The Annals of Statistics
1979, Vol. 7, No. 1, 139–149

CONTRIBUTIONS TO THE THEORY OF NONPARAMETRIC REGRESSION, WITH APPLICATION TO SYSTEM IDENTIFICATION

BY E. SCHUSTER AND S. YAKOWITZ

The University of Texas at El Paso and The University of Arizona

The objective in nonparametric regression is to infer a function $m(x)$ on the basis of a finite collection of noisy pairs $\{(X_i, m(X_i) + N_i)\}_{i=1}^n$, where the noise components N_i satisfy certain lenient assumptions and the domain points X_i are selected at random. It is known a priori only that m is a member of a nonparametric class of functions (that is, a class of functions like $C[0, 1]$ which, under customary topologies, does not admit a homeomorphic indexing by a subset of a Euclidean space).

The main theoretical contribution of this study is to derive uniform convergence bounds and uniform consistency on bounded intervals for the Nadaraya-Watson kernel estimator and its derivatives. Also, we obtain the corresponding convergence results for the Priestly-Chao estimator in the case that the domain points are nonrandom. With these developments we are able to apply nonparametric regression methodology to the problem of identifying noisy time-varying linear systems.

1. Introduction. Let (X_1, Y_1), (X_2, Y_2), $\cdots$ be 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.

The nonparametric regression problem is the problem of estimating the conditional expectation $m(x) \equiv E[Y|X = x]$. Equivalently, the nonparametric regression problem requires finding $m(\cdot)$, given observations

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

the N_i being an independent mean-zero noise variable which may depend on X_i. We assume throughout a finite variance of Y, and we will later insist on some regularity conditions for the "target" function $m(\cdot)$. The "nonparametric" property refers to the absence of a finite dimensional continuous parameterization of the space of functions containing $m(\cdot)$. Such a finite dimensional parameterization characterizes classical regression theory (described, for example, in Wilks (1962, Section 10.3)).

Algorithms for nonparametric regression (NPR) are applicable to a great many engineering activities which require estimating performance levels of a system as a function of randomly chosen input values, assuming that the system response is random or measured by instruments which introduce random error. (For example,

Received March 1977; revised January 1978.
AMS 1970 *subject classification.* 62G05.
Key words and phrases. Nonparametric regression, derivatives of regression functions, system identification.

consider consumer response at a given price level. The consumer response (expected number of sales) is only probabilistically related to the number of sales over a certain period.)

In the case of multivariate X_i values, NPR is relevant to estimating the shape of ore bodies on the basis of core samples or the pressure head in oil or water aquifers on the basis of well information. Use of NPR has been made in actuary science (Gartside (1975)) and stock market prediction (Butler (1975)). Watson (1964) obtains an estimate of blood pressure as a function of age.

The main subject of our theoretical investigation will be the Nadaraya-Watson estimator (Nadaraya (1964), Watson (1964)) which we will refer to as the kernel estimator. Rather complete references to the basic NPR schemes can be found in Stone (1977).

The kernel estimator is given by

$$(1) \qquad m_n(x) = (\Sigma_{i=1}^n Y_i k((x - X_i)/a_n)/a_n) / (\Sigma_{i=1}^n k((x - X_i)/a_n)/a_n).$$

In the above, a_n is a sequence of positive numbers which converge to 0. $k(\cdot)$ is a probability density function. The numerator in (1) is an estimator of $\int y f(x, y)\, dy$ and the denominator is a Parzen (1962) estimator of the marginal density f_X. For work on kernel estimators see also Schuster (1968), (1972), Rosenblatt (1969), Nadaraya (1970), Benedetti (1974), (1975) and Butler (1975).

We now summarize the results of this study. The following assumptions will be in force:

(i) $E[Y^2] < \infty$,
(ii) $g(x) > 0$, all x, g being the marginal density of X.
(iii) $m(x)$ has a continuous Nth derivative.
(iv) If the characteristic function of k is ψ, then $\int |u|^N |\psi(u)| du < \infty$.

In the section to follow, we give conditions under which for any $\varepsilon > 0$, $0 \leqslant j \leqslant N$, and any finite closed interval $[a, b]$, there is a constant C such that for any positive ε and n sufficiently large

$$P\left[\max_{a \leqslant x \leqslant b} |m_n^{(j)}(x) - m^{(j)}(x)| > \varepsilon\right] < C / \left(n a_n^{2j+2} \varepsilon\right).$$

In the above expression, $m_n^{(j)}$ and $m^{(j)}$ denote the jth derivatives. The main ideas of the proof follow from Schuster's (1968) Ph.D. dissertation where a more general result is given in the case $N = 0$. In Section 3 we will see that the same convergence result holds if the domain points X_i of the sample are not random, but form a partition of $[a, b]$. We believe that this is the first study to establish convergence rates for the kernel estimator when Y is not bounded or to show that the derivatives of the kernel estimator are consistent estimates of the corresponding derivatives of the target functions. Of the studies known to us, Fisher and Yakowitz (1976) obtain convergence bounds for an NPR scheme (the potential function method), and in that case, bounded convergence was obtained only for a complicated norm. Nadaraya (1970) obtains convergence bounds for the kernel estimator under the assumption that Y is bounded. Hansen and Pledger (1976) first thought about estimating derivatives of regression functions without mentioning any potential applications. We apply the derivative estimators to the problem of identifying parameters of time-varying linear systems in the final portion of this study.

2. Convergence rates for the Watson estimator and its derivatives. The following notation is used. $(X_1, Y_1), \cdots,$ 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 function is f. Furthermore, with respect to some sequence of positive $\{a_n\}$, we define

$$\begin{aligned} w_n(x) &= \Sigma_{s=1}^n Y_s k((x - X_s)/a_n)/(na_n), \\ g(x) &= \int f(x, y)\, dy, \\ w(x) &= \int y f(x, y)\, dy, \\ m(x) &= w(x)/g(x) = E(Y|X = x), \\ \psi(x) &= \int \exp(iux) k(u)\, du, \quad \text{where} \quad i^2 = -1, \end{aligned}$$

that is, $\psi(x)$ is the characteristic function of k.

$$\phi_n(u) = 1/n \Sigma_{s=1}^n Y_s \exp(iuX_s) = \int\int y e^{iux} dF_n(x, y),$$

where F_n denotes the empirical distribution function associated with $(X_1, Y_1), \cdots, (X_n, Y_n)$.

$$\begin{aligned} g_n(x) &= (na_n)^{-1} \Sigma_{i=1}^n k((x - X_i)/a_n), \\ m_n(x) &= w_n(x)/g_n(x), \quad g_n(x) > 0 \\ &= 0, \qquad \text{otherwise}. \end{aligned}$$

If h is a real function having a kth derivative at a point x, $h^{(k)}(x)$ will denote that derivative.

In the following we do not attempt to prove our theorems under the most general conditions, but rather try to establish the results under conditions covering most applications while utilizing developments in the published literature.

LEMMA 1.. *Let k be a continuous density satisfying the condition*

$$\lim_{|u| \to \infty} |uk(u)| = 0,$$

and let $\{a_n\}$ be a sequence of positive numbers converging to zero. If there exists an open interval containing the bounded, closed interval $[a, b]$ on which the real function h is continuous and if $\int_c^d |h(u)|\, du$ is finite, for $-\infty \leqslant c < a < b \leqslant d \leqslant \infty$, then

$$\lim_{n \to \infty} \sup_{x \in [a, b]} \left| \int_c^d a_n^{-1} k(u/a_n) h(x - u)\, du - h(x) \right| = 0.$$

PROOF. Let $Z_n(x, u) = a_n^{-1}k(u/a_n)\{h(x - u) - h(x)\}$ and let $I = [x - d, x - c]$, then

$$|\int_c^d a_n^{-1}k((x - u)/a_n)\{h(u) - h(x)\}du| = |\int_I Z_n(x, u)\, du|$$
$$\leqslant |\int_{I\cap\{|u|\leqslant\delta\}} Z_n(x, u)du| + |\int_{I\cap\{|u|>\delta\}} Z_n(x, u)\, du|$$
$$\leqslant \sup_{|u|\leqslant\delta}|h(x - u) - h(x)| + |\int_{I\cap\{|u|>\delta\}} u^{-1}(u/a_n)\, k(u/a_n)\, h(x - u)\, du|$$
$$+ |h(x)|\int_{|u|>\delta/a_n} k(u)du$$
$$\leqslant \sup_{|u|\leqslant\delta}|h(x - u) - h(x)| + \delta^{-1}\sup_{|t|>\delta/a_n}|tk(t)|\int_c^d|h(u)|\, du$$
$$+|h(x)|\int_{|u|>\delta/a_n} k(u)\, du.$$

Let

$$A = \int_c^d|h(x)|\, dx \quad\text{and}\quad B = \sup_{a\leqslant x\leqslant b}|h(x)|.$$

Then

$$\sup_{a\leqslant x\leqslant b}|\int_I Z_n(x, u)\, du| \leqslant \sup_{a\leqslant x\leqslant b}\sup_{|u|<\delta}|h(x - u) - h(x)|$$
$$+A\delta^{-1}\sup_{|t|>\delta/a_n}|tk(t)| + B\int_{|u|>\delta/a_n} k(u)\, du$$

which tends to zero as we first let n tend to infinity and then let δ tend to zero. Now $h(x) = \int h(x)a_n^{-1}k(u/a_n)\, du$, so the proof can be completed by observing that

$$\sup_{a\leqslant x\leqslant b}|\int_{x-c}^{\infty} h(x)a_n^{-1}k(u/a_n)\, du + \int_{-\infty}^{x-d} h(x)a_n^{-1}k(u/a_n)\, du|$$
$$\leqslant B\Big(\int_{(a-c)/a_n}^{\infty} k(u)\, du + \int_{-\infty}^{(b-d)/a_n} k(u)\, du\Big)$$

which tends to zero as n tends to infinity (recall $c < a < b < d$).

COROLLARY 1. *Let k be a univariate probability density function with* $\lim_{|u|\to\infty}|uk(u)| = 0$, *and let $k^{(s)}$ be a continuous function of bounded variation for $s = 0, 1, \cdots, N$. If $w^{(j)}$ is of bounded variation for $j = 0, 1, \cdots, N$ and if there exists an open interval containing $[a, b]$ on which $w^{(N)}$ is continuous, then*

$$\lim_{n\to\infty}\sup_{[a, b]}|E[w_n^{(j)}(x) - w^{(j)}(x)]| = 0 \qquad \text{for } j = 0, 1, \cdots, N.$$

PROOF.

$$E[w_n(x)] = a_n^{-1}E[Y_1k((x - X_1)/a_n)]$$
$$= a_n^{-1}\int w(u)\, k((x - u)/a_n)\, du$$
$$= a_n^{-1}\int w(x - v)\, k(v/a_n)\, dv \to_n w(x),$$

uniformly on $[a, b]$, by Lemma 1.

For $j = 1$,

$$E[w_n^{(1)}(x)] = a_n^{-2}\int w(u)\, k^{(1)}((x - u)/a_n)\, du$$
$$= a_n^{-1}\big[(w(x - v)\, k(v/a_n))|_{-\infty}^{\infty} + \int w^{(1)}(x - v)\, k(v/a_n)\, dv\big]$$
$$\to_n w^{(1)}(x),$$

uniformly on $[a, b]$. The argument for higher derivative follows by repeating the above technique.

The following is Lemma 2.4 of Schuster (1969), page 1189.

LEMMA 2. *Let k be a univariate probability density function with $\int |uk(u)|\, du$ finite, and let $k^{(j)}$ be a continuous function of bounded variation for $j = 0, 1, \cdots, N$. If g and its first $N + 1$ derivatives are bounded, and if ϵ_n is a sequence of positive numbers such that $\lim_{n\to\infty} a_n/\epsilon_n = 0$, then there exist positive constants C_1 and C_2 such that*

$$P\{\sup_x |g_n^{(j)}(x) - g^{(j)}(x)| > \varepsilon_n\} \leqslant C_1 \exp(-C_2 n \varepsilon_n^2 a_n^{2j+2})$$

for $j = 0, 1, \cdots, N$.

REMARK. If we only assume g and its first $N + 1$ derivatives are bounded in an open interval containing $[a, b]$, then the conclusion of Lemma 2 holds for the sup over $[a, b]$.

If we assume that the characteristic function ψ is absolutely integrable, then by the standard inversion formula we may write

$$w_n(x) = (2\pi)^{-1} \int e^{-iux} \psi(a_n u) \phi_n(u)\, du.$$

Hence,

$$w_n^{(j)}(x) = (2\pi)^{-1} \int (-iu)^j e^{-iux} \psi(a_n u) \phi_n(u)\, du, \qquad 0 \leqslant j \leqslant N. \tag{2}$$

LEMMA 3. *Suppose k and w satisfy the conditions of Corollary 1 and suppose $u^j \psi(u)$ is absolutely integrable. If $E(Y^2) < \infty$, then there exists a constant $C > 0$ such that for every positive ϵ and for n sufficiently large,*

$$P\{\sup_{[a,b]} |w_n^{(j)}(x) - w^{(j)}(x)| > \varepsilon\} \leqslant C/(n a_n^{2j+2} \varepsilon^2)$$

for $j = 0, 1, \cdots, N$.

PROOF. Using the representation (2) for $w_n^{(j)}$, we see that

$$\begin{aligned}\sup_{[a,b]} &|w_n^{(j)}(x) - E[w_n^{(j)}(x)]| \\ &= \sup_{[a,b]} |(2\pi)^{-1} \int u^j e^{-iux} \psi(a_n u) \{\phi_n(u) - E[\phi_n(u)]\}\, du| \\ &\leqslant (2\pi)^{-1} \int |u^j \psi(a_n u) \{\phi_n(u) - E[\phi_n(u)]\}|\, du.\end{aligned}$$

Thus,

$$\begin{aligned}\sup_{[a,b]} &|w_n^{(j)}(x) - E[w_n^{(j)}(x)]|^2 \\ &\leqslant (\tfrac{1}{4}\pi^2)(\int |u^j \psi(a_n u) \{\phi_n(u) - E[\phi_n(u)]\}|\, du)^2 \\ &= (\tfrac{1}{4}\pi^2) \int\int |u^j v^j \psi(a_n u) \psi(a_n v) \{\phi_n(u) - E[\phi_n(u)]\} \\ &\qquad \times \{\phi_n(v) - E[\phi_n(v)]\}|\, du\, dv\end{aligned}$$

so that by the Schwarz inequality we have that

$$E\left[\sup_{[a,b]}|w_n^{(j)}(x) - E[w_n^{(j)}(x)]|^2\right]$$

$$\leqslant \left(\tfrac{1}{4}\pi^2\right)\int\int|u^j v^j \psi(a_n u)\psi(a_n v)|E^{\frac{1}{2}}\left[|\phi_n(u) - E\phi_n(u)|^2\right] \cdot E^{\frac{1}{2}}\left[|\phi_n(v) - E\phi_n(v)|^2\right] du\, dv$$

$$= \left\{(2\pi)^{-1}\int|u^j\psi(a_n u)|E^{\frac{1}{2}}\left[|\phi_n(u) - E\phi_n(u)|^2\right] du\right\}^2.$$

Now

$$E\left[|\phi_n(u) - E\phi_n(u)|^2\right] = (n^{-2})E|\Sigma_{s=1}^n\left(Y_s e^{iuX_s} - E\left[Y_s e^{iuX_s}\right]\right)|^2$$
$$= (n^{-1})E\left[|Ye^{iuX} - E[Ye^{iuX}]|^2\right]$$
$$= (n^{-1})\left(E|Ye^{iuX}|^2 - |EYe^{iuX}|^2\right)$$
$$= (n^{-1})\left(E[Y^2] - |E[Ye^{iuX}]|^2\right) \leqslant E[Y^2]/n.$$

Thus we have shown

$$E\left[\sup_{[a,b]}|w_n^{(j)}(x) - E[w_n^{(j)}(x)]|^2\right]$$
$$\leqslant \left(\tfrac{1}{4}\pi^2\right)E[Y^2]\left(\int|u^j\psi(a_n u)|\, du\right)^2/n$$
$$= \left(\tfrac{1}{4}\pi^2\right)E[Y^2]\left(\int|u^j\psi(u)|\, du\right)^2/\left(na_n^{2+2j}\right).$$

By Corollary 1, Tchebychev's inequality, and the above inequality, we have for n sufficiently large

$$P\left\{\sup_{[a,b]}|w_n^{(j)}(x) - w^{(j)}(x)| > \varepsilon\right\}$$
$$\leqslant P\left\{\sup_{[a,b]}|w_n^{(j)}(x) - E[w_n^{(j)}(x)]| > \varepsilon/2\right\}$$
$$\leqslant (4/\varepsilon^2)E\sup_{[a,b]}|w_n^{(j)}(x) - Ew_n^{(j)}(x)|^2 \leqslant C_1/na_n^{2j+2}$$

where

$$C_1 = 4E[Y^2]\left(\int|u^j\psi(u)|\, du\right)^2/\varepsilon^2$$

and the proof is complete.

THEOREM 1. *Assume that the conditions of Lemmas* 2 *and* 3 *are in force and that* k, g, *and* w *have continuous* $N+1$st *derivatives. Assume further that* g *is bounded away from* 0 *in an interval* $[a, b]$. *Then there is a constant* C *such that for any positive* ε *and for* n *sufficiently large*

$$P\left[\sup_{a\leqslant x\leqslant b}|m_n^{(N)}(x) - m^{(N)}(x)| > \varepsilon \leqslant C/\left(na_n^{2N2+}\varepsilon^2\right).$$

PROOF. Since $m = w/g$, we may use the Leibnitz expansion $m^{(N)}(x) = \Sigma_{k=0}^N\binom{N}{k}w^{(k)}(x)(g^{-1}(x))^{(N-k)}$ and the equivalent expansion for $m_n^{(N)}(x)$ to conclude that

$$\left(m_n^{(N)}(x) - m^{(N)}(x)\right) = \Sigma_{k=0}^N\binom{N}{k}\left(w_n^{(k)}(x)\left(g_n^{-1}(x)\right)^{(N-k)} - w^{(k)}(x)\left(g^{-1}(x)\right)^{(N-k)}\right).$$

The theorem will follow when it is shown that for some $C(k)$ and for n sufficiently large,

$$\text{(3)} \qquad P\Big[\max_{a \leqslant x \leqslant b} |w_n^{(k)}(x)(g_n^{-1}(x))^{(N-k)} - w^{(k)}(x)g^{-1}(x)^{(N-k)}| \geqslant \varepsilon \Big/ \Big\{(N+1)\binom{N}{k}\Big\}\Big] < C(k)\big/\big(\varepsilon^2 n a_n^{2N+2}\big).$$

For notational convenience, we let

$$\|h\|_{[a,b]} \equiv \sup_{a \leqslant x \leqslant b} |h(x)|.$$

Observe that $(g^{-1})^{(N-k)}$ may be expressed as an additive and multiplicative combination of the terms $g^{-1}, g, g^{(1)}, \cdots, g^{(N)}$, which are hypothesized to be uniformly continuous on $[a, b]$. For this reason, the reader may verify that $(g^{-1})^{(N-k)}$, as a function of the preceding terms, satisfies the uniform Lipshitz condition which follows. Assume that for $0 \leqslant v \leqslant N$

$$\| g_1^{(v)} - g^{(v)}\|_{[a,b]} < \delta$$

where $2\delta < \min_{a \leqslant x \leqslant b} g(x)$. Then, from consideration of the mean value theorem for the expansion of $(g^{-1})^{(N-k)}$, one concludes that there exists constants α_j such that

$$\|(g^{-1})^{(N-k)} - (g_1^{-1})^{(N-k)}\|_{[a,b]} \leqslant \Sigma_{v=0}^{N} \alpha_v \| g_1^{(v)} - g^{(v)}\|_{[a,b]}$$

where $(g_1^{-1})^{(N-k)}$ has been obtained by substituting the terms $g_1^{-1}, g_1^{(1)}, \cdots, g_1^{(N)}$ into the expression for $(g^{-1})^{(N-k)}$ in terms of $g^{-1}, g^{(1)}, \cdots, g^{(N)}$ mentioned above.

Let $E(\varepsilon_1)$ denote the event that for $0 \leqslant v \leqslant N$,

$$\|w^{(v)} - w_n^{(v)}\|_{[a,b]}, \quad \| g^{(v)} - g_n^{(v)}\|_{[a,b]} < \varepsilon_1 < \delta.$$

Then we have from the preceding development and an easy error formula, that

$$\|w_n^{(k)}(g_n^{-1})^{(N-k)} - w^{(k)}(g^{-1})^{(N-k)}\|_{[a,b]} < \big(\Sigma\alpha_j + \|(g^{-1})^{(N-k)}\|_{[a,b]} + \Sigma\alpha_j \|w^{(k)}\|_{[a,b]}\big)\varepsilon_1.$$

But from Lemmas 2 and 3, for n sufficiently large,

$$P\big[E(\varepsilon_1)^C\big] \leqslant \big\{\Sigma_{j=0}^{N}\big(C_v / n a_n^{2v+2}\varepsilon_1^2\big) + C'\exp\big(-C_2 n \varepsilon_1^2 a_n^{2v+2}\big)\big\} < (C_n + 1)\big/\big(n a_n^{2N+2}\varepsilon_1^2\big).$$

In the above expression, $E(\varepsilon_1)^C$ denotes the complement of $E(\varepsilon_1)$. If

$$\varepsilon_1 \leqslant \varepsilon \Big/ \Big[(N+1)\binom{N}{k}\big(\Sigma\alpha_j + \| g^{(N-k)}\|_{[a,b]} + \Sigma\alpha_j \|w^{(k)}\|_{[a,b]}\big)\Big],$$

then inequality (3) holds and the theorem is established.

3. Nonparametric regression from noisy sampled data. Let $\{f(\cdot;x)x \in [0, 1]\}$ be a family of probability density functions. We let $w(x) = \int yf(y; x)\, dy$ and $P_n = \{x_0, x_1, \cdots, x_n\}$ where $0 \leqslant x_0 \leqslant x_1 \leqslant \cdots \leqslant x_n = 1$ is a partition of [0, 1]. Let $|P_n| = \max|x_i - x_{i-1}|$. For fixed n, we will estimate $w(\cdot)$ by independent samples $\{Y_i\}_{i=1}^n$ where Y_i has density $f(y; x_i)$. The estimator will be the analogue of w_n in Section II; that is,

$$w_n(x) = \Sigma_{i=1}^n Y_i(x_i - x_{i-1})k((x - x_i)/a_n)/a_n$$

where $\{a_n\}$, k and its characteristic function are as in the preceding section. This estimator was introduced in Priestley and Chao (1972) and was also studied by Benedetti (1974, 1975).

The development to follow parallels the analysis in Section II. We define

$$\phi_n(u) = n^{-1}\Sigma_{s=1}^n Y_s \exp(iux_s).$$

Observe that $E\phi_n(u) = n^{-1}\Sigma_{s=1}^n w(x_s)\exp(iux_s)$. We use Lemma 1 and the following Lemmas 4 and 5.

LEMMA 4. *Let k be a probability density function having a bounded continuous first derivative and satisfying $\lim_{|u|\to\infty}|uk(u)| = 0$. Assume $\{a_n\}$ is a sequence of positive numbers converging to 0 such that $na_n^2 \to \infty$. If h is any function defined and having a continuous first derivative on* [0, 1] *and if $|P_n| = 0(n^{-1})$, then*

$$\lim_{n\to\infty}\sup_{a\leqslant x\leqslant b}|a_n^{-1}\Sigma_{i=1}^n k((x - x_i)/a_n)h(x_i)(x_i - x_{i-1}) - h(x)| = 0$$

for any a, b such that $0 < a < b < 1$.

PROOF. From the definition of the Riemann integral we see

$$|a_n^{-1}\Sigma_{i=1}^n k((x - x_i)/a_n)h(x_i)(x_i - x_{i-1}) - \int_0^1 a_n^{-1}k((x - u)/a_n)h(u)\, du| \leqslant \Sigma_{i=1}^n(M_i - m_i)(x_i - x_{i-1})$$

where

$$M_i = M_i(n), \quad m_i = m_i(n)$$

denote, respectively, the maximum and minimum of

$$\{a_n^{-1}k((x - u)/a_n)h(u) : x_{i-1} \leqslant u \leqslant x_i\}, \quad i = 1, 2, \cdots, n.$$

Since a continuous function attains both its maximum and minimum on any closed interval, we can use the mean value theorem to see that

$$0 \leqslant M_i - m_i \leqslant (|k((x - \xi)/a_n)h'(\xi)| + a_n^{-1}|k'((x - \xi)/a_n)h(\xi)|)(x_i - x_{i-1})/a_n$$

for some $\xi \in [x_{i-1}, x_i]$. Since k and k' are bounded on $(-\infty, \infty)$, h and h' are bounded on [0, 1], $na_n^2 \to \infty$, and $n(x_i - x_{i-1}) = 0(1)$, we see that $M_i(n) - M_i(n) \to 0$ uniformly in i as n tends to ∞. Thus one can verify that

$$\lim_{n\to\infty}\sup_{a\leqslant x\leqslant b}|a_n^{-1}\Sigma_{i=1}^n k((x - x_i)/a_n)h(x_i)(x_i - x_{i-1}) - \int_0^1 a_n^{-1}k((x - u)/a_n)h(u)\, du| = 0.$$

Lemma 4 now follows from Lemma 1 and the triangle inequality.

LEMMA 5. *Let k and $\{a_n\}$ satisfy the conditions of Lemma 4. Suppose also that* $\lim_{u\to\infty}|u^{j+1}k^{(j)}(u)| = 0$ *for* $j = 0, 1, \cdots N$ *and that the first* $(N + 1)$ *derivatives of* k *are continuous and bounded on* $(-\infty, \infty)$ *and that* $(na_n^{N+2}) \to_n \infty$. *If* $w^{(N+1)}$ *exists and is continuous on* $[0, 1]$ *then* $\lim_{n\to\infty}\sup_{[a,b]}|E[w_n^{(j)}(x)] - w^{(j)}(x)| = 0$ *for* $j = 0, 1, \cdots, N$ *for any* a, b *with* $0 < a < b < 1$.

PROOF. Let us first consider the case $j = 0$, then

$$E[w_n(x)] = a_n^{-1}\Sigma_{i=1}^n k((x - x_i)/a_n)w(x_i)(x_i - x_{i-1})$$

which, by Lemma 4, converges uniformly on $[a, b]$ to $w(x)$. The convergence of derivatives is established, as in Lemma 1, by repeated integration by parts. For $i = 1$

$$E[w_n^{(1)}(x)] = a_n^{-2}\Sigma_{i=1}^n k'((x - x_i)/a_n)w(x_i)(x_i - x_{i-1})$$

and

$$a_n^{-2}\int_0^1 k'((x - u)/a_n)w(u)\,du$$
$$= -a_n^{-1}w(u)k((x - u)/a_n)|_{u=0}^1 + a_n^{-1}\int_0^1 k((x - u)/a_n)w^{(1)}(u)\,du.$$

Since $\lim_{u\to\infty}|uk(u)| = 0$, the first term on the right-hand side tends to zero uniformly for $x \in [a, b]$, as n tends to infinity. Proceeding as in the proof of Lemma 4 one obtains

$$\lim_{n\to\infty}\sup_{a\leqslant x\leqslant b}|a_n^{-2}\Sigma_{i=1}^n k^{(1)}((x - x_i)/a_n)w(x_i)(x_i - x_{i-1})$$
$$-\int_0^1 a_n^{-2}k^{(1)}((x - u)/a_n)w(u)\,du| = 0$$

since $na_n^3 \to \infty$. The desired conclusion for $i = 1$ then follows from Lemma 1. The proof for higher derivatives follows in a similar fashion.

THEOREM 2. *Assume that the conditions of Lemma 5 are in force and that* $u^N\psi(u)$ *is absolutely integrable. If* $0 < a < b < 1$, *and if there is a bound* V *such that*

$$\int y^2 f(y; x)\,dy \leqslant V, \qquad \text{all } x \in [0, 1],$$

then for every $\varepsilon > 0$ *there is a constant* $C > 0$ *such that for* n *sufficiently large,*

$$P[\sup_{[a,b]}|w_n^{(N)}(x) - w^{(N)}(x)| > \varepsilon] < C/(\varepsilon^2 n a_n^{2N+2}).$$

PROOF. Since the different samples are assumed independent, one can easily see that

$$E[|\phi_n(u) - E[\phi_n(u)]|^2] = n^{-2}\Sigma_{s=1}^n E[(Y_s - w(x_s))^2] \leqslant V/n.$$

The rest of the proof follows that of Lemma 3.

4. Identification applications. The study of estimation of derivatives through nonparametric regression was motivated in part by our interest in identification of time-varying linear systems using noisy measurements on a system trajectory. For example, consider the time-varying linear differential equation

$$w^{(1)}(x) = \theta(x)w(x), \qquad x \in [0, 1].$$

Let us suppose that if n measurements are to be made, they are made at equally spaced points $1/n, 2/n, \cdots, 1$ in the unit interval. Thus, the noisy values $\{Y_i\}_{i=1}^n$ are recorded, where the Y_i's are independent, have probability density functions $f(\cdot; i/n)$, and means $w(i/n)$ equal to the differential equation solution values at i/n, $0 \leqslant i \leqslant n$. We may estimate $\theta(x)$ by

$$\theta_n(x) \equiv w_n^{(1)}(x)/w_n(x), \qquad x \in [0, 1].$$

Error analysis of this identification technique may be undertaken with the results of the preceding sections. Also, higher order systems may be identified if noisy solutions associated with different forcing functions (inhomogeneous terms) are available.

We have conducted some preliminary computer experimentation without making much effort to adjust the parameters $\{a_n\}$ or to find a good kernel $k(\cdot)$. The asymptotic distribution results in Schuster (1972) indicates one might take a_n in the neighborhood of $0(n^{-\frac{1}{5}})$ to $0(n^{-\frac{1}{3}})$. This also seems to be the case in the numerical work reported in Watson (1964). Watson also reports that the choice of kernel does not seem to affect convergence results so that one should use a kernel which is relatively easy to compute. We used a standard normal density in much of our empirical work with a_n about $\sigma n^{-\frac{1}{3}}$ where σ^2 was the variance of the noise variable. These empirical studies were sufficiently encouraging to indicate that the identification applications considered here warrant additional theoretical and empirical study.

The literature (e.g., Graupe (1972) and references therein) on system identification is largely devoted to inference of parameters in time-invariant systems. Some effort has been made to apply the method of quasilinearization (e.g., Yeh and Tauxe (1971)) and gradient techniques (Seinfeld (1969)) to time-varying system identification, but these approaches are constrained to assume a finite dimensional function space for the unknown parameter. Sagar, Yakowitz, and Duckstein (1975) and Yakowitz and Noren (1976) study a time-varying system identification scheme which uses techniques of linear regression to get an estimate of the parameter based solely on local noisy solution measurements to get parameter estimates $\theta(x)$ for a grid of x values. Problems in ground water analysis, heat and pollution diffusion, and petroleum engineering call for progress in time-varying system identification. Because of the convergence of derivatives established in this paper, nonparametric regression theory could seem to provide a promising avenue for these problems, but much theoretical work remains to be done toward making nonparametric regression as efficient as possible.

REFERENCES

BENEDETTI, J. (1974). Kernel estimation of regression functions. Ph.D. dissertation, Dept. of Biostatistics, Univ. Washington.

BENEDETTI, J. (1975). Kernel estimation of regression functions. *Proc. Computer Scinece and Statist.: 8th Annual Symposium on the Interface.* Health Science Computing Facility, UCLA.

BUTLER, G. A. (1975). Heuristic regression for large commercial problems. *Proc. Computer Science and*

Statist.: 8th Annual Symposium on the Interface. Health Science Computing Facility, UCLA, 398–405.

FISHER, L. and YAKOWITZ, S. (1976). Uniform convergence of the potential function algorithm. *SIAM J. Control* **14** 95–103.

GARTSIDE, P. (1975). An approach to conditional survival estimation. *Proc. Computer Science and Statist.: 8th Annual Symposium on the Interface.* Health Science Computing Facility, UCLA, 424–431.

GRAUPE, D. (1972). *Identification of Systems.* Van Nostrand Reinholt, New York.

HANSEN, D. L. and PLEDGER, G. (1976). Consistency in concave regression. *Ann. Statist.* **4** 1038–1050.

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

NADARAYA, E. A. (1970). Remarks on nonparametric estimates for density functions and regression curves. *Theor. Probability Appl.* **15** 134–137.

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

ROSENBLATT, M. (1969). Conditional probability density and regression estimators. In *Multivariate Analysis*, **2**, 25–31. Academic Press, New York.

SAGAR, B., YAKOWITZ, S. and DUCKSTEIN, L. (1975). A direct method for the identification of the parameters of dynamic non-homogeneous aquifers. *Water Resources Research* 11(**4**) 563–370.

SCHUSTER, E. F. (1968). Estimation of a probability density function with applications in statistical inference. Ph.D. dissertation, Univ. Arizona.

SCHUSTER, E. F. (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* **40** 1187–1195.

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.

SEINFELD, J. (1969). Identification of parameters in partial differential equations. *Chem. Engineering Sci.* **24** 65–74.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–620.

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

WILKS, S. (1962). *Mathematical Statistics.* Wiley, New York.

YAKOWITZ, S. and NOREN, P. (1976) On the identification of parameters in dynamic linear partial differential equations. *J. Math. Anal. Appl.* 521–538

YEH, W. and TAUXE, G. (1971). Quasilinearization and identification of Aquifer parameters. *Water Resources Research* 7(**2**) 375–381.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS AT EL PASO
EL PASO, TEXAS 79968

SYSTEMS AND INDUSTRIAL ENGINEERING
THE UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721