Statistics and Probability Letters 79 (2009) 1809–1817

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

# Asymptotic expansion for ISE of kernel density estimators under censored dependent model

Vahid Fakoor, Sarah Jomhoori *, Hasanali Azarnoosh

*Department of statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, P.O. Box: 1159-91775, Iran*

ARTICLE INFO

*Article history:*
Received 19 September 2008
Received in revised form 4 May 2009
Accepted 6 May 2009
Available online 12 May 2009

ABSTRACT

In some long term studies, we encounter a series of dependent and censored observations. Randomly censored data consist of i.i.d. pairs of observations $(X_i, \delta_i)i = 1, \ldots, n$. If $\delta_i = 0$, $X_i$ denotes a censored observation, and if $\delta_i = 1$, $X_i$ denotes a survival time, which is the variable of interest. One of the global stochastic measures of the distance between a density and its kernel density estimator is integrated square error. In this paper, we apply the technique of strong approximation to establish an asymptotic expansion for the integrated square error of the kernel density estimate, when censored data are showing some kind of dependence.



## 1. Introduction and preliminaries

In medical follow-up or in engineering life testing studies, the life time variable may not be observable. Let $X_1, \ldots, X_n$ be a sequence of life times, having a common unknown continuous marginal distribution function $F$ with a density function $f = F'$ and hazard rate $\lambda = f/(1-F)$. The random variables are not assumed to be mutually independent. Let the random variable $X_i$ be censored on the right by the random variable $Y_i$, so that one observes only

$$Z_i = X_i \wedge Y_i \quad \text{and} \quad \delta_i = I(X_i \le Y_i),$$

where $\wedge$ denotes minimum and $I(.)$ is the indicator of the event specified in parentheses. In this random censorship model, the censoring times $Y_1, \ldots, Y_n$ are assumed to be independently and identically distributed and they are also assumed to be independent of the $X_i$'s. For easy reference, denote with $G$ the distribution of the $Y_i$'s. Since censored data traditionally occur in lifetime analysis, we assume that $X_i$ and $Y_i$ are nonnegative. The actually observed $Z_i$'s have a distribution function $H$ satisfying

$$\overline{H}(t) = 1 - H(t) = (1 - F(t))(1 - G(t)).$$

Denote by

$$F_*(t) = P(Z \le t, \delta = 1),$$

the sub-distribution function for the uncensored observations. Define

$$N_n(t) = \sum_{i=1}^{n} I(Z_i \le t, \delta = 1) = \sum_{i=1}^{n} I(X_i \le t \wedge Y_i),$$

* Corresponding author.
*E-mail addresses:* fakoor@math.um.ac.ir (V. Fakoor), sa_jo67@stu-mail.um.ac.ir, sara_jomhoori@yahoo.com (S. Jomhoori), azarnush@math.um.ac.ir (H. Azarnoosh).

0167-7152/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2009.05.003

the number of uncensored observations less than or equal to $t$, and

$$Y_n(t) = \sum_{i=1}^{n} I(Z_i \geq t),$$

the number of censored or uncensored observations greater than or equal to $t$ and also the empirical distribution functions of $\bar{H}(t)$ and $F_*(t)$ are respectively defined as

$$\overline{Y}_n(t) = n^{-1}Y_n(t), \qquad \overline{N}_n(t) = n^{-1}N_n(t).$$

Then the Kaplan–Meier estimator for $1 - F(t)$, based on the censored data is

$$1 - \hat{F}_n(t) = \prod_{s \leq t} \left(1 - \frac{\mathrm{d}N_n(s)}{Y_n(s)}\right), \quad t < Z_{(n)},$$

where $Z_{(i)}$'s are the order statistics of $Z_i$ and $dN_n(t) = N_n(t) - N_n(t-)$.

As is known (see, e.g. Gill, 1980), for a d.f. $F$ on $[0, \infty)$, the cumulative hazard function $\Lambda$ is defined by

$$\Lambda(t) = \int_0^t \frac{\mathrm{d}F(s)}{1 - F(s^-)},$$

and $\Lambda(t) = -\log(1 - F(t))$ for the case that $F$ is continuous. The empirical cumulative hazard function $\hat{\Lambda}_n(t)$ is given by

$$\hat{\Lambda}_n(t) = \int_0^t \frac{\mathrm{d}N_n(s)}{Y_n(s)},$$

which is called the Nelson–Aalen estimator of $\Lambda(t)$ in the literature.

Based on the Kaplan–Meier estimator, Blum and Susarla (1980) proposed to estimate $f$, by a sequence of kernel estimators $f_n$, defined by

$$f_n(t) = \frac{1}{h_n} \int K\left(\frac{t-s}{h_n}\right) \mathrm{d}\hat{F}_n(s), \tag{1.1}$$

where $K$ is a kernel function having finite support on $[-1, 1]$ and $h_n$ is a sequence of positive bandwidths tending to zero as $n \to \infty$. As an estimator for $\lambda$, we shall consider

$$\lambda_n(t) = \frac{1}{h_n} \int K\left(\frac{t-s}{h_n}\right) \mathrm{d}\hat{\Lambda}_n(s). \tag{1.2}$$

It is well known that the most widely accepted stochastic measure of the global performance of a kernel estimator is its integrated square error ($ISE$), defined by

$$ISE(f_n) = \int (f_n(t) - f(t))^2 \mathrm{d}t. \tag{1.3}$$

Indeed, it is often suggested that $f_n$ be constructed to minimize mean integrated square error ($MISE$), defined by

$$MISE(f_n) = \int E\,(f_n(t) - f(t))^2\, \mathrm{d}t,$$

in an asymptotic sense. The asymptotic behavior of $ISE$ has been studied extensively by many authors. In the uncensored case, Bickel and Rosenblatt (1973) employed the uniform strong approximation of the empirical process by the Brownian bridge to obtain a central limit theorem for the $ISE$ of the Rosenblatt–Parzen kernel estimators of a density function. Hall (1982) established an asymptotic expansion in probability for the integrated square error of kernel density estimator using strong approximation. Hall (1984) derived central limit theorem for the $ISE$ of density estimator using martingale theory and U-statistics approach. In the right censored case, Yang (1993) employed the martingale techniques by Gill (1983) to get a central limit theorem for the $ISE$ of the product limit kernel density estimators. Zhang (1998) applied the technique of strong approximation to establish an asymptotic expansion for $ISE$ of the kernel density estimate $f_n$. Sun and Zheng (1999) proved a central limit theorem for the $ISE$ of the kernel hazard rate estimators in left truncated and right censored data.

However, for the case that censored observations are dependent, there are hardly few results available. Jomhoori et al. (2007) studied the central limit theorem for $ISE$ of the kernel hazard rate estimator under dependent censorship. The main aim of this paper is to derive an asymptotic expansion for integrated square error of kernel density and hazard estimates, for the case in which the underlying lifetime are assumed to be $\alpha$-mixing whose definition is given below. For easy reference, let us recall the following definition.

**Definition 1.** Let $\{X_i, i \geq 1\}$ denote a sequence of random variables. Given a positive integer $m$, set

$$\alpha(m) = \sup_{k \geq 1}\{|P(A \cap B) - P(A)P(B)|;\ A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+m}^{\infty}\}, \tag{1.4}$$

where $\mathcal{F}_i^k$ denote the $\sigma$-field of events generated by $\{X_j;\ i \leq j \leq k\}$. The sequence is said to be $\alpha$-mixing (strongly mixing) if the mixing coefficient $\alpha(m) \to 0$ as $m \to \infty$.

Among various mixing conditions used in the literature, $\alpha$-mixing, is reasonably weak and has many practical applications. There exist many processes and time series fulfilling the strong mixing condition. As a simple example we can consider the Gaussian AR(1) process for which

$$Z_t = \rho Z_{t-1} + \varepsilon_t,$$

where $|\rho| < 1$ and $\varepsilon_t$'s are independently identically distributed random variables with standard normal distribution. It can be shown (see Ibragimov and Linnik, 1971, pp. 312–313) that $\{Z_t\}$ satisfies strong mixing condition. The stationary autoregressive moving average (ARMA) processes, which are widely applied in time series analysis, are $\alpha$-mixing with exponential mixing coefficient, i.e., $\alpha(n) = e^{-\nu n}$ for some $\nu > 0$. The threshold models, the EXPAR models (see Ozaki, 1979), the simple ARCH models (see Engle, 1982; Masry and Tjostheim, 1995 and Masry and Tjostheim, 1997) and their extensions (see Diebolt and Guégan, 1993) and the bilinear Markovian models are geometrically strongly mixing under some general ergodicity conditions. Auestad and Tjostheim (1990) provided excellent discussions on the role of $\alpha$-mixing for model identification in nonlinear time series analysis.

Now, for the sake of simplicity, the assumptions used in this paper are as follows.

**Assumptions.** (1) Suppose that $\{X_i, i \geq 1\}$ is a sequence of stationary $\alpha$-mixing random variables with continuous distribution function $F$.

(2) Suppose that the censoring time variables $\{Y_i, i \geq 1\}$ are i.i.d. random variables with continuous distribution function $G$ and are independent of $\{X_i, i \geq 1\}$.

(3) $\alpha(n) = O(n^{-\nu})$ for some $\nu > 3$.

(4) The kernel function $K$ is symmetric, of bounded variation on $(-1, 1)$ and satisfies the following conditions:

$$\int_{-1}^{1} K(t)\mathrm{d}t = 1,$$

$$\int_{-1}^{1} tK(t)\mathrm{d}t = 0,$$

$$\int_{-1}^{1} t^2K(t)\mathrm{d}t = \sigma^2 \neq 0,$$

$$\int_{-1}^{1} |\mathrm{d}K(t)| = \nu.$$

(5) $f$ and $\lambda$ are twice continuously differentiable on $[0, \tau]$ where $\tau = \sup\{t : H(t) < 1\}$.

The layout of the paper is as follows. In Section 2, we construct a two parameter Gaussian process that strongly approximates two empirical processes. In Section 3, we apply the strong approximation result of Section 2 to establish an asymptotic expansion of integrated square error of the kernel density and hazard rate estimates.

## 2. Strong approximation for the empirical processes

In this section, we construct a two parameter mean zero Gaussian process that strongly uniformly approximates the empirical processes $Z_{n1}(t) = \sqrt{n}(\hat{\Lambda}_n(t) - \Lambda(t))$ and $Z_{n2}(t) = \sqrt{n}(\widehat{F}_n(t) - F(t))$.

**Theorem 1.** *Suppose that Assumptions* (1)–(3) *are satisfied. On a rich probability space, there exists a two parameter mean zero Gaussian process* $\{B(u, v)\ u, v \geq 0\}$ *such that,*

$$\sup_{t\geq 0} |Z_{n1}(t) - B(t, n)| = O((\log n)^{-\lambda}) \quad a.s., \tag{2.1}$$

$$\sup_{t\geq 0} |Z_{n2}(t) - (1 - F(t))B(t, n)| = O((\log n)^{-\lambda}) \quad a.s. \tag{2.2}$$

**Proof.** First, in view of Lemma 1 of Cai (1998), $\{(X_i, Y_i); i \geq 1\}$ is a sequence of $\alpha$-mixing random variables with mixing coefficient $4\alpha(n)$. In particular, so is $\{(X_i, Y_i); X_i \geq Y_i, i = 1, \ldots, n\}$. Then, it follows from Theorem 1 of Cai (1998) that

$$\sup_{t\geq 0} |\overline{Y}_n(x) - \overline{H}(x)| = O(a_n) \quad a.s., \tag{2.3}$$

where

$$a_n = \left(\frac{\log\log n}{n}\right)^{1/2}.$$

Furthermore, it follows from Theorem 3 of Dhompongsa (1984) that there exist two Kiefer processes $\{k^{(i)}(u, v); u, v \geq 0\}, i = 1, 2$ with covariance functions

$$E[k^{(i)}(u, v)k^{(i)}(u', v')] = \Gamma^{(i)}(u, u') \min(v, v'), \quad i = 1, 2$$

and $\Gamma^{(i)}(u,u')$ is defined by

$$\Gamma^{(i)}(u,u') = \mathrm{Cov}(g_1^{(i)}(u), g_1^{(i)}(u')) + \sum_{k=2}^{\infty}[\mathrm{Cov}(g_1^{(i)}(u), g_k^{(i)}(u')) + \mathrm{Cov}(g_1^{(i)}(u'), g_k^{(i)}(u))],$$

where $g_k^{(1)}(u) = I(Z_k \le u, \delta_k = 1) - F_*(u)$ and $g_k^{(2)}(u) = I(Z_k \le u) - H(u)$, such that, for some $\lambda > 0$ depending only on $\nu$, given in assumption (3),

$$\sup_{t\in\mathbb{R}} |\overline{N}_n(t) - F_*(t) - k^{(1)}(t,n)/n| = O(b_n), \quad a.s. \tag{2.4}$$

and

$$\sup_{t\in\mathbb{R}} |\overline{Y}_n(t) - \overline{H}(t) - k^{(2)}(t,n)/n| = O(b_n), \quad a.s. \tag{2.5}$$

where

$$b_n = n^{-1/2}(\log n)^{-\lambda}.$$

Now, usual decomposition of $Z_{n1}(t)$ implies

$$\begin{aligned} Z_{n1}(t) &= \sqrt{n}[\hat{\Lambda}_n(t) - \Lambda(t)] = \sqrt{n}\left[\int_0^t \frac{\mathrm{d}\overline{N}_n(x)}{\overline{Y}_n(x)} - \int_0^t \frac{\mathrm{d}F_*(x)}{\overline{H}(x)}\right] \\ &= \sqrt{n}\int_0^t \frac{[\overline{H}(x) - \overline{Y}_n(x)]}{\overline{H}^2(x)}\mathrm{d}F_*(x) + \sqrt{n}\int_0^t \frac{\mathrm{d}[\overline{N}_n(x) - F_*(x)]}{\overline{H}(x)} + R_{n1}(t) \end{aligned}$$

where

$$\begin{aligned} n^{-1/2}R_{n1}(t) &= \int_0^t \frac{(\overline{Y}_n(x) - \overline{H}(x))^2}{\overline{Y}_n(x)\overline{H}^2(x)}\mathrm{d}F_*(x) + \int_0^t \left(\frac{1}{\overline{Y}_n(x)} - \frac{1}{\overline{H}(x)}\right)\mathrm{d}[\overline{N}_n(x) - F_*(x)] \\ &= I_1 + I_2. \end{aligned}$$

Define, for $t \ge 0$, a two parameter Gaussian process

$$B(t,n) = \frac{k^{(1)}(t,n)/\sqrt{n}}{\overline{H}(t)} - \int_0^t \frac{k^{(1)}(x,n)/\sqrt{n}}{\overline{H}^2(x)}\mathrm{d}H(x) - \int_0^t \frac{k^{(2)}(x,n)/\sqrt{n}}{\overline{H}^2(x)}\mathrm{d}F_*(x).$$

Clearly $E(B(t,n)) = 0$. Let

$$\beta_1(t,n) = \sqrt{n}(\overline{N}_n(t) - F_*(t)) - k^{(1)}(t,n)/\sqrt{n},$$

and

$$\beta_2(t,n) = \sqrt{n}(\overline{Y}_n(t) - \overline{H}(t)) - k^{(2)}(t,n)/\sqrt{n}.$$

Theorem 1 is about the order of

$$\sup_{t\ge 0} |Z_{n1}(t) - B(t,n)| = \sup_{t\ge 0} |R_{n1}(t) + R_{n2}(t)|, \tag{2.6}$$

where

$$R_{n2}(t) = \frac{\beta_1(t,n)}{\overline{H}(t)} - \int_0^t \frac{\beta_1(x,n)}{\overline{H}^2(x)}\mathrm{d}H(x) - \int_0^t \frac{\beta_2(x,n)}{\overline{H}^2(x)}\mathrm{d}F_*(x).$$

To deal with $R_{n1}(t)$, we deduce from (2.3)

$$I_1 = O(a_n^2) \quad a.s. \tag{2.7}$$

To estimate $I_2$, divide the interval $[0,\tau]$ into subintervals $[x_i, x_{i+1}]$, $i = 1, \dots, k_n$ where $k_n = O(a_n^{-1})$, and $0 = x_1 < x_2 < \cdots < x_{k_{n+1}} = \tau$ are such that $H(x_{i+1}) - H(x_i) = O(a_n)$. It is easy to check that

$$\begin{aligned} |I_2| &= \int_0^t \left(\frac{1}{\overline{Y}_n(x)} - \frac{1}{\overline{H}(x)}\right)\mathrm{d}[\overline{N}_n(x) - F_*(x)] \\ &\le 2\max_{1\le i\le k_n}\sup_{y\in[x_i,x_{i+1}]} |\bar{Y}_n^{-1}(y) - \bar{Y}_n^{-1}(x_i) - \bar{H}^{-1}(y) + \bar{H}^{-1}(x_i)| \\ &\quad + k_n \sup_{0\le x\le\tau} |\bar{Y}_n^{-1}(x) - \bar{H}^{-1}(x)| \max_{1\le i\le k_n} |\overline{N}_n(x_{i+1}) - \overline{N}_n(x_i) - F_*(x_{i+1}) + F_*(x_i)| \end{aligned}$$

$$\leq C \max_{1\leq i\leq k_n} \sup_{y\in[x_i,x_{i+1}]} |\overline{Y}_n(y) - \overline{Y}_n(x_i) - \overline{H}(y) + \overline{H}(x_i)| + C \max_{1\leq i\leq k_n} |\overline{N}_n(x_{i+1}) - \overline{N}_n(x_i) - F_*(x_{i+1}) + F_*(x_i)| + O(a_n^2)$$

$$\leq \max_{1\leq i\leq k_n} \left\{ \sup_{y\in[x_i,x_{i+1}]} |k^{(2)}(y,n) - k^{(2)}(x_i,n)|/n + |k^{(1)}(x_{i+1},n) - k^{(1)}(x_i,n)|/n \right\} + O(b_n).$$

Theorem 1.15.2 of Csörgő and Révész (1981) implies

$$I_2 = O\left(\left(\frac{\log\log k_n}{nk_n}\right)^{1/2}\right) + O(b_n) = O(b_n) \quad a.s. \tag{2.8}$$

Therefore, by combining (2.7) and (2.8), we have

$$\sup_{t\geq 0} |R_{n1}(t)| = O((\log n)^{-\lambda}) \quad a.s. \tag{2.9}$$

Next, by applying (2.4) and (2.5), we have

$$\sup_{t\geq 0} |R_{n2}(t)| = O((\log n)^{-\lambda}) \quad a.s. \tag{2.10}$$

Combining (2.6), (2.9) and (2.10) we obtain (2.1). It can be shown that

$$\widehat{F}_n(t) - F(t) = (1 - F(t))[\hat{\Lambda}_n(t) - \Lambda(t)] + O\left(\frac{\log\log n}{n}\right) \quad a.s. \tag{2.11}$$

Therefore (2.2) is proved via (2.11). □

**Remark 1.** In the $\alpha$-mixing case, we cannot achieve the same rate as in the iid case i.e. $O(n^{-1/2}(\log n)^2)$ (see Burke et al., 1988, Theorem 1). The main reason is that our approach utilizes the strong approximation introduced by Dhompongsa (1984) as a kiefer process with a negligible reminder term of order $O(n^{-1/2}(\log n)^{-\lambda})$. This is not as sharp as in iid case.

## 3. Integrated square error of the kernel estimators

In this section, we consider the *ISE* of the kernel density estimator on the interval $[0, \tau - \epsilon]$ and find an asymptotic expansion for this error in terms of the sample size $n$ and the bandwidth $h_n$. Also, we derive the same result for the *ISE* of $\lambda_n$. We shall only prove Theorem 2 in detail, since for $\lambda_n$ the arguments are similar.

For any $\epsilon > 0$, the integrated square error of $f_n$ on the interval $[0, \tau - \epsilon]$ is defined to be

$$ISE(f_n) = \int_0^{\tau-\epsilon} (f_n(t) - f(t))^2 \mathrm{d}t.$$

**Theorem 2.** *Let $h_n$ be a sequence of positive bandwidths satisfying $h_n = O(n^{-1/6})$ as $n \to \infty$. Suppose that assumptions (1)–(5) hold, then for any $\epsilon > 0$, we have*

$$ISE(f_n) = \frac{h_n^4\sigma_2^2}{4}\int_0^{\tau-\epsilon} [f''(t)]^2 \mathrm{d}t + \frac{\nu^2}{nh_n^2}\int_0^{\tau-\epsilon} \bar{F}^2(t)B^2(t,n)\mathrm{d}t + o_p(h_n^4) + o_p\left(\frac{1}{nh_n^2}\right), \tag{3.1}$$

*where $B(u, v)$ is the two parameter Gaussian process defined in Theorem 1.*

**Corollary 1.** *Under the same conditions of Theorem 2*

$$ISE(\lambda_n) = \frac{h_n^4\sigma_2^2}{4}\int_0^{\tau-\epsilon} [\lambda''(t)]^2 \mathrm{d}t + \frac{\nu^2}{nh_n^2}\int_0^{\tau-\epsilon} B^2(t,n)\mathrm{d}t + o_p(h_n^4) + o_p\left(\frac{1}{nh_n^2}\right). \tag{3.2}$$

**Remark 2.** In the iid case, when observations are subject to random right censoring, Zhang (1998), with optimal bandwidth of $h_n = O(n^{-1/5})$, established an asymptotic expansion for ISE. This condition, does not, seem to be in our work.

The proof of Theorem 2 is based on the following three lemmas. We begin with introducing some further notations. We define

$$\tilde{f}_n(t) = \frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{d}F(s),$$

$$Q_{n1} = \int_0^{\tau-\epsilon}\left[\int_{-1}^1 \bar{F}(t - h_n u)B(t - h_n u, n)\mathrm{d}K(u)\right]^2 w(t)\mathrm{d}t,$$

$$Q_{n2} = \int_0^{\tau-\epsilon}\left[\int_{-1}^1 \bar{F}(t - h_n u)B(t - h_n u, n)\mathrm{d}K(u)\right] w(t)\mathrm{d}t,$$

where $w(t)$ is some(measurable) function defined on $(0, \infty)$.

To study asymptotic expansion for ISE of Kernel density estimators, we also need to study the modulus of continuity of approximating process $B(u, v)$. In the first lemma, we prove the global modulus of continuity of the Gaussian process $B(u, v)$.

**Lemma 1.** *Let $h_n$ be a sequence of positive numbers for which*

$$\lim_{n\to\infty} \frac{\log h_n^{-1}}{\log\log n} = \infty. \tag{3.3}$$

*Then*

$$\sup_{0\le t\le\tau-\epsilon} \sup_{-1\le u\le 1} |B(t-h_nu, n) - B(t, n)| = O\left(\sqrt{2h_n \log h_n^{-1}}\right) \quad a.s. \tag{3.4}$$

**Proof.** First, we have

$$\begin{aligned}
|B(t-h_nu, n) - B(t, n)| &\le \left| \frac{k^{(1)}(t-h_nu, n)/\sqrt{n}}{\overline{H}(t-h_nu)} - \frac{k^{(1)}(t, n)/\sqrt{n}}{\overline{H}(t)} \right| + \sup_{0\le x\le\tau-\epsilon} \left| \frac{k^{(1)}(x, n)}{\sqrt{n}} \right| \left| \frac{\overline{H}(t-h_nu) - \overline{H}(t)}{\overline{H}^2(\tau)} \right| \\
&\quad + \sup_{0\le x\le\tau-\epsilon} \left| \frac{k^{(2)}(x, n)}{\sqrt{n}} \right| \left| \frac{F_*(t-h_nu) - F_*(t)}{\overline{H}^2(\tau)} \right| \\
&= I_1 + I_2 + I_3.
\end{aligned}$$

It can be shown, after simple algebra that for large $n$,

$$\begin{aligned}
\sup_{0\le t\le\tau-\epsilon} \sup_{-1\le u\le 1} \left| \frac{k^{(1)}(t-h_nu, n)/\sqrt{n}}{\overline{H}(t-h_nu)} - \frac{k^{(1)}(t, n)/\sqrt{n}}{\overline{H}(t)} \right| &= \sup_{0\le x\le\tau-\epsilon} \sup_{0\le y\le h_n} \left| \frac{k^{(1)}(x+y, n) - k^{(1)}(x, n)}{\overline{H}(\tau)\sqrt{n}} \right| \\
&\quad + \sup_{0\le t\le\tau-\epsilon} \sup_{-1\le u\le 1} \left| \frac{k^{(1)}(t, n)}{\sqrt{n}} \right| \left| \frac{1}{\bar{H}(t-h_nu)} - \frac{1}{\bar{H}(t)} \right| \\
&= I_{11} + I_{12}.
\end{aligned}$$

By the global modulus of continuity for the Kiefer processes (see Theorem 1.15.2 of Csörgő and Révész, 1981), we have

$$I_{11} = O\left(\sqrt{h_n \log h_n^{-1}}\right) \quad a.s. \tag{3.5}$$

To deal with $I_{12}$, according to the Mean Value Theorem and the law of iterated logarithm for the Kiefer processes (see Theorem A of Berks and Philipp, 1977), we have

$$I_{12} = O(h_n\sqrt{\log\log n}) \quad a.s. \tag{3.6}$$

It follows from (3.5) and (3.6) that

$$\sup_{0\le t\le\tau-\epsilon} \sup_{-1\le u\le 1} I_1 = O\left(\sqrt{h_n \log h_n^{-1}}\right) \quad a.s. \tag{3.7}$$

Likewise, we observe that

$$\sup_{0\le t\le\tau-\epsilon} \sup_{-1\le u\le 1} I_2 = O(h_n\sqrt{\log\log n}) \quad a.s., \tag{3.8}$$

$$\sup_{0\le t\le\tau-\epsilon} \sup_{-1\le u\le 1} I_3 = O(h_n\sqrt{\log\log n}) \quad a.s. \tag{3.9}$$

Therefore, Eqs. (3.7)–(3.9) imply (3.4). □

The next lemma establishes an asymptotic expansion for $Q_{n1}$.

**Lemma 2.** *Let $f(t)$ and $w(t)$ be continuous on $[0, \tau]$. Under assumptions (1)–(5) and for any $\epsilon > 0$, we have*

$$Q_{n1} = \nu^2 \int_0^{\tau-\epsilon} \bar{F}^2(t) B^2(t, n) |w(t)| \mathrm{d}t + O_p\left(\sqrt{2h_n \log h_n^{-1}}\right).$$

**Proof.** Simple algebra shows

$$\begin{aligned}
Q_{n1} &= \int_0^{\tau-\epsilon} \left\{ \int_{-1}^{1} \bar{F}(t-h_nu)[B(t-h_nu, n) - B(t, n)]\mathrm{d}K(u) + \int_{-1}^{1} \bar{F}(t-h_nu) B(t, n) \mathrm{d}K(u) \right\}^2 w(t)\mathrm{d}t \\
&= K_{n1} + K_{n2} + K_{n3},
\end{aligned} \tag{3.10}$$

where

$$K_{n1} = \int_0^{\tau-\epsilon} \left\{ \int_{-1}^{1} \bar{F}(t - h_n u)[B(t - h_n u, n) - B(t, n)]\mathrm{d}K(u) \right\}^2 w(t)\mathrm{d}t,$$

$$K_{n2} = \int_0^{\tau-\epsilon} B^2(t, n) \left\{ \int_{-1}^{1} \bar{F}(t - h_n u)\mathrm{d}K(u) \right\}^2 w(t)\mathrm{d}t,$$

$$K_{n3} = 2\int_0^{\tau-\epsilon} \left\{ \int_{-1}^{1} \bar{F}(t - h_n u)[B(t - h_n u, n) - B(t, n)]\mathrm{d}K(u) \right\} \times \left\{ \int_{-1}^{1} \bar{F}(t - h_n u, n)\mathrm{d}K(u) \right\} B(t, n)w(t)\mathrm{d}t.$$

To deal with $K_{n1}$, we apply Lemma 1

$$|K_{n1}| \le \int_0^{\tau-\epsilon} \left\{ |B(t - h_n u, n) - B(t, n)|\, |\bar{F}(t - h_n u)||\mathrm{d}K(u)| \right\}^2 |w(t)|\mathrm{d}t$$
$$= O_p(2h_n \log h_n^{-1}). \tag{3.11}$$

A Taylor expansion of $F$ yields

$$|K_{n2}| \le \int_0^{\tau-\epsilon} B^2(t, n) \left\{ \int_{-1}^{1} |\bar{F}(t - h_n u)||\mathrm{d}K(u)| \right\}^2 |w(t)|\mathrm{d}t$$
$$= \nu^2 \int_0^{\tau-\epsilon} \bar{F}^2(t)B^2(t, n)|w(t)|\mathrm{d}t + O_p(h_n). \tag{3.12}$$

Likewise, applying Lemma 1 gives

$$|K_{n3}| = O_p\left(\sqrt{2h_n \log h_n^{-1}}\right). \tag{3.13}$$

Combining (3.10) with (3.11)–(3.13) completes the proof.
□

The following lemma pertains to the asymptotic behavior for $Q_{n2}$.

**Lemma 3.** *Under the conditions of Lemma 2, we have for any $\epsilon > 0$*

$$Q_{n2} = O_p\left(\sqrt{2h_n \log h_n^{-1}}\right). \tag{3.14}$$

**Proof.** First, we can write

$$Q_{n2} = K_{n4} + K_{n5} + K_{n6}, \tag{3.15}$$

where

$$K_{n4} = \int_0^{\tau-\epsilon} \left[ \int_{-1}^{1} \left(\bar{F}(t - h_n u) - \bar{F}(t)\right) (B(t - h_n u, n) - B(t, n))\, \mathrm{d}K(u) \right] w(t)\mathrm{d}t,$$

$$K_{n5} = \int_0^{\tau-\epsilon} B(t, n) \left[ \int_{-1}^{1} \left(\bar{F}(t - h_n u) - \bar{F}(t)\right) \mathrm{d}K(u) \right] w(t)\mathrm{d}t,$$

$$K_{n6} = \int_0^{\tau-\epsilon} \left[ \int_{-1}^{1} \bar{F}(t)B(t - h_n u, n)\mathrm{d}K(u) \right] w(t)\mathrm{d}t.$$

Applying (3.4) with mean value theorem gives

$$|K_{n4}| \le M_f h_n \int_0^{\tau-\epsilon} \int_{-1}^{1} |B(t - h_n u, n) - B(t, n)||\mathrm{d}K(u)||w(t)|\mathrm{d}t$$
$$= O_p\left(\sqrt{2h_n^3 \log h_n^{-1}}\right), \tag{3.16}$$

where $M_f = \sup_{0 \le t \le \tau} |f(t)|$. Furthermore, we have

$$|K_{n5}| \le \int_0^{\tau-\epsilon} |B(t, n)| \left[ \int_{-1}^{1} |\bar{F}(t - h_n u) - \bar{F}(t)||\mathrm{d}K(u)| \right] |w(t)|\mathrm{d}t$$

$$\le M_f h_n \nu \int_0^{\tau-\epsilon} |B(t,n)||w(t)|\mathrm{d}t$$
$$= O_p(h_n). \tag{3.17}$$

The term $K_{n6}$ can be written as

$$K_{n6} = \int_0^{\tau-\epsilon} \bar{F}(t)\left[\int_{-1}^{1}(B(t-h_nu)-B(t,n))\mathrm{d}K(u)\right]w(t)\mathrm{d}t$$
$$= O_p\left(\sqrt{2h_n\log h_n^{-1}}\right).$$

This, in conjunction with (3.15)–(3.17) completes the proof. □

**Proof of Theorem 2.** Using Theorem 1 and for large $n$, we have

$$f_n(t)-\tilde{f}_n(t) = -\frac{1}{\sqrt{n}h_n}\int_0^{\infty}\sqrt{n}[\widehat{F}_n(s)-F(s)]\mathrm{d}K\left(\frac{t-s}{h_n}\right)$$
$$= \frac{1}{\sqrt{n}h_n}\int_{-1}^{1}\bar{F}(t-h_nu)B(t-h_nu,n)\mathrm{d}K(u) + O_p\left(\frac{(\log n)^{-\lambda}}{\sqrt{n}h_n}\right), \tag{3.18}$$

uniformly in $t\in[0,\tau-\epsilon]$. Since $f$ is twice continuously differentiable on $[0,\tau]$, it is easy to see that

$$\tilde{f}_n(t)-f(t) = \frac{1}{2}f''(t)h_n^2\sigma^2 + o(h_n^2), \tag{3.19}$$

uniformly in $t\in[0,\tau]$. Combining (3.18) with (3.19) yields

$$f_n(t)-f(t) = \frac{1}{\sqrt{n}h_n}\int_{-1}^{1}\bar{F}(t-h_nu)B(t-h_nu,n)\mathrm{d}K(u) + \frac{1}{2}h_n^2\sigma^2 f''(t) + O_p\left(\frac{(\log n)^{-\lambda}}{\sqrt{n}h_n}\right) + o_p(h_n^2) \tag{3.20}$$

uniformly in $t\in[0,\tau-\epsilon]$. From (3.20) we deduce that

$$ISE(f_n) = \int_0^{\tau-\epsilon}[f_n(t)-f(t)]^2\mathrm{d}t$$
$$= \frac{1}{4}h_n^4\sigma^4\int_0^{\tau-\epsilon}[f''(t)]^2\mathrm{d}t + \frac{1}{nh_n^2}D_{n1} + \frac{h_n\sigma_2^2}{\sqrt{n}}D_{n2}$$
$$+ \left[o_p(h_n^2) + O_p\left(\frac{(\log n)^{-\lambda}}{\sqrt{n}h_n}\right)\right]\left[o_p(h_n^2) + O_p\left(\frac{(\log n)^{-\lambda}}{\sqrt{n}h_n}\right) + h_n^2\sigma^4\int_0^{\tau-\epsilon}f''(t)\mathrm{d}t + \frac{2}{\sqrt{n}h_n}D_{n3}\right], \tag{3.21}$$

where

$$D_{n1} = \int_0^{\tau-\epsilon}\left[\int_{-1}^{1}\bar{F}(t-h_nu)B(t-h_nu,n)\mathrm{d}K(u)\right]^2\mathrm{d}t,$$
$$D_{n2} = \int_0^{\tau-\epsilon}f''(t)\left[\int_{-1}^{1}\bar{F}(t-h_nu)B(t-h_nu,n)\mathrm{d}K(u)\right]\mathrm{d}t,$$
$$D_{n3} = \int_0^{\tau-\epsilon}\left[\int_{-1}^{1}\bar{F}(t-h_nu)B(t-h_nu,n)\mathrm{d}K(u)\right]\mathrm{d}t.$$

Applying Lemma 2 with $w(t)=1$ yields

$$D_{n1} = \nu^2\int_0^{\tau-\epsilon}B^2(t,n)\mathrm{d}t + O_p\left(\sqrt{2h_n\log h_n^{-1}}\right). \tag{3.22}$$

Applying Lemma 3 with $w(t)=f''(t)$ and $w(t)=1$, respectively gives

$$D_{n2} = O_p\left(\sqrt{2h_n\log h_n^{-1}}\right),$$

and

$$D_{n3} = O_p\left(\sqrt{2h_n\log h_n^{-1}}\right).$$

This in conjunction with (3.21) and (3.22) completes the proof. □

## Acknowledgments

The authors would like to sincerely thank the anonymous referee for his/her valuable comments on a previous draft, which resulted in the present version of this paper. The authors wish to acknowledge partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad.

## References

Auestad, B., Tjostheim, D., 1990. Identification of nonlinear time series: First order characterisation and order determination. Biometrica 77, 669–687.
Berks, I., Philipp, W., 1977. An almost sure invariance principle for the empirical distribution function of mixing random variables. Z. Wahrscheinlichkeitstheo. Verwandte Geb. 41, 115–137.
Bickel, P., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Ann. Statist. 1, 1071–1095.
Blum, J.R., Susarla, V., 1980. In: Krishnaiah, P.K. (Ed.), Maximal Deviation Theory of Density and Failure Rate Function Estimates Based on Censored Data. In: Multivariate Analysis, vol. 5. pp. 213–222.
Burke, M.D., Csörgő, S., Horváth, L., 1988. A correction to and improvement of strong approximations of some biometric estimates under random censorship. Probab. Theory Related Fields 79, 51–57.
Cai, Z., 1998. Asymptotic properties of Kaplan–Meier estimator for censored dependent data. Statist. Probab. Lett. 37, 381–389.
Csörgő, M., Révész, P., 1981. Strong Approximation in Probability and Statistics. Academic Press, New York.
Dhompongsa, S., 1984. A note on the almost sure approximation of the empirical process of weakly dependent random variables. Yokohama Math. J. 32, 113–121.
Diebolt, J., Guégan, D., 1993. Tail behaviour of the stationary density of general nonlinear autoregressive processes of order 1. J. Appl. Probab. 30, 315–329.
Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrics 50 (4), 987–1007.
Gill, R.D., 1980. Censoring and Stochastic Integrals. In: Mathematical Centre Tracs, vol. 124. Mathematisch centrum, Amsterdam.
Gill, R.D., 1983. Large sample behavior of the product-limit estimator on the whole line. Ann. Statist. 11, 49–58.
Hall, P., 1982. Limit theorems for stochastic measures of accuracy of density estimators. Stochastic Process Appl. 13, 11–25.
Hall, P., 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. J. Multivariate Anal. 14, 1–16.
Ibragimov, I.A., Linnik, Yu.V., 1971. Independent and Stationary Sequences of Random Variables. Walters-Noordhoff, Groningen, The Netherlands.
Jomhoori, S., Fakoor, V., Azarnoosh, H.A., 2007. Central limit theorem for integrated square error of kernel hazard estimator under censored dependent model. Pakistan J. Statist. 23 (3), 193–203.
Masry, E., Tjostheim, D., 1995. Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. Econom. Theory 11, 258–289.
Masry, E., Tjostheim, D., 1997. Additive nonlinear ARX time series and projection estimates. Econom. Theory 13, 214–256.
Ozaki, T., 1979. Nonlinear time series models for nonlinear random vibrations. Technical Report, University of Manchester.
Sun, L., Zheng, Z., 1999. The asymptotics of the integrated square error for the kernel hazard rate estimators with left truncated and right censored data. Syst. Sci. Math. Sci. 12, 251–262.
Yang, S., 1993. A central limit theorem for the integrated square error of the kernel density estimators with randomly censored data. J. Statist. Plann. Inference 37, 127–143.
Zhang, B., 1998. A note on the integrated square errors of kernel density estimators under random censorship. Stochastic Process. Appl. 75, 225–234.