Hazard rate estimation in nonparametric regression with censored data

Ingrid Van Keilegom[1]
Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays 20
B-1348 Louvain-la-Neuve
Belgium

Noël Veraverbeke
Department of Mathematics
Limburgs Universitair Centrum
Universitaire Campus
B-3590 Diepenbeek
Belgium

September 20, 2000

Abstract

Consider a regression model in which the responses are subject to random right censoring. In this model, Beran studied the nonparametric estimation of the conditional cumulative hazard function and the corresponding cumulative distribution function. The main idea is to use smoothing in the covariates. Here we study asymptotic properties of the corresponding hazard function estimator obtained by convolution smoothing of Beran's cumulative hazard estimator. We establish asymptotic expressions for the bias and the variance of the estimator, which together with an asymptotic representation lead to a weak convergence result. Also, the uniform strong consistency of the estimator is obtained.

KEY WORDS: Asymptotic representation; Hazard rate; Nonparametric regression; Right censoring; Weak convergence.

[1] A first draft of this paper was written while the first author was working at Limburgs Universitair Centrum and later at Penn State University. The research grant 'Projet d'Actions de Recherche Concertées', No. 98/03–217, from the Belgian government, is also gratefully acknowledged.

1 Introduction

Consider a regression model in which the response is subject to random right censoring. This paper concerns the kernel estimation of the conditional hazard function of the response given a value of the covariate.

We first introduce some notations. At fixed design points $x_1 \leq \ldots \leq x_n$ we have nonnegative responses $Y_1, \ldots, Y_n$ such as survival or failure times. For simplicity, we assume that the support of the covariates is the interval $[0,1]$. These design points represent e.g. the dose of a drug (in case of a medical study) or some environmental condition of a machine (like temperature, humidity, ...) for an industrial study. The responses Y_i $(i = 1, \ldots, n)$ are independent random variables and the distribution function of Y_i at x_i will be denoted by $F_{x_i}(t) = P(Y_i \leq t)$. As often occurs in medical or industrial studies, the responses are subject to random right censoring, i.e. the observed random variables at design point x_i are T_i and δ_i $(i = 1, \ldots, n)$, where $T_i = \min(Y_i, C_i)$, $\delta_i = I(Y_i \leq C_i)$ and $C_1, \ldots, C_n$ are nonnegative independent censoring variables with distribution functions $G_{x_i}(t) = P(C_i \leq t)$. We assume that Y_i and C_i are independent for each i. Hence, the distribution function $H_{x_i}(t) = P(T_i \leq t)$ satisfies $1 - H_{x_i}(t) = (1 - F_{x_i}(t))(1 - G_{x_i}(t))$. At a given fixed design value $x \in [0,1]$, we write F_x, G_x, H_x for the distribution function of respectively the response Y_x at x, the censoring variable C_x at x and $T_x = \min(Y_x, C_x)$. Also we write $\delta_x = I(Y_x \leq C_x)$.

Assume that $F_x(t)$ has a density $f_x(t)$ and denote the cumulative hazard function by $\Lambda_x(t)$ and the hazard function as $\lambda_x(t)$. This paper concerns nonparametric estimation of the hazard function $\lambda_x(t)$ by kernel smoothing of an appropriate nonparametric estimator for $\Lambda_x(t)$. This problem has been extensively studied in the literature. In the simplest case of i.i.d. observations (and no censoring) we refer to Watson and Leadbetter (1964a,b), Rice and Rosenblatt (1976), Singpurwalla and Wong (1983). In the presence of right random censoring, the problem has been studied by Tanner and Wong (1983), Ramlau - Hansen (1983), Yandell (1983), McNichols and Padgett (1985), Diehl and Stute (1988), Lo, Mack and Wang (1989), Müller and Wang (1990).

In the present case of regression with randomly censored survival data, it was Beran (1981) who first studied the estimation of the cumulative hazard function $\Lambda_x(t)$ (and the corresponding cumulative distribution function) without any parametric assumptions. It is an alternative to the classical Cox (1972) regression model and the main idea is the use of smoothing over the covariate space. Beran proved consistency and Dabrowska (1987, 1989) studied the asymptotic properties of the distribution and quantile function estimators. Further results, including bootstrap approximations were obtained in Van Keilegom and Veraverbeke (1996, 1997a,b, 1998a).

We introduce the notation

$$H_x^u(t) = P(T_x \leq t, \delta_x = 1) = \int_0^t (1 - G_x(s-))dF_x(s)$$

for the subdistribution function of the uncensored observations. The cumulative hazard function Λ_x can than be rewritten as

$$\Lambda_x(t) = \int_0^t \frac{dH_x^u(s)}{1 - H_x(s-)}. \tag{1.1}$$

The idea of Beran (1981) is to replace H_x and H_x^u by empirical quantities. We will use here kernel estimators with Gasser-Müller type weights $w_{ni}(x; h_n)$. They are defined as

$$w_{ni}(x; h_n) = \frac{1}{c_n(x; h_n)} \int_{x_{i-1}}^{x_i} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) dz \qquad i = 1, \ldots, n$$

$$c_n(x; h_n) = \int_0^{x_n} \frac{1}{h_n} K\left(\frac{x-z}{h_n}\right) dz.$$

Here $x_0 = 0$, K is a known probability density function (kernel) and $\{h_n\}$ is a sequence of positive constants (bandwidth), tending to 0 as $n \to \infty$. Replacing now H_x and H_x^u in (1.1) by the following kernel type estimators

$$H_{xh}(t) = \sum_{i=1}^n w_{ni}(x; h_n) I(T_i \leq t)$$

$$H_{xh}^u(t) = \sum_{i=1}^n w_{ni}(x; h_n) I(T_i \leq t, \delta_i = 1)$$

leads to the following Nelson-Aalen type estimator for $\Lambda_x(t)$:

$$\Lambda_{xh}(t) = \int_0^t \frac{dH_{xh}^u(s)}{1 - H_{xh}(s-)} \ .$$

Our estimator for $\lambda_x(t)$ is now based on a further kernel smoothing of $\Lambda_{xh}(t)$. We therefore use a second kernel K_0 and a second bandwidth sequence $\{b_n\}$, also tending to 0 as $n \to \infty$.

The estimator for $\lambda_x(t)$ is then defined as

$$\begin{aligned}(1.2) \qquad \lambda_{xhb}(t) &= \frac{1}{b_n}\int K_0\left(\frac{t-s}{b_n}\right) d\Lambda_{xh}(s) \\ &= \frac{1}{b_n}\sum_{i=1}^{n} K_0\left(\frac{t-T_{(i)}}{b_n}\right)\frac{\delta_{(i)} w_{n(i)}(x;h_n)}{1-\sum_{j=1}^{i-1} w_{n(j)}(x;h_n)}.\end{aligned}$$

Here $T_{(1)} \leq \ldots \leq T_{(n)}$ are the ordered $T_1, \ldots, T_n$ and the $\delta_{(i)}$ and $w_{n(i)}(x;h_n)$ are the corresponding δ_i and $w_{ni}(x;h_n)$. Note that if we take the weights $w_{ni}(x;h_n)$ all equal to n^{-1}, then the estimator becomes the kernel-smoothed Nelson estimator for censored data

$$\frac{1}{b_n}\sum_{i=1}^{n} K_0\left(\frac{t-T_{(i)}}{b_n}\right)\frac{\delta_{(i)}}{n-i+1}$$

which was studied in Tanner and Wong (1983) and the references given above. Also, if there is no censoring ($T_i = Y_i, \delta_i = 1$), we obtain the well known hazard rate estimator for complete data.

The estimator (1.2) has been studied by McKeague and Utikal (1990), who used counting processes to prove the asymptotic normality of $\lambda_{xhb}(t)$. Li and Doss (1995) considered a local polynomial estimator for $\lambda_x(t)$ (instead of a local constant one) and obtained the asymptotic normality of their estimator. The results of both papers, however, are not shown for the bandwidths h_n and b_n of optimal rate. The results for the optimal bandwidths are established in Li (1997) for both type of estimators.

Nielsen and Linton (1995) considered an alternative kernel estimator for the hazard $\lambda_x(t)$. It is easily seen, however, that this estimator has the same limiting normal distribution as our estimator $\lambda_{xhb}(t)$ (compare Theorem 4.1 below with their Theorem 1). Differences might be expected in the higher order terms of e.g. asymptotic expressions of the mean squared error. It should also be noted that the Nielsen-Linton estimator allows multidimensional and time dependent covariates, while our $\lambda_{xhb}(t)$ restricts to one covariate, independent of time. On the other hand, note that the present paper does not require the smoothing over time to be done with the same bandwidth as for smoothing over the covariate space. Also note that if we take $h_n = b_n$, the conditions on the bandwidth required for our asymptotic normality result (Theorem 4.1) reduce to $nh_n^{2+\delta} \to \infty$ (for some $\delta > 0$) and $nh_n^6 \to C \geq 0$, while the corresponding result in Nielsen and Linton (1995) is slightly weaker in that $\delta = 0$ in their case. Nielsen and Linton obtained uniform weak consistency, while we prove below weak convergence for the properly normalized process.

In this paper, we first establish an a.s. asymptotic representation for $\lambda_{xhb}(t)$ (Section 2) and calculate the asymptotic bias and covariance (see Section 3), which enables us in Section 4 to obtain the asymptotic normality of $\lambda_{xhb}(t)$ under very general conditions on the bandwidths h_n and b_n, including the bandwidths of optimal rate. Next, in Section 5, we prove the weak convergence of the process $(nh_nb_n)^{1/2}(\lambda_{xhb}(t) - \lambda_x(t))$. The strong uniform consistency of $\lambda_{xhb}(t)$ (with rate of convergence) will be derived in Section 6.

2 Almost sure asymptotic representation

The study of the asymptotic properties of the proposed estimator $\lambda_{xhb}(t)$ will be facilitated by proving first an a.s. asymptotic representation.

We first introduce some further notations and conditions. For the design points $x_1, \ldots, x_n$ we denote $\underline{\Delta}_n = \min_{1\le i\le n}(x_i - x_{i-1})$ and $\overline{\Delta}_n = \max_{1\le i\le n}(x_i - x_{i-1})$. Further let $\|K\|_\infty = \sup_{u\in I\!R} K(u)$, $\|K\|_2^2 = \int_{-\infty}^{\infty} K^2(u)\,du$, $\mu_1^K = \int_{-\infty}^{\infty} uK(u)\,du$, $\mu_2^K = \int_{-\infty}^{\infty} u^2K(u)\,du$ and similarly for K_0. We will constantly use the following assumptions on the design and on the kernels K and K_0 :

(C1) $x_n \to 1$, $\overline{\Delta}_n = O(n^{-1})$, $\overline{\Delta}_n - \underline{\Delta}_n = o(n^{-1})$

(C2) (i) K is a probability density function with finite support $[-L, L]$ for some $L > 0$, $\mu_1^K = 0$, and K is Lipschitz of order 1.

(C2) (ii) K_0 is a continuous probability density function of bounded variation and with finite support $[-L_0, L_0]$ for some $L_0 > 0$, $\mu_1^{K_0} = 0$.

Note that $c_n(x; h_n) = 1$ for n sufficiently large (depending on x) since $x_n \to 1$ and K has finite support. This makes that in all proofs of asymptotic results, we will take $c_n(x; h_n) = 1$. For any (sub)distribution function L, we denote T_L for the right endpoint of its support, i.e. $T_L = \inf\{t : L(t) = L(\infty)\}$. Hence, $T_{H_x} = \min(T_{F_x}, T_{G_x})$.

In the formulation of our results, we will need typical types of smoothness conditions on functions like $H_x(t)$ and $H_x^u(t)$. We formulate them here for a general (sub)distribution function $L_x(t)$, $0 \le x \le 1$, $t \in I\!R$, and for a fixed $T > 0$:

(C3) $\dot{L}_x(t) = \dfrac{\partial}{\partial x} L_x(t)$ exists and is continuous in $(x,t) \in [0,1] \times [0,T]$

(C4) $L'_x(t) = \dfrac{\partial}{\partial t} L_x(t)$ exists and is continuous in $(x,t) \in [0,1] \times [0,T]$

(C5) $\ddot{L}_x(t) = \dfrac{\partial^2}{\partial x^2} L_x(t)$ exists and is continuous in $(x,t) \in [0,1] \times [0,T]$

(C6) $L''_x(t) = \dfrac{\partial^2}{\partial t^2} L_x(t)$ exists and is continuous in $(x,t) \in [0,1] \times [0,T]$

(C7) $\dot{L}'_x(t) = \dfrac{\partial^2}{\partial x \partial t} L_x(t)$ exists and is continuous in $(x,t) \in [0,1] \times [0,T]$

(C8) $\ddot{L}'_x(t) = \dfrac{\partial^3}{\partial x^2 \partial t} L_x(t)$ exists and is continuous in $(x,t) \in [0,1] \times [0,T]$

(C9) $L'''_x(t) = \dfrac{\partial^3}{\partial t^3} L_x(t)$ exists and is continuous in $(x,t) \in [0,1] \times [0,T]$.

In order to formulate the asymptotic representation for $\lambda_{xhb}(t)$ we define

(2.1)
$$\begin{aligned}
\beta_{xhb}(t) &= \frac{1}{b_n} \int \Lambda_x(t - b_n u)\, dK_0(u) - \lambda_x(t) \\
&\quad + \frac{1}{b_n} \int \left\{ \int_0^{t-b_n u} \frac{EH_{xh}(s) - H_x(s)}{(1 - H_x(s))^2}\, dH_x^u(s) \right. \\
&\quad \left. + \int_0^{t-b_n u} \frac{d(EH_{xh}^u(s) - H_x^u(s))}{1 - H_x(s)} \right\} dK_0(u) \\
h_{tx}(z,\delta) &= \frac{1}{b_n} \int [g_{t-b_n u,x}(z,\delta) - Eg_{t-b_n u,x}(z,\delta)]\, dK_0(u) \\
g_{tx}(z,\delta) &= \int_0^t \frac{I(z \le s) - H_x(s)}{(1 - H_x(s))^2} dH_x^u(s) + \frac{I(z \le t, \delta = 1) - H_x^u(t)}{1 - H_x(t)} \\
&\quad - \int_0^t \frac{I(z \le s, \delta = 1) - H_x^u(s)}{(1 - H_x(s))^2} dH_x(s).
\end{aligned}$$

Since

(2.2) $\lambda_{xhb}(t) - \lambda_x(t)$
$$\begin{aligned}
&= \frac{1}{b_n} \int \Lambda_{xh}(t - b_n u)\, dK_0(u) - \lambda_x(t) \\
&= \frac{1}{b_n} \int [\Lambda_{xh}(t - b_n u) - \Lambda_x(t - b_n u)]\, dK_0(u) + \frac{1}{b_n} \int \Lambda_x(t - b_n u)\, dK_0(u) - \lambda_x(t),
\end{aligned}$$

a possible way to obtain an asymptotic representation for $\lambda_{xhb}(t) - \lambda_x(t)$, is to replace $\Lambda_{xh}(t - b_n u) - \Lambda_x(t - b_n u)$ in the above expression by a representation. Such a representation is implicitly contained in the proof of Theorem 1 in Van Keilegom and Veraverbeke (1997b). However, the theorem is only valid for bandwidth sequences h_n satisfying $nh_n^5/\log n = O(1)$, not including the bandwidth of optimal rate, which is $h_n = Kn^{-1/6}$ (see next section). It can be shown that this condition can be omitted, leading however to a representation with a remainder term of the order $O(h_n^3)$ and this order will turn out to be too slow. For these reasons, we establish in the next theorem a representation directly for $b_n^{-1} \int [\Lambda_{xh}(t - b_n u) - \Lambda_x(t - b_n u)]\, dK_0(u)$.

Theorem 2.1 *Assume (C1), (C2), $H_x(t)$ satisfies (C5) - (C8) and $H_x^u(t)$ satisfies (C5) - (C7) in $[0, T]$ with $T < T_{H_x}$, $h_n \to 0$, $b_n \to 0$, $(nh_n)^{-1} \log n \to 0$. Then, for $t < T_{H_x}$,*

$$\lambda_{xhb}(t) - \lambda_x(t) = \sum_{i=1}^{n} w_{ni}(x; h_n) h_{tx}(T_i, \delta_i) + \beta_{xhb}(t) + r_{xhb}(t) \tag{2.3}$$

and for $T < T_{H_x}$,

$$\sup_{0 \le t \le T} |r_{xhb}(t)| = O\left((nh_n b_n)^{-1} \log n + (b_n^{1/2} + h_n^2 b_n^{-1})(nh_n)^{-1/2} (\log n)^{1/2} + h_n^4 b_n^{-1} \right) \quad a.s.$$

as $n \to \infty$.

Proof. First note that

$$\sup_{0 \le t \le T} \left| \Lambda_{xh}(t) - \int_0^t \frac{dH_{xh}^u(s)}{1 - H_{xh}(s)} \right| = O((nh_n)^{-1}) \quad \text{a.s.}$$

Hence, using (2.2), it suffices to consider

$$\begin{aligned} &\frac{1}{b_n} \int \left[\int_0^{t - b_n u} \frac{dH_{xh}^u(s)}{1 - H_{xh}(s)} - \int_0^{t - b_n u} \frac{dH_x^u(s)}{1 - H_x(s)} \right] dK_0(u) \\ &= \frac{1}{b_n} \int \left[\int_0^{t - b_n u} \frac{H_{xh}(s) - H_x(s)}{(1 - H_x(s))^2} dH_x^u(s) + \frac{H_{xh}^u(t - b_n u) - H_x^u(t - b_n u)}{1 - H_x(t - b_n u)} \right. \\ &\quad - \int_0^{t - b_n u} \frac{H_{xh}^u(s) - H_x^u(s)}{(1 - H_x(s))^2} dH_x(s) + R_{n1}(x, t - b_n u) + R_{n2}(x, t - b_n u) \\ &\quad \left. + R_{n3}(x, t - b_n u) \right] dK_0(u) \end{aligned} \tag{2.4}$$

where

$$R_{n1}(x,t)=\int_0^t \frac{(H_{xh}(s)-H_x(s))^2}{(1-H_x(s))^2(1-H_{xh}(s))}dH_x^u(s)$$

$$R_{n2}(x,t)=\int_0^t \left[\frac{1}{1-H_{xh}(s)}-\frac{1}{1-EH_{xh}(s)}\right]d(H_{xh}^u(s)-H_x^u(s))$$

$$R_{n3}(x,t)=\int_0^t \left[\frac{1}{1-EH_{xh}(s)}-\frac{1}{1-H_x(s)}\right]d(H_{xh}^u(s)-H_x^u(s)).$$

For $R_{n1}(x,t)$ we have

$$\begin{aligned}&\frac{1}{b_n}\sup_{0\le t\le T}\left|\int R_{n1}(x,t-b_nu)\,dK_0(u)\right|\\ \text{(2.5)}\qquad &\le \frac{K}{b_n}\sup_{0\le t\le T}|H_{xh}(t)-H_x(t)|^2(1-H_{xh}(T))^{-1}(1-H_x(T))^{-2}.\end{aligned}$$

From Lemma A1(b) and A3(a) in Van Keilegom and Veraverbeke (1997b) it follows that $\sup_t|H_{xh}(t)-H_x(t)|=O((nh_n)^{-1/2}(\log n)^{1/2}+h_n^2)$ a.s. Hence, (2.5) is $O((nh_nb_n)^{-1}\log n+h_n^4b_n^{-1})$ a.s. Next, partition $[0,T]$ into $O(b_n^{-1})$ subintervals $[t_i,t_{i+1}]$ of length b_n. Then,

$$\begin{aligned}&\frac{1}{b_n}\int R_{n2}(x,t-b_nu)\,dK_0(u)\\ &=\frac{1}{b_n}\int K_0\left(\frac{t-s}{b_n}\right)\left[\frac{1}{1-H_{xh}(s)}-\frac{1}{1-EH_{xh}(s)}\right]d(H_{xh}^u(s)-H_x^u(s))\\ &=\frac{1}{b_n}\int K_0\left(\frac{t-s}{b_n}\right)\left[\frac{1}{1-H_{xh}(s)}-\frac{1}{1-EH_{xh}(s)}-\frac{1}{1-H_{xh}(t_i)}+\frac{1}{1-EH_{xh}(t_i)}\right]\\ &\qquad\qquad d(H_{xh}^u(s)-H_x^u(s))\\ &\quad+\frac{1}{b_n}\left[\frac{1}{1-H_{xh}(t_i)}-\frac{1}{1-EH_{xh}(t_i)}\right]\int K_0\left(\frac{t-s}{b_n}\right)d(H_{xh}^u(s)-H_x^u(s))\\ &=\alpha_{n1}(x,t)+\alpha_{n2}(x,t),\end{aligned}$$

where $t_i\le t\le t_{i+1}$. Then, if $K_0(\frac{t-s}{b_n})\ne 0$, it follows that $|t_i-s|\le Kb_n$. We start with $\alpha_{n1}(x,t)$:

$$\begin{aligned}&|\alpha_{n1}(x,t)|\\ &\le Kb_n^{-1}\sup{}^*\left|\frac{1}{1-H_{xh}(t)}-\frac{1}{1-EH_{xh}(t)}-\frac{1}{1-H_{xh}(s)}+\frac{1}{1-EH_{xh}(s)}\right|\end{aligned}$$

$$
\begin{aligned}
&\quad \times\{H^u_{xh}(t+L_0b_n)-H^u_{xh}(t-L_0b_n)+H^u_x(t+L_0b_n)-H^u_x(t-L_0b_n)\} \\
&\le Kb_n^{-1}\left\{\sup{}^*\left|\frac{H_{xh}(t)-EH_{xh}(t)}{(1-EH_{xh}(t))^2}-\frac{H_{xh}(s)-EH_{xh}(s)}{(1-EH_{xh}(s))^2}\right|+O((nh_n)^{-1}\log n)\right\} \\
&\quad \times\{\sup{}^*|H^u_{xh}(t)-H^u_x(t)-H^u_{xh}(s)+H^u_x(s)|+2(H^u_x(t+L_0b_n)-H^u_x(t-L_0b_n))\} \\
&\le K\sup{}^*\frac{|H_{xh}(t)-EH_{xh}(t)-H_{xh}(s)+EH_{xh}(s)|}{(1-H_x(t))^2}+O((nh_n)^{-1}\log n) \\
&\quad +O(b_n(nh_n)^{-1/2}(\log n)^{1/2}) \\
&= O(b_n^{1/2}(nh_n)^{-1/2}(\log n)^{1/2}+(nh_n)^{-1}\log n)
\end{aligned}
$$

a.s., using Proposition A5 in Van Keilegom and Veraverbeke (1997b), where $\sup^*$ denotes the supremum over all (s,t) satisfying $|t-s|\le Kb_n$ and where the second inequality follows from the fact that $\sup_t|H_{xh}(t)-EH_{xh}(t)|=O((nh_n)^{-1/2}(\log n)^{1/2})$ a.s. (see Proposition A3(a) in Van Keilegom and Veraverbeke (1997b)). Next, we consider $\alpha_{n2}(x,t)$:

$$
\begin{aligned}
&\alpha_{n2}(x,t) \\
&= b_n^{-1}(H_{xh}(t_i)-EH_{xh}(t_i))\{(1-H_x(t_i))^{-2}+o(1)\}\int(H^u_{xh}(t-vb_n)-H^u_x(t-vb_n))\,dK_0(v) \\
&= O((nh_nb_n)^{-1}\log n+h_n^2b_n^{-1}(nh_n)^{-1/2}(\log n)^{1/2})
\end{aligned}
$$

a.s., uniformly over all t. For $R_{n3}(x,t)$ we have :

$$
\begin{aligned}
&\frac{1}{b_n}\int R_{n3}(x,t-b_nu)\,dK_0(u) \\
&= \frac{1}{b_n}\int\int_0^{t-b_nu}\left[\frac{1}{1-EH_{xh}(s)}-\frac{1}{1-H_x(s)}\right]d(H^u_{xh}(s)-H^u_x(s))\,dK_0(u) \\
&= \frac{1}{b_n}\int\int_0^{t-b_nu}\frac{EH_{xh}(s)-H_x(s)}{(1-EH_{xh}(s))(1-H_x(s))}d(H^u_{xh}(s)-H^u_x(s))\,dK_0(u) \\
&= \frac{1}{b_n}\int\Bigg\{(H^u_{xh}(t-b_nu)-H^u_x(t-b_nu))\frac{EH_{xh}(t-b_nu)-H_x(t-b_nu)}{(1-EH_{xh}(t-b_nu))(1-H_x(t-b_nu))} \\
&\quad -\int_0^{t-b_nu}(H^u_{xh}(s)-H^u_x(s))\frac{\partial}{\partial s}\left[\frac{EH_{xh}(s)-H_x(s)}{(1-EH_{xh}(s))(1-H_x(s))}\right]ds\Bigg\}\,dK_0(u) \\
&= O(h_n^2b_n^{-1}(nh_n)^{-1/2}(\log n)^{1/2}+h_n^4b_n^{-1})
\end{aligned}
$$

a.s., since

$$\frac{\partial}{\partial t}(EH_{xh}(t) - H_x(t)) = \frac{1}{2}\mu_2^K \ddot{h}_x(t)h_n^2 + o(h_n^2) + O(n^{-1}). \tag{2.6}$$

The proof of the latter is analogous to that of $EH_{xh}(t) - H_x(t) = \frac{1}{2}\mu_2^K \ddot{H}_x(t)h_n^2 + o(h_n^2) + O(n^{-1})$ (see Lemma A1(b) in Van Keilegom and Veraverbeke (1997b)). Finally, note that the main term in (2.4) equals

$$\begin{aligned}
&\frac{1}{b_n}\sum_{i=1}^n w_{ni}(x;h_n)\int g_{t-b_nu,x}(T_i,\delta_i)\,dK_0(u) \\
&= \sum_{i=1}^n w_{ni}(x;h_n)h_{tx}(T_i,\delta_i) \\
&\quad + \frac{1}{b_n}\int\left\{\int_0^{t-b_nu}\frac{EH_{xh}(s)-H_x(s)}{(1-H_x(s))^2}dH_x^u(s) + \int_0^{t-b_nu}\frac{d(EH_{xh}^u(s)-H_x^u(s))}{1-H_x(s)}\right\}dK_0(u).
\end{aligned}$$

This finishes the proof.

3 Asymptotic expressions for bias and covariance

The bias term $\beta_{xhb}(t)$ in the presentation of Theorem 2.1 consists of two parts (see (2.1)) :

$$\beta_{xhb}(t) = \beta_{xhb}^{(1)}(t) + \beta_{xhb}^{(2)}(t). \tag{3.1}$$

For the first part we can write

$$\beta_{xhb}^{(1)}(t) = \int \lambda_x(t-b_nu)K_0(u)du - \lambda_x(t) = \frac{1}{2}b_n^2\lambda_x''(t)\mu_2^{K_0} + o(b_n^2)$$

provided K_0 is a bounded density with $\mu_1^{K_0} = 0$ and λ_x is twice continuously differentiable at t. For the second part in (3.1), it follows from (2.6) that

$$\begin{aligned}
\beta_{xhb}^{(2)}(t) &= \int K_0(u)\left\{\frac{EH_{xh}(t-b_nu)-H_x(t-b_nu)}{(1-H_x(t-b_nu))^2}h_x^u(t-b_nu)\right. \\
&\qquad \left. + \frac{(EH_{xh}^u(s)-H_x^u(s))'|_{s=t-b_nu}}{1-H_x(t-b_nu)}\right\}du \\
&= \frac{1}{2}\mu_2^K h_n^2\int K_0(u)\phi_x(t-b_nu)\,du + o(h_n^2) + O(n^{-1}) \\
&= \frac{1}{2}\mu_2^K h_n^2\phi_x(t) + o(h_n^2) + O(n^{-1}),
\end{aligned}$$

where

$$\phi_x(t)=\frac{\ddot{H}_x(t)h_x^u(t)}{(1-H_x(t))^2}+\frac{\ddot{h}_x^u(t)}{1-H_x(t)}.$$

The conclusion is that the bias is given by

$$\beta_{xhb}(t)=\frac{1}{2}\lambda_x''(t)\mu_2^{K_0}b_n^2+\frac{1}{2}\phi_x(t)\mu_2^K h_n^2+o(b_n^2)+o(h_n^2)+O(n^{-1}). \tag{3.2}$$

We now deal with the covariance of the approximating sum in Theorem 2.1 :

$$\begin{aligned}
&\mathrm{Cov}\left(\sum_{i=1}^n w_{ni}(x;h_n)h_{sx}(T_i,\delta_i),\sum_{i=1}^n w_{ni}(x;h_n)h_{tx}(T_i,\delta_i)\right)\\
&=\frac{1}{b_n^2}\sum_{i=1}^n w_{ni}^2(x;h_n)\int\int \mathrm{Cov}(g_{s-b_nu,x}(T_i,\delta_i),g_{t-b_nv,x}(T_i,\delta_i))\,dK_0(u)\,dK_0(v)\\
&=-\frac{1}{b_n^2}\sum_{i=1}^n w_{ni}^2(x;h_n)\int\int_{-L_0}^{v+\frac{s-t}{b_n}} K_0(u)\frac{\partial}{\partial u}\mathrm{Cov}(g_{s-b_nu,x}(T_i,\delta_i),g_{t-b_nv,x}(T_i,\delta_i))\,du\,dK_0(v)\\
&\quad-\frac{1}{b_n^2}\sum_{i=1}^n w_{ni}^2(x;h_n)\int\int_{v+\frac{s-t}{b_n}}^{L_0} K_0(u)\frac{\partial}{\partial u}\mathrm{Cov}(g_{s-b_nu,x}(T_i,\delta_i),g_{t-b_nv,x}(T_i,\delta_i))\,du\,dK_0(v).
\end{aligned}$$

Some straightforward calculations show that

$$\begin{aligned}
&\mathrm{Cov}(g_{s-b_nu,x}(T_i,\delta_i),g_{t-b_nv,x}(T_i,\delta_i))\\
&=\int_0^{t-b_nv}\frac{H_x(y)-H_{x_i}(y)}{(1-H_x(y))^2}\int_0^y\frac{dH_x^u(z)}{(1-H_x(z))^2}dH_x^u(y)\\
&\quad+\int_0^{t-b_nv}\frac{1}{1-H_x(y)}\int_0^y\frac{dH_x^u(z)}{(1-H_x(z))^2}d(H_x^u(y)-H_{x_i}^u(y))\\
&\quad+\int_0^{t-b_nv}\frac{1}{(1-H_x(y))^2}\int_y^{s-b_nu}\frac{H_x(z)-H_{x_i}(z)}{(1-H_x(z))^2}dH_x^u(z)dH_x^u(y)\\
&\quad+\int_0^{t-b_nv}\frac{1}{(1-H_x(y))^2}\int_y^{s-b_nu}\frac{d(H_x^u(z)-H_{x_i}^u(z))}{1-H_x(z)}dH_x^u(y)+\int_0^{\min(s-b_nu,t-b_nv)}\frac{dH_{x_i}^u(y)}{(1-H_x(y))^2}\\
&\quad-\left[-\int_0^{t-b_nv}\frac{H_x(y)-H_{x_i}(y)}{(1-H_x(y))^2}dH_x^u(y)+\int_0^{t-b_nv}\frac{d(H_{x_i}^u(y)-H_x^u(y))}{1-H_x(y)}\right]\\
&\qquad\cdot\left[-\int_0^{s-b_nu}\frac{H_x(y)-H_{x_i}(y)}{(1-H_x(y))^2}dH_x^u(y)+\int_0^{s-b_nu}\frac{d(H_{x_i}^u(y)-H_x^u(y))}{1-H_x(y)}\right],
\end{aligned}$$

from which $\frac{\partial}{\partial u}\mathrm{Cov}(g_{s-b_nu,x}(T_i,\delta_i), g_{t-b_nv,x}(T_i,\delta_i))$ can be easily obtained. Next, we apply Lemma A1 in Van Keilegom and Veraverbeke (1997a), which states that for any function $\gamma : [0,1] \times [0,T]^2 \to I\!R$ for which $\gamma(\cdot, s, t) : [0,1] \to I\!R$ $(s,t \in [0,T])$ is Lipschitz with Lipschitz constant uniformly bounded on $[0,T]^2$ and for which $\gamma(x,\cdot,\cdot) : [0,T]^2 \to I\!R$ is bounded, it holds that

$$\sup_{0\le s,t\le T}\left|\sum_{i=1}^n w_{ni}^2(x;h_n)\gamma(x_i,s,t) - \frac{\|K\|_2^2}{nh_n}\gamma(x,s,t)\right| = o((nh_n)^{-1}).$$

It follows that

$$\begin{aligned}&\sum_{i=1}^n w_{ni}^2(x;h_n)\frac{\partial}{\partial u}\mathrm{Cov}(g_{s-b_nu,x}(T_i,\delta_i), g_{t-b_nv,x}(T_i,\delta_i))\\ &= \begin{cases} o((nh_n)^{-1}b_n) & (s-b_nu > t-b_nv)\\ -b_n\dfrac{\|K\|_2^2}{nh_n}\dfrac{h_x^u(s-b_nu)}{(1-H_x(s-b_nu))^2} + o((nh_n)^{-1}b_n) & (s-b_nu < t-b_nv)\end{cases}\end{aligned}$$

and hence

$$\begin{aligned}&\mathrm{Cov}\left(\sum_{i=1}^n w_{ni}(x;h_n)h_{sx}(T_i,\delta_i), \sum_{i=1}^n w_{ni}(x;h_n)h_{tx}(T_i,\delta_i)\right)\\ &= \frac{\|K\|_2^2}{nh_nb_n}\int\int_{v+\frac{s-t}{b_n}}^{L_0} K_0(u)\frac{h_x^u(s-b_nu)}{(1-H_x(s-b_nu))^2}\,du\,dK_0(v) + o((nh_nb_n)^{-1})\\ &= \frac{\|K\|_2^2}{nh_nb_n}\int K_0(v)K_0\left(v+\frac{s-t}{b_n}\right)\frac{h_x^u(t-b_nv)}{(1-H_x(t-b_nv))^2}\,dv + o((nh_nb_n)^{-1})\\ &= \frac{\|K\|_2^2}{nh_nb_n}\int K_0(v)K_0\left(v+\frac{s-t}{b_n}\right)dv\,\frac{\lambda_x(t)}{1-H_x(t)} + o((nh_nb_n)^{-1}),\end{aligned} \tag{3.3}$$

since $\lambda_x(t) = h_x^u(t)/(1-H_x(t))$.

Remark 3.1 *It follows that the approximate mean squared error of the estimator $\lambda_{xhb}(t)$ is of the form*

$$\frac{C_1}{nh_nb_n} + (C_2h_n^2 + C_3b_n^2)^2$$

for some constants C_1, C_2, C_3. Minimization gives that the optimal choice is

$$h_n = b_n = Cn^{-\frac{1}{6}}$$

for some constant $C > 0$.

4 Asymptotic normality

The representation in Theorem 2.1 entails the asymptotic normality result for $(nh_nb_n)^{1/2}(\lambda_{xhb}(t) - \lambda_x(t))$.

Theorem 4.1 *Assume (C1), (C2), $H_x(t)$ satisfies (C5) - (C8) and $H_x^u(t)$ satisfies (C5) - (C9) in $[0,T]$ with $T < T_{H_x}$, $(nh_nb_n)^{-1}(\log n)^2 \to 0$, $b_n^2 \log n \to 0$, $h_n^4 b_n^{-1} \log n \to 0$, $nh_n^9 b_n^{-1} \to 0$, $nh_nb_n^5 \to C_1 \geq 0$, $nh_n^5 b_n \to C_2 \geq 0$. Then, for $t \leq T$, as $n \to \infty$,*

$$(nh_nb_n)^{1/2}(\lambda_{xhb}(t) - \lambda_x(t)) \to N(b_x(t); s_x^2(t)),$$

where

$$\begin{aligned} b_x(t) &= \frac{1}{2} C_1^{1/2} \lambda_x''(t) \mu_2^{K_0} + \frac{1}{2} C_2^{1/2} \phi_x(t) \mu_2^K, \\ s_x^2(t) &= \|K\|_2^2 \, \|K_0\|_2^2 \frac{\lambda_x(t)}{1 - H_x(t)}. \end{aligned}$$

In particular, if $h_n = b_n = Cn^{-1/6}$ $(C > 0)$, then $b_x(t) = \frac{1}{2}C^3(\lambda_x''(t)\mu_2^{K_0} + \phi_x(t)\mu_2^K)$. (Note that one or both terms of $b_x(t)$ can be zero, depending on whether C_1 and C_2 equal zero or not.)

Proof. The stated conditions on the bandwidths h_n and b_n ensure that $(nh_nb_n)^{1/2}$ times the remainder term in the representation of Theorem 2.1 tends to zero. Let

$$\sum_{i=1}^n Z_{ni} = (nh_nb_n)^{1/2} \sum_{i=1}^n w_{ni}(x; h_n) h_{tx}(T_i, \delta_i)$$

denote the main term in the representation of $(nh_nb_n)^{1/2}(\lambda_{xhb}(t) - \lambda_x(t))$. We have that $E(Z_{ni}) = 0$. Also, by (3.3),

$$\begin{aligned} \sum_{i=1}^n E(Z_{ni}^2) &= nh_nb_n \mathrm{Var}\left(\sum_{i=1}^n w_{ni}(x; h_n) h_{tx}(T_i, \delta_i) \right) \\ &= \|K\|_2^2 \, \|K_0\|_2^2 \frac{\lambda_x(t)}{1 - H_x(t)} + o(1). \end{aligned}$$

Moreover,

$$\sum_{i=1}^n E|Z_{ni}|^3 \leq K(nh_nb_n)^{-1/2} \sum_{i=1}^n E(Z_{ni}^2) = O((nh_nb_n)^{-1/2}).$$

Hence, the Liapunov ratio equals

$$\frac{\sum_{i=1}^n E|Z_{ni}|^3}{(\sum_{i=1}^n \text{Var}(Z_{ni}))^{3/2}} = O((nh_nb_n)^{-1/2}) = o(1).$$

The expression for $b_x(t)$ follows from (3.2).

Remark 4.2 *Li (1997) proved the above asymptotic normality result for the special case where $h_n = b_n = Cn^{-1/6}$ ($C > 0$) and for a random design. There is, however, a minor mistake in the expression of the mean of the limiting process : formula (2.16) in Theorem 2 (and also formula (2.8) in Theorem 1) should be divided by two.*

5 Weak convergence

In this section we establish the weak convergence of the hazard function estimator. In order to obtain the tightness of the process $(nh_nb_n)^{1/2}(\lambda_{xhb}(\cdot) - \lambda_x(\cdot))$, the argument needs to be of the form b_nt ($t \in [0, \tilde{T}]$, $\tilde{T}$ arbitrary). This is a typical feature for processes of nonparametric density, hazard or regression function estimators and can also be found in e.g. Rosenblatt (1971) and Van Keilegom and Veraverbeke (1998b). In the latter paper, the conditional density and hazard function are estimated under the heteroscedastic model $Y = m(X) + \sigma(X)\varepsilon$, where ε is independent of X and m and σ are smooth but unknown functions. The rate of convergence of the proposed estimators is faster than in the present, completely nonparametric context (see the above paper for details).

Theorem 5.1 *Assume (C1), (C2), $H_x(t)$ satisfies (C5) - (C8) and $H_x^u(t)$ satisfies (C5) - (C9) in $[0, T]$ with $T < T_{H_x}$, $(nh_nb_n)^{-1}(\log n)^2 \to 0$, $b_n^2 \log n \to 0$, $h_n^4 b_n^{-1} \log n \to 0$, $nh_n^9 b_n^{-1} \to 0$, $nh_nb_n^5 \to C_1 \geq 0$, $nh_n^5 b_n \to C_2 \geq 0$. Then, the process*

$$(nh_nb_n)^{1/2} \left(\frac{1 - H_x(b_nt)}{\lambda_x(b_nt)} \right)^{1/2} (\lambda_{xhb}(b_nt) - \lambda_x(b_nt))$$

($x \in [0, 1]$ fixed, $t \in [0, \tilde{T}]$, $\tilde{T} > 0$ arbitrary), converges weakly to a Gaussian process $Z_x(t)$ with mean function

$$E(Z_x(t)) = \frac{1}{2} C_1^{1/2} \lambda_x''(t) \mu_2^{K_0} + \frac{1}{2} C_2^{1/2} \phi_x(t) \mu_2^K$$

and covariance function

$$Cov(Z_x(s), Z_x(t)) = \|K\|_2^2 \int K_0(v) K_0(v + s - t)\, dv.$$

Proof. For showing the weak convergence of the main term in the representation (see Theorem 2.1), use will be made of Theorem 2.11.9 in van der Vaart and Wellner (1996). We start by showing 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 $(nh_nb_n)^{1/2}(\frac{1-H_x(b_nt)}{\lambda_x(b_nt)})^{1/2}(\lambda_{xhb}(b_nt_j) - \lambda_x(b_nt_j))$ $(0 \le t_1, \dots, t_k \le \tilde{T}$ arbitrary, k arbitrary). The proof parallels completely that of Theorem 4.1, which deals with the case $k = 1$, and will therefore be omitted. For the calculation of the bias and the covariance we refer to Section 3. 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 function $g_{tx}(z, \delta)$ is bounded. We will next show that

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

for every $\delta_n \downarrow 0$, where $N_{[]}$ is the bracketing number and

$$\mathcal{F} = \Big\{(nh_nb_n)^{1/2}\Big(\frac{1-H_x(b_nt)}{\lambda_x(b_nt)}\Big)^{1/2}(\lambda_{xhb}(b_nt) - \lambda_x(b_nt)); 0 \le t \le \tilde{T}\Big\}.$$

Partition $[0, \tilde{T}]$ into $O(\varepsilon^{-1})$ subintervals $[t_j, t_{j+1}]$ of length at most $K\varepsilon$ for some $K > 0$. We will show that

$$nh_nb_n \sum_{i=1}^n \frac{w_{ni}^2(x; h_n)}{b_n^2} E\left[\sup_{t_j \le t, t' \le t_{j+1}} |W_{ni}(t) - W_{ni}(t')|^2\right] \le \varepsilon^2, \tag{5.1}$$

where

$$W_{ni}(t) = \int [g_{b_n(t-u),x}(T_i, \delta_i) - Eg_{b_n(t-u),x}(T_i, \delta_i)]\, dK_0(u).$$

This does not only imply the third condition in van der Vaart and Wellner (1996), but also the second, since the partitions are independent of n. Since the function $g_{b_nt,x}(z, \delta)$ consists of three terms, also $W_{ni}(t)$ can be decomposed into three terms. The most difficult term to deal with is the second one, which equals $Z_{ni}(t) - E[Z_{ni}(t)]$, where

$$Z_{ni}(t) = \int \frac{I(T_i \le b_n(t-u), \delta_i = 1)}{1 - H_x(b_n(t-u))}\, dK_0(u).$$

We will prove that (5.1) is satisfied if $W_{ni}(t)$ is replaced by $Z_{ni}(t)$ (the derivation for $E[Z_{ni}(t)]$ follows immediately by integration by parts). Consider

$$Z_{ni}(t) - Z_{ni}(t')$$

$$= \int \frac{I(T_i \le b_n(t-u), \delta_i = 1) - I(T_i \le b_n(t'-u), \delta_i = 1)}{1 - H_x(b_n(t-u))}\, dK_0(u)$$
$$+ \int I(T_i \le b_n(t'-u), \delta_i = 1) \left[\frac{1}{1 - H_x(b_n(t-u))} - \frac{1}{1 - H_x(b_n(t'-u))}\right] dK_0(u).$$

We concentrate on the first term, which equals

$$I(\delta_i = 1) \int_{t' - \frac{T_i}{b_n}}^{t - \frac{T_i}{b_n}} \frac{1}{1 - H_x(b_n(t-u))}\, dK_0(u)$$
$$= I(\delta_i = 1) \left\{ \frac{K_0(t - \frac{T_i}{b_n})}{1 - H_x(T_i)} - \frac{K_0(t' - \frac{T_i}{b_n})}{1 - H_x(T_i + b_n(t-t'))} \right\}$$
$$+ b_n I(\delta_i = 1) \int_{t' - \frac{T_i}{b_n}}^{t - \frac{T_i}{b_n}} \frac{K_0(u)}{(1 - H_x(b_n(t-u)))^2} h_x(b_n(t-u))\, du$$
$$= \alpha_{ni1}(x, t, t') + \alpha_{ni2}(x, t, t').$$

The term $\alpha_{ni2}(x, t, t')$ is bounded by $K\varepsilon b_n$ and hence

$$n h_n b_n^{-1} \sum_{i=1}^n w_{ni}^2(x; h_n) E \sup_{t_j \le t, t' \le t_{j+1}} |\alpha_{ni2}(x, t, t')|^2 \le K\varepsilon^2 b_n.$$

For $\alpha_{ni1}(x, t, t')$ it suffices to consider

$$b_n^{-1} E \sup_{t_j \le t, t' \le t_{j+1}} \left| K_0\left(t - \frac{T_i}{b_n}\right) - K_0\left(t' - \frac{T_i}{b_n}\right)\right|^2$$
$$= b_n^{-1} E \left[\sup_{t_j \le t, t' \le t_{j+1}} \left| K_0\left(t - \frac{T_i}{b_n}\right) - K_0\left(t' - \frac{T_i}{b_n}\right)\right|^2 I(b_n(t_{j+1} - L_0) \le T_i \le b_n(t_j + L_0))\right]$$
$$\le K\varepsilon^2.$$

This finishes the proof.

6 Uniform strong consistency result with rate of convergence

Theorem 6.1 *Assume (C1), (C2), $H_x(t)$ satisfies (C3), (C5) and (C6) and $H_x^u(t)$ satisfies (C3), (C5), (C8) and (C9) in $[0, T]$ with $T < T_{H_x}$, $h_n \to 0$, $b_n \to 0$, $nh_n \to \infty$.*

Then, as $n \to \infty$,

$$\sup_{0 \le t \le T} |\lambda_{xhb}(t) - \lambda_x(t)| = O((nh_n b_n)^{-1/2}(\log n)^{1/2} + h_n^2 + b_n^2) \quad a.s.$$

Proof. Write

$$\lambda_{xhb}(t) - \lambda_x(t) = \frac{1}{b_n}\int [\Lambda_{xh}(t - b_n u) - \Lambda_x(t - b_n u)]\, dK_0(u)$$
$$+\frac{1}{b_n}\int \Lambda_x(t - b_n u)\, dK_0(u) - \lambda_x(t).$$

The second term equals $\beta_{xhb}^{(1)}(t)$ (defined in Section 3) and hence this term is $O(b_n^2)$ uniformly in t. The first term can be written as

$$\frac{1}{b_n}\int \left[\int_0^{t-b_n u} \frac{dH_{xh}^u(s)}{1 - H_{xh}(s-)} - \int_0^{t-b_n u} \frac{dH_x^u(s)}{1 - H_x(s)}\right] dK_0(u)$$
$$= \frac{1}{b_n}\int \int_0^{t-b_n u} \frac{H_{xh}(s-) - H_x(s)}{(1 - H_{xh}(s-))(1 - H_x(s))} dH_{xh}^u(s)\, dK_0(u)$$
$$+\frac{1}{b_n}\int \int_0^{t-b_n u} \frac{d(H_{xh}^u(s) - H_x^u(s))}{1 - H_x(s)} dK_0(u)$$
$$= \alpha_{n1}(x,t) + \alpha_{n2}(x,t).$$

We start with $\alpha_{n1}(x,t)$:

$$\begin{aligned}
&|\alpha_{n1}(x,t)| \\
&= \frac{1}{b_n}\left|\int K_0\left(\frac{t-s}{b_n}\right) \frac{H_{xh}(s-) - H_x(s)}{(1 - H_{xh}(s-))(1 - H_x(s))} dH_{xh}^u(s)\right| \\
&\le \frac{K}{b_n}\sup_{0 \le s \le T} |H_{xh}(s-) - H_x(s)|\{H_{xh}^u(t + L_0 b_n) - H_{xh}^u(t - L_0 b_n)\} \\
&\le \frac{K}{b_n}\sup_{0 \le s \le T} |H_{xh}(s-) - H_x(s)|\{2\sup_{0 \le t \le T} |H_{xh}^u(t) - H_x^u(t)| + H_x^u(t + L_0 b_n) - H_x^u(t - L_0 b_n)\} \\
&= O((nh_n)^{-1/2}(\log n)^{1/2} + (nh_n b_n)^{-1}\log n + h_n^2 + h_n^2 b_n^{-1}(nh_n)^{-1/2}(\log n)^{1/2} + h_n^4 b_n^{-1})
\end{aligned}$$

a.s., since $\sup_{0 \le t \le T} |H_{xh}(t) - H_x(t)| = O((nh_n)^{-1/2}(\log n)^{1/2} + h_n^2)$ a.s. by Lemmas A1(b) and A3(a) in Van Keilegom and Veraverbeke (1997b) and since $\sup_{0 \le t \le T} |H_{xh}(t) -$

$H_{xh}(t-)| = O((nh_n)^{-1})$ a.s. The term $\alpha_{n2}(x,t)$ equals

$$\frac{1}{b_n}\int \frac{H^u_{xh}(t-b_nu)-H^u_x(t-b_nu)}{1-H_x(t-b_nu)}dK_0(u)$$
$$-\frac{1}{b_n}\int \int_0^{t-b_nu} \frac{H^u_{xh}(s)-H^u_x(s)}{(1-H_x(s))^2}dH_x(s)dK_0(u)$$
$$=\alpha_{n21}(x,t)+\alpha_{n22}(x,t).$$

Using integation by parts we can write

$$\begin{aligned}\alpha_{n22}(x,t) &= -\int K_0(u)\frac{H^u_{xh}(t-b_nu)-H^u_x(t-b_nu)}{(1-H_x(t-b_nu))^2}h_x(t-b_nu)\,du\\ &= O((nh_n)^{-1/2}(\log n)^{1/2}+h_n^2) \quad \text{a.s.}\end{aligned}$$

Since K_0 is of bounded variation, there exist increasing functions K_1 and K_2 defined on $[-L_0,L_0]$, such that $K_0=K_1-K_2$. Hence,

$$\begin{aligned}\sup_{0\le t\le T}|\alpha_{n21}(x,t)| \le{}& \frac{1}{b_n}\sup_{0\le t\le T}\left|\int \frac{H^u_{xh}(t-b_nu)-H^u_x(t-b_nu)}{1-H_x(t-b_nu)}dK_1(u)\right|\\ &+\frac{1}{b_n}\sup_{0\le t\le T}\left|\int \frac{H^u_{xh}(t-b_nu)-H^u_x(t-b_nu)}{1-H_x(t-b_nu)}dK_2(u)\right|.\end{aligned}$$

We consider the first term :

$$\begin{aligned}&\frac{1}{b_n}\int \frac{H^u_{xh}(t-b_nu)-H^u_x(t-b_nu)}{1-H_x(t-b_nu)}dK_1(u)\\ &=\frac{1}{b_n}\int (H^u_{xh}(t-b_nu)-H^u_x(t-b_nu))\left[\frac{1}{1-H_x(t-b_nu)}-\frac{1}{1-H_x(t)}\right]dK_1(u)\\ &+\frac{1}{b_n}\frac{1}{1-H_x(t)}\int (H^u_{xh}(t-b_nu)-H^u_x(t-b_nu))\,dK_1(u). \qquad (6.1)\end{aligned}$$

Clearly, the first term above is $O((nh_n)^{-1/2}(\log n)^{1/2}+h_n^2)$ a.s. Divide $[0,T]$ into subintervals $[t_j,t_{j+1}]$ ($j=1,\dots,O(a_n^{-1})$) of length Ka_n, where $a_n=c(nh_nb_n)^{-1/2}(\log n)^{1/2}$. Then, the second term of (6.1) multiplied with $1-H_x(T)$ is bounded by

$$\begin{aligned}&\frac{1}{b_n}\max_j\left|\int (H^u_{xh}(t_j-b_nu)-EH^u_{xh}(t_j-b_nu))\,dK_1(u)\right|\\ &+\frac{1}{b_n}\max_j\left|\int (EH^u_{xh}(t_j-b_nu)-H^u_x(t_j-b_nu))\,dK_1(u)\right|\\ &+\frac{1}{b_n}\max_j\left|\int (H^u_x(t_{j+1}-b_nu)-H^u_x(t_j-b_nu))\,dK_1(u)\right|. \qquad (6.2)\end{aligned}$$

The last term of the above sum is $O((nh_nb_n)^{-1/2}(\log n)^{1/2})$ upon integration by parts. The second term is $O(h_n^2 + n^{-1})$ by using integration by parts and equation (2.6). On the first term we apply Bernstein's inequality (see e.g. Serfling (1980)). We start with calculating the variance of $b_n^{-1} \int (H_{xh}^u(t_j - b_n u) - EH_{xh}^u(t_j - b_n u))\, dK_1(u)$, which equals

$$\frac{1}{b_n^2} \sum_{i=1}^n w_{ni}^2(x; h_n) \int \int \mathrm{Cov}(I(T_i \leq t_j - b_n u, \delta_i = 1), I(T_i \leq t_j - b_n v, \delta_i = 1))\, dK_1(u)\, dK_1(v)$$

$$= \frac{2}{b_n^2} \sum_{i=1}^n w_{ni}^2(x; h_n) \int \int_{-L_0}^{v} H_{x_i}^u(t_j - b_n v)(1 - H_{x_i}^u(t_j - b_n u))\, dK_1(u)\, dK_1(v)$$

$$\leq \frac{K}{nh_nb_n}$$

for some $K > 0$, after applying integration by parts. Hence, Bernstein's inequality together with Borel-Cantelli shows that the first term of (6.2) is $O((nh_nb_n)^{-1/2}(\log n)^{1/2})$ a.s.

References

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

Cox, D. R. (1972). Regression models and life-tables. *J. Royal Statist. Soc. Ser. B*, **34**, 187-202.

Dabrowska, D. M. (1987). Non-parametric regression with censored survival time data. *Scand. J. Statist.*, **14**, 181-197.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Ann. Statist.*, **17**, 1157-1167.

Diehl, S. and Stute, W. (1988). Kernel density and hazard function estimation in the presence of censoring. *J. Multiv. Anal.*, **25**, 299-310.

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 nonparametric regression for life history data using local linear fitting. *Ann. Statist.*, **23**, 787-823.

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. Th. Rel. 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.

McNichols, D.T. and Padgett, W.J. (1985). Nonparametric methods for hazard rate estimation from right-censored samples. *J. Chin. Statist. Assoc.*, **23**, 1-15.

Müller, H.-G. and Wang, J.-L. (1990). Locally adaptive hazard smoothing. *Prob. Th. Rel. Fields*, **85**, 523-538.

Nielsen, J.P. and Linton, O.B. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *Ann. Statist.*, **23**, 1735-1748.

Ramlau - Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.*, **11**, 453-466.

Rice, J. and Rosenblatt, M. (1976). Estimation of the log survivor function and hazard function. *Sankhyā Ser. A*, **38**, 60-78.

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

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics.*

Wiley, New York.

Singpurwalla, N.D. and Wong, M.-Y. (1983). Estimation of the failure rate - a survey of nonparametric methods, part I : non-bayesian methods. *Commun. Statist., - Theory Meth.*, **12**, 559-588.

Tanner, M.A. and Wong, W.H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.*, **11**, 989-993.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer-Verlag, New York.

Van Keilegom, I. and Veraverbeke, N. (1996). Uniform strong convergence results for the conditional Kaplan-Meier estimator and its quantiles. *Commun. Statist., - Theory Meth.*, **25**, 2251-2265.

Van Keilegom, I. and Veraverbeke, N. (1997a). Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process. *Commun. Statist., - Theory Meth.*, **26**, 853-869.

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

Van Keilegom, I. and Veraverbeke, N. (1998a). Bootstrapping quantiles in a fixed design regression model with censored data. *J. Statist. Planning Inf.*, **69**, 115-131.

Van Keilegom, I. and Veraverbeke, N. (1998b). Density and hazard estimation in censored regression models (submitted).

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

Watson, G.S. and Leadbetter, M.R. (1964b). Hazard analysis II. *Sankhyā Ser. A*, **26**, 110-116.

Yandell, B.S. (1983). Nonparametric inference for rates with censored survival data. *Ann. Statist.*, **11**, 1119-1135.