Z. Wahrscheinlichkeitstheorie verw. Gebiete
62, 475–483 (1983)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

On Arbitrarily Slow Rates of Global Convergence in Density Estimation

Luc Devroye*

McGill University, School of Computer Science, 805 Sherbrooks Street West, Montreal, Canada H3A 2K6

Summary. Let a density f on R^d be estimated by $f_n(x, X_1, \ldots, X_n)$ where $x \in R^d$, f_n is a Borel measurable function of its arguments, and $X_1, \ldots, X_n$ are independent random vectors with common density f. Let $p \geqq 1$ be a constant. One of the main results of this note is that for every sequence f_n, and for every positive number sequence a_n satisfying $\lim_n a_n = 0$, there exists an f such that

$$E\left(\int |f_n(x) - f(x)|^p \, dx\right) > a_n \quad \text{infinitely often.}$$

Here it suffices to look at all the f that are bounded by 2 and vanish outside $[0, 1]^d$. For $p = 1$, f can always be restricted to the class of infinitely many times continuously differentiable densities with all derivatives absolutely bounded and absolutely integrable.

1. Introduction

Assume that one has to estimate a density f on R^d from $X_1, \ldots, X_n$, a sequence of independent random vectors with common density f. A *density estimate* is a sequence (f_n) of Borel measurable mappings: $R^{d(n+1)} \to R$; for fixed n, $f(x)$ is estimated by $f_n(x) = f_n(x, X_1, \ldots, X_n)$. In this note, we take a look at the rate of convergence of $E(\int |f_n(x) - f(x)|^p \, dx)$ $(p \geqq 1)$ for all density estimates. We could for instance inquire about the uniform rate of convergence over a suitable class of densities Ξ, i.e. our criterion is

$$\sup_{f \in \Xi} E\left(\int |f_n(x) - f(x)|^p \, dx\right)$$

or

$$\sup_{f \in \Xi} E\left(\int |f_n(x) - f(x)|^p \, dx\right) \Big/ \int f^p(x) \, dx.$$

* Research carried out while the authors was visiting the Applied Research Laboratories, University of Texas at Austin

Most of the results known to date about the uniform rate of convergence are summarized in the work of Bretagnolle and Huber (1979). We will merely complement their results with a couple of interesting observations.

Our main results concern the rate of convergence for individual f. If f_n is given, and if $a_n \to 0$ is a sequence of positive numbers, can we find an f in Ξ such that

$$\limsup_n a_n^{-1} E(\int |f_n(x)-f(x)|^p \, dx)/\int f^p(x)\, dx \geqq 1?$$

Notice that the *same* f is considered throughout the sequence. Perhaps the first result in this direction was proved by Boyd and Steele (1978): for any density estimate, there exists an f in $\Xi = \{$all normal densities on R with zero mean$\}$ and a constant $c(f)>0$ such that

$$\limsup_n nE(\int |f_n(x)-f(x)|^2 \, dx) \geqq c(f).$$

We will see below that if Ξ is slightly enlarged, then any slow rate of convergence to 0 can be achieved for $E(\int |f_n(x)-f(x)|^2\, dx)$. The result of Boyd and Steele cannot be improved for normal density estimation: for example, when $d=1$, f is normal (μ, σ^2) and f_n is normal $(\hat{\mu}, \hat{\sigma}^2)$ where $\hat{\mu}, \hat{\sigma}^2$ are the usual sample-based estimates of μ and σ^2, then

$$\lim_n P\left(\int (f_n(x)-f(x))^2\, dx < \frac{x}{16\sqrt{\pi}\, n\sigma}\right) = F(x)$$

where F is the distribution function of $4V+3U$ and V, U are independent chi-square random variables with one degree of freedom (see Maniya (1969) who also has a similar result for $d>1$). Thus the rate predicted by Boyd and Steele can be achieved.

For a discussion of the best possible rates of pointwise convergence of density estimates, we refer to the work of Farrell (1967, 1972), Wahba (1975) and Stone (1981). There is an extensive literature about the rates of convergence, both pointwise and global, for particular density estimates. The only estimate that we will refer to in this note is the *kernel estimate*

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

where h_n is a sequence of positive numbers, and K is a given density on R^d. The pointwise convergence rate is studied by Wahba (1975) and Rosenblatt (1971), and its L_2 global convergence rate is investigated by Rosenblatt (1971), Nadaraya (1974), Bretagnolle and Huber (1979) (who also consider L_p convergence for $p \geqq 1$) and others too numerous to mention here. For the L_1 rate of convergence, we refer to Abou-Jaoude (1977).

The following classes of densities are important to us:

G: all densities vanishing outside $[0, 1]^d$ and bounded by 2.

$G(g)$: all densities of the form

$$\sum_{i=1}^{\infty} p_i\, g(x+x_i)$$

where g is an arbitrary fixed density with support contained in $[0,1]^d$, $(p_1, p_2, \ldots)$ is a probability vector, and $x_1, x_2, \ldots$ are points in $\mathbb{R}^d$ spread out well enough such that for all x $g(x+x_i)=0$, all i except possibly for one such i. [*Note:* If $d=1$ and $g(x)=\text{constant} \times \exp\left(-\frac{1}{x(1-x)}\right)$ on $[0,1]$, then every density in $G(g)$ is infinitely many times continuously differentiable.]

L_p: all densities on $\mathbb{R}^d$ for which $\int f^p(x)\,dx<\infty$.

Theorem 1. Let f_n be *any* density estimate, and let $p \geqq 1$ be fixed. Let $f \in L_p$.

(i)
$$\inf_n \sup_{f\in G(g)} E\left(\int |f_n(x)-f(x)|^p\,dx\right)/\int f^p(x)\,dx \geqq 1/2^{p-1};$$
$$\inf_n \sup_{f\in G} E\left(\int |f_n(x)-f(x)|^p\,dx\right)/\int f^p(x)\,dx \geqq 1/2^{p-1}.$$

(ii) Let $a_n \to 0$ be a sequence of positive numbers.
Then
$$\sup_{f\in G} \limsup_n a_n^{-1} E\left(\int |f_n(x)-f(x)|^p\,dx\right)/\int f^p(x)\,dx=\infty$$
and
$$\sup_{f\in G(g)} \limsup_n a_n^{-1} E\left(\int |f_n(x)-f(x)|\,dx\right)=\infty.$$

Remark 1. (*The optimality of* (ii)). For $p=2$, result (ii) partially strengthens the theorem of Boyd and Steele (1979) mentioned earlier. It is also not vacuous because there are density estimates for which

$$\lim_n E\left(\int |f_n(x)-f(x)|\,dx\right)=0 \tag{1}$$

for *all* f: for the histogram estimate, see Abou-Jaoude (1976); for the kernel estimate and recursive versions of it, see Devroye (1979); (1) is known to hold for the kernel estimate for *all* bounded K with integrable radial majorant and *all* f when

$$\lim_n h_n+(nh_n^d)^{-1}=0$$

(Devroye and Wagner, 1979). Note that even for the small class G no meaningful rate of convergence result is possible. When $p=1$, there exist densities in $G(g)$ for any g, that yield an arbitrarily slow rate of convergence. In other words, tail conditions alone, or smoothness conditions alone do not suffice to study the L_1 rate of convergence for *any* density estimate. For the practitioner who uses nonparametric density estimates because he does not have enough information about f in the first place, result (ii) is disastrous.

Remark 2. Result (ii) implies that G is too rich to study the L_p rate of convergence for any $p \geqq 1$ and any density estimate, and that $G(g)$ is too rich to do the same for the L_1 rate of convergence. These results do not contradict the work of Bretagnolle and Huber (1979) who showed the following: for a suitable modification of the kernel estimate (i.e., h_n is a carefully chosen function of the data $X_1, \ldots, X_n$ and the kernel K satisfies $\int K=1$, $\int x^j K=0$ for all $1 \leqq j < s$, and $\int |x|^s K < \infty$ for some $s \geqq 1$) and for $p \geqq 1$, $d=1$,

$$\sup_{\substack{f:\\ f^{(s)}\in L_p,\ f\in L_p\\ f \text{ compact support when } 1\leqq p<2}} \limsup_n n^{sp/(2s+1)} E\left(\int |f_n(x)-f(x)|^p\,dx\right)/D_{sp}(f) \leqq C_{sp} \tag{2}$$

where $C_{sp}>0$ is a constant depending upon s and p only, and

$$D_{sp}(f) = \left(\int |f^{(s)}(x)|^p\,dx\right)^{p/(2s+1)} \left(\int f^{p/2}(x)\,dx\right)^{sp/(2s+1)}.$$

For $1 \leqq p < 2$, they also require that K have compact support.

Thus, the kernel estimate can achieve a certain rate of convergence for certain classes of densities: for example, for $p=1$, the densities considered by Bretagnolle and Huber have compact support and have $f^{(s)} \in L_p$ (any density in $G(g)$ certainly satisfies the latter condition when g does). By our Theorem 1, the omission of one of their conditions will invalidate the result. For $p=2$, $s \geqq 1$, any density in $G(g)$ will satisfy (2) when $g^{(s)} \in L_p$, $g \in L_p$. Note however that Theorem 1(ii) gives no information about $G(g)$ when $p>1$.

Remark 3. (Uniform Rate of Convergence). Theorem 1(i) implies that for all $p \geqq 1$, G and $G(g)$ are too rich to study the uniform rate of convergence of any density estimate. This complements the following result of Bretagnolle and Huber (1979): let $d=1$, and let $D(p,s,c)$ be the class of all f on R for which $f^{(s)} \in L_p$, $f \in L_p$ ($s \geqq 1$ is an integer) and for which $D_{sp}(f) \leqq c$. Then for any density estimate

$$\liminf_n n^{sp/(2s+1)} \sup_{f\in D(p,s,c)} E\left(\int |f_n(x)-f(x)|^p\,dx\right) \begin{matrix} \geqq C_{sp}c, & p>1, \\ \geqq C_{sp}c-(2e)^{-4}, & p=1. \end{matrix}$$

Here $C_{sp}>0$ depends only upon s and p.

Remark 4. Rosenblatt (1971) has shown that the kernel estimate satisfies, for $d=1$,

$$E\left(\int |f_n(x)-f(x)|^2\,dx\right) \sim \frac{\alpha}{nh_n} + \frac{\beta}{4} h_n^4$$

as $n\to\infty$, $h_n \to 0$, $nh_n \to \infty$, K is bounded and symmetric and f belongs to the class $F=\{$all densities on $\mathbb{R}^1$ that are twice continuously differentiable and for which f is bounded, $f^2, f''^2 \in L_p\}$. The constants are

$$\alpha = \int K^2(x)\,dx, \qquad \beta = \left(\int x^2 K(x)\,dx\right)^2 \int f''^2(x)\,dx.$$

Thus, taking $h_n = (\alpha/(\beta n))^{1/5}$ gives the optimal L_2 rate

$$E\left(\int |f_n(x)-f(x)|^2\,dx\right) \sim \tfrac{5}{4}\alpha^{4/5}\beta^{1/5}/n^{4/5}.$$

(See also Nadaraya (1974).) Yet, at the same time, for *some* f in F,

$$E\left(\int |f_n(x)-f(x)|\,dx\right) \geqq \frac{1}{\text{logloglogog}\, n}$$

infinitely often (by Theorem 1). In other words, Rosenblatt's result (and most other L_2 results) gives us little information about how close f_n is to f (and should certainly not be used to determine a good value for h_n). The discrepancy between good L_2 rates and bad L_1 rates is due to the fact that in L_2, tails are less important. For the study of the L_1 rate of kernel estimates, F is too rich. With additional tail conditions, it is easily seen that the optimal L_1 rate is $n^{-2/5}$ (Rosenblatt, 1979; see also remark 2).

2. Proofs

We will use two families of densities throughout the proofs section.

Family 1. Let g be a density with support on $[0,1]^d$, and let $g_y(x)=g(x-y)$. Let $b\in[0,1]$ have binary expansion $b_0\cdot b_1 b_2 b_3 \ldots$, and define the density f on R^d parametrized by b as follows:

$$f(b,x)=\sum_{i=1}^{\infty} p_i\, g_{(2i,0,\ldots,0)+(b_i,0,\ldots,0)}(x)$$

where $(p_1, p_2, \ldots)$ is a fixed probability vector. Note that f is a density for each b, and that

$$\int f^p(b,x)\,dx=\left(\sum_{i=1}^{\infty} p_i^p\right)\int g^p(x)\,dx.$$

Family 2. Partition $[0,1]^d$ into sets A_i, A_i', $i=1,2,\ldots$ where $\int_{A_i} dx=\int_{A_i'} dx=p_i/2$, and $(p_1, p_2, \ldots)$ is a fixed probability vector. Let b be as for family 1, and define the density parametrized by b as follows:

$$f(b,x)=2\sum_{i=1}^{\infty} I_{b_i A_i+(1-b_i)A_i'}(x)$$

where I is the indicator function. Clearly, $f\in G$. Also, $\int f^p(b,x)\,dx=2^{p-1}$, all $p>0$.

Proof of Theorem 1. For fixed $b\in[0,1]$, the density parametrized by b as in families 1 or 2 will be denoted by f_b or $f_b(x)$. B is a uniform $[0,1]$ random variable; given B, let $X_1, \ldots, X_n$ be independent random vectors with common density f_B. All the integrals that follow are with respect to dx. Let us define for family 1,

$$C_i=((2i,0,\ldots,0)+[0,1]^d)\cup((2i+1,0,\ldots,0)+[0,1]^d)$$

where "+" is the translation operator on sets, and for family 2, $C_i=A_i\cup A_i'$. Let $N_i=\sum_{j=1}^{n} I_{C_i}(X_j)$ where I is the indicator function. Note that by construction $N=(N_1, N_2, \ldots)$ is independent of B. Let B', B'' be random variables equal to B except in their i-th digits, where we force $B_i'=0$, $B_i''=1$. We have for all $p\geqq 1$:

$$\sup_{0\leqq b\leqq 1} E\left(\int |f_n-f_B|^p \mid B=b\right)\geqq E\left(\int |f_n-f_B|^p\right)=E\left(\sum_{i=1}^{\infty}\int_{C_i} |f_n-f_B|^p\right). \tag{3}$$

On $N_i=0$, B_i and $X_1, \ldots, X_n$ are conditionally independent. Thus,

$$\begin{aligned}
&E(I_{N_i=0} \int_{C_i} |f_n - f_B|^p \mid X_1, \ldots, X_n) \\
&= E(I_{N_i=0} I_{B_i=0} \int_{C_i} |f_n - f_{B'}|^p + I_{N_i=0} I_{B_i=1} \int_{C_i} |f_n - f_{B''}|^p \mid X_1, \ldots, X_n) \\
&= \tfrac{1}{2} I_{N_i=0} E(\int_{C_i} |f_n - f_{B'}|^p + \int_{C_i} |f_n - f_{B''}|^p \mid X_1, \ldots, X_n) \\
&\geqq I_{N_i=0} 2^{-p} E(\int_{C_i} |f_{B'} - f_{B''}|^p). \qquad (4)
\end{aligned}$$

For family 1, we have

$$\int_{C_i} |f_{B'} - f_{B''}|^p = 2 p_i^p \int g^p, \qquad (5)$$

and for family 2,

$$\int_{C_i} |f_{B'} - f_{B''}|^p = 2^p \int_{C_i} dx = 2^p p_i. \qquad (6)$$

If we repeat (4–6) for all i, take expectations in (4) and substitute into (3), we have for family 1

$$\begin{aligned}
&\sup_{0 \leqq b \leqq 1} E(\int |f_n - f_B|^p \mid B=b) / \int f^p \\
&\geqq \left(\frac{1}{2^{p-1}}\right) \sum_{i=1}^{\infty} P(N_i=0)\, p_i^p \Big/ \sum_{i=1}^{\infty} p_i^p \\
&= \left(\frac{1}{2^{p-1}}\right) \sum_{i=1}^{\infty} p_i^p (1-p_i)^n \Big/ \sum_{i=1}^{\infty} p_i^p. \qquad (7)
\end{aligned}$$

For family 2, the left-hand-side of (7) is at least equal to

$$\sum_{i=1}^{\infty} p_i (1-p_i)^n. \qquad (8)$$

Note that (7) and (8) are valid for all n and all f_n. Consider for example $p_i = K^{-1}$, $1 \leqq i \leqq K$, and $p_i=0$, $i>K$. Then (7) equals $(1-1/K)^n/2^{p-1}$ and the supremum of this over all K is $1/2^{p-1}$. Also, (8) equals $(1-1/K)^n/2^{p-1}$. This proves Theorem 1(i).

We will first show that when $a_n \leqq 1/8$ for all n, there exists a sequence $(p_1, p_2, \ldots)$ such that

$$\sum_{i=1}^{\infty} p_i (1-p_i)^n \geqq a_n, \quad \text{all } n. \qquad (9)$$

This can be shown by construction. We will in fact show (9) for $a'_n = \max_{m \geqq n} a_m + 1/(4(n+1))$. Note that a'_n tends to 0 strictly monotonically and that $a'_1 \leqq 1/4$. We find integers $1=k_1<k_2<\ldots$ and positive numbers p_i such that $p_1 = 1-2a'_1$, and for $n \geqq 2$, $k_{n-1} < i \leqq k_n$:

$$p_i \leqq 1/(2n), \qquad \sum_{i=k_{n-1}+1}^{k_n} p_i = 2(a'_{n-1} - a'_n).$$

Note that

$$\sum_{i=1}^{\infty} p_i = p_1 + \sum_{n=2}^{\infty} 2(a'_{n-1} - a'_n) = 1 - 2a'_1 + 2a'_1 = 1.$$

Also, for $n \geqq 2$,

$$\sum_{i=1}^{\infty} p_i (1-p_i)^n \geqq (1-1/(2n))^n \sum_{p_i \leqq 1/(2n)} p_i \geqq (1/2) \sum_{p_i \leqq 1/(2n)} p_i$$

$$\geqq (1/2) \sum_{i=k_{n-1}+1}^{\infty} p_i = (1/2) \sum_{i=n}^{\infty} 2(a'_{i-1} - a'_i) = a'_{n-1} \geqq a'_n \geqq a_n.$$

For $n=1$, we have $p_1(1-p_1) = 2a'_1(1-2a'_1) \geqq a'_1 \geqq a_1$. This concludes the proof of (9).

Let us define $J_n(b) = E(\int |f_n - f_B|^p \mid B=b)$ and $\bar{J}_n(b) = \sup_{m \geqq n} J_m(b)/a_m$. If $\bar{J}_n(b) \not\to 0$ for some b, then for that b, we have

$$\limsup_n J_n(b)/a_n > 0 \tag{10}$$

and we are done. Thus, let us assume that $\bar{J}_n(b) \to 0$ for all $b \in [0,1]$. We will now prove that this leads to a contradiction. Let $D_n = \{b: \bar{J}_n(b) > 1\}$. Since D_n decreases monotonically to the empty set, we have $\int_{D_n \cap [0,1]} dx = o(1)$ by the Lebesque dominated convergence theorem. Let D'_n be the complement of D_n. Clearly, by Fatou's lemma,

$$\begin{aligned} 0 &= \sup_{0 \leqq b \leqq 1} \limsup_n \frac{J_n(b)}{a_n} I_{D'_n}(b) \\ &\geqq E\left(\limsup_n \frac{J_n(B)}{a_n} I_{D'_n}(B)\right) \\ &\geqq \limsup_n E\left(\frac{J_n(B)}{a_n} I_{D'_n}(B)\right). \end{aligned} \tag{11}$$

Consider family 2. In (4) we make a few changes: introduce $X = (X_1, X_2, \ldots)$. Then,

$$\begin{aligned} & E(I_{N_i=0}\, a_n^{-1} I_{D'_n}(B) \int_{C_i} |f_n - f_B|^p \mid X) \\ & \geqq I_{N_i=0} (2a_n)^{-1} E(I_{\bar{J}_n(B') \leqq 1} \int_{C_i} |f_n - f_{B'}|^p + I_{\bar{J}_n(B'') \leqq 1} \int_{C_i} |f_n - f_{B''}|^p \mid X) \\ & \geqq I_{N_i=0} (1/2^p a_n) E(I_{\max(\bar{J}_n(B'), \bar{J}_n(B'')) \leqq 1} \int_{C_i} |f_{B'} - f_{B''}|^p \mid X) \\ & \geqq I_{N_i=0} (p_i/a_n) E(I_{\max(\bar{J}_n(B'), \bar{J}_n(B'')) \leqq 1} \mid X) \end{aligned}$$

and the expected value of the last expression is

$$\begin{aligned} & p_i a_n^{-1} P(N_i = 0, \max(\bar{J}_n(B'), \bar{J}_n(B'')) \leqq 1) \\ & \geqq p_i a_n^{-1} (P(N_i=0) - P(N_i=0, \bar{J}_n(B') > 1) - P(N_i=0, \bar{J}_n(B'') > 1)) \\ & \geqq p_i a_n^{-1} (P(N_i=0) - 2P(N_i=0, \bar{J}_n(B) > 1) - 2P(N_i=0, \bar{J}_n(B) > 1)) \\ & = p_i a_n^{-1} (P(N_i=0) - 4P(N_i=0, D_n)). \end{aligned} \tag{12}$$

Let $A_n=[k_{n-1}, \infty)$ where k_n is as defined earlier, and set

$$Z_n=\sum_{i\in A_n} p_i I_{N_i=0}.$$

We have shown that

$$E\left(\frac{J_n(B)}{a_n} I_{D_n'}(B)\right) \geqq E\left(\frac{Z_n}{a_n}\right)-4E\left(\frac{Z_n}{a_n} I_{D_n}(B)\right)$$
$$\geqq E\left(\frac{Z_n}{a_n}\right)-4\sqrt{E\left(\frac{Z_n^2}{a_n^2}\right) P(B\in D_n)}$$

(by Schwarz's inequality). Since $P(B\in D_n)=o(1)$, we obtain our desired contradiction if $E(Z_n/a_n) \not\to 0$ and $E(Z_n^2)=O(E^2(Z_n))$. But

$$E\left(\frac{Z_n}{a_n}\right)=\frac{1}{a_n}\sum_{i>k_{n-1}} p_i(1-p_i)^n \not\to 0$$

by our construction. Also,

$$\begin{aligned} E(Z_n^2) &= \sum_{i\in A_n} p_i^2(1-p_i)^n+\sum_{i\neq j,\, i,\, j\in A_n} p_i p_j(1-p_i-p_j)^n \\ &\leqq 2\sum_{i\in A_n} p_i^2(1-p_i)^{2n}+\sum_{i\neq j,\, i,\, j\in A_n} p_i(1-p_i)^n p_j(1-p_j)^n \\ &\leqq 2E^2(Z_n) \end{aligned}$$

where we used the fact that on A_n, $(1-p_i)^n \geqq \frac{1}{2}$, and that in any case $1-p_i-p_j \leqq (1-p_i)(1-p_j)$. Since $\int f^p(b,x)\,dx=2^{p-1}$ for all b, and since we can always replace a_n by $\sqrt{a_n}$, we have proved Theorem 1(ii) for family 2. For family 1, we note that $(7)=(8)$ for $p=1$ and that $\int f^p(b,x)\,dx =(\sum p_i^p)\, g^p(x)\,dx \leqq \int g^p(x)\,dx<\infty$. Thus, the argument for family 2 can be mimicked, and this concludes the proof of Theorem 1.

Acknowledgement. The author gratefully acknowledges the helpful remarks of one referee. This research was carried out under sponsorship of the office of Naval Research contract N00014-81-K-0145 while the author was visiting the University of Texas during the summer of 1981.

References

Abou-Jaoude, S.: Conditions nécessaires et suffisantes de convergence $L1$ en probabilité de l'histogramme pour une densité. Ann. Inst. Henri Poincaré **12**, 213–231 (1976)

Abou-Jaoude, S.: Thèse. Université de Paris VI (1977)

Boyd, D.W., Steele, J.M.: Lower bounds for nonparametric density estimation rates. Ann. Statist. **6**, 932–934 (1978)

Bretagnolle, J., Huber, C.: Estimation des densités: risque minimax. Z. Wahrscheinlichkeitstheorie verw. Gebiete **47**, 119–137 (1979)

Devroye, L.: On the pointwise and the integral convergence of recursive estimates of probability densities. Utilitas Mathematica **15**, 113–128 (1979)

Devroye, L., Wagner, T.J.: The $L1$ convergence of kernel density estimates. Ann. Statist. **7**, 1136–1139 (1979)

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

Farrell, R.H.: On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. Ann. Math. Statist. **43**, 170–180 (1972)

Maniya, G.M.: The square error of the density estimate of a multidimensional normal distribution for a given sample. Theor. Probability Appl. **14**, 149–153 (1969)

Nadaraya, E.A.: On the integral mean square error of some nonparametric estimates for the density function. Theor. Probability Appl. **19**, 133–141 (1974)

Rosenblatt, M.: Curve estimates. Ann. Math. Statist. **42**, 1815–1842 (1971)

Rosenblatt, M.: Global measures of deviation for kernel and nearest neighbor density estimates. In: Smoothing Techniques For Curve Estimation, Th. Gasser and M. Rosenblatt Eds., Lecture Notes in Mathematics #**757**, pp. 181–190. Berlin-Heidelberg-New York: Springer 1979

Stone, C.J.: Optimal rates of convergence for nonparametric estimators. Ann. Statist. **8**, 1348–1360 (1980)

Wahba, G.: Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation. Ann. Statist. **3**, 15–29 (1980)

Received October 1, 1981; in revised form September 2, 1982