Ann. Inst. Statist. Math.
Vol. 54, No. 1, 19–28 (2002)

KERNEL DENSITY AND HAZARD RATE ESTIMATION FOR CENSORED DATA UNDER α-MIXING CONDITION

ECKHARD LIEBSCHER

Institute of Mathematics, Technical University Ilmenau, 98684 Ilmenau/Thür, Germany,
e-mail: lieb@mathematik.tu-ilmenau.de

(Received December 7, 1999; revised July 24, 2000)

Abstract. We derive rates of uniform strong convergence for kernel density estimators and hazard rate estimators in the presence of right censoring. It is assumed that the failure times (survival times) form a stationary α-mixing sequence. Moreover, we show that, by an appropriate choice of the bandwidth, both estimators attain the optimal strong convergence rate known from independent complete samples. The results represent an improvement over that of Cai's paper (cf. Cai (1998*b*, *J. Multivariate Anal.*, **67**, 23–34)).

Key words and phrases: Kernel density estimator, right censorship, strong convergence, hazard rate estimator.

1. Introduction

In this paper we consider the model of randomly right censorship. Let $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ be two independent sequences of nonnegative random variables. We regard $X_1, X_2, \ldots$ as failure times (or survival times) and suppose that $\{X_k\}$ form a stationary α-mixing sequence such that each X_i has the distribution function F. Assume that the distribution function F is absolutely continuous with a bounded density function f. The censoring times $Y_1, Y_2, \ldots$ are assumed to be independent and to have the common distribution function G. Let $Z_i = \min(X_i, Y_i)$ and $\delta_i = I(Z_i = X_i)$. $I(\cdot)$ denotes the indicator function. In contrast to statistics for complete data, we observe only the pairs $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$ and the estimators are based on these pairs.

The Kaplan-Meier estimator F_n^* for the distribution function F is defined by

$$F_n^*(x) = \begin{cases} 1 - \prod_{i: 1 \le i \le n, Z_{(i)} \le x} \left(\dfrac{n-i}{n-i+1}\right)^{\delta_{(i)}} & \text{for} \quad x < Z_{(n)}, \\ 1 & \text{otherwise} \end{cases} \tag{1.1}$$

(cf. Kaplan and Meier (1958)). $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$ denote the order statistics of $Z_1, \ldots, Z_n$, and $\delta_{(i)}$ is the concomitant of $Z_{(i)}$.

There is a wide range of literature on the Kaplan-Meier estimator (1.1) for censored independent observations. We refer to papers by Efron (1967), Breslow and Crowley (1974), Peterson (1977), Földes and Rejtö (1981), and Gu and Lai (1990). Martingale methods for analyzing properties of the Kaplan-Meier estimator are described in the monograph by Gill (1980). Ying and Wei (1994), Lecoutre and Ould-Saïd (1995*b*) and Cai (1998*a*) study the convergence of the Kaplan-Meier estimator for dependent data.

In their papers, Voelkel and Crowley (1984), Belyaev (1991) and Ying and Wei (1994) give examples of applications for censored dependent failure times.

Under the model of right censorship with independent failure times, kernel density estimators are studied by Mielniczuk (1986), Diehl and Stute (1988), Hentzschel and Liebscher (1990), and Xiang (1994). There are only few papers dealing with density estimation for dependent data under a censoring model. The recent paper by Cai (1998*b*) provides a rate of strong convergence for such kernel estimators when the failure times form a stationary α -mixing sequence. But the convergence rate proved by Cai is not the best possible one. The goal of this paper is to improve the result by Cai (1998*b*) and to derive rates of uniform strong convergence for the kernel estimator of the density f and the kernel estimator of the hazard rate. We prove that, under appropriate assumptions and by an appropriate choice of the bandwidth, the density estimator and the hazard rate one attain the optimal strong convergence rate. This rate coincides with the optimized convergence rate (optimized with respect to the bandwidth) in the case of independent complete data. The strong convergence of a hazard rate estimator for censored data was already examined by Lecoutre and Ould-Saïd (1995*a*) under α-mixing assumption. A generalization of our results to more complicated censoring models is straightforward. For example, one may consider censoring models where, in addition, failure and censoring times are dependent (cf. Tsai (1986), Liebscher and Schäbe (1997)). In this case, there is a problem of identifiability. Convergence properties of kernel density estimators for complete data arising from a stationary mixing sequence are investigated in the author's papers (Liebscher (1995, 1996, 2001)) where further references are given.

The paper is organized as follows: Section 2 contains the convergence rate of the kernel density estimator. In Section 3 we provide the convergence rate of the hazard rate estimator. The proofs are deferred to Section 4.

2. Main results

Let us define the α-mixing property of a sequence $\{\xi_k, k = 1, 2, \ldots\}$ of random variables. $\mathcal{F}_l^m$ denotes the σ-algebra generated by $\xi_l, \xi_{l+1}, \ldots, \xi_m$. The α-mixing coefficient of $\{\xi_k\}$ is given by

$$\alpha_m := \sup_{k=1,2,\ldots} \sup \left\{ |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_{m+k}^{\infty}, B \in \mathcal{F}_1^k \right\}.$$

If $\alpha_m \to 0$ as m tends to infinity, we say that the sequence $\{\xi_k\}$ is *α-mixing or strong mixing*. Properties of α-mixing sequences and examples of such sequences are analysed thoroughly in the monograph by Doukhan (1994).

Throughout this section, $\{X_k\}$ is assumed to be a stationary α -mixing sequence of random variables with mixing coefficients $\{\alpha_k\}$. Using the Kaplan-Meier estimator (1.1), the kernel estimator for the density f is given by

$$\hat{f}_n(x) = b(n)^{-1} \int_0^{+\infty} K\left(\frac{t-x}{b(n)}\right) dF_n^*(t) \qquad (x \in [0, \infty)) \tag{2.1}$$

(cf. for example Mielniczuk (1986)). We suppose that the bandwidth $b(n)$ satisfies

$$b(n) > 0, \quad \lim_{n\to\infty} b(n) = 0 \quad \text{and} \quad \lim_{n\to\infty} nb(n)\ln(n)^{-1} = \infty.$$

Let $H(t) = 1 - (1 - F(t))(1 - G(t))$ be the distribution function of Z_i and $t_H := \sup\{t : H(t) < 1\}$. The kernel function $K : \mathbb{R} \to \mathbb{R}$ and the sequence $\{X_k\}$ are assumed to satisfy the following conditions:

CONDITION. $\mathcal{K}(p)$, $p \geq 1$. K is a Lipschitz-continuous function with $K(t) = 0$ for $|t| > 1$,

$$\int_{-1}^{1} K(t)dt = 1, \qquad \int_{-1}^{1} t^j K(t)dt = 0 \quad (j = 1, \ldots, p-1).$$

CONDITION. $\mathcal{J}$. For all integers $j \geq j_0$, the joint density f_j of X_1 and X_{j+1}exists on $\mathbb{R}$. Furthermore, for some $\eta > 0$,

$$\begin{aligned} &f_j(x,y) \leq C_1 \quad \forall j \geq j_0, \quad \forall x,\ y \in [0, t_H] : |x - y| < \eta, \\ &f(x) \leq C_2 \quad \forall x \in [0, t_H] \end{aligned}$$

with positive constants C_1, C_2.

The following theorem gives the rate of uniform convergence of the kernel density estimator $\hat{f}_n$ on the interval $I = [\underline{a}, \bar{a}]$ with $\underline{a} > 0$, $\bar{a} < t_H$.

THEOREM 2.1. *Suppose that* $\alpha_k = O(k^{-r})$ $(r > 3)$ *is fulfilled, and conditions* $\mathcal{J}$ *and* $\mathcal{K}(p)$ *are satisfied for some integer* $p \geq 1$. *Assume that* f *has a bounded derivative of order* p *on* $[0, t_H]$. *Then*

$$\sup_{x \in I} |\hat{f}_n(x) - f(x)| = O\left(\sqrt{\frac{\ln(n)}{nb(n)}} + v_n + b^p(n)\right) \quad a.s. \tag{2.2}$$

where $v_n = (n^{1-r}\ln(n)^{r+1}\ln\ln^2(n))^{1/(r+2)}b(n)^{-1} \to 0$ *as* $n \to \infty$. *Moreover, choosing* $b(n) = \text{const} \cdot (n/\ln(n))^{-1/(2p+1)}$, *we obtain the optimized convergence rate*

$$\sup_{x \in I} |\hat{f}_n(x) - f(x)| = O\left(\left(\frac{n}{\ln(n)}\right)^{-p/(2p+1)}\right) \quad a.s. \tag{2.3}$$

provided that $r > 4 + 3/p$.

The second term of the right hand side of (2.2) is additional if compared with corresponding results in the case of uncensored independent samples. It describes the impact of the dependence of $X_1, X_2, \ldots$. Moreover, the convergence rate (2.3) coincides with the optimized convergence rate in the case of uncensored independent samples. In our considerations we restrict ourselves to the case $r > 3$ since this assumption is required to get the law of iterated logarithm for the empirical process and the Kaplan-Meier estimator (cf. Cai (1998*b*), Lemmas 2 and 3).

Now we consider the hazard rate estimator. If $\lambda(t) = f(t)/(1 - F(t))$ denotes the hazard rate of distribution F, then the hazard rate can be rewritten as follows

$$\lambda(t) = \frac{f^*(t)}{1 - H(t)}, \qquad f^*(t) = f(t)(1 - G(t)).$$

Therefore, a natural definition of the estimator for the hazard rate is given by

$$\hat{\lambda}_n(x) = \frac{\hat{f}_n^*(x)}{1 - H_n(x)} \qquad (x \in (0, t_H))$$

where

$$\hat{f}_n^*(x) = \frac{1}{nb(n)} \sum_{i=1}^{n} K\left(\frac{x - Z_i}{b(n)}\right) \delta_i \qquad \text{and} \qquad H_n(x) := n^{-1} \sum_{i=1}^{n} I(Z_i \le x).$$

H_n is the empirical distribution function of H and hence a consistent estimator for H. In the next section, we show that $\hat{f}_n^*$ is also a reasonable estimator for f^*. The following theorem provides the uniform convergence rate of the hazard rate estimator $\hat{\lambda}_n$.

THEOREM 2.2. *Let the assumptions of Theorem* 2.1 *be satisfied. Suppose that* f^* *has a bounded derivative of order* p *on* $[0, t_H]$. *Then*

$$\sup_{x \in I} |\hat{\lambda}_n(x) - \lambda(x)| = O\left(\sqrt{\frac{\ln(n)}{nb(n)}} + v_n + b^p(n)\right) \qquad a.s.$$

(v_n, I *as in Theorem* 2.1).

This convergence rate coincides with that of Theorem 2.1. By an appropriate choice of $b(n)$, we obtain the optimal convergence rate $O((n/\ln(n))^{-p/(2p+1)})$ for $\hat{\lambda}_n$ similarly to (2.3). In the one-dimensional case, Theorem 2.2 improves Theorem 3 in Lecoutre and Ould-Saïd (1995*a*). A generalization of Theorem 2.2 to higher dimensions is straightforward (cf. Lecoutre and Ould-Saïd (1995*a*)).

3. Proofs

Throughout this section, we assume that the settings of Sections 1 and 2 are valid and conditions $\mathcal{J}$ and $\mathcal{K}(1)$ are fulfilled. For brevity we write $\sup_x$ instead of $\sup_{x \in I}$, $I = [\underline{a}, \bar{a}]$, and $\bar{L}(t)$ instead of $1 - L(t)$ for any function L. Let $K_b(t) = b(n)^{-1} K(tb(n)^{-1})$. In the following N denotes the sub-distribution function for the failures alone, i.e.

$$N(x) := \mathbb{P}\{X_i \le x, \delta_i = 1\} = \int_0^x \bar{G}(t) dF(t) \qquad (x \in (0, \infty)).$$

The corresponding empirical function is defined by $N_n(x) := n^{-1} \sum_{i=1}^{n} I(Z_i \le x) \cdot \delta_i$. Then, for t with $H_n(t) < 1$, we have (cf. Peterson (1977))

$$F_n^*(t) = \int_0^t \frac{\bar{F}_n^*(s)}{\bar{H}_n(s)} dN_n(s).$$

Let v be a real number with $\bar{a} < v < t_H$. We introduce

$$f_n(x) := \int_0^\infty K_b(x - t) dF(t).$$

The proof of Theorem 2.1 is based on the following decompositions:

$$
\begin{aligned}
|\hat{f}_n(x) - f_n(x)| &\le |\hat{f}_n(x) - f_n(x)| + |f_n(x) - f(x)| \qquad \text{and} \\
|\hat{f}_n(x) - f_n(x)| &= \left| \int_0^\infty K_b(x-t)d(F_n^*(t) - F(t)) \right| \qquad (3.1)\\
&\le J_{1n}(x) + J_{2n}(x) + J_{3n}(x) \qquad \text{a.s.} \quad (x \in I)
\end{aligned}
$$

where

$$
\begin{aligned}
J_{1n}(x) &:= \left| \int_0^v K_b(x-t)\bar{G}(t)^{-1} d\left(N_n(t) - N(t)\right) \right|, \\
J_{2n}(x) &:= \left| \int_0^v K_b(x-t)\left(\bar{H}_n(t)^{-1}\bar{F}_n^*(t) - \bar{G}(t)^{-1}\right) dN_n(t) \right|, \\
J_{3n}(x) &:= \left| \int_v^{t_H} K_b(x-t) d\left(F_n^*(t) - F(t)\right) \right|.
\end{aligned}
$$

Since $f_n(x)$ coincides with the expectation of the estimator in the complete sample case (i.e. without censoring), we can utilize standard results about the convergence rate of $|f_n(x) - f(x)|$. In the sequel we prove that the rate of uniform strong convergence of the terms $J_{1n}(x)$ to $J_{3n}(x)$ is equal to $\rho_n := n^{-1/2}\sqrt{\ln(n)}b(n)^{-1/2} + v_n$ or faster, v_n as in Theorem 2.1. Obviously, $\sup_x J_{3n}(x) = 0$ holds for sufficiently large n since $b(n) < v - \bar{a}$ and $\sup_{u \ge (v-\bar{a})/b(n)} |K(u)| = 0$ for $n \ge n_1$.

LEMMA 3.1.

$$
\sup_x J_{1n}(x) = O(\rho_n) \qquad a.s. \quad and
$$

$$
\sup_x \left| \int_0^v |K_b(x-t)| \cdot \bar{G}(t)^{-1} d\left(N_n(t) - N(t)\right) \right| = O\left(\rho_n\right) \qquad a.s.
$$

For the proof of Lemma 3.1, we need the Bernstein-type inequality proved by the author (cf. Liebscher (2001), Proposition 5.1), and an inequality for variances of sums of random variables.

PROPOSITION 3.1. *Let $\{\xi_i\}_{i\in\mathbb{N}}$ be a stationary α-mixing sequence of real random variables with mixing coefficients $\{\bar{\alpha}_i\}$. Assume that $\mathbb{E}\xi_1 = 0$ and $|\xi_k| \le S < +\infty$ a.s. $(k = 1, \ldots, n)$. Then, for $n, m \in \mathbb{N}, 0 < m \le n/2$, for $\varepsilon > 0$,*

$$
\mathbb{P}\left\{ \left| \sum_{i=1}^n \xi_i \right| > \varepsilon \right\} \le 4\exp\left\{ -\frac{\varepsilon^2}{16}\left(nm^{-1}D_m + \frac{1}{3}\varepsilon S m \right)^{-1} \right\} + 32\frac{S}{\varepsilon}n\bar{\alpha}_m
$$

where $D_m = \max_{j \le 2m} \mathbb{D}^2(\sum_{i=1}^j \xi_i)$.

Define $U_{ni}(x) = K_b(x - X_i) \cdot \bar{G}(X_i)^{-1} \cdot \delta_i I(X_i \le v) n^{-1}$ $(i = 1, \ldots, n)$, $\tilde{U}_{ni}(x) = U_{ni}(x) - \mathbb{E}U_{ni}(x)$. Obviously, we have

$$
\begin{aligned}
J_{1n}(x) &= \left| \sum_{i=1}^n \tilde{U}_{ni}(x) \right| \qquad (x \in I) \quad \text{and} \\
\sup_x |U_{ni}(x)| &\le \bar{G}(v)^{-1} \sup_{t \in [-1,1]} |K(t)| n^{-1} b(n)^{-1}. \qquad (3.2)
\end{aligned}
$$

Since $\{X_k\}$ is a stationary α-mixing sequence of random variables (mixing coefficients $\{\alpha_k\}$), the sequence $\{(X_k, Y_k), k = 1, 2, \ldots\}$ has the same properties and mixing coefficients $\{\breve{\alpha}_k\}$ with $\breve{\alpha}_k \leq 4\alpha_k$ by Lemma 1 of Cai (1998*b*). Hence $\{U_{nk}(x), k = 1, 2, \ldots\}$ is a stationary α-mixing sequence of random variables with mixing coefficients $\{\tilde{\alpha}_k\}$, $\tilde{\alpha}_k \leq 4\alpha_k$ for any fixed n and x.

LEMMA 3.2. *Assume that $\alpha_k = O(k^{-r})$ $(r \geq 2)$. Then*

$$\max_{j \leq 2m} \sup_x \mathbb{D}^2 \left(\sum_{i=1}^{j} U_{ni}(x) \right) = O(mn^{-2} b(n)^{-1}).$$

PROOF. We have

$$\begin{aligned}
\sup_x \mathbb{D}^2 U_{nk}(x) &\leq n^{-2} \sup_x \int_0^v K_b^2(x-s) \cdot \bar{G}(s)^{-2} \mathbb{E}(I(X_i \leq Y_i) \mid X_i = s) dF(s) \\
&= n^{-2} \sup_x \int_0^v K_b^2(x-s) \cdot \bar{G}(s)^{-1} f(s) ds \\
&\leq n^{-2} \bar{G}(v)^{-1} b(n)^{-1} \sup_x \int_{(x-v)/b(n)}^{x/b(n)} K^2(t) f(x - tb(n)) dt \\
&\leq n^{-2} \bar{G}(v)^{-1} b(n)^{-1} \int_{-1}^{1} K^2(t) dt \sup_{t \in [0,\infty)} f(t) \qquad (k = 1, \ldots, n).
\end{aligned}$$

Moreover,

$$\begin{aligned}
&|\operatorname{cov}(U_{nk}(x), U_{nl}(x))| \\
&\quad = n^{-2} \left| \int_0^v \int_0^v K_b(x-t) \bar{G}(t)^{-1} \mathbb{P}\{Y_k \geq t\} K_b(x-u) \bar{G}(u)^{-1} \mathbb{P}\{Y_l \geq u\} \right. \\
&\qquad\qquad \left. \cdot (f_{l-k}(t,u) - f(t)f(u)) du dt \right| \\
&\quad = n^{-2} \left| \int_0^v \int_0^v K_b(x-t) K_b(x-u) (f_{l-k}(t,u) - f(t)f(u)) du dt \right| \\
&\quad \leq n^{-2} \int_{(x-v)/b(n)}^{x/b(n)} \int_{(x-v)/b(n)}^{x/b(n)} |K(t)||K(u)| \\
&\qquad \cdot |f_{l-k}(x - tb(n), x - ub(n)) - f(x - tb(n)) f(x - ub(n))| du dt
\end{aligned}$$

for $l > k$. Since condition $\mathcal{J}$ is valid and $\underline{a}/b(n) \geq 1$, $(\bar{a} - v)/b(n) \leq -1, 2b(n) \leq \eta$ for $n \geq n_2$, we get

$$\max_{\substack{k,l=1,\ldots,n \\ l \geq k+j_0}} \sup_x |\operatorname{cov}(U_{nk}(x), U_{nl}(x))| = O(n^{-2}).$$

An application of Lemma 2.3 in Liebscher (1996) yields

$$\max_{j \leq 2m} \sup_x \mathbb{D}^2 \left(\sum_{i=1}^{j} U_{ni}(x) \right) = O(mn^{-2}(b(n)^{-2/r} + b(n)^{-1}))$$

which implies Lemma 3.2. □

PROOF OF LEMMA 3.1. We prove only the first part since the proof of the second part follows analogously. An application of Proposition 3.1, Lemma 3.2 and (3.2) leads to

$$\sup_x \mathbb{P}\{J_{1n}(x) > \varepsilon\} \le 4 \cdot \exp\{-C_3\varepsilon^2 \cdot (n^{-1}b(n)^{-1} + \varepsilon m n^{-1}b(n)^{-1})^{-1}\} + \frac{C_4}{\varepsilon b(n) m^r}$$

for any $\varepsilon > 0$ with positive constants C_3, C_4 not depending on n, ε or m. Let $\{a_n\}$ and $\{\lambda_n\}$ be sequences of positive real numbers such that $\lambda_n \le a_n b(n)$, $\lim_{n\to\infty} a_n = 0$. The compact interval I can be covered with closed intervals $D_1, \ldots, D_\nu \subset I$ having length λ_n and centres $u_1, \ldots, u_\nu$ such that $\nu \le \text{const} \cdot \lambda_n^{-1}$. Here we drop the dependence of ν on n . Further it follows that, for $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left\{\max_{j=1,\ldots,\nu} J_{1n}(u_j) > \varepsilon a_n\right\}\\
&\quad\le \nu \cdot \max_{j=1,\ldots,\nu} \mathbb{P}\{J_{1n}(u_j) > \varepsilon a_n\}\\
&\quad\le C_5 \cdot \lambda_n^{-1}(\exp\{-C_6\varepsilon^2 a_n^2(n^{-1}b(n)^{-1} + \varepsilon m n^{-1} b(n)^{-1} a_n)^{-1}\} + \varepsilon^{-1}a_n^{-1}b(n)^{-1}m^{-r})\\
&\quad\le C_5 \cdot \lambda_n^{-1}(\exp\{-C_6\varepsilon^2 a_n^2(n^{-1}b(n)^{-1} + \varepsilon \ln(n)^{-1})^{-1}\} + n\ln(n)^{-1}m^{-r-1})
\end{aligned}$$

provided that $n^{-1}b(n)^{-1}ma_n^{-1} \le \ln(n)^{-1}$. C_5, $C_6 > 0$ are constants not depending on n, m or ε. We choose $\lambda_n := \lceil n^2 \ln\ln^2(n) m^{-1-r}\rceil$ such that $\sum_{n=1}^\infty \lambda_n^{-1} n \ln(n)^{-1} m^{-1-r} < +\infty$. Hence the series $\sum_{n=1}^\infty \mathbb{P}\{\sup_x J_{1n}(x) > \varepsilon a_n\}$ converges for some large $\varepsilon > 0$ provided that

$$n^{-1}b(n)^{-1}a_n^{-2} \le \ln(n)^{-1}, \qquad n^{-1}b(n)^{-1}ma_n^{-1} \le \ln(n)^{-1}.$$

An application of the Borel-Cantelli lemma yields

$$\max_{j=1,\ldots,\nu} J_{1n}(u_j) = O(a_n) \quad \text{a.s.,} \quad a_n := n^{-1/2}b(n)^{-1/2}\sqrt{\ln(n)} + n^{-1}\ln(n)b(n)^{-1}m. \tag{3.3}$$

Let $\tilde{K}(t) = 1$ for $t \in [-2, 2]$, $\tilde{K}(t) = 0$ otherwise. We have $\sup_{x\in D_j} |x - u_j| \le \lambda_n$ and $\lambda_n \le b(n)$ for $n \ge n_3$. By the Lipschitz property of K,

$$\begin{aligned}
&\sup_{x\in D_j} |J_{1n}(x) - J_{1n}(u_j)| \\
&\quad= \sup_{x\in D_j}\left|\int_0^v (K_b(x-t) - K_b(u_j - t))\bar{G}(t)^{-1}d(N_n(t) - N(t))\right|\\
&\quad\le C_7 b(n)^{-2}\lambda_n \int_0^v \tilde{K}((u_j - t)/b(n))\bar{G}(t)^{-1}d(N_n(t) + N(t)).
\end{aligned} \tag{3.4}$$

$C_7 > 0$ is a constant not depending on j, n. Analogously to (3.3), one proves

$$\max_{j=1,\ldots,\nu}\left|b(n)^{-1}\int_0^v \tilde{K}((u_j - t)/b(n))\bar{G}(t)^{-1}d(N_n(t) - N(t))\right| = O(a_n) \qquad \text{a.s.} \tag{3.5}$$

Obviously,

$$\begin{aligned}
\max_{j=1,\ldots,\nu}\int_0^v \tilde{K}((u_j - t)/b(n))\bar{G}(t)^{-1}dN(t) &= b(n)\max_{j=1,\ldots,\nu}\int_{(u_j-v)/b(n)}^{u_j/b(n)} \tilde{K}(t)f(u_j - tb(n))dt\\
&\le 4b(n)\sup_{t\in[0,v]} f(t).
\end{aligned}$$

Therefore, combining (3.4) and (3.5), we obtain

$$\max_{j=1,\ldots,\nu} \sup_{x\in D_j} |J_{1n}(x) - J_{1n}(u_j)| = O(b(n)^{-1}\lambda_n(a_n+1)) \qquad \text{a.s.}$$

and, together with (3.3),

$$\sup_x J_{1n}(x) = O(b(n)^{-1}\lambda_n + a_n) \qquad \text{a.s.}$$

A minimization of the convergence rate with respect to m leads to $m := \lceil (n^3 \ln \ln^2(n) \ln(n)^{-1})^{1/(r+2)} \rceil$. This completes the proof of Lemma 3.1. □

The next step is to verify the convergence rate of $J_{2n}(x)$.

LEMMA 3.3. *For any $\tau \in (0, t_H)$, we have*

$$\sup_{t\in[0,\tau]} |F_n^*(t) - F(t)| = O\left(\sqrt{\frac{\ln\ln(n)}{n}}\right) \qquad a.s.$$

PROOF. This lemma is a consequence of Lemma 2 of Cai (1998*b*) and the derivations on page 33 of Cai's paper. □

LEMMA 3.4. $\sup_x J_{2n}(x) = O(\rho_n)$ *a.s.*

PROOF By virtue of Theorem 3.2 of Cai and Roussas (1992),

$$\sup_{t\in[0,v]} |H_n(t) - H(t)| = O\left(\sqrt{\frac{\ln\ln(n)}{n}}\right) \qquad \text{a.s.} \tag{3.6}$$

and

$$\bar{H}_n(v) \geq \bar{H}(v) - O\left(\sqrt{\frac{\ln\ln(n)}{n}}\right) \qquad \text{a.s.} \tag{3.7}$$

Observe that

$$|\bar{H}_n(t)^{-1}\bar{F}_n^*(t)\bar{G}(t) - 1| \leq \bar{H}_n(v)^{-1}\left(|\bar{F}_n^*(t) - \bar{F}(t)|\bar{G}(t) + |\bar{H}_n(t) - \bar{H}(t)|\right)$$

($t \in [0, v]$). Then, by (3.6), (3.7) and Lemma 3.3,

$$\begin{aligned}
&\sup_{t\in[0,v]} |\bar{H}_n(t)^{-1}\bar{F}_n^*(t)\bar{G}(t) - 1| \\
&\quad\leq \left(\bar{H}(v) - \sup_{s\in[0,v]} |H_n(s) - H(s)|\right)^{-1} \\
&\qquad\cdot \left(\sup_{t\in[0,v]} |F_n^*(t) - F(t)| + \sup_{t\in[0,v]} |H_n(t) - H(t)|\right) \\
&\quad= O\left(\sqrt{\frac{\ln\ln(n)}{n}}\right) \qquad \text{a.s.}
\end{aligned}$$

and

$$\sup_x J_{2n}(x) \le O\left(\sqrt{\frac{\ln\ln(n)}{n}}\right) \cdot \sup_x \int_0^v |K_b(x-t)| \cdot \bar{G}(t)^{-1} dN_n(t) \qquad \text{a.s.} \tag{3.8}$$

Moreover, we conclude

$$\sup_x \int_0^v |K_b(x-t)| \cdot \bar{G}(t)^{-1} dN(t) = \sup_x \int_0^v |K_b(x-t)| f(t) dt = O(1). \tag{3.9}$$

Lemma 3.4 is now a consequence of Lemma 3.1, (3.8) and (3.9). □

Proof of Theorem 2.1. By a standard argument, we derive $\sup_x |f_n(x) - f(x)| = O(b^p(n))$. Theorem 2.1 follows from (3.1) and Lemmas 3.1 and 3.4. □

At the end of this section, we derive the convergence rate for the hazard rate estimator. Using $U_{ni}(x) = K_b(x - X_i)\delta_i n^{-1}$ instead of the definition in Lemma 3.1, one can prove the following lemma in the same manner as Lemma 3.1.

Lemma 3.5.

$$\sup_x |\hat{f}_n^*(x) - \mathbb{E}\hat{f}_n^*(x)| = \sup_x \left| \int_0^\infty K_b(x-t) d(N_n(t) - N(t)) \right| = O(\rho_n) \qquad a.s.$$

Proof of Theorem 2.2. Observe that

$$\begin{aligned}
&\sup_x |\hat{\lambda}_n(x) - \lambda(x)| \\
&\le \sup_x \left(\left| \frac{\hat{f}_n^*(x) - f^*(x)}{\bar{H}(x)} \right| + |\hat{f}_n^*(x)| \left| \frac{\bar{H}_n(x) - \bar{H}(x)}{\bar{H}(x)\bar{H}_n(x)} \right| \right) \\
&\le \bar{H}(\bar{a})^{-1} \sup_x |\hat{f}_n^*(x) - f^*(x)| \\
&\quad + \left(\sup_x |\hat{f}_n^*(x) - f^*(x)| + \sup_x f(x) \right) \bar{H}(\bar{a})^{-1} \\
&\quad \cdot \left(\bar{H}(\bar{a}) - \sup_x |H_n(x) - H(x)| \right)^{-1} \sup_x |H_n(x) - H(x)|
\end{aligned} \tag{3.10}$$

where $f^*(t) = f(t)\bar{G}(x)$. By Taylor expansion, we arrive at

$$\sup_x |\mathbb{E}\hat{f}_n^*(x) - f^*(x)| = \sup_x \left| \int_{-1}^1 K(t)(f^*(x - tb(n)) - f^*(x)) dt \right| = O(b^p(n)).$$

Hence Theorem 2.2 is a consequence of Lemma 3.5, (3.6) and (3.10). □

Acknowledgements

The author wishes to thank the anonymous referees for their valuable suggestions.

References

Belyaev, Yu. K. (1991). Asymptotical properties of nonparametric point estimators based on complexly structured reliability data with right-censoring, *Statistics*, **22**, 589–612.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship, *Ann. Statist.*, **2**, 437–453.

Cai, Z. (1998*a*). Asymptotic properties of Kaplan-Meier estimator for censored dependent data, *Statist. Probab. Lett.*, **37**, 381–389.

Cai, Z. (1998*b*). Kernel density and hazard rate estimation for censored dependent data, *J. Multivariate Anal.*, **67**, 23–34.

Cai, Z. and Roussas, G. G. (1992). Uniform strong estimation under α-mixing, with rates, *Statist. Probab. Lett.*, **15**, 47–55.

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

Doukhan, P. (1994). *Mixing: Properties and Examples*, Lecture Notes in Statist, **85**, Springer, New York.

Efron, B. (1967). The two sample problem with censored data, *Proc. 5th Berkeley Symp. on Math. Statist. Prob.*, 831–853, University of California Press, Berkeley,.

Földes, A. and Rejtö, L. (1981). A LIL type result for the product limit estimator, *Z. Wahrsch. Verw. Gebiete*, **56**, 75–84.

Gill, R. (1980). *Censoring and Stochastic Integrals*, Mathematical Centre Tracts, Vol. 124, Mathematics Centrum, Amsterdam.

Gu, M. G. and Lai, T. L. (1990). Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation, *Ann. Probab.*, **18**, 160–189.

Hentzschel, J. and Liebscher, E. (1990). Density estimators for complete and censored samples, *Wissenschaftliche Zeitschrift der HfV Dresden*, Special Issue **57**, 32–57 (in German).

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

Lecoutre, J. P. and Ould-Saïd, E. (1995*a*). Hazard rate estimation for strong-mixing and censored processes, *J. Nonparametr. Statist.*, **5**, 83–89.

Lecoutre, J. P. and Ould-Saïd, E. (1995*b*). Convergence of the conditional Kaplan-Meier estimate under strong mixing, *J. Statist. Plann. Inference*, **44**, 359–369.

Liebscher, E. (1995). Strong convergence of sums of φ-mixing random variables, *Math. Methods Statist.*, **4**, 216–229.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation, *Stochastic Process. Appl.*, **65**, 69–80.

Liebscher, E. (2001). Estimation of the density and the regression function under mixing conditions, *Statist. Decisions*, **19**, 9–26.

Liebscher, E. and Schäbe, H. (1997). A generalization of the Kaplan-Meier estimator to Harris-recurrent Markov chains, *Statist. Papers*, **38**, 63–75.

Mielniczuk, J. (1986). Some asymptotic properties of kernel estimators of a density function in case of censored data, *Ann. Statist.*, **14** , 766–773.

Peterson, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions, *J. Amer. Statist. Assoc.*, **72**, 854–858.

Tsai, W. Y. (1986). Estimation of survival curves from dependent censorship models via a generalized self-consistent property with nonparametric Bayesian estimation application, *Ann. Statist.*, **14**, 238–249.

Voelkel, J.G. and Crowley, J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations, *Ann. Statist.*, **12**, 142–160.

Xiang, X. (1994). Law of the logarithm for density and hazard rate estimation for censored data, *J. Multivariate Anal.*, **49**, 278–286.

Ying, Z. and Wei, L. J. (1994). The Kaplan-Meier estimate for dependent failure time observations, *J. Multivariate Anal.*, **50**, 17–29.