Nonparametric Estimation of Density Function for Sample Process under Mixing Condition

Zhu–Yu Li[1], **Zudi Lu**[2] **and Gen–Xiang Chai**[3]
[1]SFB 303, University of Bonn, Bonn, Germany, 53113
[2]Insititute of Statistics, Catholic University of Louvain, Louvain-la-Neuve, 1348, Belgium
[3]Department of Applied Mathematics, Tongji University, Shanghai, 200092, China

Abstract: Nonparametric density estimation is a useful tool for examining the structure of data, in particular, for the stationary time series, since usually it is really difficult to find the real marginal density of the series. Some papers contributed to this aspect for α-mixing stationary sequence can be found in the literature, e.g., Robinson (1983), Tran (1989, 1990). However, just as Tran et al (1996) stressed, yet there are a great number of processes which may not be α-mixing. In this paper, we will adopt a nonparametrical method to estimate unknown density function of a sample data process which is based on relaxing ϕ -mixing assumptions (see Billingsley (1968) and Bierens (1983)). Uniformly weak and strong consistency and the convergence rates of the estimator we adopted will be disussed, and some numerical examples will be given.

AMS Classifications: 62G07, 62G20, 62M10.

Keywords: Relaxing ϕ-Mixing Sample, Stationary Sample Process, Nonparametric Statistics, Kernel Density Estimators, Asymptotic Consistency, Rate of Convergence.
Short running title: Density estimation under mixing condition

1. Introduction

Probability density estimation and nonparametric, nonliear regression are probably the two most widely studied nonparametric estimation problems. Many techniques have been developed under independent observations. In recent years, extending these techniques to dependent observations and time series data which are not Gaussian cases is much attractive. Robinson (1983) gives several reasons why a nonparametric approach to time series analysis is of interest. For example, nonparamteric methods can be used to estimate the finite dimensional densities and these can be used for detecting Gaussianity or non-Gaussianity of the process. Nonparametric approaches can also be extended for prediciton, and the estimation of the regression function for the processes satisfying weak dependence conditions, such as mixing. In this paper, we will use general kernel estimator to estimate the unknown density function of a sample stationary, v-stable process based on a ϕ-mixing sequence, and study its consistency and the rate of convergence. Some numerical examples will be given to show how the estimation works on practical data.

Before we start our paper, first let us recall the definition of the ϕ-mixing process. Let $\{z_j,\ -\infty < j < +\infty\}$ be a sequence of random variables defined on an Euclidean space. Suppose that $\{z_j\}$ is stationary. Denote by $\mathcal{F}_{-\infty}^{j}$ the σ field generated by the rondom vectors $z_j, z_{j-1}, \cdots,$ and by $\mathcal{F}_{j+k}^{+\infty}$ the σ field generated by the rondom vectors $z_{j+k}, z_{j+k+1}, \cdots$. If there is a nonnegative real function ϕ_k satisfying $\lim_{k\to\infty} \phi_k = 0$ such that for every set $A \in \mathcal{F}_{-\infty}^{j}$ and every $B \in \mathcal{F}_{j+k}^{+\infty}$, there is

$$|P(A \cap B) - P(A)P(B)| \leq \phi_k P(A),$$

for each $j\,(-\infty < j < \infty)$ and each $k \geq 1$, then $\{z_j\}$ is said to be a ϕ-mixing process (Billingsley, 1968). In particular, we note that an independent process is ϕ-mixing with $\phi_0 = 1,\ \phi_k = 0,\ k \geq 1$. Although ϕ-mixing condition is much weaker than usual iid assumption on data analysis till now we have been using in nonparametric statistical density estimation, some processes of interest in time series analysis, for example, the normal $AR(1)$ process, are not ϕ-mixing. In the following, we discuss some estimation problem based on relaxing ϕ-mixing assumption process. Let $\{z_j\}$ be a ϕ-mixing process in R^p and $\{x_j\}$ be defined as

$$x_j = g(z_j, z_{j-1}, z_{j-2}, \cdots), \quad j = 1, 2, \cdots, \tag{1.1}$$

where g is a Borel measurable function on $R^p \times R^p \times \cdots \longrightarrow R^1$. It is known that $\{x_j\}$ is still the stationary process, but it need not be a ϕ-mixing any more. In Bierens(1983), he introduced the following sample transformation. Assume $E|x_j|^2 < \infty$ and for any integer $m \geq 1$, let

$$x_j^{(m)} = E(x_j | z_j, z_{j-1}, z_{j-2}, \cdots, z_{j-m+1}), \quad j = 1, 2, \cdots. \tag{1.2}$$

It is easy to know that $\{x_j^m\}$ is ϕ^*-mixing,

$$\phi_n^* = \begin{cases} \phi_{n-m} & n \geq m, \\ 1 & n < m. \end{cases}$$

Denote

$$v(m) = E|x_j - x_j^{(m)}|^2.$$

Definition 1. The process $\{x_j\}$ is said v-stable (in L^2) with respect to the process $\{z_j\}$ if

$$\lim_{m \to \infty} v(m) = 0.$$

A simple example of v-stable process is the stationary $AR(1)$ process, $x_j = \theta x_{j-1} + z_j$, where $|\theta| < 1$ and $\{z_j\}$ is ϕ-mixing staionary with zear mean and finite variance.

Let $x_1, x_2, \cdots, x_n$ be a realization of v-stable process defined above and have an unknown common density function $f(x)$. In the following, we will introduce a nonparametrical kernel estimator of f, and discuss its statistical properties based on v-stable process. Define the kernel density estimator of $f(x)$ by

$$f_n(x) = (nh_n)^{-1} \sum_{i=1}^{n} K(\frac{x - x_i}{h_n}). \tag{1.3}$$

Here $K(\cdot)$ is a known kernel density function on R^1, and $h_n > 0$ is a smoothing parameter, or a sequence of bandwidths tending to zero as n tends to infinity. In Section 2, the weak consistency of $f_n(x)$ by the similar method used in Bierens(1983) will be discussed. The pointwise strong consistency and uniform strong consistency will be given in Section 3 by using an inequality in Chai(1989), which will be given as a useful formulation. The rates of convergence will be considered in Section 4, and some simulation results will be given in Section 5.

At the end of this section, we refer to some references in the literature concerned with the kernel density estimation for the stationary sequences. Under α-mixing context, see Robinson (1983), Tran (1989, 1990) and the references therein. However, just as a recent paper of Tran et al (1996) stressed, there exist a great number of processes which may not be α-mixing. In (1.1), when $\{z_t\}$ is an iid sequence and g is a linear function, the kernel density function of $\{X_t\}$ is discussed recently by Hallin and Tran (1996), which is clearly a special case of the context of this paper. We remark that in nonlinear time series analysis, the stationary solutions of many models are nonlinear functions of the iid sequences. The following examples highlight this point.

Example 1. Consider a random coefficient model

$$X_t = (\phi + e_t)X_{t-1} + \varepsilon_t, \tag{1.4}$$

where $\{e_t\}$ and $\{\varepsilon_t\}$ are two independent iid sequences with $Ee_t = E\varepsilon_t = 0$, $Ee_t^2 = \sigma^2 < \infty$, $E\varepsilon_t^2 = \sigma_1^2 < \infty$. Then if $\phi^2 + \sigma^2 < 1$, the second-order stationary solution of the model is

$$X_t = \varepsilon_t + (\phi + e_t)\varepsilon_{t-1} + (\phi + e_t)(\phi + e_{t-1})\varepsilon_{t-2} + \cdots,$$

by which $\{X_t\}$ is a v-stable sequence with

$$z_t = (\varepsilon_t, e_t), \qquad v(m) = (\phi^2 + \sigma^2)^m \sigma_1^2/[1 - (\phi^2 + \sigma^2)].$$

Example 2. Consider a bilinear model

$$X_t = (\phi + ae_t)X_{t-1} + e_t - be_{t-1}, \tag{1.5}$$

where $\{e_t\}$ is an iid sequence with $Ee_t = 0$, $Ee_t^2 = \sigma^2 < \infty$. Then if $\phi^2 + a^2\sigma^2 < 1$, the second-order stationary solution of the model is

$$X_t = e_t - be_{t-1} + (\phi + ae_t)(e_{t-1} - be_{t-2}) + (\phi + ae_t)(\phi + ae_{t-1})(e_{t-2} - be_{t-3}) + \cdots.$$

Clearly, $\{X_t\}$ is a v-stable sequence with

$$z_t = e_t, \qquad v(m) = (\phi^2 + a^2\sigma^2)^m(1 + b^2)\sigma^2/[1 - (\phi^2 + a^2\sigma^2)].$$

2. Weak consistency of $f_n(x)$

Lemma 1. *For nonnegative integer n, let $v(n) = O(n^{-\lambda})$, $\lambda > 0$. Assume that the kernel density function $K(x)$ has an absolutely integrable characteristic function $\beta(t)$, say*

$$\beta(t) = \int_R e^{itx}K(x)dx,$$

satisfying $\int_R |t||\beta(t)|dt < \infty$. Let $\rho_n = \frac{1}{n}\sum_{i=1}^n \phi_i^{1/2}$, then

$$E\sup_{x\in R}|f_n(x) - Ef_n(x)| = O(\sqrt{\rho_n^*}h_n^{-1}), \tag{2.1}$$

where $\rho_n^* = \max\{\rho_n,\ h_n^{-2/(\lambda+1)}(\sqrt{n})^{-2\lambda/(\lambda+1)}\}$.

Proof. The proof of this lemma is completely similar to that of Lemma 4 of Bierens (1983) and hence is cancelled.

Remark 1. We here remark that there is a slight error in (4.2) of Bierens (1983), in which $\rho_n^* = \max\{\rho_n,\ h_n^{-2/(\lambda+1)}(\sqrt{n})^{-2/(\lambda+1)}\}$. However, substituting his (4.6) into his (4.5), one should take that

$$\rho_n^* = \max\{\rho_n,\ h_n^{-2/(\lambda+1)}(\sqrt{n})^{-2\lambda/(\lambda+1)}\},$$

as in (2.1) of this paper.

Lemma 2. *Let $K(\cdot)$ be p.d.f., and $h_n \downarrow 0$ as $n \to \infty$. Then*
(i) if f is continuous at x,

$$|Ef_n(x) - f(x)| \to 0; \tag{2.3a}$$

(ii) if f uniformly continuous on R,

$$\sup_{x\in R} |Ef_n(x) - f(x)| \to 0, \tag{2.3b}$$

as $n \to \infty$.

Proof: The proof of this lemma is easy and hence cancelled.

Theorm 1. *Under conditions of Lemma 1 and Lemma 2, let*

$$\sum_{m=0}^{\infty} \phi_m^{1/2} < \infty, \qquad \frac{1}{h_n}(h_n(\sqrt{n})^{\lambda})^{-\frac{1}{\lambda+1}} \to 0,$$

then (i) under Lemma 2 (i),

$$|f_n(x) - f(x)| \xrightarrow{P} 0;$$

(ii) under Lemma 2 (ii),

$$\sup_{x\in R} |f_n(x) - f(x)| \xrightarrow{P} 0.$$

Proof. Only prove the second result. Recall the definition of ρ_n, $\rho_n = O(\frac{1}{n})$. From Lemma 1, we get

$$E\sup_{x\in R} |f_n(x) - Ef_n(x)| = O(\max\{(\sqrt{n}h_n)^{-1},\ h_n^{-1}(h_n(\sqrt{n})^{\lambda})^{-\frac{1}{\lambda+1}}\}) \to 0,$$

when $n \to \infty$. From Lemma 2, $\forall \epsilon > 0$, when n is large enough, there is

$$\begin{aligned} P(\sup_x |f_n(x) - f(x)| > \epsilon) &\leq P(\sup_x |f_n(x) - Ef_n(x)| > \epsilon/2) \\ &\leq 2E \sup_x |f_n(x) - Ef_n(x)|/\epsilon. \end{aligned}$$

Then, from Lemma 1 we get

$$P(\sup_{x \in R} |f_n(x) - f(x)| > \epsilon) \to 0, \quad n \to \infty.$$

3. Strong Consistency of $f_n(x)$

Let $\{\xi_n\}_{n=1}^{\infty}$ be a ϕ-mixing, stationary random variable sequence. $E\xi_1 = 0$, $|\xi_1| \leq p$ a.s.. Let $S_n = \sum_{i=1}^{n} \xi_i$. For each n, there are $l = l(n)$, $k = k(n)$, such that

$$2lk \leq n < 2(l+1)k. \tag{3.1}$$

Lemma 3. *Under the above notations and assumptions, if* $\sum_{n=1}^{\infty} \phi_n < \infty$, *then for* $\forall \epsilon > 0$ *and* $t > 0$ *satisfying*

$$0 < 2tkp < 1/2, \tag{3.2}$$

there is always

$$P(S_n > \epsilon) \leq e \cdot e^{-t\epsilon + c_0 tl}(1 + c_1 \phi_k)^l, \tag{3.3}$$

in which c_0, c_1 *are constants only related with* p, ϕ. If $c = 1 + 4\sum_{n}^{\infty} \phi_n > 1$, $c_0 = pc, c_1 = 2e^{1/2}$ (Chai, 1989).

Proof. This is Theorm 1 of Chai (1989).

In the following, we use c to denote a generic constant which may be different at each place.

Corollary. *Let* $\{\eta\}_{i=1}^{\infty}$ *be a stationary,* ϕ^**-mixing sequence of random variables, in which,*

$$\phi_m^* = \begin{cases} \phi_{m-l^*}, & m \geq l^*, \\ 1, & m < l^*. \end{cases}$$

and ϕ is the same as in Lemma 3. $E\eta_1 = 0$, $|\eta_1| \leq p$, *a.s.* $l^* = cn^{1/3}$. *Then* $\forall \epsilon > 0$, *when n is large enough, there is*

$$P(\frac{1}{n}|\sum_{i=1}^{n} \eta_i| > \epsilon) \leq c_2 \exp\{-c_3 \sqrt[3]{n}\epsilon\}, \tag{3.4}$$

where *constants* $c_2,\ c_3$ *are free from* $n,\ \epsilon$.

Proof. From the symmetric of (3.4), we only need prove $\exists c_2'>0,\ c_3'>0,\ \forall \epsilon>0$ such that

$$P(\frac{1}{n}\sum_{i=1}^{n}\eta_i>\epsilon)\le c_2'\exp\{-c_3'\epsilon\sqrt[3]{n}\}, \tag{3.5}$$

where constants $c_2',\ c_3'$ may be different from $c_2,\ c_3$, and free from $n,\ \epsilon$. When n is large enough, from Lemma 3, take l,k,t satisfying (3.1) and (3.2), then for any fixed $l^*\ge 1$, there is

$$\begin{aligned} P(\frac{1}{n}\sum_{i=1}^{n}\eta_i>\epsilon) &\le e\cdot e^{-tn\epsilon+(1+4\sum_{m=1}^{\infty}\phi_m^*)tlp}(1+c_1\phi_k^*)^l \\ &= e\cdot e^{-tn\epsilon+(1+4l^*+4\sum_{n=1}^{\infty}\phi_n)tlp}(1+c_1\phi_{k-l^*})^l. \end{aligned}$$

Let $k=O(n^{2/3}),\ l=O(n^{1/3}),\ t=O(n^{-2/3})$ (such that (3.1) and (3.20) hold), $l^*=O(n^{1/3})$, then

$$(1+4l^*+4\sum_{n=1}^{\infty}\phi_n)tlp=O((1+4\sum_{n=1}^{\infty}\phi_n)tlpl^*)=O(1),$$

and since $\sum\phi_n<\infty$, $\phi_k\le 1/k$ for k large enough,

$$(1+c_1\phi_{k-l^*})^l=O((1+\frac{c_1}{n^{2/3}-l^*})^{n^{1/3}})=O(1).$$

Thus we get (3.5) and then (3.4).

Theorem 2. *Let* $K(\cdot)\in R$ *in (1.2) satisfy Lipschitz condition, that is,* $\exists c>0$, *such that* $|K(x)-K(y)|\le c|x-y|$ *for any* $x,y\in R$. *And let* $\sum_{n=1}^{\infty}\phi_n<\infty$, $v(l)=E(x_1-x_1^{(l)})^2=O(l^{-\lambda})$ *with* $\lambda>0$. *If* $l=l(n)=O(n^{1/3})$, $h_n\to 0$, *when* $n\to\infty$, *such that* $\sum_{n=1}^{\infty}(n^{\lambda/6}h_n^2)^{-1}<\infty$, *and* $\forall c>0$, $\sum_{n=1}^{\infty}exp\{-c\sqrt[3]{n}h_n\}<\infty$. *Then we have*

(i) For any fixed x *at which* f *is continuous,*

$$|f_n(x)-f(x)|\to 0,\quad a.s.\quad n\to\infty. \tag{3.6}$$

(ii) If the variation of $K(\cdot)$ *is bounded, and* f *is uniformly continuous on* R, *then*

$$\sup_{x\in R}|f_n(x)-f(x)|\to 0,\quad a.s.\quad n\to\infty. \tag{3.7}$$

Proof. Denote
$$f_n^l(x)=\frac{1}{nh_n}\sum_{i=1}^n K(\frac{x-x_i^{(l)}}{h_n}),$$
then
$$\begin{aligned}|f_n(x)-Ef_n(x)| &\le |f_n(x)-f_n^l(x)| \\ &+ |f_n^l(x)-Ef_n^l(x)| \\ &+ |Ef_n^l(x)-Ef_n(x)|.\end{aligned}\tag{3.8}$$
Since
$$\sup_{x\in R}|f_n(x)-f_n^l(x)|=\frac{1}{nh_n}\sup_{x\in R}|\sum_{i=1}^n(K(\frac{x-x_i}{h_n})-K(\frac{x-x_i^{(l)}}{h_n}))|\le\frac{c}{nh_n^2}\sum_{i=1}^n|x_i-x_i^{(l)}|,$$
therefore
$$\begin{aligned}P(\sup_{x\in R}|f_n(x)-f_n^l(x)|\ge\epsilon) &\le P(\sum_{i=1}^n|x_i-x_i^{(l)}|\ge c^{-1}n\epsilon h_n^2)\\ &\le \frac{c}{n\epsilon h_n^2}\sum_{i=1}^n E|x_i-x_i^{(l)}|\\ &\le \frac{c}{n\epsilon h_n^2}\sum_{i=1}^n E^{1/2}(|x_i-x_i^{(l)}|^2)\\ &= \frac{c}{\epsilon h_n^2}v(l)^{1/2}.\end{aligned}$$
Let $l=l(n)=c'n^{1/3},\ 0<c'<1,$ then
$$\sum_{n=1}^\infty P(\sup_{x\in R}|f_n(x)-f_n^{l(n)}|\ge\epsilon)=O_\epsilon(\sum_{n=1}^\infty(n^{\lambda/6}h_n^2)^{-1}).$$
For any $\epsilon>0$,
$$\sum_{n=1}^\infty P(\sup_{x\in R}|f_n(x)-f_n^{l(n)}(x)|\ge\epsilon)<\infty,$$
hence
$$\sup_{x\in R}|f_n(x)-f_n^{l(n)}(x)|\to 0,\quad a.s.\tag{3.9}$$
Next, consider
$$|Ef_n^{l(n)}(x)-Ef_n(x)| \le \frac{c}{nh_n^2}\sum_{i=1}^n E|x_i-x_i^{(l(n))}|$$

$$\begin{aligned}
&\leq \frac{c}{nh_n^2}\sum_{i=1}^{n} E^{1/2}(|x_i - x_i^{(l(n))}|^2) \\
&\leq c\frac{v(l(n))^{1/2}}{h_n^2} = O((n^{\lambda/6}h_n^2)^{-1}),
\end{aligned}$$

which is free from x, thus

$$\sup_{x\in R}|Ef_n^{l(n)}(x) - Ef_n(x)| \to 0, \qquad n\to\infty.$$

Finally,

$$|f_n^{l(n)}(x) - Ef_n^{l(n)}| = |\frac{1}{nh_n}\sum_{i=1}^{n}(K(\frac{x-x_i^{(l(n))}}{h_n}) - EK(\frac{x-x_i^{(l(n))}}{h_n}))|,$$

and for any fixed x, a random variable sequence $\{K(\frac{x-x_i^{(l(n))}}{h_n}) - EK(\frac{x-x_i^{(l(n))}}{h_n})\}_{i=1}^{n}$ satisfies conditions in Corollary of Lemma 3. Hence $\forall\epsilon > 0$, when n is large enough

$$\begin{aligned}
P(|f_n^{l(n)}(x) - Ef_n^{l(n)}(x)| > \epsilon) &= P(\frac{1}{nh_n}|\sum_{i=1}^{n}(K(\frac{x-x_i^{(l(n))}}{h_n}) - EK(\frac{x-x_i^{(l(n))}}{h_n}))| > \epsilon) \\
&\leq c_1\exp\{-c_2\sqrt[3]{n}h_n\epsilon\}.
\end{aligned}$$

Thus, from

$$\sum_{n=1}^{\infty} P(|f_n^{l(n)}(x) - Ef_n^{l(n)}(x)| > \epsilon) < \infty,$$

and when x is fixed, we get

$$|f_n^{l(n)}(x) - Ef_n^{l(n)}(x)| \to 0, \qquad a.s.. \tag{3.10}$$

Therefore, together with (3.8) - (3.10), and from (2.3), we get

$$|f_n(x) - f(x)| \to 0, \qquad a.s. \qquad n\to\infty.$$

Next, we prove (3.7). Observe

$$\sup_{x\in R}|f_n^{l(n)}(x) - Ef_n^{l(n)}(x)| = \sup_{x\in R}|\frac{1}{nh_n}|\sum_{i=1}^{n}(K(\frac{x-x_i^{(l(n))}}{h_n}) - EK(\frac{x-x_i^{(l(n))}}{h_n}))|. \tag{3.11}$$

Let F_n be the notation of empirical function for $x_1^{(l(n))}, \cdots, x_n^{(l(n))}$, and F be a marginal distribution function of $x_1^{(l(n))}, \cdots, x_n^{(l(n))}$. Then (3.11) equals

$$\sup_{x\in R} |\frac{1}{h_n} \int K(\frac{x-u}{h_n}) d(F_n(u) - F(u))|.$$

Since $F_n(u) - F(u)$ is a bounded variation function, from integration by parts, (3.11) can be done as

$$\begin{aligned}
\sup_{x\in R} |\frac{1}{h_n} K(\frac{x-u}{h_n})[F_n(u) - F(u)]|_{-\infty}^{+\infty} \quad &- \quad \frac{1}{h_n} \int (F_n(u) - F(u)) dK(\frac{x-u}{h_n})| \\
&\leq \quad \sup_{x\in R} \frac{1}{h_n} \int |F_n(u) - f(u)| d|K(\frac{x-u}{h_n})| \\
&\leq \quad \frac{1}{h_n} \sup_{u\in R} |F_n(u) - F(u)| V(K), (3.12)
\end{aligned}$$

where $V(K)$ denotes the complete variation of $K(\cdot)$, and is bounded. From well known Gnedenko Theorem, to prove $\sup_{u\in R} |F_n(u) - F(u)| \to 0$, $a.s.$ $(n \to \infty)$, we only need to prove for each u, there is

$$|F_n(u) - F(u)| \to 0, \quad a.s. \quad n \to \infty.$$

But we know

$$\frac{1}{h_n} |F_n(u) - F(u)| = \frac{1}{nh_n} |\sum_{i=1}^{n} (I(x_i^{(l(n))} \leq u) - EI(x_i^{(l(n))} \leq u))|.$$

Here, random variable sequence $\{I(x_i^{(l(n))} \leq u) - EI(x_i^{(l(n))} \leq u)\}_{i=1}^{n}$ satisfies conditions in the Corollary of Lemma 3, therefore,

$$\begin{aligned}
P(\frac{1}{h_n} |F_n(u) - F(u)| \geq \epsilon) \quad &= \quad P(\frac{1}{nh_n} |\sum_{i=1}^{n} (I(x_i^{(l(n))} \leq u) - EI(x_i^{(l(n))} \leq u))| \geq \epsilon) \\
&\leq \quad c_1 \exp\{-c_2 \sqrt[3]{n} h_n \epsilon\},
\end{aligned}$$

thus

$$\sum_{n=1}^{\infty} P(\frac{1}{h_n} |F_n(u) - F(u)| \geq \epsilon) \leq \sum_{n=1}^{\infty} c_1 \exp\{-c_2 \sqrt[3]{n} h_n \epsilon\} < \infty.$$

Then

$$\frac{1}{h_n} |F_n(u) - F(u)| \to 0, \quad a.s. \quad n \to \infty,$$

thus

$$\sup_{x\in R}|f_n^{l(n)}(x)-Ef_n^{l(n)}(x)|\to 0,\quad a.s.\quad n\to\infty.$$

From Lemma 2, we get (3.7).

4. Convergence Rates of $f_n(x)$

Lemma 4. *If f is bounded, uniformly continuous function with bounded 2nd order derivative, and*

$$\int uK(u)du=0,\qquad \int u^2K(u)du<\infty,$$

then

$$\sup_x|Ef_n(x)-f(x)|=O(h_n^2), \tag{4.1}$$

Proof. Clearly,

$$\begin{aligned}|Ef_n(x)-f(x)| &= |\frac{1}{nh_n}\int\sum_{i=1}^n K(\frac{x-u}{h_n})f(u)du-f(x)|\\ &= |\int K(u)f(x-h_nu)du-f(x)|\\ &= |\int K(u)(f(x-h_nu)-f(x))du|.\end{aligned}$$

According to the Taylor expansion,

$$f(x-h_nu)-f(x)=(-h_nu)f'(x)+\frac{(-h_nu)^2}{2}f''(\xi),$$

$\xi\in(x-h_nu,x)$, thus

$$\begin{aligned}|Ef_n(x)-f(x)| &= |\int K(u)((-h_nu)f'(x)+\frac{h_n^2u^2}{2}f''(\xi))du|\\ &= |\frac{h_n^2}{2}\int u^2K(u)f''(\xi)du|.\end{aligned}$$

Therefore,

$$\sup_x|Ef_n(x)-f(x)|\le\frac{h_n^2}{2}\sup_{y\in R}|f''(y)|\int u^2K(u)du.$$

From given conditions, we get (4.1).

Theorem 3. *In Theorem 1, let* $\lambda > 0$, $h_n = n^{-\lambda/(6\lambda+8)}$, $K(\cdot)$, f *satisfy conditions in Lemma 4. Then*

$$\sup_x |f_n(x) - f(x)| = O_p(\epsilon_n), \tag{4.2}$$

in which

$$\epsilon_n = O(n^{-\frac{\lambda}{3\lambda+4}}). \tag{4.3}$$

Proof. From Lemma 4,

$$\sup_x |Ef_n(x) - f(x)| = O(h_n^2) = O(\epsilon_n),$$

and then from Lemma 1

$$E\sup_x |f_n(x) - Ef_n(x)| = O(\sqrt{\rho_n^*}h_n^{-1}).$$

Because

$$\begin{aligned} P(\sup_x |f_n(x) - Ef_n(x)| > c\epsilon_n) &\leq c^{-1}\epsilon_n^{-1}E\sup_x |f_n(x) - Ef_n(x)| \\ &\leq c^{-1}\ \epsilon_n^{-1}h_n^{-1}max\{n^{-1/2},\ h_n^{\frac{-1}{\lambda+1}}n^{-\frac{\lambda}{2(\lambda+1)}}\} \\ &= c^{-1}\ \epsilon_n^{-1}h_n^{-1-\frac{1}{\lambda+1}}n^{-\frac{\lambda}{2(\lambda+1)}}, \end{aligned}$$

from (4.3) we get

$$P(\sup_x |f_n(x) - Ef_n(x)| > c\epsilon_n) \to 0, \qquad n \to \infty,\ \ c \to \infty.$$

Then finish the proof.

Theorem 4. *In Theorem 2, if* $\lambda > 9$, $h_n = n^{-1/9}$, *then (i) for fixed* x *at which* $f(x)$ *is continuous, there is*

$$f_n(x) - f(x) = o(r_n) \quad a.s.; \tag{4.4a}$$

(ii) under Theorem 2 (ii),

$$\sup_{x\in R} |f_n(x) - f(x)| = o(r_n) \quad a.s. \tag{4.4b}$$

where $r_n = n^{-2/9}(\log\log n)\log n$.

Proof. From Lemma 4,

$$\sup_x r_n^{-1}(Ef_n(x) - f(x)) = O(r_n^{-1}h_n^2). \tag{4.5}$$

In the following, we need to get

$$r_n^{-1}(f_n(x) - Ef_n(x)) \to 0, \quad a.s. \quad n \to \infty.$$

That is, we need to prove $\forall \epsilon > 0$,

$$\sum_{n=1}^{\infty} P\{r_n^{-1}(f_n(x) - Ef_n(x)) > \epsilon\} < \infty.$$

From the proof of Theorem 2, we know

$$\begin{aligned} P(r_n^{-1}|f_n(x) - Ef_n(x)| > \epsilon) &= P(|f_n(x) - Ef_n(x)| > r_n\epsilon) \\ &\le \frac{c}{\epsilon}(r_n n^{\lambda/6} h_n^2)^{-1} + c_1 \exp\{-c_2 \sqrt[3]{n} h_n r_n \epsilon\}. (4.6) \end{aligned}$$

When $h_n = n^{-1/9}$, $r_n = n^{-2/9}(\log\log n)\log n$, there are

$$r_n^{-1}h_n^2 \to 0, \quad n \to \infty,$$

$$\sqrt[3]{n} h_n r_n = n^{1/3} n^{-1/9} n^{-2/9} (\log\log n)\log n \quad \to \infty.$$

and

$$\begin{aligned} \exp\{-c_2 \sqrt[3]{n} h_n r_n \epsilon\} &= \exp\{-c_2\epsilon(\log\log n)\log n\} \\ &= e^{-c_2\epsilon(\log\log n)\log n} \\ &= e^{\log(n^{-c_2\epsilon\log\log n})} \\ &= n^{-c_2\epsilon\log\log n}. \end{aligned}$$

And then for n large enough, fixed ϵ, we get $c_2\epsilon\log\log n > 1$, thus

$$\sum_{n=1}^{\infty} \exp\{-c_2 \sqrt[3]{n} h_n r_n \epsilon\} < \infty. \tag{4.7}$$

Because

$$\sum_{n=1}^{\infty}(r_n n^{\lambda/6} h_n^2)^{-1} = \sum_{n=1}^{\infty}(n^{-\frac{2}{9}}(\log\log n)(\log n) n^{\frac{\lambda}{6}} n^{-\frac{2}{9}})^{-1} = \sum_{n=1}^{\infty}(n^{\frac{\lambda}{6}-\frac{4}{9}}(\log\log n)\log n)^{-1},$$

when λ is suitably large, say $\lambda > 9$, there is

$$\sum_{n=1}^{\infty}(r_n n^{\lambda/6} h_n^2)^{-1} < \infty. \tag{4.8}$$

From (4.5) - (4.8), we get (4.4).

Since f is uniformly continuous on R, we can immediately extend result of the above to

$$\sup_{x \in R} |f_n(x) - f(x)| = o(r_n), \quad a.s..$$

5. Numerical Example

A simulation of Example 2 in Section 1 has been made. Here constants $a = b = 1/2$, and $\phi = 0.4$ in (1.5), that is

$$X_t = (0.4 + 0.5e_t)X_{t-1} + e_t - 0.5e_{t-1}, \tag{5.1}$$

in which $\{e_t\}$ come from iid uniform distribution on [-1/2, 1/2]. Fig. 1.1 shows the realization of the stationary solution of (5.1) with the size of samples = 1000. Fig. 1.2 to Fig. 1.4 indicate the marginal density estimations of X_t by using estimator (1.3) with Gaussian kernel and different window widths, from which the asymetric property of the density of X_t is clear.

A simple practical example is given by using foreign exchange data of Canadian Dollar vs. Australian Dollar and British Pound from January 2, 1991 to June 4, 1996. The number of total observations for each pair is 1359. Since the real data processes are not stationary, we use the first difference in the logarithms to get stationarity. We call the new processes as log-difference stationary processes. These are shown in Figure 2. To use (1.3), we employ the following three kinds of kernels:

(i) Triangular kernel on $[-1, 1]$:

$$K_1(u) = \begin{cases} 1 - |u|, & \text{if } |u| \le 1, \\ 0, & \text{otherwise;} \end{cases} \tag{5.2}$$

(ii) Epanechnikov kernel:

$$K_2(u) = \begin{cases} \frac{3}{4}(1-u^2), & \text{if } |u| < 1, \\ 0, & \text{otherwise;} \end{cases} \tag{5.3}$$

(iii) Gaussian kernel:

$$K_3(u) = (2\pi)^{-1} \exp\{-\frac{1}{2}u^2\}, \qquad -\infty < u < \infty. \tag{5.4}$$

From the literature, we know that the choice of kernel function, K, is of essentially negligible concern, but the choice of smoothing parameter is crucial. It is well known that the optimal smoothing parameter h_n in (1.3) is

$$h_n = C(K)\{\int f''(x)^2 dx\}^{-1/5} n^{-1/5}, \tag{5.5}$$

$C(K) = \{\int t^2 K(t)dt\}^{-2/5}\{\int K(t)^2 dt\}^{1/5}$. From (5.5), we know that h_n itself depends on the unknown density being estimated. If we consider an easy and natural approach to f, a standard family of distribution is used to assign a value to the term $\int f''(x)^2 dx$ in h_n for an ideal smooth parameter. Without lose of generality, for example, consider f coming from a normal distribution with zero mean and finite variance σ^2 (Silverman, 1986). Then

$$\int f''(x)^2 dx = (3/8)\pi\sigma^{-5} \approx 0.212\sigma^{-5}. \tag{5.6}$$

Therefore, from (5.2) - (5.6), smoothing parameters could be obtained as

$h_1 = 2.58\hat{\sigma}_1 n^{-1/5}$,

$h_2 = 2.34\hat{\sigma}_2 n^{-1/5}$,

$h_3 = 1.06\hat{\sigma}_3 n^{-1/5}$,

where $\hat{\sigma}_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$ is a sample variance of the given data sequence.

Figure 3 gives the density estimation for each log-difference stationary process in three kinds of kernels. They show approximately normality.

Acknowledgments

Authors would like to thank Jürgen Arns and Frank Heid, Department of Economics II, University of Bonn, for their cooperation on data and figures in the paper.

References

Bierens, H. J. (1983). Uniform Consistency of Kernel Estimators of a Regression Function Under Generalized Conditions. *Journal of the American Statistical Association,* **78**, 699-707.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley & Sons, New York.

Chai, G. X. (1989). Consistency of the Nearest- Neighbor Density Estimator of a Stationary Sequence. *ACTA MATHEMATICA SINICA,* **32**, 423-432.

Chung, K. L. (1974). *A Course in Probability Theory.* Academic Press, New York, London.

Hallin, M. and Tran, L. T. (1996) Kernel Density Estimation for Linear Processes: Asymptotic Normality and Optimal Bandwidth Derivation. *Annals of Institute of Statistical Mathematics,* **48,** 429-449.

Li, Z. Y. (1995). A Study of Nonparametric Estimation of Error Distribution in Linear Model based on L_1-norm. *Hiroshima Mathematical Journal*, **25**, 171-205.

Li, Z.Y., Zhang, W.Y., Chai, G.X. and Suryawanshi, S. (1995). An Estimation of Error Distribution in Nonparametric Regression. Thec. Report, No. 95-25, Center for Multivariate Analysis, Pennsylvania State University.

Prakasa Rao, B.L.S. (1996). *On Some Recent Advances in Nonparametric Functional Estimation.* Sectional Presidential Address, **83rd** Session.

Robinson, P. M. (1983). Nonparamteric Estimators for Time Series. *Journal of Time Series Analysis,* **4**, 185-197.

Tran, L. T. (1989). Recursive density estimation under dependence. *IEEE Trans. Inform. Theory,* **35**, 1103-1108.

Tran, L. T. (1990). Kernel density estimation under dependence. *Statist. and Prob. Letters,* **10**, 193-201.

Tran, L. T., Roussas, G., Yakowitz, S. and Truong Van B. (1996). Fixed-design regression for linear time series, *Ann. Statist.,* **24**, 975-991.