The Annals of Statistics
1977, Vol. 5, No. 1, 143–154

CONSISTENCY PROPERTIES OF NEAREST NEIGHBOR DENSITY FUNCTION ESTIMATORS

BY DAVID S. MOORE[1] AND JAMES W. YACKEL

Purdue University

Let $X_1, X_2, \cdots$ be R^p-valued random variables having unknown density function f. If K is a density on the unit sphere in R^p, $\{k(n)\}$ a sequence of positive integers such that $k(n) \to \infty$ and $k(n) = o(n)$, and $R(k, z)$ is the distance from a point z to the $k(n)$th nearest of $X_1, \cdots, X_n$, then $f_n(z) = (nR(k, z)^p)^{-1} \sum K((z - X_i)/R(k, z))$ is a nearest neighbor estimator of $f(z)$. When K is the uniform kernel, f_n is an estimator proposed by Loftsgaarden and Quesenberry. The estimator f_n is analogous to the well-known class of Parzen–Rosenblatt bandwidth estimators of $f(z)$. It is shown that, roughly stated, any consistency theorem true for the bandwidth estimator using kernel K and also true for the uniform kernel bandwidth estimator remains true for f_n. In this manner results on weak and strong consistency, pointwise and uniform, are obtained for nearest neighbor density function estimators.

1. Introduction. Let $X_1, X_2, \cdots$ be independent and identically distributed random variables taking values in R^p and having bounded density function f with respect to Lebesgue measure on R^p. A class of estimators of f which has been widely studied since the work of Rosenblatt (1956) and Parzen (1962) has the form

$$\hat{f}_n(z) = \frac{1}{nr(n)^p} \sum_{i=1}^n K\left(\frac{z - X_i}{r(n)}\right)$$

where

$$K(u) \quad \text{is a bounded density on} \quad R^p \tag{1.1}$$

and $\{r(n)\}$ is a sequence of positive numbers such that

$$r(n) \to 0 \quad \text{and} \quad nr(n)^p \to \infty \quad \text{as} \quad n \to \infty\,. \tag{1.2}$$

Call the estimators $\hat{f}_n$ *bandwidth estimators* of f. Among the results for $\hat{f}_n$ available in the literature under various conditions on K, $\{r(n)\}$ and f are: pointwise consistency and uniform consistency in probability (Cacoullos (1966)); pointwise consistency and uniform consistency with probability 1 (Nadaraya (1965) for $p = 1$ and Van Ryzin (1969) for $p \geqq 1$); asymptotic normality of $\hat{f}_n(z_i)$ for a fixed finite set of z_i (Cacoullos (1966)); weak convergence results for processes

Received August 1975; revised June 1976.

[1] Research of this author was sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. 72-2350B.

AMS 1970 *subject classifications.* 60F15, 62G05.

Key words and phrases. Nonparametric density estimation, multivariate density estimation, nearest neighbor estimators, bandwidth estimators, uniform consistency, strong consistency, weak consistency.

related to $\hat{f}_n(z)$ (Bickel and Rosenblatt (1973)); and a study of a sequential version of $\hat{f}_n(z)$ (Davies and Wegman (1975)). A survey of related literature is given by Wegman (1972).

When K is the uniform density on the unit sphere in R^p, the bandwidth estimator is simply empiric measure divided by Lebesgue measure for the sphere $\hat{S}_n(z)$ of radius $r(n)$ centered at z. Call this estimator $\hat{g}_n(z)$. A conceptually similar estimator of $f(z)$ was studied by Loftsgaarden and Quesenberry (1965). Their estimator $g_n(z)$ is empiric measure divided by Lebesgue measure for the sphere $S_k(z)$ centered at z and having radius $R(k, z)$ equal to the distance from z to the $k(n)$th nearest of $X_1, \cdots, X_n$. Here $\{k(n)\}$ is a sequence of positive integers such that

$$(1.3) \qquad k(n) \to \infty \quad \text{and} \quad k(n)/n \to 0 \quad \text{as} \quad n \to \infty .$$

Call g_n a *nearest neighbor estimator* of f. Loftsgaarden and Quesenberry showed that g_n is pointwise consistent in probability at continuity points of f. Wagner (1973) established pointwise consistency w.p. 1 under an additional assumption equivalent to $k(n)/\log n \to \infty$. In fact, by applying Theorem 6 of Kiefer (1972), one can establish pointwise consistency w.p. 1 of g_n under the weaker condition $k(n)/\log\log n \to \infty$, and show that this condition is the weakest possible.

Uniform consistency w.p. 1 of g_n in the case $p = 1$ when f is uniformly continuous and has known support was proved by Moore and Henrichon (1969) under the condition $k(n)/\log n \to \infty$. (They state only uniform consistency in probability, but their proof establishes convergence w.p. 1: apply the Borel–Cantelli lemma to the estimate $P_n \leqq (n + 1)a(\varepsilon)^{-k(n)}$ appearing on page 1501.) Since this proof uses a linear ordering of the X_i which does not generalize to the multivariate case $p > 1$, uniform consistency w.p. 1 of g_n for multivariate densities remains an open question. These nearest neighbor estimators have been widely used (e.g., [3], [6]) and many practitioners prefer them to bandwidth estimators.

Analogy with $\hat{f}_n$ suggests defining a general nearest neighbor density estimator by

$$f_n(z) = \frac{1}{nR(k, z)^p} \sum_{i=1}^n K\left(\frac{z - X_i}{R(k, z)}\right)$$

which allows unequal weights to be given to the observations. Thus g_n is the uniform kernel case of f_n, just as $\hat{g}_n$ is the uniform kernel case of $\hat{f}_n$. A class of estimators superficially similar to f_n was studied by Wagner (1975) for the case $p = 1$. Wagner's estimators replace $R(k, z)$ by a random radius Γ_n which is independent of z. They therefore lack the feature that only the observations nearest to z control estimation of $f(z)$.

This paper is devoted to a study of consistency properties of f_n, with results for g_n being obtained as a special case. The proofs require a restriction which reflects the nearest neighbor spirit, namely that along any ray emanating from z, nearer observations be given no less weight than more distant observations.

That is,

$$(1.4) \qquad K(cu) \geqq K(u) \quad \text{for any} \quad 0 \leqq c \leqq 1 \quad \text{and} \quad u \quad \text{in} \quad R^p .$$

Unlike the bandwidth estimators $\hat{f}_n$, the nearest neighbor estimators f_n may fail to be consistent unless some restriction is placed on the distance from z of X_i which are given positive weight. An example appears in Section 4. A natural restriction is that only the $k(n)$ observations nearest to z be used in estimating $f(z)$, i.e.,

$$(1.5) \qquad K(u) = 0 \quad \text{for all} \quad u \quad \text{in} \quad R^p \quad \text{with} \quad |u| > 1 .$$

Somewhat greater generality is achieved in Section 4.

Rather than attempting to prove specific consistency results for f_n, a consistency-equivalence relationship between f_n and a corresponding $\hat{f}_n$ is obtained. Heuristic considerations suggest that when $k(n) \sim \alpha nr(n)^p$ for some $\alpha > 0$, the nearest neighbor estimators $\{f_n\}$ for kernel K and $\{k(n)\}$ should have the same consistency properties as the bandwidth estimators $\{\hat{f}_n\}$ for the same kernel K and bandwidths $\{r(n)\}$. This observation is made rigorous for uniform kernels in Section 3 and for general kernels in Section 4. Specifically, given a kernel K, a number $\alpha > 0$, and an integer $k(n)$, denote by $\hat{f}(n, \alpha, z)$ the bandwidth estimator with kernel K and bandwidth $r(n, \alpha)$ defined by $k(n) = \alpha nr(n, \alpha)^p$. Denote by $\hat{g}(n, \alpha, z)$ the uniform kernel estimator with bandwidth $r(n, \alpha)$. The following theorem is a consequence of the results of Section 4.

THEOREM 1.1. *Let $\varepsilon > 0$, K satisfying* (1.1), (1.4) *and* (1.5), *and $\{k(n)\}$ be given. Then there exists $\eta > 0$ and a finite set of positive numbers $\alpha_1, \cdots, \alpha_M$ such that $|f_n(z) - f(z)| > \varepsilon$ implies that either $|\hat{f}(n, \alpha_i, z) - f(z)| > \eta$ or $|\hat{g}(n, \alpha_i, z) - g(z)| > \eta$ for at least one i in $\{1, \cdots, M\}$.*

Since the choice of $\eta, \alpha_1, \cdots, \alpha_M$ is uniform in n, z and sample points ω, it follows that *any consistency result* (*pointwise or uniform*, *in probability or with probability* 1) *valid for the bandwidth estimator $\hat{f}_n$ with kernel K and bandwidths $\{r(n)\}$, and also valid for the uniform kernel bandwidth estimator with bandwidths $\{r(n)\}$, remains valid for the nearest neighbor estimator f_n with kernel K and $k(n) \sim nr(n)^p$.* The only qualification to this statement is that the conditions on $\{r(n)\}$ must be satisfied also by $\{cr(n)\}$ for any $c > 0$.

Theorem 1.1 allows a large literature on consistency results for $\hat{f}_n$ to be restated for f_n. In particular, weak pointwise and uniform consistency follow from Cacoullos (1966) and strong pointwise and uniform consistency from Van Ryzin (1969). As an example, here is the nearest neighbor version of Van Ryzin's Theorem 2.

COROLLARY 1.1. *Let K satisfy* (1.1), (1.4), (1.5) *and have absolutely integrable characteristic function $c(t)$ satisfying $\int |c(\delta t) - c(t)|\, dt < M|\delta - 1|$ for δ sufficiently* *near* 1 *and some* $M > 0$. *Let* $\{k(n)\}$ *satisfy* (1.3) *and*

$$k(n)/k(n+1) \to 1$$
$$k(n)^2/n \to \infty$$
$$\sum_{n=1}^{\infty} k(n)^{-2} < \infty$$
$$\sum_{n=1}^{\infty} \frac{n^{1-1/p}}{k(n)^{2-1/p}} \left| \left(\frac{n+1}{k(n+1)}\right)^{1/p} - \left(\frac{n}{k(n)}\right)^{1/p} \right| < \infty .$$

Then

$$\sup_z |f_n(z) - f(z)| \to 0$$

with probability 1 *if* f *is uniformly continuous.*

Uniform kernel bandwidth estimators appear in Theorem 1.1 as essential auxiliaries, used in the proof to provide bounds on $R(k, z)$ when $f(z)$ is small. Since the literature on strong uniform consistency for bandwidth estimators contains little that applies to the uniform kernel case, Section 2 is devoted to remarks on this subject. In particular, the uniform kernel does not satisfy the main condition (absolutely integrable characteristic function) of Van Ryzin's Theorem 2. In order to obtain Corollary 1.1, Van Ryzin's result must be shown to hold for the uniform kernel. This is done in Theorem 2.1. The argument which is there applied to Van Ryzin's Theorem 2 can also be applied to Corollary 1.1 to establish strong uniform consistency of g_n when $\{k(n)\}$ satisfies the conditions of Corollary 1.1. A strong uniform consistency result for the Loftsgaarden–Quesenberry estimator therefore follows from Corollary 1.1.

The results of Section 4 are more general than those of Theorem 1.1 in several respects, the most important of which is that the roles of bandwidth and nearest neighbor estimators may be interchanged. For example, it now follows from Moore and Henrichon (1969) that when $p = 1$ and f has known support, the uniform kernel bandwidth estimator $\hat{g}_n$ is uniformly consistent w.p. 1 if $r(n) \to 0$ and $nr(n)/\log n \to \infty$. This is strictly weaker than existing conditions on $\{r(n)\}$ for strong uniform consistency of bandwidth estimators; it is satisfied by $r(n) \sim cn^{-\delta}$ for $0 < \delta < 1$, as opposed to $0 < \delta < \frac{1}{2}$ for the conditions of Nadaraya (1965) and Van Ryzin (1969). The principles of this paper will also allow future consistency results for either class of estimators to be restated for the other. They do not, however, allow distribution results for $\hat{f}_n$ to be carried over to f_n. Proofs of asymptotic normality for f_n and of some other results appear in [13].

The proofs of Sections 3 and 4 are combinatorial in nature and do not require that the Euclidean metric be used, that f be a density function with respect to Lebesgue measure, or even that the X_i take values in a Euclidean space. To emphasize the generality of the proofs and facilitate possible applications, Sections 3 and 4 use a setting much more general than is required to establish the results stated above.

2. Uniform consistency of bandwidth estimators. The uniform consistency results of Cacoullos (1966) and Van Ryzin (1969) require that the kernel K have absolutely integrable characteristic function, and hence do not apply to the uniform kernel. Nadaraya (1965) treats only the univariate case. This section generalizes the results of these authors to meet the needs of Theorem 1.1. First, all of the consistency theorems of Cacoullos and Van Ryzin apply to the uniform kernel K_0 on the unit sphere in R^p because K_0 can be approximated by kernels sufficiently smooth to possess absolutely integrable characteristic functions. Only the result required in the proof of Corollary 1.1 will be explicitly stated.

THEOREM 2.1. *Let K_0 be the uniform density on the unit sphere in R^p and suppose $\{r(n)\}$ satisfies the conditions of Theorem 2 of* [17]. *Then*

$$\hat{f}_n(z) = \frac{1}{nr(n)^p} \sum_{i=1}^n K_0\left(\frac{z - X_i}{r(n)}\right)$$

converges uniformly to $f(z)$ with probability 1 if $f(z)$ is uniformly continuous.

PROOF. From Lemma 4 on page 102 of Esseen (1945) it follows that for any $\delta > 0$ there exist densities K_1 and K_2 on R^p, having absolutely integrable characteristic functions and satisfying all other regularity conditions in [17], such that

$$C_1 K_1(u) \leqq K_0(u) \leqq C_2 K_2(u)$$

for all u in R^p and for constants $C_1 \geqq (1 - \delta)^p$ and $C_2 \leqq (1 + \delta)^p$. If $\hat{f}_n^{\,1}$ and $\hat{f}_n^{\,2}$ denote the bandwidth estimators based on $r(n)$ and (respectively) K_1, K_2, then for all n, z and sample points ω,

$$C_1 \hat{f}_n^{\,1}(z) \leqq \hat{f}_n(z) \leqq C_2 \hat{f}_n^{\,2}(z) .$$

Since Theorem 2 of [17] gives uniform strong consistency for $\hat{f}_n^{\,1}$ and $\hat{f}_n^{\,2}$ and $\delta > 0$ is arbitrary, $\hat{f}_n$ is also uniformly strongly consistent.

Nadaraya gives a strong consistency result for $p = 1$ under conditions on $\{r(n)\}$ which are similar to those of Van Ryzin, but he requires that the kernel K be of bounded variation. Theorem 2.2 below generalizes this result to $p > 1$ and restates the condition on $\{r(n)\}$ in a more usable form. Functions of bounded variation in R^p will be defined as follows (see Section 254 of Hobson (1927)). For any rectangle I in R^p, denote by Δ_I the usual rectangle difference operator. (That is, if a random variable X has df F on R^p, then $P[X \text{ in } I] = \Delta_I(F)$.) A function K is of *bounded variation* on R^p, if the supremum of $\Delta_{I(1)}(K) + \cdots + \Delta_{I(m)}(K)$ over all partitions of R^p into finitely many rectangles $I(1), \cdots, I(m)$ is finite. For $p > 1$ this definition is quite restrictive: the uniform density on the unit cube in R^p is of bounded variation, but the uniform density on the unit sphere is not. Thus our methods do not prove Theorem 2.2 for nearest neighbor estimators, since the corresponding result for uniform kernel bandwidth estimators is not covered by Theorem 2.2. Those methods do, however, prove the analog of Theorem 2.2 for nearest neighbor estimators with kernels supported on the unit cube in R^p (i.e., supported on the unit sphere in the "maximum component" metric).

THEOREM 2.2 *Let $K(u)$ satisfy* (1.1) *and be of bounded variation on R^p; let $\{r(n)\}$ satisfy $r(n) \to 0$ and $nr(n)^{2p}/\log n \to \infty$ as $n \to \infty$. Then*

$$\sup_z |\hat{f}_n(z) - f(z)| \to 0$$

with probability 1 *if f is uniformly continuous.*

PROOF. Nadaraya establishes this result for $p = 1$ under the condition that $\sum_{n=1}^{\infty} e^{-snr(n)^2} < \infty$ for all $s > 0$. His proof generalizes to $p > 1$ under the condition that

$$\sum_{n=1}^{\infty} e^{-snr(n)^{2p}} < \infty \qquad \text{for all} \quad s > 0\,, \tag{2.1}$$

provided that the bound on $n^{\frac{1}{2}} \sup |F_n - F|$ (F_n the empiric df and F the df of X_i) given in Theorem 1 of Kiefer and Wolfowitz (1958) is used to replace the bound due to Smirnov for $p = 1$ and used by Nadaraya. Integration by parts in R^p (see Section 351 of Hobson (1927)) is used, and the Kiefer–Wolfowitz bound must be applied in each dimension less than or equal to p. It follows from Section II.6 of Hardy and Riesz (1952) that (2.1) is equivalent to $\log n/nr(n)^{2p} \to 0$ as $n \to \infty$.

3. Equivalence results for uniform kernels. Let $X_1, X_2, \cdots$ be independent and identically distributed random variables taking values in a normed linear space $\mathscr{X}$ and having bounded density function f with respect to a positive measure μ on the Borel σ-field of $\mathscr{X}$. We assume that μ has the property that the measure of the sphere of radius r centered at z is a continuous function of r for any z. (Not necessarily the same for all z.) In R^p this assumption is satisfied by all μ absolutely continuous with respect to Lebesgue measure and by most nonatomic singular measures.

Suppose $\{h(n)\}$ is a sequence of real numbers such that $h(n) \to 0$ and $nh(n) \to \infty$. (In the Euclidean space setting of Section 1, $h(n) = r(n)^p$.) To avoid trivial complications in the proofs, assume that $k(n) = nh(n)$ is an integer for each n. If $||\cdot||$ is a norm in $\mathscr{X}$ and $\alpha > 0$, let

$$\hat{S}(n, \alpha, z) = \{x \text{ in } \mathscr{X} : ||z - x|| \leqq r(n, \alpha, z)\}$$

where $r(n, \alpha, z)$ is chosen so that $\mu\{\hat{S}(n, \alpha, z)\} = \alpha h(n)$. If $K(n, \alpha, z)$ is the (random) number of $X_1, \cdots, X_n$ contained in $\hat{S}(n, \alpha, z)$, a uniform kernel bandwidth estimator of $f(z)$ is (since $\mu\{\hat{S}(n, \alpha, z)\} = \alpha h(n)$)

$$\hat{g}(n, \alpha, z) = \frac{K(n, \alpha, z)}{n\alpha h(n)}\,. \tag{3.1}$$

We denote by $[c]$ the greatest integer in c. For $\beta > 0$, let $R(k, \beta, z)$ be the (random) distance from z to the $[\beta k(n)]$th nearest among $X_1, \cdots X_n$ and let

$$S(k, \beta, z) = \{x \text{ in } \mathscr{X} : ||z - x|| \leqq R(k, \beta, z)\}\,.$$

$S(k, \beta, z)$ contains exactly $[\beta k(n)]$ of $X_1, \cdots, X_n$, so a uniform kernel nearest neighbor estimator of $f(z)$ is

$$g(n, \beta, z) = \frac{[\beta k(n)]}{n\mu\{S(k, \beta, z)\}} . \tag{3.2}$$

When $\mathscr{X}$ is R^p, $\|\cdot\|$ is the Euclidean norm, $\alpha = \beta = 1$, and μ is λ/c, where λ is Lebesgue measure on R^p and c is the constant such that $\lambda\{x \text{ in } R^p : |x| \leqq r\} = cr^p$, then $\hat{g}$ and g are the special cases of the estimators $\hat{f}_n$ and f_n of Section 1 for K, the uniform density on the unit sphere.

We now state a theorem which, since it holds for each x, n and sample point ω, allows any (pointwise or uniform, weak or strong) consistency result for $\hat{g}$ to be transferred to g, and vice versa. For any $\delta > 0$, define

$$L(j, \delta) = \{x : (j-1)\delta \leqq f(x) < j\delta\} \tag{3.3}$$

for $j = 1, 2, \cdots, J(\delta)$ where $\delta J(\delta) \geqq \sup f(x)$.

THEOREM 3.1. *Let ε and δ be given, with $0 < \delta < \varepsilon$, and let $k(n) = nh(n)$.*

(a) *For any $\beta > 0$, $|g(n, \beta, x) - f(x)| > \varepsilon$ implies that $|\hat{g}(n, \alpha, x) - f(x)| \geqq \varepsilon - \delta$ for at least one of a finite set of values of α depending only on β, ε, δ and $J(\delta)$.*

(b) *If n is so large that $k(n)^{-1} \leqq (\varepsilon - \delta)/2$, then for any $\alpha > 0$, $|\hat{g}(n, \alpha, x) - f(x)| > \varepsilon$ implies that $|g(n, \beta, x) - f(x)| \geqq (\varepsilon - \delta)/2$ for at least one of a finite set of values of β depending only on α, ε, δ and $J(\delta)$.*

PROOF. (a) Without loss of generality, take $\beta = 1$. Suppose that $|g(n, 1, x) - f(x)| > \varepsilon$ for some n, x and sample point ω. Take x to be in $L(j, \delta)$ and since x, n and ω are fixed throughout the argument, write g_n for $g(n, 1, x)$ and S_k for $S(k, 1, x)$, so that $g_n = k(n)/n\mu\{S_k\}$.

Case 1. $g_n > f(x) + \varepsilon \geqq (j-1)\delta + \varepsilon$. Then

$$\mu\{S_k\} < \frac{k(n)}{n((j-1)\delta + \varepsilon)} = \frac{h(n)}{(j-1)\delta + \varepsilon} .$$

Choose $\alpha = ((j-1)\delta + \varepsilon)^{-1}$. Since $\mu\{\hat{S}(n, \alpha, x)\} = \alpha h(n) > \mu\{S_k\}$, it follows that $\hat{S}(n, \alpha, x) \supseteqq S_k$ and that $K(n, \alpha, x) \geqq k(n)$. Therefore

$$\hat{g}(n, \alpha, x) = \frac{K(n, \alpha, x)}{n\alpha h(n)} \geqq \alpha^{-1} = (j-1)\delta + \varepsilon .$$

Since x is in $L(j, \delta)$, $\hat{g}(n, \alpha, x) - f(x) > \varepsilon - \delta$.

Case 2. $g_n < f(x) - \varepsilon < j\delta - \varepsilon$. Then

$$\mu\{S_k\} > \frac{k(n)}{n(j\delta - \varepsilon)} = \frac{h(n)}{j\delta - \varepsilon} .$$

Choose $\alpha = (j\delta - \varepsilon)^{-1}$. Since $\mu\{\hat{S}(n, \alpha, x)\} = \alpha h(n) < \mu\{S_k\}$, we have $K(n, \alpha, x) \leqq k(n)$. Therefore

$$\hat{g}(n, \alpha, x) = \frac{K(n, \alpha, x)}{n\alpha h(n)} \leqq \alpha^{-1} = j\delta - \varepsilon .$$

Since x is in $L(j, \delta)$, $\hat{g}(n, \alpha, x) - f(x) \leqq -(\varepsilon - \delta)$. This completes the proof of (a). Note that no more than $2J(\delta) - 1$ values of α are needed to handle $\mathscr{X} = \bigcup L(j, \delta)$.

(b) Take $\alpha = 1$ and suppose that for some n, ω and some x in $L(j, \delta)$, $|\hat{g}(n, 1, x) - f(x)| > \varepsilon$. Write $\hat{S}_n$ for $\hat{S}(n, 1, x)$, $K(n)$ for $K(n, 1, x)$ and $\hat{g}_n = K(n)/n\mu\{\hat{S}_n\} = K(n)/nh(n)$ for $\hat{g}(n, 1, x)$.

Case 1. $\hat{g}_n > f(x) + \varepsilon \geqq (j-1)\delta + \varepsilon$. Then

$$K(n) > ((j-1)\delta + \varepsilon)nh(n) .$$

Choose $\beta = (j-1)\delta + \varepsilon$, so that $S(k, \beta, x) \subseteqq \hat{S}_n$ and

$$g(n, \beta, x) = \frac{[\beta k(n)]}{n\mu\{S(k, \beta, x)\}} > \frac{\beta k(n) - 1}{nh(n)} = \beta - k(n)^{-1}$$
$$\geqq j\delta + (\varepsilon - \delta)/2 .$$

Therefore $g(n, \beta, x) - f(x) > (\varepsilon - \delta)/2$.

Case 2. $\hat{g}_n < f(x) - \varepsilon \leqq j\delta - \varepsilon$. Then $K(n) < (j\delta - \varepsilon)nh(n)$, so that for $\beta = j\delta - \varepsilon$, $S(k, \beta, x) \supseteqq \hat{S}_n$ and $\mu\{S(k, \beta, x)\} \geqq h(n)$. So

$$g(n, \beta, x) = \frac{[\beta k(n)]}{n\mu\{S(k, \beta, x)\}} \leqq \frac{[\beta k(n)]}{nh(n)} \leqq \beta = j\delta - \varepsilon$$

and $g(n, \beta, x) - f(x) \leqq -(\varepsilon - \delta)$. Once again we have used at most $2J(\delta) - 1$ values of β.

4. Equivalence results for general kernels. Let $K(u)$ be a density on $\mathscr{X}$ and $\hat{S}(n, \alpha, x)$ be the sphere centered at z having specified measure $\alpha h(n)$ and radius $r(n, \alpha, z)$ as in Section 3. Then a general bandwidth estimator of $f(z)$ has the form

$$\hat{f}(n, \alpha, z) = \frac{1}{n\alpha h(n)} \sum_{i=1}^{n} K\left(\frac{z - X_i}{r(n, \alpha, z)}\right) . \tag{4.1}$$

If $S(k, \beta, z)$ is the sphere centered at z and having radius $R(k, \beta, z)$ equal to the distance from z to the $[\beta k(n)]$th nearest X_i, a general nearest neighbor estimator of $f(z)$ has the form

$$f(n, \beta, z) = \frac{1}{n\mu\{S(k, \beta, z)\}} \sum_{i=1}^{n} K\left(\frac{z - X_i}{R(k, \beta, z)}\right) . \tag{4.2}$$

Throughout this section we assume that $\hat{f}_n$ and f_n use the same kernel $K(u)$, which satisfies (1.4) for u in $\mathscr{X}$. We also continue to assume that $k(n) = nh(n)$. We continue to reserve $\hat{g}$ and g for the special cases defined by (3.1) and (3.2), in which the kernel is the uniform density on the unit spere $\{x : ||x|| \leqq 1\}$.

The estimators f_n need not share the consistency properties of $\hat{f}_n$ for general K. For example, suppose that the support of f is a compact set C in R^p of diameter d and that z lies a distance $\Delta > 0$ from C. Then $R = R(k, \beta, z) \geqq \Delta$ and $\mu\{S(k, \beta, z)\} \leqq \mu\{S\}$ for all n, β, where S is the sphere with center z and radius

$\Delta + d$. By (1.4),

$$K\left(\frac{z-u}{R}\right) \geqq K\left(\frac{z-u}{\Delta}\right) \geqq \eta = \inf_{x \text{ in } C} K\left(\frac{z-x}{\Delta}\right),$$

so if K is any density such that $\eta > 0$,

$$f(n, \beta, z) \geqq \frac{n\eta}{n\mu\{S\}} = \frac{\eta}{\mu\{S\}} > 0$$

for all n and β, while $f(z) = 0$. It is clear that lack of uniform convergence of $R(k, \beta, z)$ to zero will similarly prevent uniform consistency of f_n on the support of f, if f can be arbitrarily small on its support.

Theorem 4.1 below concerns equivalence of convergence properties for f_n and $\hat{f}_n$ on sets where f is bounded away from 0. It follows from Theorem 4.1 that for K such that $\hat{f}_n$ is consistent, f_n is consistent on the support of f; if K is such that $\hat{f}_n$ is uniformly consistent, f_n is uniformly consistent on $\{x : f(x) > \eta\}$ for any $\eta > 0$. Consistency only for f with known support is undesirable in a nonparametric density estimator. Corollary 4.1 shows that when K satisfies (1.5), f_n fully shares the consistency properties of $\hat{f}_n$. Finally, we describe a method of truncating the estimators f_n for more general kernels so as to preserve equivalence of convergence with $\hat{f}_n$.

An auxiliary lemma is required, of which part (a) is immediate from the definition of g and part (b) follows directly from the proof of Theorem 3.1 (b).

LEMMA 4.1. (a) *If* $|g(n, 1, x) - f(x)| \leqq \varepsilon$ *for* x *in* $L(j, \delta)$ *and* $\varepsilon > \delta > 0$, *then*

$$\frac{h(n)}{j\delta + \varepsilon} \leqq \mu\{S(k, 1, x)\} \leqq \frac{h(n)}{(j-1)\delta - \varepsilon}.$$

(b) *If* $|\hat{g}(n, 1, x) - f(x)| \leqq \varepsilon$ *for* x *in* $L(j, \delta)$ *and* $\varepsilon > \delta > 0$, *then*

$$\mu\{S(k, \beta'', x)\} \leqq \mu\{\hat{S}(n, 1, x)\} \leqq \mu\{S(k, \beta', x)\}$$

where $\beta' = (j-1)\delta + \varepsilon$ *and* $\beta'' = j\delta - \varepsilon$.

THEOREM 4.1. *Let* $\varepsilon > 0$ *be given and choose* $0 < \delta < \varepsilon/12$.

(a) *For any* $\beta > 0$ *and* x *in* $\{z : f(z) \geqq 2\delta\}$, $|f(n, \beta, x) - f(x)| > \varepsilon$ *implies that* $|\hat{f}(n, \alpha, x) - f(x)| \geqq \delta/2$ *or* $|\hat{g}(n, \alpha, x) - f(x)| \geqq \delta/2$ *for at least one of a finite set of values of* α *not depending on* n, x, *or the sample point* ω.

(b) *If* n *is so large that* $k(n) > \delta^{-1}$, *then for any* $\alpha > 0$ *and* x *in* $\{z : f(z) \geqq \delta\}$, $|f(n, \alpha, x) - f(x)| > \varepsilon$ *implies that* $|f(n, \alpha, x) - f(x)| \geqq \delta/2$ *or* $|g(n, \beta, x) - f(x)| \geqq \delta/4$ *for at least one of a finite set of values of* β *not depending on* n, x *or the sample point* ω.

PROOF. We give only the proof of (a). Without loss of generality take $\beta = 1$ and write f_n for $f(n, 1, x)$, R_k for $R(k, 1, x)$, and S_k for $S(k, 1, x)$. There are two cases: either $|g(n, 1, x) - f(x)| > 3\delta/2$, so that by Theorem 3.1 (a) $|\hat{g}(n, \alpha, x) - f(x)| > 3\delta/2 - \delta = \delta/2$ for at least one of a finite set of values of α; or $|g(n, 1, x) -$

$f(x)| \leqq 3\delta/2$, so that by Lemma 4.1 (a).

$$\frac{h(n)}{(j+\frac{3}{2})\delta} \leqq \mu\{S_k\} \leqq \frac{h(n)}{(j-\frac{5}{2})\delta} \tag{4.3}$$

for x in $L(j, \delta)$. We need only show that (4.3) implies that $|\hat{f}(n, \alpha, x) - f(x)| \geqq \delta/2$ for either $\alpha = ((j+\frac{3}{2})\delta)^{-1}$ or $\alpha = ((j-\frac{5}{2})\delta)^{-1}$ when x is in $L(j, \delta)$ and $j \geqq 3$.

Case 1. $f_n > f(x) + \varepsilon \geqq (j-1)\delta + \varepsilon$. Then by (4.2) and (4.3),

$$\sum K\left(\frac{x - X_i}{R}\right) > \frac{((j-1)\delta + \varepsilon)nh(n)}{(j+\frac{3}{2})\delta}. \tag{4.4}$$

Choosing $\alpha = ((j-\frac{5}{2})\delta)^{-1}$ gives by (4.3) that $\mu\{\hat{S}(n, \alpha, x)\} \geqq \mu\{S_k\}$, therefore that $r(n, \alpha, x) \geqq R_k$, and hence that

$$\begin{aligned} \hat{f}(n, \alpha, x) &= \frac{1}{n\alpha h(n)} \sum K\left(\frac{x - X_i}{r(n, \alpha, x)}\right) \\ &\geqq \frac{1}{n\alpha h(n)} \sum K\left(\frac{x - X_i}{R_k}\right) && \text{by (1.4)} \\ &> \frac{((j-1)\delta + \varepsilon)(j-\frac{5}{2})\delta}{(j+\frac{3}{2})\delta} && \text{by (4.4)} \\ &\geqq (j+\tfrac{1}{2})\delta \end{aligned}$$

for $j \geqq 3$ and $0 < \delta < \varepsilon/12$. Thus $\hat{f}(n, \alpha, x) - f(x) > \delta/2$ since x is in $L(j, \delta)$.

Case 2. $f_n < f(x) - \varepsilon < j\delta - \varepsilon$. This can only hold if $j\delta > \varepsilon$, so $j \geqq 12$ here. By (4.3),

$$\sum K\left(\frac{x - X_i}{R_k}\right) < \frac{(j\delta - \varepsilon)nh(n)}{(j-\frac{5}{2})\delta}. \tag{4.5}$$

Choosing $\alpha = ((j+\frac{3}{2})\delta)^{-1}$ gives by (4.3) that $r(n, \alpha, x) \leqq R_k$ and

$$\begin{aligned} \hat{f}(n, \alpha, x) &\leqq \frac{1}{n\alpha h(n)} \sum K\left(\frac{x - X_i}{R_k}\right) && \text{by (1.4)} \\ &< \frac{(j\delta - \varepsilon)(j+\frac{3}{2})\delta}{(j-\frac{5}{2})\delta} && \text{by (4.5)} \\ &< (j-\tfrac{3}{2})\delta \end{aligned}$$

for δ as above. So $\hat{f}(n, \alpha, x) - f(x) < -\delta/2$. This concludes the proof of part (a). Part (b) is proved by similar reasoning using Theorem 3.1 (b) and Lemma 4.1 (b). Note that the set of values of α used in the proof does not depend on the kernel K.

COROLLARY 4.1. *Suppose that* $K(u) = 0$ *for all* u *in* $\mathscr{X}$ *with* $\|u\| > 1$. *If we choose* $\delta < \varepsilon/4K(0)$, *then Theorem* 4.1 (a) *holds for all* x *in* $\mathscr{X}$. *If we choose* $\delta < (\frac{3}{2}K(0) + 1)^{-1}\varepsilon$, *then Theorem* 4.1 (b) *holds for all* x *in* $\mathscr{X}$.

PROOF. We give only the proof of the extension of Theorem 4.1 (a) to x in

$L(1, \delta) \cup L(2, \delta)$. The only case not covered by the proof of Theorem 4.1 (a) is that for which $|g(n, 1, x) - f(x)| \leqq 3\delta/2$ and $f_n > f(x) + \varepsilon$. We will show that for $\alpha = K(0)/((j-1)\delta + \varepsilon)$, $\hat{f}(n, \alpha, x) - f(x) > \varepsilon - 2\delta$ for x in $\{z : f(z) < 2\delta\}$. Thus adding this α to the finite set of Theorem 4.1 (a) is sufficient.

Since $||u|| > 1$ implies $K(u) = 0$, $K(u) \leqq K(0)$ for all u, and $||(x - X_i)/R_k|| \leqq 1$ for exactly $k(n)$ of the X_i, we have

$$k(n)K(0) \geqq \sum K\left(\frac{x - X_i}{R_k}\right) > n\mu\{S_k\}((j-1)\delta + \varepsilon)\,. \tag{4.6}$$

Therefore $\mu\{S_k\} < \alpha k(n)/n = \mu\{\hat{S}(n, \alpha, x)\}$, so that $r(n, \alpha, x) \geqq R_k$ and

$$\begin{aligned} \hat{f}(n, \alpha, x) &= \frac{1}{n\alpha h(n)} \sum K\left(\frac{x - X_i}{r(n, \alpha, x)}\right) \\ &\geqq \frac{1}{n\alpha h(n)} \sum K\left(\frac{x - X_i}{R_k}\right) && \text{by (1.4)} \\ &> \frac{\mu\{S_k\}((j-1)\delta + \varepsilon)^2}{h(n)K(0)} && \text{by (4.6)} \\ &\geqq \frac{((j-1)\delta + \varepsilon)^2}{(j + \frac{3}{2})\delta K(0)} && \text{by (4.3)} \\ &\geqq \varepsilon && \text{for } j = 1, 2 \end{aligned}$$

if $\delta \leqq \varepsilon/4K(0)$. Since $f(x) < 2\delta$, $\hat{f}(n, \alpha, x) - f(x) > \varepsilon - 2\delta$ as claimed.

In Corollary 4.1 the radius of S_k was controlled by restricting the support of K. An alternative is to truncate f_n when S_k is too large. As in the proof of Corollary 4.1 it is only necessary to handle the case x in $L(1, \delta) \cup L(2, \delta)$ and $|g(n, 1, x) - f(x)| \leqq 3\delta/2$, in which case by (4.3) $\mu\{S_k\} \geqq 2h(n)/7\delta$. Define

$$\begin{aligned} \bar{f}(n, \delta, x) &= f(n, 1, x) && \text{if } \mu\{S(k, 1, x)\} < 2h(n)/7\delta \\ &= 0 && \text{otherwise.} \end{aligned}$$

Then Theorem 4.1 (a) holds with f replaced by $\bar{f}$ at all x in $\mathscr{X}$. If $\hat{f}$ and $\hat{g}$ converge to f (pointwise or uniformly), one can use this result to show that $f^*(n, x) = \bar{f}(n, \delta_n, x)$ converges in the same sense when $\delta_n \to 0$ and $h(n)/\delta_n \to 0$. In practice, use of $\bar{f}(n, \delta, x)$ guarantees that convergence properties of $\hat{f}$ are shared where $f(x) \geqq 2\delta$ and that the error does not exceed 2δ where $f(x) < 2\delta$.

REFERENCES

[1] Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviation of density function estimators. *Ann. Statist.* **1** 1071–1095.

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

[3] Cover, T. M. and Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Trans. Information Theory* **13** 21–27.

[4] Davies, H. I. and Wegman, E. J. (1975). Sequential nonparametric density estimation. *IEEE Trans. Information Theory* **21** 619–628.

[5] ESSEEN, G. (1945). Fourier analysis of distribution functions. *Acta Math.* **77** 1–125.

[6] FU, K. S. and HENRICHON, E. G. (1968). On mode estimation in pattern recognition. *IEEE Proc. Seventh Symp. Adaptive Processes*, 3.a.1–3.a.10.

[7] HARDY, G. H. and RIESZ, M. (1952). *The General Theory of Dirichlet's Series.* Cambridge Univ. Press.

[8] HOBSON, E. W. (1927). *The Theory of Functions of a Real Variable.* Cambridge Univ. Press.

[9] KIEFER, J. C. (1972). Iterated logarithm analogues for sample quantiles when $p_n \downarrow 0$. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 227–244, Univ. of California Press.

[10] KIEFER, J. C. and WOLFOWITZ, J. (1958). On the deviations of the empiric distribution function of vector chance variables. *Trans. Amer. Math. Soc.* **87** 173–186.

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

[12] MOORE, D. S. and HENRICHON, E. G. (1969). Uniform consistency of some estimates of a density function. *Ann. Math. Statist.* **40** 1499–1502.

[13] MOORE, D. S. and YACKEL, J. W. (1976). Large sample properties of nearest neighbor density function estimators. To appear in *Statistical Decision Theory and Related Topics* (S. S. Gupta and D. S. Moore, eds.). Academic Press, New York.

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

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

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

[17] VAN RYZIN, J. (1969). On strong consistency of density estimates. *Ann. Math. Statist.* **40** 1765–1772.

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

[19] WAGNER, T. J. (1975). Nonparametric estimates of probability densities. *IEEE Trans. Information Theory* **21** 438–440.

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907