

# Communications in Statistics - Theory and Methods <br> Publication details, including instructions for authors and subscription information: http:// www.tandfonline.com/loi/ Ista20 

# Some Asymptotic Results of Kernel Density Estimators Under Random Left-Truncation and Dependent Data 

Vahid Fakoor ${ }^{\text {a }}$ \& Sarah Jomhoori ${ }^{\text {b }}$
${ }^{\text {a }}$ Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran
${ }^{\mathrm{b}}$ Department of Statistics, Faculty of Sciences, Birjand University, Iran
Available online: 03 Nov 2011

To cite this article: Vahid Fakoor \& Sarah J omhoori (2011): Some Asymptotic Results of Kernel Density Estimators Under Random Left-Truncation and Dependent Data, Communications in Statistics - Theory and Methods, 40:23, 4249-4262

To link to this article: http:// dx. doi.org/ 10.1080/ 03610926.2010.508297

## PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions
This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# Some Asymptotic Results of Kernel Density Estimators Under Random Left-Truncation and Dependent Data 

VAHID FAKOOR ${ }^{1}$ AND SARAH JOMHOORI ${ }^{2}$<br>${ }^{1}$ Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran<br>${ }^{2}$ Department of Statistics, Faculty of Sciences, Birjand University, Iran


#### Abstract

Problems with truncated data arise frequently in survival analyses and reliability applications. The estimation of the density function of the lifetimes is often of interest. In this article, the estimation of density function by the kernel method is considered, when truncated data are showing some kind of dependence. We apply the strong Gaussian approximation technique to study the strong uniform consistency for kernel estimators of the density function under a truncated dependent model. We also apply the strong approximation results to study the integrated square error properties of the kernel density estimators under the truncated dependent scheme.


Keywords Integrated square error; Kiefer process; Strong Gaussian approximation; Strong mixing; Strong uniform consistency; Truncated data.

Mathematics Subject Classification 62G07; 62G20.

## 1. Introduction and Preliminaries

In medical follow-up or in engineering life testing studies, one may not be able to observe the variable of interest, referred to hereafter as the lifetime. Among the different forms in which incomplete data appear, right censoring and left truncation are two common ones. Left truncation may occur if the time origin of the lifetime precedes the time origin of the study. Only subjects that fail after the start of the study are being followed, otherwise they are left truncated. Woodroofe (1985) reviewed examples from astronomy and economy where such data may occur.

Let $\mathbf{X}_{1}, \mathbf{X}_{2}, \ldots, \mathbf{X}_{\mathrm{N}}$ be a sequence of the lifetime variables which may not be mutually independent, but have a common unknown distribution function (d.f.) $F$ with a density function $f$. Let $\mathbf{T}_{1}, \mathbf{T}_{\mathbf{2}}, \ldots, \mathbf{T}_{\mathrm{N}}$ be a sequence of independent and identically distributed random variables with continuous d.f. $G$, they are also

Received February 15, 2010; Accepted July 9, 2010
Address correspondence to Vahid Fakoor, Department of Statistics, School of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran; E-mail: fakoor@ math.um.ac.ir
assumed to be independent of the random variables $\mathbf{X}_{\mathbf{i}}$ 's. In the left-truncation model, $\left(\mathbf{X}_{\mathbf{i}}, \mathbf{T}_{\mathbf{i}}\right)$ is observed only when $\mathbf{X}_{\mathbf{i}} \geq \mathbf{T}_{\mathbf{i}}$. Let $\left(X_{1}, T_{1}\right), \ldots,\left(X_{n}, T_{n}\right)$ be the actually observed sample (i.e., $\left.X_{i} \geq T_{i}\right)$, and put $\gamma:=\mathbf{P}\left(\mathbf{T}_{1} \leq \mathbf{X}_{1}\right)>0$, where $\mathbf{P}$ is the absolute probability (related to the $N$-sample). Note that $n$ itself is a random variable and that $\gamma$ can be estimated by $n / N$ (although this estimator cannot be calculated since $N$ is unknown). Assume, without loss of generality, that $\mathbf{X}_{\mathbf{i}}$ and $\mathbf{T}_{\mathbf{i}}$ are nonnegative random variables, $i=1, \ldots, N$. For any d.f. $L$ denote the left and right endpoints of its support by $a_{L}=\inf \{x: L(x)>0\}$ and $b_{L}=\sup \{x: L(x)<1\}$, respectively. Then under the current model, as discussed by Woodroofe (1985), we assume that $a_{G} \leq a_{F}$ and $b_{G} \leq b_{F}$. Define

$$
\begin{equation*}
C(x)=\mathbf{P}\left(\mathbf{T}_{\mathbf{1}} \leq x \leq \mathbf{X}_{\mathbf{1}} \mid \mathbf{T}_{\mathbf{1}} \leq \mathbf{X}_{1}\right)=\mathbb{P}\left(T_{1} \leq x \leq X_{1}\right)=\gamma^{-1} G(x)(1-F(x)), \tag{1.1}
\end{equation*}
$$

where $\mathbb{P}(\cdot)=\mathbf{P}(\cdot \mid n)$ is the conditional probability (related to the $n$-sample) and consider its empirical estimate by

$$
\begin{equation*}
C_{n}(x)=n^{-1} \sum_{i=1}^{n} I\left(T_{i} \leq x \leq X_{i}\right) \tag{1.2}
\end{equation*}
$$

where $I(\cdot)$ is the indicator function. Then the product-limit (PL) estimator $\widehat{F}_{n}$ of $F$ is given by

$$
\begin{equation*}
\widehat{F}_{n}(x)=1-\prod_{X_{i} \leq x}\left(1-\frac{1}{n C_{n}\left(X_{i}\right)}\right) . \tag{1.3}
\end{equation*}
$$

The cumulative hazard function $\Lambda(x)$ is defined by

$$
\begin{equation*}
\Lambda(x)=\int_{0}^{x} \frac{d F(u)}{1-F(u)} \tag{1.4}
\end{equation*}
$$

Let

$$
\begin{equation*}
F^{*}(x)=\mathbf{P}\left(\mathbf{X}_{\mathbf{1}} \leq x \mid \mathbf{T}_{\mathbf{1}} \leq \mathbf{X}_{\mathbf{1}}\right)=\mathbb{P}\left(X_{1} \leq x\right)=\gamma^{-1} \int_{0}^{x} G(u) d F(u) \tag{1.5}
\end{equation*}
$$

be the d.f. of the observed lifetimes. Its empirical estimator is given by

$$
F_{n}^{*}(x)=n^{-1} \sum_{i=1}^{n} I\left(X_{i} \leq x\right) .
$$

On the other hand, the d.f. of the observed $T_{i}$ 's is given by

$$
G^{*}(x)=\mathbf{P}\left(\mathbf{T}_{1} \leq x \mid \mathbf{T}_{1} \leq \mathbf{X}_{1}\right)=\mathbb{P}\left(T_{1} \leq x\right)=\gamma^{-1} \int_{0}^{\infty} G(x \wedge u) d F(u)
$$

and is estimated by

$$
G_{n}^{*}(x)=n^{-1} \sum_{i=1}^{n} I\left(T_{i} \leq x\right)
$$

It then follows from (1.1) and (1.2) that

$$
\begin{equation*}
C(x)=G^{*}(x)-F^{*}(x), \quad C_{n}(x)=G_{n}^{*}(x)-F_{n}^{*}(x-) \tag{1.6}
\end{equation*}
$$

In the independence framework with no truncation, the kernel estimate $f_{n}$ of a real univariate density $f$ introduced by Rosenblatt (1956) and defined by

$$
f_{n}(t)=\sum_{i=1}^{n} \frac{1}{n h_{n}} K\left(\frac{t-X_{i}}{h_{n}}\right),
$$

where $X_{1}, \ldots, X_{n}$ are independent observations from the density $f, K$ is a kernel function, and $h_{n}$ is a sequence of (positive) "bandwidths" tending to zero as $n \rightarrow$ $\infty$. Parzen (1962) showed that under some mild smoothness conditions on $K$ (and $f), f_{n}(t)$ is in any respect a consistent estimator of $f(t)$ for each $t \in \mathbb{R}$. The weak and strong uniform consistency properties of $f_{n}$ have been considered by several authors, including Nadaraya (1965), Schuster (1969), and Van Ryzin (1969). In these articles, the condition placed on the bandwidth for strong uniform consistency includes $\sum \exp \left(-c n h_{n}{ }^{2}\right)<\infty$ for all positive $c$. Silverman (1978) established the strong uniform consistency for $f_{n}-f$ using the strong approximation technique developed by Komlós et al. (1975) for the ordinary empirical process.

Under random left truncation model, for the case in which the lifetime observations are mutually independent, the estimation for density has been studied extensively by many authors during recent years, for example, Uzunoğullari and Wang (1992), Gijbels and Wang (1993), Sun (1997), Sun and Zhou (1998), and Arcones and Giné (1995).

Using the strong representation of the PL estimator in the form of an average of random variables plus a reminder term, Sun and Zhou (2001), established uniform consistency (with rate) and asymptotic normality of the kernel estimators of density function when the truncated data are subjected to strong mixing condition (see definition below).

In this article, we apply the strong Gaussian approximation technique to establish some asymptotic results of the kernel density estimators, including the strong uniform consistency and asymptotic expansion for the integrated square error (ISE) of the kernel density estimators when the truncated data are subjected to a kind of dependence whose definition is given below.

Definition 1.1. Let $\left\{X_{i}, i \geq 1\right\}$ denote a sequence of random variables. Given a positive integer $m$, set

$$
\begin{equation*}
\alpha(m)=\sup _{k \geq 1}\left\{|P(A \cap B)-P(A) P(B)| ; A \in \mathscr{F}_{1}^{k}, B \in \mathscr{F}_{k+m}^{\infty}\right\}, \tag{1.7}
\end{equation*}
$$

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

Among various mixing conditions used in the literature, $\alpha$-mixing, is reasonably weak and has many practical applications. There exists many processes and time
series fulfilling the strong mixing condition. As a simple example, we can consider the Gaussian $\operatorname{AR}(1)$ 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 $\left\{Z_{t}\right\}$ 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., $\alpha(n)=e^{-v n}$ for some $v>0$. The threshold models, the EXPAR models (see Ozaki, 1979), the simple ARCH models (see Engle, 1984; Masry and Tjostheim, 1995, 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 article are as follows.

## Assumptions.

(1) Suppose that $\left\{X_{i}, i \geq 1\right\}$ is a sequence of stationary $\alpha$-mixing random variables with continuous distribution function $F$, survival function $S(\cdot)$ and mixing coefficient $\alpha(n)=O\left(e^{-(\log n)^{1+\nu}}\right)$, for some $v>0$.
(2) Suppose that the truncated time variables $\left\{Y_{i}, i \geq 1\right\}$ are i.i.d. random variables with continuous distribution function $G$ and are independent of $X_{i}$ 's, also let $a_{G}<a_{F}$.
(3) $f$ is continuous with bounded second derivative on $[0, b]$, where $0 \leq b<b_{F}$.
(4) Suppose that the symmetric kernel function $K$ satisfies $\int_{-1}^{1} K(t) d t=1$, $\int_{-1}^{1} t K(t) d t=0, \int_{-1}^{1} t^{2} K(t) d t=\sigma^{2} \neq 0$, and $K(t)=0$ if $t \notin(-1,1)$ and is of bounded variation on $(-1,1)$ with total variation denoted by $V_{K}$.
The layout of the article is as follows. In Sec. 2.11, we provide some asymptotic behaviors of kernel density estimation. The counterpart of these results for the censored dependent model was established by Fakoor et al. (2009) and Fakoor (2010). Some proofs of the results are deferred to the Appendix.

## 2. Asymptotic Behaviors

We first introduce the following Gaussian process, which plays an important role to present our results.

$$
\begin{align*}
& \text { Let } g_{j}(s)=I\left(X_{j} \leq s\right)-F^{*}(s), j \geq 0 \\
& \Gamma\left(s, s^{\prime}\right)=\operatorname{Cov}\left(g_{1}(s), g_{1}\left(s^{\prime}\right)\right)+\sum_{j=2}^{\infty}\left[\operatorname{Cov}\left(g_{1}(s), g_{j}\left(s^{\prime}\right)\right)+\operatorname{Cov}\left(g_{1}\left(s^{\prime}\right), g_{j}(s)\right)\right] . \tag{2.1}
\end{align*}
$$

Define, for $0 \leq t \leq b<b_{F}$, two-parameter mean zero Gaussian process

$$
\begin{equation*}
B(t, n):=\frac{k(t, n) / \sqrt{n}}{C(t)}+\int_{0}^{t} \frac{k(u, n) / \sqrt{n}}{C^{2}(u)} d C(u) \tag{2.2}
\end{equation*}
$$

where $\{k(s, t), 0 \leq s, t \leq b\}$ is a Kiefer process in Theorem 3 of Dhompongsa (1984) with covariance function

$$
\Gamma^{*}\left(t, t^{\prime}, s, s^{\prime}\right)=\min \left(t, t^{\prime}\right) \Gamma\left(s, s^{\prime}\right)
$$

and $\Gamma\left(s, s^{\prime}\right)$ given by (2.1).
We now restate below a strong approximation by Bolbolian Ghalibaf et al. (2010) for the PL process $\alpha_{n}(t):=\sqrt{n}\left[\widehat{F}_{n}(t)-F(t)\right]$ by a two-parameter Guassian process at the rate $O\left((\log n)^{-\lambda}\right)$, for some $\lambda>0$. The statements are conditional on the observed sample size $n$.

Theorem 2.1. (Bolbolian et al., 2010) Let $b<b_{F}$. Suppose that Assumptions (1) and (2) are satisfied. On a rich probability space, there exists a two-parameter mean zero Gaussian process $B(u, v)$ for $u, v \geq 0$, such that,

$$
\begin{equation*}
\sup _{0 \leq t \leq b}\left|\alpha_{n}(t)-S(t) B(t, n)\right|=O\left((\log n)^{-\lambda}\right) \quad \text { a.s., } \tag{2.3}
\end{equation*}
$$

for some $\lambda>0$.

### 2.1. Strong Uniform Consistency

It is the purpose of this section to study the strong uniform consistency for $f_{n}-f$, using the strong Gaussian approximation technique obtained in Theorem 2.1 for the PL process. Our approach is first to apply the strong approximation technique to establish the strong uniform consistency of $f_{n}-\tilde{f}_{n}$, where

$$
\begin{equation*}
\tilde{f}_{n}(t)=\frac{1}{h_{n}} \int_{o}^{\infty} K\left(\frac{t-s}{h_{n}}\right) d F(s) . \tag{2.4}
\end{equation*}
$$

Theorem 2.2. Let $h_{n}$ be a sequence of positive bandwidths tending to zero as $n \rightarrow \infty$. Suppose that Assumptions (1)-(4) hold and that

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}=0 \tag{2.5}
\end{equation*}
$$

for some $\lambda>0$. Then, for any $b<b_{F}$,

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \sup _{0 \leq t \leq b}\left|f_{n}(t)-f(t)\right|=0 \quad \text { a.s. } \tag{2.6}
\end{equation*}
$$

Proof. See the Appendix.
An inspection of the proof of Theorem 2.2 gives the rate of strong uniform consistency for $f_{n}-\widetilde{f}_{n}$.

Lemma 2.1. Under the same conditions as in Theorem 2.2, we have

$$
\sup _{0 \leq t \leq b}\left|f_{n}(t)-\tilde{f}_{n}(t)\right|=O\left(\sqrt{\frac{\log \log n}{n}}\right)+O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}\right) \text { a.s. }
$$

Remark 2.1. If the bandwidth $h_{n}$ is chosen to be $h_{n} \sim \alpha n^{-\beta}$ with $\alpha>0$ and $0<\beta \leq \frac{1}{2}$, then condition (2.5) is satisfied.

Remark 2.2. In the independence framework with no truncation, for suitable kernels, Silverman (1978) showed that the condition $h_{n}^{-1}=o(n / \log n)$ as $n \rightarrow \infty$ is sufficient for strong uniform consistency of kernel density estimates. In the $\alpha$-mixing case with truncation, we cannot achieve the same rate as in the iid case.

Using strong Gaussian approximation in Theorem 2.1 for the PL process, we can find a two parameter mean zero Gaussian process which strongly uniformly approximate the empirical density process. Let

$$
\begin{equation*}
\psi_{n}(t, s)=\frac{1}{h_{n}} K\left(\frac{t-s}{h_{n}}\right) \tag{2.7}
\end{equation*}
$$

Theorem 2.3. Suppose that Assumptions (1)-(4) hold. Then, for any $b<b_{F}$

$$
\sup _{0 \leq t \leq b}\left|\sqrt{n}\left(f_{n}(t)-f(t)\right)-\beta(t, n)\right|=O\left(\frac{(\log n)^{-\lambda}}{h_{n}}+\sqrt{n} h_{n}^{2}\right) \quad \text { a.s., }
$$

where

$$
\beta(t, n)=-\int_{0}^{\infty} S(x) B(x, n) d \psi_{n}(t, s)
$$

Proof. Applying Theorem 2.1, we have

$$
\begin{aligned}
f_{n}(t)-f(t) & =\left(f_{n}(t)-\widetilde{f}_{n}(t)\right)+\left(\widetilde{f}_{n}(t)-f(t)\right) \\
& =\int_{0}^{\infty} \psi_{n}(t, x) d\left[\widehat{F}_{n}(x)-F(x)\right]+\left(\widetilde{f}_{n}(t)-f(t)\right) \\
& =-\frac{1}{\sqrt{n}} \int_{0}^{\infty} Z_{n}(x) d \psi_{n}(t, x)+\left(\widetilde{f}_{n}(t)-f(t)\right) \\
& \stackrel{\text { a.s. }}{=}-\frac{1}{\sqrt{n}} \int_{0}^{\infty} S(x) B(x, n) d \psi_{n}(t, x)+O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}\right)+\left(\widetilde{f}_{n}(t)-f(t)\right) .
\end{aligned}
$$

By a two-term Taylor expansion for $\tilde{f}_{n}-f$, we obtain the result.
Remark 2.3. Theorem 2.3 suggests the optimal rate $h_{n} \sim\left(n^{-1 / 2}(\log n)^{-\lambda}\right)^{1 / 3}$ for such approximation.

### 2.2. Integrated Square Error

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

$$
\operatorname{ISE}\left(f_{n}\right)=\int\left(f_{n}(t)-f(t)\right)^{2} d t
$$

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

$$
\operatorname{MISE}\left(f_{n}\right)=\int E\left(f_{n}(t)-f(t)\right)^{2} d t
$$

in an asymptotic sense. In this section, we consider the ISE of the kernel density estimator on the interval $[0, b]$ and find an asymptotic expansion for this error in terms of sample size $n$ and the bandwidth $h_{n}$. For any $b<b_{F}$, the integrated square error of $f_{n}$ on the interval $[0, b]$ is defined to be

$$
\operatorname{ISE}\left(f_{n}\right)=\int_{0}^{b}\left(f_{n}(t)-f(t)\right)^{2} d t
$$

Theorem 2.4. Let $h_{n}$ be a sequence of positive bandwidths satisfying $h_{n}=O\left(n^{-1 / 6}\right)$ as $n \rightarrow \infty$. Suppose that Assumptions (1)-(4) hold, then for $b<b_{F}$, we have

$$
\begin{align*}
\operatorname{ISE}\left(f_{n}\right)= & \frac{h_{n}^{4} \sigma_{2}^{2}}{4} \int_{0}^{b}\left[f^{\prime \prime}(t)\right]^{2} d t+\frac{V_{K}^{2}}{n h_{n}^{2}} \int_{0}^{b} \bar{F}^{2}(t) B^{2}(t, n) d t \\
& +o_{p}\left(h_{n}^{4}\right)+o_{p}\left(\frac{1}{n h_{n}^{2}}\right) \tag{2.8}
\end{align*}
$$

where $B(u, v)$ is the two-parameter Gaussian process defined in (2.2).
The proof of Theorem 2.4 is based on the following lemmas. We begin with introducing some further notations. We define

$$
\begin{aligned}
& Q_{n 1}=\int_{0}^{b}\left[\int_{-1}^{1} S\left(t-h_{n} u\right) B\left(t-h_{n} u, n\right) d K(u)\right]^{2} w(t) d t, \\
& Q_{n 2}=\int_{0}^{b}\left[\int_{-1}^{1} S\left(t-h_{n} u\right) B\left(t-h_{n} u\right) d K(u)\right] w(t) d t,
\end{aligned}
$$

where $w(t)$ is some (measurable) function defined on $(0, \infty)$. The next Lemma establishes an asymptotic expansion for $Q_{n 1}$.

Lemma 2.2. Let $f(t)$ and $w(t)$ are continuous on $[0, b]$. Under Assumptions (1)-(4), we have

$$
Q_{n 1}=V_{K}^{2} \int_{0}^{b} S^{2}(t) B^{2}(t, n)|w(t)| d t+O_{p}\left(\sqrt{h_{n} \log h_{n}^{-1}}\right)
$$

Proof. See the Appendix.
The following Lemma pertains to the asymptotic behavior for $Q_{n 2}$.
Lemma 2.3. Under the conditions of Lemma 2.2, we have

$$
\begin{equation*}
Q_{n 2}=O_{p}\left(\sqrt{h_{n} \log h_{n}^{-1}}\right) \tag{2.9}
\end{equation*}
$$

Proof. See the Appendix.

Proof of Theorem 2.4. Using (2.3) and for large $n$, we have

$$
\begin{equation*}
f_{n}(t)-\tilde{f}_{n}(t)=\frac{1}{\sqrt{n} h_{n}} \int_{-1}^{1} S\left(t-h_{n} u\right) B\left(t-h_{n} u, n\right) d K(u)+O_{p}\left(\frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}\right) \tag{2.10}
\end{equation*}
$$

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

$$
\begin{equation*}
\tilde{f}_{n}(t)-f(t)=\frac{1}{2} f^{\prime \prime}(t) h_{n}^{2} \sigma^{2}+o\left(h_{n}^{2}\right) \tag{2.11}
\end{equation*}
$$

uniformly in $t \in[0, b]$. Combining (2.10) with (2.11) yields

$$
\begin{align*}
f_{n}(t)-f(t)= & \frac{1}{\sqrt{n} h_{n}} \int_{-1}^{1} S\left(t-h_{n} u\right) B\left(t-h_{n} u, n\right) d K(u) \\
& +\frac{1}{2} h_{n}^{2} \sigma^{2} f^{\prime \prime}(t)+O_{p}\left(\frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}\right)+o_{p}\left(h_{n}^{2}\right) \tag{2.12}
\end{align*}
$$

uniformly in $t \in[0, b]$. From (2.12) we deduce that

$$
\begin{align*}
\operatorname{ISE}\left(f_{n}\right)= & \int_{0}^{b}\left[f_{n}(t)-f(t)\right]^{2} d t \\
= & \frac{1}{4} h_{n}^{4} \sigma^{4} \int_{0}^{b}\left[f^{\prime \prime}(t)\right]^{2} d t+\frac{1}{n h_{n}^{2}} D_{n 1}+\frac{h_{n} \sigma_{2}^{2}}{\sqrt{n}} D_{n 2} \\
& +\left[o_{p}\left(h_{n}^{2}\right)+O_{p}\left(\frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}\right)\right]\left[o_{p}\left(h_{n}^{2}\right)+O_{p}\left(\frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}\right)\right. \\
& \left.+h_{n}^{2} \sigma^{4} \int_{0}^{b} f^{\prime \prime}(t) d t+\frac{2}{\sqrt{n} h_{n}} D_{n 3}\right], \tag{2.13}
\end{align*}
$$

where

$$
\begin{aligned}
& D_{n 1}=\int_{0}^{b}\left[\int_{-1}^{1} S\left(t-h_{n} u\right) B\left(t-h_{n} u, n\right) d K(u)\right]^{2} d t \\
& D_{n 2}=\int_{0}^{b} f^{\prime \prime}(t)\left[\int_{-1}^{1} S\left(t-h_{n} u\right) B\left(t-h_{n} u, n\right) d K(u)\right] d t \\
& D_{n 3}=\int_{0}^{b}\left[\int_{-1}^{1} S\left(t-h_{n} u\right) B\left(t-h_{n} u, n\right) d K(u)\right] d t
\end{aligned}
$$

Applying Lemma 2.2 with $w(t)=1$ yields

$$
\begin{equation*}
D_{n 1}=V_{K}^{2} \int_{0}^{b} B^{2}(t, n) d t+O_{p}\left(\sqrt{2 h_{n} \log h_{n}^{-1}}\right) \tag{2.14}
\end{equation*}
$$

Applying Lemma 2.3 with $w(t)=f^{\prime \prime}(t)$ and $w(t)=1$, respectively gives

$$
D_{n 2}=O_{p}\left(\sqrt{h_{n} \log h_{n}^{-1}}\right)
$$

and

$$
D_{n 3}=O_{p}\left(\sqrt{h_{n} \log h_{n}^{-1}}\right)
$$

This in conjunction with (2.13) and (2.14) completes the proof.

## Appendix

To study strong uniform consistency of kernel density estimators, we also need to study the modulus of continuity of approximating process $B(u, v)$. In the next Lemma, we prove the global modulus of continuity of the Gaussian process $B(u, v)$.

Lemma A.1. Let $h_{n}$ be a sequence of positive numbers for which

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \frac{\log h_{n}^{-1}}{\log \log n}=\infty \tag{3.1}
\end{equation*}
$$

Then, for any $b<b_{F}$

$$
\begin{equation*}
\sup _{0 \leq \leq \leq b-1 \leq u \leq 1} \sup _{1 \leq}\left|B\left(t-h_{n} u, n\right)-B(t, n)\right|=O\left(\sqrt{h_{n} \log h_{n}^{-1}}\right) \text { a.s. } \tag{3.2}
\end{equation*}
$$

Proof. First, we have

$$
\begin{aligned}
\left|B\left(t-h_{n} u, n\right)-B(t, n)\right| \leq & \left|\frac{k\left(t-h_{n} u, n\right) / \sqrt{n}}{C\left(t-h_{n} u\right)}-\frac{k(t, n) / \sqrt{n}}{C(t)}\right| \\
& +\left(\inf _{0 \leq x \leq b} C^{2}(x)\right)^{-1} \sup _{0 \leq x \leq b}\left|\frac{k(x, n)}{\sqrt{n}}\right|\left|C^{-1}\left(t-h_{n} u\right)-C^{-1}(t)\right| \\
= & I_{1}+I_{2} .
\end{aligned}
$$

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

$$
\begin{aligned}
\sup _{0 \leq t \leq b} \sup _{-1 \leq u \leq 1} I_{1} \leq & \left(\inf _{0 \leq x \leq b} C(x)\right)^{-1} \sup _{0 \leq t \leq b-1 \leq u \leq 1} \sup _{0 \leq 1}\left|k\left(t-h_{n} u, n\right) / \sqrt{n}-k(t, n) / \sqrt{n}\right| \\
& +\left(\inf _{0 \leq x \leq b} C^{2}(x)\right)^{-1} \sup _{0 \leq x \leq b}\left|\frac{k(x, n)}{\sqrt{n}}\right| \sup _{0 \leq t \leq b-1 \leq u \leq 1} \sup _{01}\left|C\left(t-h_{n} u\right)-C(t)\right| \\
= & I_{11}+I_{12} .
\end{aligned}
$$

By the global modulus of continuity for the Kiefer processes (cf., e.g., Theorem 1.15.2 in Csörgő and Révész, 1981), we have

$$
\begin{equation*}
I_{11}=O\left(\sqrt{h_{n} \log h_{n}^{-1}}\right) \quad \text { a.s. } \tag{3.3}
\end{equation*}
$$

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 Berkes and Philipp, 1977),
we have

$$
\begin{equation*}
I_{12}=O\left(h_{n} \sqrt{\log \log n}\right) \quad \text { a.s. } \tag{3.4}
\end{equation*}
$$

It follows from (3.3) and (3.4) that

$$
\begin{equation*}
\sup _{0 \leq t \leq b-1 \leq u \leq 1} \sup _{1} I_{1}=O\left(\sqrt{h_{n} \log h_{n}^{-1}}\right) \quad \text { a.s. } \tag{3.5}
\end{equation*}
$$

Likewise, we observe that

$$
\begin{equation*}
\sup _{0 \leq t \leq b-1 \leq u \leq 1} \sup _{2} I_{2}=O\left(h_{n} \sqrt{\log \log n}\right) \text { a.s. } \tag{3.6}
\end{equation*}
$$

Combining (3.5)-(3.6), completes the proof.
Lemma A.2. Assuming the same conditions as in Theorem 2.2, we have

$$
\lim _{n \rightarrow \infty} \sup _{0 \leq t \leq b}\left|f_{n}(t)-\tilde{f}_{n}(t)\right|=0 \quad \text { a.s. }
$$

Proof. According to (2.3), there exists a two parameter Gaussian process $B(t, n)$ such that, for large $n$ and $t \in[0, b]$, we have

$$
\begin{align*}
f_{n}(t)-\tilde{f}_{n}(t)= & -\frac{1}{\sqrt{n} h_{n}} \int_{0}^{\infty} Z_{n}(x) d K\left(\frac{t-x}{h_{n}}\right) \\
\stackrel{\text { a.s. }}{=} & \frac{1}{\sqrt{n} h_{n}} \int_{-1}^{1} S\left(t-u h_{n}\right) B\left(t-u h_{n}, n\right) d K(u)+O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}\right) \\
= & \frac{1}{\sqrt{n} h_{n}} S(t) \int_{-1}^{1}\left[B\left(t-u h_{n}, n\right)-B(t, n)\right] d K(u) \\
& +\frac{1}{\sqrt{n} h_{n}} \int_{-1}^{1}\left[S\left(t-u h_{n}\right)-S(t)\right]\left[B\left(t-u h_{n}, n\right)-B(t, n)\right] d K(u) \\
& +\frac{1}{\sqrt{n} h_{n}} B(t, n) \int_{-1}^{1}\left[S\left(t-u h_{n}\right)-S(t)\right] d K(u)+O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}\right) \\
= & I_{1 n}(t)+I_{2 n}(t)+I_{3 n}(t)+O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n} h_{n}}\right) . \tag{3.7}
\end{align*}
$$

To deal with $I_{1 n}$, we apply Lemma A.1, so we have

$$
\begin{equation*}
\sup _{0 \leq \leq \leq b}\left|I_{1 n}(t)\right|=O\left(\sqrt{\frac{\log h_{n}^{-1}}{n h_{n}}}\right) \text { a.s. } \tag{3.8}
\end{equation*}
$$

Let $M_{f}=\sup _{0 \leq t \leq b} f(t)$, then it follows from the Mean Value Theorem that

$$
\begin{equation*}
\left|S\left(t-h_{n} u\right)-S(t)\right| \leq M_{f} h_{n} \tag{3.9}
\end{equation*}
$$

for $u \in[-1,1]$ and $t \in[0, b]$. Now applying Lemma A. 1 yields

$$
\begin{equation*}
\sup _{0 \leq t \leq b}\left|I_{2 n}(t)\right|=O\left(\sqrt{\frac{h_{n} \log h_{n}^{-1}}{n}}\right) \text { a.s. } \tag{3.10}
\end{equation*}
$$

According to the law of iterated logarithm for the Kiefer process (see Theorem A of Berkes and Philipp, 1977), we have

$$
\begin{equation*}
\sup _{0 \leq t \leq b}|B(t, n)|=O(\sqrt{\log \log n}) \quad \text { a.s. } \tag{3.11}
\end{equation*}
$$

It follows from (3.9) and (3.11)

$$
\begin{equation*}
\sup _{0 \leq t \leq b}\left|I_{3 n}(t)\right|=O\left(\sqrt{\frac{\log \log n}{n}}\right) \text { a.s. } \tag{3.12}
\end{equation*}
$$

Combining (3.7), (3.8), (3.10), and (3.12), we conclude

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \sup _{0 \leq t \leq b}\left|f_{n}(s)-\tilde{f}_{n}(s)\right|=0 \quad \text { a.s. } \tag{3.13}
\end{equation*}
$$

Proof of Theorem 2.2. Since $f$ is continuous on $[0, b], f$ is uniformly continuous on $[0, b]$, and hence it is easy to show by the dominated convergence theorem that

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \sup _{0 \leq \leq \leq b}\left|\tilde{f}_{n}(s)-f(s)\right|=0 \tag{3.14}
\end{equation*}
$$

Therefore, the result is a straightforward consequence of (3.14), (3.13) and the equality

$$
f_{n}-f=f_{n}-\tilde{f}_{n}+\tilde{f}_{n}-f
$$

Proof of Lemma 2.2. Simple algebra shows

$$
\begin{align*}
Q_{n 1}= & \int_{0}^{b}\left\{\int_{-1}^{1} S\left(t-h_{n} u\right)\left[B\left(t-h_{n} u, n\right)-B(t, n)\right] d K(u)\right. \\
& \left.\quad+\int_{-1}^{1} S\left(t-h_{n} u\right) B(t, n) d K(u)\right\}^{2} w(t) d t \\
=K_{n 1}+ & K_{n 2}+K_{n 3} \tag{3.15}
\end{align*}
$$

where

$$
\begin{aligned}
K_{n 1}= & \int_{0}^{b}\left\{\int_{-1}^{1} S\left(t-h_{n} u\right)\left[B\left(t-h_{n} u, n\right)-B(t, n)\right] d K(u)\right\}^{2} w(t) d t \\
K_{n 2}= & \int_{0}^{b} B^{2}(t, n)\left\{\int_{-1}^{1} S\left(t-h_{n} u\right) d K(u)\right\}^{2} w(t) d t \\
K_{n 3}= & 2 \int_{0}^{b}\left\{\int_{-1}^{1} S\left(t-h_{n} u\right)\left[B\left(t-h_{n} u, n\right)-B(t, n)\right] d K(u)\right\} \\
& \times\left\{\int_{-1}^{1} S\left(t-h_{n} u, n\right) d K(u)\right\} B(t, n) w(t) d t .
\end{aligned}
$$

To deal with $K_{n 1}$, we apply Lemma A. 1

$$
\begin{align*}
\left|K_{n 1}\right| & \leq \int_{0}^{b}\left\{\left|B\left(t-h_{n} u, n\right)-B(t, n)\right|\left|S\left(t-h_{n} u\right)\right||d K(u)|\right\}^{2}|w(t)| d t \\
& =O_{p}\left(h_{n} \log h_{n}^{-1}\right) \tag{3.16}
\end{align*}
$$

A Taylor expansion of $F$ yields

$$
\begin{align*}
\left|K_{n 2}\right| & \leq \int_{0}^{b} B^{2}(t, n)\left\{\int_{-1}^{1}\left|S\left(t-h_{n} u\right)\right||d K(u)|\right\}^{2}|w(t)| d t \\
& =V_{K}^{2} \int_{0}^{b} S^{2}(t) B^{2}(t, n)|w(t)| d t+O_{p}\left(h_{n}\right) \tag{3.17}
\end{align*}
$$

Likewise, applying Lemma A. 1 gives

$$
\begin{equation*}
\left|K_{n 3}\right|=O_{p}\left(\sqrt{h_{n} \log h_{n}^{-1}}\right) . \tag{3.18}
\end{equation*}
$$

Combining (3.15) with (3.16)-(3.18) completes the proof.
Proof of Lemma 2.3. First, we can write

$$
\begin{equation*}
Q_{n 2}=K_{n 4}+K_{n 5}+K_{n 6}, \tag{3.19}
\end{equation*}
$$

where

$$
\begin{aligned}
& K_{n 4}=\int_{0}^{b}\left[\int_{-1}^{1}\left(S\left(t-h_{n} u\right)-S(t)\right)\left(B\left(t-h_{n} u, n\right)-B(t, n)\right) d K(u)\right] w(t) d t \\
& K_{n 5}=\int_{0}^{b} B(t, n)\left[\int_{-1}^{1}\left(S\left(t-h_{n} u\right)-S(t)\right) d K(u)\right] w(t) d t \\
& K_{n 6}=\int_{0}^{b}\left[\int_{-1}^{1} S(t) B\left(t-h_{n} u, n\right) d K(u)\right] w(t) d t .
\end{aligned}
$$

Applying (3.2) with mean value theorem gives

$$
\begin{align*}
\left|K_{n 4}\right| & \leq M_{f} h_{n} \int_{0}^{b} \int_{-1}^{1}\left|B\left(t-h_{n} u, n\right)-B(t, n)\|d K(u)\| w(t)\right| d t \\
& =O_{p}\left(\sqrt{h_{n}^{3} \log h_{n}^{-1}}\right) \tag{3.20}
\end{align*}
$$

where $M_{f}=\sup _{0 \leq t \leq b}|f(t)|$. According to (3.11), we have

$$
\begin{align*}
\left|K_{n 5}\right| & \leq \int_{0}^{b}|B(t, n)|\left[\int_{-1}^{1}\left|S\left(t-h_{n} u\right)-S(t)\right||d K(u)|\right]|w(t)| d t \\
& \leq M_{f} h_{n} V_{K} \int_{0}^{b}|B(t, n) \| w(t)| d t \\
& =O_{p}\left(h_{n} \sqrt{\log \log n}\right) . \tag{3.21}
\end{align*}
$$

The term $K_{n 6}$ can be written as

$$
\begin{aligned}
K_{n 6} & =\int_{0}^{b} S(t)\left[\int_{-1}^{1}\left(B\left(t-h_{n} u\right)-B(t, n)\right) d K(u)\right] w(t) d t \\
& =O_{p}\left(\sqrt{h_{n} \log h_{n}^{-1}}\right)
\end{aligned}
$$

This, in conjunction with (3.19)-(3.21), completes the proof.

## Acknowledgment

The authors would like to sincerely thank the anonymous referee for his/her valuable comments. The first author was supported by a grant from Ferdowsi Universtiy of Mashhad (No. MS89138FAK).

## References

Arcones, M. A., Giné, E., (1995). On the law of the iterated logarithm for canonical Ustatistics and processes. Stochastic Process Appl. 58:217-245.
Auestad, B., Tjostheim, D. (1990). Identification of nonlinear time series: first order characterisation and order determination. Biometrica 77:669-687.
Berkes, I., Philipp, W. (1977). An almost sure invariance principle for the empirical distribution function of mixing random variables. Z. Wahrsch. verw. Gebiete 41:115-137.
Bolbolian Ghalibaf, M., Fakoor, V., Azarnoosh, H. A. (2010). Strong Gaussian approximations of product-limit and quantile processes for truncated data under strong mixing. Statist. Probab. Lett. 80:581-586.
Csörgó, M., Révész, P. (1981). Strong Approximation in Probability and Statistics. New York: Academic Press.
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.
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.
Engle, R. F. (1984). Autoregressive conditional heteroscedasticity with estimates of the variance of united Kingdom inflation. Econometrics 50(4):987-1007.
Fakoor, V., Jomhoori, S., Azarnoosh, H. A. (2009). Asymptotic expansian for ISE of kernel density estimators under censored dependent model. Statist. Probab. Lett. 79:1809-1817.
Fakoor, V. (2010). Strong uniform consistency of kernel density estimators under censored dependent model. Statist. Probab. Lett. 80:318-323.
Gijbels, I., Wang, J. L. (1993). Strong representations of the survival function estimator for truncated and censored data with applications. J. Multivariate Anal. 47:210-229.
Ibragimov, I. A., Linnik, Yu. V. (1971). Independent and Stationary Sequences of Random Variables. The Netherlands: Walters-Noordhoff, Groningen.
Komlós, J., Major, P., Tusnády, G. (1975). An approximation of partial sums of independent random variables and the sample d.f. Z. Wahrsch. Verw. Gebiete 32:111-132.
Masry, E., Tjostheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. Econ. Theor. 11:258-289.
Masry, E., Tjostheim, D. (1997). Additive nonlinear ARX time series and projection estimates. Econ. Theor. 13:214-256.
Nadaraya, E. A. (1965). On non-parametric estimates of density functions and regression curves. Theor. Probab. Appl. 10:186-190.

Ozaki, T. (1979). Nonlinear time series models for nonlinear random vibrations. Technical Report. University of Manchester, Manchester, England.
Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. 33:1065-1076.
Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function. Ann. Math. Statist. 27:832-837.
Schuster, E. F. (1969). Estimation of a probability density and its derivatives. Ann. Math. Statist. 40:1187-1196.
Silverman, B. M. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist. 6:177-184.
Sun, L. (1997). Bandwidth choice for hazard rate estimators from left truncated and right censored data. Statist. Probab. Lett. 36:101-114.
Sun, L., Zhou, Y. (1998). Sequential confdence bands for densities under truncated and censored data. Statist. Probab. Lett. 40:31-41.
Sun, L., Zhou, Y. (2001). Survival function and density estimation for truncated dependent data. Statist. Probab. Lett. 52:47-57.
Uzunoḡullari, U., Wang, J. L. (1992). A comparison of hazard rate estimators for left truncated and right censored data. Biometrika 79:293-310.
Van Ryzin, J. (1969). On strong consistency of density estiamtes. Ann. Math. Statist. 40:1765-1772.
Woodroofe, M. (1985). Estimating a distribution function with truncated data. Ann. Statist. 13:163-177.

