The Annals of Statistics
1984, Vol. 12, No. 2, 612–623

UNIFORM CONSISTENCY OF A CLASS OF REGRESSION FUNCTION ESTIMATORS[1]

BY W. HÄRDLE AND S. LUCKHAUS

University of Frankfurt and University of Heidelberg

We study a wide class of nonparametric regression function estimators including kernel estimators and robust smoothers. Under different assumptions on the kernel and the sequence of bandwidths, we obtain weak uniform consistency rates on a bounded interval. The uniform consistency is shown in a "stochastic design model" and in a "fixed design model".

1. Introduction. Let (X_1, Y_1), (X_2, Y_2), $\cdots$ be independent bivariate random data sampled either with stochastic design rv's $X_1, X_2, \cdots$ or with fixed design points $x_1, x_2, \cdots$. In the stochastic design model (X_1, Y_1), (X_2, Y_2), $\cdots$ are independent bivariate random variables identically distributed as a bivariate random variable (X, Y) whose joint cumulative distribution function is F and whose joint probability density is $f(x, y)$. In the fixed design model (noisy sampled data) we have an underlying family of probability density functions $\{f(\cdot; x): x \in [0, 1]\}$ and $\mathscr{P}_n = \{x_1, x_2, \cdots x_n\}$ where $0 \leq x_1 \leq x_2, \cdots, \leq x_n = 1$ is a partition of $[0, 1]$ determined by the experimenter.

The nonparametric regression problem is the problem of estimating the regression curve of Y on X. Equivalently, the nonparametric regression problem requires finding $m(x) = m_{\psi,F}(x)$, given observations

$$\{X_i, m(X_i) + N_i\}_{i=1}^n.$$

The function ψ is used here as an indexing parameter, since, as is shown in examples below, the shape of ψ determines the regression curve $m(x)$. Different choices of ψ yield the conditional mean or the conditional median for instance. The N_i being an independent noise variable which may depend on X_i and $m_{\psi,F}$ being the trend satisfying

$$E_x\psi(Y - m(x)) = 0 \tag{1.1}$$

where $E_x(\cdot) = E(\cdot \mid X = x)$ in the stochastic design case or $E_x(\cdot) = \int \cdot f(y; x)\, dy$ in the fixed case and $\psi(\cdot)$ is a monotone continuous function.

We propose to estimate $m(x)$ by $m_n(x)$ a solution (with respect to θ) of

$$\textstyle\sum_{i=1}^n \alpha_i(x)\psi(Y_i - \theta) = 0 \tag{1.2}$$

where $\alpha_i(x) = \alpha_i^{(n)}(x)$ are (localizing) weights depending on X. In the present

---

Received June 1982; revised May 1983.

[1] This research was made possible by the Deutsche Forschungsgemeinschaft. Sonderforschungsbereich 123 "Stochastische Mathematische Modelle" and Air Force Scientific Research Contract AFOSR-F49620-82-C-0009.

AMS 1980 *subject classification*. Primary 62G15; secondary 60E15, 62F25.

Key words and phrases. Regression, robust smoothing, kernel estimators, uniform convergence rates.

paper we derive—under mild conditions on the weight sequence $\alpha_i^{(n)}(\cdot)$—the uniform consistency of $m_n(\cdot)$ on the interval $I = [0, 1]$. We show that

$$\text{(UC)} \qquad r_n^{-1} \sup_{0 \le t \le 1} | m_n(t) - m(t) | = O_p(1)$$

with rate $r = r_n$. In the derivation of this result we shall need bounds on moments of sums of independent rv's, as given by Whittle (1960), Theorem 2.

The quite general setup of $m_n(x)$ as the solution of (1.2) and $m(x)$ as the solution of (1.1) allows us by tuning $\alpha_i(\cdot)$ and $\psi(\cdot)$ to obtain a wide class of estimators and regression functions as will be shown in the following examples.

One of the following examples (Example 5) will give a partial answer to a question raised by C. J. Stone in his special invited paper on optimal rates of convergence (Stone, 1982, page 1044, Question 4).

EXAMPLE 1. Take $\psi(u) = u$ in both (1.1) and (1.2) and define

$$\alpha_i(x) = n^{-1} h^{-1} K((x - X_i)/h)$$

for kernel $K(\cdot)$ and a sequence of bandwidth $h = h(n)$ tending to zero. The resulting regression curve in the stochastic design case is

$$m_{\psi,F}(x) = m(x) = E(Y \mid X = x)$$

and the estimator is

$$m_n^*(x) = (nh)^{-1} \sum_{i=1}^n K((x - X_i)/h) Y_i / [(nh)^{-1} \sum_{j=1}^n K((x - X_j)/h)].$$

The estimator was proposed independently by Nadaraya (1964) and Watson (1964). Rosenblatt (1969) and Collomb (1977, 1979) computed bias and variance rates. Schuster (1972) demonstrated the multivariate normality at a finite number of distinct points; Schuster and Yakowitz (1979) derived uniform consistency of $m_n^*(x)$ on a finite interval. Recently, Johnston (1979) in his thesis proved a uniform consistency result (with rates) for the related estimator

$$m_n^*(x) \cdot [(nh)^{-1} \sum_{i=1}^n K((x - X_i)/h)/f_X(x)],$$

$f_X(\cdot)$ denoting the marginal density of X. Further (uniform) consistency results for $m_n^*(x)$ were obtained by Major (1973), Konakov (1977), Nadaraya (1973, 1974), Stone (1977) among others. A bibliographic review on the estimation of $m(x) = E(Y \mid X = x)$ may be found in Collomb (1981).

EXAMPLE 2. Take $\psi(u) = u$ and define in the fixed design case

$$\alpha_i(x) = h^{-1} \int_{s_{i-1}}^{s_i} K\left(\frac{x - u}{h}\right) du$$

where $s_0 = 0$, $s_{j-1} \le x_j \le s_j$, $s_n = 1$, $\int K(u)\, du = 1$ and $h = h(n)$ is as above a sequence of bandwidths tending to zero as $n \to \infty$. Since $\sum_{i=1}^n \alpha_i(x) = 1$, the resulting estimator is

$$\bar{m}_n(x) = \sum_{i=1}^n \alpha_i(x) Y_i,$$

first discussed by Gasser and Müller (1979) and recently considered by Cheng and Lin (1981) (with $s_j \equiv x_j$ in $\alpha_j(x)$). The Priestley and Chao (1972) estimator does not fall in the class of estimators here, but is, as shown by Cheng and Lin (1981), also uniform consistent obtaining the same rate as $\bar{m}_n(x)$.

In the following example we will assume symmetry of $f(y \mid x)$. Note that for the results of this paper neither symmetry of $f(y \mid x)$ nor antisymmetry of ψ are required. This assumption is only made to obtain in a convenient way the conditional mean from equation (1.1).

EXAMPLE 3. Take $f(y \mid x)$ respectively $f(y; x)$ be symmetric and ψ a bounded, antiasymmetric function. Then again (1.1) gives for the stochastic design case

$$m_{\psi,F}(x) = m(x) = E(Y \mid X = x)$$

respectively

$$m(x) = \int y f(y; x)\, dy$$

in the fixed design case. The regression curve is thus a quantity $m_{\psi,F}(x)$ which minimizes (w.r.t. θ)

$$\int \rho(Y - \theta) f(y \mid x)\, dy$$

where we assume ρ to be positive, even, convex and differentiable with derivative $\rho' = \psi$. This is exactly the notation of a M-functional (Bickel and Lehmann, 1975, page 1053) and shows that $m_n(x)$ from (1.2) (with weights $\{\alpha_i(x)\}$ as in example 1 or example 2) is a robust estimator of $m(x)$.

In the stochastic design case $m_n(x)$ is a solution (with respect to θ) of

$$n^{-1}h^{-1} \textstyle\sum_{i=1}^{n} K((x - X_i)/h)\psi(Y_i - \theta) = 0.$$

In the fixed design case the estimator is a solution (with respect to θ) of

$$h^{-1} \sum_{i=1}^{n} \left[\int_{i-1}^{s_i} K\!\left(\frac{x - u}{h}\right) du \right] \psi(Y_i - \theta) = 0.$$

Pointwise consistency and asymptotic normality along with some numerical results are shown in Härdle (1983) and Härdle and Gasser (1982). In the last paper it is also shown that the robust estimator $m_n(x)$ proves to be useful in the evaluation of Laser spectra (Raman spectra). If we take for instance

$$\psi(u) = \max\{-\kappa, \min\{u, \kappa\}\}, \quad \kappa \geq 0$$

we obtain a Huber-type (Huber, 1964) robust nonparametric regression function estimator. Bias and variance rates for this Huber-type estimator with a uniform window, i.e. $K(u) = I_{[-.5,.5]}(u)$ were computed by Stuetzle and Mittal (1979).

EXAMPLE 4. Taking $\psi(u) = \alpha u^{\alpha-1}$, $u \geq 0$ and $\psi(u) = -\alpha(-u)^{\alpha-1}$, $u < 0$, $1 < \alpha < 2$ allows us by tuning α to steer from the (local) least square estimator, which is $m_n^*(x)$ as $\alpha = 2$, to the (local) median (as $\alpha \to 1$) and vice versa. The whole class of these estimators will also be covered by our theorems.

EXAMPLE 5. Take $\psi(u) = ½ - I(u \le 0)$, a ψ-function leading to the conditional median $m(x) = \text{med}(Y \mid X = x)$ as the regression curve. Stone (1982) raised the question if $\{n^{-r}\}$ ($r = (p - m)/(2p + d)$ in his notation) is still an achievable rate of convergence. The results of this paper give a partial answer to that question. We show that for $p = 1$, $d = 1$, $m = 0$ a subclass of his $\{T(\theta)\}$ indeed, $\{(n^{-1}\log n)^r\}$, the optimal rate of his Theorem 1 (for the type of distance considered here) is achieved. To see this in the "stochastic design model", note that assumption (A4) of Section 2 is trivially fulfilled and assumption (A3) is satisfied if there exists a constant c_0 such that $f(m(x) \mid x) > 2c_0$, $x \in I$. Assumptions (A1) and (A5) are only technical and (A2) is the definition of $m(x) = \text{med}(Y \mid X = x)$. Assume now that m is continuously differentiable so that the modulus of continuity $\omega_m(\delta)$ is linear in δ. Then, Theorem 1 below says that with $r_n \sim h_n$ and $r_n \sim (\log n)^{1/2}(nh_n)^{-1/2}$ uniform consistency of m_n can be achieved with rate $r_n = n^{-1/3} (\log n)^{1/3}$ which is the optimal rate given in Stone (1982). Quite analogous conclusions can be drawn in the fixed design model.

We present the result (UC) for $\alpha_i(x)$ as in example 1 and example 2 for the stochastic design case in Theorem 1 and a following remark. Theorem 2 shows (UC) in the fixed design case with $\alpha_i(x)$ as in example 2. All theorems require a certain amount of smoothness of $m(\cdot)$, expressed through the behaviour of the modulus of continuity of m which we denote by ω_m. These results are improvements over some previous work. Our assumptions are weaker than those of Major (1973) in that Y is not required to be bounded a.s. and our results are stronger than those of Schuster and Yakowitz (1978) because we were able to compute uniform convergence rates for $m_n^*(x)$ as in Mack and Silverman (1982).

2. Results. We will make the following assumptions on the kernel function and on moments of $[\psi(Y - m(x) + s)]$.

(A1) The kernel K is positive, continuously differentiable with compact support

$$[-A, A] \quad \text{and} \quad \int_{-A}^{0} K(u)\, du = \int_{0}^{A} K(u)\, du = \frac{1}{2}.$$

(A2) ψ is a monotone, locally bounded function with $E_x\psi(Y - m(x)) = 0$.

(A3) There are constants $c_0, c_1 > 0$ such that for every $x \in I = [0, 1]$

$$| E_x\psi(Y - m(x) + s) | > c_0 | s |, \quad | s | \le c_1.$$

(A4,k) For some $k \ge 2$ let $\sup_{x\in I} E_x | \psi(Y - m(x) \pm c_1) |^k < \infty$.

(A4,∞) ψ is bounded, $\sup_{u\in\mathbb{R}} | \psi(u) | \le B_\psi < \infty$.

(A5) The marginal density of X is bounded from above and below

$$0 < a \le f_X(u) \le b < \infty \quad \text{for all} \quad u \in I.$$

(A6) There exists a constant C_0 such that for every $x \in I$

$$| E_x\psi(Y - m(x) + s) | < C_0 | s |.$$

Some remarks about the assumptions should be made. The first assumption is very common in nonparametric regression and needs no further explanation (Collomb, 1981). The second assumption is just the proper (implicit) definition of the regression function. Assumption (A3) needs some more motivation. Assume for simplicity that we have a homoscedastic error structure that is $f(y \mid x) = f(y - m(x))$ and $f(y \mid x)$ is symmetric. If we have that $\psi(u) = u$ then (A3) is trivially fulfilled. For the nonlinear ψ functions, (A3) is satisfied if $\mid \int \psi(y + s)/sf(y)\, dy \mid > c_0$ for small s. So (A3) can be interpreted as a criterion for $E_x\psi'(y + s)$, (s small) staying away from zero, provided it exists at all. Assumption (A6) is trivially fulfilled for $\psi(u) = u$. For nonlinear ψ functions (A6) is obviously fulfilled if $\mid \int \psi(y + s)/sf(y)\, dy \mid < C_0$, which can be interpreted as an upper bound for $E_x\psi'(y - m(x) + s)$, s small. We have chosen this quite technical way of formulation to include the conditional median corresponding to $\psi(u) = 1/2 - I(u \le 0)$ which is nondifferentiable at $u = 0$. The assumption (A4,k) will be used for unbounded ψ functions only, (A4,∞) is just the definition of a bounded ψ function making m_n a robust estimator of m.

As already mentioned, the modulus of continuity of m will be denoted by

$$\omega_m(\delta) = \sup_{x \in I} \sup_{|x - x'| < \delta} |m(x) - m(x')|.$$

As long as there is no confusion, the index "n" will be dropped in the sequel.

The following theorems will split up into a statement on unbounded ψ functions (i.e. containing as a special case the Nadaraya-Watson estimator) and a statement on bounded ψ functions. The theorems tell us how we have to choose the sequence $h = h(n)$ in dependence of the sample size n and the rate $r = r_n$ in order to obtain (UC).

We begin with the uniform consistency in the stochastic design case.

THEOREM 1. *Let the data be generated with stochastic design* $\{X_i\}_{i=1}^n$, *and let* $\alpha_i(t) = (nh)^{-1}K((t - X_i)/h)$. *Assume that* (A1)–(A5) *hold and let*

$$\omega_m(2Ah) < r, \quad nh^2/\log n \ge d > 0.$$

If (A4,k) *holds let*

$$nh^{1+2/(k-1)}r^{2+2/(k-1)} \longrightarrow \infty$$

and if (A4,∞) *holds*

$$nhr^2/\log n \ge \xi_1$$

ξ_1 *depending on* c_0, c_1, B_ψ, a, b. *Then* $m_n(x)$ *satisfies* (UC). *If in addition* (A6) *holds, only*

$$nr^2h^{1+2/(k-1)} \longrightarrow \infty$$

and (A4,k) *suffice to establish* (UC).

REMARK. It can be shown that (UC) also holds for the situation described in example 2 for the stochastic design case. Very similar arguments that are used to prove Theorem 2 yield that if

$$\begin{aligned} &nh^{1+2/(k-1)}r^{2+2/(k-1)}/\log n \to \infty &&\text{in case of (A4,k),} \\ &nhr^2/(\log n)^2 \geq \xi_2 &&\text{in case of (A4,}\infty) \end{aligned}$$

the uniform consistency (UC) with rate $r = r_n$ follows.

THEOREM 2. *Let the data be generated with fixed design points* $\{x_i\}_{i=1}^n$, *satisfying* $\sup_i |x_i - x_{i-1}| = O(n^{-1})$ *and set* $\alpha_i(t)$ *as in example* 2. *Assume that* (A1)–(A5) *hold and* $\omega_m(2Ah) < r$. *If* (A4,k) *holds, let*

$$nh^{1+2/(k-1)}r^{2+2/(k-1)} \to \infty$$

and if (A4,∞) *holds*

$$nhr^2/\log n \geq \xi_2$$

ξ_2 *depending on* c_1, c_0, B_ψ, $\int K^2$. *Then* $m_n(x)$ *satisfies* (UC). *If in addition* (A6) *holds, the condition*

$$nh^{1+2/(k-1)}r^2 \to \infty$$

together with (A4,k) *suffice to establish* (UC).

3. Proofs. To show that the class of estimators defined through (1.2) satisfies (UC) for the various choices of ψ-functions and weights $\{\alpha_i(x)\}_{i=1}^n$, we have to show that

$$P\{\sup_{x\in I} |m_n(x) - m(x)| > r_n\}$$

is arbitrarily small. Now by monotonicity of ψ, this can be estimated by

$$P(\Omega_n) + P(\Omega_n')$$

where

$$\Omega_n = \{\sup_{x\in I} g_n(x, -r) \geq 0\}, \quad \Omega_n' = \{\inf_{x\in I} g_n(x, r) \leq 0\}$$

and

$$g_n(x, s) = \sum_{i=1}^n \alpha_i(x)\psi(Y_i - m(x) + s).$$

By the symmetry of the problem it will suffice to consider $P(\Omega_n)$.

The principal idea of the proof is to lay an equidistant mesh $0 = t_0 < t_1 < \cdots < t_{\ell_n} = 1$, where $\ell_n \ll n$, to sum the probabilities at the meshpoints and to use the mean value theorem applied to $\alpha_i(t)$ between them. More precisely we have

$$\begin{aligned} P(\Omega_n) \leq\ & \ell_n \sup_{t=t_j} P[\{g_n(t, -r_n/2) \geq -\eta_n\} \cap M_n] \\ &+ \ell_n \sup_{t=t_j} P[\{\sup_{|u-t|<\ell_n^{-1}} g_n(u, -r_n/2) > \eta_n + g_n(t, -r_n/2)\} \cap M_n] \\ &+ P(M_n^c) = \ell_n[\sup_{t=t_j} U_{1n}(t) + \sup_{t=t_j} U_{2n}(t)] + U_{3n}, \end{aligned}$$

where ℓ_n, η_n are arbitrary sequences to be specified later and M_n is an arbitrary set to be chosen for the different cases (stochastic/fixed design and the particular

$\{\alpha_i(x)\}_{i=1}^n$). We will also make use of the following fact that in the fixed design

$$\sum_{i=1}^n \alpha_i^2(t) = O(n^{-1}h^{-1}) \quad \text{uniformly in} \quad t$$

(Gasser and Müller, 1979) and in the stochastic design

$$\sum_{i=1}^n E\alpha_i^2(t) = O(n^{-1}h^{-1}) \quad \text{uniformly in} \quad t$$

(Johnston, 1979).

PROOF OF THEOREM 1. Suppose that (A4,k) holds, then with $\eta_n = \beta r_n$, β small enough to satisfy the assumptions of Lemma 1, we obtain from (A.1)

$$\sup_{t=t_j} U_{1n}(t) \le \nu_1 r^{-k}(nh)^{-k/2},$$

and if $\ell_n^{-1} < Ah$ we have from (A.3)

$$\sup_{t=t_j} U_{2n}(t) \le \nu_2(r\ell_n)^{-k}[h^{-k} + (nh^3)^{-k/2}],$$

and if (A6) holds

$$\sup_{t=t_j} U_{2n}(t) \le \nu_2'(r\ell_n)^{-k}[h^{-k}r^{-k} + (nh^3)^{-k/2}]$$

where ν denote large constants and M_n is chosen as in Lemma 3. Then with $\ell_n^2 = nh^{-1}$ (such that $h^{-1} \ll \ell_n \ll n$) we have from Lemma 3

$$\begin{aligned} P(\Omega_n) &\le \mu_1\ell_n\{(nhr^2)^{-k/2} + (\ell_n^2r^2h^2)^{-k/2} + [nh(\ell_n^2r^2h^2)]^{-k/2}\} \\ &\quad + \mu_2\ell_n\exp(-\mu_3nh^2) \\ &\le \mu_4(nh^{-1})^{1/2}(nhr^2)^{-k/2} + \mu_2\exp(\log n - \mu_3nh^2) \end{aligned}$$

which is small by the assumptions of the theorem. A similar inequality shows that if (A6) is fulfilled, $P(\Omega_n)$ can be made arbitrarily small. Suppose now that (A4,∞) holds and choose ℓ_n such that

$$\sum_{i=1}^n |\alpha_i'(t)| \le \sup|K'|\, n^{-1}h^{-2}4b\varepsilon^{-1}n(Ah + \ell_n^{-1}) \le \eta_n\ell_n/(2B_\psi)$$

to fulfill the assumptions of Lemma 2. This can be made for $\eta_n = \beta r_n$ and $r\ell_nh \ge \mu_5$, μ_5 a large constant. So we get by Lemma 1 and Lemma 2

$$P(\Omega_n) \le \mu_6\exp(-\mu_7r^2nh - \log r - \log h) + \mu_8\exp(-\varepsilon^{-1}nh^2 - \log r - \log h)$$

which is small by the assumptions of the theorem.

PROOF OF THEOREM 2. Define $M_n = \{(x_1, x_2, \cdots, x_n)\colon \sup_{2\le i\le n-1}|s_{i+1} - s_{i-1}| < \gamma/n\}$. If γ is chosen large enough we have that $M_n^c = \phi$ by assumption on the fixed design. Since

$$\alpha_i(t) = h^{-1}\int_{s_{i-1}}^{s_i} K\left(\frac{t-u}{h}\right) du$$

we have

$$\sum_{i=1}^{n} \alpha_i(t) = \int_{(t-Ah,t+Ah)\cap I} K(u)\, du \geq \frac{1}{2}$$

$$\sum_{i=1}^{n} |\alpha_i(t)|^2 \leq h^{-1}\sup|s_{i+1} - s_{i-1}| \int |K^2|$$

$$\sum_{i=1}^{n} |\alpha_i'(t)| \leq h^{-1} \int |K'|.$$

Choosing $\eta_n = \beta r_{n'}$ β small enough, we get from (A1) and (A4,k)

$$U_{1n}(t) \leq \mu_9 r^{-k}[1/(nh)]^{k/2}$$

$$U_{2n}(t) \leq \mu_{10}[(\ell_n rh)^{-k} + (1/(nh^3))^{k/2}(r\ell_n)^{-k}].$$

Respectively if (A6) holds

$$U_{2n}(t) \leq \mu_{11}[(\ell_n h)^{-k} + (1/(nh^3))^{k/2}(r\ell_n)^{-k}]$$

taking $\ell_n^2 = nh^{-1}$ respectively $\ell_n^2 = nh^{-1}r^2$ (for the (A6) case) shows that $P(\Omega_n)$ is small. Now in the case that ψ is bounded we see that

$$\ell_n = \varepsilon^{-1}B_\psi \int |K'|/(c_0 h), \quad \varepsilon \text{ small}$$

ensures

$$U_{2n}(t) = 0$$

and

$$U_{1n}(t) \leq \mu_{12} \exp(-\mu_{13} r^2 nh - \log h).$$

APPENDIX

It is shown here how the terms $U_{1n}(t)$, $U_{2n}(t)$, U_{3n} may be estimated in the different cases (stochastic design, fixed design). Lemma 1 and Lemma 2 are shown for the fixed design case only. The proofs for the stochastic design case are essentially the same by conditioning on $\{X_1, \cdots, X_n\}$.

LEMMA 1. *Suppose that the modulus of continuity of $m(\cdot)$ satisfies $\omega_m(Ah_n) \leq r_n/4$ and let $\eta_n \leq c_0\delta r_n/8$ where δ is a small constant, c_0 is the constant of* (A3) *and $M_n \subset \{\sum_{i=1}^{n} \alpha_i(t) > \delta\}$. Then if* (A4,k) *holds*

$$U_{1n}(t) \leq \eta_n^{-k}\Lambda_k^{(1)}[\sup_{x\in M_n}(\textstyle\sum_{i=1}^{n} \alpha_i^2(t))^{k/2}] \times \sup_{0\leq x\leq 1} E_x(|\psi(Y - m(x) \pm c_1)^k). \tag{A.1}$$

Otherwise if (A4,∞) *holds*

$$U_{1n}(t) \leq \exp[-\Lambda_\infty^{(1)}\eta_n^2(B_\psi^2 \textstyle\sum_{i=1}^{n} \alpha_i^2(t))^{-1}] \tag{A.2}$$

where $\Lambda_\cdot^{(1)}$ denote constants.

PROOF. Using the assumption on ω_m and the monotonicity of ψ near the origin we have

$$E_{x_i}\psi(Y_i - m(t) - r_n/2) < E_{x_i}\psi(Y_i - m(x_i) - r_n/4)$$

for all $i \in \{j: |t - x_j| < Ah\}$ and therefore

$$\sum_{i=1}^n \alpha_i(t)[\psi(Y_i - m(t) - r_n/2) - E_{x_i}\psi(Y_i - m(t) - r_n/2)] > c_0 r_n/4 \sum_{i=1}^n \alpha_i(t) + g_n(t, -r_n/2)$$

by assumption (A3). So by Chebychev's inequality and Theorem 2 of Whittle (1960) we have that

$$\begin{aligned} U_{1n}(t) &\le P\{\textstyle\sum_{i=1}^n \alpha_i(t)[\psi(Y_i - m(t) - r_n/2) - E_{x_i}\psi(Y_i - m(t) - r_n/2)] > \eta_n\} \\ &\le \eta_n^{-k}\lambda_k[\textstyle\sum_{i=1}^n \alpha_i^2(t)]^{k/2} \\ &\quad \times \sup_{|t-x_i|<Ah} E_{x_i}[|\psi(Y - m(t) - r_n/2) - E_{x_i}\psi(Y_i - m(t) - r_n/2)|^k] \end{aligned}$$

in the case that (A4,k) is used. Otherwise, if (A4,∞) holds, by an easy extension of Whittle's Theorem 2

$$U_{1n}(t) \le \exp[-\eta_n^2(4eB_\psi^2 \textstyle\sum_{i=1}^n \alpha_i^2(t))^{-1}]$$

for bounded ψ functions which shows that (A.1), (A.2) hold. □

The next lemma estimates $U_{2n}(t)$.

LEMMA 2. *Suppose that the modulus of continuity of* $m(\cdot)$ *satisfies* $\omega_m(\ell_n^{-1} + Ah) < r_n/2$. *Then, if* (A4,k) *holds*

$$\begin{aligned} U_{2n}(t) &\le \eta_n^{-k}\ell_n^{-k}\Lambda_k^{(2)}\{\sup_{\mathbf{x}\in M_n, u\in I}|\textstyle\sum_{i=1}^n \alpha_i'(u)|^k \sup_{u\in I} E_x|\psi(Y - m(x) - r_n)|^k \\ &\quad + \sup_{\mathbf{x}\in M_n, u\in I}[\textstyle\sum_{i=1}^n [\alpha_i'(u)]^2]^{k/2} \sup_{u\in I} E_x|\psi(Y - m(x) \pm c_1)|^k\}, \end{aligned}$$

$\Lambda_k^{(2)}$ *a constant.*

On the other hand if (A4,∞) is true, then $U_{2n}(t) = 0$ *provided that*

$$M_n \subset \mathscr{M}_n = \{\textstyle\sum_{i=1}^n |\alpha_i'(t)| < \eta_n \ell_n/[2B_\psi]\}.$$

PROOF. By the assumption on the modulus of continuity of $m(\cdot)$ and the mean value theorem we conclude that

$$U_{2n}(t) \le P\left\{\int_{\Gamma_n(t)} \left| \textstyle\sum_{i=1}^n \alpha_i'(u)\psi(Y_i - m(t) - r_n/2) \right| du > \eta_n\right\}$$

where $\Gamma_n(t) = \{u: |u - t| \le \ell_n^{-1}\}$. This already shows that $U_{2n}(t) = 0$ if $M_n \subset \mathscr{M}_n$ and (A4,∞) holds.

We now further estimate the RHS of the inequality above using Chebyshev's inequality. We then have

$$U_{2n}(t) \le \Lambda_k^{(2)} \eta_n^{-k} \Big\{ E \Big| \int_{\Gamma_n(t)} \sum_{i=1}^n \alpha_i'(u) E_{x_i} \Big| \psi(Y_i - m(t) - r_n/2) | \, du |^k$$

$$+ E \Big| \int_{\Gamma_n(t)} \Big| \sum_{i=1}^n \alpha_i'(u) [\psi(Y_i - m(t) - r_n/2)$$

$$- E_{x_i} \Big| \psi(Y_i - m(t) - r_n/2) |] | \, du \Big|^k \Big\}$$

$$= V_{1n} + V_{2n}, \quad \text{say}.$$

Now by Hölder's inequality (with $p = k$) and Theorem 2 of Whittle (1960), we have

$$V_{2n} \le \left[\int_{\Gamma_n(t)} du \right]^{k-1} E \int_{\Gamma_n(t)} | \sum_{i=1}^n \alpha_i'(u) [\psi(Y_i - m(t) - r_n/2)$$

$$- E_{x_i} | \psi(Y_i - m(t) - r_n/2) |] |^k \, du$$

$$\le [2\ell_n^{-1}]^k \{ \sup_{u \in I, \mathbf{x} \in M_n} [\sum_{i=1}^n [\alpha_i'(u)]^2]^{k/2} 2^k$$

$$\times \sup_{u \in I, |u - x_i| < Ah + \ell_n^{-1}} E_{x_i} | \psi(Y_i - m(u) - r_n/2) |^k \}.$$

Applying now the assumption on the modulus of continuity, we have the desired upper bound for both V_{1n} and V_{2n} (after an application of Hölder's inequality to V_{1n}, too). □

In the following lemma we estimate the term U_{3n} for different sets M_n.

LEMMA 3. *Let*

$M_n = \{(X_1, \cdots, X_n): \sum_{i=1}^n \alpha_i(t_j) > a/4$ *and*

$\#\{ |X_i - t_j| < Ah + \ell_n^{-1} \} < 4bn(Ah + \ell_n^{-1})\varepsilon^{-1}$ *for* $j = 0, \cdots, \ell_n, 0 < \varepsilon \le 1\}$

in the stochastic design case. Then

$$U_{3n} \le \Lambda^{(3)} \ell_n \exp[-\lambda_3 \varepsilon^{-1} n (h_n^2 + \ell_n^{-2})],$$

where $\Lambda^{(3)}$, λ_3 *are constants.*

PROOF. Since $E\alpha_i(t) = n^{-1} \int_{[-Ah+t, Ah+t] \cap I} K(u) f_x(t + uh) \, du \ge a(2n)^{-1}$, $\sum_{i=1}^n \alpha_i(t) \le a/4$ implies $| \sum_{i=1}^n [\alpha_i(t) - E\alpha_i(t)] | > a/4$. Now by Whittle's theorem we have

$$P(| \sum_{i=1}^n [\alpha_i(t) - E\alpha_i(t)] | > a/4) \le \exp\left(-\lambda_1 \frac{\sup K^2}{\inf f_X^2} n \right), \quad \lambda_1 = \text{const}.$$

On the other hand

$$\#\{|X_i - t| < Ah + \ell_n^{-1}\} = \sum_{i=1}^n I_{\Delta_n(t)}(X_i) = n^{-1} \sum_{i=1}^n \bar{Z}_i$$

where

$$\Delta_n(t) = \{u: |u - t| < Ah + \ell_n^{-1}\}.$$

Since $n^{-1}E\bar{Z}_i(t) = \int_{\Delta_n(t)} f_X(u)\, du \le 2b\,(Ah + \ell_n^{-1})$ we have by Whittle's theorem that

$$P(n^{-1} \sum_{i=1}^n \bar{Z}_i \ge 4bn(Ah + \ell_n^{-1})\varepsilon^{-1}) \le \exp[-\lambda_2 n(Ah + \ell_n^{-1})], \quad \lambda_2 = \text{const.} \quad \square$$

Acknowledgment. The authors would like to thank C. Jennen, R. Lerche and B. Silverman for many fruitful comments and suggestions. We also want to thank the referees for substantial improvements of the paper.

REFERENCES

BICKEL, P. J. and LEHMANN, E. L. (1975). Descriptive statistics for nonparametric models II. *Ann. Statist.* **3** 1045–1069.

CHENG, KUANG FU and LIN, PI-ERH. (1981). Nonparametric estimation of a regression function. *Z. Wahsch. verw. Gebiete* **57** 223–233.

COLLOMB, G. (1979). Quelques propriétés de la methode du noyau pour l'estimation non-paramétrique de la régression en un point fixé. *C. R. Acad. Sci. Paris* **285A** 289–292.

COLLOMB, G. (1979). Conditions nécessaires et suffisantes de convergence uniforme d'un estimateur de la régression, estimation des dérivées de la régression. *C. R. Acad. Sci. Paris* **288A** 161–164.

COLLOMB, G. (1981). Estimation non-paramétrique de la régression: revue bibliographique. *ISR* 75–93.

GASSER, T. and MÜLLER, H. G. (1979). Kernel estimation of regression functions. *Springer Lecture Notes in Math.* **757**.

HÄRDLE, W. and GASSER, T. (1983). Robust nonparametric function fitting. *J. Roy. Statist. Soc. (B),* to appear.

HÄRDLE, W. (1984). Robust regression function estimation. *J. Multivariate Anal.*, to appear.

HUBER, P. J. (1964). Robust estimation of location. *Ann. Math. Statist.* **35** 73–101.

JOHNSTON, G. J. (1979). Smooth nonparametric regression analysis. Ph.D. dissertation, University of North Carolina, Chapel Hill.

KONAKOV, V. D. (1977). On a global measure of deviation for an estimate of the regression line. *Theor. Probab. Appl.* **22** 858–868.

MACK, Y. P. and SILVERMAN, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. verw. Gebiete* **61** 405–415.

MAJOR, P. (1973). On a non-parametric estimation of the regression function. *Stud. Sci. Math. Hung.* **8** 347–361.

NADARAYA, E. A. (1964). On estimating regression. *Theor. Probab. Appl.* **9** 141–142.

NADARAYA, E. A. (1973). Some limit theorems related to nonparametric estimates of regression curve (in Russian). *Bull. Acad. Sci. Georgian S.S.R.* 71n n°, 57–60.

NADARAYA, E. A. (1974). The limit distribution of the quadratic deviation of nonparametric estimates of the regression function. *Soobshch. Akad. Nauk. Gruz. SSR* 74, 33–36 (in Russian).

PRIESTLEY, M. B. and CHAO, M. T. (1972). Nonparametric function fitting. *J. Roy. Statist. Soc. B* **34** 385–392.

ROSENBLATT, M. (1969). Conditional probability density and regression estimates. In *Multivariate Analysis* II, ed. Krishnaiah.

SCHUSTER, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.* **43** 84–88.

SCHUSTER, E. F. and YAKOWITZ, S. (1979). Contributions to the theory of nonparametric regression, with application to system identification. *Ann. Statist.* **7** 139–145.
STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620.
STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.
STUETZLE, W. and MITTAL, Y. (1979). Some comments on the asymptotic behavior of robust smoothers in *Smoothing Techniques for Curve Estimation* (ed. T. Gasser and M. Rosenblatt). *Springer Lecture Notes* **757**, Heidelberg.
WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theor. Probab. Appl.* **5** 302–305.

DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF FRANKFURT
D-6000 FRANKFURT/MAIN
WEST GERMANY

SONDERFORSCHUNGSBEREICH 123
IM NEUENHEIMER FELD 293
UNIVERSITY OF HEIDELBERG
D-6900 HEIDELBERG
WEST GERMANY