The Annals of Statistics
1984, Vol. 12, No. 3, 917–926

ASYMPTOTIC NORMALITY OF NEAREST NEIGHBOR REGRESSION FUNCTION ESTIMATES

BY WINFRIED STUTE

University of Giessen

Let (X, Y) be a random vector in the plane. We show that a smoothed N.N. estimate of the regression function $m(x) = \mathbb{E}(Y \mid X = x)$ is asymptotically normal under conditions much weaker than needed for the Nadaraya-Watson estimate. It also turns out that N.N. estimates are more efficient than kernel-type estimates if (in the mean) there are few observations in neighborhoods of x.

Introduction and main results. Assume that (X, Y) is a random vector in $\mathbb{R}^2$. If Y has finite expectation $\mathbb{E}(Y)$, the regression function $m(x) = \mathbb{E}(Y \mid X = x)$, $x \in \mathbb{R}$, of Y on X exists and is (almost surely in x) uniquely defined in view of the equation $m(X) = \mathbb{E}(Y \mid X)$. Let $(X_1, Y_1), (X_2, Y_2), \cdots$ be independent random observations with the same distribution as (X, Y). It is required to construct an estimate $m_n(x_0) = m_n(x_0, X_1, Y_1, \cdots, X_n, Y_n)$ of $m(x_0)$, which behaves well even when only little information on the distribution of (X, Y) is available. The point x_0 may be interpreted as a future value taken on by some X for which Y is not yet observed.

Nadaraya (1964) and Watson (1964) independently proposed the estimate

$$m_n^*(x_0) = \frac{\sum_{i=1}^n Y_i K((x_0 - X_i)/a_n)}{\sum_{i=1}^n K((x_0 - X_i)/a_n)}$$

where K is an appropriate kernel function (integrating to one) and $a_n \to 0$ is a sequence of bandwidths. That m_n^* is a reasonable estimate for m may be seen by arguments involving multivariate densities (cf. Watson, 1964), or by the fact (cf. Devroye and Wagner, 1980) that

$$(na_n)^{-1} \sum_{i=1}^n Y_i K\left(\frac{x_0 - X_i}{a_n}\right) \to m(x_0)f(x_0)$$

and

$$(na_n)^{-1} \sum_{i=1}^n K\left(\frac{x_0 - X_i}{a_n}\right) \to f(x_0) \quad \text{in probability.}$$

Here f denotes the (marginal) density of X, assumed to be positive at x_0. Schuster (1972), under conditions requiring the existence of f and finiteness of $\mathbb{E}(\mid Y \mid^3)$,

Received August 1983; revised March 1984.
AMS 1980 *subject classifications*. Primary 62J02, 62E20; secondary 62G05.
Key words and phrases. Nearest neighbor estimates, regression function, nonparametric estimation, asymptotic normality.

showed that

$$(na_n)^{1/2}[m_n^*(x_0) - m(x_0)] \to N(0, \sigma^2) \quad \text{in distribution},$$

where $\sigma^2 = \mathrm{Var}(Y \mid X = x_0) \int K^2(u)\, du/f(x_0)$. See also Rosenblatt (1969). In fact, even joint convergence in distribution of m_n^* at finitely many points $x_1, \cdots, x_k$ holds, with $m_n^*(x_1), \cdots, m_n^*(x_k)$ being asymptotically independent.

More generally, one might consider estimates of the form

$$m_n(x_0) = \sum_{i=1}^n Y_i W_{ni} = \sum_{i=1}^n Y_i W_{ni}(x_0, X_1, \cdots, X_n),$$

where W_{ni}, $1 \le i \le n$, is an array of weights with $\sum_{i=1}^n W_{ni} \to 1$. Stone (1977) obtained necessary and sufficient conditions on the weights for the convergence

$$\mathbb{E}\left(\int | m_n(x) - m(x) | \mu(dx)\right) \to 0 \quad \text{as} \quad n \to \infty, \tag{1}$$

where μ is the distribution of X. In particular, (1) was shown to hold true for certain nearest neighbor type estimates.

In this paper the asymptotic normality of such estimates is derived under conditions much weaker than needed for the Nadaraya-Watson estimate. Specifically, only finiteness of $\mathbb{E}(Y^2)$ is required, while X need not have a density at all. With the same meaning of K and a_n, we shall consider the estimates

$$m_n(x_0) = (na_n)^{-1} \sum_{i=1}^n Y_i K\left(\frac{F_n(x_0) - F_n(X_i)}{a_n}\right).$$

Here F_n is the empirical distribution function (e.d.f.) of $X_1, \cdots, X_n$. Technically, m_n only depends on the ranks of $X_1, \cdots, X_n$. This amounts to the effect that if X has a continuous d.f. F, the original problem of estimating $m(x_0)$ may be transformed into one of estimating a regression function at $F(x_0)$, with the X-sample uniformly distributed on $[0, 1]$. The mean square convergence of m_n to m has been studied by Yang (1981).

That m_n is in fact a (smoothed) nearest neighbor type estimate may be seen when $K = 1_{[-1/2, 1/2]}$. In this case $m_n(x_0)$ is the average number of Y_i's for which, when $X_i \ge x_0$ (say), there exist no more than $k_n = na_n/2$ X_j-values with $x_0 \le X_j$ and such that $X_j < X_i$.

Since

$$(na_n)^{-1} \sum_{i=1}^n K\left(\frac{F_n(x_0) - F_n(X_i)}{a_n}\right) \to 1 \quad \text{in probability},$$

the estimate

$$m_{n0}(x_0) = \frac{\sum_{i=1}^n Y_i K((F_n(x_0) - F_n(X_i))/a_n)}{\sum_{i=1}^n K((F_n(x_0) - F_n(X_i))/a_n)}$$

has the same asymptotic behavior as m_n. Empirical investigations suggest, however, that m_{n0} is superior to m_n when the sample size is small. This seems to be true due to the fact that the weights of m_{n0} sum up to one. The kernel function

K is allowed to take on negative values, too. There are no reasonable grounds for a restriction to nonnegative kernels, as is sometimes done in density estimation.

Mack (1981) proved asymptotic normality of a somewhat different type of N.N.-estimate, under more restrictive assumptions on the distribution of (X, Y). See also Royall (1966).

After these introductory remarks we shall now state our main result.

THEOREM. *Assume that X has a continuous d.f. F, and let $\mathbb{E}(Y^2) < \infty$. Let K be a twice continuously differentiable kernel function vanishing outside some bounded interval. For each bandsequence $(a_n)_n$ with $a_n \to 0$ and $na_n^3 \to \infty$ we then have*

$$(na_n)^{1/2}[m_n(x_0) - \bar{m}_n(x_0)] \to N(0, \sigma_0^2) \quad \text{in distribution} \tag{2}$$

for μ-almost all $x_0 \in \mathbb{R}$, where

$$\bar{m}_n(x_0) = a_n^{-1} \int yK\left(\frac{F(x_0) - F(x)}{a_n}\right)H(dx, dy),$$

H is the d.f. of (X, Y) and

$$\sigma_0^2 = \operatorname{Var}(Y \mid X = x_0) \int K^2(u)\, du.$$

To show that $m_n(x_0) - m(x_0)$ has the same limit distribution as $m_n(x_0) - \bar{m}_n(x_0)$, one needs to prove

$$(na_n)^{1/2}(\bar{m}_n(x_0) - m(x_0)) \to 0 \quad \text{as} \quad n \to \infty. \tag{3}$$

While the convergence of the "stochastic component" $m_n - \bar{m}_n$ could be proved under minimal assumptions on the underlying distribution, (3) typically needs some further smoothness conditions which guarantee $\bar{m}_n(x_0) \to m(x_0)$ at a satisfactory rate. In our case, smoothness of the function

$$m \circ F^{-1}(u) \equiv \mathbb{E}(Y \mid F(X) = u)$$

in a neighborhood of $F(x_0)$ suffices. This should be compared with the Nadaraya-Watson estimate, in which case F was assumed to admit a density.

COROLLARY. *Under the assumptions of the Theorem, let K be such that $\int K(u)\, du = 1$ and $\int uK(u)\, du = 0$. If $m \circ F^{-1}$ is twice continuously differentiable in a neighborhood of $0 < F(x_0) < 1$ and $na_n^5 \to 0$, then (3) holds, so that*

$$(na_n)^{1/2}[m_n(x_0) - m(x_0)] \to N(0, \sigma_0^2) \quad \text{in distribution} \tag{4}$$

whenever (2) is satisfied.

If $m \circ F^{-1}$ is d-times differentiable ($d \geq 2$), the condition $na_n^5 \to 0$ may be weakened to $na_n^{2d+1} \to 0$, if one chooses a kernel K for which $\int u^i K(u)\, du = 0$ for $i = 1, 2, \cdots, d - 1$. For $d > 2$ this may be only achieved if one admits negative values for K.

With the interpretation that x_0 is some future observation on X, the condition $0 < F(x_0) < 1$ may be taken for granted, the set of x_0 with $F(x_0) = 0$ or $F(x_0) = 1$ being a μ-null set.

A comparison with σ^2 shows that σ_0^2 does not depend on the (unknown, possibly nonexisting) marginal density of X, while, if X has a density f, $\sigma_0^2/\sigma^2 = f(x_0)$. Hence, if $f(x_0) < 1$, $m_n(x_0)$ is more efficient than $m_n^*(x_0)$. The situation is similar to nearest neighbor density estimation (cf. Moore and Yackel, 1977). In each case a small value of $f(x_0)$ results in fewer X-observations in neighborhoods of x_0, so that in fixed radii estimation only a small portion of Y-data points is involved.

Lemmas and proofs. In the following, let H_n denote the (bivariate) empirical d.f. of the sample $(X_1, Y_1), \cdots, (X_n, Y_n)$. We then have

$$m_n(x_0) = a_n^{-1} \int yK\left(\frac{F_n(x_0) - F_n(x)}{a_n}\right) H_n(dx, dy).$$

Since K is twice differentiable, Taylor expansion yields

$$\begin{aligned} &m_n(x_0) \\ &= a_n^{-1} \int yK\left(\frac{F(x_0) - F(x)}{a_n}\right) H_n(dx, dy) \\ &\quad + a_n^{-2} \int y[F_n(x_0) - F_n(x) - F(x_0) + F(x)]K'\left(\frac{F(x_0) - F(x)}{a_n}\right) H_n(dx, dy) \\ &\quad + a_n^{-3} \int y[F_n(x_0) - F_n(x) - F(x_0) + F(x)]^2 \frac{K''(\Delta) H_n(dx, dy)}{2} \\ &\equiv I_1 + I_2 + I_3, \end{aligned}$$

where Δ is between $a_n^{-1}[F_n(x_0) - F_n(x)]$ and $a_n^{-1}[F(x_0) - F(x)]$.

LEMMA 1. *$(na_n)^{1/2} I_3 \to 0$ in probability as $n \to \infty$.*

PROOF. Since K vanishes outside some finite interval, say $(-1, 1)$ w.l.o.g., the above expansion of $m_n(x_0)$ holds true with integration restricted to those x for which $|F_n(x_0) - F_n(x)| < a_n$. The Dvoretzky-Kiefer-Wolfowitz (1956) bound for the tails of $\sup_x |F_n(x) - F(x)|$ yields that for given $\varepsilon > 0$ there exists some finite C such that for $n \in \mathbb{N}$, $\sup_x |F_n(x) - F(x)| \le Cn^{-1/2}$ up to an event of probability less than or equal to ε. On this set the inequality $|F_n(x_0) - F_n(x)| < a_n$ entails

$$|F(x_0) - F(x)| \le a_n + 2Cn^{-1/2} \le C_1 a_n \quad \text{for some} \quad C_1 < \infty.$$

Lemma 2.3 in Stute (1982) asserts that

$$\sup_{x:|F(x_0)-F(x)| \le C_1 a_n} (na_n^{-1})^{1/2} |F_n(x_0) - F_n(x) - F(x_0) + F(x)|$$

is stochastically bounded as $n \to \infty$. The assertion of the lemma now follows from the facts that K'' is bounded and $\limsup_{n\to\infty} \int |y| H_n(dx, dy) < \infty$ with probability one, upon observing that $\varepsilon > 0$ was arbitrary and $na_n^3 \to \infty$. □

Next, we shall show that $(na_n)^{1/2} I_2$ is asymptotically equivalent to

$$-(na_n)^{1/2} a_n^{-1} m(x_0) \int K\left(\frac{F(x_0) - F(x)}{a_n}\right)[F_n(dx) - F(dx)].$$

For this, define

$$Z_n = n^{-1} a_n^{-3/2} \textstyle\sum_{i=1}^n [Y_i - m(X_i)] \cdot [\alpha_n(x_0) - \alpha_n(X_i)] K'\left(\dfrac{F(x_0) - F(X_i)}{a_n}\right)$$

with

$$\alpha_n(x) = n^{1/2}[F_n(x) - F(x)], \quad x \in \mathbb{R}$$

denoting the empirical process pertaining to $X_1, \cdots, X_n$. Furthermore, let $\mathscr{F} = \sigma(X_1, X_2, \cdots)$ be the σ-field generated by the X-data.

LEMMA 2. *$Z_n \to 0$ in probability as $n \to \infty$.*

PROOF. To prove the lemma, we shall show $\lim_{n\to\infty} \mathbb{E}(Z_n^2) = 0$. For this, observe that conditionally on $\mathscr{F}$, the summands of Z_n are centered and uncorrelated. Hence

$$\mathbb{E}(Z_n^2 \mid \mathscr{F})$$
$$= a_n^{-3} n^{-2} \textstyle\sum_{i=1}^n \mathbb{E}([Y_i - m(X_i)]^2 \mid \mathscr{F}) \left\{(\alpha_n(x_0) - \alpha_n(X_i)) K'\left(\dfrac{F(x_0) - F(X_i)}{a_n}\right)\right\}^2.$$

Let F_{n-1}^i and α_{n-1}^i be the e.d.f. and empirical process of the sample $X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n$, respectively. Since

$$F_n(x) = F_{n-1}^i(x) - n^{-1} F_{n-1}^i(x) + n^{-1} 1_{(-\infty, x]} \circ X_i, \quad 1 \le i \le n, \quad x \in \mathbb{R}$$

we obtain

$$\alpha_n(x_0) - \alpha_n(X_i) = (n/(n-1))^{1/2} [\alpha_{n-1}^i(x_0) - \alpha_{n-1}^i(X_i)] + R_n,$$

where $|R_n| \le 4n^{-1/2}$. From $(a + b)^2 \le 2(a^2 + b^2)$ we therefore infer

$$[\alpha_n(x_0) - \alpha_n(X_i)]^2 \le 2\left[\frac{n}{n-1} (\alpha_{n-1}^i(x_0) - \alpha_{n-1}^i(X_i))^2 + \frac{16}{n}\right].$$

It follows that

$$\mathbb{E}(Z_n^2) \le 2a_n^{-3} n^{-1} \int h(x)\left[2 \mid F(x_0) - F(x) \mid + \frac{16}{n}\right]\left\{K'\left(\frac{F(x_0) - F(x)}{a_n}\right)\right\}^2 F(dx), \quad n \ge 2,$$

with $h(x) = \mathbb{E}((Y - m(X))^2 \mid X = x)$. Our assumptions on K guarantee that K' is bounded. The assertion of the Lemma therefore follows from the integrability of h and the convergence $na_n^3 \to \infty$. □

Next, consider the function

$$k(x, y) = m(x)K'\left(\frac{F(x_0) - F(x)}{a_n}\right)\{1_{(-\infty, x_0]}(y) - 1_{(-\infty, x]}(y)\}$$

with corresponding "von Mises" statistic

$$T_n = n \int k(x, y)[F_n(dy) - F(dy)][F_n(dx) - F(dx)] = \int k(x, y)\alpha_n(dy)\alpha_n(dx).$$

Lemma B on page 223 in Serfling (1980) yields

$$\mathbb{E}(T_n^2) = O(1) \quad \text{as} \quad n \to \infty. \tag{5}$$

That the function k depends on n (through a_n) is immaterial for the proof of (5) due to the boundedness of K'. It follows from $na_n^3 \to \infty$ that

$$a_n^{-3/2} \int k(x, y)\alpha_n(dy)[F_n(dx) - F(dx)] \to 0 \quad \text{in probability.}$$

In view of Lemma 2 we thus obtain that $(na_n)^{1/2} I_2$ is asymptotically equivalent to

$$a_n^{-3/2} \int m(x)[\alpha_n(x_0) - \alpha_n(x)]K'\left(\frac{F(x_0) - F(x)}{a_n}\right)F(dx).$$

In the following we use the well-known representation $\alpha_n(x) = \bar{\alpha}_n(F(x))$, $x \in \mathbb{R}$, of α_n in terms of a uniform empirical process $\bar{\alpha}_n$.

LEMMA 3. *We have*

$$a_n^{-3/2} \int |m(x) - m(x_0)| \, |\alpha_n(x_0) - \alpha_n(x)| \left| K'\left(\frac{F(x_0) - F(x)}{a_n}\right)\right| F(dx) \to 0$$

in probability for μ-almost all $x_0 \in \mathbb{R}$.

PROOF. For continuous F, the above integral is equal to

$$a_n^{-3/2} \int_0^1 |m(F^{-1}(u)) - m(x_0)| \, |\bar{\alpha}_n(F(x_0)) - \bar{\alpha}_n(u)| \left| K'\left(\frac{F(x_0) - u}{a_n}\right)\right| du$$

$$\leq a_n^{-3/2} \sup_{u: |F(x_0) - u| \leq a_n} |\bar{\alpha}_n(F(x_0)) - \bar{\alpha}_n(u)|$$

$$\cdot \int_0^1 |m(F^{-1}(u)) - m(x_0)| \left| K'\left(\frac{F(x_0) - u}{a_n}\right)\right| du,$$

where as before we made the assumption $K \equiv 0$ and hence $K' \equiv 0$ outside $(-1, 1)$. From Lemma 2.3 in Stute (1982) we get

$$\sup_{u: |F(x_0) - u| \leq a_n} |\bar{\alpha}_n(F(x_0)) - \bar{\alpha}_n(u)| = O_{\mathbb{P}}(a_n^{1/2}) \quad \text{as} \quad n \to \infty.$$

Furthermore,

$$a_n^{-1} \int_0^1 | m(F^{-1}(u)) - m(x_0) | \left| K'\left(\frac{F(x_0) - u}{a_n}\right) \right| du$$

$$\leq \int | m(F^{-1}(F(x_0) - ua_n)) - m(x_0) | \, | K'(u) | \, du.$$

Observe that $F^{-1}(F(x_0)) = x_0$ for μ-almost all $x_0 \in \mathbb{R}$. By Theorem 2 in Stein (1970), page 62–63,

$$\lim_{n\to\infty} \int | m(F^{-1}(s - ua_n)) - mF^{-1}(s) | \, | K'(u) | \, du = 0$$

for Lebesgue almost all s, say for all $s \in$ A. By continuity of F, we thus have

$$\mu(\{x_0: F(x_0) \in A, F^{-1}(F(x_0)) = x_0\}) = 1.$$

This proves the Lemma. □

LEMMA 4. *$(na_n)^{1/2} I_2$ is asymptotically equivalent to*

$$-a_n^{-1/2} m(x_0) \int K\left(\frac{F(x_0) - F(x)}{a_n}\right) \alpha_n(dx).$$

PROOF. Lemma 3 shows that $(na_n)^{1/2} I_2$ is asymptotically equivalent to

$$a_n^{-3/2} m(x_0) \int [\alpha_n(x_0) - \alpha_n(x)] K'\left(\frac{F(x_0) - F(x)}{a_n}\right) F(dx)$$

$$= -a_n^{-3/2} m(x_0) \int \alpha_n(x) K'\left(\frac{F(x_0) - F(x)}{a_n}\right) F(dx)$$

$$= -a_n^{-1/2} m(x_0) \int K\left(\frac{F(x_0) - F(x)}{a_n}\right) \alpha_n(dx),$$

where the last equality follows upon integrating by parts. □

We are now in the position to give the

PROOF OF THE THEOREM. According to Lemma 4 it remains to show

$$I_4 \equiv \left(\frac{n}{a_n}\right)^{1/2} \int [y - m(x_0)] K\left(\frac{F(x_0) - F(x)}{a_n}\right) [H_n(dx, dy) - H(dx, dy)] \to N(0, \sigma_0^2)$$

in distribution. This may be achieved along classical lines. For each n, I_4 is a standardized sum of i.i.d. random variables with

$$\begin{aligned}\operatorname{Var}(I_4) &= a_n^{-1}\Big\{\int (y - m(x_0))^2 K^2\Big(\frac{F(x_0) - F(x)}{a_n}\Big) H(dx, dy) \\ &\qquad - \Big[\int (y - m(x_0)) K\Big(\frac{F(x_0) - F(x)}{a_n}\Big) H(dx, dy)\Big]^2\Big\} \\ &= a_n^{-1}\Big\{\int h(x) K^2\Big(\frac{F(x_0) - F(x)}{a_n}\Big) F(dx) \\ &\qquad - \Big[\int (m(x) - m(x_0)) K\Big(\frac{F(x_0) - F(x)}{a_n}\Big) F(dx)\Big]^2\Big\},\end{aligned}$$

with $h(x) = \mathbb{E}((Y - m(x_0))^2 \mid X = x)$. As in the proof of Lemma 3, it follows that $\operatorname{Var}(I_4) \to h(x_0) \int K^2(u)\, du$ for μ-almost all $x_0 \in \mathbb{R}$. So it suffices to show that the array defining the I_4's satisfies the Lindeberg condition (for μ-almost all x_0). Since the centering constants are asymptotically negligible, it is easy to see that it remains to prove

$$a_n^{-1} \int_{\{|y - m(x_0)| \geq \delta(na_n)^{1/2}\}} (y - m(x_0))^2 K^2\Big(\frac{F(x_0) - F(x)}{a_n}\Big) H(dx, dy) \to 0 \tag{6}$$

$$\text{for all } \delta > 0 \text{ as } n \to \infty.$$

For this, put

$$h_a(x) = \mathbb{E}((Y - m(x_0))^2 1_{\{|Y - m(x_0)| > a\}} \mid X = x), \quad a > 0,$$

and assume that this conditional expectation is obtained from integration w.r.t. a regular conditional distribution. In particular, $h_a(x) \downarrow 0$ for each x as $a \uparrow \infty$. Since $na_n \to \infty$, (6) will follow if

$$\limsup_{n\to\infty} a_n^{-1} \int h_a(x) K^2\Big(\frac{F(x_0) - F(x)}{a_n}\Big) F(dx) \tag{7}$$

can be made arbitrarily small upon choosing a large enough. As in the proof of Lemma 3, it follows from standard results in differentiation theory that (7) is equal to $h_a(x_0) \int K^2(u)\, du$ for μ-almost all $x_0 \in \mathbb{R}$. In view of the above remark, this may be made small by letting $a \uparrow \infty$, whence (6). This completes the proof of the Theorem. □

PROOF OF THE COROLLARY. We have

$$\bar{m}_n(x_0) = a_n^{-1} \int m(x) K\Big(\frac{F(x_0) - F(x)}{a_n}\Big) F(dx)$$

and

$$a_n^{-1} \int K\Big(\frac{F(x_0) - F(x)}{a_n}\Big) F(dx) = \int_{(F(x_0) - 1)/a_n}^{F(x_0)/a_n} K(x)\, dx.$$

Since K has bounded support, $0 < F(x_0) < 1$ and $a_n \to 0$, the last integral is equal to one for all $n \geq n_0$, say. Hence

$$\bar{m}_n(x_0) - m(x_0) = a_n^{-1} \int [m(x) - m(x_0)] K\left(\frac{F(x_0) - F(x)}{a_n}\right) F(dx)$$

$$= \int [m \circ F^{-1}(F(x_0) - ua_n) - m \circ F^{-1}(F(x_0))] K(u)\, du$$

$$= I_5$$

provided that $F^{-1}(F(x_0)) = x_0$. Since $\int uK(u)\,du = 0$, Taylor expansion of the above integrand yields $I_5 = O(a_n^2)$, so that the Corollary is an immediate consequence of the Theorem and the assumption $na_n^5 \to 0$. □

REMARK 1. As is apparent from the proof of the Corollary, the condition $na_n^5 \to 0$ is needed only to guarantee that the deterministic error term $\bar{m}_n(x_0) - m(x_0)$ is asymptotically negligible. The optimal choice "$na_n^5 \to$ positive const." may be treated likewise, to the effect that $(na_n)^{1/2}(m_n - m)$ is asymptotically biased.

REMARK 2. As in Schuster (1972) the Cramér-Wold device may be applied to show that $(na_n)^{1/2}(m_n - m)$ converges jointly in distribution at finitely many points $x_1, \cdots, x_k$, with $m_n(x_1), \cdots, m_n(x_k)$ being asymptotically independent.

REMARK 3. The results of this paper may be extended to the case when X is multivariate, by applying the results of Stute (1984). Beyond technicalities there will be one major difference to the univariate case, however, in that transforming to the uniform distribution now leads to a distribution with uniform marginals, but otherwise depending on the distribution of X.

REFERENCES

DEVROYE, L. P. and WAGNER, T. J. (1980). On the L_1 convergence of kernel estimators of regression functions with applications in discrimination. *Z. Wahrsch. verw. Gebiete* **51** 15–25.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

MACK, Y. P. (1981). Local properties of k-NN regression estimates. *SIAM J. Alg. Disc. Meth.* **2** 311–323.

MOORE, D. S. and YACKEL, J. W. (1977). Large sample properties of nearest neighbor density function estimators. In *Statistical Decision Theory and Related Topics II.* (S. S. Gupta and D. S. Moore, eds.) 269–279. Academic, New York.

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

ROSENBLATT, M. (1969). Conditional probability density and regression estimates. In *Multivariate Analysis II.* (Krishnaiah, ed.) 25–31.

ROYALL, R. M. (1966). A class of nonparametric estimates of a smooth regression function. Ph.D. thesis, Stanford University.

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.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

STEIN, E. M. (1970). Singular integrals and differentiability properties of functions. Princeton University Press. Princeton, New Jersey.

STONE, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.* **5** 595–645.

STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

STUTE, W. (1984). The oscillation behavior of empirical processes: The multivariate case. *Ann. Probab.* **12** 361–379.

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

YANG, S. S. (1981). Linear functions of concomitants of order statistics with applications to nonparametric estimation of a regression function. *J. Amer. Statist. Assoc.* **76** 658–662.

MATHEMATISCHES INSTITUT DER
JUSTUS-LIEBIG UNIVERSITÄT GIEßEN
ARNDTSTRAßE 2
D-6300 GIEßEN
WEST GERMANY