The Annals of Statistics
1990, Vol. 18, No. 2, 891–906

ASYMPTOTIC DISTRIBUTION OF ROBUST ESTIMATORS FOR NONPARAMETRIC MODELS FROM MIXING PROCESSES

By Graciela Boente and Ricardo Fraiman

Universidad de Buenos Aires and Conicet and Universidad de la República and Universidad de Buenos Aires

In this paper we obtain the asymptotic distribution of robust nonparametric autoregression estimators for dependent observations. Weights based on kernel and nearest-neighbor methods are considered.

1. Introduction. Let $\{(X_t, Y_t): t \geq p+1\}$ be a strictly stationary process, $X_t \in \mathbb{R}^p$, $Y_t \in \mathbb{R}$. We consider nonparametric estimators of the regression function $\Phi(x) = E(Y_t|X_t = x)$. Since we can take $X_t = (Y_{t-1}, \dots, Y_{t-p})$, this includes estimation of the predictor function.

In recent years several dependence conditions for stochastic processes have been used in order to study the asymptotic behavior of nonparametric estimates of the predictor function. More precisely, certain mixing conditions have been considered. Roughly speaking, all that these mixing conditions say is that the dependence between the random variables is weaker the farther they are apart. The first paper on this subject was by Rosenblatt (1956) who introduced the notion of strong mixing processes: Let $\{Z_j: j \geq 1\}$ be a stochastic process and denote by $M_{a,b}$ the σ-algebra generated by the random variables $\{Z_t: a \leq t \leq b\}$, $1 \leq a \leq \infty$. The process is said to satisfy a strong mixing or α-mixing condition if there exists a sequence $\alpha(n)$ of positive numbers such that $\lim_{n\to\infty}\alpha(n) = 0$ and for any $A \in M_{1,t}$, $B \in M_{t+n,+\infty}$ we have

$$|P(A \cap B) - P(A)P(B)| \leq \alpha(n).$$

However, the more often studied mixing condition is a stronger one, the φ-mixing condition [Billingsley (1968)]. The sequence is said to be φ-mixing or uniform strongly mixing if there exist coefficients $\varphi(n)$ such that $\lim_{n\to\infty}\varphi(n) = 0$ and the inequality

$$|P(A \cap B) - P(A)P(B)| \leq \varphi(n)P(A)$$

holds for any $A \in M_{1,t}$, $B \in M_{t+n,+\infty}$.

This condition is considerably stronger than the α-mixing condition. In particular, for a Gaussian stationary process, the φ-mixing condition is equivalent to m-dependence [i.e., there exists $m \in N$ such that $\varphi(n) = 0$ for all $n > m$] as proved in Theorem 17.3.2 of Ibragimov and Linnik (1971), while α-mixing processes generally include a pth-order autoregressive model.

Received October 1987; revised May 1989.

AMS 1980 *subject classifications*. Primary 62F35; secondary 62G05.

Key words and phrases. Robust nonparametric autoregression, M-estimators, kernel estimates, nearest-neighbor methods, α-mixing processes.

Another dependence measure, weaker than the φ-mixing, is the ρ-mixing condition which is based on the maximal correlation and is defined as follows. The process $\{Z_t: t \geq 1\}$ is said to be ρ-mixing if

$$|P(A \cap B) - P(A)P(B)| \leq \rho(n)[P(A)P(B)]^{1/2}$$

holds for any $A \in M_{1,t}$, $B \in M_{t+n,+\infty}$ and some sequence $\{\rho(n): n \geq 1\}$ decreasing to 0.

A first reference to nonparametric methods in time-series analysis is Watson (1964), who applied kernel methods to a meteorological prediction problem. The kernel estimate of $\Phi(x)$ is defined by

$$\Phi_T(x) = \sum_{t=p+1}^{T} w_{tT}(x)Y_t, \tag{1.1}$$

where

$$w_{tT}(x) = K((X_t - x)/h_T) \Big/ \sum_{\tau=p+1}^{T} K((X_\tau - x)/h_T),$$

K is a nonnegative integrable function on $\mathbb{R}^p$ and $h_T > 0$. Asymptotic properties of such estimators and predictors were obtained by Roussas (1969), Bosq (1980), Doukhan and Ghindès (1980, 1983), Collomb (1982, 1984), Robinson (1983), Yakowitz (1985) and Doukhan, León and Portal (1985). Collomb (1984) considered uniform consistency for kernel estimators under the φ-mixing condition. Roussas (1969) Doukhan and Ghindès (1980) and Yakowitz (1985) obtained pointwise consistency for Markov processes satisfying the G2 condition which is basically a φ-mixing condition. These consistency results where extended to less restrictive dependence structures by Peligrad (1988) who assumed a ρ-mixing condition and also required for φ-mixing processes weaker assumptions on the bandwidth selection. Truong and Stone (1988) obtained the optimal rates of convergence in probability for α-mixing processes and Roussas (1988) studied strong consistency results for all kinds of mixing conditions. Truong and Stone (1988) and also Roussas (1988) eliminated the condition of boundedness on the response variables. This condition was also relaxed to a moment condition by Sarda and Vieu (1986, 1988) for φ-mixing processes. Robinson (1983) obtained weak consistency and asymptotic normality of the kernel estimate for α-mixing processes. See also Bradley (1983) for some normality results.

Collomb (1980, 1985) extended nearest-neighbor methods from density estimation to nonparametric regression in the i.i.d. case and for φ-mixing processes, respectively. These estimators are defined through

$$\hat{\Phi}_T(x) = \sum_{t=p+1}^{T} W_{tT}(x)Y_t, \tag{1.2}$$

where

$$W_{tT}(x) = K((X_t - x)/H_T) \Big/ \sum_{\tau=p+1}^{T} K((X_\tau - x)/H_T),$$

H_T is the distance between x and its k-nearest neighbor among $X_{p+1}, \dots, X_T$, $k = k_T$ is a fixed integer and K is as in (1.1). Mack (1981) studied the asymptotic distribution of these estimates for the i.i.d. case.

Both methods are highly sensitive to the effect of just one isolated disparate observation Y_t, particularly if X_t is close to x, since they are weighted averages of the observations. In order to obtain robust nonparametric estimates M- and R-type methods have been considered.

In the i.i.d. case Tsybakov (1983) and Härdle (1984) studied pointwise asymptotic properties of a M-type version of the Nadaraya–Watson method when scale is known. Later on, Härdle and Tsybakov (1988) extended their previous results to scale equivariant kernel estimates obtaining a central limit theorem for simultaneous regression and scale estimation. See also Boente and Fraiman (1989b). R-type kernel and nearest-neighbor with kernel estimates have been considered by Cheng and Cheng (1987), where by R-type estimates we mean robust estimators obtained from rank tests [see for instance Huber (1981)].

M-estimators with kernel weights were adapted to time-series models by Robinson (1984) who established a central limit theorem when scale is known assuming an α-mixing dependence structure. A similar approach was considered by Collomb and Härdle (1984), who obtained uniform convergence of this family of estimates for φ-mixing processes.

In Boente and Fraiman (1989a) strong consistency of robust scale equivariant estimators for nonparametric regression models based on kernel and nearest-neighbor methods was obtained for φ- and α-mixing processes.

In this paper we study the asymptotic distribution of both families of estimates under some regularity conditions, for α-mixing processes. Our results for kernel weights are closely related to those of Robinson (1984). We consider the case when scale is unknown by using a consistent scale estimator. The asymptotic distribution for the robust estimates based on kernel weights is also used as an auxiliary result in order to obtain the asymptotic normality of the robust nearest-neighbor estimates.

Let ψ be a real function and (X, Y) be a random vector with the same distribution as (X_{p+1}, Y_{p+1}). Denote by $F(y|X = x)$ a regular version of the conditional distribution function of $Y|X = x$ and define $g(x)$ as the solution of

$$\int \psi((y - g(x))/s(x)) \, dF(y|X = x) = 0, \tag{1.3}$$

where $s(x)$ is any robust scale measure, for example, $s(x) = MAD_C(x) = \text{med}(|Y - m(x)| \, |X = x)$, where $m(x) = \text{med}(Y|X = x)$ is the median of the conditional distribution function.

If $F(y|X=x)$ is symmetric around $\tau(x)$ and ψ is an odd, strictly increasing, bounded and continuous function, we have that $g(x)=\tau(x)$. The robust nonparametric estimators of $g(x)$ related to kernel and nearest-neighbor weights are defined as the unique solution $g_T(x)$ and $\hat{g}_T(x)$ of

$$\sum_{t=p+1}^{T} w_{tT}(x)\psi((Y_t - g_T(x))/s_T(x)) = 0 \tag{1.4}$$

or

$$\sum_{t=p+1}^{T} W_{tT}(x)\psi((Y_t - \hat{g}_T(x))/\hat{s}_T(x)) = 0, \tag{1.5}$$

respectively, where w_{tT} and W_{tT} are defined in (1.1) and (1.2), respectively. The scale measures are, for instance, $s_T(x) = \operatorname{med}(|Y - m_T(x)|\,|X=x)$ and $\hat{s}_T(x) = \operatorname{med}(|Y - \hat{m}_T(x)|\,|X=x)$, where the medians are evaluated corresponding to the empirical conditional distribution function based on kernel and on nearest-neighbor methods, respectively, i.e.,

$$F_T(y|X=x) = \sum_{t=p+1}^{T} w_{tT}(x)1_A(Y_t), \tag{1.6}$$

$$\hat{F}_T(y|X=x) = \sum_{t=p+1}^{T} W_{tT}(x)1_A(Y_t), \tag{1.7}$$

where $A=(-\infty, y]$ and 1_A denotes the indicator function of the set A. More generally, any robust estimator consistent to $s(x)$ can be used.

In Section 2 the asymptotic normality of the proposed estimates is stated. In Theorem 1 we require the kernel's bandwidth to verify the condition $\lim_{T\to\infty} Th_T^{p+2} = \beta$, $0 \le \beta < \infty$, instead of $\lim_{T\to\infty} Th_T^{p+2} = 0$ and the asymptotic bias is calculated. Analogously, in Theorem 2, we require the sequence k_T to verify that $\lim_{T\to\infty} k_T T^{-2/(p+2)} = \gamma$, $0 \le \gamma < \infty$. The bias of both estimates is the same as for their linear relatives and the relationship between both asymptotic variances is the same as for kernel and nearest-neighbor density estimates.

In Section 3 proofs and some auxiliary results are given.

2. Asymptotic distribution. We will need the following assumptions:

H1. $\psi\colon \mathbb{R} \to \mathbb{R}$ is an odd, strictly increasing, bounded and continuous function such that $\lim_{t\to\infty}\psi(t) = a > 0$.

H2. The function ψ is twice continuously differentiable with second derivative ψ'' verifying that there exist positive constants c, M and ε such that $|\psi''(t)| \le c|t|^{-(2+\varepsilon)}$ for $|t| \ge M$.

For instance $\psi(t) = \operatorname{arctg}(t)$ verifies H1 and H2.

H3. $E[\psi'((Y_t - g(x))/s(x))|X_t = x] \ne 0$.

H4. The process $\{(X_t, Y_t): t \geq p+1\}$ is a strictly stationary α-mixing process, with the mixing coefficients $\alpha(n)$ verifying:

$$N \sum_{N+1}^{\infty} \alpha(j) \to 0 \quad \text{as } N \to \infty.$$

H5. The vector X_t has a density f continuous and positive at x.

H6. (a) For all $s \geq 1$ the density $f_s(u, v)$ of (X_t, X_{t+s}) is bounded uniformly in s, or

(b) For all $s \geq p+1$ the density $f_s(u, v)$ of (X_t, X_{t+s}) is bounded uniformly in s and $X_t = (Z_{t-1}, \dots, Z_{t-p})$, $Y_t = Z_t$, where $\{Z_t\}$ is a strictly stationary α-mixing process.

H7. The kernel $K: \mathbb{R}^p \to \mathbb{R}$ is bounded, nonnegative, $\int K(u)\, du = 1$ and $|u|^p K(u) \to 0$ as $|u| \to \infty$.

H8. There exists $0 \leq \beta < \infty$ such that $h_T T^{1/(p+2)} \to \beta$ as $T \to \infty$.

H9. There exists a continuous, symmetric distribution function F_0 such that the conditional distribution $F(y|X = x) = F_0((y - g(x))/s(x))$ with g and s such that

(a) g verifies a Lipschitz condition of order 1, and there exists

$$\lim_{\varepsilon \to 0} (g(x + \varepsilon u) - g(x))/\varepsilon = g'(x, u),$$

(b) s verifies a Lipschitz condition of order $\frac{1}{2}$, i.e., $|s(u) - s(x)| \leq C|u - x|^{1/2}$ for some $C > 0$, and $\lim_{\varepsilon \to 0}(s(x + \varepsilon u) - s(x))/\varepsilon^{1/2} = 0$.

Note that without loss of generality we may assume that the scale function of $F(y|X = u)$ is $s(u) = MAD_C(u)$.

H10. The kernel K is twice continuously differentiable and verifies:

(a) $0 < \int |K_1(u)|\, du < \infty$, $\int K_1^2(u)\, du < \infty$ and $|u|^p K_1(u) \to 0$ as $|u| \to \infty$, where $K_1(u) = \sum_{j=1}^{p} (\partial K/\partial u_j)(u) u_j$.

(b) $|u|^{p+1} K_2(u) \to 0$ as $|u| \to \infty$, where $K_2(u) = \sum_{i,j} (\partial^2 K/\partial u_i\, \partial u_j)(u) u_i u_j$ and $u = (u_1, \dots, u_p)$.

H11. There exists $0 \leq \beta < \infty$ such that $k_T^{1/p} T^{(1/(p+2) - 1/p)} \to \beta(f(x)\lambda(V_1))^{1/p}$, where $\lambda(V_1)$ denotes the Lebesgue measure of the unit ball.

THEOREM 1. *Assume $H1$ to $H9$ and that $g_T(x) \to g(x)$ in probability. Then if $s_T(x)$ is any sequence of scale estimators such that $s_T(x) \to s(x)$ in probability*

$$(Th_T^p)^{1/2}(g_T(x) - g(x)) \to_w N\left(b_1, \sigma_1^2 \int \psi^2(u)\, dF_0(u) \Big/ \left(\int \psi'(u)\, dF_0(u)\right)^2\right)$$

holds, where

$$b_1 = \beta^{(1+p/2)} \int g'(x, u) K(u)\, du, \qquad \sigma_1^2 = s^2(x) \int K^2(u)\, du / f(x)$$

and $\to_w$ stands for weak convergence.

THEOREM 2. *Assume H1 to H11 and that* $\hat{g}_T(x) \to g(x)$ *in probability. Then, if* $\hat{s}_T(x) \to s(x)$ *in probability we have that* $k_T^{1/2}(\hat{g}_T(x) - g(x))$ *is asymptotically normally distributed with mean* $b = b_1(f(x)\lambda(V_1))^{1/2}$ *and variance* $\sigma^2 = \sigma_1^2 f(x)\lambda(V_1)V(\psi)$, *where* b_1 *and* σ_1^2 *are given in Theorem* 1, *and* $V(\psi) = \int \psi^2(u)\,dF_0(u)/(\int \psi'(u)\,dF_0(u))^2$.

REMARK 2.1. Conditions and a proof for the almost sure convergence of $g_T(x)$ and $\hat{g}_T(x)$ to $g(x)$ may be found in Theorems 2.1 and 3.1 of Boente and Fraiman (1989a). However, in order to obtain the asymptotic distribution of the estimates we just need the convergence in probability, which follows straightforwardly by a second-order Taylor expansion as in the proof of Lemma 4 of Section 3. Therefore we will just assume it on Theorems 1 and 2.

REMARK 2.2. The asymptotic variance in the case of kernel weights is given by $V_1 = \sigma_1^2 V(\psi)$ with

$$V(\psi) = \frac{\int \psi^2((u - g(x))/s(x))\,dF_{Y|X=x}(u)}{\left(\int \psi'((u - g(x))/s(x))\,dF_{Y|X=x}(u)\right)^2}$$

and

$$\sigma_1^2 = s^2(x)\int K^2(u)\,du/f(x)$$

and by

$$V_2 = V_1 f(x)\lambda(V_1) = s^2(x)\int K^2(u)\,du\;\lambda(V_1)V(\psi)$$

when we use nearest-neighbor weights. $s_T^2(x)$ [respectively $\hat{s}_T^2(x)$] provides a consistent estimator of $s^2(x)$ and $f_T(x) = (Th_T^p)^{-1}\sum_{t=p+1}^T K((X_t - x)/h_T)$ is a consistent estimator of $f(x)$, as was shown by Robinson (1983). Then in order to estimate V_1 (respectively V_2) it is enough to give a consistent estimator of $V(\psi)$. For kernel weights define

$$B_{1T} = \sum_{t=p+1}^{T} w_{tT}(x)\psi^2((Y_t - g_T(x))/s_T(x))$$

and

$$B_{2T} = \sum_{t=p+1}^{T} w_{tT}(x)\psi'((Y_t - g_T(x))/s_T(x)) = \tilde{\lambda}_T(x, g_T(x), s_T(x)).$$

Then $A_T = B_{1T}/B_{2T}^2$ is a consistent estimate of $V(\psi)$. Effectively, in the proof of Lemma 4, it is shown that $\tilde{\lambda}_T(x, g_T(x), s_T(x))$ converges to $\tilde{\lambda}(x, g(x), s(x)) = \int \psi'(u)\,dF_0(u)$ since $g_T(x)$ and $s_T(x)$ are consistent estimates of $g(x)$ and $s(x)$. The same argument can be applied to ψ^2 since ψ is strictly increasing and bounded. Thus B_{1T} converges to $\int \psi^2((u - g(x))/s(x))\,dF_{Y|X=x}(u)$ in probability, and therefore $A_T s_T^2(x) \int K^2(u)\,du / f_T(x)$ is a consistent estimate of the asymptotic variance.

An analogous argument hold for nearest-neighbor weights.

REMARK 2.3. The problem of bandwidth selection has been considered by several authors. In the i.i.d. case, Härdle and Marron (1985) considered a cross-validation method for the classical kernel estimate, which tends to select a bandwidth that yields a good estimate of the regression function, and they showed the asymptotic optimality with respect to the mean square error. Vieu and Hart (1988) adapted this proposal to φ-mixing processes. Wong (1983) also considered the smoothness parameter by cross-validation on the average square error and showed consistency in the i.i.d. case. Recently, Härdle, Hall and Marron (1988) and Härdle and Bowman (1988) studied the problem of the smoothing parameter selection for fixed carriers in the i.i.d. case, using cross-validation with a weighted least squares criterion.

In our framework, both the smoothing parameter h_T (or the number of nearest neighbors k_T) and β can be selected by minimizing the asymptotic mean square error as was done by Härdle (1986) in the i.i.d. case for his robust proposal. Since the objective function is the same, the problem is reduced to using consistent estimators of the scale, $s(x)$ and the density function, which are given, for instance, in Remark 2.2 and in Boente and Fraiman (1986), respectively. This selection of the parameters involved will lead to consistent and asymptotically normally distributed estimators.

3. Proofs. In order to prove Theorem 1 we will use the following lemma due to Robinson (1983), Lemma 7.1.

LEMMA 1. *Let* $\{V_{tT},\ 1 \le t \le T,\ T \ge 1\}$ *be a triangular array of random variables zero mean and* $\{a_T\colon T \ge 1\}$ *a sequence of positive constants such that*:

(i) *For each* T, V_{tT}, $t = 1, \dots, T$, *are identically distributed random variables and* V_{tT} *is measurable with respect to the* σ*-field generated by* (X_t, Y_t) *with* $\{(X_t, Y_t)\colon t \ge 1\}$ *verifying* H4.

(ii) *There exists* $C > 0$ *such that* $P(|V_{tT}| \le C) = 1$ *for all* $1 \le t \le T$, $T \ge 1$.

(iii) $a_T \to 0$ *and* $Ta_T \to \infty$ *as* $T \to \infty$.

(iv) *There exists* $\sigma^2 > 0$ *such that* $E(V_{tT}^2)/a_T \to \sigma^2$ *as* $T \to \infty$.

(v) *There exists* $C_1 > 0$ *independent of* T *such that* $E(|V_{tT} V_{t+s,T}|) \le C_1 a_T^2$ *for* $s \ge 1$, $1 \le t \le T$ *and* T *large enough*.

Then $S_T = (Ta_T)^{-1/2} \sum_{t=1}^{T} V_{tT}$ *converges in distribution to a normal variable with zero mean and variance* σ^2.

In the following lemma we give a short proof of the asymptotic distribution of the linear kernel estimates, using Lemma 1, in order to include the case where $Th_T^{p+2} \to \beta$, $0 \le \beta < \infty$, as $T \to \infty$.

LEMMA 2. *Let* $\{(X_t, Z_t): t \geq 1\}$ *be a stationary random process verifying* H4 *such that* $|Z_t| \leq M$ *for all* $t \geq 1$. *Denote by* $F(z|X_1 = u)$ *the conditional distribution of* Z_1 *given* $X_1 = u$, *by* $\Phi(u) = E(Z_1|X_1 = u)$ *and by* $\sigma^2(u) = E((Z_1 - \Phi(x))^2|X_1 = u)$. *Let us suppose that*:

(i) Φ *is Lipschitz and* $\lim_{\varepsilon \to 0}(\Phi(x + \varepsilon u) - \Phi(x))/\varepsilon = \Phi'(x, u)$.
(ii) σ^2 *is continuous in a neighborhood of* x.

Let $\Phi_T(x) = \sum_{t=1}^T w_{tT}(x) Z_t$ *with* $w_{tT}(x)$ *defined as in* (1.1). *Then* H5 *to* H8 *imply that* $(Th_T^p)^{1/2}(\Phi_T(x) - \Phi(x))$ *is asymptotically normally distributed with mean* $b_2 = \beta^{(1+p/2)}\int \Phi'(x, u)K(u)\,du$ *and variance* $\sigma_2^2 = \sigma^2(x)K_0$ *with* $K_0 = \int K^2(u)\,du/f(x)$.

PROOF. Since $(Th_T^p)^{-1}\sum_{t=1}^T K((X_t - x)/h_T)$ converges to $f(x)$ in probability, it is enough to show that

$$\text{(a)} \qquad (Th_T^p)^{-1/2} \sum_{t=1}^T K((X_t - x)/h_T)(Z_t - \Phi(X_t)) \to_w N\big(0, \sigma^2(x) f(x) \textstyle\int K^2(u)\,du\big)$$

and

$$\text{(b)} \qquad S_T = (Th_T^p)^{-1/2} \sum_{t=1}^T K((X_t - x)/h_T)(\Phi(X_t) - \Phi(x)) \to b_2 f(x) \quad \text{in probability.}$$

Under H6(b), (a) follows from Theorem 5.1 of Robinson (1983). Under H6(a), (a) follows easily applying Lemma 1 to

$$V_{tT} = K((X_t - x)/h_T)(Z_t - \Phi(X_t)).$$

As

$$E(S_T) = (Th_T^p)^{1/2} \int K(u)(\Phi(uh_T + x) - \Phi(x)) f(uh_T + x)\,du,$$

H7, H8(i) and the dominated convergence theorem entail that

$$\lim_{T \to \infty} E(S_T) = f(x)\beta^{(p+2)/2} \int K(u)\Phi'(x, u)\,du.$$

Therefore, in order to prove (b), it is enough to show that the variance of S_T verifies that $\lim_{T \to \infty} V(S_T) = 0$. Denote by $W_{tT} = K((X_t - x)/h_T)(\Phi(X_t) - \Phi(x))$. As $|W_{tT}| \leq 2M|K|_\infty = C_1$, where $|K|_\infty = \sup_{u \in \mathbb{R}^p}|K(u)|$, Theorem 17.2.1 of Ibragimov and Linnik (1971) implies that $|\mathrm{Cov}(W_{tT}, W_{t+r,T})| \leq \alpha(r)C_1$. On the other hand, using H7 and assumption (i), standard arguments led to the following inequalities:

$$|\mathrm{Cov}(W_{tT}, W_{t+r,T})| \leq C_2 h_T^2 (h_T^p)^2 \quad \text{for } r > p,$$

$$|\mathrm{Cov}(W_{tT}, W_{t+r,T})| \leq C_2 h_T^2 (h_T^p) \quad \text{for } 1 \leq r \leq p,$$

and

$$V(W_{tT}) \le C_2 h_T^2 h_T^p,$$

for some positive constant C_2. Therefore, as

$$V(S_t) \le (Th_T^p)^{-1} \sum_{t=1}^{T} V(W_{tT}) + 2(Th_T^p)^{-1} \sum_{t=1}^{T-1} \sum_{r=1}^{T-t} |\mathrm{Cov}(W_{tT}, W_{t+r,T})|$$

$$\le C_2 h_T^2 + 2C_2(N-p) h_T^p h_T^2 + 2C_1 (h_T^p)^{-1} \sum_{N+1}^{\infty} \alpha(r) + 2C_2 p h_T^2,$$

taking $N = [(h_T^p)^{-1}] + 1$, H4 implies the desired result. □

We will denote by $\psi_\sigma(t) = \sigma\psi(t/\sigma)$.

LEMMA 3. *Under* H1, H2 *and* H4 *to* H7, *if* $h_T \to 0$ *and* $Th_T^p \to \infty$ *as* $T \to \infty$ *and* $F(y|X = x)$ *is symmetric around* $g(x)$ *we have*

$$(Th_T^p)^{1/2} \sum_{t=p+1}^{T} w_{tT}(x)\left[\psi_{\sigma(T)}(Y_t - g(x)) - \psi_\sigma(Y_t - g(x))\right] \to 0 \quad \text{as } T \to \infty,$$

in probability for any sequence $\sigma(T) = \sigma_T(x)$ *such that* $\sigma_T(x) \to \sigma(x) = \sigma > 0$ *as* $T \to \infty$ *in probability.*

PROOF. By H5 and H7 we have that $(Th_T^p)^{-1}\sum_{t=p+1}^{T} K((X_t - x)/h_T)$ converges to $f(x)$ in probability. Therefore, it is enough to show that

$$(Th_T^p)^{-1/2} \sum_{t=p+1}^{T} K((X_t - x)/h_T)\left[\psi_{\sigma(T)}(Y_t - g(x)) - \psi_\sigma(Y_t - g(x))\right] \to 0$$

in probability as $T \to \infty$. For any $s > 0$ define $H_s(u) = \psi_{\sigma+s}(u) - \psi_\sigma(u)$, $I_s(u) = \psi_{\sigma-s}(u) - \psi_\sigma(u)$, $J_T^+(s) = (Th_T^p)^{-1/2}\sum_{t=p+1}^{T} K((X_t - x)/h_T) H_s(Y_t - g(x))$, $\bar{J}_T^+(s) = J_T^+(s) - E(J_T^+(s))$, $J_T^-(s) = (Th_T^p)^{-1/2}\sum_{t=p+1}^{T} K((X_t - x)/h_T) I_s(Y_t - g(x))$ and $\bar{J}_T^-(s) = J_T^-(s) - E(J_T^-(s))$.

It suffices to show that

$$\lim_{T\to\infty} \lim_{d\to 0} \sup_{0\le s\le d} E(J_T^+(s)) = 0 \tag{3.1}$$

and

$$\lim_{T\to\infty} \lim_{d\to 0} P\left(\sup_{0\le s\le d} |\bar{J}_T^+(s)| > \varepsilon \right) = 0 \tag{3.2}$$

holds and the same result with $J_T^-(s)$ instead of $J_T^+(s)$. (3.1) is straightforward using H1, H7 and the dominated convergence theorem.

In order to show (3.2) it is enough to prove that the sequence $\bar{J}_T^+(s)$ of random variables on the space $C([0,1])$ of continuous functions on $[0,1]$ is tight. According to Theorem 12.3 of Billingsley (1968) if suffices to verify that

1. The sequence $\{\bar{J}_T^+(0)\}$ is tight.
2. There exist constants $\gamma > 0$ and $\alpha > 1$ and a nondecreasing function F, on $[0, 1]$ such that

$$E\left(\left|\bar{J}_T^+(s_2) - \bar{J}_T^+(s_1)\right|^\gamma\right) \le (F(s_2) - F(s_1))^\alpha$$

for all $0 \le s_1 < s_2$ and T large enough.

(1) follows since $\bar{J}_T^+(0) = 0$.

As in Lemma A of Fraiman (1980) we have

$$|H_{s_2}(u) - H_{s_1}(u)| \le F(s_2) - F(s_1), \tag{3.3}$$

where $F(t) = at - C(\sigma + t)^{-1}$, C is a positive constant and a is given in H1.

Denote by $V_{tT} = K((X_t - x)/h_T)[H_{s_2}(Y_t - g(x)) - H_{s_1}(Y_t - g(x))]$. Then we have that (3.3) implies that $|V_{tT}| \le \sup\{|K(u)|: u \in R^p\}(F(s_2) - F(s_1)) = C_1$. Therefore Theorem 17.2.1 of Ibragimov and Linnik (1971) implies

$$|\mathrm{Cov}(V_{tT}, V_{t+r,T})| \le \alpha(r) C_1^2. \tag{3.4}$$

On the other hand,

$$\begin{aligned}
|\mathrm{Cov}(V_{tT}, V_{t+r,T})| &\le (F(s_2) - F(s_1))^2 \\
&\quad \times \{E[K((X_t - x)/h_T) K((X_{t+r} - x)/h_T)] \\
&\qquad + E^2 K((X_t - x)/h_T)\} \\
&\le h_T^{2p}(F(s_2) - F(s_1))^2 \{C_2 + (\textstyle\int K(u) f(uh_T + x)\, du)^2\}
\end{aligned}$$

holds for $r > p$ from H5 and H6 with C_2 a positive constant.

From a slight modification of Bochner's theorem used by Parzen (1962) we have

$$\int K(u) f(uh_T + x)\, du \to f(x) \quad \text{as } T \to \infty,$$

which entails that there exists positive constants C_3, C_4 and $T_0 \in N$ such that

$$\begin{aligned}
|\mathrm{Cov}(V_{tT}, V_{t+r,T})| &\le C_3 h_T^{2p}(F(s_2) - F(s_1))^2 \quad \text{for } r > p, \\
|\mathrm{Cov}(V_{tT}, V_{t+r,T})| &\le C_4 h_T^{p}(F(s_2) - F(s_1))^2 \quad \text{for } 1 \le r \le p,
\end{aligned} \tag{3.5}$$

and

$$\begin{aligned}
E(V_{tT}^2) &\le (F(s_2) - F(s_1))^2 E[K^2((X_t - x)/h_T)] \\
&\le C_4 h_T^p (F(s_2) - F(s_1))^2 \quad \text{for } T \ge T_0.
\end{aligned} \tag{3.6}$$

Finally, as

$$\begin{aligned}
&E\left(\left(\bar{J}_T^+(s_2) - \bar{J}_T^+(s_1)\right)^2\right)\\
&= (Th_T^p)^{-1}\operatorname{Var}\left(\sum_{t=p+1}^{T} V_{tT}\right)\\
&= (Th_T^p)^{-1}\sum_{t=p+1}^{T}\operatorname{Var}(V_{tT}) + 2(Th_T^p)^{-1}\sum_{t=p+1}^{T-1}\sum_{r=1}^{T-t}\operatorname{Cov}(V_{tT}, V_{t+r,T})\\
&\le (Th_T^p)^{-1}\sum_{t=p+1}^{T} E(V_{tT}^2)\\
&\quad + 2(Th_T^p)^{-1}\sum_{t=p+1}^{T}\left\{\sum_{r=1}^{N}|\operatorname{Cov}(V_{tT}, V_{t+r,T})| + \sum_{r=N+1}^{T}|\operatorname{Cov}(V_{tT}, V_{t+r,T})|\right\},
\end{aligned}$$

(3.4), (3.5) and (3.6) imply that for $T \ge T_0$,

$$\begin{aligned}
&E\left(\left(\bar{J}_T^+(s_2) - \bar{J}_T^+(s_1)\right)^2\right)\\
&\quad\le (F(s_2) - F(s_1))^2\left[C_4(p+1) + C_3(N-p)h_T^p + \|K\|_\infty h_T^{-p}\sum_{N+1}^{\infty}\alpha(r)\right].
\end{aligned}$$

Let $N = [h_T^{-p}] + 1$, then we have $Nh_T^p \le 2$, $N \le T$ as $Th_T^p \to \infty$ and therefore H4 implies that for T large enough

$$\begin{aligned}
&E\left(\left(\bar{J}_T^+(s_2) - \bar{J}_T^+(s_1)\right)^2\right)\\
&\quad\le (F(s_2) - F(s_1))^2\left[C_4(p+1) + 2C_3 + \|K\|_\infty N\sum_{N+1}^{\infty}\alpha(r)\right]\\
&\quad\le C(F(s_2) - F(s_1))^2
\end{aligned}$$

with $C = C_4 + 3C_3$.

A similar argument shows that (3.1) and (3.2) holds for $J_T^-(s)$. □

REMARK 3.1. Note that Lemma 3 also holds if we replace ψ by ψ^2 and by ψ' since (3.3) holds in these cases.

LEMMA 4. *Under* H3 *and the assumptions of Lemma* 3, *if* $F(y|X = x)$ *is a continuous function of* x, *symmetric around* $g(x)$ *and* $g_T(x) \to g(x)$ *and* $s_T(x) \to s(x)$ *is probability, we have that* $(Th_T^p)^{1/2}(g_T(x) - g(x))$ *has the same asymptotic distribution as*

$$s(x)\left[\tilde{\lambda}(x, g(x), s(x))\right]^{-1}(Th_T^p)^{1/2}\sum_{t=p+1}^{T} w_{tT}(x)\psi((Y_t - g(x))/s(x)),$$

where $\tilde{\lambda}(x, u, \sigma) = \int\psi'((y-u)/\sigma)\,dF(y|X = x)$.

PROOF. Denote by $\lambda_T(x, u, \sigma) = \sum_{t=p+1}^{T} w_{tT}(x)\psi((Y_t - u)/\sigma)$ and by $\tilde{\lambda}_T(x, u, \sigma) = \sum_{t=p+1}^{T} w_{tT}(x)\psi'((Y_t - \mu)/\sigma)$. The mean value theorem entails

$$\begin{aligned} 0 &= (Th_T^p)^{1/2}\lambda_T(x, g_T(x), s_T(x))s_T(x) \\ &= (Th_T^p)^{1/2}\lambda_T(x, g(x), s_T(x))s_T(x) \\ &\quad -(Th_T^p)^{1/2}(g_T(x) - g(x))A_T, \end{aligned} \tag{3.7}$$

with

$$A_T = \tilde{\lambda}_T(x, g(x), s_T(x)) + (g_T(x) - g(x))\gamma_T(x, \xi_T(x), s_T(x))/s_T(x)$$

and

$$\xi_T(x) = (1 - \theta_T)g(x) + \theta_T g_T(x)$$

with $0 < \theta_T < 1$,

$$\gamma_T(x, u, \sigma) = (1/2) \sum_{t=p+1}^{T} w_{tT}(x)\psi''((Y_t - u)/\sigma).$$

From H2 we have that $\gamma_T(x, \xi_T(x), s_T(x))$ is bounded and therefore the second term in A_T converges to 0 in probability.

On the other hand, $\tilde{\lambda}_T(x, g(x), s_T(x)) - \tilde{\lambda}_T(x, g(x), s(x)) \to 0$ in probability by Remark 3.1. Finally, if we show that $\tilde{\lambda}_T(x, g(x), s(x)) \to \tilde{\lambda}(x, g(x), s(x))$ in probability, the conclusion of Lemma 4 follows from (3.7) and Lemma 3.

Since $(Th_T^p)^{-1}\sum_{t=p+1}^{T} K_1((X_t - x)/h_T)$ converges to $f(x)$ in probability, it is enough to show that $(Th_T^p)^{-1}\sum_{t=p+1}^{T} K_1((X_t - x)/h_T)\psi'((Y_t - g(x))/s(x))$ converges to $f(x)\tilde{\lambda}(x, g(x), s(x))$ in probability, which follows straightforwardly from Markov's inequality, by majorizing the variance as in (3.5) and (3.6). □

PROOF OF THEOREM 1. As $g_T(x)$ converges to $g(x)$ in probability we have that by Lemma 4 it suffices to show that

$$\begin{aligned} &(Th_T^p)^{1/2} \sum_{t=p+1}^{T} w_{tT}(x)Z_t \\ &\quad \to_w N\left(b_1\int\psi'(u)\,dF_0(u)/s(x), \sigma_1^2\int\psi^2(u)\,dF_0(u)/s^2(x)\right), \end{aligned} \tag{3.8}$$

where $Z_t = \psi((Y_t - g(x))/s(x))$, which follows from Lemma 2 since $\Phi(x) = 0$, $\sigma^2(x) = \int\psi^2(u)\,dF_0(u)$ and $\Phi'(x, u) = g'(x, u)\int\psi'(t)\,dF_0(t)/s(x)$. □

PROOF OF THEOREM 2. Define as in Lemma 4,

$$\lambda_T(x, u, \sigma) = \sum_{t=p+1}^{T} W_{tT}(x)\psi((Y_t - u)/\sigma)$$

and

$$\tilde{\lambda}_T(x,u,\sigma) = \sum_{t=p+1}^{T} W_{tT}(x)\psi'((Y_t - u)/\sigma).$$

The mean value theorem entails

$$(3.9)\quad \begin{aligned} 0 = k_T^{1/2}\lambda_T(x,\hat{g}_T(x),\hat{s}_T(x))\hat{s}_T(x) &= k_T^{1/2}\lambda_T(x,g(x),\hat{s}_T(x))\hat{s}_T(x) \\ &\quad - k_T^{1/2}(\hat{g}_T(x) - g(x))A_T^*, \end{aligned}$$

where

$$A_T^* = \tilde{\lambda}_T(x,g(x),\hat{s}_T(x)) + (\hat{g}_T(x) - g(x))\gamma_T(x,\xi_T(x),\hat{s}_T(x))/\hat{s}_T(x),$$
$$\xi_T(x) = (1-\theta_T)g(x) + \theta_T\hat{g}_T(x)$$

with $0 < \theta_T < 1$ and

$$\gamma_T(x,u,\sigma) = (1/2)\sum_{t=p+1}^{T} W_{tT}(x)\psi''((Y_t - u)/\sigma).$$

As in Lemma 4, the second term in A_T^* converges to 0 in probability.

Then (3.9) implies that $k_T^{1/2}(\hat{g}_T(x) - g(x))$ has the same asymptotic distribution as

$$s(x)\left(\int\psi'(u)\,dF_0(u)\right)^{-1} k_T^{1/2}\sum_{t=p+1}^{T} W_{tT}(x)\psi((Y_t - g(x))/\hat{s}_T(x)),$$

if we show that $\tilde{\lambda}_T(x,g(x),\hat{s}_T(x))$ converges to $\tilde{\lambda}(x,g(x),s(x))$ in probability.

Therefore it is enough to show that

$$\text{(i)}\qquad k_T^{1/2}\sum_{t=p+1}^{T} W_{tT}(x)\psi((Y_t - g(x))/\hat{s}_T(x)) \to^w N(b_3,\sigma_3^2),$$

where $b_3 = (f(x)\lambda(V_1))^{1/2}b_1\int\psi'(u)\,dF_0(u)/s(x)$, $\sigma_3^2 = f(x)\lambda(V_1)\sigma_1^2B_0/s^2(x)$ and $B_0 = \int\psi^2(u)\,dF_0(u)$
and

$$\text{(ii)}\qquad \tilde{\lambda}_T(x,g(x),\hat{s}_T(x)) \to \tilde{\lambda}(x,g(x),s(x)) \quad \text{in probability.}$$

In order to prove (i), let $h_T^p = k_T/(Tf(x)\lambda(V_1))$. Boente and Fraiman (1986) established that $k_T^{1/2}((h_T/H_T) - 1)$ is asymptotically normally distributed and that $(TH_T^p)^{-1}\sum_{t=p+1}^{T}K((X_t - x)/H_T) \to f(x)$ in probability. Thus it is enough to show that

$$S_T = k_T^{1/2}(Th_T^p)^{-1}\sum_{t=p+1}^{T} K((X_t - x)/H_T)\psi((Y_t - g(x))/\hat{s}_T(x))$$

converges to a normal distribution with mean $b_4 = b_3f(x)$ and variance $\sigma_4^2 = \sigma_3^2f(x)^2$.

A second-order Taylor expansion gives $S_T = S_{T1} + S_{T2} + S_{T3}$, where

$$S_{T1} = k_T^{1/2}(Th_T^p)^{-1} \sum_{t=p+1}^{T} K((X_t - x)/h_T)\psi((Y_t - g(x))/\hat{s}_T(x)),$$

$$S_{T2} = k_T^{1/2}((h_T/H_T) - 1)(Th_T^p)^{-1} \times \sum_{t=p+1}^{T} K_1((X_t - x)/h_T)\psi((Y_t - g(x))/\hat{s}_T(x)),$$

$$S_{T3} = k_T^{1/2}((h_T/H_T) - 1)^2(Th_T^p)^{-1} \times \sum_{t=p+1}^{T} K_2((X_t - x)/\xi_T)\psi((Y_t - g(x))/\hat{s}_T(x)),$$

where $\min(h_T, H_T) \le \xi_T \le \max(h_T, H_T)$.

The proof of (i) will be complete if we show that

(a) $S_{T1} \to_w N(b_4, \sigma_4^2)$,
(b) $S_{T2} \to 0$ and $S_{T3} \to 0$ in probability as $T \to \infty$.

As $U_T = (f(x)\lambda(V_1))^{1/2}(Th_T^p)^{-1}\sum_{t=p+1}^{T} K((X_t - x)/h_T)$ converges in probability to $(f(x)\lambda(V_1))^{1/2} f(x)\int K(u)\,du$ and $S_{T1} = (Th_T^p)^{1/2}\sum_{t=p+1}^{T} w_{tT}(x)\psi((Y_t - g(x))/\hat{s}_T(x))U_T$, (a) follows from Lemma 3 and (3.8).

Since $k_T^{1/2}((h_T/H_T) - 1)$ is asymptotically normally distributed and ψ is bounded, in order to prove (b) it is enough to show that

$$(Th_T^p)^{-1} \sum_{t=p+1}^{T} K_1((X_t - x)/h_T)\psi((Y_t - g(x))/\hat{s}_T(x)) \to 0 \quad \text{in probability} \tag{3.10}$$

and that

$$(Th_T^p)^{-1} \sum_{t=p+1}^{T} |K_2((X_t - x)/\xi_T)| \quad \text{is bounded in probability.} \tag{3.11}$$

(3.11) can be obtained from H10(b) in a similar way as in Boente and Fraiman (1986), Theorem 5.

Denote by $\lambda_T^1(x, u, \sigma) = \sum_{t=p+1}^{T} w_{tT}^1(x)\psi((Y_t - u)/\sigma)$, where $w_{iT}^1(x) = K_1((X_i - x)/h_T)/\sum_{t=p+1}^{T} K_1((X_t - x)/h_T)$. By H10(a) and Lemma 3 we have that

$$\lambda_T^1(x, g(x), \hat{s}_T(x)) - \lambda_T^1(x, g(x), s(x)) \to 0 \quad \text{in probability.}$$

Then (3.10) follows if we show that

$$Q_T = (Th_T^p)^{-1} \sum_{t=p+1}^{T} K_1((X_t - x)/h_T)\psi((Y_t - g(x))/s(x)) \to 0 \quad \text{in probability.}$$

Since H3 holds, $E(Q_T) \to 0$ and straightforward calculations lead to $\mathrm{Var}(Q_T) \to 0$.

(ii) After a second-order Taylor expansion as in (i), the same argument used in Lemma 4 can be applied to conclude the proof. □

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

BOENTE, G. and FRAIMAN, R. (1986). Strong order of convergence and asymptotic distribution of nearest neighbor density estimates from dependent observations. Trabajos de Matemática 115. Publicaciones del I.A.M. Conicet.

BOENTE, G. and FRAIMAN, R. (1989a). Robust nonparametric regression estimation for dependent observations. *Ann. Statist.* **17** 1242–1256.

BOENTE, G. and FRAIMAN R. (1989b). Robust nonparametric regression. *J. Multivariate Anal.* **29** 180–198.

BOSQ, D. (1980). Une méthode non paramétrique de prédiction d'un processus stationnaire. Prédictions d'une mesure aléatoire. *C. R. Acad. Sci. Paris Sér. A* **290** 711–713.

BRADLEY, R. C. (1983). Asymptotic normality of some kernel-type estimators of probability density. *Statist. Prob. Lett.* **1** 295–300.

CHENG, K. F. and CHENG, P. E. (1987). Robust nonparametric estimation of a regression function. *Sankhyá Ser. B* **49** 9–22.

COLLOMB, G. (1980). Estimation de la régression par la méthode des k points les plus proches avec noyau: Quelques propiétés de convergence ponctuelle. *Lecture Notes in Math.* **831** 159–175.

COLLOMB, G. (1982). Prédiction non paramétrique: Étude de l'erreur quadratique du prédictogramme. *C. R. Acad. Sci. Paris Sér. A* **294** 59–62.

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

COLLOMB, G. (1985). Nonparametric time series analysis and prediction: Uniform almost sure convergence of window and *k-NN* autoregression estimates. *Statistics* **16** 297–307.

COLLOMB, G. and HÄRDLE, W. (1984). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stochastic Process. Appl.* **23** 77–89.

DOUKHAN, P. and GHINDÈS, M. (1980). Estimations dans le processus "$x_{n+1} = f(x_n) + \varepsilon_n$." *C. R. Acad. Sci. Paris Sér. A* **290** 921–923.

DOUKHAN, P. and GHINDÈS, M. (1983). Estimation dans le processus autorégressif d'ordre 1. *Stochastic Process. Appl.* **15** 271–293.

DOUKHAN, P., LEÓN, J. and PORTAL, F. (1985). Principes d'invariance faible pour la mesure empirique d'une suite de variables aléatoires mélangeante. Unpublished manuscript.

FRAIMAN, R. (1980). Estimadores robustos para regresión no lineal. Doctoral dissertation. Univ. Buenos Aires.

HÄRDLE, W. (1984). Robust regression function estimation. *J. Multivariate Anal.* **14** 169–180.

HÄRDLE, W. (1986). A two-stage simultaneous estimation and bandwidth selection rule. Unpublished manuscript.

HÄRDLE, W. and BOWMAN, A. W. (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *J. Amer. Statist. Assoc.* **83** 102–110.

HÄRDLE, W., HALL, P. and MARRON, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *J. Amer. Statist. Assoc.* **83** 86–101.

HÄRDLE, W. and MARRON, J. S. (1985). Optimal bandwidth choice in nonparametric regression estimation. *Ann. Statist.* **13** 1465–1481.

HÄRDLE, W. and TSYBAKOV, A. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist.* **16** 120–135.

HUBER, P. (1981). *Robust Statistics*. Wiley, New York.

IBRAGIMOV, I. A. and LINNIK, Y. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noerdhoff, Groningen.

MACK, Y. P. (1981). Local properties of *k-NN* regression estimates. *SIAM J. Algebraic Discrete Methods* **2** 311–323.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

PELIGRAD, M. (1988). Properties of uniform consistency of the kernel estimates of density and regression function. Unpublished manuscript.

ROBINSON, P. M. (1983). Nonparametric estimators for time series. *J. Time Series Anal.* **4** 185–206.

ROBINSON, P. M. (1984). Robust nonparametric autoregression. *Lecture Notes in Statist.* **26** 247–255.

ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42** 43–47.

ROUSSAS, G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.* **40** 1386–1400.

ROUSSAS, G. (1988). Nonparametric regression under mixing conditions. Unpublished manuscript.

SARDA, P. and VIEU, P. (1986). Régression nonparamétrique et prédiction d'un processus Markovien. *Cahiers Centre Études Rech. Opér.* **28** 203–209.

SARDA, P. and VIEU, P. (1988). Vitesses de convergence d'estimateurs non-paramétriques d'une régression et de ses dérivées. *C. R. Acad. Sci. Paris Sér. A* **306** 83–86.

TRUONG, Y. K. and STONE, C. (1988). Nonparametric function estimation involving time series. Unpublished manuscript.

TSYBAKOV, A. (1983). Robust estimates of a function. *Problems Control Inform. Theory* **18** 39–52.

VIEU, P. and HART, J. D. (1988). Nonparametric regression under dependence: A class of asymptotically optimal data-driven bandwidths. Unpublished manuscript.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyá Ser. A* **26** 359–372.

WONG, W. H. (1983). On the consistency of cross-validation in kernel nonparametric regression. *Ann. Statist.* **11** 1136–1141.

YAKOWITZ, S. J. (1985). Nonparametric density estimation and prediction for Markov sequences. *J. Amer. Statist. Assoc.* **80** 215–221.

DEPARTAMENTO DE MATEMATICA
FACULTAD DE CIENCIAS EXACTAS
Y NATURALES
PABELLON 1, CIUDAD UNIVERSITARIA
BUENOS AIRES, 1428
ARGENTINA