NONPARAMETRIC TIME SERIES PREDICTION.
I. KERNAL ESTIMATORS BASED ON LOCAL AVERAGES

by

Young K. Truong

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1832

July 1987

NONPARAMETRIC TIME SERIES PREDICTION.
I. KERNEL ESTIMATORS BASED ON LOCAL AVERAGES

By YOUNG K. TRUONG
Department of Biostatistics
University of North Carolina, Chapel Hill

Abstract. The local and global asymptotic properties of nonparametric estimators of conditional expectations constructed based on local averages are obtained for stationary time series which satisfies weak mixing conditions. For this class of nonparametric estimators, the results presented in this paper constitute a generalization of results on optimal rates of convergence (based on random samples) of Stone (1980, *Ann. Statist.* **8** 1348–1360) and Stone (1982, *Ann. Statist.* **10** 1040–1053) to time series. Futhermore, these results also generalize those of Collomb (1984, *Z. Wahrsch. verw. Gebiete* **66** 441–460) to unbounded time series under weaker mixing conditions.

Keywords. Kernel estimators, local averages, nonparametric regression, rates of convergence, cumulant, mixing.

(1973), Priestley (1979). While the study of non-linear models in time series is still in its early stages, what has been learned so far is sufficient to indicate that this is a very rich and potentially rewarding field. Analysis of particular series have shown that non-linear models can provide better fits to the data (as one would expect) and, more importantly, that the structure underlying the data can not be captured by linear models.

So far, the study of non-linear models has been restricted to a few specific forms. For example, Priestley (1980), Tong and Lim (1980), Nicholls and Quinn (1980), and Haggan and Ozaki (1980, 1981) consider various non-linear filters of, possibly independent, identically distributed Gaussian random variables. In practice it may be difficult to decide a priori, which, if any, of these models is best suited to a given set of data.

Asymptotic results for the conditional expectation has been established by Doukhan and Ghindes (1980), Collomb (1984), Bierens (1983), Robinson (1983) and Truong and Stone (1987b) under various mixing conditions. In Robinson (1983), pointwise consistency and a central limit theorem was obtained for kernel estimators based on local averages under the α-mixing condition. Collomb (1984) and Bierens (1983) considered the uniform consistency and rate of convergence for kernel estimators based on local averages under the ϕ-mixing condition, which is considerably stronger than the α-mixing condition. Collomb and Härdle (1984) considered the uniform rate of convergence (also under ϕ-mixing) for a class of robust nonparametric estimators that did not include local medians. Under the α-mixing, Truong and Stone (1987b) considered the pointwise and L^2 rates of convergence that did not cover the uniform rate.

The present approach is to assume those mixing conditions that have been widely used in Brillinger (1981). These conditions are as weak as the α-mixing condition, and the proofs based upon them are a lob) simpler. In fact, with these mixing conditions, the methods in Truong and Stone (1987b) can be modified to yield stronger results on uniform rates of convergence. These results, the pointwise and the global rates of convergence, for nonparametric estimators of conditional expectations constructed by kernel methods based on local averages are described in Section 3.

3. NONPARAMETRIC TIME SERIES PREDICTION

Results on the global rate of convergence of nonparametric estimators of conditional expectations based on a realization of a discrete time stationary time series will be treated in this section. Recall that d is the dimensionality of the explainatory variable $\mathbf{X}$ and let U denote a nonempty bounded open neighborhood of the origin of $\mathbf{R}^d$. Let $\{(\mathbf{X}_i, Y_i), i = 0, \pm 1, \ldots\}$ be an $(d+1)$ vector-valued strictly stationary series and set $\theta(\mathbf{x}) = E(Y_0|\mathbf{X}_0 = \mathbf{x})$. Let $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ denote a realization of this process.

Assumption 1. *There is a positve constant M_0 such that*

$$|\theta(\mathbf{x}) - \theta(\mathbf{x}')| \le M_0 \|\mathbf{x} - \mathbf{x}'\| \qquad \text{for } \mathbf{x}, \mathbf{x}' \in U,$$

where $\|\mathbf{x}\| = (x_1^2 + \cdots + x_d^2)^{1/2}$ for $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbf{R}^d$.

Assumption 2. *The distribution of $\mathbf{X}_0$ is absolutely continuous and its density $f(\cdot)$ is bounded away from zero and infinity. That is, there is a positive constant M_1 such that $M_1^{-1} \le f(\mathbf{x}) \le M_1$ for $\mathbf{x} \in U$.*

The following technical condition is required for bounding the variance of various terms in the proof. (See Lemmas 5–8.)

Assumption 3. *The conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_0$ is absolutely continuous and its density $h(\cdot|\mathbf{x})$ is bounded away from zero and infinity. That is $M_1^{-1} \le h(\mathbf{y}|\mathbf{x}) \le M_1$ for $\mathbf{x}$ and $\mathbf{y} \in U$.*

Collomb (1984) derived asymptotic properties for nonparametric estimators of conditional expectations based on bounded stationary time series. In order to extend the argument to include the unbounded time series, the following moment condition is required. See also Assumption 7.

Assumption 4. *There is a positive constant $\nu > 2$ such that*

$$\sup_{\mathbf{x} \in U} E(|Y_0|^\nu | \mathbf{X}_0 = \mathbf{x}) < \infty .$$

Let $(Z_1, \ldots, Z_k)$ denote a vector of k random variables with $E|Z_j|^k < \infty$, for $j = 1, \ldots, k$. The rth order joint cumulant, $\text{cum}(Z_1, \ldots, Z_k)$, of $(Z_1, \ldots, Z_k)$ is given by

$$\text{cum}(Z_1, \ldots, Z_k) = \sum (-1)^{p-1}(p-1)! E\left(\prod_{j \in \nu_1} Z_j\right) \cdots E\left(\prod_{j \in \nu_p} Z_j\right)$$

where the sum extends over all partitions $(\nu_1, \ldots, \nu_p)$, $p = 1, 2, \ldots, k$, of $(1, \ldots, k)$. Equivalently (Brillinger, 1981), $\text{cum}(Z_1, \ldots, Z_k)$ is given by the coefficient of $i^k t_1 \cdots t_k$ in the Taylor series expansion of $\log E[\exp(i \sum_1^k Z_j t_j)]$ about the origin. An important special case of this definition occurs when $Z_j = Z, j = 1, \ldots, k$. This then gives the cumulant of order k of a univriate random variable. See Brillinger (1981) for methods on computing cumulants.

A weak dependence condition on the stationary sequence will now be described. Let $f(\cdot)$ denote a real-valued, measurable functions on $\mathbf{R}^{d+1}$. Set

$$c(i, j) = \sup_{|f| \leq 1} \text{cum}\{f(\mathbf{X}_i, Y_i), f(\mathbf{X}_j, Y_j)\}.$$

and

$$c(t_1, t_2, \ldots, t_k) = \sup_{|f| \leq 1} \text{cum}\{f(\mathbf{X}_{t_1}, Y_{t_1}), f(\mathbf{X}_{t_2}, Y_{t_2}), \ldots, f(\mathbf{X}_{t_k}, Y_{t_k})\}.$$

If the series $(\mathbf{X}_t, Y_t)$, $t = 0, \pm 1, \ldots$ is strictly stationary, then

$$c(t_1, t_2, \ldots, t_k) = c(t_1 + u, t_2 + u, \ldots, t_k + u).$$

for $t_1, \ldots, t_k, u = 0, \pm 1, \ldots$. In this case, we sometimes use the asymmetric notation

$$c(t_1, t_2, \ldots, t_{k-1}) = c(t_1, t_2, \ldots, t_{k-1}, 0).$$

The stationary sequence is said to be mixing if $c(t_1, \ldots, t_k) \to 0$ as $t_1, \ldots, t_k \to \infty$. See Brillinger (1981).

Assumption 5. *$\sum_{j \geq N} c^{1-\nu/2}(j) = O(N^{-1})$ where ν is same as in Assumption 4.*

The following condition is similar to Assumption 2.6.3 of Brillinger (1981).

Assumption 6. $\sum_k C_k z^k/k! < \infty$ *for z in a neighborhood of zero, where*

$$C_k = \sum_{v_1=-\infty}^{\infty} \cdots \sum_{v_{k-1}=-\infty}^{\infty} |c(v_1, \ldots, v_{k-1})|.$$

In order to deal with unbounded series, another mixing condition is needed. (See Lemma 8 and Theorem 3.) Set

$$c^*(t_1, t_2, \ldots, t_k) = \sup_{|f|\leq 1} \text{cum}\{f(\mathbf{X}_{t_1})[Y_{t_1} - \theta(\mathbf{X}_{t_1})], \ldots, f(\mathbf{X}_{t_k})[Y_{t_k} - \theta(\mathbf{X}_{t_k})]\}.$$

Assumption 7. $\sum_k C_k^* z^k/k! < \infty$ *for z in a neighborhood of zero, where*

$$C_k^* = \sum_{v_1=-\infty}^{\infty} \cdots \sum_{v_{k-1}=-\infty}^{\infty} |c^*(v_1, \ldots, v_{k-1})|.$$

Given positive numbers a_n and b_n, $n \geq 1$, let $a_n \sim b_n$ mean that a_n/b_n is bounded away from zero and infinity. Given random variables V_n, $n \geq 1$, let $V_n = O_{pr}(b_n)$ mean that the random variables $b_n^{-1} V_n$, $n \geq 1$ are bounded in probability or, equivalently, that

$$\lim_{c\to\infty} \limsup_n P(|V_n| > cb_n) = 0.$$

The kernel estimators of conditional expectations and conditional medians will now be described. For each $n \geq 1$, let $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ be a realization of the (strictly) stationary time series and let δ_n denote a sequence of positive numbers such that tends to zero. Set $I_n(\mathbf{x}) = \{i : 1 \leq i \leq n \text{ and } \|\mathbf{X}_i - \mathbf{x}\| \leq \delta_n\}$ and $N_n(\mathbf{x}) = \#(I_n(\mathbf{x}))$. Also set $\hat{\theta}_n(\mathbf{x}) = N_n(\mathbf{x})^{-1} \sum_{I_n(\mathbf{x})} Y_i$ as an estimator of $\theta(\mathbf{x})$, which is a kernel smoohter based on local averges. Put $r = (2 + d)^{-1}$.

Theorem 1. *Suppose that Assumptions 1–5 hold and $\delta_n \sim n^r$. Then*

$$|\hat{\theta}_n(0) - \theta(0)| = O_{pr}(n^{-r}).$$

The proof of this theorem, which will be given in Section 5, is basically a refinement of the corresponding one given in Stone (1980), with additional arguments involving asymptotic independence (see Lemmas 1–8).

Let C be a fixed compact subset of U having a nonempty interior and let $g(\cdot)$ be a real-valued function on $\mathbf{R}^d$. Set

$$\|g\|_q = \left\{ \int_C |g(\mathbf{x})|^q \, d\mathbf{x} \right\}^{\frac{1}{q}}, \qquad 1 \leq q < \infty;$$
$$\|g\|_\infty = \sup_{\mathbf{x} \in C} |g(\mathbf{x})|.$$

Theorem 2. *Suppose that Assumptions 1–6 hold and $\delta_n \sim n^r$. Then there exists a $c > 0$ such that*

$$\lim_n P(\|\hat{\theta}_n - \theta\|_2 \geq c n^r) = 0\,.$$

The proof of this theorem will be given in Section 5. The argument is a refinement of the corresponding one for Theorem 1.

Theorem 3. *Suppose that Assumptions 1-7 hold and $\delta_n \sim (n^{-1}\log n)^r$. Then there exists a $c > 0$ such that*

$$\lim_n P\left(\|\hat{\theta}_n - \theta\|_\infty \geq c(n^{-1}\log n)^r \right) = 0\,.$$

The proof of this theorem will be given in Section 5.

4. DISCUSSION

For $n \geq 1$, let $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ be a random sample of size n from the distribution of $(\mathbf{X}, Y)$ and let k denote a non-negative integer. Let $\theta(\cdot)$ be the regression function of Y on $\mathbf{X}$ and suppose that $\theta(\cdot)$ has bounded $(k+1)$th derivative. Set $r = p/(2p+d)$ where $p = k + 1$. Stone (1980, 1982) showed that if $1 \leq q < \infty$, then n^{-r} is the optimal rate of convergence in both pointwise and L^q norms; while $(n^{-1}\log n)^{-r}$ is the optimal rate of convergence in L^∞ norm. To find an estimator of $\theta(\cdot)$ that achieves these optimal rates of concergence, given $\mathbf{x}$, let $\hat{P}_n(\cdot;\mathbf{x})$ be the polynomial on $\mathbf{R}^d$ of degree k that minimizes

$$\sum\nolimits_{I_n(\mathbf{x})} [\, Y_i - \hat{P}_n(\mathbf{X}_i;\mathbf{x}) \,]^2$$

and set $\hat{\theta}_n(\mathbf{x}) = \hat{P}_n(\mathbf{x};\mathbf{x})$ (if $q = \infty$, define $\hat{\theta}_n$ as above over a finite subset of C and then extend it to all of C by suitable interpolation). Note that this estimator can be easily obtained by solving the corresponding normal equation. The corresponding generalization to time series is immediate and it is an interesting open question whether the asymptotic properties just described still hold for $p > 1$ in this context.

To robustify the above procedure, let $\hat{P}_n(\cdot;\mathbf{x})$ be a polynomial on $\mathbf{R}^d$ of degree k which minimizes

$$\sum\nolimits_{I_n(\mathbf{x})} \rho(Y_i - \hat{P}_n(\mathbf{X}_i;\mathbf{x}))$$

and set $\hat{\theta}_n(\mathbf{x}) = \hat{P}_n(\mathbf{x};\mathbf{x})$. As a special case, Truong and Stone (1987a, 1987b) consider the asymptotic properties of kernel estimators based on local medians in the context of random sample and time series. The problem on L^∞ rate of convergence for stationary time series satisfying weak α-mixing condition remains open.

One drawback that the nonparametric approach has is the high *dimensionality*, which can be thought of in terms of the *variance* in estimation. In other words: A *huge* data set may be required for nonparametric estimation of a function of many variables; otherwise the variance of the estimator may be unacceptably large. This drawback is serious especially in time series analysis where the future usually depends on much of the past.

A possible solution would be to use *additivity* as in Stone (1985) to alleviate *curse of dimensionality*. More formally, let $\theta(\cdot)$ be the regression function defined on $\mathbf{R}^d$ and suppose that θ is additive; that is, that there is smooth functions $\theta_1(\cdot), \ldots, \theta_d(\cdot)$ defined on $\mathbf{R}^1$ such that

$$\theta(x_1, \ldots, x_d) = \mu + \theta_1(x_1) + \cdots + \theta_d(x_d)\,,$$

where $\mu = E(Y)$. Using B-splines, an estimator of $\theta(\cdot)$ can be constructed to achieve the optimal rates of convergence n^{-r}, where r now is equal to $p/(2p+1)$. The rates of convergence here do not depend on the dimensional parameter d. Another nice feature about this estimator is that it is smoother and is as flexible as ordinary nonparametric procedures constructed by the kernel method.

The corresponding methodology is generalized immediately to time series, and it is an interesting open problem to determine whether the asymptotic properties described above (with r independent of d) also hold in this context.

5. PROOF OF THEOREMS

Let f and g be real-valued, measurable functions defined on $\mathbf{R}^{d+1}$. Set $U = f(\mathbf{X}_i, Y_i)$ and $V = g(\mathbf{X}_j, Y_j)$. Put $c \equiv c(|i - j|)$. The following lemma follows immediately from the definition.

Lemma 1. *Suppose that* $|f(\cdot)| < C_1$ *and* $|g(\cdot)| < C_2$. *Then*

$$|E(UV) - E(U)E(V)| \le C_1 C_2 c.$$

The following two lemmas are identical to that under the α-mixing condition. See Hall and Heyde (1980).

Lemma 2. *Suppose that* $E|U|^p < \infty$ *for some* $p > 1$ *and* $|g(\cdot)| < C$. *Then*

$$|E(UV) - E(U)E(V)| \le 3C\|U\|_p c^{1-\frac{1}{p}}.$$

Proof. Suppose $c > 0$, since otherwise U and V are independent and the inequality is trivial. Set $C_1 = c^{-1/p}\|U\|_p$, $U_1 = U 1_{\{|U| \le C_1\}}$ and $U_2 = U - U_1$.

$$\begin{aligned}|E(UV) - E(U)E(V)| &\le |E(U_1V) - E(U_1)E(V)| + |E(U_2V) - E(U_2)E(V)| \\ &\le CC_1c + 2CE|U_2|. \qquad (5.1)\end{aligned}$$

But

$$E|U_2| = E[\,|U|1_{\{|U|>C_1\}} \le C_1^{-p+1}E|U|^p. \qquad (5.2)$$

The Lemma now follows from (5.1) and (5.2).

Lemma 3. *Suppose that* $E|U|^p < \infty$, $E|V|^q < \infty$ *where* $p, q > 1$ *and* $p^{-1} + q^{-1} < 1$. *Then*

$$|E(UV) - E(U)E(V)| \le 5\|U\|_p\|V\|_q c^{1-p^{-1}-q^{-1}}.$$

Proof. Suppose $c > 0$. Set $C = c^{-1/q}\|V\|_q$, $V_1 = V1_{\{|V| \le C\}}$ and $V_2 = V - V_1$. Then using Lemma 2 and Hölder's inequality,

$$\begin{aligned}|E(UV) - E(U)E(V)| &\le |E(UV_1) - E(U)E(V_1)| + |E(UV_2) - E(U)E(V_2)| \\ &\le 3C\|U\|_p c^{1-p^{-1}} + 2\|U\|_p\|V_2\|_p(p-1)^{-1}, \qquad (5.3)\end{aligned}$$

while

$$E|V_2|^{p(p-1)^{-1}} = E[\,|V|^{p(p-1)^{-1}} 1_{\{|V|>C\}}\,] \le C^{-q+p(p-1)^{-1}} E|Y|^q. \tag{5.4}$$

The lemma follows from (5.3) and (5.4).

For each $i = 1, \ldots, n$, set $K_i = 1_{\{\|\mathbf{X}_i\| \le \delta_n\}}$. The following lemma is an immediate consequence of Assumptions 2 and 3.

Lemma 4. *There is a positive constant C_1 such that*

$$E(K_i K_{i+j}) \le C_1 \delta_n^{2d}.$$

Lemma 5. $\mathrm{Var}(\sum_i K_i) = O(n\delta_n^d)$.

Proof. By Lemma 1, $|\mathrm{Cov}(K_i, K_{i+j})| \le c(j)$. Thus by Assumption 5 and Lemma 4,

$$\begin{aligned}\mathrm{Var}(\textstyle\sum_i K_i) &= n\mathrm{Var}(K_1) + 2\textstyle\sum_i\sum_j \mathrm{Cov}(K_i, K_{i+j}) \\ &= O\left(n\delta_n^d + n\textstyle\sum_1^n \min\left(c(j), \delta_n^{2d}\right)\right) = O(n\delta_n^d).\end{aligned}$$

The following result follows from Tchebychev's inequality, Lemma 5 and Assumption 2.

Lemma 6. *There is a positive constant k_1 such that*

$$\lim_n P(\textstyle\sum_i K_i \le k_1 n \delta_n^d) = 0.$$

Lemma 7. $\mathrm{Var}(\sum_i K_i[\,Y_i - \theta(\mathbf{X}_i)\,]) = O(n\delta_n^d)$.

Proof. (Robinson, 1983) Let B be a positive constant and set

$$Y_i' = Y_i 1_{\{|Y_i| \le B\}};\ \ Y_i'' = Y_i 1_{\{|Y_i| \ge B\}}.$$

$$\theta'(\mathbf{X}_i) = E[\,Y_i' \,|\, \mathbf{X}_i\,],\ \ \theta''(\mathbf{X}_i) = E[\,Y_i'' \,|\, \mathbf{X}_i\,].$$

Then $Y_i = Y_i' + Y_i''$ and $\theta(\mathbf{X}_i) = \theta'(\mathbf{X}_i) + \theta''(\mathbf{X}_i)$.

Set $Z_i = Y_i' - \theta'(\mathbf{X}_i)$. Observe that $|Z_i| \le 2B$ and $E(Z_i|\mathbf{X}_i) = 0$. By the argument used in the proof of Lemma 5,

$$\begin{aligned}\mathrm{Var}(\textstyle\sum_i K_i Z_i) &= n\mathrm{Var}(K_1 Z_1) + 2\textstyle\sum_i \sum_j \mathrm{Cov}(K_i Z_i, K_{i+j} Z_{i+j}) \\ &= O\left(n\delta_n^d + n\textstyle\sum_1^n \min(\alpha(j), \delta_n^{2d})\right) = O(n\delta_n^d). \qquad (5.5)\end{aligned}$$

Set $W_i = Y_i'' - \theta''(\mathbf{X}_i)$. Applying Holder's inequality twice,

$$\begin{aligned}&E(K_i|W_i|K_{i+j}|W_{i+j}|) \\ &= E\left[(K_i|W_i|^\nu)^{\frac{1}{\nu}}(K_{i+j}|W_{i+j}|^\nu)^{\frac{1}{\nu}}(K_i K_{i+j})^{1-\frac{2}{\nu}} K_i^{\frac{1}{\nu}} K_{i+j}^{\frac{1}{\nu}}\right] \\ &\le \{E[K_i|W_i|^\nu]\}^{\frac{2}{\nu}}\{E[K_i K_{i+j}]\}^{1-\frac{2}{\nu}}. \qquad (5.6)\end{aligned}$$

By Lemma 3,

$$E(K_i|W_i|K_{i+j}|W_{i+j}|) \le 5\{E(K_i|W_i|^\nu)\}^{\frac{2}{\nu}}\{c(j)\}^{1-\frac{2}{\nu}}. \qquad (5.7)$$

According to Assumption 2,

$$\begin{aligned}E(K_i|W_i|^s) &= E(K_i E(|W_i|^s|K_i)) \\ &\le M_1 \sup_{\|\mathbf{y}\|\le\delta_n} Q(\mathbf{y}) \int K_i(\mathbf{x})\,d\mathbf{x} = O(\delta_n^d) \text{ for } 1 \le s \le \nu, \qquad (5.8)\end{aligned}$$

where $Q(\mathbf{y}) = E(|W_i|^s|\mathbf{X}_i = \mathbf{y})$ is bounded in $\mathbf{y} \in U$ by Assumption 4. By (5.6)–(5.8), Lemma 4 and Assumption 5(i) (note that $E(W_i|\mathbf{X}_i) = 0$),

$$\begin{aligned}\mathrm{Var}(\textstyle\sum_i K_i W_i) &= n\mathrm{Var}(K_1 W_1) + 2\textstyle\sum_i \sum_j \mathrm{Cov}(K_i W_i, K_{i+j} W_{i+j}) \\ &= O\left(n\delta_n^d + n(\delta_n^d)^{\frac{2}{\nu}}\textstyle\sum_1^n \min\left\{c^{1-\frac{2}{\nu}}(j), (\delta_n^{2d})^{1-\frac{2}{\nu}}\right\}\right) = O(n\delta_n^d). \qquad (5.9)\end{aligned}$$

It follows from (5.5) and (5.9) that

$$\begin{aligned}\mathrm{Var}(\textstyle\sum_i K_i|Y_i - \theta(\mathbf{X}_i)|) &\le 2\{\mathrm{Var}(\textstyle\sum_i K_i Z_i) + \mathrm{Var}(\textstyle\sum_i K_i W_i)\} \\ &= O(n\delta_n^d),\end{aligned}$$

which completes the proof of Lemma 7.

Lemma 8. *Set* $Z_i = K_i[Y_i - \theta(\mathbf{X}_i)]$, $i = 1, \ldots, n$ *and* $\sigma_n^2 = \mathrm{Var}\sum_i Z_i$. *Then there are positive constants* A_1 *and* A_2 *such that*

$$P(\sum_i Z_i \geq cn\delta_n^{d+1}) \leq A_1 \exp(-c^2 A_2 n\delta_n^{d+2}).$$

Proof. The cumulant of $\sum_i Z_i$ has the form

$$\sum_{t_1} \cdots \sum_{t_k} c^*(t_1 - t_k, \ldots, t_{k-1} - t_k) \leq nC_k^*,$$

where C_k^* is defined in Assumption 7. Therefore (since Z_i has mean zero)

$$|\log E(\alpha \exp \textstyle\sum_i Z_i) - \alpha^2\sigma_n^2/2| \leq n \displaystyle\sum_3^\infty C_k^*|\alpha|^k/k!$$

By Assumption 7 and Lemma 7, there are positive constants A_1 and A_3 such that for α sufficiently small,

$$E(\alpha \exp \textstyle\sum_i Z_i) \leq A_1 \exp(\alpha^2\sigma_n^2/2) \leq A_1 \exp(\alpha^2 A_3 n\delta_n^d/2).$$

By Markov's inequality and put $\alpha = c\delta_n/A_3$, there is a positive constant A_2 such that

$$\begin{aligned} P(\textstyle\sum_i Z_i \geq cn\delta_n^{d+1}) &\leq \exp(-\alpha cn\delta_n^{d+1}) E(\alpha \exp \textstyle\sum_i Z_i) \\ &\leq A_1 \exp(-\alpha cn\delta_n^{d+1} + \alpha^2 A_3 n\delta_n^d/2) \\ &\leq A_1 \exp(-c^2 A_2 n\delta_n^{d+2}). \end{aligned}$$

This completes the proof of Lemma 8.

For the following lemma, let A_4 denote a positive constant and let V_i be a bounded real-valued, measurable function of $(\mathbf{X}_i, Y_i)$ such that $E(V_i) = \mu$ and $\mathrm{Var}(\sum_i V_i) \leq A_4 n\mu$, $i = 1, \ldots, n$.

Lemma 9. *Suppose Assumption 6 holds. Then there are positive constants* A_5 *and* A_6 *such that*

$$P(\textstyle\sum_i V_i \leq \frac{1}{2}n\mu) \leq A_5 \exp(-A_6 n\mu).$$

Proof. The cumulant of $\sum_i V_i$ has the form

$$\sum_{t_1} \cdots \sum_{t_k} c(t_1 - t_k, \ldots, t_{k-1} - t_k) \leq nC_k,$$

where C_k is defined in Assumption 6. Therefore

$$|\log E(\alpha \exp \textstyle\sum_i V_i) - \alpha^2 \mathrm{Var}(\sum_i V_i)/2| \leq n \displaystyle\sum_3^\infty C_k |\alpha|^k / k!$$

By Assumption 6, there is a positive constant A_5 such that for α sufficiently small,

$$E(\alpha \exp \textstyle\sum_i V_i) \leq A_5 \exp(\alpha^2 \mathrm{Var}(\sum_i V_i)/2) \leq A_5 \exp(\alpha^2 A_4 n\mu/2).$$

The desired result follows from Markov's inquality.

Lemma 10. *There is a positive constant A_7 such that*

$$\lim_n P(N_n(\mathbf{x}) \geq A_7 n\delta_n^d \text{ for } \mathbf{x} \in C) = 1.$$

Proof. Given positive numbers a_n and b_n, $n \geq 1$, let $a_n \simeq b_n$ mean that $a_n/b_n \to 1$ as $n \to \infty$. We assume $C = [-\frac{1}{2}, \frac{1}{2}]^d$. Write C as the disjoint union of M_n^d cubes $C_{n\alpha}$ of length $\simeq \delta_n/d$ where $M_n \simeq d\delta_n^{-1}$ and $\alpha = 1, \ldots, M_n^d$. Set $K_{i\alpha} = 1_{\{\mathbf{X}_i \in C_{n\alpha}\}}$, $\mu = \mu_\alpha = E(K_{i\alpha}) \sim \delta_n^d$ and $N_\alpha = \#\{i : 1 \leq i \leq n; \mathbf{X}_i \in C_{n\alpha}\} = \sum_i K_{i\alpha}$. To prove the lemma, it suffices to show that

$$\lim_n P(N_\alpha \geq \tfrac{1}{2} n\delta_n^d \text{ for } \alpha = 1, \ldots, M_n^d) = 1.$$

Indeed, according to Lemma 5, $\mathrm{Var}(N_\alpha) = O(n\delta_n^d)$. Thus, by Lemma 9, there are positive constants A_8 and A_9 such that

$$P(N_\alpha \leq \tfrac{1}{2} n\mu) \leq A_8 \exp(-A_9 n\delta_n^d).$$

Hence

$$\begin{aligned} P(N_\alpha \geq \tfrac{1}{2} n\mu \text{ for } \alpha = 1, \ldots, M_n^d) &\geq 1 - M_n^d \max_\alpha P(N_\alpha \leq \tfrac{1}{2} n\mu) \\ &\geq 1 - O(\delta_n^{-d}) \exp(-A_9 n\delta_n^d) \to 1 \quad \text{as } n \to \infty, \end{aligned}$$

since $M_n^d \sim \delta_n^d$ and $n\mu \sim n\delta_n^d \sim \delta_n^{-2}$ (or $\sim \delta_n^{-2}\log n$). This completes the proof of Lemma 10.

Proof of Theorem 1. According to Assumption 1

$$|\theta(\mathbf{X}_i) - \theta(\mathbf{0})| \le M_0\delta_n \qquad \text{for } i \in I_n.$$

Thus

$$\left|N_n^{-1}\textstyle\sum_{I_n}[\,\theta(\mathbf{X}_i) - \theta(\mathbf{0})\,]\right| = O_{pr}(n^{-r}). \tag{5.10}$$

On the other hand,

$$\begin{aligned}
&P(N_n^{-1}|\textstyle\sum_{I_n}[\,Y_i - \theta(\mathbf{X}_i)]| \ge cn^{-r}) \\
&\le P(N_n^{-1}|\textstyle\sum_{I_n}[\,Y_i - \theta(\mathbf{X}_i)]| \ge cn^{-r}; N_n > k_1 n\delta_n^d) + P(N_n \le k_1 n\delta_n^d) \\
&\le P(|\textstyle\sum_{I_n}[\,Y_i - \theta(\mathbf{X}_i)]| \ge k_1 cn^{-r}n\delta_n^d) + P(N_n \le k_1 n\delta_n^d).
\end{aligned}$$

Hence, by Lemma 3, Lemma 4 and Tchebychev's inequality,

$$\left|N_n^{-1}\textstyle\sum_{I_n}[\,Y_i - \theta(\mathbf{X}_i)\,]\right| = O_{pr}(n^{-r}). \tag{5.11}$$

The conclusion of Theorem 1 follows from (5.10) and (5.11).

Proof of Theorem 2. We may assume that C is contained in the interior of the cube $C_0 = [\,-\frac{1}{2}, \frac{1}{2}\,]^d \subset U$. According to Assumption 1, there is a positive constant k_1 such that $|\theta(\mathbf{X}_i) - \theta(\mathbf{x})| \le k_1\|\mathbf{X}_i - \mathbf{x}\| \le k_1\delta_n$, for $i \in I_n(\mathbf{x})$ and $\mathbf{x} \in C$. Thus there is a positive constant k_2 such that

$$\lim_n P\left(\left|N_n(\mathbf{x})^{-1}\textstyle\sum_{I_n(\mathbf{x})}[\,\theta(\mathbf{X}_i) - \theta(\mathbf{x})\,]\right| \ge k_2\delta_n \ \text{ for some } \mathbf{x} \in C\right) = 0. \tag{5.12}$$

Set $Z_n(\mathbf{x}) = \sum_{i\in I_n(\mathbf{x})}[\,Y_i - \theta(\mathbf{X}_i)\,]$. By Assumption 4 and Lemma 7,

$$E[\,Z_n^2(\mathbf{x})\,] = O(n\delta_n^d) \qquad \text{uniformly over } \mathbf{x} \in C.$$

Consequently,

$$E\left[\int_C |Z_n(\mathbf{x})|^2\,d\mathbf{x}\right] = \int_C E[\,|Z_n(\mathbf{x})|^2\,]\,d\mathbf{x} = O(n\delta_n^d). \tag{5.13}$$

By Lemma 10, there is a positive constant k_3 such that

$$\lim_n P(\Omega_n) = 1, \tag{5.14}$$

where $\Omega_n = \{N_n(\mathbf{x}) \geq k_3 n\delta_n^d \text{ for } \mathbf{x} \in C\}$. By (5.13) and (5.14),

$$\begin{aligned}
&P\left(\left\{\int_C \left|N_n(\mathbf{x})^{-1}\textstyle\sum_{I_n(\mathbf{x})}[Y_i - \theta(\mathbf{X}_i)]\right|^2 d\mathbf{x}\right\}^{\frac{1}{2}} \geq c(n^{-1}\delta_n^{-d})^{\frac{1}{2}}\right) \\
&\leq P(\Omega_n^c) + P\left(\int_C |Z_n(\mathbf{x})|^2 \, d\mathbf{x} \geq c^2 k_3^2 n\delta_n^d\right) \\
&= P(\Omega_n^c) + \frac{O(1)n\delta_n^d}{c^2 n\delta_n^d} = o(1) \qquad \text{as } n, c \to \infty.
\end{aligned} \tag{5.15}$$

It follows from (5.12) and (5.15) that

$$\lim_{c\to\infty} \lim_n P\left(\|\hat{\theta}_n - \theta\|_2 \geq c\left(\delta_n + (n^{-1}\delta_n^{-d})^{\frac{1}{2}}\right)\right) = 0.$$

The Conclusion of Theorem 2 now follows by choosing δ_n so that $\delta_n = (n^{-1}\delta_n^{-d})^{\frac{1}{2}}$, or equivalently, $\delta_n = n^{-r}$.

Proof of Theorem 3. By Lemma 8–10,

$$\begin{aligned}
P(N_n(\mathbf{x})^{-1}\textstyle\sum_{I_n(\mathbf{x})}[Y_i - \theta(\mathbf{X}_i)] \geq c\delta_n) &\leq P(\Omega_n^c) + P\left(|Z_n(\mathbf{x})| \geq ck_3 n\delta_n^{d+1}\right) \\
&\leq O(\delta_n^{-d}) \exp(-A_9 n\delta_n^d) + A_1 \exp(-c^2 A_2 \log n).
\end{aligned}$$

Let C_n be a finite subset of C such that $\#(C_n) \leq n^{k_4}$ for some fixed positive constant k_4. It follows from $n\delta_n^{d+2} \sim \log n$ that there is a positive constant c such that

$$P(\max_{C_n} |N_n(\mathbf{x})^{-1}\textstyle\sum_{I_n(\mathbf{x})}[Y_i - \theta(\mathbf{X}_i)]| \geq c\delta_n) \leq 2n^{k_4} \exp(-c^2 \log n).$$

Consequently,

$$\lim_n P(\max_{C_n} |N_n(\mathbf{x})^{-1}\textstyle\sum_{I_n(\mathbf{x})}[Y_i - \theta(\mathbf{X}_i)]| \geq c\delta_n) = 0. \tag{5.16}$$

It follows from (5.12) and (5.16) that there exists a $c > 0$ such that

$$\lim_n P(\max_{C_n} |\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{X})| \geq c\delta_n) = 0. \tag{5.17}$$

Set $L_n = [n^{2r}]$. Let C_n be the collection of $(2L_n + 1)^d$ points in C each of whose coordinates is of the form $j/(2L_n)$ for some integer j such that $|j| \leq L_n$. Then C can be written as the union of $(2L_n)^d$ subcubes, each having length $(2L_n)^{-1}$ and all of its vertices in C_n. For each $\mathbf{x} \in C$ there is a subcube Q_w with center w such that $\mathbf{x} \in Q_w$. And $\mathbf{x}$ can be written as a convex combination $\sum \lambda_v v$ of the vertices of one of these subcubes. Set $\bar{\theta}(\mathbf{x}) = \sum \lambda_v \theta(v)$. Then there is a positive constant k_5 such that

$$\max_{\mathbf{x}\in C} |\bar{\theta}(\mathbf{x}) - \theta(\mathbf{x})| \leq \frac{k_5}{L_n} = o((n^{-1}\log n)^r).$$

Set $\bar{\theta}_n(\mathbf{x}) = \sum \lambda_v \hat{\theta}_n(v)$, where $\mathbf{x} = \sum \lambda_v v$. Then

$$\begin{aligned} |\bar{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| &\leq \sum \lambda_v |\hat{\theta}_n(v) - \theta(v)| + |\bar{\theta}(\mathbf{x}) - \theta(\mathbf{x})| \\ &\leq \max_{C_n} |\hat{\theta}_n(\mathbf{x}) - \theta(\mathbf{x})| + o((n^{-1}\log n)^r), \end{aligned}$$

so by (5.17)

$$\lim_n P(\|\hat{\theta}_n - \theta\|_\infty \geq c\delta_n) = 0.$$

This completes the proof of Theorem 3.

REFERENCES

AKAIKE, H. (1974a) Markov representation of stochastic processes and its application to the analysis of the autoregressive moving-average process. *Ann. Inst. Statist. Math.* **26**, 263–387.

AKAIKE, H. (1974b) Stochastic theory of minimal realisation. *I.E.E.E. Trans. Automatic Control AC*-**19**, 263–387.

BIERENS, H. J. (1983) Uniform consistency of kernel estimators of a regression function under generalized conditions. *J. Amer. Statist. Assoc.* **78**, 699–707.

BLOOMFIELD, P. and STEIGER, (1983) *Least Absolute Deviations.* Birkhauser, Boston.

BRILLINGER, D. R. (1981) *Time series: Data analysis and theory.* Holden-Day, San Francisco.

COLLOMB, B. G. (1984) Propriétés de convergence presque complète du prédicteur à noyau *Z. Wahrsch. verw. Gebiete* **66**, 441–460.

COLLOMB, B. G. and HÄRDLE, W. (1984) Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations. *Preprint.*

DOUKHAN, P. and GHINDES, M. (1980) Estimations dans le processus $X_{n+1} = f(X_n) + \epsilon_n$. *C. R. de l'Academie des Sciences, Paris.* **291**, 61–64.

DUNSMUIR, W. and HANNAN, E. J. (1976) Vector linear time series models. *Adv. Appl. Prob.* **8**, 339–364.

HAGGAN, V. and OZAKI, T. (1980) Amplitude-dependent exponential AR model fitting for non-linear random vibrations. In *Time Series*, edited by O. D. Anderson. North-Holland: Amsterdam.

HAGGAN, V. and OZAKI, T. (1981) Modelling non-linear random vibrations using an amplitude-dependent autoregressive model. *Biometrika* **68**, 189–196.

HALL, P. and HEYDE, C. C. (1980) *Martingale limit theory and its applications.* Academic Press, New York.

HANNAN, E. J. (1973) The asymptotic theory of linear time series models. *J. Appl. Prob.* **10**, 130–145.

NICHOLLS, D. F. and QUINN, B. G. (1980) The estimation of random coefficient autoregressive models. I. *J. Time Series Anal.* **1**, 37–46.

PRIESTLEY, M. B. (1979) *Time series and spectral analysis.* Academic Press, New York.

PRIESTLEY, M. B. (1980) State-dependent models: A general approach to non-linear time series analysis. *J. Time Series Anal.* **1**, 47–71.

ROBINSON, P. M. (1983) Nonparametric estimators for time series. *J. Time Series Anal.* **4**, 185–207.

STONE, C. J. (1980) Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348–1360.

STONE, C. J. (1982) Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10**, 1040–1053.

STONE, C. J. (1985) Additive regression and other nonparametric models. *Ann. Statist.* **13**, 689–705.

TONG, H. and LIM, K. S. (1980) Threshold Autoregression, limit cycles and cyclical data. *J. Roy. Statist. Soc. B* **42**, 245–292.

TRUONG, K. Y. and STONE, C. S. (1987a) Asymptotic properties of nonparametric time series prediction Institute of Statistics Mimeo Series No. 1830.

TRUONG, K. Y. and STONE, C. S. (1987b) Asymptotic properties of kernel estimators based on local medians. Institute of Statistics Mimeo Series No. 1831.