The Annals of Statistics
1980, Vol. 8, No. 2, 231–239

DISTRIBUTION-FREE CONSISTENCY RESULTS IN NONPARAMETRIC DISCRIMINATION AND REGRESSION FUNCTION ESTIMATION[1]

BY LUC P. DEVROYE AND T. J. WAGNER

McGill University and University of Texas, Austin

Let (X, Y) be an $\mathbb{R}^d \times \mathbb{R}$-valued random vector and let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be a random sample drawn from its distribution. We study the consistency properties of the kernel estimate $m_n(x)$ of the regression function $m(x) = E\{Y|X = x\}$ that is defined by

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

where k is a bounded nonnegative function on $\mathbb{R}^d$ with compact support and $\{h_n\}$ is a sequence of positive numbers satisfying $h_n \to_n 0$, $nh_n^d \to_n \infty$. It is shown that $E\{\int |m_n(x) - m(x)|^p \mu(dx)\} \to_n 0$ whenever $E\{|Y|^p\} < \infty (p \geq 1)$. No other restrictions are placed on the distribution of (X, Y). The result is applied to verify the Bayes risk consistency of the corresponding discrimination rules.

1. Introduction and summary. In this paper we present consistency results for the nonparametric regression function estimation problem. Assume that (X, Y), $(X_1, Y_1), \cdots, (X_n, Y_n)$ are independent identically distributed $\mathbb{R}^d \times \mathbb{R}$-valued random vectors with $E\{|Y|\} < \infty$. The purpose is to estimate the regression function

$$m(x) = E\{Y|X = x\}$$

from the *data*, $(X_1, Y_1), \cdots, (X_n, Y_n)$. The estimate studied here is the *kernel estimate*

$$m_n(x) = \frac{\Sigma_{i=1}^n Y_i k((X_i - x)/h_n)}{\Sigma_{i=1}^n k((X_i - x)/h_n)} \tag{1}$$

first proposed by Watson (1964) and Nadaraya (1964). In (1) k is a nonnegative function on $\mathbb{R}^d$ and $\{h_n\}$ is a sequence of positive numbers. The pointwise consistency of (1) is discussed by Watson (1964), Nadaraya (1964, 1965), Rosenblatt (1969), Schuster (1972), Greblicki (1974) and Noda (1976). The uniform consistency is treated in the papers by Nadaraya (1965, 1970), Greblicki (1974) and Devroye (1978a).

Received July 1978; revised December 1978.
[1]This research was sponsored by AFOSR Grant 77-3385.
AMS 1970 *subject classifications.* Primary 62H30; secondary 62G05, 60F15.
Key words and phrases. Nonparametric discrimination, consistency, regression function, Bayes risk consistency, kernel estimate.

In this note we are interested in the convergence to 0 of

$$J_{np} = E\{\int |m_n(x) - m(x)|^p \mu(dx)\}, \tag{2}$$

where μ is the probability measure of X.

THEOREM 1. *If*

$$E\{|Y|^p\} < \infty \qquad p \geq 1, \tag{3}$$

$$h_n \to_n 0, \tag{4}$$

$$nh_n^d \to_n \infty, \tag{5}$$

and if k *satisfies*

(6)
- (i) k *is a nonnegative function on* $\mathbb{R}^d$ *bounded by* $k^* < \infty$;
- (ii) k *has compact support* A;
- (iii) $k \geq \beta I_B$ *for some* $\beta > 0$ *and some closed sphere* B *centered at the origin and having positive radius* (I *is the indicator function*),

then $J_{np} \to_n 0$.

Notice that no conditions whatever are put on the underlying distribution of (X, Y) except for the necessary condition $E\{|Y|^p\} < \infty$. Recently, Stone (1977) showed that the nearest neighbor regression function estimates also have $J_{np} \to_n 0$ for all possible distributions of (X, Y). (For other consistency properties, see Royall (1966), Cover (1968), and Devroye (1978b.) This same property is also shared by some regression function estimates that are based upon the partitioning of $\mathbb{R}^d$ into blocks (Gordon and Olshen (1978), see also Mahalanobis (1961), Parthasarathy and Bhattacharya (1961), Anderson (1966)).

2. Development. The proof of Theorem 1 is based on the inequalities developed in Lemmas 1 and 2. In (1) and subsequent expressions, we arbitrarily define 0/0 to be 0.

LEMMA 1. *Let* k *be a function on* $\mathbb{R}^d$ *satisfying* (6). *Then there exists a constant* $0 < \gamma < \infty$ *only depending upon* k *such that for* all $z \in \mathbb{R}^d$, all $r > 0$ *and* all *probability measures* μ *on the Borel sets of* $\mathbb{R}^d$,

$$\int \frac{k((z-x)/r)}{\int k((y-x)/r)\mu(dy)} \mu(dx) \leq \gamma.$$

PROOF OF LEMMA 1. We define the set $a + bC$, $a \in \mathbb{R}^d$, $b \in \mathbb{R}$, $C \subseteq \mathbb{R}^d$ by $\{x | x = a + by, y \in C\}$. We will use the symbols k^*, β, A, B defined in (6). The sphere B has a positive radius $\rho > 0$.

First we find a finite cover $\{A_1, \cdots, A_N\}$ of A by translates of $B/2$. Because this is a cover we have

$$k((z-x)/r) \leqslant k^* I_{z-rA}(x) \leqslant \Sigma_{i=1}^N k^* I_{z-rA_i}(x).$$

Further, for every $x \in z - rA_i$, $z - rA_i \subseteq x + rB$. Thus,

$$\int k((y-x)/r)\mu(dy) \geqslant \beta\mu(x+rB) \geqslant \beta\mu(z-rA_i)$$

for such x. Consequently,

$$\int \frac{k((z-x)/r)}{\int k((y-x)/r)\mu(dy)}\mu(dx) \leqslant \Sigma_{i=1}^N \frac{k^*\mu(z-rA_i)}{\beta\mu(z-rA_i)} = Nk^*/\beta,$$

independently of z, r and μ.

LEMMA 2. *Let k be a function satisfying* (6), *and let* (X, Y), (X_1, Y_1), $\cdots$, (X_n, Y_n) *be a sequence of independent, identically distributed random vectors from* $\mathbb{R}^d \times \mathbb{R}$ *where X has probability measure μ.*

For all $r > 0$ and all $n \geqslant 1$,

$$E\{\Sigma_{i=1}^n |Y_i| k((X_i - X)/r)/\Sigma_{j=1}^n k((X_j - X)/r)\} \leqslant 7\gamma E\{|Y|\} \tag{8}$$

where γ is the constant of Lemma 1.

PROOF OF LEMMA 2. We need only consider $E\{|Y|\} < \infty$. We assume that $n \geqslant 8$ since (8) is clearly satisfied for $n \leqslant 7$ because γ in Lemma 1 is greater than 1. Let

$$N(x) = \Sigma_{j=2}^n k((X_j - x)/r)$$

so that $N_1(x) = E\{N(x)\} = (n-1)\int k((y-x)/r)\mu(dy)$. Noting that Var $\{N(x)\} \leqslant (n-1)k^*\int k((y-x)/r)\mu(dy)$, we have

$$\begin{aligned} P\{N(x) < N_1(x)/2\} &= P\{N(x) - N_1(x) < -N_1(x)/2\} \\ &\leqslant \min\{1, 4k^*/N_1(x)\}, \end{aligned} \tag{9}$$

by Chebyshev's inequality. Since $N_1(X_1) = E\{N(X_1)/X_1\}$, we rewrite (8) as

$$\begin{aligned} &\Sigma_{i=1}^n E\{|Y| k((X - X_i)/r)/\left(\Sigma_{j \neq i} k((X_j - X_i)/r) + k((X - X_i)/r)\right)\} \\ &\quad = nE\{|Y| k((X - X_1)/r)/\left(\Sigma_{j=2}^n k((X_j - X_1)/r) + k((X - X_1)/r)\right)\} \\ &\quad \leqslant nE\{|Y| k((X - X_1)/r)/\left(N_1(X_1)/2 + k((X - X_1)/r)\right)\} \\ &\qquad + nE\{|Y| I_{[N(X_1) < N_1(X_1)/2;\, k((X - X_1)/r) > 0]}\}. \end{aligned}$$

The first term in the last expression is upper bounded by

$$\begin{aligned} &nE\{|Y| \min\{1, 2k((X - X_1)/r)/N_1(X_1)\}\} \\ &\quad = nE\{|Y| E\{\min\{1, 2k((X - X_1)/r)/N_1(X_1)\} | X\}\} \\ &\quad \leqslant nE\{|Y| \sup_z E\{\min(1, 2k((z - X_1)/r)/N_1(X_1))\}\} \\ &\quad = nE\{|Y|\} \sup_z \int \min\left\{1, \frac{2k((z-x)/r)}{N_1(x)}\right\} \mu(dx). \end{aligned}$$

The second term is upper bounded by

$$nE\left\{|Y|I_A\left(\frac{X-X_1}{r}\right)P\{N(X_1)<N_1(X_1)/2|X_1\}\right\}$$

$$\leqslant nE\left\{|Y|I_A\left(\frac{X-X_1}{r}\right)\min\left\{1,\frac{4k^*}{N_1(X_1)}\right\}\right\}$$

$$\leqslant nE\left\{|Y|\min\left\{1,\frac{4k^*I_A((X-X_1)/r)}{N_1(X_1)}\right\}\right\}$$

$$\leqslant nE\{|Y|\}\sup_z\int\min\left\{1,\frac{4k^*I_A((z-x)/r)}{N_1(x)}\right\}\mu(dx)$$

where we used (9) and the previous argument. Combining both inequalities with Lemma 1, we upper bound the left-hand-side of (8) by:

$$6\frac{n}{n-1}\gamma E\{|Y|\}\leqslant 7\gamma E\{|Y|\}.$$

COROLLARY. *By Jensen's inequality, we have for all* $p\geqslant 1$,

$$E\left\{\left[\Sigma_{i=1}^n|Y_i|k((X_i-X)/r)/\Sigma_{j=1}^n k((X_j-X)/r)\right]^p\right\}$$

$$\leqslant E\left\{\Sigma_{i=1}^n|Y_i|^p k((X_i-X)/r)/\Sigma_{j=1}^n k((X_j-X)/r)\right\}\leqslant 7\gamma E\{|Y|^p\}.$$

The constant $\gamma>0$ is a covering constant because it is proportional to the number of spheres $B/2$ needed to cover a compact set A. In the next lemma another facet of the covering problem is used.

LEMMA 3. *Let k be a nonnegative function on $\mathbb{R}^d$ satisfying* (6) (iii). *If* (5) *holds, then*

$$P\{n\int k((y-X)/h_n)\mu(dy)<c\}\to_n 0$$

for <u>*all*</u> $c>0$ *and* <u>*all*</u> *probability measures* μ *for* X.

PROOF OF LEMMA 3. Since $k\geqslant\beta I_B$ for some $\beta>0$ and some closed sphere B centered at the origin with radius $\rho>0$, we have for $x\in\mathbb{R}^d$,

$$n\int k((y-x)/h_n)\mu(dy)\geqslant n\beta\mu(x+h_nB).$$

A sphere of radius r can be covered by $\max(4, 4dr/s)^d$ closed spheres of radius $s/2$. To see this, construct a set C of points $a_i=is/d$, $i=0,\pm1,\pm2,\cdots$, with $|a_i|\leqslant r$. Add to C the end points $-r$ and $+r$. The grid C^d has at most $(2+2dr/s)^d\leqslant\max(4,4dr/s)^d$ points. We show that the spheres with radius $s/2$ centered at the points in C^d cover the sphere $S(0, r)$ centered at 0 with radius r. For each $x\in S(0,r)$ we have $\|x-a\|\leqslant\Sigma_{i=1}^d s/2d=s/2$ for some $a\in C^d$.

Find r so large that $1 - \mu(S(0, r)) < \varepsilon$. If $S_1, S_2, \cdots$ are the spheres of radius $\rho h_n/2$ covering $S(0, r)$, then $x \in S_i$ implies $S_i \subset x + h_n B$. Consequently,

$$
\begin{aligned}
P\{n \textstyle\int k((y - X)/h_n)\mu(dy) < c\} \\
&\leqslant P\{n\beta\mu(X + h_n B) < c\} \\
&\leqslant P\{X \notin S(0, r)\} + P\{X \in S(0, r);\, n\beta\mu(X + h_n B) < c\} \\
&< \varepsilon + \Sigma_i P\{X \in S_i;\, n\beta\mu(X + h_n B) < c\} \\
&\leqslant \varepsilon + \Sigma_{i\,:\,P\{X \in S_i\} < c/\beta n} P\{X \in S_i\} \\
&\leqslant \varepsilon + 4^d c/\beta n + (4\, dr/\rho h_n)^d c/\beta n \\
&\leqslant 2\varepsilon \quad \text{for } n \text{ large enough.}
\end{aligned}
$$

The lemma follows by the arbitrariness of $\varepsilon > 0$.

We now prove Theorem 1. One of the facts crucial to the proof is the denseness of all bounded continuous functions in $L_p(\mu)$, a property also exploited by Stone (1977) in his consistency proof for nearest neighbor estimates.

Proof of Theorem 1. For any function $g : \mathbb{R}^d \to \mathbb{R}$ we have $|m_n(x) - m(x)|^p \leqslant (U_1^p(x) + U_2^p(x) + U_3^p(x) + U_4^p(x))4^{p-1}$ where now

$$N(x) = \Sigma_i k((X_i - x)/h_n),$$

$$U_1(x) = N(x)^{-1}|\Sigma_i(Y_i - m(X_i))k((X_i - x)/h_n)|,$$

$$U_2(x) = N(x)^{-1}\Sigma_i|m(X_i) - g(X_i)|k((X_i - x)/h_n),$$

$$U_3(x) = |N(x)^{-1}\Sigma_i g(X_i)k((X_i - x)/h_n) - g(x)|,$$

and

$$U_4(x) = |g(x) - m(x)|.$$

We will show that $E\{\int U_i^p(x)\mu(dx)\} < \varepsilon$ for n large enough, $i = 1, 2, 3, 4$.

Since m^p is μ-integrable, we can find a function g that is bounded, continuous, and zero outside a compact set such that $\int U_4^p(x)\mu(dx) < \varepsilon$ (Dunford and Schwartz, 1957, page 298).

By the corollary to Lemma 2

$$E\{U_2^p(X)\} = \int E\{U_2^p(x)\}\,\mu(dx) \leqslant 7\gamma E\{|m(X) - g(X)|^p\} < \varepsilon$$

by the choice of g.

Let g be fixed and put $c_g = \sup_x|g(x)|^p$. We can find $\delta > 0$ so small that

$$\sup_{x \in \mathbb{R}^d;\, y \in x + \delta A}|g(y) - g(x)| < (\varepsilon/2)^{1/p},$$

where A is the support of k. Thus $U_3^p(x) < \varepsilon/2$ when $h_n < \delta$ and $N(x) > 0$. Thus, if $h_n < \delta$,

$$\begin{aligned} E\{U_3^p(X)\} &\leq c_g \int P\{N(x) = 0\}\,\mu(dx) + \varepsilon/2 \\ &\leq c_g \int (1 - \mu(x + h_n A))^n \mu(dx) + \varepsilon/2 \\ &\leq c_g \int e^{-n\mu(x+h_nA)}\mu(dx) + \varepsilon/2 \\ &\leq c_g\big(e^{-\lambda} + P\{n\mu(X + h_nA) < \lambda\}\big) + \varepsilon/2 \end{aligned}$$

where $\lambda > 0$ is picked large enough to make $c_g e^{-\lambda}$ smaller than $\varepsilon/4$. Letting n grow large and applying Lemma 3 with $k = I_A$ to the term $P\{n\mu(X + h_nA) < \lambda\}$ shows that $E\{U_3^p(x)\} < \varepsilon$ for all n large enough.

Next, we use the fact that conditioned on $X_1, \cdots, X_n$, the random variables $Y_1 - m(X_1), \cdots, Y_n - m(X_n)$ are independent. Assume first that $|Y_1| \leq c_t < \infty$ a.s. (so that $|Y_1 - m(X_1)| \leq 2c_t$ a.s.). Then for $p \leq 2$

$$\begin{aligned} \left(E\left\{\int U_1^p(x)\mu(dx)\right\}\right)^{2/p} &\leq \int E\{U_1^2(x)\}\,\mu(dx) \\ &= \int E\{E\{U_1^2(x)|X_1, \cdots, X_n\}\}\,\mu(dx) \\ &= \int E\Big\{\Sigma_i E\{(Y_i - m(X_i))^2 | X_i\} \\ &\qquad k^2((X_i - x)/h_n)/N^2(x)\Big\}\,\mu(dx) \\ &\leq 4c_t^2 \int E\{\min\{1; k^*/N(x)\}\}\,\mu(dx) \\ &\leq c_t^2 k^*/c_s + 4c_t^2 \int P\{N(x) < c_s\}\,\mu(dx) \\ &\leq 4c_t^2 k^*/c_s + 4c_t^2 \int P\{N(x) < E\{N(x)\}/2\}\,\mu(dx) \\ &\quad + 4c_t^2 \int I_{(0,2c_s)}(E\{N(x)\})\mu(dx). \end{aligned}$$

For any c_s, c_t, the last term tends to 0 as $n \to \infty$ by Lemma 3. The first term can be made arbitrarily small by choosing c_s large enough. For the middle term, which is estimated as in the proof of Lemma 2 (see (9)) by

$$4c_t^2 \int \min\{1; 4k^*/E\{N(x)\}\}\,\mu(dx) \leq 4c_t^2 4k^*/c_s + 4c_t^2 \int I_{(0,c_s)}(E\{N(x)\})\mu(dx),$$

we have already demonstrated that it is small for large n and large c_s.

For $p > 2$, use the facts that

$$U_1^p(x) = U_1^2(x)U_1^{p-2}(x) \leq (2c_t)^{p-2}U_1^2(x)$$

and

$$\int E\{U_1^p(x)\}\,\mu(dx) \leq (2c_t)^{p-2}\int E\{U_1^2(x)\}\,\mu(dx),$$

and proceed similarly.

To complete the proof of Theorem 1, we only have to show that $E\{U_1^p(x)\}$ can be made arbitrarily small even if Y_1 is not a.s. bounded. Assume that $c_t > 0$ is a constant, and let $Y_i = Y_i' + Y_i''$ where

$$Y_i' = Y_i I_{[-c_t, c_t]}(Y_i)$$

and

$$Y_i'' = Y_i I_{[-c_t, c_t]^c}(Y_i).$$

Further, let $m'(x) = E\{Y_1'|X_1 = x\}$, $m''(x) = E\{Y_1''|X_1 = x\}$, and notice that $m(x) = m'(x) + m''(x)$ for almost all $x(\mu)$. We have for almost all $x(\mu)$:

$$\begin{aligned} U_1^p(x) &\leq 2^{p-1}\big(N(x)^{-1}|\Sigma_{i=1}^n(Y_i' - m'(X_i))k((X_i - x)/h_n)|\big)^p \\ &\quad + 2^{p-1}\big(N(x)^{-1}\Sigma_{i=1}^n|Y_i'' - m''(X_i)|k((X_i - x)/h_n)\big)^p \\ &\equiv U_1'^p(x) + U_1''^p(x). \end{aligned}$$

It is clear from the previous argument (since $|Y_1' - m(X_1')| \leq 2c_t$ a.s.) that for any $c_t > 0$, $E\{\int U_1'^p(x)\mu(dx)\}$ can be made arbitrarily small. For the last term we use

$$\begin{aligned} E\{\textstyle\int U_1''^p(x)\mu(dx)\} &\leq 7\gamma E\{|Y_1'' - m''(X_1)|^p\}2^{p-1} \\ &\leq 7\gamma 2^{2p-2}\big(E\{|Y_1|^p I_{[-c_t, c_t]^c}(Y_1)\} + E\{|m''(X_1)|^p\}\big) \\ &\leq 4\gamma 2^{2p} E\{|Y_1|^p I_{[-c_t, c_t]^c}(Y_1)\} \\ &\to 0 \quad \text{as} \quad c_t \to \infty \end{aligned}$$

by the finiteness of $E\{|Y_1|^p\}$.

3. Bayes risk consistency in discrimination. In discrimination Y takes values from a known finite set $\{1, \cdots, M\}$ and the problem, as before, is to estimate Y from X. The estimate $g_n(X)$ is a measurable function of X with values in $\{1, \cdots, M\}$ and the performance of the estimate with the data is now measured by the probability of error,

$$L_n = P\{g_n(X) \neq Y|X_1, Y_1, \cdots, X_n, Y_n\}.$$

Clearly, L_n cannot be smaller than the Bayes probability of error

$$L^* = \inf_{g : \mathbb{R}^d \to \{1, \cdots, M\}} P\{g(X) \neq Y\}.$$

If we define

$$p_i(x) = P\{Y = i|X = x\}, \qquad 1 \leq i \leq M, \quad x \in \mathbb{R}^d,$$

then all discrimination rules g satisfying

$$g(x) \neq i \quad \text{whenever} \quad p_i(x) < \max_{1 \leq l \leq M} p_e(x)$$

have probability of error L^*. The unknown regression functions p_i can be estimated by any method. Writing p_{ni} for the estimate of p_i, we can in turn pick g_n such that

(11) $$g_n(x) \neq i \quad \text{whenever} \quad p_{ni}(x) < \max_{1 \leq l \leq M} p_{nl}(x).$$

For *all* discrimination rules satisfying (11) we have

(12) $$0 \leq L_n - L^* \leq 2\Sigma_{i=1}^M \int |p_i(x) - p_{ni}(x)|\,\mu(dx).$$

Inequality (12) is easy to show (see, e.g., Stone (1977)). If we write $a(x)$ for $p_i(x)$ with $i = g_n(x)$, and $a_n(x)$ for $p_{ni}(x)$ when $i = g_n(x)$, then it is true that $L^* = \int(1 - \max_i p_i(x))\mu(dx)$, $L_n = \int(1 - a(x))\mu(dx)$, $a_n(x) = \max_i p_{ni}(x)$ and

$$\begin{aligned} L_n - L^* &= \int(\max_i p_i(x) - a(x))\mu(dx) \\ &= \int(\max_i p_i(x) - \max_i p_{ni}(x))\mu(dx) \\ &\quad + \int(a_n(x) - a(x))\mu(dx) \\ &\leqslant 2\Sigma_{i=1}^{M}\int|p_i(x) - p_{ni}(x)|\,\mu(dx). \end{aligned}$$

This proves (12). The inequality (12) links in a very simple way the distance in L_1 between the p_{ni} and the p_j with $L_n - L^*$. For instance, if we use (1) as our regression function estimate (i.e., to estimate p_j, replace Y_i in (1) by $I_{[Y_i=j]}$), then the condition (11) reduces to

(13) $$g_n(x) \neq i$$

whenever

$$\Sigma_{j\,:\,Y_j=i}k((X_j - x)/h_n) < \max_{1\leqslant l\leqslant M}\Sigma_{j\,:\,Y_j=l}k((X_j - x)/h_n).$$

If k is the indicator function of the unit sphere centered at the origin, then (13) is equivalent to taking a majority vote with those Y_j for which $\|X_j - x\| \leqslant h_n$. This simple rule can be traced back to the work of Fix and Hodges (1951). The following theorem is a direct corollary of inequality (12) and Theorem 1.

THEOREM 2. (*Bayes risk consistency*). *All discrimination rules satisfying* (13) *are Bayes risk consistent* (*that is*, $E\{L_n\} \to_n L^*$) *if* (4–6) *hold.*

Theorem 2 is entirely distribution-free: no restrictions are put on the distribution of (X, Y). This result may seem a bit surprising because (13) was originally obtained in the literature for the Parzen density estimate under the assumption that X has a density (Van Ryzin (1966), Glick (1972, 1976), Greblicki (1974, 1977), Devroye and Wagner (1976, 1977)). In all but the last of the cited papers, additional continuity conditions were put on the density of X to prove Bayes risk consistency.

REFERENCES

ANDERSON, T. W. (1966). Some nonparametric multivariate procedures based on statistically equivalent blocks. In *Multivariate Analysis* (P. R. Krishnaiah, ed.) 5–27, Academic Press, New York.

COVER, T. M. (1968). Estimation by the nearest neighbor rule. *IEEE Trans. on Information Theory* **IT-14** 50–55.

DEVROYE, L. P. (1978a). The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Transactions on Information Theory* **IT-24** 142–151.

DEVROYE, L. P. (1978b). The uniform convergence of the Nadaraya-Watson regression function estimate. *Canad. J. Statist.* **6** 179–191.

DEVROYE, L. P. and WAGNER, T. J. (1976). Nonparametric discrimination and density estimation. Technical Report 183, Electronics Research Center, The University of Texas at Austin.

DEVROYE, L. P. and WAGNER, T. J. (1977). On the L_1 convergence of kernel regression function estimators with applications in discrimination. *Wahrscheinlichkeitstheorie und Verw. Gebiete.* To appear.

DUNFORD, N. and SCHWARTZ, J. T. (1957). *Linear Operators, Part I. General Theory*. Interscience, New York.

FIX, E. and HODGES, J. L. (1951). Discriminatory analysis. Nonparametric discrimination: consistency properties. Report 4. Project No. 21-49-004. USAF School of Aviation Medicine, Randolph Field, Texas.

GLICK, N. (1972). Sample-based classification procedures derived from density estimators. *J. Amer. Statist. Assoc.* **67** 116–122.

GLICK, N. (1976). Sample-based classification procedures related to empiric distributions. *IEEE Trans. on Information Theory* **IT-22** 454–461.

GORDON, L. and OLSHEN, R. A. (1978). Asymptotically efficient solutions to the classification problem. *Ann. Statist.* **6** 515–533.

GREBLICKI, W. (1974). Asymptotically optimal probabilistic algorithms for pattern recognition and identification. Monografie No. 3 Prace Naukowe Instytutu Cybernetyki Technicznej Politechniki Wroclawskiej No. 18, Wroclaw, Poland.

GREBLICKI, W. (1978). Asymptotically optimal procedures with density estimates. *IEEE Trans. on Information Theory* IT-24, 250–251.

MAHALANOBIS, P. C. (1961). A method of fractile graphical analysis. *Sankhyā, Series A* **23** 41–64.

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

NADARAYA, E. A. (1965). On nonparametric estimates of density functions and regression curves. *Theor. Probability Appl.* **10** 186–190.

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

NODA, K. (1976). Estimation of a regression function by the Parzen kernel-type density estimators. *Ann. Inst. Statist. Math.* **28** 221–234.

PARTHASARATHY, K. R. and BHATTACHARYA, P. K. (1961). Some limit theorems in regression theory. *Sankhyā, Series A* **23** 91–102.

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

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

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.

STEIN, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions*. Princeton Univ. Press.

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

VAN RYZIN, J. (1966). Bayes risk consistency of classification procedures using density estimation. *Sankhyā, Series A* **28** 161–170.

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

SCHOOL OF COMPUTER SCIENCE
McGILL UNIVERSITY
P. O. BOX 6070, STATION A
MONTREAL, CANADA H3C 3G1

DEPARTMENT OF ELECTRICAL ENGINEERING
UNIVERSITY OF TEXAS
AUSTIN, TEXAS 78712