Stochastic Processes and their Applications 36 (1990) 107-116
North-Holland

NONPARAMETRIC REGRESSION ESTIMATION UNDER MIXING CONDITIONS

George G. ROUSSAS
Division of Statistics, University of California, Davis, CA 95616, USA

Received 1 June 1987
Revised 31 May 1989

For $j=1,2,\ldots$, let $\{Z_j\}=\{(X_j, Y_j)\}$ be a strictly stationary sequence of random variables, where the X's and the Y's are $\mathbb{R}^p$-valued and $\mathbb{R}^q$-valued, respectively, for some integers $p, q \geq 1$. Let ϕ be an integrable Borel real-valued function defined on $\mathbb{R}^q$ and set $r(x)=\mathscr{E}[\phi(Y)\,|\,X=x]$, $x\in\mathbb{R}^p$. The function ϕ need not be bounded. The quantity $r(x)$ is estimated by $r_n(x)=R_n(x)/\hat{f}_n(x)$, where $\hat{f}_n(x)$ is a kernel estimate for the probability density function f of the X's and $R_n(x)=(nh^p)^{-1}\sum_{j=1}^{n}\phi(Y_j)\cdot K((x-X_j)/h)$. If the sequence $\{Z_j\}$ enjoys any one of the standard four kinds of mixing properties, then, under suitable additional assumptions, $r_n(x)$ is strongly consistent, uniformly over compacts. Rates of convergence are also specified.

AMS 1980 Subject Classifications: Primary 62G05; Secondary 62J02, 62M05, 62M09, 62H12.

nonparametric regression estimates * kernel estimates * stationarity * mixing

1. Introduction

Let $\{Z_j\}, j=1,2,\ldots$, be a strictly stationary sequence of real-valued random variables (r.v.'s). The problem under consideration is that of nonparametrically estimating the conditional expectation of Z_{j+1}, given the immediately u previous r.v.'s $Z_{j-u+1},\ldots,Z_j$, under suitable regularity conditions. This problem may be cast in a slightly different framework as follows. Consider the strictly stationary sequence of pairs of r.v.'s $\{(X_j, Y_j)\}$, $j\geq 1$, where the X's are $\mathbb{R}^p$-valued and the Y's are real-valued. Then, under appropriate regularity conditions, the problem is that of providing a nonparametric estimate for the regression function $\mathscr{E}(Y\,|\,X=x)$; the r.v.'s X and Y are distributed as the r.v.'s X_1 and Y_1, respectively.

In a recent paper [4], Collomb and Härdle, pursuing previous works of their own as well as those of others (see, for example, [2, 3, 6, 7, 11, 12, 13]), proposed, among other things, robust estimates for the above mentioned regression function working along the lines of M-estimates. Basic among their assumption is the property of

This research was partially supported by a grant from the Office of Research and Technology of Greece.

0304-4149/90/$3.50 © 1990—Elsevier Science Publishers B.V. (North-Holland)

ϕ-mixing for the underlying sequence. They note, however, that, for Gaussian stationary autoregressive processes, the ϕ-mixing condition is equivalent to m-dependence (see [9, Theorem 17.3.2]). They proceed then with the advice that future research along those lines be based on weaker mixing conditions. It is in response to this suggestion that this paper is written.

Actually, what is done here is to exploit results recently obtained in connection with the problem of estimating the (common) probability density function (p.d.f.) of r.v.'s, forming a strictly stationary sequence and subject to any one of the standard four kinds of mixing, by means of the kernel method. As will be seen, these results allow us to construct an estimate of the regression function under consideration—or, more properly, of a quantity which includes this regression function as a special case. This estimate is shown to be strongly consistent, uniformly over a certain compact subset of $\mathbb{R}^p$. Rates of convergence are also given. The setting, however, is that of the conventional approach. The relevant arguments, suitably modified and reinforced by assumptions stronger than those employed herein, may also be used for constructing M-estimates the way is done in [4]. The relevant derivations are intended to be presented in a subsequent paper.

Other results of the same general nature which have been recently obtained, under various settings, are those reported in [8, 14, 20]. Thus, in [8], times series are considered which exhibit a deterministic trend, and, within a certain framework, this trend is estimated by the so-called M-smoothers. Consistency in the probability sense and asymptotic normality of the proposed estimate is established. In [14], time series of vector-valued r.v.'s are considered and it is assumed that they satisfy one of the modes of ϕ_i-mixing for $i=1$ or 2 or 4. Nonparametric density and regression estimators are constructed, and they are shown to be pointwise consistent in quadratic mean and also in the probability sense. Finally, in [20], for Markovian processes which are geometrically ergodic, the regression $\mathscr{E}(X_2 \mid X_1 = x)$ is estimated by means of the nearest-neighbour method, and the proposed estimate is shown to be consistent in quadratic mean. Analogous results are obtained for certain non-Markovian processes.

2. Notation, definitions and assumptions

Regarding notation and assumptions, we are going to draw heavily on those used in [10, 16, 18]; see also [19]. In an attempt to minimize the references, the basics will be presented below. In the literature, the four kinds of mixing often used are usually referred to as ψ-mixing, ϕ (or uniform)-mixing, ρ-mixing (or mixing based on the maximal correlation) and α (or strong)-mixing. Notationally, it is convenient to label them as ϕ_i-mixing, $i=1,\ldots,4$, respectively, so that ϕ_i-mixing implies ϕ_{i+1}-mixing, $i=1,2,3$. Thus, let $Z_j=(X_j, Y_j)$, $j=1,2,\ldots$, be pairs of r.v.'s defined on some probability space $(\Omega, \mathscr{A}, P)$ and such that the X's take values on $\mathbb{R}^p$ and the Y's take values on $\mathbb{R}^q$ for some p, $q \geqslant 1$ integers. It will be assumed below that the sequence $\{Z_j\}$ is strictly stationary, and, for $m<n$, let $\mathscr{F}_m^n$ be the σ-field induced in Ω by the r.v.'s Z_j, $m\leqslant j\leqslant n$. Then:

Definition 2.1. The sequence $\{Z_j\}$ is said to be ϕ_i-mixing, $i=1,\ldots,4$, if, for every $A\in\mathscr{F}_1^k$ and every $B\in\mathscr{F}_{k+n}^{\infty}$, the following inequalities are satisfied:

For ϕ_1-mixing:

$$|P(A\cap B)-P(A)P(B)|\leqslant\phi_1(n)P(A)P(B)\quad\text{with } \phi_1(n)\downarrow 0,\ \text{as } n\to\infty.$$

For ϕ_2-mixing:

$$|P(A\cap B)-P(A)P(B)|\leqslant\phi_2(n)P(A)\quad\text{with } \phi_2(n)\downarrow 0,\ \text{as } n\to\infty.$$

For ϕ_3-mixing:

$$|P(A\cap B)-P(A)P(B)|\leqslant\phi_3(n)[P(A)P(B)]^{1/2}\quad\text{with } \phi_3(n)\downarrow 0,\ \text{as } n\to\infty.$$

For ϕ_4-mixing:

$$|P(A\cap B)-P(A)P(B)|\leqslant\phi_4(n)\quad\text{with } \phi_4(n)\downarrow 0,\ \text{as } n\to\infty.$$

Rather than seeking to construct an estimate of the regression function $\mathscr{E}(Y\,|\,X=x)$ (for the case that $q=1$), it would be advisable to consider conditional expectations of the form $\mathscr{E}[\phi(Y)\,|\,X=x]$, where ϕ is an integrable Borel real-valued function defined on $\mathbb{R}^q$. Then, referring to the initial set-up of real-valued r.v.'s $Z_j, j\geqslant 1$, one may estimate by suitable choices of ϕ, interesting quantities such as $\mathscr{E}(Z_{j+s}^m\,|\,Z_{j-u+1},\ldots,Z_j)$, $m>0$ and $s\geqslant 0$ both integers, or a suitable function of a finite number of Z's beyond the present Z_j. It is to be noted, however, that neither the function ϕ nor the Y's in the setting $Z_j=(X_j, Y_j)$, the X's being $\mathbb{R}^p$-valued and the Y's being $\mathbb{R}^q$-valued, need be bounded, contrary to what is customarily assumed in the literature.

It would be advisable to group together the assumptions to be used in this paper. These assumptions are explicitly stated for the case of ϕ_4-mixing, and the results are also established for this case only, for the sake of ease of presentation. Details for the remaining kind of mixing may be found in [17].

Assumptions. (A1) The r.v.'s $Z_j=(X_j, Y_j), j=1,2,\ldots,$ with the X's being $\mathbb{R}^p$-valued and the Y's being $\mathbb{R}^q$-valued, p, q, $\geqslant 1$ integers, form a strictly stationary sequence $\{Z_j\}$.

(A2) The sequence $\{Z_j\}$ is ϕ_4-mixing and the mixing coefficient ϕ_4 satisfies the following summability requirement: $\phi_4^*=\sum_{j=1}^{\infty}\phi_4(j)<\infty$.

(A3) Let $\alpha=\alpha(n)$ be a positive integer and let $\mu=\mu(n)$ be the largest positive integer for which $2\alpha\mu\leqslant n$. Then

$$\limsup[1+6\,\mathrm{e}^{1/2}\phi_4^{1/(\mu+1)}(\alpha)]^{\mu}<\infty.$$

(A4) (i) The sequence $\{h\}=\{h_n\}\downarrow 0$ and K is a bounded p.d.f. defined on $\mathbb{R}^p$ into $\mathbb{R}$.

(ii) $\|x\|^p K(x) \to 0$, as $\|x\| \to \infty$, where the norm $\|x\|$ of $x = (x_1, \ldots, x_p)$ is defined by $\|x\| = \max(|x_1|, \ldots, |x_p|)$.

Assumption (A5) below is meant to take care of the unboundedness of ϕ.

(A5) (i) ϕ is a real-valued Borel function defined on $\mathbb{R}^q$ such that $\mathscr{E}|\phi(Y)|^s < \infty$ for some $s > 1$.

(ii) Let $0 < B_n \uparrow \infty$ be truncating constants as in (3.4) below. Then B_n also satisfy the convergence requirement $\sum_{n=1}^{\infty} B_n^{-s} < \infty$.

$$\text{(iii)} \quad \sup\left[\int_{\mathbb{R}^q} |\phi(y)|^s f_{X,Y}(x, y)\, \mathrm{d}y;\ x \in \mathbb{R}^p\right] = C < \infty,$$

where $f_{X,Y}$ is, of course, the joint p.d.f. of X and Y; X and Y are distributed as X_1 and Y_1, respectively.

(A6) (i) For any two points x and x' in $\mathbb{R}^p$ and for some positive constant C (independent of these points):

$$|K(x) - K(x')| \leq C\|x - x'|.$$

$$\text{(ii)} \quad \int_{\mathbb{R}^p} \|x\| K(x)\, \mathrm{d}x < \infty.$$

(A7) For any point x in $\mathbb{R}^p$, there are (generic) positive constants $C(x)$ such that, for all $x' \in \mathbb{R}^p$ and with J being as in (A8):

$$\text{(i)} \quad |f(x) - f(x')| \leq C(x)\|x - x'\|, \qquad \sup[C(x);\ x \in J] = C < \infty;$$

$$\text{(ii)} \quad |\psi(x) - \psi(x')| \leq C(x)\|x - x'\|, \qquad \sup[C(x);\ x \in J] = C < \infty,$$

where $\psi(x) = r(x) f(x)$.

(A8) There exists a compact subset J of $\mathbb{R}^p$ such that

$$\inf[f(x);\ x \in J] > 0.$$

Remark 2.1. The condition stated in assumption (A3) appears in a natural manner in establishing exponential bounds for probabilities of large deviations. It is therefore closely tied to the particular way of proving those results. (See proofs of Lemmas 2.3 and 3.3 in [10].) Alternative conditions leading to the same results would be equally appropriate.

In order to avoid unnecessary repetitions, it is to be mentioned here once and for all that all limits are taken as $\{n\}$ or subsequences thereof tend to ∞, unless otherwise explicitly stated.

3. Some preliminary results

Set

$$r(x) = \mathscr{E}[\phi(Y)|X = x], \quad x \in \mathbb{R}^p. \tag{3.1}$$

The quantity $r(x)$ is to be estimated by $r_n(x)$, where

$$r_n(x) = R_n(x)/\hat{f}_n(x), \tag{3.2}$$

and

$$\hat{f}_n(x) = \frac{1}{nh^p}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right), \qquad R_n(x) = \frac{1}{nh^p}\sum_{j=1}^{n} \phi(Y_j)K\left(\frac{x - X_j}{h}\right). \tag{3.3}$$

The function ϕ is not assumed to be bounded. The problems arising from this unboundedness are dealt with by means of suitable truncation. In handling the details, Proposition 1 in [11] was particularly helpful. Denote by

$$\phi^B(y) = \phi(y)I_{\{|\phi(y)|<B_n\}}(y), \quad 0 < B_n \uparrow \infty, \tag{3.4}$$

and set

$$R_n^B(x) = \frac{1}{nh^p}\sum_{j=1}^{n} \phi^B(Y_j)K\left(\frac{x - X_j}{h}\right). \tag{3.5}$$

Now, for $k > 0$, denote by C_k the cube in $\mathbb{R}^p$ which is the cartesian product of p copies of $[-k, k]$. Then, working as in [16] (see also [1]), divide $[-k, k]$ into b_n subintervals each of length δ_n, and let J_{nl}, $l = 1, \ldots, N$, be the sets into which C_k is divided. Let x_{nl} be arbitrary points in J_{nl}. Pick k sufficiently large, so that $C_k \supset J$. Then, clearly,

$$\begin{aligned}
|r_n(x) - r(x)| &= \hat{f}_n^{-1}(x)|R_n(x) - r(x)\hat{f}_n(x)| \\
&\leqslant \hat{f}_n^{-1}(x)\{|R_n(x) - R_n^B(x)| \\
&\qquad + |\mathscr{E}R_n(x) - \mathscr{E}R_n^B(x)| + |R_n^B(x) - R_n^B(x_{nl})| \\
&\qquad + |\mathscr{E}R_n^B(x) - \mathscr{E}R_n^B(x_{nl})| \\
&\qquad + |\mathscr{E}R_n(x) - \psi(x)| + r(x)|\hat{f}_n(x) - f(x)| \\
&\qquad + |R_n^B(x_{nl}) - \mathscr{E}R_n^B(x_{nl})|\},
\end{aligned} \tag{3.6}$$

where it is recalled that

$$\psi(x) = r(x)f(x). \tag{3.7}$$

The behavior of the terms on the right hand side of (3.6) is exhibited in the following results.

Proposition 3.1. *Let $R_n(x)$ and $R_n^B(x)$ be defined by* (3.3) *and* (3.5), *respectively, and suppose that assumptions* (A1) *and* (A5)(i)-(ii) *are satisfied. Then, for each ω outside a null set N, there exists a positive integer $n_0(\omega)$ such that*

$$R_n(x) = R_n^B(x) \quad \textit{for } n \geqslant n_0(\omega) \textit{ and for all } x \in \mathbb{R}^p.$$

Proof. For every $\varepsilon > 0$, one has

$$P\left[\left|\frac{\phi(Y_n)}{B_n}\right| > \varepsilon\right] = P[|\phi(Y_n)| > \varepsilon B_n] \leqslant \varepsilon^{-s}\mathscr{E}|\phi(Y)|^s B_n^{-s},$$

so that $\phi(Y_n)/B_n \to 0$ a.s. Since $0 < B_n \uparrow \infty$, for $\omega \in N^c$ with $P(N^c) = 1$, one obtains

$$|\phi[Y_j(\omega)]| < B_n, \quad j = 1, \ldots, n, \quad n \geqslant n_0(\omega),$$

some positive integer. The definition of $R_n(x)$ and $R_n^B(x)$ completes the proof. $\square$

Also, we have:

Proposition 3.2. *Under assumptions* (A1), (A4)(i), (A5)(i) *and* (A5)(iii), *it follows that*

$$|\mathscr{E}R_n(x) - \mathscr{E}R_n^B(x)| \leqslant CB_n^{1-s} \quad \textit{for all } x \in \mathbb{R}^p.$$

Proof. In fact,

$$\begin{aligned}
|\mathscr{E}R_n(x) - \mathscr{E}R_n^B(x)| &= \left| \mathscr{E}\left[\frac{1}{nh^p} \sum_{j=1}^{n} \phi(Y_j) I_{\{|\phi(Y_j)| \geqslant B_n\}}(Y_j) K\left(\frac{x - X_j}{h}\right) \right] \right| \\
&\leqslant h^{-p} \mathscr{E} \left| \phi(Y) I_{\{|\phi(Y)| \geqslant B_n\}}(Y) K\left(\frac{x - X}{h}\right) \right| \\
&= h^{-p} \int_{\mathbb{R}^p} \int_{(|\phi(y)| \geqslant B_n)} |\phi(y)| K\left(\frac{x - t}{h}\right) f_{X,Y}(t, y) \, \mathrm{d}y \, \mathrm{d}t \\
&= h^{-p} \int_{\mathbb{R}^p} K\left(\frac{x - t}{h}\right) \left[\int_{(|\phi(y)| \geqslant B_n)} |\phi(y)| f_{X,Y}(t, y) \, \mathrm{d}y \right] \mathrm{d}t \\
&\leqslant h^{-p} \int_{\mathbb{R}^p} K\left(\frac{x - t}{h}\right) \left[\sup_{\tau \in \mathbb{R}^p} \int_{(|\phi(y)| \geqslant B_n)} |\phi(y)| f_{X,Y}(\tau, y) \, \mathrm{d}y \right] \mathrm{d}t \\
&= \int_{\mathbb{R}^p} K(u) \left[\sup_{\tau \in \mathbb{R}^p} \int_{(|\phi(y)| \geqslant B_n)} |\phi(y)| f_{X,Y}(\tau, y) \, \mathrm{d}y \right] \mathrm{d}u \\
&= \sup_{\tau \in \mathbb{R}^p} \int_{(|\phi(y)| \geqslant B_n)} |\phi(y)| f_{X,Y}(\tau, y) \, \mathrm{d}y \\
&= \sup_{\tau \in \mathbb{R}^p} \int_{(|\phi(y)| \geqslant B_n)} |\phi(y)|^{1-s} |\phi(y)|^s f_{X,Y}(\tau, y) \, \mathrm{d}y \\
&\leqslant B_n^{1-s} \sup_{\tau \in \mathbb{R}^p} \int_{(|\phi(y)| \geqslant B_n)} |\phi(y)|^s f_{X,Y}(\tau, y) \, \mathrm{d}y \\
&\leqslant B_n^{1-s} \sup_{\tau \in \mathbb{R}^p} \int_{\mathbb{R}^q} |\phi(y)|^s f_{X,Y}(\tau, y) \, \mathrm{d}y \\
&\leqslant CB_n^{1-s}.
\end{aligned}$$

Thus,

$$|\mathscr{E}R_n(x) - \mathscr{E}R_n^B(x)| \leqslant CB_n^{1-s} \quad \text{for all } x \in \mathbb{R}^p. \qquad \square$$

One may also establish the following result regarding the quantity $R_n(x)$ and its truncated version $R_n^B(x)$. More precisely, one has:

Proposition 3.3. *Under assumptions* (A1), (A4)(i), (A5)(i), (A6) *and* (A7)(ii), *the following results hold true*:

(i) $|\mathscr{E}R_n(x)-\psi(x)|\leqslant C(x)h$ *for all* $x\in\mathbb{R}^p$.

(ii) *For any points* x, $x'\in\mathbb{R}^p$, *both* $|R_n^B(x)-R_n^B(x')|$ *and* $|\mathscr{E}R_n^B(x)-\mathscr{E}R_n^B(x')|$ *are bounded by* $CB_nh^{-(p+1)}\ \|x-x'\|$.

Proof. (i) It is easily seen that

$$\mathscr{E}R_n(x)=\frac{1}{h^p}\int_{\mathbb{R}^p} r(x-u)K\left(\frac{u}{h}\right)f(x-u)\,\mathrm{d}u=\int_{\mathbb{R}^p}\psi(x-ht)K(t)\,\mathrm{d}t,$$

so that

$$|\mathscr{E}R_n(x)-\psi(x)|=\left|\int_{\mathbb{R}^p}\psi(x-ht)K(t)\,\mathrm{d}t-\int_{\mathbb{R}^p}\psi(x)K(t)\,\mathrm{d}t\right|$$

$$\leqslant\int_{\mathbb{R}^p}|\psi(x)-\psi(x-ht)|K(t)\,\mathrm{d}t\leqslant C(x)h\int_{\mathbb{R}^p}\|t\|K(t)\,\mathrm{d}t=C(x)h.$$

(ii) It suffices to consider the first quantity only. One has

$$|R_n^B(x)-R_n^B(x')|\leqslant\frac{1}{nh^p}\sum_{j=1}^{n}\left|\phi^B(Y_j)\left[K\left(\frac{x-X_j}{h}\right)-K\left(\frac{x'-X_j}{h}\right)\right]\right|$$

$$\leqslant\frac{1}{h^p}B_nC\frac{1}{h}\|x-x'\|=CB_nh^{-(p+1)}\|x-x'\|.\qquad\square$$

Now let ψ_n be a (positive) norming factor and let the quantities α, h and δ_n be chosen, so that

$$\psi_nh\to 0,\qquad \psi_n\delta_nh^{-(p+1)}\to 0\quad\text{and}\quad\{\psi_n/\alpha h^p\}\text{ is bounded away from }0.\tag{3.8}$$

Then, under the assumptions made in this paper, Theorem 3.1 in [10] applies and gives, for each x_{nl}, $l=1,\ldots,N$,

$$P[\psi_n|\hat f_n(x_{nl})-\mathscr{E}\hat f_n(x_{nl})|>\varepsilon]\leqslant C\exp[-C\varepsilon^2\psi_n^{-2}nh^{2p}],\tag{3.9}$$

where C is a generic constant. Then, working as in the proof of Theorem 3.1 in [16], utilizing (3.9) and the fact that $N\leqslant C\delta_n^{-p}$, one has

$$P[\sup_{x\in J}\psi_n|\hat f_n(x)-f(x)|>\varepsilon]\leqslant\sum_{l=1}^{N}P[\psi_n|\hat f_n(x_{nl})-\mathscr{E}\hat f_n(x_{nl})|>\tfrac{1}{3}\varepsilon]$$

$$\leqslant C\delta_n^{-p}\exp(-C\varepsilon^2\psi_n^{-2}nh^{2p}).\tag{3.10}$$

Finally, let the quantities ψ_n, α, h and δ_n be as in (3.8) and also satisfy the condition

$$\sum_{n=1}^{\infty}\delta_n^{-p}\exp(-C\psi_n^{-2}nh^{2p})<\infty\quad(C>0).\tag{3.11}$$

Then, by the Borel-Cantelli lemma, one has:

Proposition 3.4. *Suppose assumptions* (A1)-(A4)(i), (A6) *and* (A7)(i) *are satisfied. Then, for a (positive) norming factor* ψ_n, *one has*

$$\sup[\psi_n|\hat{f}_n(x)-f(x)|;\ x\in J]\to 0 \quad a.s.,$$

provided α, h, δ_n *and* ψ_n *are chosen, so that* (3.8) *and* (3.11) *are fulfilled.* □

Finally, the boundedness of $R_n^B(x)$, $x\in\mathbb{R}^p$, allow us to suitably apply inequality (3.9), use the fact that $N\leqslant C\delta_n^{-p}$, and the Borel-Cantelli lemma in order to obtain:

Proposition 3.5. *Suppose assumptions* (A1)-(A4)(i) *are satisfied. Then*

$$\max_l \psi_n|R_n^B(x_{nl})-\mathscr{E}R_n^B(x_{nl})|\to 0 \quad a.s.,$$

provided α, h, δ_n, B_n *and* ψ_n *are chosen so that*

$$\sum_{n=1}^{\infty} \delta_n^{-p}\exp(-C\psi_n^{-2}nB_n^{-2}h^{2p})<\infty \quad (C>0),$$

the convergences in (3.12) *below hold and* $\{\psi_nB_n/\alpha h^p\}$ *is bounded away from* 0, *where*

$$\psi_nh\to 0, \qquad \psi_nB_n^{1-s}\to 0, \qquad \psi_nB_n\delta_nh^{-(1+p)}\to 0. \quad □ \tag{3.12}$$

4. Main result and concluding remarks

The main result of this paper is the following theorem whose proof is a consequence of the preliminary results established. More precisely, one has:

Theorem 4.1. *Under assumptions* (A1)-(A8),

$$\sup[\psi_n|r_n(x)-r(x)|;\ x\in J]\to 0 \quad a.s.,$$

provided α, h, δ_n, B_n *and* ψ_n *are chosen, so that*

$$\sum_{n=1}^{\infty} \delta_n^{-p}\exp(-C\psi_n^{-2}nB_n^{-2}h^{2p})<\infty \quad (C>0),$$

the convergences in (3.12) *hold true and* $\{\psi_n/\alpha h^p\}$ *is bounded away from* 0.

Proof. By (A7), r is continuous, and by (A8) and Proposition 3.4, $\sup_{x\in J}\hat{f}_n^{-1}(x)$ is bounded a.s. for all sufficiently large n. Next, multiply both sides of (3.6) by ψ_n, take the suprema over $x\in J$, utilize Propositions 3.1-3.3 and the fact that $x\in J_{nl}$ implies $\|x-x_{nl}\|\leqslant\delta_n$, in order to obtain, a.s. and for all sufficiently large n,

$$\begin{aligned}\sup_{x\in J}\psi_n|r_n(x)-r(x)| &\leqslant C\psi_nB_n^{1-s}+C\psi_nB_n\delta_nh^{-(p+1)}+C\psi_nh\\ &\quad +C\sup_{x\in J}\psi_n|\hat{f}_n(x)-f(x)|+\max_l\psi_n|R_n^B(x_{nl})-\mathscr{E}R_n^B(x_{nl})|.\end{aligned}$$

Taking the limits (as $n \to \infty$), the desired result follows by means of Propositions 3.4 and 3.5. $\square$

Remark 4.1. Under assumptions (A1)–(A5) alone, it is seen that $r_n(x) \to r(x)$ a.s. for any continuity point x of f, provided α, h and B_n are chosen, so that

$$\sum_{n=1}^{\infty} \delta_n^{-p} \exp(-CnB_n^{-2}h^{2p}) < \infty \quad (C > 0)$$

and $\{1/\alpha h^p\}$ is bounded away from 0. For details, the interested reader is referred to Theorem 3.1 in [17].

Remark 4.2. In order to gain some insight as to what kind of h, α, B_n, δ_n, and ψ_n satisfy the assumptions made in Theorem 4.1, let us suppose that the choice of α equal to the integral part of $\frac{1}{2}n^r$, $\alpha = [\frac{1}{2}n^r]$, for some $0 < r < 1$, satisfies assumption (A3), and let $h = n^{-\theta} (\theta > 0)$ and $B_n = n^{\lambda}$ $(\lambda > 0)$. Also take $\delta_n = n^{-\gamma}$ $(\gamma > 0)$ and $\psi_n = n^{\delta}$ $(\delta > 0)$. Then

$$\sum_{n=1}^{\infty} \delta_n^{-p} \exp(-C\psi_n^{-2}nB_n^{-2}h^{2p}) < \infty \quad (C > 0),$$

the convergences in (3.12) are satisfied and $\{\psi_n/\alpha h^p\}$ is bounded away from 0, provided

$$r \leqslant \delta + \theta p, \qquad \delta < \min\{\theta, \lambda(s-1)\}, \qquad \delta + \lambda + \theta p < \min\{\tfrac{1}{2}, \gamma - \theta\}.$$

Remark 4.3. Under ϕ_2-mixing, assumptions (A2) and (A3) become

$$\phi_2^* = \sum_{j=1}^{\infty} \phi_2(j) < \infty \quad \text{and} \quad \limsup[1 + 2\,\mathrm{e}^{1/2}\phi_2(\alpha)]^{\mu} < \infty,$$

respectively. In such a setting, consider the problem of estimating $r(x)$ on the basis of X_j, $j = 1, \ldots, n+1$, coming from a strictly stationary Markov process satisfying condition (D_0) (see, for example, [5, pp. 192 and 221]). In this case, $\phi_2(n) = \gamma\rho^n$ for some $\gamma > 0$ and some $\rho \in (0, 1)$, and $\lim[1 + 2\,\mathrm{e}^{1/2}\phi_2(\alpha)]^{\mu} = 1$ for the choice of α suggested in the previous remark. Then $\sup_{x \in J}[\psi_n|r_n(x) - r(x)|] \to 0$ a.s. under the conditions of Theorem 4.1 on the quantities α, h, δ_n, B_n and ψ_n. This strengthens a result obtained in [15].

Addendum

While this paper was at a late stage of the refereeing process, the author became aware of a manuscript by Professor Lanh Tat Tran of Indiana University, Bloomington, entitled "Kernel density and regression estimation for dependent random variables, and time series" and dealing with the same problem. Under a different approach and set of assumptions, an improved rate of almost sure convergence was obtained. The paper is in the process of being refereed, and the results are too complicated to be summarized here in a meaningful way. The interested reader may request a preprint from its author.

Acknowledgement

Thanks are due to a referee (or referees) and an Associate Editor whose thoughtful comments led to an improvement of an earlier version of this paper.

References

[1] P.K. Bhattacharya, Estimation of probability density and its derivatives, Sankhyā Ser. A 29 (1967) 373-382.
[2] G. Collomb, Propriétés de convergence presque complète du prédicteur à noyau, Z. Wahrsch. Verw. Gebiete 66 (1984) 441-460.
[3] G. Collomb, Nonparametric time series analysis and prediction: uniform almost sure convergence of the window and $k-N$, N, autoregression estimates, Math. Oper. and Stat. Ser. Statistics 16 (1985) 297-307.
[4] G. Collomb and W. Härdle, Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations, Stochastic Process. Appl. 23 (1986) 77-89.
[5] J.L. Doob, Stochastic Processes (Wiley, New York, 1953).
[6] W. Härdle, Robust regression function estimation, J. Multivariate Anal. 14 (1984) 169-180.
[7] W. Härdle and S. Luckhaus, Uniform consistency of a class of regression estimators, Ann. Statist. 12 (1984) 612-623.
[8] W. Härdle and P.-D. Tuan, Some theory on M-smoothing of time series, J. Time Series Anal. 7 (1986) 191-204.
[9] I.A. Ibragimov and Yu.V. Linnik, Independent and stationary sequences of random variables (Wolters-Noordhoff, Groningen, The Netherlands, 1971).
[10] D. Ioannides and G.G. Roussas, Probability bounds for sums in triangular arrays of random variables under mixing conditions, in: K. Matusita, ed., Statistical Theory and Data Analysis II (North-Holland, Amsterdam, 1988) pp. 293-308.
[11] Y.P. Mack and B.W. Silverman, Weak and strong uniform consistency of kernel regression estimates, Z. Wahrsch. Verw. Gebiete 61 (1982) 405-415.
[12] P.M. Robinson, Nonparametric estimators for time series, J. Time Series Anal. 4 (1983) 185-207.
[13] P.M. Robinson, Robust nonparametric autoregression, in: J. Franke, W. Härdle, D. Martin, eds., Lecture Notes in Statistics, Vol. 26 (Springer, Berlin, 1984).
[14] 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, Part A (1986) 539-549.
[15] G.G. Roussas, Nonparametric estimation of the transition distribution function of a Markov process, Ann. Math. Statist. 40 (1969) 1386-1400.
[16] G.G. Roussas, Nonparametric estimation in mixing sequences of random variables, J. Statist. Plann. Inference 18 (1987) 135-149.
[17] G.G. Roussas, Nonparametric regression estimation under mixing conditions. Tech. Rept. No. 101, Division of Statistics, University of California (Davis, 1987).
[18] G.G. Roussas, A moment inequality of S_{n,k_n} for triangular arrays of random variables under mixing conditions, with applications, in: K. Matusita, ed., Statistical Theory and Data Analysis II (North-Holland, Amsterdam, 1988) pp. 273-292.
[19] G.G. Roussas and D. Ioannides, Moment inequalities for mixing sequences of random variables, Stochastic Anal. Appl. 5(1) (1987) 61-120.
[20] S. Yakowitz, Nearest-neighbour methods for time series analysis, J. Time Series Anal. 8 (1987) 235-247.