for all location and scale functionals. Hence, working with the functionals $m(x)$ and $\sigma(x)$ in (2.2) constitutes no restriction of generality. The estimation of these functions consists in replacing the unknown distribution $F(\cdot|x)$ with the following nonparametric estimator which is due to Beran (1981):

$$\tilde{F}(y|x) = 1 - \prod_{Z_i \leqslant y, \Delta_i = 1} \left\{ 1 - \frac{W_i(x, a_n)}{\sum_{j=1}^{n} I(Z_j \geqslant Z_i) W_j(x, a_n)} \right\}, \tag{2.3}$$

where $W_i(x, a_n)$ are the Nadaraya–Watson weights

$$W_i(x, a_n) = \frac{K_1((x - X_i)/a_n)}{\sum_{j=1}^{n} K_1((x - X_j)/a_n)},$$

with K_1 a known probability density function (kernel) and $\{a_n\}$ a sequence of positive constants tending to zero as n tends to infinity, called a bandwidth sequence. This estimator reduces to the usual Kaplan and Meier (1958) estimator if all weights $W_i(x, a_n)$ equal n^{-1} (i.e. if there are no covariates). This leads to

$$\hat{m}(x) = \int_0^1 \tilde{F}^{-1}(s|x)J(s)\,\mathrm{d}s, \qquad \hat{\sigma}^2(x) = \int_0^1 \tilde{F}^{-1}(s|x)^2 J(s)\,\mathrm{d}s - \hat{m}^2(x) \tag{2.4}$$

as estimators for $m(x)$ and $\sigma(x)$. Let $\hat{E}_i = (Z_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$ and define

$$\hat{F}_e(y) = 1 - \prod_{\hat{E}_{(i)} \leqslant y, \Delta_{(i)} = 1} \left(1 - \frac{1}{n - i + 1}\right), \tag{2.5}$$

where $\hat{E}_{(i)}$ is the ith order statistic of $\hat{E}_1, \ldots, \hat{E}_n$ and $\Delta_{(i)}$ is the corresponding censoring indicator. Relation (2.1) now suggests

$$\hat{F}(y|x) = \hat{F}_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)}\right) \tag{2.6}$$

as an estimator for $F(y|x)$. The estimator is an alternative for the Beran estimator $\tilde{F}(y|x)$, which often does not behave well in the right tail if the censoring is heavy. The estimator $\hat{F}(y|x)$, however, behaves well in the right tail for all values of x, provided that there is a region of the covariate space where the censoring of Y is light. This is because in that case, the right tail of $F_e(\cdot)$ can be well estimated and hence, by (2.1), the same is true for the right tail of $F(\cdot|x)$ for all values of x, including those for which the censoring of Y is heavy. The asymptotic properties of this estimator were studied in Van Keilegom and Akritas (1999). We now estimate the conditional density function $f(y|x)$ and the hazard function $\lambda(y|x) = f(y|x)/(1 - F(y|x))$ by smoothing the estimator $\hat{F}(y|x)$ with a second kernel function K_2 (in order to make the results less technical we work with the same bandwidth as used for smoothing over the covariate space):

$$\hat{f}(y|x) = a_n^{-1}\int K_2\left(\frac{y-t}{a_n}\right) \mathrm{d}\hat{F}(t|x), \tag{2.7}$$

$$\hat{\lambda}(y|x) = \frac{\hat{f}(y|x)}{1 - \hat{F}(y|x)}. \tag{2.8}$$

The primary objective of this paper is to prove an asymptotic representation and the weak convergence of the estimators $\hat{f}(y|x)$ and $\hat{\lambda}(y|x)$.

The above estimator $\hat{f}(y|x)$ is obtained by first evaluating $F_e(y)$ in $(y - m(x))/\sigma(x)$ (which leads to $F(y|x)$) and then smoothing $F(y|x)$. We could also start by smoothing $F_e(y)$, and then evaluating the result in $(y - m(x))/\sigma(x)$. This leads to the following alternative estimator:

$$\hat{f}_{\mathrm{alt}}(y|x) = \frac{1}{\hat{\sigma}(x)}\hat{f}_e\left(\frac{y - \hat{m}(x)}{\hat{\sigma}(x)}\right),$$

where $\hat{f}_e(y) = a_n^{-1}\int K((y-t)/a_n)\,\mathrm{d}\hat{F}_e(t)$ is an estimator for the density $f_e(y)$ of ε. It can be seen that the estimators $\hat{f}(y|x)$ and $\hat{f}_{\mathrm{alt}}(y|x)$ are not asymptotically equivalent, due to an extra contribution in the asymptotic representation of $\hat{f}_{\mathrm{alt}}(y|x)$. In this paper we do not consider the estimator $\hat{f}_{\mathrm{alt}}(y|x)$, but we note that its asymptotic properties can be obtained in a very similar way as for $\hat{f}(y|x)$.

The following functions enter in the asymptotic representation for $\hat{f}(y|x)$ and $\hat{\lambda}(y|x)$:

$$
\begin{aligned}
\xi_e(z,\delta,y) &= (1-F_e(y))\left\{-\int_{-\infty}^{y\wedge z}\frac{\mathrm{d}H_{e1}(s)}{(1-H_e(s))^2}+\frac{I(z\leqslant y,\delta=1)}{1-H_e(z)}\right\},\\
\xi(z,\delta,y|x) &= (1-F(y|x))\left\{-\int_{-\infty}^{y\wedge z}\frac{\mathrm{d}H_1(s|x)}{(1-H(s|x))^2}+\frac{I(z\leqslant y,\delta=1)}{1-H(z|x)}\right\},\\
\eta(z,\delta|x) &= \int_{-\infty}^{+\infty}\xi(z,\delta,v|x)J(F(v|x))\,\mathrm{d}v\sigma^{-1}(x),\\
\zeta(z,\delta|x) &= \int_{-\infty}^{+\infty}\xi(z,\delta,v|x)J(F(v|x))\frac{v-m(x)}{\sigma(x)}\,\mathrm{d}v\sigma^{-1}(x),\\
\gamma_1(y|x) &= \int_{-\infty}^{y}\frac{h_e(s|x)}{(1-H_e(s))^2}\,\mathrm{d}H_{e1}(s)+\int_{-\infty}^{y}\frac{\mathrm{d}h_{e1}(s|x)}{1-H_e(s)},\\
\gamma_2(y|x) &= \int_{-\infty}^{y}\frac{sh_e(s|x)}{(1-H_e(s))^2}\,\mathrm{d}H_{e1}(s)+\int_{-\infty}^{y}\frac{\mathrm{d}(sh_{e1}(s|x))}{1-H_e(s)},\\
\varphi(x,z,\delta,y) &= -(1-F_e(y))\eta(z,\delta|x)\gamma_1(y|x)-(1-F_e(y))\zeta(z,\delta|x)\gamma_2(y|x).
\end{aligned}
$$

Further, we estimate $H_e(y)$ and $H_{e1}(y)$ by the empirical distribution functions

$$
\hat{H}_e(y)=n^{-1}\sum_{i=1}^{n}I(\hat{E}_i\leqslant y)\quad\text{and}\quad\hat{H}_{e1}(y)=n^{-1}\sum_{i=1}^{n}I(\hat{E}_i\leqslant y,\Delta_i=1).
$$

Let T (or T_x) be any value less than the upper bound of the support of $H_e(\cdot)$ (or $H(\cdot|x)$), such that $\inf_{x\in R_X}(1-H(T_x|x))>0$, and define $\Omega=\{(x,y);\ (y-m(x))/\sigma(x)\leqslant T\}$. For a (sub)distribution function $L(y|x)$ we will use the notation

$$
L'(y|x)=\frac{\partial}{\partial y}L(y|x),\qquad \dot{L}(y|x)=\frac{\partial}{\partial x}L(y|x),
$$

and similar notation will be used for higher-order derivatives. Further, let $\|K_i\|_2^2=\int K_i^2(u)\,\mathrm{d}u$ and $\mu_2^{K_i}=\int u^2K_i(u)\,\mathrm{d}u$, $i=1,2$.

The assumptions we need to make in the proofs of the main results are the following:

Assumption 1.

(*i*) *$na_n^5(\log a_n^{-1})^{-1}=O(1)$, and $na_n^{3+2\delta}(\log a_n^{-1})^{-1}\to\infty$ for some $\delta>0$.*
(*ii*) *The support R_X of X is bounded, convex and its interior is not empty.*
(*iii*) *For $i=1,2$, the probability density function K_i has compact support $[-L_i,L_i]$, $\int uK_i(u)\,\mathrm{d}u=0$ and K_i is twice continuously differentiable.*

Assumption 2.

(*i*) *There exist $0\leqslant s_0\leqslant s_1\leqslant 1$ such that $s_1\leqslant\inf_xF(T_x|x)$, $s_0\leqslant\inf\{s\in[0,1];J(s)\neq 0\}$, $s_1\geqslant\sup\{s\in[0,1];J(s)\neq 0\}$ and $\inf_{x\in R_X}\inf_{s_0\leqslant s\leqslant s_1}f(F^{-1}(s|x)|x)>0$.*
(*ii*) *J is twice continuously differentiable, $\int_0^1J(s)\,\mathrm{d}s=1$ and $J(s)\geqslant 0$ for all $0\leqslant s\leqslant 1$.*

Assumption 3.

(*i*) *The distribution F_X is three times continuously differentiable and* $\inf_{x \in R_X} f_X(x) > 0$.

(*ii*) *The functions m and* σ *are twice continuously differentiable and* $\inf_{x \in R_X} \sigma(x) > 0$.

(*iii*) $\mathrm{E}|E|^5 = \mathrm{E}|(Z - m(X))/\sigma(X)|^5 < \infty$.

Assumption 4. *The functions $\eta(z, \delta|x)$ and $\zeta(z, \delta|x)$ are twice continuously differentiable with respect to x and their first and second derivatives (with respect to x) are bounded uniformly in $x \in R_X$, $z < T_x$ and δ.*

Assumption 5. *For $L(y|x)$ equal to $H(y|x)$, $H_1(y|x)$, $H_e(y|x)$ or $H_{e1}(y|x)$, $L'(y|x)$ is continuous in (x, y) and $\sup_{x,y}|y^2 L'(y|x)| < \infty$, and the same holds for all other partial derivatives of $L(y|x)$ with respect to x and y up to order 4.*

3. Main results

Suppose throughout this section that Assumptions 1–5 hold.

Theorem 3.1.

$$\hat{f}(y|x) - f(y|x) = (na_n)^{-1} \sum_{i=1}^{n} \int \xi_e\left(E_i, \Delta_i, \frac{y - va_n - m(x)}{\sigma(x)}\right) \mathrm{d}K_2(v)$$

$$+ (na_n)^{-1} \sum_{i=1}^{n} K_1\left(\frac{x - X_i}{a_n}\right) g_{x,y}(Z_i, \Delta_i) + a_n^2 b_f(y|x) + r_n(y|x),$$

where

$$g_{x,y}(z, \delta) = f_X^{-1}(x)\sigma^{-1}(x)\left\{\eta(z, \delta|x) f'_e\left(\frac{y - m(x)}{\sigma(x)}\right)\right.$$

$$\left. + \zeta(z, \delta|x)\left[\frac{y - m(x)}{\sigma(x)} f'_e\left(\frac{y - m(x)}{\sigma(x)}\right) + f_e\left(\frac{y - m(x)}{\sigma(x)}\right)\right]\right\},$$

$$b_f(y|x) = \frac{1}{2}\mu_2^{K_2}\sigma^{-1}(x)\int\left[\frac{\partial}{\partial y}\mathrm{E}\left(\varphi\left(z, Z, \Delta, \frac{y - m(x)}{\sigma(x)}\right)\middle| X = u\right) f_X(u)\right]''_{u=z} \mathrm{d}z$$

$$+ \frac{1}{2}\mu_2^{K_2} f''(y|x),$$

and $\sup\{|r_n(y|x)|;\ (x, y) \in \Omega\} = o_P((na_n)^{-1/2}) + o_P(a_n^2)$.

Next we show the the local weak convergence of this density estimator in a neighbourhood $y + a_n t$ ($-\tilde{T} \leqslant t \leqslant \tilde{T}$, $\tilde{T} > 0$ arbitrary) of a fixed point y. By choosing $t = 0$, the asymptotic normality of $\hat{f}(y|x)$ follows from this result. We need to restrict attention to this local type of weak convergence, since otherwise the tightness of the process cannot be established. This is a typical feature of processes of nonparametric density, hazard or regression function estimators and can also be found in, for example, Rosenblatt (1971).

Note that Theorems 3.2 and 3.4 below are formulated both for the optimal bandwidth case ($K > 0$) and the non-optimal case ($K = 0$).

Theorem 3.2. *If* $na_n^5 = K$ *for some* $K \geqslant 0$, *then the process* $(na_n)^{1/2}(\hat{f}(y + a_n t|x) - f(y + a_n t|x))$ ($x \in R_X$ *and* $y \leqslant T\sigma(x) + m(x)$ *fixed,* $-\tilde{T} \leqslant t \leqslant \tilde{T}$, $\tilde{T} > 0$ *arbitrary*) converges weakly to a Gaussian process $Z_f(y,\, t|x)$ *with mean function*

$$\mathrm{E}(Z_f(y,\, t|x)) = K^{1/2} b_f(y|x) + \frac{1}{2} K^{1/2} \mu_2^{K_1} [\mathrm{E}(g_{x,y}(Z,\, \Delta)|X = u) f_X(u)]''_{u=x}$$

and covariance function

$$\begin{aligned}\mathrm{cov}(Z_f(y,\, t|x),\, Z_f(y,\, t'|x)) &= f_X(x)\|K_1\|_2^2 \,\mathrm{var}(g_{x,y}(Z,\, \Delta)|X = x) \\ &\quad + \sigma^{-1}(x) \int K_2(w) K_2(w + t - t')\, \mathrm{d}w \frac{h_{e1}((y - m(x))/\sigma(x))}{(1 - G_e((y - m(x))/\sigma(x)))^2}.\end{aligned}$$

Theorem 3.3.

$$\begin{aligned}\hat{\lambda}(y|x) - \lambda(y|x) &= (na_n)^{-1}(1 - F(y|x))^{-1} \sum_{i=1}^n \int \xi_e\left(E_i,\, \Delta_i,\, \frac{y - v a_n - m(x)}{\sigma(x)} \right) \mathrm{d}K_2(v) \\ &\quad + (na_n)^{-1} \sum_{i=1}^n K_1\left(\frac{x - X_i}{a_n}\right) k_{x,y}(Z_i,\, \Delta_i) + a_n^2 b_\lambda(y|x) + r_n(y|x),\end{aligned}$$

where

$$k_{x,y}(z,\, \delta) = (1 - F(y|x))^{-1} g_{x,y}(z,\, \delta) + f(y|x)(1 - F(y|x))^{-2} h_{x,y}(z,\, \delta),$$

$$h_{x,y}(z,\, \delta) = \left[\eta(z,\, \delta|x) + \zeta(z,\, \delta|x)\frac{y - m(x)}{\sigma(x)}\right] f_e\left(\frac{y - m(x)}{\sigma(x)}\right) f_X^{-1}(x),$$

$$\begin{aligned}b_\lambda(y|x) &= (1 - F(y|x))^{-1} b_f(y|x) + \frac{1}{2}\mu_2^{K_1}(1 - F(y|x))^{-2} f(y|x) \\ &\quad \times \int \left\{ \mathrm{E}\left[\varphi\left(z,\, Z,\, \Delta,\, \frac{y - m(x)}{\sigma(x)}\right) \middle| X = u\right] f_X(u) \right\}''_{u=z} \mathrm{d}z,\end{aligned}$$

and $\sup\{|r_n(y|x)|;\ (x,\, y) \in \Omega\} = o_P((na_n)^{-1/2}) + o_P(a_n^2)$.

Finally, we give the weak convergence of the hazard estimator $\hat{\lambda}(y|x)$. The proof is very similar to that for the density function (see Theorem 3.2) and is therefore not given.

Theorem 3.4. *If* $na_n^5 = K$ *for some* $K \geqslant 0$, *then the process* $(na_n)^{1/2}(\hat{\lambda}(y + a_n t|x)$

$-\lambda(y+a_n t|x))$ ($x \in R_X$ *and* $y \leqslant T\sigma(x)+m(x)$ *fixed,* $-\tilde{T} \leqslant t \leqslant \tilde{T}$, $\tilde{T}>0$ *arbitrary*) converges weakly to a Gaussian process $Z_\lambda(y, t|x)$ *with mean function*

$$\mathrm{E}(Z_\lambda(y,\, t|x)) = K^{1/2} b_\lambda(y|x) + \frac{1}{2} K^{1/2} \mu_2^{K_1} [\mathrm{E}(k_{x,y}(Z,\, \Delta)|X=u) f_X(u)]''_{u=x}$$

and covariance function

$$\begin{aligned}\operatorname{cov}(Z_\lambda(y,\, t|x),\, Z_\lambda(y,\, t'|x)) &= f_X(x)\|K_1\|_2^2 \operatorname{var}(k_{x,y}(Z,\, \Delta)|X=x)\\ &\quad + \sigma^{-1}(x) \int K_2(w) K_2(w+t-t')\,\mathrm{d}w \frac{h_{e1}((y-m(x))/\sigma(x))}{(1-H_e((y-m(x))/\sigma(x)))^2}.\end{aligned}$$

4. Simulations

In this section we carry out a number of simulations, in which the small-sample performance of the estimator $\hat{f}(y|x)$ is compared with that of the completely nonparametric estimator

$$\tilde{f}(y|x) = a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) \mathrm{d}\tilde{F}(t|x),$$

which is based on Beran's estimator $\tilde{F}(y|x)$ defined in (2.3), instead of the estimator $\hat{F}(y|x)$. Assume that the covariate X is uniformly distributed on the interval [0, 1]. We consider Weibull survival and censoring distributions

$$\begin{aligned}(Y|X=x) &\sim \mathrm{Weibull}(c_0 + c_1 x + c_2 x^2,\, d),\\ (C|X=x) &\sim \mathrm{Weibull}(e_0 + e_1 x + e_2 x^2,\, d),\end{aligned}$$

i.e. $F(y|x) = 1 - \exp(-(c_0 + c_1 x + c_2 x^2) y^d)$ and similarly for $G(y|x)$, for certain values for the parameters. From the conditional independence of Y and C for given X, it follows that $(Z|X=x) \sim \mathrm{Weibull}(c_0 + e_0 + (c_1+e_1)x + (c_2+e_2)x^2,\, d)$ and that

$$P(\Delta = 0|x) = \frac{e_0 + e_1 x + e_2 x^2}{c_0 + e_0 + (c_1+e_1)x + (c_2+e_2)x^2}. \tag{4.1}$$

It is easily shown that if $m(x)$ equals the conditional mean and $\sigma(x)$ the conditional standard deviation, then

$$P(\varepsilon \leqslant y|x) = 1 - \exp(-\{y[\Gamma(1+2d^{-1}) - \Gamma^2(1+d^{-1})]^{1/2} + \Gamma(1+d^{-1})\}^d), \tag{4.2}$$

which is independent of x. Hence, model (1.1) is satisfied for these distributions. We simulated samples of size $n = 150$ and chose $x = 0.5$. The results are obtained from 1000 simulation runs. The kernel function K is the biquadratic kernel $K(u) = \frac{15}{16}(1-u^2)^2 I(|u| \leqslant 1)$.

Although, in the asymptotic theory, we used the same bandwidth for smoothing over x and y, we prefer here to work with two bandwidths, as in practice the range of x might be very different from that of y. Let a_n be the bandwidth for x, and let b_n be the bandwidth for y. In order to select these bandwidths in an appropriate way for a fixed x and y, we consider a grid of values of a_n and b_n, approximate the mean square error (MSE) of $\tilde{f}(y|x)$ (or $\hat{f}(y|x)$) for each pair (a_n, b_n) by the MSE of the 1000 simulation outcomes, and select the pair (a_n, b_n) for which the MSE of $\tilde{f}(y|x)$ (or $\hat{f}(y|x)$) is minimal.

For the location and scale functions $m(x)$ and $\sigma(x)$, defined in (2.2), we need to choose a proper score function J. Since $\tilde{F}^{-1}(s|x)$ is only defined for $s \leqslant \tilde{F}(+\infty|x)$, $J(s)$ must be zero from that point on. We therefore worked with trimmed means and variances determined by $J(s) = I(0 \leqslant s \leqslant b)/b$, where $b = \min_i \tilde{F}(+\infty|X_i)$.

The first set of parameters we consider is the following: $c_0 = 1$, $c_2 = 20$, $e_0 = 0.1$ $e_1 = 0$, $e_2 = 25$ and $d = 3$. The parameter c_1 is determined such that the probability of censoring at $x = 0.5$ (given by (4.1)) equals 0.3 or 0.6. Figure 1 shows how the probability of censoring varies with x for both situations. The simulation results for this set-up are shown in Table 1 (where $\xi(p|x) = F^{-1}(p|x)$). The bias, variance and MSE of the two estimators are obtained from the 1000 simulation runs. We see that for both situations and for all values of y, the estimator $\hat{f}(y|x)$ behaves considerably better than the completely nonparametric estimator $\tilde{f}(y|x)$, both with respect to the bias and the variance.

Next, we consider the parameters $c_0 = 1$, $c_1 = 2$, $c_2 = 3$, $e_0 = 1$, $e_1 = 2$, $e_2 = 3$ and $d = 3$. This leads to a constant censoring rate of 0.5 for all x. The results for this set-up are summarized in Table 2. Here too the estimator $\hat{f}(y|x)$ outperforms $\tilde{f}(y|x)$. It should be

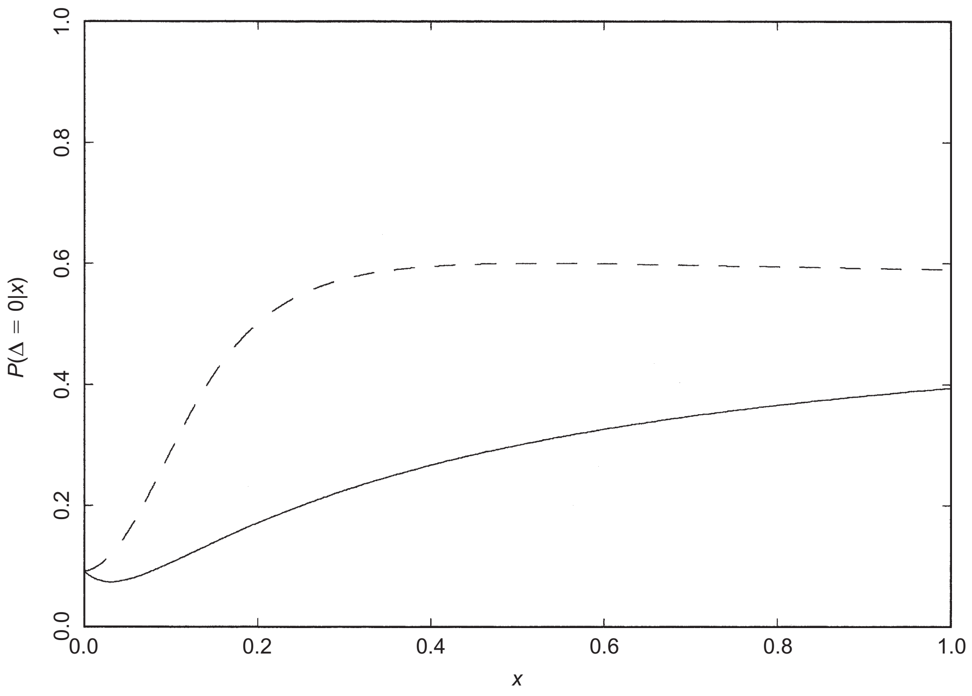

Figure 1. Graph of the probability of censoring versus x when $c_0 = 1$, $c_2 = 20$, $e_0 = 0.1$, $e_1 = 0$, $e_2 = 25$, $d = 3$ and $P(\Delta = 0|0.5) = 0.3$ (solid curve) and 0.6 (dashed curve).

Table 1. Mean square error of the two estimators when $c_0 = 1$, $c_2 = 20$, $e_0 = 0.1$, $e_1 = 0$, $e_2 = 25$ and $d = 3$

			Bias		Variance		MSE	
$P(\Delta = 0 \mid x)$	y	$f(y\|x)$	$\tilde{f}(y\|x)$	$\hat{f}(y\|x)$	$\tilde{f}(y\|x)$	$\hat{f}(y\|x)$	$\tilde{f}(y\|x)$	$\hat{f}(y\|x)$
0.3	$\xi(0.20\|x)$	2.169	−0.2133	−0.1598	0.0587	0.0607	0.1042	0.0862
	$\xi(0.40\|x)$	2.825	−0.2933	−0.2161	0.1219	0.1004	0.2080	0.1471
	$\xi(0.60\|x)$	2.780	−0.3549	−0.1943	0.0998	0.0948	0.2257	0.1325
	$\xi(0.80\|x)$	2.024	−0.2024	−0.1252	0.0425	0.0265	0.0835	0.0422
	$\xi(0.90\|x)$	1.285	−0.0158	−0.0148	0.0030	0.0017	0.0033	0.0019
	$\xi(0.95\|x)$	0.766	−0.0043	−0.0047	0.0001	0.0001	0.0001	0.0001
0.6	$\xi(0.20\|x)$	1.428	−0.1900	−0.1685	0.0479	0.0498	0.0840	0.0782
	$\xi(0.40\|x)$	1.861	−0.3451	−0.2390	0.0915	0.0848	0.2105	0.1419
	$\xi(0.60\|x)$	1.831	−0.4171	−0.2318	0.0845	0.0636	0.2585	0.1173
	$\xi(0.80\|x)$	1.333	−0.2687	−0.1438	0.0377	0.0190	0.1099	0.0397
	$\xi(0.90\|x)$	0.846	−0.0387	−0.0102	0.0043	0.0020	0.0058	0.0021
	$\xi(0.95\|x)$	0.504	0.0043	−0.0035	0.0002	0.0002	0.0002	0.0002

Table 2. Mean square error of the two estimators when $c_0 = 1$, $c_1 = 2$, $c_2 = 3$, $e_0 = 1$, $e_1 = 2$, $e_2 = 3$ and $d = 3$

		Bias		Variance		MSE	
y	$f(y\|x)$	$\tilde{f}(y\|x)$	$\hat{f}(y\|x)$	$\tilde{f}(y\|x)$	$\hat{f}(y\|x)$	$\tilde{f}(y\|x)$	$\hat{f}(y\|x)$
$\xi(0.20\|x)$	1.237	−0.1104	−0.1164	0.0171	0.0149	0.0293	0.0284
$\xi(0.40\|x)$	1.612	−0.1769	−0.1294	0.0371	0.0412	0.0684	0.0580
$\xi(0.60\|x)$	1.586	−0.2055	−0.1441	0.0396	0.0316	0.0818	0.0524
$\xi(0.80\|x)$	1.154	−0.1420	−0.0965	0.0163	0.0157	0.0364	0.0250
$\xi(0.90\|x)$	0.733	−0.0151	−0.0179	0.0030	0.0035	0.0032	0.0038
$\xi(0.95\|x)$	0.437	0.0904	0.1062	0.0048	0.0020	0.0130	0.0133

noted, however, that the difference in performance between the two estimators is more extreme in Table 1 than in Table 2. The reason for this lies in the fact that in the set-up of Table 1 there is a region where the censoring is quite light (namely for small x-values, see Figure 1), while in the second set-up the censoring is uniform over x. The data in the region of light censoring are used by the estimator $\hat{f}(y|x)$, but not by $\tilde{f}(y|x)$ (since $\tilde{f}(y|x)$ only uses data in a small window around $x = 0.5$). The use of data from a region where there is not much censoring can considerably improve the estimator, especially in the right tail of the distribution, where uncensored data are typically rare. This explains why the existence of a region of light censoring has an influence on the performance of $\hat{f}(y|x)$.

Finally, the above procedure for optimizing the choice of a_n and b_n by minimizing the MSE of the simulation outcomes can be adapted to real data by creating bootstrap samples and by minimizing the MSE over a large number of these bootstrap samples. Van Keilegom and Veraverbeke (1997) propose a bootstrap procedure for the general context of nonparametric regression with censored data which can be used for this purpose. In view of the complexity of the asymptotic variance, this bootstrap procedure can also be used for the construction of confidence intervals and for testing hypotheses concerning the density and the hazard function.

Appendix

We start with a lemma in which it is shown that a certain process of the random variables $\hat{E}_i = (Z_i - \hat{m}(X_i))/\hat{\sigma}(X_i)$, $i = 1, \ldots, n$, is asymptotically equivalent to the same process evaluated at $E_i = (Z_i - m(X_i))/\sigma(X_i)$, $i = 1, \ldots, n$. This lemma is crucial to the proofs of the main results, since it allows results for E to be carried over to analogous results for $\hat{E}$. The proof of this lemma is complex and technical. The main reason for this complexity is that it is hard to replace $\hat{m}$ and $\hat{\sigma}$ respectively by m and σ, given that they appear inside an indicator function, which is not differentiable. We sketch the proof of the lemma. Details can be found in Van Keilegom and Veraverbeke (2000).

Lemma A.1. *Suppose that Assumptions* 1–5 *hold. Then*

$$\sup_{-\infty<y<+\infty}\left|(na_n)^{-1}\sum_{i=1}^{n}\int\{I(\hat{E}_i \leqslant y - va_n) - I(E_i \leqslant y - va_n) - P(\hat{E} \leqslant y - va_n|\mathcal{X}_n) + P(E \leqslant y - va_n)\}\,\mathrm{d}K_2(v)\right| = o_P((na_n)^{-1/2}), \tag{A.1}$$

where $P(\hat{E} \leqslant y|\mathcal{X}_n)$ *is the distribution of* $\hat{E} = (Z - \hat{m}(X))/\hat{\sigma}(X)$ *conditioning on* (X_j, Z_j, Δ_j), $j = 1, \ldots, n$.

Proof. We use the notation $y_v = y - va_n$, $d_{n1}(x) = (\hat{m}(x) - m(x))/\sigma(x)$ and $d_{n2}(x) = (\hat{\sigma}(x) - \sigma(x))/\sigma(x)$ throughout the proof. First it can be seen that, up to a term of order $o_P((na_n)^{-1/2})$ uniformly in y, the expression between modulus bars in (A.1) is equal to

$$\begin{aligned}(na_n)^{-1}\sum_{i=1}^{n}\int\{&I(E_i \leqslant d_{n1}(X_i) + y_v(1 + d_{n2}(X_i))) - P(E \leqslant d_{n1}(X) + y_v(1 + d_{n2}(X)))\\ &- J_i(y_v, d_{n1}(X_i), d_{n2}(X_i)) + \mathrm{E}[J(y_v, d_{n1}(X), d_{n2}(X))]\\ &- I(E_i \leqslant y_v) + P(E \leqslant y_v) + J_i(y_v, 0, 0) - \mathrm{E}[J(y_v, 0, 0)]\}\,\mathrm{d}K_2(v),\end{aligned} \tag{A.2}$$

where $J_i(y, a, b)$ equals 0 for $y \leqslant (E_i - a)/(1 + b) - a_n^{1/2}$, 1 for $y \geqslant (E_i - a)/(1 + b)$

$+a_n^{1/2}$, and a linear function starting at 0 and ending at 1 on the interval $[(E_i - a)/(1+b) - a_n^{1/2},\ (E_i - a)/(1+b) + a_n^{1/2}]$.

The proof is based on results in van der Vaart and Wellner (1996). Let

$$Z_{ni}(y, d_1, d_2) = n^{-1/2} a_n^{1/2-\delta} \int \{ I(E_i \leqslant d_1(X_i) + y_v(1 + d_2(X_i))) - J_i(y_v, d_1(X_i), d_2(X_i)) - I(E_i \leqslant y_v) + J_i(y_v, 0, 0)\} \, \mathrm{d}K_2(v),$$

$i = 1, \ldots, n$. We consider $\sum_{i=1}^n [Z_{ni} - \mathrm{E}(Z_{ni})]$ as a process over the class

$$\mathcal{F} = \{(y, d_1, d_2);\ -\infty < y < +\infty,\ d_i(X) = e_n \tilde{d}_i(X) \text{ and } \tilde{d}_i(X) \in C_1^\delta(R_X),\ i = 1, 2\},$$

where $e_n = (n a_n^{1+\delta})^{-1/2} (\log a_n^{-(1+\delta)})^{1/2}$, $\delta > 0$, and where $C_1^\delta(R_X)$ is the class of all differentiable functions d defined on the domain R_X of X such that

$$\|d\|_\delta = \sup_x |d(x)| + \sup_{x,x'} \frac{|d(x) - d(x')|}{|x - x'|^\delta} \leqslant 1.$$

It can be shown that, for $i = 1, 2$, $P(e_n^{-1} d_{ni} \in C_1^\delta(R_X)) \to 1$ as $n \to \infty$. We will show that $\sum_{i=1}^n (Z_{ni} - \mathrm{E} Z_{ni})$ converges weakly to a Gaussian process. From this is follows that (A.2) is $O_P(n^{-1/2} a_n^{-1/2+\delta}) = o_P((n a_n)^{-1/2})$ uniformly in y.

To show the weak convergence of the given process, we will verify the conditions of Theorem 2.11.9 in van der Vaart and Wellner (1996). Proving the convergence of the marginals of the process $\sum_{i=1}^n (Z_{ni} - \mathrm{E} Z_{ni})$ is achieved by showing that Lyapunov's ratio

$$\frac{\sum_{i=1}^n \mathrm{E}|Z_{ni} - \mathrm{E}[Z_{ni}]|^3}{\left[\sum_{i=1}^n \mathrm{var}(Z_{ni})\right]^{3/2}} \leqslant \frac{\sup_{y, d_1 d_2} |Z_{n1}(y, d_1, d_2)|}{\left[\sum_{i=1}^n \mathrm{var}(Z_{ni})\right]^{1/2}} \tag{A.3}$$

tends to zero as n tends to infinity. Lengthy calculations show that $\sum_{i=1}^n \mathrm{var}(Z_{ni}) = O(n^{-1} a_n^{-4-3\delta} \log a_n^{-(1+\delta)})$ and hence (A.3) is $o(1)$. It remains to calculate the bracketing number $N_{[\,]}(\varepsilon, \mathcal{F}, L_2^n)$, which is the minimal number of sets N_ε in a partition of $\mathcal{F}$ into sets $\mathcal{F}_{\varepsilon j}^n$ such that, for every partitioning set $\mathcal{F}_{\varepsilon j}^n$,

$$\sum_{i=1}^n \mathrm{E}\{\sup |Z_{ni}(y, d_1, d_2) - Z_{ni}(y', d_1', d_2')|^2\} \leqslant \varepsilon^2,$$

where the supremum is taken over all $(y, d_1, d_2), (y', d_1', d_2') \in \mathcal{F}_{\varepsilon j}^n$. Partition the line into $m_1 = O(\varepsilon^{-17/3})$ subintervals y_i, $i = 1, \ldots, m_1$: let $y_0 = -\infty$, $y_1 = -\varepsilon^{-2/3}$, $y_{m_1 - 1} = \varepsilon^{-2/3}$, $y_{m_1} = +\infty$ and divide the interval $[y_1, y_{m_1-1}]$ into $m_1 - 2$ equally spaced subintervals of length $O(\varepsilon^5)$. Next, note that in Corollary 2.7.2 of van der Vaart and Wellner (1996), it is stated that $m_2 = N_{[\,]}(\varepsilon, C_1^\delta(R_X), L_2(P)) \leqslant \exp(K\varepsilon^{-1/\delta})$ and $m_3 = N_{[\,]}(\varepsilon^{5/3}, C_1^\delta(R_X), L_2(P)) \leqslant \exp(K\varepsilon^{-5/(3\delta)})$. Let $d_{j1}^L \leqslant d_{j1}^U$, $j = 1, \ldots, m_2$, be the functions defining the m_2 ε-brackets for $C_1^\delta(R_X)$ and let $d_{k2}^L \leqslant d_{k2}^U$, $k = 1, \ldots, m_3$), be the functions defining the m_3 $\varepsilon^{5/3}$-brackets for $C_1^\delta(R_X)$. Lengthy and technical calculations show that the brackets for the considered process are given by $[y_i, y_{i+1}] \times [d_{j1}^L, d_{j1}^U] \times [d_{k2}^L, d_{k2}^U]$ $(i = 1, \ldots, m_1, j = 1, \ldots, m_2, k = 1, \ldots, m_3)$. Hence, the bracketing number for the process considered is $O(\varepsilon^{-17/3} \exp(K\varepsilon^{-5/(3\delta)}))$, which satisfies $\int_0^{\delta_n} \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{F}, L_2^n)}\, \mathrm{d}\varepsilon \to 0$ for all $\delta_n \downarrow 0$. This shows that the conditions of Theorem 2.11.9 in van der Vaart and Wellner (1996) are satisfied, from which the weak convergence of the process $\sum_{i=1}^n (Z_{ni} - \mathrm{E} Z_{ni})$ follows. $\square$

Proof of Theorem 3.1. Write

$$
\begin{aligned}
\hat{f}(y|x) - f(y|x) &= a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) \mathrm{d}\hat{F}(t|x) - f(y|x) \\
&= a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) \mathrm{d}[\hat{F}(t|x) - F(t|x) + a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) \mathrm{d}F(t|x) - f(y|x) \\
&= -a_n^{-1} \int [\hat{F}(t|x) - F(t|x)]\, \mathrm{d}K_2\left(\frac{y-t}{a_n}\right) + a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) \mathrm{d}F(t|x) - f(y|x) \\
&= a_n^{-1} \int [\hat{F}(y - va_n|x) - F(y - va_n|x)]\, \mathrm{d}K_2(v) + a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) \mathrm{d}F(t|x) - f(y|x).
\end{aligned}
\tag{A.4}
$$

The second term in (A.4) equals

$$
\begin{aligned}
a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) [f(t|x) - f(y|x)]\, \mathrm{d}t &= f'(y|x) a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) (t-y)\, \mathrm{d}t \\
&\quad + \frac{1}{2} f''(y|x) a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) (t-y)^2\, \mathrm{d}t \\
&\quad + a_n^{-1} \int K_2\left(\frac{y-t}{a_n}\right) o(|t-y|^2)\, \mathrm{d}t \\
&= \frac{1}{2} a_n^2 \mu_2^{K_2} f''(y|x) + o(a_n^2).
\end{aligned}
$$

The derivation of the first term in (A.4) is quite lengthy and requires a substantial number of technical lemmas. We restrict ourselves to a sketch of the proof, and refer the interested reader to the Appendix of the technical report of Van Keilegom and Veraverbeke (2000), where a detailed derivation can be found. Note that the first term of (A.4) equals

$$
\begin{aligned}
a_n^{-1} \int \Bigg\{ &\left[\hat{F}_e\left(\frac{y - va_n - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - va_n - \hat{m}(x)}{\hat{\sigma}(x)}\right) \right] \\
&+ \left[F_e\left(\frac{y - va_n - \hat{m}(x)}{\hat{\sigma}(x)}\right) - F_e\left(\frac{y - va_n - m(x)}{\sigma(x)}\right) \right] \Bigg\}\, \mathrm{d}K_2(v).
\end{aligned}
\tag{A.5}
$$

The second term of (A.5) can be written as

$$
(na_n)^{-1} \sum_{i=1}^n K_1\left(\frac{x - X_i}{a_n}\right) g_{x,y}(Z_i, \Delta_i) + o_P((na_n)^{-1/2}),
$$

uniformly for $(x, y) \in \Omega$, after applying a representation for $\hat{m}(x)$ and $\hat{\sigma}(x)$ given by Proposition 4.6 in Van Keilegom and Akritas (1999). For the first term of (A.5), Lemma A.1 plays a crucial role, as it implies, after some technical calculations, that, up to a bias term of order $O(a_n^2)$, the estimator $\hat{F}_e$ (which is constructed with $\hat{E}_i$, $i = 1, \ldots, n$) can be replaced by the analogous estimator based on E_i, $i = 1, \ldots, n$. Hence, a representation for the latter Kaplan–Meier estimator given in Lo and Singh (1986) implies that the first term of (A.5) can be written as

$$(na_n)^{-1} \sum_{i=1}^{n} \int \xi_e \left(E_i, \Delta_i, \frac{y - va_n - m(x)}{\sigma(x)} \right) \mathrm{d}K_2(v)$$
$$+ \frac{1}{2} a_n^2 \mu_2^{K_2} \sigma^{-1}(x) \int \left[\frac{\partial}{\partial y} \mathrm{E} \left\{ \varphi \left(z, Z, \Delta, \frac{y - m(x)}{\sigma(x)} \right) \middle| X = u \right\} f_X(u) \right]''_{u=z} \mathrm{d}z$$
$$+ o_P((na_n)^{-1/2}) + o_P(a_n^2),$$

uniformly for $(x, y) \in \Omega$, from which the result follows. □

Proof of Theorem 3.2. To show the weak convergence of the main term in the representation, use will be made of Theorem 2.11.9 in van der Vaart and Wellner (1996). We start by calculating the covariance of the process. That

$$\mathrm{cov}\left\{ K_1\left(\frac{x - X}{a_n}\right) g_{x, y + a_n t}(Z, \Delta),\ K_1\left(\frac{x - X}{a_n}\right) g_{x, y + a_n t'}(Z, \Delta) \right\}$$
$$= a_n f_X(x) \|K_1\|_2^2 \,\mathrm{var}\{g_{x,y}(Z, \Delta) | X = x\} + O(a_n^2)$$

follows easily from a Taylor expansion. Next, let $e_v(t) = (y + (t - v)a_n - m(x))/\sigma(x)$ and $e = (y - m(x))/\sigma(x)$. Using the expression for the covariance of the function ξ_e given in Lo and Singh (1986), we have

$$\mathrm{cov}\left[\int \xi_e(E, \Delta, e_v(t))\, \mathrm{d}K_2(v),\ \int \xi_e(E, \Delta, e_v(t'))\, \mathrm{d}K_2(v) \right]$$
$$= \iint \mathrm{E}\{\xi_e(E, \Delta, e_v(t)) \xi_e(E, \Delta, e_w(t'))\}\, \mathrm{d}K_2(v)\, \mathrm{d}K_2(w)$$
$$= \iint (1 - F_e(e_v(t)))(1 - F_e(e_w(t'))) \int_{-\infty}^{e_v(t) \wedge e_w(t')} \frac{\mathrm{d}H_{e1}(s)}{(1 - H_e(s))^2}\, \mathrm{d}K_2(v)\, \mathrm{d}K_2(w)$$
$$= \iint_{-\infty}^{w + t - t'} (1 - F_e(e_v(t)))\, \mathrm{d}K_2(v) (1 - F_e(e_w(t'))) \int_{-\infty}^{e_w(t')} \frac{\mathrm{d}H_{e1}(s)}{(1 - H_e(s))^2}\, \mathrm{d}K_2(w)$$
$$+ \iint_{-\infty}^{v - t + t'} (1 - F_e(e_w(t')))\, \mathrm{d}K_2(w) (1 - F_e(e_v(t))) \int_{-\infty}^{e_v(t)} \frac{\mathrm{d}H_{e1}(s)}{(1 - H_e(s))^2}\, \mathrm{d}K_2(v). \qquad \text{(A.6)}$$

The first term of (A.6) equals

$$\int K_2(w+t-t')(1-F_e(e_w(t')))^2 \int_{-\infty}^{e_w(t')} \frac{\mathrm{d}H_{e1}(s)}{(1-H_e(s))^2}\,\mathrm{d}K_2(w)$$

$$-\frac{a_n}{\sigma(x)}\int\int_{-\infty}^{w+t-t'} K_2(v)f_e(e_v(t))\,\mathrm{d}v(1-F_e(e_w(t')))\int_{-\infty}^{e_w(t')}\frac{\mathrm{d}H_{e1}(s)}{(1-H_e(s))^2}\,\mathrm{d}K_2(w).$$

Applying integration by parts to both terms above, we obtain

$$-\int K_2(w)K_2'(w+t-t')(1-F_e(e_w(t')))^2\int_{-\infty}^{e_w(t')}\frac{\mathrm{d}H_{e1}(s)}{(1-H_e(s))^2}\,\mathrm{d}w$$

$$-\frac{a_n}{\sigma(x)}\int K_2(w)K_2(w+t-t')\,\mathrm{d}w(1-F_e(e))f_e(e)\int_{-\infty}^{e}\frac{\mathrm{d}H_{e1}(s)}{(1-H_e(s))^2}$$

$$+\frac{a_n}{\sigma(x)}\int K_2(w)K_2(w+t-t')\,\mathrm{d}w(1-F_e(e))^2\frac{h_{e1}(e)}{(1-H_e(e))^2}+O(a_n^2).$$

Using the same techniques, we find that the second term of (A.6) equals

$$\int K_2(w)K_2'(w+t-t')(1-F_e(e_w(t')))^2\int_{-\infty}^{e_w(t')}\frac{\mathrm{d}H_{e1}(s)}{(1-H_e(s))^2}\,\mathrm{d}w$$

$$+\frac{a_n}{\sigma(x)}\int K_2(w)K_2(w+t-t')\,\mathrm{d}w(1-F_e(e))f_e(e)\int_{-\infty}^{e}\frac{\mathrm{d}H_{e1}(s)}{(1-H_e(s))^2}+O(a_n^2).$$

Hence, (A.6) is equal to

$$\frac{a_n}{\sigma(x)}\int K_2(w)K_2(w+t-t')\,\mathrm{d}w\frac{h_{e1}((y-m(x))/\sigma(x))}{(1-G_e((y-m(x))/\sigma(x)))^2}+O(a_n^2).$$

Next we need to calculate $\mathrm{cov}(K_1((x-X)/a_n)g_{x,y+a_nt}(Z,\Delta), \int\xi_e(E,\Delta,e_v(t'))\,\mathrm{d}K_2(v))$. Since the calculations are similar to those above, we confine ourselves to giving an outline. Since the function $g_{x,y}(z,\delta)$ is formed from the functions $\eta(z,\delta|x)$ and $\zeta(z,\delta|x)$, which contain the function $\xi(z,\delta,s|x)$, we need to calculate the covariance of $\xi(Z,\Delta,s|x)$ and $\xi_e(E,\Delta,y)$, given X. Next, integration by parts over the v variable shows that the order of the covariance is $O(a_n^2)$.

Let us now calculate the bias of the second term of the representation in Theorem 3.1 (the first one is unbiased). This follows from a Taylor expansion:

$$a_n^{-1}\int K_1\left(\frac{x-u}{a_n}\right)\mathrm{E}(g_{x,y+a_nt}(Z,\Delta)|X=u)\,\mathrm{d}F_X(u)$$

$$=\frac{1}{2}a_n^2\mu_2^{K_1}[\mathrm{E}(g_{x,y}(Z,\Delta)|X=u)f_X(u)]''|_{u=x}+o(a_n^2),$$

since $\mathrm{E}(g_{x,y}(Z,\Delta)|X=x)=0$. Next, we show the convergence of the finite-dimensional distributions. By the Cramér–Wold device, we need to show the convergence of any linear combination of the functions $(na_n)^{1/2}(\hat{f}(y+a_nt_j|x)-f(y+a_nt_j|x))$ $(-\tilde{T}\leqslant t_1,\ldots,t_k\leqslant\tilde{T}$

arbitrary, k arbitrary). We do this by verifying Lyapunov's condition. Since the functions $g_{x,y}$ and ξ_e are bounded, the Lyapunov ratio is easily seen to be $O((na_n)^{-1/2}) = o(1)$.

It remains to verify the three displayed conditions in Theorem 2.11.9 in van der Vaart and Wellner (1996). The first one is obviously satisfied, since the functions $g_{x,y}(z, \delta)$ and $\xi_e(z, \delta, y)$ are bounded. We will next show that

$$\int_0^{\delta_n} \sqrt{\log N_{[\,]}(\varepsilon, \mathcal{F}, L_2(P))}\, \mathrm{d}\varepsilon \to 0$$

for every $\delta_n \downarrow 0$, where $N_{[\,]}$ is the bracketing number, $\mathcal{F} = \{W_n(t); -\tilde{T} \leqslant t \leqslant \tilde{T}\}$, with

$$\begin{aligned} W_n(t) &= \int \xi_e\left(E, \Delta, \frac{y + (t - v)a_n - m(x)}{\sigma(x)}\right) \mathrm{d}K_2(v) + K_1\left(\frac{x - X}{a_n}\right) g_{x, y+a_n t}(Z, \Delta) \\ &= W_{n1}(t) + W_{n2}(t), \end{aligned}$$

P is the probability measure corresponding to the joint distribution of (Z, Δ, X), and $L_2(P)$ is the L_2 norm. Partition $[-\tilde{T}, \tilde{T}]$ into $O(\varepsilon^{-1})$ subintervals $[t_j, t_{j+1}]$ of length at most $K\varepsilon$ for some $K > 0$. We will show that

$$a_n^{-1} \mathrm{E} \sup_{t_j \leqslant t, t' \leqslant t_{j+1}} |W_n(t) - W_n(t')|^2 \leqslant \varepsilon^2. \tag{A.7}$$

This implies not only the third condition in van der Vaart and Wellner (1996) but also the second, since the partitions are independent of n. It is easily seen that $|W_{n2}(t) - W_{n2}(t')| \leqslant K\varepsilon a_n$ whenever $|t - t'| \leqslant K\varepsilon$, and hence (A.7) is satisfied for W_n replaced by W_{n2}. Since the function $\xi_e(z, \delta, y)$ consists of three terms, $W_{n1}(t)$ can also be decomposed into three terms. The most difficult term to deal with is the second one, which equals $Z_n(t) - \mathrm{E}[Z_n(t)]$, where

$$Z_n(t) = \int \frac{I(E \leqslant e_v(t), \Delta = 1)}{1 - G_e(e_v(t))}\, \mathrm{d}K_2(v).$$

We will prove that (A.7) is satisfied if $W_n(t)$ is replaced by $Z_n(t)$ (the derivation for $\mathrm{E}[Z_n(t)]$ follows immediately, using integration by parts). Consider

$$\begin{aligned} &Z_n(t) - Z_n(t') \\ &\quad = \int \frac{I(E \leqslant e_v(t), \Delta = 1) - I(E \leqslant e_v(t'), \Delta = 1)}{1 - G_e(e_v(t))}\, \mathrm{d}K_2(v) \\ &\qquad + \int I(E \leqslant e_v(t'), \Delta = 1)\left[\frac{1}{1 - G_e(e_v(t))} - \frac{1}{1 - G_e(e_v(t'))}\right] \mathrm{d}K_2(v). \end{aligned}$$

We concentrate on the first term which, using the notation $E_n(x, y) = (E\sigma(x) - y + m(x)/a_n$ and choosing $t' \leqslant t$, equals

$$
\begin{aligned}
&I(\Delta=1)\int_{t'-E_n(x,y)}^{t-E_n(x,y)} \frac{1}{1-G_e(e_v(t))}\,\mathrm{d}K_2(v)\\
&\quad = I(\Delta=1)\left\{\frac{K_2(t-E_n(x,\,y))}{1-G_e(E)} - \frac{K_2(t'-E_n(x,\,y))}{1-G_e(E+\sigma^{-1}(x)(t-t')a_n}\right\}\\
&\quad\quad + \frac{a_n}{\sigma(x)} I(\Delta=1)\int_{t'-E_n(x,y)}^{t-E_n(x,y)} \frac{K_2(v)}{(1-G_e(e_v(t)))^2}\, g_e(e_v(t))\,\mathrm{d}v.
\end{aligned}
$$

The last term above is not more than $K\varepsilon a_n$, which is sufficiently small for the required condition to be satisfied. For the first term above it suffices to consider

$$
\begin{aligned}
&a_n^{-1}\mathrm{E}\sup_{t_j\leqslant t,t'\leqslant t_{j+1}} |K_2(t-E_n(x,\,y)) - K_2(t'-E_n(x,\,y))|^2\\
&= a_n^{-1}\mathrm{E}\left[\sup_{t_j\leqslant t,t'\leqslant t_{j+1}} |K_2(t-E_n(x,y)) - K_2(t'-E_n(x,y))|^2 I(t_j - L_2 \leqslant E_n(x,y) \leqslant t_{j+1}+L_2)\right]\\
&\leqslant K\varepsilon^2,
\end{aligned}
$$

where $[-L_2,\, L_2]$ is the support of the kernel K_2. This completes the proof. $\square$

Proof of Theorem 3.3. Write

$$
\begin{aligned}
\hat\lambda(y|x) - \lambda(y|x) &= \frac{\hat f(y|x)}{1-\hat F(y|x)} - \frac{f(y|x)}{1-F(y|x)}\\
&= \frac{\hat f(y|x) - f(y|x)}{1-F(y|x)} + \frac{f(y|x)[\hat F(y|x) - F(y|x)]}{(1-F(y|x))^2}\\
&\quad + \frac{(\hat f(y|x) - f(y|x))(\hat F(y|x) - F(y|x))}{1-\hat F(y|x)} + \frac{f(y|x)[\hat F(y|x) - F(y|x)]^2}{(1-\hat F(y|x))(1-F(y|x))^2}.
\end{aligned}
$$

The representation for the first term above follows from Theorem 3.1. For the second term we use a slight generalization of the representation for $\hat F(y|x)$ given in Theorem 3.3 in Van Keilegom and Akritas (1999). The latter theorem assumes that $na_n^5 \to 0$, which implies that all bias terms are $o((na_n)^{-1/2})$. Here we assume weaker conditions on the bandwidth, which means that a bias term has to be added in the representation. In order to show that the third and fourth terms are of lower order, we will prove that

$$
\sup_{(x,y)\in\Omega} |\hat F(y|x) - F(y|x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \text{ a.s.,}
$$

$$
\sup_{(x,y)\in\Omega} |\hat f(y|x) - f(y|x)| = O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \text{ a.s.}
$$

We will show the latter result; the former can be obtained in a similar way.

The representation for $\hat f(y|x)$ established in Theorem 3.1 has two main terms, say

$A_1(y|x)$ and $A_2(y|x)$. The term $A_1(y|x)$ contains the function $g_{x,y}(z, \delta)$, which is formed from the functions $\eta(z, \delta|x)$ and $\zeta(z, \delta|x)$. These functions are integrals of the function $\xi(z, \delta, y)$. Now, since

$$(na_n)^{-1} \sum_{i=1}^{n} K_1\left(\frac{x - X_i}{a_n}\right) \xi(Z_i, \Delta_i, y|x) \tag{A.8}$$

$$= (na_n)^{-1} \sum_{i=1}^{n} K_1\left(\frac{x - X_i}{a_n}\right) \left[\int_{-\infty}^{y} \frac{I(Z_i \leqslant s) - H(s|x)}{(1 - H(s|x))^2} \, \mathrm{d}H_1(s|x)\right.$$

$$\left. + \frac{I(Z_i \leqslant y, \Delta_i = 1) - H_1(y|x)}{1 - H(y|x)} - \int_{-\infty}^{y} \frac{I(Z_i \leqslant s, \Delta_i = 1) - H_1(s|x)}{(1 - H(s|x))^2} \, \mathrm{d}H(s|x)\right]$$

is $O((na_n)^{-1/2}(\log a_n^{-1})^{1/2})$ a.s. uniformly in x and $y \leqslant T_x$ by Lemma A.1 in Van Keilegom and Akritas (1999), it follows that $A_1(y|x)$ is uniformly of the right order. For $A_2(y|x)$ we first note that $n^{-1}\sum_{i=1}^{n}\xi_e(E_i, \Delta_i, y)$ can be decomposed into three terms in a similar way as was done in (A.8) for the function $\xi(z, \delta, y|x)$. We consider the second term, which is the most difficult one. Denote $e_v(y) = (y - va_n - m(x))/\sigma(x)$ and $e(y) = (y - m(x))/\sigma(x)$. Then,

$$(na_n)^{-1} \sum_{i=1}^{n} \int \frac{I(E_i \leqslant e_v(y), \Delta_i = 1) - H_{e1}(e_v(y))}{1 - H_e(e_v(y))} \, \mathrm{d}K_2(v)$$

$$= (na_n)^{-1} \sum_{i=1}^{n} \int (1 - H_e(e_v(y)))^{-1} [I(E_i \leqslant e_v(y), \Delta_i = 1) - H_{e1}(e_v(y))$$

$$- I(E_i \leqslant e(y), \Delta_i = 1) + H_{e1}(e(y))] \, \mathrm{d}K_2(v)$$

$$= O((na_n)^{-1/2}(\log a_n^{-1})^{1/2}) \text{ a.s.}$$

by Lemma 2.4 in Stute (1982). This completes the proof.

Acknowledgement

This research was supported by the Projet d'Actions de Recherche Concertées no. 98/03-217 from the Belgian Government.

References

Beran, R. (1981) Nonparametric regression with randomly censored survival data. Technical report, University of California, Berkeley.

Dette, H. and Gefeller, O. (1995) The impact of different definitions of nearest neighbour distances for censored data on nearest neighbour kernel estimators of the hazard rate. *J. Nonparamet. Statist.*, **4**, 271–282.

Gefeller, O. and Dette, H. (1992) Nearest neighbour kernel estimation of the hazard function from censored data. *J. Statist. Comput. Simulation*, **43**, 93–101.

Gray, R.J. (1996) Hazard rate regression using ordinary nonparametric regression smoothers. *J. Comput. Graph. Statist.*, **5**, 190–207.

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

Li, G. (1997) Optimal rate local smoothing in a multiplicative intensity counting process model. *Math. Methods Statist*., **6**, 224–244.

Li, G. and Doss, H. (1995) An approach to nonparmetric regression for life history data using local linear fitting. *Ann. Statist.*, **23**, 787–823.

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

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 Related Fields*, **80**, 461–473.

McKeague, I.W. and Utikal, K.J. (1990) Inference for a nonlinear counting process regression model. *Ann. Statist*., **18**, 1172–1187.

Müller, H.-G. and Wang, J.-L. (1994) Hazard estimation under random censoring with varying kernels and bandwidths. *Biometrics*, **50**, 61–76.

Rosenblatt, M. (1971) Curve estimates. *Ann. Math. Statist*., **42**, 1815–1842.

Stute, W. (1982) The oscillation behavior of empirical processes. *Ann. Probab.*, **10**, 86–107.

van der Vaart, A.W. and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

Van Keilegom, I. and Akritas, M.G. (1999) Transfer of tail information in censored regression models. *Ann. Statist.*, **27**, 1745–1784.

Van Keilegom, I. and Veraverbeke, N. (1997) Estimation and bootstrap with censored data in fixed design nonparametric regression. *Ann. Inst. Statist. Math.*, **49**, 467–491.

Van Keilegom, I. and Veraverbeke, N. (2000) Density and hazard estimation in censored regression models. Technical report 0036, Institut de Statistique, Université Catholique de Louvain.

Received November 2000 and revised April 2002