The Annals of Statistics
1985, Vol. 13, No. 3, 1041–1049

A NOTE ON THE L_1 CONSISTENCY OF VARIABLE KERNEL ESTIMATES[1]

Dedicated to the Memory of Gerard Collomb

BY LUC DEVROYE

McGill University

A sample $X_1, \cdots, X_n$ of i.i.d. R^d-valued random vectors with common density f is used to construct the density estimate

$$f_n(x) = (1/n) \sum_{i=1}^n H_{ni}^{-d} K((x - X_i)/H_{ni}),$$

where K is a given density on R^d, and the H_{ni}'s are positive functions of n, i and $X_1, \cdots, X_n$ (but not of x). The H_{ni}'s can be thought of as locally adapted smoothing parameters. We give sufficient conditons for the weak convergence to 0 of $\int | f_n - f |$ for all f. This is illustrated for the estimate of Breiman, Meisel and Purcell (1977).

1. Introduction. Most consistent nonparametric density estimates have a built-in smoothing parameter. Numerous schemes have been proposed (see, e.g., references found in Rudemo, 1982; or Devroye and Penrod, 1984) for selecting the smoothing parameter as a function of the data only (a process called *automatization*), and for introducing locally adaptable smoothing parameters. In this note, we give conditions which insure that estimators of the form

$$f_n(x) = (1/n) \sum_{i=1}^n K_{H_{ni}}(x - X_i) \tag{1}$$

are weakly convergent in $L_1(R^d)$ to the common density f of $X_1, \cdots, X_n$, a sample of independent random vectors. In (1), K is a given density on R^d (*kernel*), $K_u(x) = u^{-d}K(x/u)$, $u > 0$, and $H_{ni} = H_{ni}(X_1, \cdots, X_n)$, $1 \le i \le n$, is a positive-valued function of i, n and $X_1, \cdots, X_n$. The H_{ni}'s can be thought of as locally adapted smoothing parameters, and (1) generalizes the *kernel estimate* (Rosenblatt, 1956; Parzen, 1962; Cacoullos, 1966). Note that the H_{ni}'s do not depend upon x, so that f_n is a density in x. Among estimators of the form (1), we cite the *Breiman-Meisel-Purcell estimate* (Breiman et al., 1977), or *variable kernel estimate*, where

$H_{ni} = \alpha$ times the distance between X_i and its kth nearest neighbor among $X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n$,

$\alpha > 0$ is a constant, and k_n is a sequence of positive integers.

The purpose of this note is (i) to obtain the L_1 convergence of (1) for *all* f under fairly weak conditions on the H_{ni}'s, and (ii) to prove that the variable kernel estimate converges in L_1 for *all* f under suitable conditions on the sequence k_n. We do not make any claims about rates of convergence; to obtain some sort

Received September 1984; revised March 1985.

[1] The research was supported by NSERC Grant A3456.

AMS 1980 *subject classifications*. Primary 60F15; secondary 62G05.

Key words and phrases. Nonparametric density estimation, consistency, variable kernel estimate, nearest neighbor, embedding.

of insurance against nonconsistency is all we want here. But this is precisely where the technical difficulties arise. For sufficiently smooth f, it is relatively straightforward to prove that (1) is convergent in L_1. To extend this result towards all f, it is not enough to invoke the theorem about the denseness of uniformly continuous functions in $L_1(R^d)$. Here, we propose a simple embedding argument that can be useful in other applications too.

THEOREM 1. *Let $\mathscr{F}$ be the collection of all densities on R^d, and let $\mathscr{F}_0$ be a collection of densities that is dense in $\mathscr{F}$ in the L_1 sense. Assume that there exists a sequence of functions h_n: $R^d \to [0, \infty)$ such that*

$$(2) \qquad \lim_{n\to\infty} h_n(x) = 0, \quad \text{for almost all} \quad x(f), \quad \text{all} \quad f \in \mathscr{F}_0;$$

$$(3) \qquad \lim_{n\to\infty} n \inf_x h_n^d(x) = \infty, \quad \text{for all} \quad f \in \mathscr{F}_0;$$

$$(4) \qquad \lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} \sup_{y\in S_{x\varepsilon}} |(h_n(y) - h_n(x))/h_n(x)| = 0, \quad \text{for almost all} \quad x(f), \quad \text{all} \quad f \in \mathscr{F}_0,$$

where $S_{x\varepsilon}$ is the closed sphere in R^d centered at x with radius ε. Assume furthermore that K decreases along rays (i.e., $K(ux) \le K(x)$, $u \ge 1$, $x \in R^d$), that

for all i,

$$(5) \qquad H_{ni}(X_1, \cdots, X_n) = H_{n1}(X_i, X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n),$$

and $H_{n1}(x_1, x_2, \cdots, x_n)$ is invariant under permutations of $x_2, \cdots, x_n$,

and that

$$(6) \qquad H_{n1}(x, X_2, \cdots, X_n)/h_n(x) \to 1 \quad \text{in probability}, \quad \text{for almost all} \quad x(f), \quad \text{all} \quad f \in \mathscr{F}.$$

Then, for estimate (1),

$$(7) \qquad \lim_{n\to\infty} E\left(\int |f_n - f|\right) = 0, \quad \text{for all} \quad f \in \mathscr{F}.$$

REMARK. The condition that K be a density which is decreasing along rays is not very restrictive. It is satisfied for the optimal kernels in R^d, and for all kernels K that are nonincreasing functions of $\|x\|$.

EXAMPLE 1. When $H_{ni} = H_n$ for all i, where H_n is a function of n and the data, invariant under permutations of the data, (7) follows if for some sequence of positive numbers h_n, we have $H_n/h_n \to 1$ in probability, and

$$(8) \qquad \lim_{n\to\infty} h_n = 0; \quad \lim_{n\to\infty} n h_n^d = \infty.$$

This result is strictly contained in a more general result of Devroye and Penrod (1984), but the proof is quite a bit shorter.

EXAMPLE 2. (The kernel estimate). When $H_{ni} = h_n$, where h_n is a sequence of positive numbers, then the conditions of Theorem 1 are satisfied when h_n is as in (8), and K decreases along rays. It is known that (8) is necessary and sufficient for weak convergence in the sense of (7) (Devroye, 1983; see also Abou-Jaoude, 1977; and Devroye and Wagner, 1979). Furthermore, the condition that K be decreasing along rays can be dropped altogether (Devroye, 1983).

EXAMPLE 3 (The variable kernel estimate). For the variable kernel estimate, the permutation invariance condition (5) is satisfied. In Theorem 1, take $\mathscr{F}_0 = \{$all continuous densities with compact support$\}$ (which is dense in $\mathscr{F}$ in the L_1 sense), and

$$h_n(x) = \alpha(k_n/nC_d f(x))^{1/d},$$

where C_d is the volume of the unit sphere in R^d. (The definition of $h_n(x)$ when $f(x) = 0$ is irrelevant, so we can set $h_n(x) = 1$ as well when $f(x) = 0$.) Clearly, (2) and (3) are equivalent to

$$\text{(9)} \qquad \lim_{n\to\infty}(k_n/n) = 0, \quad \lim_{n\to\infty}k_n = \infty.$$

Condition (4) holds for all x with $f(x) > 0$, by the continuity of f. Thus, we need only verify condition (6). We observe now that if f_n^* denotes the *nearest neighbor density estimate* based on $X_2, \cdots, X_n$ (Fix and Hodges, 1951; Loftsgaarden and Quesenberry, 1965), then we can write

$$\text{(10)} \qquad f_n^*(x) = k_n/nC_d(H_{n1}(x, X_2, \cdots, X_n)/\alpha)^d,$$

and thus, $H_{n1}(x, X_2, \cdots, X_n)/h_n(x) = (f(x)/f_n^*(x))^{1/d}$. Thus, (6) is equivalent to the almost everywhere convergence of the nearest neighbor estimate. In the literature, only convergence at continuity points of f is given (Wagner, 1973; Moore and Yackel, 1977; Devroye and Wagner, 1976; Mack and Rosenblatt, 1979). Thus, we include a short proof of this result here (see Theorem 2 below, and its proof in Section 3). The full statement about the L_1 consistency of the variable kernel estimate is given in Theorem 3.

THEOREM 2. *Let $f_n^*(x)$ be $k_n/(nC_dD_n^d(x))$ where $D_n(x)$ is the distance between x and its k_nth nearest neighbor among $X_1, \cdots, X_n$, and k_n is a sequence of integers satisfying* (9). *Then $f_n^*(x) \to f(x)$ in probability for almost all x.*

THEOREM 3. *Let f_n be the variable kernel estimate with arbitrary constant $\alpha > 0$, with kernel K decreasing along rays, and with k_n as in* (9). *Then, for* all f,

$$\lim_{n\to\infty}E\left(\int |f_n - f|\right) = 0.$$

2. Proof of Theorem 1. Throughout this section, the conditions of Theorem 1 are assumed to hold. We will need Scheffé's theorem (Scheffé, 1947), which states that if g_n is a sequence of densities converging at almost all x to f, then $\int |g_n - f| \to 0$ as $n \to \infty$.

LEMMA 1. *It suffices to prove* (7) *for all kernels K that decrease along rays, are continuous and vanish outside a compact set.*

PROOF OF LEMMA 1. Consider f_n as in (1) with kernel K, and $f_n^\dagger$ as in (1) with kernel $K^\dagger$. Then

$$\int |f_n - f_n^\dagger| \le \frac{1}{n} \sum_{i=1}^n \int |K_{H_{ni}}(x - X_i) - K_{H_{ni}}^\dagger(x - X_i)|\, dx = \int |K - K^\dagger|.$$

Thus, it suffices to show that the kernels of Lemma 1 are dense (in the L_1 sense) in the class of kernels of Theorem 1. This can be done by construction. First, we construct a function K^* as follows:

$$K^*(x) = \int_A K(y)\, dy \bigg/ \int_A dy,$$

where

$$A = (S_{\|x\|(1+\delta)} - S_{\|x\|}) \cap B_\delta, \quad S_u = \text{sphere } S_{0u},$$

and B_δ is the cone of opening δ centered at 0 around the axis joining 0 and x, and $\delta > 0$ is a small positive constant.

Each K_δ^* is continuous except possibly at 0, and each K_δ^* decreases along rays. Futhermore, by the Lebesque density theorem (see, e.g., Wheeden and Zygmund, 1977), $K_\delta^* \to K$ as $\delta \to 0$ for almost all x. Thus, by Scheffé's theorem, $\lim_{\delta\downarrow 0} \int |K - K^*/\int K^*| = 0$. The construction is complete if we can take care of the continuity at 0 and the compact support without upsetting the continuity or monotonicity conditions. First approximate K_δ^* by $\min(K_\delta^*, M)$ where M is a large positive number. Then multiply this new function with a function $L(x)$ satisfying all the conditions of Lemma 1, and taking the value 1 on S_M for a large constant M. This function can be forced to vanish outside S_{2M} and to be continuous in-between. This concludes the proof of Lemma 1.

LEMMA 2. *It suffices to prove* (7) *for kernels as in Lemma* 1, *and for the (artificial) estimator*

$$g_n(x) = (1/n) \sum_{i=1}^n K_{h_n(X_i)}(x - X_i). \tag{11}$$

REMARK. Estimator (11) is quite a lot easier to handle than (1) because the summands are independent. Clearly, it is in the proof of Lemma 2 that we will use conditions (6) and (5) about the H_{ni}'s.

PROOF OF LEMMA 2. Define the function $\omega(u)$ by $\int |K - K_u|$, and note that by the continuity of K and Scheffé's theorem $\lim_{u\to 1} \omega(u) = 0$. Also, $\omega(u) \le 2$, for all u. Now,

$$\begin{aligned} \int |f_n - g_n| &\le \frac{1}{n} \sum_{i=1}^n \int |K_{H_{ni}}(x - X_i) - K_{h_n(X_i)}(x - X_i)|\, dx \\ &= \frac{1}{n} \sum_{i=1}^n \int |K(x) - K_{h_n(X_i)/H_{ni}}(x)|\, dx = \frac{1}{n} \sum_{i=1}^n \omega\left(\frac{h_n(X_i)}{H_{ni}}\right). \end{aligned} \tag{12}$$

By condition (5), each $h_n(X_i)/H_{ni}$ is distributed as $h_n(X_1)/H_{n1}$, and thus, $E(\int |f_n - g_n|) \to 0$ for all f if

$$\lim_{n\to\infty} E(\omega(h_n(X_1)/H_{n1})) = 0,$$

for all f. By the Lebesgue dominated convergence theorem, it is clearly sufficient that $h_n(x)/H_{n1}(x, X_2, \cdots, X_n) \to 1$ in probability for almost all x and all f, but this is precisely condition (6).

LEMMA 3. *It suffices to prove that for the estimator* (11) *with kernels as in Lemma* 1, *we have*

$$\lim_{n\to\infty} E\left(\int |g_n - f|\right) = 0, \quad \textit{for all} \quad f \in \mathscr{F}_0. \tag{13}$$

REMARK. Lemma 3 is crucial. It tells us that we need only prove the consistency of g_n on a nice subclass of densities that is dense in $\mathscr{F}$, such as the class of all uniformly continuous densities with compact support. The proof of Lemma 3 is based upon embedding.

PROOF OF LEMMA 3. *The embedding device.* Let $f_n(x, X_1, \cdots, X_n) \in L_1(R^d)$ be a density estimate of f based upon a sample $X_1, \cdots, X_n$ of i.i.d. random vectors with common density f. Then, for another density g and corresponding sample $X_1', \cdots, X_n'$,

$$\begin{aligned} &\int |f_n(x, X_1, \cdots, X_n) - f(x)|\, dx \\ &\quad \le \int |f_n(x, X_1, \cdots, X_n) - f_n(x, X_1', \cdots, X_n')|\, dx \\ &\qquad + \int |f_n(x, X_1', \cdots, X_n') - g(x)|\, dx + \int |g(x) - f(x)|\, dx. \end{aligned} \tag{14}$$

In (14), the dependence between $(X_1, \cdots, X_n)$ and $(X_1', \cdots, X_n')$ is unrestricted. Next, define $\Delta = \int (f - \min(f, g))$. By geometrical considerations, $\int |f - g| = 2\Delta$, $\int \min(f, g) = 1 - \Delta$ and $\int (g - \min(f, g)) = \Delta$. Define also the densities

$$\psi_{\min} = \min(f, g)/(1 - \Delta),$$

$$\psi_f = (f - \min(f, g))/\Delta, \quad \psi_g' = (g - \min(f, g))/\Delta.$$

Next, consider three independent samples of i.i.d. random vectors:

$$U_1, U_2, \cdots, U_n \quad \text{(common density } \psi_{\min});$$

$$V_1, V_2, \cdots, V_n \quad \text{(common density } \psi_f);$$

$$W_1, W_2, \cdots, W_n \quad \text{(common density } \psi_g).$$

Also, let N be a binomial (n, Δ) random variable independent of the three samples, and let $(\sigma_1, \cdots, \sigma_n)$ be a random permutation of $(1, \cdots, n)$, independent of N and the three samples. If we identify

$$(X_1, \cdots, X_n) = (U_1, \cdots, U_{n-N}, V_1, \cdots, V_N),$$
$$(X_1', \cdots, X_n') = (U_1, \cdots, U_{n-N}, W_1, \cdots, W_N),$$

then it is clear that $(X_{\sigma_1}, \cdots, X_{\sigma_n})$ is distributed as a sample of i.i.d. random vectors drawn from f, and that $(X'_{\sigma_1}, \cdots, X'_{\sigma_n})$ is distributed as a sample of i.i.d. random vectors drawn from g.

Let g_n be the estimator (11). Then

$$\int | g_n(x, X_{\sigma_1}, \cdots, X_{\sigma_n}) - g_n(x, X'_{\sigma_1}, \cdots, X'_{\sigma_n}) | \, dx$$
$$\leq \frac{1}{n} \sum_{i=1}^N \int | K_{h_n(V_i)}(x - V_i) - K_{h_n(W_i)}(x - W_i) | \, dx \leq \frac{2N}{n}.$$

Since (11) is permutation invariant, we can drop the random permutation to make the notation simpler. Thus, by (14),

$$E\left(\int | g_n(x, X_1, \cdots, X_n) - f(x) | \, dx\right)$$
$$\leq \frac{2E(N)}{n} + E\left(\int | g_n(x, X_1', \cdots, X_n') - g(x) | \, dx\right) + \int | g(x) - f(x) | \, dx \tag{15}$$
$$= 2\int | g - f | + E\left(\int | g_n(x, X_1', \cdots, X_n') - g(x) | \, dx\right).$$

By (15), and the denseness of $\mathscr{F}_0$, (13) would imply $\lim_{n\to\infty} E(\int | g_n - f |) = 0$ for *all* f, which is all that is needed (Lemma 2).

Theorem 1 is proved if we can show

LEMMA 4. (13) *holds for all kernels as in Lemma* 1, *and all sequences of functions* h_n *satisfying* (2)–(4).

PROOF OF LEMMA 4. It suffices to show that $g_n - f \to 0$ in probability at all points x at which $f(x) > 0$, and conclude from Glick's extension of Scheffé's theorem that $\int | g_n - f | \to 0$ in probability, and thus that $E(\int | g_n - f |) \to 0$. Assume that we have shown that $E(g_n) \to f$ for all x with $f(x) > 0$. Then, note that

$$g_n(x) - E(g_n(x)) = (1/n) \sum_{i=1}^n (K_{h_n(X_i)}(x - X_i) - E(K_{h_n(X_i)}(x - X_i)))$$

is a zero mean random variable with variance not exceeding

$$\frac{1}{n} E(K^2_{h_n(X_1)}(x - X_1)) \leq \| K \|_\infty E\left(\frac{K_{h_n(X_1)}(x - X_1)}{nh_n^d(X_1)}\right) \leq \| K \|_\infty \frac{E(g_n(x))}{n \inf_y h_n^d(y)}.$$

In view of (3), the variance tends to 0, and thus, by Chebyshev's inequality, $g_n - E(g_n) \to 0$ in probability when $f(x) > 0$.

We will now prove that $E(g_n) \to f$ when $f > 0$. Let K vanish outside S_{0c} and let S denote the support of f. The point x is fixed throughout. For arbitrary $\varepsilon > 0$, we find n_0 and σ such that for $y \in S_{x\delta}$, $n \ge n_0$,

$$|h_n^d(y) - h_n^d(x)|/h_n^d(x) < \varepsilon, \quad |f(y) - f(x)|/f(x) < \varepsilon$$

(use Condition (4)). Thus, for $y \in S \cap S_{x\delta}$,

$$\frac{1}{(1+\varepsilon)h_n^d(x)} K\left(\frac{x-y}{h_n(x)(1-\varepsilon)^{1/d}}\right) \le K_{h_n(y)}(x-y) \le \frac{1}{(1-\varepsilon)h_n^d(x)} K\left(\frac{x-y}{h_n(x)(1+\varepsilon)^{1/d}}\right).$$

Thus,

$$\begin{aligned} E(g_n) &= \int f(y)K_{h_n(y)}(x-y)\,dy \ge \int_{S\cap S_{x\delta}} f(y)K_{h_n(y)}(x-y)\,dy \\ &\ge f(x)(1-\varepsilon)\int_{S\cap S_{x\delta}} K_{h_n(y)}(x-y)\,dy \ge \frac{f(x)(1-\varepsilon)^2}{1+\varepsilon}. \end{aligned} \tag{16}$$

Also,

$$\begin{aligned} E(g_n) &\le \int_{S\cap S_{x\delta}} f(y)K_{h_n(y)}(x-y)\,dy + \int_{S\cap S_{x\delta}^c} f(y)K_{h_n(y)}(x-y)\,dy \\ &\le \frac{f(x)(1+\varepsilon)^2}{1-\varepsilon} + \|f\|_\infty \|K\|_\infty \int_{y\in S, \delta<\|x-y\|\le ch_n(y)} h_n^{-d}(y)\,dy. \end{aligned} \tag{17}$$

The last integral in (17) does not exceed

$$\int_{y\in S,\delta<\|x-y\|\le ch_n(y)} \frac{c^d}{\|x-y\|^d}\,dy. \tag{18}$$

The function $\|x-y\|^{-d}I_{[y\in S,\|x-y\|>\delta]}$ is integrable. Since for almost all y, $h_n(y) \to 0$ (condition (2)), we conclude by the Lebesgue dominated convergence theorem that (18) is $o(1)$. Combining (16) and (17) shows that $E(g_n) \to f$ whenever $f > 0$ and $f \in \mathscr{F}_0$. This concludes the proof of Lemma 4 and Theorem 1.

3. Proof of Theorem 2. Fix x, and let A_n denote the sphere centered at x with radius $D_n(x)$. Let μ be the probability measure corresponding to f, and let λ be Lebesgue measure. We will use the following convenient (but unorthodox) decomposition: $f_n^*(x) = Y_nZ_n$ where $Y_n = (k_n/n\mu(A_n))$ and $Z_n = \mu(A_n)/\lambda(A_n)$. From the probability integral transform and properties of uniform order statistics, we recall that $\mu(A_n)$ is beta$(k_n, n+1-k_n)$ distributed. Thus, the distribution of Y_n is conveniently distribution-free. If W denotes a beta$(k_n, n+1-k_n)$ random variable, then we have

$$1/Y_n = (n/(n+1))(W/E(W)),$$

where

$$E(W) = k_n/(n+1), \quad \mathrm{Var}(W) = k_n(n+1-k_n)/(n+1)^2(n+2).$$

Thus, $E(1/Y_n) = n/(n+1)$ and $\mathrm{Var}(1/Y_n) = (n/(n+1))^2(n+1-k_n)/(k_n(n+2)) \le 1/k_n$. Thus, $1/Y_n \to 1$ in probability if $\lim_{n\to\infty} k_n = \infty$.

To treat Z_n, we let S be the support set of f, and let B be the collection of Lebesgue points for f (i.e., the points at which $\mu(S_{xr})/\lambda(S_{xr}) \to f(x)$ as $r \downarrow 0$). By the Lebesgue density theorem, $\lambda(B^c) = 0$ (see, e.g., Wheeden and Zygmund, 1977). Assume first that $x \notin S$. Since S is closed, we can find $\varepsilon > 0$ such that $S_{x\varepsilon} \subseteq S^c$. Thus, $\lambda(A_n) \ge \lambda(S_{x\varepsilon}) > 0$, and thus

$$E(\mu(A_n)/\lambda(A_n)) \le k_n/((n+1)\lambda(S_{x\varepsilon})) \to 0.$$

If $x \in S$, then, by definition, for every $\varepsilon > 0$, $\mu(S_{x\varepsilon}) = p > 0$. Thus,

$$\begin{aligned} P(D_n(x) > \varepsilon) &= P(N < k_n) && \text{(where } N \text{ is Binomial}(n, p)) \\ &= P(N - E(N) < k_n - np) \\ &\le \frac{np(1-p)}{np(1-p) + (np - k_n)^2} && \text{(by Cantelli's inequality)} \\ &\le \frac{1-p}{1-p+np/4} && \text{(when } k_n \le np/2) \\ &= o(1). \end{aligned}$$

Thus, $D_n(x) \to 0$ in probability for $x \in S$. Therefore, $Z_n \to f(x)$ in probability for $x \in S \cap B$. We conclude that $Y_n Z_n \to f(x)$ in probability except perhaps on a set of zero Lebesgue measure.

REFERENCES

ABOU-JAOUDE, S. (1977). La convergence L_1 et L_∞ de certains estimateurs d'une densité de probabilité. Thèse de Doctorat d'État, Université Paris VI, Paris.

BREIMAN, L., MEISEL, W. and PURCELL, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics* **19** 135–144.

CACOULLOS, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* **18** 178–189.

DEVROYE, L. (1983). The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.* **11** 896–904.

DEVROYE, L. and PENROD, C. S. (1982). The strong uniform convergence of multivariate variable kernel estimates. Applied Research Laboratories, Technical Report 82-15, University of Texas at Austin.

DEVROYE, L. and PENROD, C. S. (1984). The consistency of automatic kernel density estimates. *Ann. Statist.* **12** 1231–1249.

DEVROYE, L. and WAGNER, T. J. (1976). Nonparametric discrimination and density estimation. Information Systems Research Laboratory, Technical Report 183, University of Texas at Austin.

DEVROYE, L. and WAGNER, T. J. (1979). The L_1 convergence of kernel density estimates. *Ann. Statist.* **7** 1136–1139.

FIX, E. and HODGES, J. L. (1951). Discriminatory analysis. Nonparametric discrimination: consistency properties. Report 4, Project number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas.

GLICK, N. (1974). Consistency conditions for probability estimators and integrals of density estimators. *Utilitas Math.* **6** 61–74.

HABBEMA, J. D. F., HERMANS, J. and REMME, J. (1978). Variable kernel density estimation in discriminant analysis. In *COMPSTAT.* (L. C. A. Corsten and J. Hermans, eds.) Physica-Verlag, Wien.

KRZYZAK, A. (1983). Classification procedures using multivariate variable kernel density estimate. *Pattern Recognition Letters* **1** 293–298.

LOFTSGAARDEN, D. O. and QUESENBERRY, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36** 1049–1051.

MACK, Y. P. and ROSENBLATT, M. (1979). Multivariate k-nearest neighbor density estimates. *J. Multivariate Anal.* **9** 1–15.

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.) Academic Press, New York.

MOORE, D. S. and YACKEL, J. W. (1977). Consistency properties of nearest neighbor density estimates. *Ann. Statist.* **5** 143–154.

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

RAATGEVER, J. W. and DUIN, R. P. W. (1978). On the variable kernel model for multivariate nonparametric density estimation. In *COMPSTAT.* (L. C. A. Corsten and J. Hermans, eds.) Physica-Verlag, Wien.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

RUDEMO, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.* **9** 1–15.

SCHEFFÉ, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.* **18** 434–458.

WAGNER, T. J. (1973). Strong consistency of a nonparametric estimate of a density function. *IEEE Trans. Systems Man Cybernet.* **3** 289–290.

WHEEDEN, R. L. and ZYGMUND, A. (1977). *Measure and Integral.* Marcel Dekker, New York.

SCHOOL OF COMPUTER SCIENCE
MCGILL UNIVERSITY
805 SHERBROOKE STREET WEST
MONTREAL, CANADA H3A 2K6