The Annals of Mathematical Statistics
1972, Vol. 43, No. 1, 170-180

ON THE BEST OBTAINABLE ASYMPTOTIC RATES OF CONVERGENCE IN ESTIMATION OF A DENSITY FUNCTION AT A POINT[1]

BY R. H. FARRELL

Cornell University

Estimation of the value $f(0)$ of a density function evaluated at 0 is studied, $f: \mathbb{R}_m \to \mathbb{R}$, $0 \in \mathbb{R}_m$. Sequences of estimators $\{\gamma_n, n \geqq 1\}$, one estimator for each sample size, are studied. We are interested in the problem, given a set C of density functions and a sequence of numbers $\{a_n, n \leqq 1\}$, how rapidly can a_n tend to zero and yet have

$$\liminf_{n\to\infty} \inf_{f \in C} P_f(|\gamma_n(X_1, \cdots, X_n) - f(0)| \leqq a_n) > 0\,?$$

In brief, by "rate of convergence" we will mean the rate which a_n tends to zero. For a continuum of different choices of the set C specified by various Lipschitz conditions on the kth partial derivatives of f, $k \geqq 0$, lower bounds for the possible rate of convergence are obtained. Combination of these lower bounds with known methods of estimation give best possible rates of convergence in a number of cases.

1. Introduction. This paper is to be considered a continuation of Farrell (1967). Throughout $\{X_n, n \geqq 1\}$ is a sequence of independently and identically distributed m-dimensional vector valued random variables such that X_1 has a density function f relative to m-dimensional Lebesgue measure. In Farrell, *op. cit.*, and in this paper, the problem considered is that of estimating the value of $f(0)$. In some recent literature, for example Leadbetter [3], the problem of estimating the entire density function is considered, i.e., given $m = 1$, simultaneously estimate the value of $f(x)$ for all $x \in \mathbb{R}_m$. However for the problem considered in this paper we restrict attention to the point estimation problem.

There are many different reasonable methods of constructing consistent sequences of estimators. The kernel method was introduced by Rosenblatt [6], and basic properties of the kernel method, when used for point estimation, were obtained by Parzen [4]. It was shown by Parzen that when $m = 1$, and f has two continuous derivatives, then many different kernels give rise to estimator sequences for which the square error converges to zero at a rate, a constant times $n^{-\frac{4}{5}}$. The constant is the product of a factor depending only on the kernel and a factor depending only on the particular f. It is thus possible to choose an optimal kernel, something apparently first done by Epsnecnikov [1]. The

Received November 19, 1970.

[1] The research supported in part by the Office of Naval Research under Contract Nonr 401 (50) of NSF Grant GP9297. Reproduction in whole or in part is permitted for any purpose of the United States Government.

same aymptotic rate seems invariably to come up when other methods are used. For example Weiss and Wolfowitz [12], who used a maximum likelihood method. The fact that different approches to the estimation problem (including M.I.S.E. = mean integrated square error) give the same asymptotic rate suggests that in fact this might be best possible, which we show to be the case in this paper.

Using mean integrated square error and working on the problem of estimating the entire function, Leadbetter [3] and Watson and Leadbetter [10] studied question, the if f is known, what kernel minimizes the asymptotic M.I.S.E. Pickands [5] has used Watson and Leadbetter, *op. cit.*, to study the asymptotic efficiency of a large class of kernels. Schuster [8], using Russian work as a start, has necessary and sufficient condition that an estimator sequence converge uniformly to f with probability one. Density estimators may be constructed using orthogonal series, e.g., see Schwartz [9] and Watson [11]. The series methods yield the same asymptotic rates as do the other methods.

Several recent articles have begun the process of generalizing the one-dimensional methods to estimation of density functions of several variables. The paper by Epsnecnikov [1] is of interest here in that his asymptotic rates of convergence are best possible, as will be shown in the sequel.

Most authors consider the problem of estimating *one* density. The methods they develop in fact give a uniform rate of convergence over a class of density functions. Rosenblatt [7] as well as Farrell [2] consider problems involving the uniform asymptotic rates of convergence over a class of density functions. Rosenblatt, *op. cit.*, has initiated the consideration of a class of density functions specified by Lipschitz condition rather than by bounds on the derivatives or asymptotic rates in the tails of the Fourier tranform of f. Rosenblatt's type of condition is considered further in this paper and bounds are obtained on the asymptotic rate of convergence which may be obtained.

The results of this paper are concerned with the question of how good, asymptotically, a sequence of estimators may be. Our results are stated for confidence intervals. Use of standard inequalities allow derivation of bounds for measures of loss other than zero-one loss functions. Our results depend on assumptions that a given estimator sequence is uniformly "good" over a specified class of density functions. The assumptions, made explicit below, hold for kernel type methods if the kernel has compact support (the optimal kernel obtained by Epsnecnikov [1] has compact support), and satisfies an orthogonality condition.

We now describe the classes of density functions to be considered. In order to do this some definitions are needed. We let $\eta : [0, \infty) \to [0, \infty)$ be a continuous function satisfying

(1.1 a) the derivative η' exists on $[0, \infty)$ and $\eta(0) = \eta'(0) = 0$.

(1.1 b) η' is a strictly increasing concave function .

(1.1 c) $\lim_{x\to 0} \eta'(bx)/\eta'(x) = \mu(b)$ exists, is finite, and $\lim_{b\to\infty} \mu(b) = \infty$.

In the sequel $||x|| = (x_1^2 + \cdots + x_m^2)^{\frac{1}{2}}$, where $x^t = (x_1, \cdots, x_m)$. A function f of m real variables will be said to be in the class $C_{k\eta}$ if

(1.2 a) all the kth order partial derivatives of f exist and are continuous;

(1.2 b) there exists a polynomial p of degree k in the variables $x_1, \cdots, x_m$ such that for all x, $|f(x) - p(x)| \leqq 2(k!)^{-1}||x||^k\eta'(||x||)$.

A limiting case of condition (1.2) is the function $\eta(x) = x$, $x > 0$. We shall say that $f \in C_{k\alpha}$ if

(1.3 a) f and its first $k - 1$ derivatives are continuous;

(1.3 b) the $(k - 1)$st partial derivatives of f are absolutely continuous relative to each of the m possible kth order partial derivative of f;

(1.3 c) the kth order partial derivatives exist almost everywhere and satisfy

$$\left|\frac{\partial^k f}{\partial x_1^{k_1} \cdots \partial x_m^{k_m}}\right| \leqq \alpha , \qquad k_1 + \cdots + k_m = k .$$

In the sequel, Case I will refer to the estimation problem restricted to a set $C_{k\eta}$, $k \geqq 0$. Case II will refer to the estimation problem restricted to a set $C_{k\alpha}$, $k \geqq 1$. The proofs given in Section 3 for Theorems 1.1 and 1.2 are given only for Case I since similar proofs give the corresponding results for Case II.

When $m = 1$, in terms of Farrell [2], the class of density functions considered in that paper is a subset of $C_{0\alpha}$. In this paper, Case II, we are concerned with $m \geqq 1$ and $k \geqq 1$. When $m = 1$ it is known, see for example Parzen [9], that given $k \geqq 1$, one may write a sequence of estimators $\{\gamma_n, n \geqq 1\}$ such that if $n \geqq 1$ then γ_n is a measurable function of n variables and

(1.4) $$1 = \lim_{a\to\infty} \lim_{n\to\infty} \inf_{f \in C_{k\alpha}} P_f(|\gamma_n(X_1, \cdots, X_n) - f(0)| \leqq an^{-k/(2k+1)}) .$$

This uniformity may be established for a large class of kernel functions including those of compact support that are orthogonal to $x, x^2, \cdots, x^{k-1}$. This may be established by use of Lemma 1.4 given below. Using results of Epsnecnikov [1] a similar result probably can be established when $m \geqq 2$ for which the appropriate asymptotic rate of convergence is $n^{-k/(2k+m)}$. We have not checked this guess.

The theorems below establish, subject to a uniformity condition, that for density functions of m variables in $C_{k\alpha}$, the rate $n^{-k/(2k+m)}$ is the best possible rate. To describe our results for Case I, some additional notation is needed. We define a function λ by

(1.5) $$\beta = (\lambda(\beta))^{2k+m-2}(\eta(\lambda(\beta)/2))^2 .$$

It is easy to chek that $x^p(\eta(x))^2$ is a strictly increasing function that takes all values from 0 to ∞. Thus $\lambda(\beta)$ is uniquely defined. We show that relative to $C_{k\eta}$ the best possible rate is $(n(\lambda(n^{-1}))^m)^{-\frac{1}{2}}$. In case $\eta(\beta) = \beta^\tau$ a calculation shows this rate to be $n^{-(k-1+\tau)/(2k-2+m+2\tau)}$. Case II is a special subcase in which $\tau = 1$.

In the sequel we suppose always that if $n \geqq 1$, $\gamma_n(X_1, \cdots, X_n)$ is the estimate based on $X_1, \cdots, X_n$.

THEOREM 1.1. *Suppose $\{a_n, n \geqq 1\}$ is a real number sequence such that*

$$\liminf_{n\to\infty} \inf_{f \in C_{k\eta}} P_f(|\gamma_n(X_1, \cdots, X_n) - f(0)| \leqq a_n) = 1 . \tag{1.6}$$

Then,

$$\liminf_{n\to\infty} n(\lambda(n^{-1}))^m a_n^2 = \infty . \tag{1.7}$$

In Case II, $k \geqq 1$, (1.7) still holds.

THEOREM 1.2. *Suppose $\{a_n, n \geqq 1\}$ is a real number sequence such that*

$$\limsup_{n\to\infty} \sup_{f \in C_{k\eta}} a_n^{-2} E_f(\gamma_n(X_1, \cdots, X_n) - f(0))^2 < \infty . \tag{1.8}$$

Then,

$$\liminf_{n\to\infty} n(\lambda(n^{-1}))^m a_n^2 > 0 . \tag{1.9}$$

In Case II, $k \geqq 1$, (1.9) still holds.

THEOREM 1.3. *Suppose in Case I the hypotheses of Theorem 1.2 hold and that $\eta(t) = t^\tau$ for some $\tau \geqq 1$. Suppose*

$$\limsup_{n\to\infty} a_n n^{(k+\tau-1)/(2k+m+2\tau-2)} < \infty . \tag{1.10}$$

Let $k^* \leqq k$ and $\eta^*(t) = t^{\tau^*}$ where $\tau^* \leqq \tau$. Assume $k + \tau > k^* + \tau^*$. There exists a sequence of functions $\{f_n, n \geqq 1\}$ in $C_{k^*\eta^*}$ such that

$$\lim_{n\to\infty} P_{f_n}(|\gamma_n(X_1, \cdots, X_n) - f_n(0)| \leqq a_n) = 0 . \tag{1-11}$$

Theorem 1.3 says in effect that a sequence of estimators good relative to the class $C_{k\eta}$ will not give uniformly good results over the larger class $C_{k^*\eta^*}$. Thus the statistician who constructs his sequence of estimators relative to the wrong class of functions is likely to obtain bad estimates.

Theorem 1.1 to 1.3 assume that the estimators used satisfy a uniformity condition. The kernel method of Parzen, if the kernel satisfies conditions listed above, automatically has the requisite uniformity expressed in (1.4). This is easily shown using the following Lemma 1.4.

LEMMA 1.4. *Let $\{x_n, n \geqq 1\}$* and *$\{y_n, n \geqq 1\}$ be real number sequences in the interval $[a, b]$ such that $\lim_{n\to\infty} (x_n - y_n) = 0$. If $\{f_n, n \geqq 1\}$ is a sequence of functions in $C_{k\alpha}$ then $\lim_{n\to\infty} (f_n(x_n) - f_n(y_n)) = 0$.*

PROOF. Let $m \geqq 1$ be the least integer such that for some constant α_m, $\max_{n\geqq 1} \sup_{x \in [a,b]} |f_n^{(m)}(x)| \leqq \alpha_m$, where $f_n^{(m)}$ is the mth derivative of f_n. By definition of $C_{k\alpha}$, it follows that $m \leqq k$. We show an assumption that $m \geqq 2$ is contradictory. If $m \geqq 2$ then there exists a subsequence $\{f_{n1}, n \geqq 1\}$ of the functions and sequence $\{w_n, n \geqq 1\}$ in $[a, b]$ such that

$$\lim_{n\to\infty} |f_{n1}^{(m-1)}(w_n)| = \infty . \tag{1.12}$$

Since $|f_{n1}^{(m-1)}(x) - f_{n1}^{(m-1)}(y)| \leqq \alpha_m|x - y|$, $n \geqq 1$, the relation (1.12) implies $\lim_{n\to\infty} |f_{n1}^{(m-1)}(x)| = \infty$ uniformly in $x \in [a, b]$. Choose $\varepsilon_1 = \pm 1$ so that $\lim_{n\to\infty} \varepsilon_1 f_{n1}^{(m-1)}(x) = \infty$ uniformly in $x \in [a, b]$. Then

$$\lim_{n\to\infty} \varepsilon_1(f_{n1}^{(m-2)}(y) - f_{n1}^{(m-2)}(x)) = \lim_{n\to\infty} \int_x^y \varepsilon_1 f_{n1}^{(m-1)}(t)\, dt = \infty . \tag{1.13}$$

This implies that in one of the intervals $[a, (2a + b)/3]$ and $[(a + 2b)/3, b]$, for some number $\varepsilon_2 = \pm 1$, that $\lim_{n\to\infty} \varepsilon_m f_{n1}^{(m-2)}(x) = \infty$ uniformly in x. Repeating this argument m times we see there is a nondegenerate subinterval $[a', b'] \subset [a, b]$ and $\varepsilon_m = \pm 1$ such that $\lim_{n\to\infty} f_{n1}^{)0)}(x) = \lim_{n\to\infty} \varepsilon_m f_{n1}(x) = \infty$ uniformly in $x \in \varepsilon[a', b']$. However, $f_{n1} \geqq 0$ and $\int_{-\infty}^{\infty} f_{n1}(t)\, dt = 1$, $n \geqq 1$. Therefore $\varepsilon_m = 1$ and we obtain the contradiction that

$$1 = \lim_{n\to\infty} \int_{-\infty}^{\infty} f_{n1}(t)\, dt \geqq \lim_{n\to\infty} \int_{a'}^{b'} f_{n1}(t)\, dt = \infty .$$

Consequently $m = 1$.

We complete the proof of the lemma by denying the conclusion of the lemma and obtaining a contradiction. Using the hypotheses of the lemma, by taking a subsequence if necessary, we may suppose $|f_n(x_n) - f_n(y_n)| \geqq \varepsilon$. Using the mean value theorem, there exists a real number sequence $\{x_{n1}, n \geqq 1\}$ such that if $n \geqq 1$ then x_{n1} is between x_n and y_n and such that

$$|x_n - y_n|\, |f_n'(x_{n1})| \geqq \varepsilon . \tag{1.14}$$

Since $\lim_{n\to\infty}(x_n - y_n) = 0$, the inequality (1.14) implies that $\lim_{n\to\infty} |f_n'(x_{n1})| = \infty$. This clearly contradicts the conclusion of the first paragraph that $m = 1$. □

The method of this paper is to investigate modifications $f + e$ of a density function f such that $e(0)$ is "large" but $f + e$ changes the probabilities of events "very little." In Section 2 we construct a sequence of functions $\{e_{k\delta}, k \geqq 1, 0 < \delta < \infty\}$. In Section 3 by considering density functions of the form $f + e_{k\delta}$ we obtain proofs of the three theorems.

2 a. Constructions for Case II. In this section we construct recursively a sequence $\{g_{k\delta}, k \geqq 0\}$ of functions satisfying

(2.1 a) if $x \notin [-2^k\delta, 2^k\delta]$ then $g_{k\delta}(x) = 0$.

(2.1 b) $g_{k\delta}$ is an odd function of the variable.

(2.1 c) the $(k - 1)$st derivative of $g_{k\delta}$ is absolutely continuous.

(2.1 d) $\sup_{x\in\mathbb{R}} |g_{k\delta}(x)| = 2^{\gamma}\delta^k$, where $\gamma = (k - 1)(k - 2)/2$.

(2.1 e) There exists a real number sequence $\{c_k, k \geqq 1\}$ such that

$$\int_{-\infty}^{\infty} g_{k\delta}^2(x)\, dx = c_k \delta^{(2k+1)} .$$

Here k is an integer index, and δ a real number $\geqq 0$. The main technique of computation in the sequel is the Fourier transform.

Define

(2.2) $$g_{0\delta}(x) = 1 \quad \text{if} \quad -\delta \leqq x < 0\,, \qquad \text{and}$$
$$g_{0\delta} = -1 \quad \text{if} \quad 0 \leqq x < \delta\,.$$

Proceeding recursively, $g_{(k-1)\delta}$ has been defined then

(2.3) $$g^*_{k\delta}(x) = \int_{-2^{(k-1)}\delta}^{x} g_{(k-1)\delta}(t)\,dt\,, \quad \text{if} \quad -2^{(k-1)}\delta \leqq x \leqq 2^{(k-1)}\delta\,, \quad \text{and} \quad g^*_{k\delta}(x) = 0 \quad \text{otherwise}\,.$$

Define

(2.4) $$g_{k\delta}(x) = g^*_{(k-1)\delta}(x + 2^{(k-1)}\delta) - g^*_{(k-1)}(x - 2^{(k-1)}\delta)\,.$$

The following four lemmas require only simple calculations and their proofs are omitted.

LEMMA 2.1. *If* $k \geqq 1$ and $x \in \mathbb{R}$ *then*

(2.4 a) $$g^*_{k\delta}(x) \geqq 0\,, \qquad \text{and}$$

(2.4 b) $$g^*_{k\delta}(2^k\delta) = 0\,.$$

LEMMA 2.2. *If* $k \geqq 1$ *then* $g_{k\delta}$ *has* $(k-1)$ *continuous derivatives and the* $(k-1)$st *derivative is absolutely continuous. The function* $g_{k\delta}$ *is an odd function.*

LEMMA 2.3. *The Fourier transform of* $g_{0\delta}$ *is*

(2.5) $$\hat{g}_{0\delta}(t) = 2(it)^{-1}(1 - \cos t\delta)\,.$$

LEMMA 2.4. *The Fourier transform of* $g_{k\delta}$, $k \geqq 1$ *is*

(2.6) $$\hat{g}_{k\delta}(t) = 2^{k+1}i^{-1}t^{-(k+1)}(\sin t\delta)\cdots(\sin 2^{k-1}t\delta)(1 - \cos t\delta)\,.$$

LEMMA 2.5. $g_{k\delta}$ *has its maximum at* $t = -2^{k-1}\delta$ *and* $g_{k\delta}(-2^{k-1}\delta) = 2^{\gamma}\delta^k$ *where* $\gamma = (k-1)(k-2)/2$.

PROOF. If $t < 0$ then $g_{k\delta}(t) \geqq 0$, and if $t \geqq 0$ then $g_{k\delta}(t) \leqq 0$. Thus the maximum of $g_{k\delta}$ is the maximum of

(2.7) $$g^*_{k\delta}(t + 2^{k-1}\delta) = \int_{-2^{k-1}}^{2^{k-1}\delta+t} g_{(k-1)\delta}(x)\,dx\,.$$

Since $g_{(k-1)\delta}$ is an odd function which is positive for negative arguments and negative for positive arguments, the value of t maximizing (2.7) is $t = -2^{k-1}\delta$.

In order to compute the value of the maximum let $\lambda = 2^{k-2}\delta$. Make two changes of variable and use the fact that the functions $g_{k\delta}$ have compact support to obtain

(2.8) $$g_{k\delta}(-2\lambda) = \int_{-\lambda}^{\lambda} dy \int_{-\lambda}^{z} g_{(k-2)\delta}(z)\,dz\,.$$

We may evaluate the integral (2.8) by computing the Fourier transform of

(2.9) $$\int_{-\lambda}^{z} g_{(k-2)\delta}(z)\,dz$$

and evaluating the Fourier transform at zero. This process yields

(2.10) $$g_{k\delta}(-2\lambda) = \lim_{t\to 0} -(it)^{-1}\hat{g}_{(k-2)\delta}(t) = 2^{\gamma}\delta^{k} ,$$

where

$$\gamma = (k-1)(k-2)/2 .$$

In this calculation we have used formula (2.6). □

We complete this section by stating and proving

LEMMA 2.6. *There exists a sequence of positive real numbers* $\{c_k, k \geqq 1\}$ *such that*

(2.11) $$\int_{-\infty}^{\infty} (g_{k\delta}(x))^2\, dx = c_k \delta^{(2k+1)} .$$

PROOF. We use Parseval's identity for the Fourier transform, see Wiener [13]. Thus, using (2.6) and making the change of variable $s = t\delta$ in the integral $\int |\hat{g}_{k\delta}(t)|^2\, dt$ yields (2.11).

2 b. Constructions for Case I. In this section we construct a sequence of functions $\{g_{k\delta}, k \geqq 1\}$, $\delta \geqq 0$, used to modify density functions in Case I. We suppose given $\eta : [0, \infty) \to [0, \infty)$ such that

(2.12) $\eta(0) = 0$, the derivative η' exists and is continuous, $\eta'(0) = 0$, η' is a strictly increasing concave function.

A standard result about concave functions then states that

(2.13) if $x, y \in [0, \infty)$ then $|\eta'(x) - \eta'(y)| \leqq \eta'(|x-y|)$.

In the sequel we will need to assume

(2.14) there exists a real valued function μ such that if $b \in (0, \infty)$ then $\lim_{x\to\infty} \eta(bx)/\eta(x) = \mu(b) > 0$, and $\lim_{b\to\infty} \mu(x) = \infty$.

The sequence of functions to be constructed satisfy

(2.15 a) if $x \notin [-2^k\delta, 2^k\delta]$ then $g_{k\delta}(x) = 0$.

(2.15 b) $g_{k\delta}$ is an odd function.

(2.15 c) the kth derivative of $g_{k\delta}$ is continuous and satisfies $|g_{k\delta}^{(k)}(x + p\delta)| \leqq \eta'(|x|)$ for all $x \in R$ and integers p.

(2.15 d) $$\sup_{x\in R} |g_{k\delta}(x)| = 2^{\gamma+1}\delta^{k-1}\eta(\delta/2) , \qquad \gamma = (k-1)(k-2)/2 .$$

(2.15 e) there exists a real number sequence $\{c_k', k \geqq 1\}$ such that if $k \geqq 0$ then

$$\int_{-\infty}^{\infty} (g_{\delta k}(x))^2\, dx \leqq c_k' \delta^{2k-1}(\eta(\delta/2))^2 .$$

The proofs required parallel those of Section 2a. Consequently we omit most detail. We note that we now define

(2.16 a) if $x > 0$ then $g_{0\delta}(x) = -g_{0\delta}(-x)$;

(2.16 b) if $0 \leqq x \leqq \delta/2$ then $g_{0\delta}(x) = -\eta'(x)$.

(2.16 c) if $\delta/2 \leqq x \leqq \delta$ then $g_{0\delta}(x) = -\eta'(\delta - x)$;

(2.16 d) if $\delta < x$ then $g_{0\delta}(x) = 0$.

The functions $g^*_{k\delta}$ and $g_{k\delta}$, $k \geqq 1$ are then defined recursively by the relations (2.3) and (2.4). Then (2.15 a) and (2.15 b) follow immediately. (2.15 e) is an immediate consequence of (2.15 a) and (2.15 d). It is easily verified that if $x \in R$ then the kth derivative of $g_{k\delta}$ exists, is continuous, and satisfies (2.15 c).

We now verify (2.15 d). From (2.10) and (2.6) we require the value

$$(2.17) \qquad \lim_{t\to\infty} (it)^{-1}[2^{k-2}t^{-(k-2)}(\sin \delta t) \cdots (\sin 2^{k-3}\delta t)\hat{g}_{0\delta}(t)] = 2^{\gamma}\delta^{k-2} \lim_{t\to 0} (it)^{-1}\hat{g}_{0\delta}(t) = 2^{\gamma+1}\delta^{k-1}\eta(\delta/2) .$$

2c. Functions of several variables. In this section we suppose x and x_0 are in $\mathbb{R}_m$ and that $x^t = (x_1, \cdots, x_m)$ and $x_0^t = (x_{10}, \cdots, x_{m0})$. We define

$$(2.18) \qquad e_{k\delta}(x) = m^{-1} \textstyle\sum_{i=1}^m g_{k\delta}(x_i - 2^{k-1}\delta) .$$

Then there exists a constant c_1 such that

$$(2.19) \qquad \text{if } ||x|| > c_1\delta \text{ then } e_{k\delta}(x) = 0 .$$

As defined the functions $e_{k\delta}$ have their maximum at 0. From (2.15 d),

$$(2.20) \qquad e_{k\delta}(0) = 2^{\gamma+1}\delta^{k-1}\eta(\delta/2) .$$

A combination of (2.19) and (2.20) imply there exists a constant c_2 such that

$$(2.21) \qquad \int (e_{k\delta}(x))^2\, dx \leqq c_2\delta^{2k-2+m}(\eta(\delta/2))^2 .$$

We now show that $e_{k\delta}$ is in $C_{k(\eta/2)}$. Using Taylor's formula with remainder one obtains at once for an expansion about $(x_{10} - 2^{k-1}\delta, \cdots, x_{m0} - 2^{k-1}\delta)$ that

$$(2.22) \qquad | e_{k\delta}(x) - m!^{-1} \textstyle\sum_{i=1}^m \sum_{j=1}^k (j!)^{-1}g_{k\delta}^{(j)}(x_{i0} - 2^{k-1}\delta)(x_i - x_{i0})^j| \leqq (mk!)^{-1} \sum_{i=1}^m |x_i - x_{i0}|^k |g_{k\delta}^{(k)}(\theta(x_i, x_{i0})) - g_{k\delta}^{(k)}(x_{i0} - 2^{k-1}\delta)| ,$$

where $\theta(x_i, x_{i0})$ is a number between $x_i - 2^{k-1}\delta$ and $x_{i0} - 2^{k-1}\delta$. We use the fact that $|g_{k\delta}^{(k)}(x + p\delta)| \leqq \eta'(|x|)$ for all real x and integers p, as stated in (2.15 c). Setting $x_0 = 0$ we obtain from (2.22) that

$$(2.23) \qquad \begin{aligned} |e_{k\delta}(x) - m^{-1} \textstyle\sum_{i=1}^m \sum_{j=1}^k (j!)^{-1}g_{k\delta}^{(j)}(-2^{k-1}\delta)x_i{}^j| &\leqq (k!\, m)^{-1} \textstyle\sum_{i=1}^m |x_i|^k |g_{k\delta}^{(k)}(\theta(x_i, 0))| \\ &\leqq (k!)^{-1}\eta'(||x||)m^{-1} \textstyle\sum_{i=1}^m |x_i|^k \\ &\leqq (k!)^{-1}||x||^k\eta'(||x||) . \end{aligned}$$

The mixed partial derivatives of $e_{k\delta}$ are all zero. We find $(\partial^k/\partial x_i^k)e_{k\delta}(x) = g_{k\delta}^{(k)}(x_i)$. Therefore the kth order partial derivatives of $e_{k\delta}$ are all continuous. Thus $e_{k\delta} \in C_{k(\eta/2)}$.

Finally, since $e_{k\delta}$ has compact support, it is integrable and

$$\int e_{k\delta}(x)\,dx = 0\ . \tag{2.24}$$

3. Proofs of the theorems.

PROOF OF THEOREM 1.1. Case I may be treated as a special subcase of Case II with $\eta(x) = x$ for all $x \in \mathbb{R}$. Thus we do not treat the two cases separately.

The set $C_{k(\eta/2)}$ contains a function nonzero and constant near zero. Thus let $\varepsilon > 0$ and $a > 0$ and $f \in C_{k(\eta/2)}$ such that if $||x|| \leqq \varepsilon_0$ then $f(x) = a$. There exists $\delta_0 > 0$ such that if $0 \leqq \delta \leqq \delta_0$ then h_δ defined by

$$h_\delta(x) = f(x) + e_{k\delta}(x)\ , \qquad e_{k\delta} \in C_{k(\eta/2)}\ , \tag{3.1}$$

is a density function that is in $C_{k\eta}$. The constant δ_0 is to be chosen that $c_1\delta_0 < \varepsilon$, where c_1 is given in (2.19). Then the support of $e_{k\delta}$ is in the sphere $||x|| < \varepsilon$. Thus

$$\text{if } ||x|| \leqq \varepsilon \text{ then } h_\delta(x) = a + e_{k\delta}(x), \text{ and if } ||x|| > \varepsilon \text{ then } h_\delta(x) = f(x). \tag{3.2}$$

Let x_n be the indicator function of the event $|\gamma_n(X_1, \cdots, X_n) - h_\delta(0)| \leqq a_n$. Then

$$\begin{aligned} &P_{h_\delta}(|\gamma_n(X_1, \cdots, X_n) - h_\delta(0)| \leqq a_n) \\ &\quad = \int \cdots \int x_n(x_1, \cdots, x_n) \textstyle\prod_{i=1}^n (h_\delta(x_i)/f(x_i)) \prod_{i=1}^n f(x_i) \prod_{i=1}^n dx_i \\ &\quad \leqq (P_f(|\gamma_n(X_1, \cdots, X_n) - h_\delta(0)| \leqq a_n))^{\frac{1}{2}} \\ &\qquad \times (\int \cdots \int \textstyle\prod_{i=1}^n (h_\delta(x_i))^2(\prod_{i=1}^n f(x_i))^{-1} \prod_{i=1}^n dx_i)^{\frac{1}{2}}\ . \end{aligned} \tag{3.3}$$

Using (3.2) and (2.24) we obtain

$$\int (h_\delta(x)/f(x))^2 f(x)\,dx = 1 + a^{-1} \int (e_{k\delta}(x))^2\,dx\ . \tag{3.4}$$

Using (2.19) and (2.21) and defining $c_3 = a^{-1}c_2$ we obtain

$$a^{-1} \int (e_{k\delta}(x))^2\,dx \leqq c_3\delta^{2k+m-2}(\eta(\delta/2))^2\ . \tag{3.5}$$

Defining a function λ by (1.5) we make the choice

$$\delta = \lambda(\beta) \text{ where } \beta = c_3^{-1}b'/n, \text{ and } b' \text{ is an arbitrary constant.} \tag{3.6}$$

Abbreviating notation in an obvious way we obtain,

$$P_{h_{\lambda(\beta)}}(|\gamma_n - h_{\lambda(\beta)}(0)| \leqq a_n) \leqq (P_f(|\gamma_n - h_{\lambda(\beta)}(0)| \leqq a_n))^{\frac{1}{2}}(1 + b'/n)^{n/2}\ . \tag{3.7}$$

The value

$$h_{\lambda(\beta)}(0) = a + 2^{\gamma+1}(\lambda(\beta))^{k-1}\eta(\lambda(\beta)/2)\ , \tag{3.8}$$

where $\gamma = (k-1)(k-2)/2$. Our hypothesis is that

$$\lim_{n\to\infty} P_f(|\gamma_n - f(0)| \leqq a_n) = 1\ , \tag{3.9}$$

uniformly in $f \in C_{k\mu}$. Thus the left side of (3.7) tends to one so that

(3.10) $$\liminf_{n\to\infty} P_f(|\gamma_n - h_{\lambda(\beta)}(0)| \leqq a_n) > 0 .$$

Comparison of (3.9) and (3.10) show that there must exist an n_0 such that if $n \geqq n_0$ then the events $|\gamma_n - f(0)| \leqq a_n$ and $|\gamma_n - h_{\lambda(\beta)}(0)| \leqq a_n$ have non-void intersection. Thus

(3.11) $$2a_n \geqq 2^{\gamma+1}(\lambda(\beta))^{k-1}\eta(\lambda(\beta)/2) .$$

Let $b = c_3^{-1}b'$, so that $\beta = b/n$. Square both sides of (3.11) and multiply by λ^m to obtain

(3.12) $$n(\lambda(b/n))^m(a_n 2^{-\gamma})^2 \geqq b .$$

Use (1.5) with $\beta = b/n$ and $\beta = 1/n$, divide the two identities, to obtain

(3.13) $$b = (\lambda(bn^{-1})/\lambda(n^{-1}))^{2k+m-2}(\eta(\lambda(bn^{-1})/2)/\eta(\lambda(n^{-1})/2))^2 .$$

Substitution of (3.13) into (3.12) and cancellation yields

(3.14) $$n(\lambda(n^{-1}))^m(a_n 2^{-\gamma})^2 \geqq (\lambda(bn^{-1})/\lambda(n^{-1}))^{2k-2}(\eta(\lambda(bn^{-1})/2)\eta(\lambda(n^{-1})/2))^2 .$$

Let $a(b) = \liminf_{n\to\infty} \lambda(bn^{-1})/\lambda(n^{-1})$. Then one easily shows

(3.15) $$\liminf_{n\to\infty} \eta(\lambda(bn^{-1})/2)/\eta(\lambda(n^{-1})/2) \geqq a(b)\mu(a(b)-) .$$

See (1.1 c) for a definition of μ.

Thus using (3.14) and (3.15) we obtain

(3.16) $$\liminf_{n\to\infty} n(\lambda(n^{-1}))^m(a_n 2^{-\gamma})^2 \geqq (a(b))^{2k}(\mu(a(b)-))^2 .$$

Again from (3.13) we obtain

(3.17) $$b = (a(b))^{2k+m}(\mu(a(b)+))^2 .$$

Since $\mu(a(b)) > 0$ and μ is a nondecreasing function, it follows that

(3.18) $$\lim_{b\to\infty} a(b) = \infty .$$

Let b_0 be a constant such that if $b \geqq b_0$ then $a(b) \geqq 1$. Then, since $2k \geqq 0$, if $b \geqq b_0$ we obtain from (3.16) that

(3.19) $$\liminf_{n\to\infty} n(\lambda(n^{-1}))^m(a_n 2^{-\gamma})^2 \geqq (\mu(a(b)-))^2 .$$

The constant $b > 0$ is arbitrary. Hence (1.7) follows from (3.19) together with (1.1 c). □

PROOF OF THEOREM 1.2. The hypothesis (1.8) implies that if $\{b_n, n \geqq 1\}$ is a sequence tending to ∞ then

(3.20) $$\lim_{n\to\infty} \inf_{f\in C_{k\eta}} P_f(|\gamma_n - f(0)| \leqq a_n b_n) = 1 .$$

By Theorem 1.1,

(3.21) $$\liminf_{n\to\infty} n(\lambda(n^{-1}))^m(a_n b_n)^2 = \infty .$$

This clearly implies that

(3.22) $$\liminf_{n\to\infty} n(\lambda(n^{-1}))^m a_n^2 > 0 .$$

PROOF OF THEOREM 1.3. In this proof we set

(3.23) $$h_\delta(x) = f(x) + e_{k^*\delta}(x) , \qquad k^* \leqq k,\ \tau^* \leqq \tau, \text{ and}$$

let $\delta = \delta(n)$ be given by (3.6), which for the special choice of τ is

(3.24) $$\delta(n) = (b/n)^{(k^*+m+2\tau^*-2)^{-1}} .$$

Let $\{d_n, n \geqq 1\}$ be a subsequence of the integers on which

(3.25) $$\liminf_{n\to\infty} P_{h_{\delta(d_n)}}(|\gamma_{d_n} - h_{\delta(d_n)}(0)| \leqq a_{d_n}) > 0 .$$

Let $\{b_n, n \geqq 1\}$ be a real number sequence increasing to ∞. Repeating steps (3.3) to (3.10) we obtain

(3.26) $$\liminf_{n\to\infty} P_f(|\gamma_{d_n} - h_{\delta(d_n)}(0)| \leqq a_{d_n} b_{d_n}) > 0 , \qquad \text{and}$$
$$\liminf_{n\to\infty} P_f(|\gamma_{d_n} - f(0)| \leqq a_{d_n} b_{d_n}) = 1 .$$

Therefore, as in (3.11), there exists n_0 such that if $n \geqq n_0$ then

(3.27) $$2a_{d_n} b_{d_n} \geqq e_{k^*\delta(d_n)}(0) = 2^{\gamma+1}(\delta(d_n))^{k^*-1}\eta(\delta(d_n)/2) .$$

Substitution of (1.10) for a_{d_n} and (3.24) for $\delta(d_n)$ shows (3.27) to be contradictory. Thus (1.11) holds with $f_n = h_{\delta(n)}$. □

REFERENCES

[1] EPSNECNIKOV, V. A. (1969). Nonparametric estimation of a multivariate probability density. *Teor. Verojatnost i Primenen* **1** 156–162.

[2] FARRELL, R. H. (1967). On the lack of a uniformly consistent sequence of estimators of a density function in certain cases. *Ann. Math. Statist.* **38** 471–474.

[3] LEADBETTER, M. R. (1963). On the non-parametric estimation of probability densities. Technical Report No. 11, Research Triangle Institute.

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

[5] PICKANDS, J. (1969). Efficient estimation of a probability density. *Ann. Math. Statist.* **40** 854–864.

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

[7] ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 0000–0000.

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

[9] SCHWARTZ, S. C. (1967). Estimation of a probability density by an orthogonal series. *Ann. Math. Statist.* **38** 1261–1265.

[10] WATSON, G. S. (1969). Density estimation by orthogonal series. *Ann. Math. Statist.* **40** 1496–1498.

[11] WATSON, G. S. and LEADBETTER, M. R. (1963). On estimating a probability density function I. *Ann. Math. Statist.* **34** 480–491.

[12] WEISS, L. and WOLFOWITZ, J. (1967). Estimation of a density at a point. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **7** 327–335.

[13] WIENER, N. (1933). *The Fourier Integral and Certain of its Applications.* Cambridge University Press; Dover Books (0000).