Stochastic Processes and their Applications 38 (1991) 323-333
North-Holland

Strong consistency and rates for recursive nonparametric conditional probability density estimates under (α, β)-mixing conditions

Zongwu Cai

Department of Mathematics, Hangzhou University, Hangzhou, Zhejiang 310028, People's Republic of China

Received 9 January 1990
Revised 9 August 1990

Let $\{X_j: j \geqslant 1\}$ be a real-valued stationary process. Recursive kernel estimators of the joint probability density functions, and of conditional probability density functions of X_j, given past behavior, are considered. Their strong consistency, along with rates, are given for process $\{X_j; j \geqslant 1\}$ satisfying (α, β)-mixing conditions. Here, we improve the rates of a.s. convergence in Masry (1987, 1989) without imposing considerably faster rate of decay on the mixing coefficients.

conditional probability density estimate * a.s. covergence rates * (α, β)-mixing stationary process

1. Introduction

Let $\{X_j; j \geqslant 1\}$ be a real-valued stationary process on a probability space $(\Omega, \mathscr{F}, P)$. Given a single realization $\{X_j; 1 \leqslant j \leqslant n\}$ of the process, inferences are to be made about the process. Of particular interest is the nonparametric estimation of the finite dimensional probability density functions of the process, and conditional probability density function of the process given past behavior. Many authors discussed the limit behavior, for instance, the weak consistency and asymptotic normality of such estimators were given by Robinson (1983, 1986), and strong consistency and rates of such recursive estimators were established by Masry (1987, 1989). But a.s. convergence rates of the estimators for ρ-mixing processes in Masry (1987, 1989) were much lower than those in Wegman and Davis (1979) for i.i.d. observations.

The purpose of this paper is the establishment of strong consistency and almost sure convergence rates for recursive estimators of joint and conditional probability density functions for stationary (α, β)-mixing processes. The work improves therefore some results of Masry (1987, 1989) but differs method of analysis. Our almost sure convergence rates for (α, β)-mixing processes are very close to those in Wegman and Davis (1979) for independent data. Here it is worth pointing out such a fact that we obtain a useful inequality for the maximum of the partial sum of the process. Our main method bases on the inequality.

For each integer $d \geq 1$ and integers $1 \leq i_1 < i_2 < \cdots < i_d$, let $f_d(x) = f(x_1, \ldots, x_d; i_1, \ldots, i_d)$ be the joint probability density function of the random variables $X_{i_1}, X_{i_2}, \ldots, X_{i_d}$, which is assumed to exist. For any integer p, $1 \leq p < d$, the conditional probability density function of $X'_j = (X_{j+i_{p+1}}, \ldots, X_{j+i_d})$ given $X''_j = (X_{j+i_1}, \ldots, X_{j+i_p})$ is denoted by

$$f(x_2 | x_1) = f(x; i_1, \ldots, i_d)/f(x_1; i_1, \ldots, i_p)$$

where $x_2 \in \mathbb{R}^{d-p}$ and $x = (x_1, x_2) \in \mathbb{R}^d$, put $f_p^*(x_1) = f(x_1; i_1, \ldots, i_p)$.

The estimators $\hat{f}_n(x)$ and $\hat{f}_n(x_2 | x_1)$ are defined in Section 2, their strong consistency and a.s. rates are stated and discussed therein. Meanwhile, some technical lemmas are given without proof, and an inequality for the maximum of the partial sum for (α, β)-mixing processes is given and proved. The derivations are given in Section 3.

2. Main results

Throughout the paper, we assume that for each $l \geq 1$, $K_l(x)$ is a real nonnegative function on $\mathbb{R}^l$ satisfying

$$\int_{\mathbb{R}^l} K_l(x)\, \mathrm{d}x = 1, \tag{2.1a}$$

$$\sup_{x \in \mathbb{R}^l} (1 + \|X\|^l)|K_l(x)| \leq M_0 \tag{2.1b}$$

for some positive number $M_0 < \infty$. Let $\{h_j; j \geq 1\}$ be a sequence of positive numbers such that

$$h_j \searrow 0 \quad \text{as } j \to \infty \quad \text{and} \quad nh_n^d/\log\log n \to \infty \text{ as } n \to \infty \tag{2.2}$$

and put

$$K_{l,j}(x) = h_j^{-1} K_l(x/h_j). \tag{2.3}$$

On the basis of a single realization $\{X_j; 1 \leq j \leq n\}$, we estimate $f_d(x)$, $d \geq 1$, by

$$\hat{f}_n(x) = \frac{1}{n - i_d} \sum_{j=1}^{n-i_d} K_{d,j}(x - X_j^*), \tag{2.4}$$

where $X_j^* = (X_{j+i_1}, \ldots, X_{j+i_d}) = (X'_j, X''_j)$ and it is naturally assumed that $n > i_d$. It is easy to see that

$$\hat{f}_n(x) = \frac{n - i_d - 1}{n - i_d} \hat{f}_{n-1}(x) + \frac{1}{n - i_d} K_{d,n-i_d}(x - X_{n-i_d}^*). \tag{2.5}$$

The conditional probability density function $f(x_2 | x_1)$ is estimated by

$$\hat{f}_n(x_2 | x_1) = \hat{f}_n(x)/\tilde{f}_n(x_1), \tag{2.6}$$

where

$$\tilde{f}_n(t) = \frac{1}{n - i_p} \sum_{j=1}^{n-i_p} K_{p,j}(t - X'_j), \quad t \in \mathbb{R}^p. \tag{2.7}$$

Let $\mathscr{F}_i^k$ be a σ-algebra of events generated by $\{X_j;\ i \leqslant j \leqslant k\}$ and $L_t(\mathscr{F}_a^b)$ denote the collection of all t-order random variables which are $\mathscr{F}_a^b$-measurable. For $0 \leqslant \alpha$, $\beta \leqslant 1$, $\alpha + \beta = 1$, the stationary process $\{X_j; j \geqslant 1\}$ is said to be (α, β)-mixing (see Bradley and Bryc, 1985; Shao, 1989) if

$$\sup_{\substack{\xi \in L_{\gamma\alpha}(\mathscr{F}_1^k) \\ \eta \in L_{\gamma\beta}(\mathscr{F}_{k+n}^{\infty})}} |E\xi\eta - E\xi E\eta| \leqslant \lambda(n) \cdot \|\xi\|_{1/\alpha} \|\eta\|_{1/\beta}$$

where $\lambda(n) \searrow 0$ as $n \to \infty$, which are called the (α, β)-mixing coefficients. It is not difficult to see that $(1, 0)$-mixing is uniform mixing (φ-mixing), and $(\frac{1}{2}, \frac{1}{2})$-mixing is usual ρ-mixing (see Shao, 1989).

2.1. Joint density estimators

The main results for the variance-like term $\hat{f}_n(x) - E\hat{f}_n(x)$ are given below. First, some notations are given, put $\tau_n = nh_n^d/\log\log n$, $n \geqslant 1$, and let f be a function on $\mathbb{R}^l$, $l \geqslant 1$, set $C(f) = \{x;\ f$ is continuous at point $x\}$ and $\gamma = \min\{\alpha^{-1}, \beta^{-1}\}$.

Theorem 2.1. *Let $\{X_j; j \geqslant 1\}$ be a stationary (α, β)-mixing process.*

(a) *If there exists a positive number $\delta > 5$ such that*

$$\lambda^{\gamma}(\tau_n) = \mathrm{O}(\log^{-\delta} n) \quad \text{as } n \to \infty \tag{2.8}$$

then, we have for all $x \in C(f_d)$,

$$\hat{f}_n(x) - f_d(x) \to 0 \quad \text{as } n \to \infty \tag{2.9}$$

almost surely.

(b) *If there exist two real numbers $\delta > 0$ and $\tau > 5$ such that*

$$\mu_n = (nh_n^d/(\log\log n)^{1+\delta})^{1/2} \nearrow \infty \quad \text{as } n \to \infty \tag{2.10}$$

and

$$\lambda^{\gamma}(\mu_n) = \mathrm{O}(\log^{-\tau} n) \quad \text{as } n \to \infty \tag{2.11}$$

then, for all $x \in C(f_d)$ we have

$$(nh_n^d/(\log\log n)^{1+\delta})^{1/2}(\hat{f}_n(x) - E\hat{f}_n(x)) \to 0 \tag{2.12}$$

as $n \to \infty$ almost surely.

Remark 2.1. For the i.i.d. case (and $d = 1$) we have (see Wegman and Davis, 1979)

$$\limsup_{n \to \infty} ((nh_n/\log\log n)^{1/2})(\hat{f}_n(x) - E\hat{f}_n(x)) = \theta\left(2f(x) \int_{\mathbb{R}} K^2(u)\,\mathrm{d}u\right)^{1/2} \tag{2.13}$$

almost surely, if $x \in C(f)$ and θ is a positive constant depending on $\{h_j; j \geqslant 1\}$ only. A comparison of (2.12) of Theorem 2.1 (with $d = 1$) and (2.13) shows that our a.s. convergence rate for (α, β)-mixing processes is very close to that in (2.13) for independent data.

Remark 2.2. The consistency of $\tilde{f}_n(x_1)$, needed in Subsection 2.2, if $x_1 \in C(f_p^*)$, follows immediately from (2.9) by replacing d by p, i_d by i_p, and x by x_1 in (2.9).

Remark 2.3. If $\tau_n = \mathrm{O}(n^\alpha)$ or $\mu_n = \mathrm{O}(n^\alpha)$ for some $\alpha > 0$, it is obvious that we only require the rate of decay of the ρ-mixing coefficients $\mathrm{O}(\log^{-\theta} n)$ for some $\theta > \frac{5}{2}$ in Theorem 2.1, then our conditions are weaker than those in Masry (1987, 1989).

2.2. Conditional density estimator

For brevity we only present the corresponding results to Theorem 2.1.

Theorem 2.2. *Let $\{X_j; j \geq 1\}$ be a stationary (α, β)-mixing process.*

(a) *If there exists a number $\delta > 5$ such that* (2.8) *holds, then we have for all $x \in C(f_d)$ and $x_1 \in C(f_p^*)$ with $f_p^*(x_1) > 0$,*

$$\hat{f}_n(x_2 | x_1) - f(x_2 | x_1) \to 0 \quad \text{as } n \to \infty \tag{2.14}$$

almost surely.

(b) *If there exist two real numbers $\delta > 0$ and $\tau > 5$ such that* (2.10) *and* (2.11) *hold, and assume that $f_d(x)$ and $f_p^*(x_1)$ are twice differentiable and their second partial derivatives are bounded and continuous on $\mathbb{R}^d$ and $\mathbb{R}^p$, respectively. Again, assume that the kernel functions K_d and K_p satisfy* (2.1) *and*

$$\int_{\mathbb{R}^l} u_j K_l(u)\, \mathrm{d}u = 0, \quad j = 1, 2, \ldots, l, \tag{2.15a}$$

$$\int_{\mathbb{R}^l} \|u\|^2 K_l(u)\, \mathrm{d}u < \infty, \tag{2.15b}$$

for $l = p$, d. The bandwidth parameter $\{h_j; j \geq 1\}$ is assumed to satisfy $\sum_{j=1}^\infty h_j^2 = \infty$, then

$$\begin{aligned} &\hat{f}_n(x_2 | x_1) - f(x_2 | x_1) \\ &\quad = \mathrm{O}\left(\sum_{j=1}^n h_j^2 / n \right) + \mathrm{O}((n h_n^d / (\log \log n)^{1+\delta})^{-1/2}) \quad \text{as } n \to \infty \end{aligned} \tag{2.16}$$

almost surely.

Remark 2.4. If $h_n \sim n^{-1/(4+d)}$, then (2.16) becomes

$$(n^{4/(4+d)} / (\log \log n)^{1+\delta})^{1/2} (\hat{f}_n(x_2 | x_1) - f(x_2 | x_1)) \to 0 \tag{2.17}$$

as $n \to \infty$ almost surely.

Remark 2.5. A comparison of our results and those of Masry (1987, 1989), shows that this work improves that of Masry's papers.

Throughout the paper, we shall use the denotation that C is a positive number but it can denote different values in different places. We will use the following result (see Sun, 1984, Theorem 1).

Lemma 2.1. *Assume that $K(x)$ is an integrable function on $\mathbb{R}^d$ ($d \geq 1$) satisfying the assumption* (2.1). *Then, for any integrable function $g(x)$ on $\mathbb{R}^d$, if g is continuous at point x, we have*

$$\lim_{h \to 0} \int_{\mathbb{R}^d} h^{-d} K((x-u)/h) g(u)\,\mathrm{d}u = g(x). \qquad \square$$

Lemma 2.2. *Let $\{X_j; j \geq 1\}$ be a (α, β)-mixing process. Assume that $p, q \geq 1$, $p^{-1} + q^{-1} = 1$, then for all $\xi \in L_p(\mathscr{F}_1^k)$ and $\eta \in L_q(\mathscr{F}_{k+n}^\infty)$ we have*

$$|E\xi\eta - E\xi E\eta| \leq 4\lambda^{\gamma_0}(n) \|\xi\|_p \cdot \|\eta\|_q$$

where $\gamma_0 = \min\{(\alpha p)^{-1}, (\beta q)^{-1}\}$.

Proof. See Shao (1989, Lemma 2.3.1) or Bradley and Bryc (1985, Theorem 4.1). $\square$

Lemma 2.3. *Let $\{X_j; j \geq 1\}$ be a sequence of (α, β)-mixing random variables with $EX_j = 0$, $E|X_j|^p < \infty$ ($p \geq 2$), $j = 1, 2, \ldots$, for any $k \geq 0$ and $n \geq 1$, put $T_k(n) = \sum_{j=k+1}^{k+n} X_j$. Then there exists a positive constant $C = C(p, \lambda(\cdot))$ such that*

$$E|T_k(n)|^p \leq C\left\{n^{p/2} \cdot \exp\left\{3p \sum_{j=0}^{[\log n]} \lambda^{\gamma/2}(2^j)\right\} \max_{k+1 \leq j \leq k+n} (EX_j^2)^{p/2} + n \cdot \exp\left\{C \sum_{j=0}^{[\log n]} \lambda^{\gamma/2}(2^j)\right\} \max_{k+1 \leq j \leq k+n} E|X_j|^p\right\}$$

for all $k \geq 0$ and $n \geq 1$.

Proof. See Shao (1989, Lemma 5.3.3). $\square$

The following theorem establishes a useful inequality of the maximum of the partial sum S_n with $S_n = \sum_{j=1}^n X_j$, $S_0 = 0$.

Theorem 2.3. *Let $\{X_j; j \geq 1\}$ be a sequence of (α, β)-mixing random variables satisfying $EX_n = 0$, $n = 1, 2, \ldots$, and $\sum_{n=1}^\infty \lambda^{\gamma/2}(2^n) < \infty$. Assume that there exist three sequences of positive constants $\{a_n\}$, $\{b_n\}$ and $\{c_n\}$ such that*

(a) $\max_{1 \leq j \leq n} |X_j| \leq b_n$ *a.s.*,

(b) $\max_{1 \leq j \leq n} EX_j^2 \leq c_n$,

(c) $\max_{1 \leq j \leq n} EX_j^4 \leq a_n$, *for all* $n \geq 1$. *Let* $\{p_n; n \geq 1\}$ *and* $\{q_n; n \geq 1\}$ *be two sequences of positive integers satisfying* $1 \leq q_n \leq p_n$, $p_n + q_n \leq n$, *for all* $n \geq 1$, *and again let* $\{\sigma_n^2; n \geq 1\}$ *be a sequence of positive numbers satisfying* $\sigma_n^2 \to \infty$ *as* $n \to \infty$. *Then, for any real numbers* $0 < \varepsilon_0 < \frac{1}{4}$ *and* x *satisfying*

$$\tfrac{1}{2}p_n b_n < x < 2\varepsilon_0(C^* nc_n + \varepsilon_0\sigma_n^2)/((1-4\varepsilon_0)p_n b_n) \tag{2.18}$$

it follows that there exist two positive constants C *independent of* n $(n \geq 1)$ *and* $C(\varepsilon_0)$ *such that*

$$\begin{aligned} P\left\{\max_{1\leq j\leq n} |S_j| \geq x\right\} \leq C\{&\exp\{-C(\varepsilon_0)x^2/(C^* nc_n + \varepsilon_0\sigma_n^2)\} \\ &+ x^{-2} nc_n \lambda^{\gamma}(q_n)\log^4(4n) \\ &+ \sigma_n^{-4} n(a_n + p_n c_n^2)\lambda^{\gamma}(q_n)\log^4(4n)\} \end{aligned} \tag{2.19}$$

for all $n \geq 1$, *where* C^* *is a positive constant independent of* n.

3. Proofs

Proof of Theorem 2.3. Set $p = p_n$, $q = q_n$ and $K = K_n = [n/(p+q)]$, $n = 1, 2, \ldots$, where $[x]$ denotes the integer part of x. Define, for all $i = 0, 1, \ldots, K-1$,

$$Y_K = \sum_{j=1+K(q+p)}^{n} X_j, \qquad Y_i = \sum_{j=1+i(p+q)}^{i(p+q)+p} X_j \quad \text{and} \quad Z_i = \sum_{j=1+i(p+q)+p}^{(i+1)(p+q)} X_j.$$

Obviously, $S_n = \sum_{i=0}^{K} Y_i + \sum_{i=0}^{K-1} Z_i$ and one has by (2.18) that

$$\begin{aligned} P\left\{\max_{1\leq j\leq n} |S_j| \geq x\right\} &\leq P\left\{\max_{0\leq i\leq K}\left|\sum_{j=0}^{i} Y_j\right| \geq (\tfrac{1}{2}-\varepsilon_0)x\right\} \\ &\quad + P\left\{\max_{0\leq i\leq K-1}\left|\sum_{j=0}^{i} Z_j\right| \geq \varepsilon_0 x\right\} \triangleq I_1 + I_2. \end{aligned} \tag{3.1}$$

Let $\mathscr{F}_i$ denote the σ-lagebra of events generated by $\{X_j; 1 \leq j \leq p + i(p+q)\}$ for $i = 0, 1, \ldots, K-1$ and $\mathscr{F}_{-1} = \{\Omega, \emptyset\}$. Define

$$V_i = Y_i - E(Y_i \mid \mathscr{F}_{i-1}) \quad \text{and} \quad U_i = \sum_{j=0}^{i} V_j$$

for $0 \leq i \leq K$, then clearly $\{V_i, \mathscr{F}_i; 0 \leq i \leq K\}$ is a martingale difference sequence, and

$$\begin{aligned} I_1 &\leq P\left\{\max_{0\leq i\leq K} |U_i| \geq (\tfrac{1}{2} - 2\varepsilon_0)x\right\} \\ &\quad + P\left\{\max_{1\leq i\leq K}\left|\sum_{j=1}^{i} E(Y_j \mid \mathscr{F}_{j-1})\right| \geq \varepsilon_0 x\right\} \triangleq I_3 + I_4. \end{aligned} \tag{3.2}$$

As in the proof of (2.3.58) in Shao (1989), it follows by Lemma 2.2 and Lemma 2.3 that there exists a positive constant C such that

$$E\left(\sum_{j=u+1}^{u+m} E(Y_j|\mathscr{F}_{j-1})\right)^2 \leq C\left\{mq\lambda^{\gamma}(q)\log^2(2m) \max_{u<j\leq u+m} EY_j^2\right\} \tag{3.3}$$

for all $0\leq u\leq K$ and $1\leq u+m\leq K$. Using Corollary 4 of Móricz (1982), we obtain

$$E\left(\max_{l\leq m}\left(\sum_{j=i+1}^{i+l} E(Y_j|\mathscr{F}_{j-1})\right)^2\right)\leq Cmq\lambda^{\gamma}(q)\log^4(4m) \max_{i\leq j\leq i+m} EY_j^2 \tag{3.4}$$

for all $0\leq i\leq K$ and $1\leq i+m\leq K$. Hence, one has by Lemma 2.2 and the Chebyshev's inequality that

$$I_4\leq Cx^{-2}nc_n\lambda^{\gamma}(q)\log^4(4n). \tag{3.5}$$

By (3.2), it follows that

$$I_3\leq P\left\{\max_{0\leq i\leq K} U_i\geq(\tfrac{1}{2}-2\varepsilon_0)x\right\}+P\left\{\max_{0\leq i\leq K}(-U_i)\geq(\tfrac{1}{2}-2\varepsilon_0)x\right\}\triangleq I_5+I_6. \tag{3.6}$$

In order to estimate I_5, set

$$I_7=P\left\{\max_{0\leq i\leq K} U_i\geq(\tfrac{1}{2}-2\varepsilon_0)x,\ \max_{0\leq i\leq K}\left(\sum_{j=1}^{i}(E(Y_j^2|\mathscr{F}_{j-1})-EY_j^2)\right)\leq\varepsilon_0\sigma_n^2\right\}$$

and

$$I_8=P\left\{\max_{1\leq i\leq K}\sum_{j=1}^{i}(E(Y_j^2|\mathscr{F}_{j-1})-EY_j^2)\geq\varepsilon_0\sigma_n^2\right\}.$$

Clearly,

$$I_5\leq I_7+I_8. \tag{3.7}$$

As in the proof of (3.5), one has by Lemma 2.2, the Chebyshev's inequality and Corollary 4 of Móricz (1982) that

$$I_8\leq Cn\sigma_n^{-4}\lambda^{\gamma}(q)\log^4(4n)(a_n+p_nc_n^2). \tag{3.8}$$

Set now $C_0(n)=2p_nb_n$ and for $\lambda>0$, $m=1,2,\ldots,K$,

$$T_m=\exp\left(\lambda U_m-\tfrac{1}{2}\lambda^2(1+\tfrac{1}{2}C_0(n)\lambda)\sum_{j=0}^{m}E(V_j^2|\mathscr{F}_{j-1})\right),$$

$$T_0=1.$$

Thus

$$\max_{j\leq n}|V_j|\leq C_0(n)$$

almost surely, it follows by Lemma 2.3 that there exists a constant $C^*>0$ such that

$$I_7 \leq P\left\{\max_{0\leq i\leq K} T_i > \exp(\lambda(\tfrac{1}{2}-2\varepsilon_0)x - \lambda^2/2(1+\tfrac{1}{2}C_0(n)\lambda)(nC^*c_n+\varepsilon_0\sigma_n^2))\right\}. \tag{3.9}$$

By Lemma 4.5.1 of Stout (1974, P.299), if $C_0(n)\lambda \leq 1$, we know that $\{T_m, \mathscr{F}_m; 0\leq m\leq K\}$ is a nonnegative supermartingale. Using well-known supermartingale inequality, by (2.18) and (3.9) for $\lambda = (\frac{1}{2}-2\varepsilon_0)x/(C^*nc_n+\varepsilon_0\sigma_n^2)$, we have

$$I_7 \leq \exp(-\tfrac{1}{16}(1-4\varepsilon_0)^3 x^2 (C^*nc_n^2+\varepsilon_0\sigma_n^2)^{-1}). \tag{3.10}$$

Similarly, the estimator of I_6 can be obtained. Combining (3.2), (3.5)-(3.8) and (3.10), we have

$$\begin{aligned} I_1 \leq C\{&\exp(-\tfrac{1}{16}(1-4\varepsilon_0)^3 x^2 (C^*nc_n^2+\varepsilon_0\sigma_n^2)^{-1}) \\ &+ n\sigma_n^{-4}\lambda^{\gamma}(q)\log^4(4n)(a_n+pc_n^2)+x^{-2}nc_n\lambda^{\gamma}(q)\log^4(4n)\}. \end{aligned}$$

Since $\lambda(n)\downarrow 0$ as $n\to\infty$, we can also show that the estimator of I_2 is as same as that of I_1. This, jointly with (3.1), completes the proof of (2.19). □

Proof of Theorem 2.1. Assume that x is a continuous point of f_d. First, fix x and put

$$\xi_j(x) = K_{d,j}(x-X_j^*) - EK_{d,j}(x-X_j^*)$$

for $j=1,2,\ldots,n-i_d$ and

$$S_n(x) = \sum_{j=1}^{n-i_d} \xi_j(x).$$

Obviously, one has by (2.1) and (2.4) that

$$S_n(x) = (n-i_d)(\hat{f}_n(x) - E\hat{f}_n(x)) \tag{3.11}$$

and

$$\max_{1\leq j\leq n-i_d} |\xi_j(x)| \leq 2M_0 h_n^{-d}$$

almost surely. By Lemma 3.1, there exist two positive constants $C_1(x)$ and $C_2(x)$ such that

$$\max_{1\leq j\leq n-i_d} E\xi_j^2(x) \leq C_1(x) h_n^{-d}$$

and

$$\max_{1\leq j\leq n-i_d} E\xi_j^4(x) \leq C_2(x) h_n^{-3d}.$$

Set $b_n = 2M_0h_n^{-d}$, $a_n = C_2(x)h_n^{-3d}$, $c_n = C_1(x)h_n^{-d}$. For an arbitrary $0<\varepsilon<1$, put

$$\varepsilon_0 = \tfrac{1}{8}, \qquad \sigma_n^2 = nh_n^{-d},$$

$$x = \varepsilon((n-i_d)^2 h_n^{-d}(\log\log n)^{1+\delta}/n)^{1/2}$$

and

$$q_n = [\mu_n], \qquad p_n = [2\mu_n],$$
$$\gamma(n) = ((n - i_d)^2 h_n^{-d} (\log\log n)^{1+\delta}/n)^{-1/2}.$$

It is easy to show that there exists a positive constant C such that

$$P\left\{ \gamma(n) \max_{1 \leq j \leq n} |S_j(x)| \geq \varepsilon \right\}$$
$$\leq C\{\exp(-C(\log\log n)^{1+\delta}) + \log^{-1} n(\log\log n)^{-1-\delta} + n^{-1}\log^{-1} n(h_n^{-d} + 2\mu_n)\}$$

by Theorem 2.3. Hence, for n sufficiently large we have

$$P\left\{ \gamma(n) \max_{1 \leq j \leq n} |S_j(x)| \geq \varepsilon \right\} \leq C \log^{-1} n(\log\log n)^{-1-\delta}.$$

Putting $n_k = 2^k$, $k = 1, 2, \ldots$, by the Borel-Cantelli lemma, one has

$$\gamma(n_k) \max_{1 \leq j \leq n_k} |S_j(x)| \to 0 \quad \text{as } k \to \infty$$

almost surely. For n sufficiently large, there exists k such that $n_{k-1} < n \leq n_k$ and

$$\max_{n_{k-1} \leq n \leq n_k} \gamma(n)/\gamma(n_k) \leq 4.$$

Hence

$$\gamma(n)|S_n(x)| \leq 4\gamma(n_k) \max_{1 \leq j \leq n_k} |S_j(x)|.$$

This, jointly with (3.11), suffices to prove (2.12). Similarly, we can show by Theorem 2.3 that

$$\hat{f}_n(x) - E\hat{f}_n(x) \to 0 \quad \text{as } n \to \infty$$

almost surely. We obtain (2.9) by Lemma 2.1. Thus, we have completed the proof of the theorem. □

Proof of Theorem 2.2. (a) Under the assumptions of this part, for all $x \in C(f_d)$, we have

$$\hat{f}_n(x) - f_d(x) \to 0 \quad \text{as } n \to \infty$$

almost surely. It follows by Lemma 2.3 that for all $x_1 \in C(f_p^*)$,

$$\tilde{f}_n(x_1) - f_p^*(x_1) \to 0 \quad \text{as } n \to \infty$$

almost surely. Thus, (a) follows by using the identity

$$\hat{a}/\hat{b} - a/b = ((\hat{a} - a) - a(\hat{b} - b)/b)/\hat{b}$$

with

$$\hat{a} = \hat{f}_n(x), \qquad a = f_d(x), \qquad \hat{b} = \tilde{f}_n(x_1), \qquad b = f_p^*(x_1).$$

(b) Under the assumptions of this part, by Theorem 2.1 it suffices to show that

$$(nh_n^d/(\log\log n)^{1+\delta})^{1/2}(\hat{a}-E\hat{a})\to 0 \quad \text{as } n\to\infty \tag{3.12}$$

almost surely and

$$(nh_n^p/(\log\log n)^{1+\delta})^{1/2}(\hat{b}-E\hat{b})\to 0 \quad \text{as } n\to\infty \tag{3.13}$$

almost surely. Since

$$\begin{aligned}\hat{a}/\hat{b}-a/b &= ((\hat{a}-a)-a(\hat{b}-b)/b)/\hat{b}\\ &= ((\hat{a}-E\hat{a})-a(\hat{b}-E\hat{b})/b)/\hat{b}+((E\hat{a}-a)-a(E\hat{b}-b)/b)/\hat{b}.\end{aligned} \tag{3.14}$$

Also, under the assumptions of this part, it follows by Lemma 2.1(b) of Masry (1989) that for n sufficiently large,

$$E\hat{a}-a=\mathrm{O}\left(\sum_{j=1}^{n} h_j^2/n\right) \tag{3.15}$$

and

$$E\hat{b}-b=\mathrm{O}\left(\sum_{j=1}^{n} h_j^2/n\right). \tag{3.16}$$

Now (b) follows by (3.12)-(3.16). Thus, we have completed the proof of the theorem. □

Acknowledgment

The author is indebted to the referee and the editor for their careful reading of an earlier version of the manuscript and for their constructive suggestions.

References

R.C. Bradley and W. Bryc, Multilinear forms and measures of dependence between random variables, J. Multivariate Anal. 16 (1985) 335-367.

Z.W. Cai, Strong consistency and rates for probability density estimators of weakly dependent random variables, Chinese J. Sys. Sci. Math. Sci. 10 (1990) 360-370.

Z.W. Cai, Uniform strong consistency and rates for kernel estimators of a distribution function and a regression function under weakly dependent observations, to appear in: Chinese J. Appl. Statist. Probab.

E. Masry, Nonparametric estimation of conditional probability density and expectation of stationary processes; strong consistency and rates, Stochastic Process. Appl. 32 (1989) 109-127.

E. Masry and L. Györfi, Strong consistency and rates for recursive probability density estimators of stationary processes, J. Multivariate Anal. 22 (1987) 79-83.

F.A. Móricz, A general moment inequality for the maximum of partial sums of single series, Acta Sci. Math. 44 (1982) 67-75.

P.M. Robinson, Nonparametric estimators for time series, J. Time Ser. Anal. 4 (1983) 185-207.

P.M. Robinson, On the consistency and finite sample properties of nonparametric kernel time series regression, autoregression and density estimators, Ann. Inst. Statist. Math. 38 (1986) 539-549.

Q.M. Shao, Limit theorems of sums for dependent and independent random variables, Ph.D. thesis, Tech. & Sci. Univ. of China (Anhui, 1989).
W.F. Stout, Almost Sure Convergence (Academic Press, New York, 1974).
Z.G. Sun, Asympotic unbiased and strong consistency for density function estimator, Acta Math. Sinica 27 (1984) 769-782.
E.J. Wegman and H.I. Davis, Remarks on some recursive estimators of a probability density, Ann. Statist. 7 (1979) 316-327.