The Annals of Statistics
1981, Vol. 9, No. 2, 347–355

ESTIMATION OF A MULTIVARIATE DENSITY FUNCTION USING DELTA SEQUENCES[1]

BY V. SUSARLA AND G. WALTER

Michigan State University and University of Wisconsin-Milwaukee

This paper studies some asymptotic properties of density estimates $\hat{f}$ of f based on d-variate delta sequences. The mean-square consistency, almost sure consistency, and asymptotic normality of $\hat{f}$ have been obtained as corollaries to the L_1 convergence properties of these delta sequences. Estimators based on kernel functions, orthogonal series, and some histogram methods can be obtained as special cases of $\hat{f}$.

1. Introduction. Since the pioneering work of Rosenblatt (1956) and Parzen (1962) on the kernel estimates of density f, many other methods have been proposed to estimate f. For details concerning these methods, see the references Wegman (1972 a, b), Wertz (1978), and Wertz and Schneider (1979). Some of the methods of estimation of a univariate density f can be extended to the multivariate case (for example, see Cacoullos (1966), and Deheuvels (1977)). A somewhat more general method was studied by several authors using the so-called "delta sequence" approach including, among others, Bleuez and Bosq (1976 a, b), Bosq (1977), Delecroix (1977), Földes and Révész (1974), Walter and Blum (1979), Watson and Leadbetter (1965), and Winter (1975). This method includes the kernel method, the orthogonal series method of Čencov (1962), some of the polynomial approximation methods, and some of the histogram methods. In this paper, we propose to extend this general method of density estimation to the multivariate case.

The outline of the rest of the paper is as follows: Sections 2 and 3 study some of the asymptotic properties (needed for Sections 4 and 5) of two types of delta sequences in d-dimensions. Section 4 proposes estimators $\hat{f}$ of f based on these delta sequences in Sections 2 and 3, and presents some of their properties. Section 5 discusses an asymptotic normality result for $\hat{f}$.

Throughout, $\mathbf{x} = (x_1, \ldots, x_d)$ denotes a generic point in the d-dimensional Euclidean space R^d and $d\mathbf{x}$ denotes $\pi_{i=1}^d dx_i$. For an open set $\Omega \subset R^d$, let $C_0^\infty(\Omega)$ denote the space of infinitely differentiable functions on Ω having compact support. Unless otherwise stated, all the integrals are over the full range of the values of arguments involved.

2. Positive delta sequences. In this section, we introduce the basic definition of one type of multivariate delta sequence and prove a number of convergence theorems.

DEFINITION 2.1. A sequence of functions $\{\delta_m\}$ in $L^\infty(\Omega \times \Omega)$ is said to be a *delta sequence* on Ω if for each $\phi \in C_0^\infty(\Omega)$, and $\mathbf{x} \in \Omega$,

$$\int_\Omega \delta_m(\mathbf{x}, \mathbf{y})\phi(\mathbf{y})\, d\mathbf{y} \to \phi(\mathbf{x}) \quad \text{as} \quad m \to \infty.$$

Received January, 1979; revised January, 1980.

[1] Research supported in part by NSF under grant MCS77-26809 and in part by NIH grant GM23129.

AMS 1980 *subject classifications*. Primary 62F10, 62G05, 41A25; secondary 46F99.

Key words and phrases: Density estimation, delta sequences, mean square convergence, asymptotic normality.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*. www.jstor.org

EXAMPLE 2.1. Let A_m be a nonsingular linear transformation from R^d to R^d such that $|A_m| = O(m^d)$. Let $\delta_m(\mathbf{x}, \mathbf{y}) = |A_m| k(A_m(\mathbf{y} - \mathbf{x}))$ for a density function k such that $k(\mathbf{v}) = o(|\mathbf{v}|^{-d+2})$ as $|\mathbf{v}| \to \infty$. Then $\{\delta_m\}$ satisfies Definition 2.1 (see (1.1), Deheuvels (1977)). Deheuvels (1977), generalizing (among others) the estimators of Cacoullos (1966), and Parzen (1962) (see also Wertz (1978)), provides several asymptotic properties including mean square consistency and the integrated mean square errors of density estimators based on $\{\delta_m\}$.

DEFINITION 2.2. A delta sequence $\{\delta_m\}$ on R^d is said to be of *positive type* if $\delta_m \geq 0$ and for each $\mathbf{x}$ in R^d,

(i) $\int \delta_m(\mathbf{x}, \mathbf{y})\, d\mathbf{y} = 1$,

(ii) $\sup_{r>0} r\{\int_{|\mathbf{x}-\mathbf{y}|>r} \delta_m(\mathbf{x}, \mathbf{y})\, d\mathbf{y}\} = O(m^{-1})$,

(iii) $\| \delta_m(\mathbf{x}, \cdot) \|_\infty = O(m^d)$ as $m \to \infty$,

and for each $\eta > 0$,

(iv) $\sup\{\delta_m(\mathbf{x}, \mathbf{y}) \mid |\mathbf{x} - \mathbf{y}| > \eta\} \to 0$.

NOTE 2.1. The order relations in (ii), (iii), and (iv) could depend on $\mathbf{x}$.

The following two propositions, whose proofs are straightforward, give simple sufficient conditions for a sequence of functions on R^d to be a delta sequence of this type.

PROPOSITION 2.1. *Let $\{\delta_m\}$ be any sequence of nonnegative functions satisfying* (i), *and additionally, be such that*

$$\delta_m(\mathbf{x}, \mathbf{y}) \leq cm^d/\{1 + m|\mathbf{x} - \mathbf{y}|\}^{d+1}.$$

Then $\{\delta_m\}$ is a delta sequence of positive type.

PROPOSITION 2.2. *Let $\{\delta_{m,1}(x_1, y_1)\}$, $\{\delta_{m,d}(x_d, y_d)\}$ each be a delta sequence of positive type, then $\delta_m(\mathbf{x}, \mathbf{y}) = \pi_{i=1}^{d} \delta_{m,i}(x_i, y_i)$ is a delta sequence of positive type.*

The conclusion of the first part of the following lemma is similar to the results available in the literature (see Wertz (1978)) and deals with a.e. convergence properties of delta sequences while the second part provides a rate at which such convergence holds.

LEMMA 2.1. *Let $f \in L^p(R^d)$, $1 \leq p \leq \infty$ and let $\{\delta_m\}$ be a delta sequence of positive type.*

(i) *Then*

$$\int \delta_m(\mathbf{x}, \mathbf{t}) f(\mathbf{t})\, d\mathbf{t} \to f(\mathbf{x}) \quad \text{a.e. } \textit{Lebesgue}(\mathbf{x}).$$

(ii) *If $p > 1$, and $|f(\mathbf{x} + \mathbf{t}) - f(\mathbf{x})| \leq c\,|\mathbf{t}|^\gamma$ for a $0 < \gamma < 1$, then*

$$\left| \int \delta_m(\mathbf{x}, \mathbf{t}) f(\mathbf{t})\, d\mathbf{t} - f(\mathbf{x}) \right| = O(m^{-\delta}) \qquad \textit{where } \delta = \min\{q^{-1}, \gamma\} \qquad \textit{with } q^{-1} = 1 - p^{-1}.$$

PROOF. Part (i) is similar to Theorem 1.25 of Stein and Weiss (1971) in our notation and its proof will be omitted. The proof of part (ii) has some analogue to the proof of the same theorem; but since we are interested in rates of convergence we must look more carefully at various null sequences involved in the proof of the theorem.

Let $\eta > 0$, then we have

$$\left| \int \delta_m(\mathbf{x}, \mathbf{t})\{f(\mathbf{t}) - f(\mathbf{x})\}\, d\mathbf{t} \right| \leq \left| \int_{|\mathbf{t}-\mathbf{x}|<\eta} \right| + \left| \int_{|\mathbf{t}-\mathbf{x}|\geq\eta} \right| = I_1 + I_2 \tag{2.1}$$

By a change of variables to polar coordinates, we obtain

$$I_1 \le c_1 \int_0^{\nu} \epsilon_m(\mathbf{x}, r) r^{d-1} r^{\gamma} dr = c_1 \left\{ \int_0^{1/m} + \int_{1/m}^{\eta} \right\} = c_1(I_1' + I_1'') \tag{2.2}$$

for sufficiently small η and for $1 \le m\eta$ where c_1 is a constant and

$$\epsilon_m(\mathbf{x}, r) = \int_{S_{\mathbf{x}}^{d-1}} \delta_m(\mathbf{x}, \mathbf{x} + r\mathbf{t}')\, d\delta(\mathbf{t}'). \tag{2.3}$$

Here $S_{\mathbf{x}}^{d-1}$ denotes the surface of the unit sphere with center at $\mathbf{x}$ and $d\delta(\mathbf{t}')$ represents the element of surface area on $S_{\mathbf{x}}^{d-1}$. In terms of this notation, Condition (ii) of Definition 2.2 becomes

$$\int_r^{\infty} \epsilon_m(\mathbf{x}, s) s^{d-1}\, ds = O(m^{-1} r^{-1}) \quad \text{uniformly in} \quad r > 0. \tag{2.4}$$

By condition (iii) of Definition 2.2, we have $I_1' = O(m^{-\gamma})$ and by integration by parts, we have

$$\begin{aligned} I_1'' &= \left\{ r^{\gamma} \int_r^{\infty} \epsilon_m(\mathbf{x}, s) s^{d-1}\, ds \right\}_{1/m}^{\eta} + \gamma \int_{1/m}^{\eta} r^{\gamma-1} \left(\int_r^{\infty} \epsilon_m(\mathbf{x}, s) s^{d-1}\, ds \right) dr \\ &\le O(m^{-1}) + O(m^{-\gamma}) + c_2 m^{-1} \int_{1/m}^{n} r^{\gamma-2} \\ &= O(m^{-1}) + O(m^{-\gamma}) \qquad (\text{or } O(m^{-1} \log m) \text{ if } \gamma = 1) \end{aligned}$$

by (2.4). Hence we have $I_1 = O(m^{-\gamma})$ or $(O(m^{-\gamma} \log m)$ if $\gamma = 1)$. We now show that I_2 also satisfies a similar order relation. By Hölder's inequality,

$$\begin{aligned} I_2 &\le \int_{|\mathbf{t}| \ge \eta} |f(\mathbf{x} + \mathbf{t})| \delta_m(\mathbf{x}, \mathbf{x} + \mathbf{t})\, d\mathbf{t} + \int_{|\mathbf{t}| > \eta} |f(\mathbf{x})| \delta_m(\mathbf{x}, \mathbf{x} + \mathbf{t})\, d\mathbf{t} \\ &\le \|f\|_p \|\psi_\eta \delta_m\|_q + |f(\mathbf{x})| \|\psi_\eta \delta_m\|_1 \qquad (2.5) \\ &\le \|f\|_p \|\psi_\eta \delta_m\|_\infty^{1/p} \|\psi_\eta \delta_m\|_1^{1/q} + |f(\mathbf{x})| \|\psi_\eta \delta_m\|_1 \end{aligned}$$

where $\psi_\eta = I\{\mathbf{t} \in R^d; |\mathbf{t}| > \eta\}$. By condition (iv) of Definition 2.2, $\|\psi_\eta \delta_m\|_\infty = 0(1)$ and by condition (ii)

$$\|\psi_\eta \delta_m\|_1 = \int_{|\mathbf{t}| \ge \eta} \delta_m(\mathbf{x}, \mathbf{x} + \mathbf{t})\, d\mathbf{t} = O(m^{-1} \eta^{-1}).$$

Hence from (2.5), $I_2 = O(m^{-1/q})$. This rate together with that for I_1 proves part (ii) of the result.

REMARK 2.1. Part (i) of the lemma still holds if we do not assume that $\delta_m \ge 0$, but merely require that $\|\delta_m\|_1$ be bounded. Such a result is analogous to Theorem 1A of Parzen (1962) for $d = 1$, Lemmas 1 through 4 of Földes and Révész (1974), the i.i.d. case of Theorem of Földes (1974) and Theorem 2.1 of Cacoullos (1966) and Corollaire 2.2 of Deheuvels (1977) for $d \ge 1$. Földes and Földes and Révész assume, in addition to the other usual conditions on f, that $\int |x|^{\alpha} f(x)\, dx < \infty$ for an $\alpha > 0$ which we do not have in our theorem and that their Φ_n (which is similar to our δ_m) satisfy a uniform (in n) Lipschitz condition. Also, Bleuez and Bosq (1976, Proposition 3) consider density estimators based on kernel and orthogonal functions and obtain necessary and sufficient conditions for the uniform mean square convergence of $\hat{f}$. Our pointwise result can be extended to a uniform result by assuming an appropriate uniform condition on f.

Listed below are two additional examples which satisfy the condition of Lemma 2.1.

EXAMPLE 2.2. The multivariate Fejer kernel of Fourier series $\delta_m(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{d} \delta_m(\mathbf{x}_i, \mathbf{y}_i)$, where

$$\delta_m(x_1, y_1) = \frac{(\sin\{(m+1)(x_1 - y_1)/2\})^2}{2\pi(\sin\{x_1 - y_1)/2\})^2} I_{[|y_1 - x_1| \le \pi]},$$

defines a positive delta sequence by Proposition 2.1 (See Winter (1975) for $d = 1$). Such estimates would be appropriate for estimating densities which have compact support.

EXAMPLE 2.3. Let K be a fixed compact subset of R^d. For each m, partition K into $K_{m,1}, \ldots, K_{m,m^d}$ disjoint sets each of which has diameter $\le c/m$ for some constant c depending on K. Let $\psi_{m,j}$ denote the indicator function of $K_{m,j}$, $j = 1, \ldots, m^d$. Then

$$\delta_m(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{m^d} \psi_{m,j}(\mathbf{x})\psi_{m,j}(\mathbf{y})$$

is a positive delta sequence satisfying the definition. A special case of this delta sequence obtains Example 3 of Földes and Révész (1974). Similar estimates have been used in classification problems by Gordon and Olshen (1978), and Van Ryzin (1966).

3. Higher order methods. In order to obtain improved rates of convergence, it is known that it is desirable to use nonpositive delta sequences. It is known in the univariate case that density estimators based on such sequences have mean square rates of convergence which approach $O(n^{-1})$. (See Walter and Blum (1979) and Wertz (1978).) However, it was also shown there that the same property which allows this faster rate keeps them from being asymptotically unbiased for *some continuous densities*. This latter property seems to us to be such a disadvantage so as to preclude their use except in exceptional cases. Thus we present only two examples in the multivariate case and do not provide here the general theory, which can be obtained.

EXAMPLE 3.1. (*The Fourier transform delta sequence*). Define

$$(3.1) \qquad \delta_m(\mathbf{x}, \mathbf{0}) = \frac{1}{(2\pi)^d} \int_{|\mathbf{t}| \le m} e^{-i\mathbf{x}\mathbf{t}} d\mathbf{t}, \qquad \delta_m(\mathbf{x}, \mathbf{y}) = \delta_m(\mathbf{x} - \mathbf{y}, \mathbf{0}).$$

The rate of convergence of $\int \delta_m f$ to f is $O(m^{-P+d/2})$ if f is assumed to satisfy the condition $(1 + |\mathbf{x}|^p)\tilde{f}$ is in $L^2(R^d)$ for $p > 1$, where $\tilde{f}$ is the Fourier transform of f. This extends the estimator of Blum and Susarla (1977) to the multivariate case.

EXAMPLE 3.2. Let Ω be a bounded region in R^d, and $q \in C^{2p}(\bar{\Omega})$. Let Φ_n be a complete orthonormal set of eigenfunctions of the operator $\triangle - q(x)$ ($\triangle$ is the Laplacian operator) which vanish on the boundary of Ω. Then define delta sequence $\{\delta_m\}$ by

$$\delta_m(\mathbf{x}, \mathbf{y}) = \sum_{k=0}^{m} \phi_k(\mathbf{x})\, \phi_k(\mathbf{y}).$$

If $f \in C^{2p}$, where $4p > d + 5$ and has support in the interior of Ω, then

$$\left| f(\mathbf{x}) - \int_\Omega \delta_m(\mathbf{x}, \mathbf{y}) f(\mathbf{y})\, d\mathbf{y} \right| = \left| \sum_{k=m+1}^{\infty} \phi_k(\mathbf{x}) \langle \phi_k, f\rangle \right|$$

$$= \left| \sum_{k=m+1}^{\infty} \frac{\phi_k(\mathbf{x})}{\lambda_k^p} \langle (\triangle - q)^p \phi_k, f\rangle \right|$$

$$= \left| \sum_{k=m+1}^{\infty} \frac{\phi_k(\mathbf{x})}{\lambda_k^p} \langle \phi_k, (\triangle - q)^p f\rangle \right|$$

$$\leq c(\mathbf{x})\lambda_{m+1}^{5/4+d/4-p} \| (\triangle - q)^p f \|_2^2$$

where $\{\lambda_k\}$ are the eigenvalues of $\triangle - q$. This last inequality follows from the facts that $\lambda_k \sim ck^{2/d}$ (see page 169 of Titchmarsh (1958)), $|\phi_n(\mathbf{x})| < c(\mathbf{x})\lambda_k$ and Schwarzs inequality. Hence the bias has rate $O(m^{(d+5-4p)/2d})$.

If $q = 0$, and $\Omega = (\mathbf{o}, \pi)$, we obtain trigonometric series estimator. (See Bleuez and Bosq (1976b)). This may also be extended to $\Omega = R^d$ for $q(\mathbf{x}) = |\mathbf{x}|^2$. In this case, we obtain Hermite series estimators (see Čencov (1962) and Schwartz (1964)) associated with

$$\delta_m(\mathbf{x}, \mathbf{y}) = \textstyle\sum_{\mathbf{k}=\mathbf{o}}^{\mathbf{m}} h_{\mathbf{k}}(\mathbf{x}) h_{\mathbf{k}}(\mathbf{y}) \tag{3.2}$$

where $\mathbf{m} = (m, \ldots, m)$. By using an inequality in Walter (1977), the bias rate can be shown to be $O(m^{d-p-1})$ for f such that $(\triangle - |\mathbf{x}|^2)^p f \in L^2(R^d)$.

4. Rates of convergence results in density estimation problem. The intent of this section is to apply the results of Sections 2 and 3 to the multivariate density estimation problem using i.i.d. random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ which have the common density f with respect to Lebesgue measure on R^d. Our main theorems concern the rates of convergence of $M(\mathbf{x}) = E[(\hat{f}(\mathbf{x}) - f(\mathbf{x}))^2]$ to zero where

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \textstyle\sum_{j=1}^{n} \delta_m(\mathbf{x}_j, \mathbf{x}). \tag{4.1}$$

Here $\{\delta_m\}$ with $m = m(n)$ is a delta sequence satisfying the conditions either of Sections 2 or of 3. The proofs of various parts of the theorem will follow the usual method of obtaining rates for $\mathrm{Var}(\hat{f}(\mathbf{x}))$, and for $\hat{B}(\mathbf{x}) = (E[\hat{f}(\mathbf{x})] - f(\mathbf{x}))^2$ separately, and then choosing the parameter m in δ_m so that these two rates are equal. The common rate would then be our rate result for $M(\mathbf{x})$.

THEOREM 4.1. *If $\{\delta_m\}$ is a positive type delta sequence, and $m = o(n^{1/d})$, then $M(\mathbf{x}) \to 0$ a.e. Lebesgue $(\mathbf{x})$.*

THEOREM 4.2. *Let $\{\delta_m\}$ be a positive type delta sequence, $f \in L^p(R^d)$, $p > 1$, and satisfy a Lipschitz condition of order γ at $\mathbf{x}$; if $\delta = \min\{\gamma, 1 - p^{-1}\}$, then for $m = [n^{1/(d+2\delta)}]$, $M(\mathbf{x}) = O(n^{-2\delta/(d+2\delta)})$.*

THEOREM 4.3. *Let f be a rapidly decreasing function R^d and let $\{\delta_m\}$ be given by (3.1) or (3.2). Then for each $r \geq d + 1$ and $m = [n^{1/2r}]$, $\| M \|_\infty = O(n^{-1+d/2r})$.*

Note that for rapidly decreasing functions (i.e., $f \in C^\infty$ such that $|\mathbf{x}|^r f^{(s)}(\mathbf{x})$ is bounded for all $r \geq 0$ and $s = (s_1, \ldots, s_d)$), both the condition that $(1 + |\mathbf{x}|^p)\, \tilde{f} \in L^2$ and the condition $(\triangle - |\mathbf{x}|^2)^p f \in L^2$ are satisfied for all p, where $\tilde{f}$ is the Fourier transform of f.

REMARK 4.1. The rate in Theorem 4.2 can *not* be better than $O(n^{-2/(2+d)})$ while the rates in Theorem 4.3 approach $O(n^{-1})$ as $r \to \infty$. In the case of Theorem 4.1 or 4.2, we can use the immediately verifiable fact that $n\, \mathrm{Var}(\hat{f}(\mathbf{x})) = O(\| \delta_m(\mathbf{x}, \cdot) \|_\infty)$ (since $E[\hat{f}(\mathbf{x})] \to f(\mathbf{x})$ and $\delta_m \geq 0$) while we used the slightly weaker rate $O(\| \delta_m^2(\mathbf{x}, \cdot) \|_\infty)$ for $n\, \mathrm{Var}(\hat{f}(\mathbf{x}))$ in Theorem 4.3. Since this is the essence of all the proofs of available results similar to Theorems 4.1 and 4.2 (for example, Cacoullos (1966), Deheuvels (1977), Földes (1974), Földes and Révész (1974), etc.) and since the rates for the bias depend on Lemma 2.1, the contents of Remark 2.1 apply here as well.

We now obtain rates for $\| \hat{f}(\mathbf{x}) - E[\hat{f}(\mathbf{x})] \|_\infty \to_{\text{a.s.}} 0$ (a.s. stands for almost surely) and $\| \hat{f}(\mathbf{x}) - f(\mathbf{x}) \|_\infty \to_{\text{a.s.}} 0$. We derive the rates for the first convergence as a corollary to Theorem 2 of Kiefer (1961) concerning the L_∞ behavior of the empiric distribution less its expectation while the second result follows as a corollary to the first convergence result and rates obtained for the bias of $\hat{f}_n(\mathbf{x})$ in Sections 2 and 3. As in Theorems 4.1, 4.2 and 4.3, we state the rate results for $\| \hat{f}(\mathbf{x}) - E[\hat{f}(\mathbf{x})] \|_\infty \to_{\text{a.s.}} 0$ for various delta sequences satisfying the additional condition

$$\left\| \frac{\partial^d}{\partial t_1, \ldots, \partial t_d} \delta_m(\mathbf{x}, \mathbf{t}) \right\|_1 = O(m^d) \tag{4.2}$$

where $m = m(n) \to \infty$ and m could depend on $\mathbf{x}$. The integration by parts used in the proof below has been used by Nadaraya (1965) in the case when $d = 1$, and $\{\delta_m\}$ is the sequence generated by kernel extimators.

THEOREM 4.4. *Let* $\{\delta_m\}$ *be a delta sequence satisfying* (4.2.) Then

$$| \hat{f}(\mathbf{x}) - E[\hat{f}(\mathbf{x})] | = O(m^d \sqrt{lln/n}) \qquad \text{a.s.}$$

PROOF. With $\hat{F}_n$ denoting the empiric distribution function of $\mathbf{x}_1, \ldots, \mathbf{x}_n$ whose common distribution function is F, it can be seen that

$$\hat{f}(\mathbf{x}) - E[\hat{f}(\mathbf{x})] = \int_{R^d} \delta_m(\mathbf{x}, \mathbf{t}) \, d(\hat{F}_n(\mathbf{t}) - F(\mathbf{t})). \tag{4.3}$$

Integration by parts followed by an application of (4.2) shows that

$$\begin{aligned} lhs \text{ of } (4.3) \,| &< \sup\left\{ | \hat{F}_n(\mathbf{t}) - F(\mathbf{t}) | \;\middle|\; \mathbf{t} \in R^d \right\} \times \left\| \frac{\partial^d}{\partial t_1, \ldots, \partial t_d} \delta_m(\mathbf{x}, \mathbf{t}) \right\|_1 \\ &= \| \hat{F}_n - F \|_\infty O(m^d) \qquad \text{a.s.} \end{aligned} \tag{4.4}$$

But Theorem 2 of Kiefer (1961) shows that the first factor on the right-hand side of (4.4) is a.s. $O(lln/n)^{1/2}$ thereby completing the proof.

In order to obtain rates for $\hat{B}(\mathbf{x}) = \hat{f}(\mathbf{x}) - f(\mathbf{x}) \to_{\text{a.s.}} 0$, we use the above theorem and the rates obtained for $B(\mathbf{x}) = E[\hat{f}(\mathbf{x})] - f(\mathbf{x})$, which is the bias of the estimator $\hat{f}(\mathbf{x})$ of $f(\mathbf{x})$. The latter are obtainable from the results of Section 2 just as in the proof of Theorem 4.2.

COROLLARY 4.1. *If* (4.2) *holds and* $m = O(n^{1/ad})$ *for an* $a > 1/2$, *then* $\hat{B}(\mathbf{x}) \to_{\text{a.s.}} 0$ a.e. *Lesbegue* ($\mathbf{x}$).

COROLLARY 4.2. *Let* (4.2) *hold. Let the conditions of Theorem* 4.2 *hold. Then* $\hat{B}(\mathbf{x}) = O(lln)^{1/2}/n^{\delta/2(\delta+d)}$ a.s. *with* $m = [n^{(1+2\delta/2(d+\delta)}]$.

The proofs of the above corollaries are not hard. For example, to obtain Corollary 4.2, observe that $| \hat{B}(\mathbf{x}) | \le | \hat{f}(\mathbf{x}) - E[\hat{f}(\mathbf{x})] | + | B(\mathbf{x}) | \le c\{n^{-1/2} m^d (lln)^{1/2} + \mathrm{m}^{-\delta}\}$ a.s. for some constant c.

REMARK 4.2. Similar results hold for the delta sequences given in Examples 3.1 and 3.2. However, they do not necessarily satisfy (4.2), although similar conditions, different for each example, can be shown to hold. We compare the above result to some of the available results in this direction. Both Földes (1974), and Földes and Révész (1974), obtain exponential upper bounds for $P[| \hat{f}(\mathbf{x}) - f(\mathbf{x}) | \ge \epsilon]$ ($\epsilon > 0$) for $d = 1$ case when $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are i.i.d. and in the case of dependent (but stationary) $\mathbf{x}_1, \ldots, \mathbf{x}_n$ under several conditions including the finiteness of αth absolute moment of $\mathbf{x}_1$ and a uniform Lipschitz condition on delta sequence $\{\delta_m\}$. Moreover, it is hard to recover a rate from their theorem 2 because all the null sequences hidden in their results (Theorem 2 of Földes and Révész or Theorem

1 in Földes) are also involved in a very complex way on the conditions of the theorems. We obtain our results (though weaker) under fairly minimal conditions on $\{\delta_m\}$. Moreover, inequality (4.4) can be used to obtain an exponential upper bond for $P[|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \geq \epsilon]$ by using Theorem 2 of Kiefer (1961). On the other hand, Bleuez and Bosq (1976a, b) appear to have stronger convergence rates under easily verifiable conditions, but for special cases only. Bertrand-Retali (1974) provides necessary and sufficient conditions for a.s. convergence of histogram type density estimators. Up graded versions of Theorems 4.1, 4.2, and 4.3, and Corollaries 4.1 and 4.2 to obtain uniform (in $\mathbf{x}$) conclusions can be used to estimate the mode of a multivariate density. (For a recent paper in this direction, see Sager (1978).)

5. Asymptotic normality of $\hat{\mathbf{f}}$ of (4.1). Since $\hat{f}$, defined by (4.1), is a sum of n i.i.d. random variables for each fixed n, the Basic Lemma (page 277, Loéve (1960)) can be used to show that $\{\hat{f}(\mathbf{x}) - E[\hat{f}(\mathbf{x})]\}/\text{S.d.}(\hat{f}(\mathbf{x}))$ converges in law ($\rightarrow_{\mathscr{L}}$) to the standard normal distribution ($N(0, 1)$). There are several variations of the asymptotic normality result for $\hat{f}$ depending upon the conditions of the delta sequence $\{\delta_m\}$, and the unknown density f. We shall assume that $\{\delta_m\}$ has most of its mass concentrated near the set where $\mathbf{x} = \mathbf{t}$ leading to

DEFINITION 5.1. A positive delta sequence $\{\delta_m\}$ (see Definition 2.2) is said to be *regular* if $\exists$ an interval J in $\mathscr{R}^d$ containing $\mathbf{0}$ and a constant C such that

$$\delta_m(\mathbf{x}, \mathbf{t}) \geq C\, m^d I_J(m(\mathbf{x} - \mathbf{t})) \qquad \text{a.e. (Lebesgue).} \tag{5.1}$$

We then have the following basic asymptotic normality result.

THEOREM 5.1. *Let $\{\delta_m\}$ be regular and let $m = O(n^{(1/d)-\epsilon})$ for some $0 < d\epsilon < 1$. Then $\{\hat{f}(\mathbf{x}) - E[\hat{f}(\mathbf{x})]\}/\text{s.d.}(\hat{f}(\mathbf{x})) \rightarrow N(0, 1)$ for $\mathbf{x}$ such that $f(\mathbf{x}) > 0$.*

PROOF. By the above quoted result of Loéve (1960), it suffices to show that $E[|\delta_m(\mathbf{x}, \mathbf{X}_1)] - E[\delta_m(\mathbf{x}, \mathbf{X}_1)]|^3)/\sqrt{n}\ \text{s.d.}^3(\delta_m(\mathbf{x}, \mathbf{X}_1)) \rightarrow 0$. But this ratio is bounded by $n^{-1/2}\|\delta_m(\mathbf{x}, \cdot)\|_\infty/\text{s.d.}(\delta_m(\mathbf{x}, \mathbf{X}_1)) = (= L_n$, say, since $m = m(n) \rightarrow \infty$ as $n \rightarrow \infty)$. Now $L_n \rightarrow 0$ provided $\limsup n^\alpha L_n < \infty$ for some $1 > 2\alpha > 0$, which will be shown below. Observe that

$$\limsup n^\alpha L_n = 1/\liminf \{\text{s.d.}(\delta_m(\mathbf{x}, \mathbf{X}_1))/\|\delta_m(\mathbf{x}, \cdot)\|_\infty n^{(1/2-\alpha)}\},$$

and

$$\frac{\liminf (\text{s.d.}(\delta_m(\mathbf{x}, \mathbf{X}_1)))^2 n^{1-2\alpha}}{\|\delta_m(\mathbf{x}, \cdot)\|_\infty^2} \cong \frac{\liminf n^{1-2\alpha} \int \delta_m^2(\mathbf{x}, \mathbf{t}) f(\mathbf{t})\, dt}{\|\delta_m(\mathbf{x}, \cdot)\|_\infty^2} \tag{5.2}$$

provided $n^{1-2\alpha}/m^{2d} \rightarrow 0$, since $E[\delta_m(\mathbf{x}, \mathbf{X}_1)] \rightarrow f(\mathbf{x})$ by (i) of Lemma 2.1, and $\|\delta_m(\mathbf{x}, \cdot)\|_\infty = O(m^d)$. The proof will be complete if the right-hand side of (5.2) $\geq c > 0$ for some constant c. Since $\{\delta_m\}$ is regular,

$$\int \delta_m^2(\mathbf{x}, \mathbf{t}) f(\mathbf{t})\, d\mathbf{t} \geq Cm^{2d} \int I_J(m(\mathbf{x} - \mathbf{t}) f(\mathbf{t})\, d\mathbf{t}$$

$$= Cm^d \int_J f(\mathbf{x} - \mathbf{s}/m)\, ds$$

$$= Cm^d \int_J (f(\mathbf{x} - \mathbf{s}/m) - f(\mathbf{x}))\, d\mathbf{s} + Cm^d \int_J f(\mathbf{x})\, d\mathbf{s}$$

$$= C_1 m^d (f(\mathbf{x}) - \epsilon_1)$$

for $m \geq m_0$ and an $0 < \epsilon_1 < f(\mathbf{x})$. Hence the right-hand side of (5.2) $\geq C_2 n^{1-2\alpha} m^{-d}$. If $m = [n^{(1-2\alpha)/d}]$, then $n^{1-2\alpha} m^{-d} \geq 1$ for $\alpha = d\epsilon/2 (< \frac{1}{2})$ and the proof is complete.

REMARK 5.1. We now compare the conditions of Theorem 5.1 specialized to the d = *1 case* to the conditions (H1), (H2), and (H3) (among others) of Delecroix (1977) specialized to the independent case. Under (5.1) and $m = O(n^{1-\epsilon})$, it can be seen that (H1) and (H3) are obviously satisfied. Moreover, as noted in the above proof, Lemma 2.1 shows that even (H2) is satisfied. Also, $\alpha(i) = 0$ for $i = 1, 2, \cdots$ in Delecroix's theorem in the independent case. One can replace (5.1) (slightly stronger than is necessary) obviously by the condition $\int \delta_m^2(\mathbf{x}, \mathbf{t}) f(\mathbf{t})\, d\mathbf{t} \geq cm^\alpha$ which is a generalization of (H3) to delta sequences. Finally, our theorem can be generalized to the α-mixing case by following the method of proof of Theorem 1 of Delecroix (1977).

REMARK 5.2. Both the mean square consistency result, and the asymptotic normality result, hold under (5.1). $E[\hat{f}(\mathbf{x})]$ can be replaced by $f(\mathbf{x})$ in the above theorem if $n^\alpha(E[\hat{f}(\mathbf{x})] - f(\mathbf{x}))/\| \delta_m(\mathbf{x}, \cdot) \|_\infty \to 0$ which is implied by a Lipschitz condition on f at $\mathbf{x}$.

A CONCLUDING REMARK. In the problem of estimation of a regression function, we have verified that an appropriate modification of the density estimators based on delta sequences of Section 2 satisfies the sufficient conditions on the sequence of weights occurring in Theorem 1 of Stone (1977). Consequently, we note here that the density estimators proposed here can be adapted to obtain estimators for the regression function, and such regression function estimators can be shown to have mean square errors going to zero with a rate. As in Prakasa Rao (1978), the results obtained here can be adopted to obtain sequential multivariate density estimators.

Acknowledgements. The authors wish to thank the referee for a number of helpful suggestions and for pointing out references overlooked in the original draft of the paper.

REFERENCES

BLEUEZ, J. and BOSQ, D. (1976a). Conditions nécessaires et suffisantes de convergence pour une classe d'estimateurs de la densité. *C.R. Acad. Sci. Paris Ser.* A **282** 63–66.

BLEUEZ, J. (1976b). Conditions nécessaires et suffisantes de convergence de l'estimateur de la densité par la méthode des fonctions orthogonales *C.R. Acad Sci. Paris Ser.* A **282** 1023–1026.

BLUM, J. and SUSARLA, V. (1977). A Fourier inversion method for the estimation of a density and its derivative. *J. Austrail. Math. Soc.* **22** 166–171.

BOSQ, D. (1977). Study of a class of density estimators. In *Recent Developments in Statistics* (J.R. Barra, et al., eds.) North Holland, Amsterdam.

CACOULLOS, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math.* **18** 179–189.

ČENCOV, N. (1962). Evaluation of an unknown distribution density from observations. (Russian) *Dokl. Akad. Nauk. SSSR* **147** 45–48. (Translated as: *Soviet Math. Dokl.* **3** 1559–1562.)

DEHEUVELS, P. (1977). Estimation non parametrique de la densite par histrogrammes generalises (II). *Publ. Inst. Univ. Paris.* **22** 1–23.

DELECROIX, M. (1977). Central limit theorems for density estimators of a L_2-mixing process. In *Recent Developments in Statistics* (J.R. Barra, et al., eds.) North Holland; Amsterdam.

FÖLDES, A. (1974). Density estimation for dependent sample. *Studia Sci. Math. Hungar.* **9** 81–92.

FÖLDES, A. and RÉVÉSZ, P. (1974). A general method for density estimation. *Studia Sci. Math. Hungar.* **9** 443–452.

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

KIEFER, J. (1961). On large deviations of the empiric D.F. of vector chance variables and a law of the iterated logarithm. *Pacific J. of Math.* **11** 649-660.

LOÉVE, M. (1963). *Probability Theory.* (3rd ed.) Van Nostrand, Princeton.

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

PARZEN, E. (1962). On estimation of a probability density and mode. *Ann. Math. Statist.* **35** 1065-1076.

PRAKAS, RAO, B. L. S. (1978). Sequential nonparametric estimation of density via delta-sequences. To appear in Sankhyā.

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

SAGER, T. (1978). Estimation of a multivariate mode. *Ann. Statist* **6** 802-812.

SCHWARTZ, S. C. (1964). Estimation of a probability density by an orthogonal series. *Ann. Math. Statist.* **38** 1261-1265.

STEIN, E. and WEISS, G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton Univ.

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

TITCHMARSH, E. C. (1958). *Eigenfunction Expansions Associated with Second-order Differential Equations.* Clarendon, Oxford.

VAN RYZIN, J. (1966). Bayes risk consistency of classification procedures using density estimation. *Sankhyā Ser. A* **28** 261-270.

WAHBA, G. (1971). A polynomial algorithm for density estimation. *Ann. Math. Statist.* **42** 1870-1886.

WALTER, G. (1977). Properties of Hermite Series estimation of probability density. *Ann. Statist.* **5** 1258-1264.

WALTER, G. and BLUM, J. (1979). Probability density estimation using delta sequences. *Ann. Statist.* **7** 328-340.

WATSON, G. and LEADBETTER, M. (1965). Hazard analysis II. *Sankhyā Ser. A.* **26** 101-116.

Wegman, E. J. (1971). Nonparametric probability density estimation. I. A summary of available methods. *Technometrics* **14** 533-546.

WEGMAN, E. J. (1972). Nonparametric probability density estimation. II. A comparison of density estimation methods. *J. Statist. Comp. and Simulation.* **1** 225-246.

WERTZ, W. (1978). *Statistical Density Estimation, A Survey.* Van den Hoeck and Ruprecht.

WERTZ, W. and SCHNEIDER, B. (1979). Statistical density estimation: a bibliography. *Intern. Statist. Rev.* **47** 155-175.

WINTER, B. (1975). Rate of strong consistency of two nonparametric density estimators. *Ann. Statist.* **3** 759-766.

DEPARTMENT OF STATISTICS AND PROBABILITY
WELLS HALL
MICHIGAN STATE UNIVERSITY
EAST LANSING, MICHIGAN 48824

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF WISCONSIN-MILWAUKEE
P.O. BOX 413
MILWAUKEE, WISCONSIN 53201