The Annals of Statistics
1979, Vol. 7, No. 2, 328–340

PROBABILITY DENSITY ESTIMATION USING DELTA SEQUENCES

BY G. WALTER AND J. BLUM

University of Wisconsin-Milwaukee and University of Arizona

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with common density function f. A method of density estimation based on "delta sequences" is studied and mean square rates established. This method generalizes certain others including kernel estimators, orthogonal series estimators, Fourier transform estimators, and the histogram. Rates are obtained for densities in Sobolev spaces and for densities satisfying Lipschitz conditions. The former generalizes some results of Wahba who also showed the rates obtained are the best possible. The rates obtained in the latter case have been shown to be the best possible by Farrell. This is shown independently by giving examples for which the rates are exact. Finally, a necessary and sufficient condition for asymptotic unbiasedness for continuous densities is given.

1. Introduction. Many methods for the estimation of an unknown density function f by means of functions of i.i.d. random variables $X_1, X_2, \cdots, X_n \cdots$ have been proposed in recent years. These include the kernel methods studied by Rosenblatt (1956), Parzen (1962), Schuster (1969), Singh (1974) and Susarla and Kumar (1975); the orthogonal series methods used by Schwartz (1967), Kronmal and Tarter (1968) and Walter (1976); the interpolation methods considered by Van Ryzin (1973) and Wahba (1971), (1975); and the characteristic function approach studied by Blum and Susarla (1976). Farrell (1972) and Wahba (1975) have shown that in certain cases these estimates are best possible.

In this paper we study a simple method which generalizes many of the methods mentioned above. It is based on the use of "delta sequences," i.e., sequences of functions that converge to the generalized function δ.

Certain types of these sequences were already studied by Watson and Leadbetter (1964) who called them "δ-function sequences." They established, among other things, the asymptotic unbiasedness and the asymptotic variance of estimators based on them but did not consider rates. Winter (1975) obtained the rate of strong consistency as well as the rate of asymptotic bias for estimators associated with delta sequences arising from the Fejér kernel of Fourier series. We shall consider more general classes of delta sequences and shall primarily concern ourselves with mean square rates of the estimators.

We shall consider several different types of delta sequence estimators with several examples of each. The first type is appropriate for global properties of densities and consists of delta sequences which converge at a certain rate in the dual of a Sobolev space $W_p^{(s)}$. This is the space of functions with $s-1$ absolutely continuous derivatives each of which, as well as the sth derivative, is in L^p. For

Received May 1976; revised March 1978.
AMS 1970 *subject classifications.* Primary 62F10, 41A25; secondary 46F99.
Key words and phrases. Density estimation, delta sequences, mean square convergence.

these estimators we calculate a mean square rate similar to that obtained by Wahba (1975) for densities in $W_p^{(s)}$. The other types are more appropriate for local properties. In that case the estimators are shown to have a mean square rate of convergence which depends on the Lipschitz continuity of the density at a single point.

Some of our results are sharp since they give rates which have been shown by Farrell (1972) and Wahba (1975) to be the best possible. We also present a simple example for which the rate obtained ($0(n^{-\frac{4}{5}})$) is exact. Finally, we prove a theorem which shows that the estimators which give the best rates of convergence are not asymptotically unbiased.

2. General delta-sequence estimators. In this section we consider delta sequences having only the property that they converge to the delta function. Without further restrictions we can only obtain very weak results. In subsequent sections we shall show that the examples presented here satisfy the further conditions needed for stronger results.

DEFINITION 1. Let J be an open interval of the real line R. A sequence $\{\delta_m(x, t)\}$ of bounded measurable functions on $J \times J$ is a *delta sequence* on J if for each $x \in J$ and each C^∞ function ϕ with support in J we have

$$\lim_{m\to\infty}\int_J \delta_m(x, t)\phi(t)dt = \phi(x).$$

Let $\{X_n\}$ be a sequence of i.i.d. random variables with density $f(x)$. We shall associate an estimator of f with the sequence $\delta_m(x, t)$ by letting

$$\hat{f}_{n,m}(x) = \frac{1}{n}\Sigma_{k=1}^{n}\delta_m(x, X_k). \tag{1}$$

Subsequently we shall adjust m and n so as to obtain a single sequence of estimators.

EXAMPLES.

(i) *A kernel estimator.* Let

$$\delta_m(x, t) = m\chi_{m^{-1}}(x - t) \qquad x, t \in R$$

where $\chi_{m^{-1}}$ is the indicator function of $[0, m^{-1}]$. This is the canonical example of a Parzen kernel estimator [5].

(ii) *Histogram estimator.* Let

$$\delta_m(x, t) = m\Sigma_{j=1}^{m}\chi_j(x)\chi_j(t) \qquad x, t \in (0, 1)$$

where χ_1 is the indicator function of $(0, 1/m)$, and χ_j is the indicator function of $[(j - 1/m), j/m)$ for $j = 2, \ldots, m$. This is easily seen to be the usual histogram estimator.

(iii) *Orthogonal series estimators.* Let $\{\psi_m(x)\}$ be a complete orthonormal system on (a, b) consisting of eigenfunctions of a compact operator on $L^2(a, b)$. Let

$$\delta_m(x, t) = \Sigma_{j=1}^{m}\psi_j(x)\psi_j(t), \qquad x, t \in (a, b).$$

These are delta sequences (see [15], page 500). They include trigonometric functions, Legendre polynomials, and Hermite functions. In the case of trigonometric functions we obtain the Dirichlet kernel. These include the sequences studied by Schwartz [8] and Walter [16], but not those studied by Kronmal and Tarter [4] or Winter [18].

We can also smooth the orthogonal functions as follows. Let $\{\alpha_{m,j}\}$ be the matrix corresponding to a regular summability method (see [19], page 74). Let

$$\delta_m(x, t) = \Sigma_{j=1}^{\infty}\alpha_{m,j}\Sigma_{k=1}^{j}\psi_k(x)\psi_k(t) \qquad x, t \in (a, b).$$

This is again a delta sequence. If the $\{\alpha_{m,j}\}$ corresponds to $C - 1$ summability and the $\{\psi_k\}$ are the cosine functions we have the sequence studied in [4]. If they are the usual trigonometric functions we have the sequence studied in [18].

(iv) *An interpolation estimator.* Let $\{\delta_m\}$ be a sequence of functions converging to the delta function and let $t_0, t_1, \cdots, t_m$ be $m + 1$ distinct points in (a, b). Finally, let $L_m(t) = (t - t_0)(t - t_1) \cdots (t - t_m)$. Then

$$\Delta_m(x, t) = \frac{L_m(x)}{(x - t)L_m(t)}\Sigma_{j=0}^{m}\delta_m(t - t_j), \qquad x, t \in (a, b)$$

is a delta sequence provided the $\{t_j\}$ are chosen, e.g., to be the zeroes of a Jacobi polynomial. This follows from the fact that the $\int\Delta_m(x, t)f(t)dt$ are approximations to the Lagrange interpolating polynomials which converge for such choices of the $\{t_j\}$.

(v) *Fourier transform estimators.* The estimator used by Blum and Susarla [1] involves the inverse Fourier transform of the delta of the characteristic function. It may be put in terms of the delta sequence

$$\delta_m(x, t) = \frac{1}{2\pi}\int_{-m}^{m} e^{is(x-t)}ds, \qquad x, t \in R.$$

As in the case (iii) we may smooth $\{\delta_m\}$ via a regular summability method. In particular, $C - 1$ summability yields

$$\delta_m(x, t) = \frac{2}{\pi}\frac{\sin^2 m(x - t)}{m(x - t)^2}, \qquad x, t \in R.$$

(vi) *An estimator based on Landau's sequence.* This provides a polynomial approximation which differs from orthogonal polynomials and interpolation. It is given by

$$\delta_m(x, t) = [1 - (t - x)^2]^m / \int_{-1}^{1}(1 - t^2)^m\, dt, \qquad , t \in (-1, 1)$$

and, as with the Fejér kernel of trigonometric series, is nonnegative.

(vii) *Estimators of density type.* Let $\{Y_m\}$ be a sequence of i.i.d. random variables with mean 0 and finite variance, having a bounded density. Let $g_m(y)$ be the density of the sample mean. Then the functions

$$\delta_m(x, t) = g_m(x - t), \qquad x, t \in R$$

may be shown to be a delta sequence by using the Chebyshev inequality.

3. Mean square rates based on global properties. In order to establish the mean square rate of convergence of our proposed estimator we must use some sort of rate of convergence of the delta sequence itself. The type of convergence given in Definition 1 can be shown to be characterized by the L^p convergence of appropriate antiderivatives (see [3], for example). We shall use this type of convergence to define the rate.

DEFINITION 2. Let $\{\delta_m(x, t)\}$ be a delta sequence on $J = (a, b)$ such that
(i) $\delta_m(x, \cdot) \in L^2(a, b)$ for $x \in J$;
(ii) $\|\delta_m(x, \cdot)\|_2^2 = 0(m)$ uniformly for $x \in J$.
Let s be a positive integer; denote by $\delta_m^{(-s)}(x, t)$ the antiderivative of order s with respect to t of $\delta_m(x, t)$ which, together with its first $(s - 1)$ derivatives, is zero at a; denote by $\delta_x^{(-s)}(t)(= \delta^{(-s)}(x - t))$ the function $(x - t)_+^{s-1}/(s - 1)!$. Suppose for such an s there exist numbers $q \geqslant 1$ and $0 < \beta \leqslant s + (1/q) - 1$ such that
(iii) $\delta_m^{(-s)}(x, \cdot) - \delta_x^{(-s)}(\cdot) \in L^q(a, b)$ for $x \in J$;
(iv) $\|\delta_m^{(-s)}(x, \cdot) - \delta_x^{(-s)}(\cdot)\|_q = 0(m^{-\beta})$ uniformly for $x \in J$;
(v) $|\delta_m^{(-k)}(x, b) - \delta_x^{(-k)}(b)| = 0(m^{-\beta})$ uniformly for $x \in J$ and for $k = 1, \cdots, s$.
Then $\{\delta_m\}$ is said to have (s, q) rate $m^{-\beta}$.

NOTE. Condition (ii) is needed to standardize the rate of increase of m, since a subsequence of a delta sequence is again a delta sequence.

THEOREM 1. *Let $s \geqslant 1$, and let f be a density with $f \in W_p^{(s)}(J)$. Let $\{\delta_m\}$ be a delta sequence of (s, q) rate $m^{-\beta}$, where q satisfies $(1/p) + (1/q) = 1$. Then*

$$E[\hat{f}_n(x) - f(x)]^2 = 0(n^{-1+1/(1+2\beta)}) \tag{2}$$

uniformly on J, where $\hat{f}_n$ is defined to be $\hat{f}_{n,m}$ of equation (1) *with $m = [n^{1/(1+2\beta)}]$.*

In order to prove the theorem we break up the mean square error into the usual variance and bias terms. We have

$$E[\hat{f}_{n,m}(x) - f(x)]^2 = \frac{1}{n}\left[\int_J \delta_m^2(x, t)f(t)\, dt - \left(\int_J \delta_m(x, t)f(t)\, dt\right)^2\right] + \left[\int_J \delta_m(x, t)f(t)\, dt - f(x)\right]^2. \tag{3}$$

Now $\int_J \delta_m^2(x, t)f(t)\, dt \leqslant \|f\|_\infty \|\delta_m(x, \cdot)\|_2^2 = 0(m)$, where $\|f\|_\infty$ is finite since $|f^p(x) - f^p(a)| \leqslant p\int_a^x f(x)^{p-1}|f'(x)|\, dx \leqslant p\|f\|_p^{p/q}\|f'\|_p$. Similarly

$$\left(\int_J \delta_m(x, t)f(t)\, dt\right)^2 \leqslant \|f\|_2^2 \|\delta_m(x, \cdot)\|_2^2 = 0(m).$$

Note that $\|f\|_2^2 < \infty$ since $f \in L^\infty(J) \cap L^1(J)$.

To estimate the bias term we integrate by parts s times. We obtain

$$\begin{aligned} \left|\int_J \delta_m(x, t)f(t)\, dt - f(x)\right| &= \left|\int_J [\delta_m(x, t) - \delta(x - t)]f(t)\, dt\right| \\ &\leqslant \left|\int_J [\delta_m^{(-s)}(x, t) - \delta^{(-s)}(x - t)]f^{(s)}(t)\, dt\right| + 0(m^{-\beta}) \\ &\leqslant \|\delta_m^{(-s)}(x, \cdot) - \delta_x^{(-s)}(\cdot)\|_q \|f^{(s)}\|_p + 0(m^{-\beta}), \end{aligned} \tag{4}$$

by Hölder's inequality. By hypothesis the last sum is again $0(m^{-\beta})$. Combining the two estimates we obtain $E[\hat{f}_{m,n}(x) - f(x)]^2 = 0(m/n) + 0(m^{-2\beta})$. By choosing $m = [n^{1/(1+2\beta)}]$ we obtain the desired result.

REMARK (i). The rates obtained in this theorem are the same as those obtained by Wahba [13] for certain kernel, orthogonal series, histogram and other methods. Thus the theorem may be viewed as a partial extension and unification of the result of Sections 4-6 of [13].

REMARK (ii). According to Wahba [13] again, a rate of $n^{-2\beta/(1+2\beta)-\varepsilon}$ cannot be achieved uniformly over bounded sets in $W_p^{(s)}$ for any $\varepsilon > 0$ at any point $x \in R$. Although our theorem does not state that the convergence is uniform over such bounded sets this can easily be seen to be true since each occurrence of f is dominated by the Sobolev norm. Hence the rates are best possible in these cases as well.

REMARK (iii). The requirement that $\beta \leqslant s + (1/q) - 1$ arises from the same considerations. If $\beta > s + (1/q) - 1$, a rate which contradicts the result of [13] would be obtained. We shall see that all of our examples satisfy this requirement.

This theorem can be applied directly to some of the examples in Section 2. The first of these, the rectangular delta sequence of (i) with $J = R$ and $q \geqslant 1$ satisfies

$$\|\delta_m^{(-1)} - \delta^{(-1)}\|_q \leqslant Cm^{-\frac{1}{q}}$$

and therefore has $(1, q)$ rate $m^{-1/q}$. For $s > 1$ no antiderivative is in L^q for $q \geqslant 1$. By the theorem the error is $0(n^{-2/(2+q)})$.

For some of the other examples the calculation of the rate is somewhat more complicated but can be simplified by considering dominating (Parzen) kernels. These kernels are themselves delta sequences for which we calculate the rate.

PROPOSITION 1. *Let $K(x)$ be a nonnegative function in $L^\infty(R)$ satisfying*

(a) $\int K = 1$

(b) $\int |s| K(s)\, ds < \infty$ *for* $q = 1$ *or* (b′) $\int (\log_+ |s|) |s|^q K^q(s)\, ds < \infty$ *for* $q > 1$.

Then the delta sequence given by

$$\delta_m(x, t) = mK(m(x - t))$$

has $(1, q)$ *rate* $m^{-1/q}$.

These are clearly delta sequences which satisfy the first and third condition of the definition. The second condition follows from the inequality

$$(5) \quad \int_{-\infty}^{\infty} \left| \int_{-\infty}^{x} mK(mt)dt - 1_+(x) \right| dx$$
$$= \int_{-\infty}^{0} \int_{-\infty}^{mx} K(s) ds\, dx + \int_0^\infty \int_{mx}^\infty K(s) ds\, dx$$
$$= -\int_{-\infty}^{0} \frac{t}{m} K(t) dt + \int_0^\infty \frac{t}{m} K(t) dt$$
$$\leqslant \frac{1}{m} \int_{-\infty}^{\infty} |t| K(t) dt$$

for $q = 1$, and similarly

$$(6)\qquad \int_0^\infty \left|\int_{mx}^\infty K\right|^q dx = \frac{1}{m}\left\{\int_0^1\left|\int_t^\infty K\right|^q dt + \int_1^\infty\left|\int_t^\infty K\right|^q\right\}$$
$$\leq \frac{1}{m} + \frac{1}{m}\int_1^\infty\left|\int_t^\infty sK(s)\frac{1}{s}\,ds\right|^q dt$$
$$\leq \frac{1}{m} + \frac{1}{m}\int_1^\infty\left\{\int_t^\infty |sK(s)|^q ds\left|\int_t^\infty \frac{1}{s^p}\right|^{\frac{q}{p}} ds\right\} dt$$
$$\leq \frac{1}{m} + \frac{C}{m}\int_1^\infty\int_1^s t^{(1-p)q/p}\,dt\, s^qK^q(s)\,ds$$
$$\leq \frac{1}{m} + \frac{C}{m}\int_1^\infty \log s\, s^qK^q(s)\,ds$$

for $q > 1$.

DEFINITION 3. A delta sequence $\{\delta_m\}$ is *dominated* by a kernel K if for $x, t \in R$

(a) $\delta_m(x, t) \geq 0$;

(b) $\int_{-\infty}^{\infty}\delta_m(x, t)dt = 1$;

(c) $\delta_m(x, t) \leq CmK(m(t - x))$;

where K satisfies condition (a) and (b) of Proposition 1.

COROLLARY 1. *Let $\{\delta_m\}$ be a dominated delta sequence on R whose dominating kernel satisfies the hypothesis of Proposition 1 for some $q \geq 1$; then $\{\delta_m\}$ has $(1, q)$ rate $m^{-1/q}$.*

That the rate of such delta sequences is the same as the corresponding rate for the kernels, follows from the fact that

$$(7)\qquad \int_{-\infty}^{\infty}\left|\int_{-\infty}^{s+x}\delta_m(x, t)dt - 1_+(s)\right|ds$$
$$= \int_{-\infty}^0\int_{-\infty}^{s+x}\delta_m(x, t)\,dt\,ds + \int_0^\infty\int_{s+x}^\infty\delta_m(x, t)\,dt\,ds$$
$$\leq C\int_{-\infty}^0\int_{-\infty}^{s+x}mK(m(t - x))\,dt\,ds + C\int_0^\infty\int_{s+x}^\infty mK(m(t - x))\,dt\,ds$$
$$= Cm\left\{\int_{-\infty}^0\int_{-\infty}^s K(mt)\,dt\,ds + \int_0^\infty\int_s^\infty K(mt)\,dt\,ds\right\}$$

in the case K satisfies (b) and a similar result if it satisfies (b′).

If $\{\delta_m\}$ is defined on a proper subinterval J^2 of R^2, we may extend it to all of R^2 by the simple expedient of making it zero outside of J^2. Thus this result could apply to many of the examples given in Section 2.

Example (i) was already a kernel. Example (ii), the histogram, is dominated by the kernel in example (i). Example (iii) is not positive in general, but when the $\{\phi_j\}$ are the trigonometric functions and the summability method is $(C - 1)$, i.e., when we have the Fejér kernel, the associated dominating kernel is $1/\pi(1 + t^2)$. Example (v) is similar to example (ii); the $(C - 1)$ kernel of the Fourier transform is also dominated by $1/\pi(1 + t^2)$. Example (vi) is dominated by the same kernel provided m is replaced by m^2 so that the rate is properly normalized.

Hence each of these examples has (1, q) rate $m^{-1/q}$ for $q > 1$ and in the first two for $q = 1$ as well. The mean square rate of the associated estimator is, by Theorem 1, $0(n^{-2/(q+2)})$.

Delta sequence of densities (example (vii)) may be shown directly to satisfy the conditions of Definition 2. We have for condition (iv) and $s = q = 1$,

$$(8)\quad \int_{-\infty}^{\infty}\left|\int_{-\infty}^{x} g_m - 1_+(x)\right|dx$$
$$= \int_{-\infty}^{0}\int_{-\infty}^{x} g_m\, dx + \int_0^{\infty}\int_x^{\infty} g_m\, dx$$
$$= \int_0^{\infty}\left\{\int_{-\infty}^{-x} + \int_x^{\infty}\right\} g_m\, dx \leqslant \int_{m^{-\frac{1}{2}}}^{\infty}\frac{\sigma}{x^2 m}\, dx + \int_0^{m^{-\frac{1}{2}}} dx = 0(m^{-\frac{1}{2}})$$

by Chebyshev's inequality. Condition (ii) is shown to hold by using the characteristic functions. If $\phi(\omega)$ is the characteristic function of Y, then $\phi^m(\omega/m)$ is that of $\overline{Y}_m$. The density of $\overline{Y}_m$ is uniformly bounded by the L^1 norm of $\phi^m(\omega/m)$ which satisifies

$$\int_{-\infty}^{\infty}|\phi^m(\omega/m)|\,d\omega \leqslant m\int_{-\infty}^{\infty}|\phi^m(t)|\, dt$$
$$\leqslant m\|\phi\|_{\infty}^{m-2}\|\phi\|_2^2 \leqslant m\|g\|_2^2$$

provided that $g \in L^2$. Since g was bounded and is in L^1 it is also in L^2 and hence Plancherel's equality holds. The other conditions follow easily, and hence the sequence $\{\delta_m(x, y)\} = \{g_m(x - y)\}$ has (1, 1) rate $m^{-\frac{1}{2}}$.

For the Dirichlet delta sequence of example (iii)

$$(9)\quad \delta_m(t) = \frac{\sin(m + 1/2)(t)}{2\pi \sin 1/2t} \qquad -\pi < t < \pi$$
$$= 0 \qquad \text{otherwise}.$$

We have $\|\delta_m\|_\infty \leqslant (m + 1/2)/\pi$ and

$$\delta_m(t) = \frac{1}{\pi}\left\{\frac{1}{2} + \Sigma_{k=1}^{m}\cos kt\right\} = \frac{\sin(m + 1/2)t}{2\pi\sin 1/2t},$$

and hence that

$$\delta_m^{(-1)}(x) = \frac{1}{\pi}\int_{-\pi}^{x}\left\{\frac{1}{2} + \Sigma_{k=1}^{m}\cos kt\right\} dt = \frac{x + \pi}{2\pi} + \frac{1}{\pi}\Sigma_{k=1}^{m}\frac{\sin kx}{k}.$$

Thus

$$\delta_m^{(-1)}(x) - 1_+(x) = T(x) + \frac{1}{\pi}\Sigma_{k=1}^{m}\frac{\sin kx}{k}$$

where

$$T(x) = \frac{x + \pi}{2\pi} \qquad -\pi \leqslant x < 0$$
$$= \frac{x - \pi}{2\pi} \qquad 0 < x \leqslant \pi.$$

But $T(x)$ has Fourier series $-1/\pi\Sigma_{k=1}^{\infty}(\sin kx)/k$ whence it follows that

$$\left\|T + \frac{1}{\pi}\Sigma_{k=1}^{m}\frac{\sin kx}{k}\right\|_2^2 = \frac{1}{\pi}\Sigma_{k=m+1}^{\infty}\frac{1}{k^2} \leqslant \frac{1}{\pi}m^{-1}$$

and therefore that

$$\|\delta_m^{(-1)} - 1_+\|_2 \leqslant \pi^{-\frac{1}{2}} m^{-\frac{1}{2}}.$$

For higher order integrals we see that

$$\|\delta_m^{(-s)} - \delta^{(-s)}\|_2 = 0\left(m^{-s+\frac{1}{2}}\right) \tag{10}$$

by the same formula. Hence the $(s, 2)$ rate is $m^{-s+1/2}$ for any $s = 1, 2, \cdots$.

The same procedure may be used for other orthogonal series on bounded intervals and for the Fourier transform (see [15] for the former).

4. Local convergence rates. The results of Section 3 are global in the sense that the convergence is uniform and the densities are assumed to belong to a Sobolev space, $W_p^{(s)}$. We may also obtain pointwise results for a general class of delta sequences, which we shall call delta sequences of Fejér type and whose behavior approximates that of the Fejér kernel. In doing so we shall restrict ourselves to convergence at the point 0, and require only that our sequences be composed of functions of a single variable.

DEFINITION 4. A dominated delta sequence on R is said to be a *Fejér type* if its dominating kernel is $\pi^{-1}(1 + t^2)^{-1}$.

NOTE. By Corollary 1 such sequences have $(1, q)$ rate $m^{-1/q}$ for $q > 1$.

This condition is satisfied by the Fejér kernel, by Landau's delta sequence (vi), by any positive Parzen kernel with compact support, by the Fourier transform delta sequence, and by the histogram (ii).

THEOREM 2. *Let $f(x)$ be a bounded density which satisfies a Lipschitz condition of order λ, $0 < \lambda < 1$, at $x = 0$. Let $\{\delta_m\}$ be a delta sequence of Fejér type. Then the estimator given by $\hat{f}_n = \hat{f}_{mn}$ where $m = [n^{1/(1+2\lambda)}]$ and*

$$\hat{f}_{mn} = \frac{1}{n}\Sigma_{i=1}^n \delta_m(X_i) \tag{11}$$

where $X_1, X_2, \cdots$ are i.i.d. *random variables with density $f(x)$, satisfies*

$$E\left[\hat{f}_n - f(0)\right]^2 = 0\left(n^{-1+\frac{1}{2\lambda+1}}\right). \tag{12}$$

The proof emulates that of the convergence theorem for C-1 means of Fourier series ([19], page 90). The mean square error is given by

$$E\left[\hat{f}_n - f(0)\right]^2 = \frac{\int \delta_m^2 f - \{\int \delta_m f\}^2}{n} + \left[\int \delta_m f - f(0)\right]^2 \tag{13}$$

where $m = [n^{1/(1+2\lambda)}]$. By Definition 4, $\|\delta_m\|_\infty = 0(m)$, and hence the first term in (13) is dominated by Cm/n since $\int \delta_m f \leqslant \|f\|_\infty$. The proof of the fact that

$$\left|\int \delta_m f - f(0)\right| = 0(m^{-\lambda})$$

is similar to other proofs of the same result for the Fejér kernel. We split the integral into five parts and show that each satisfies the correct inequality.

(14) $$\int \delta_m f - f(0) = \int \delta_m(x)(f(x) - f(0))\, dx$$
$$= \left\{ \int_{-\infty}^{-\delta} + \int_{-\delta}^{-\frac{1}{m}} + \int_{-\frac{1}{m}}^{\frac{1}{m}} + \int_{\frac{1}{m}}^{\delta} + \int_{\delta}^{\infty} \right\} \delta_m(x)(f(x) - f(0))\, dx$$

where δ is chosen such that $|f(x) - f(0)| \leqslant M|x|^\lambda$ for $|x| \leqslant \delta$. Going from right to left we show first that

(15) $$\left|\int_\delta^\infty \delta_m(x)(f(x) - f(0))\, dx\right| \leqslant \int_\delta^\infty |\delta_m(x)|(|f(x)| + |f(0)|)\, dx$$
$$\leqslant \frac{C}{m}\|f\|_1 + C\int_{m\delta}^\infty \frac{dt}{1+t^2}|f(0)| = 0(m^{-1}).$$

For the next integral we have, for $m \geqslant 1/\delta$,

(16) $$\int_{1/m}^{\delta} |\delta_m(x)|\, |f(x) - f(0)|\, dx$$
$$\leqslant \frac{C}{m}\int_{1/m}^{\delta} \frac{|f(x) - f(0)|}{x^2}\, dx \leqslant \frac{C}{m} M \int_{1/m}^{\delta} \frac{x^\lambda}{x^2}\, dx$$
$$= \frac{CM}{m}\left(\frac{\delta^{\lambda-1}}{\lambda - 1} - \frac{m^{-(\lambda-1)}}{\lambda - 1}\right) = 0(m^{-1}) + 0(m^{-\lambda}).$$

The middle integral gives us no trouble either since

(17) $$\int_{-\frac{1}{m}}^{\frac{1}{m}} |\delta_m(x)|\, |f(x) - f(0)|\, dx \leqslant CmM\int_{-\frac{1}{m}}^{\frac{1}{m}} |x|^\lambda\, dx = 0(m^{-\lambda}).$$

The remaining integrals are treated the same as the first two. Hence the bias term satisfies the desired inequality and the conclusion follows.

This result has been shown by Farrell [2] to be the best possible. Indeed his class $C_{0\eta}$ with $\eta(x) = x^{1+\lambda}$ consists of those functions satisfying a uniform Lipschitz condition of order λ. He shows (Theorem 1.1) that the rate of mean square convergence for any estimators cannot be better than $n^{-1+1/(1+2\lambda)}$.

It might be observed that in both theorems the best mean square rate we have been able to obtain for those delta sequences of nonnegative functions is $0(n^{-\frac{2}{3}})$ in general. However, the kernel function $K(t) = \chi_{(-\frac{1}{2}, \frac{1}{2})}(t)$ has an associated delta sequence whose (2, 1) rate is m^{-2}. Hence the error for the estimator obtained from this positive delta sequence is at least as good as $0(n^{-\frac{4}{5}})$. Again Farrell makes the observation that this is the best possible rate for positive Parzen kernels. This also follows from Wahba [13] in the special case of positive kernels.

We shall present an independent proof of these results by constructing simple examples. However, we can only do so for delta sequences satisfying:

DEFINITION 5. Let $\{\delta_m\}$ be a dominated delta sequence of functions with common compact support. It is said to be *regular* if there exist a nonempty interval

(a, b), $a \leqslant 0 \leqslant b$, and a number $C > 0$ such that

$$\delta_m(t) \geqslant Cm\chi_{(a,b)}(mt) \qquad 0 < t < \infty$$

where $\chi_{(a,b)}$ is the indicator function of (a, b).

PROPOSITION 2. *Let* $\{\delta_m\}$ *be regular. Then* (i) *there exists a density* $f(x)$ *satisfying a uniform Lipschitz condition of order* λ, $0 < \lambda \leqslant 1$, *such that*

$$E\left[\hat{f}_{mn} - f(0)\right]^2 \geqslant C_1 mn^{-1} + C_2 m^{-2\lambda};$$

(ii) *there exists a density* $g \in C^\infty$ *such that*

$$E[\hat{g}_{mn} - g(0)]^2 \geqslant C_1 mn^{-1} + C_2 m^{-4};$$

(iii) *if the support of* δ_m *is on* $[0, \infty)$, *there exists a density* $h \in C^\infty$ *such that*

$$E\left[\hat{h}_{mn} - h(0)\right]^2 \geqslant C_1 mn^{-1} + C_2 m^{-2}$$

for some C_1, $C_2 > 0$.

Part (i) is weaker than Farrell's result since it applies only to certain delta sequences, while parts (ii) and (iii) are slight generalizations of his observations.

The proof is obtained by merely exhibiting an appropriate f, g, and h. The first may be shown to be $f(t) = \alpha t^\lambda + \beta$ for $0 \leqslant t < \gamma$ and $f(t) = \beta$ for $-\gamma < t < 0$, where $[-\gamma, \gamma]$ is the support of δ_m. The values of $f(t)$ outside of $[-\gamma, \gamma]$ may be adjusted in any way to make f a density and ensure as well that it satisfies the Lipschitz condition. Then we have, if $b > 0$,

$$(18) \qquad \left|\int \delta_m(t) f(t)\, dt - f(0)\right| = \left|\int_0^\gamma \delta_m(t) \alpha t^\lambda dt\right|$$

$$\geqslant C\alpha m \int_0^{b/m} t^\lambda dt = \frac{C\alpha b^{1+\lambda}}{1+\lambda} m^{-\lambda},$$

and similarly if $a < 0$.

For the second part we take $g(t) = \alpha t^2 + \beta$ again on the common support of the δ_m. A similar calculation yields a bias term of $0(m^{-2})$. For the third, we use $h(t) = \alpha t + \beta$.

For the Fejér delta sequence itself ($\delta_m(t) = (\sin^2 mt/2)/(2m \sin^2 t/2)$) the given rate is the best possible in a stronger sense. Unless f is a constant on the interval $(-\pi, \pi)$, the difference

$$\int f(x - t)\delta_m(t)\, dt - f(x)$$

cannot be $0(1/m)$ uniformly for $|x| < \pi$ (see [19], page 122). Hence there is no hope of coming up with any set of conditions that would improve the rate beyond $0(n^{-\frac{2}{3}})$ for this sequence.

5. An alternative theorem. We have seen in Section 4 that the mean square rate of convergence of estimators arising from regular delta sequences is limited by $0(n^{-\frac{4}{5}})$. This includes many of our examples and is true in particular for the Fejér sequence. However, we have also seen that the rate for the Dirichlet sequence estimator approaches $0(n^{-1})$ for sufficiently differentiable densities. Thus it appears that the latter and similar estimators ought to be a better choice. However, this advantage is deceptive since it no longer holds if the density is merely continuous. In fact, estimators based on the Dirichlet and similar delta sequences are not asymptotically unbiased for continuous densities while those based on delta sequences of Fejér type are. This is a consequence of the following result.

THEOREM 3. *Let $\{\delta_m\}$ be a sequence of piecewise continuous functions in $L^1(\mathbb{R})$ such that*

(i) $\int \delta_m = 1$

(ii) $\int_{|x|>\gamma} |\delta_m(x)|\, dx \to 0$ *as* $m \to \infty$ *for* $\gamma > 0$

(iii) $\delta_m(x) \to 0$ *uniformly in* $|x| \geqslant \gamma$ *for* $\gamma > 0$.

Then

(iv) $\int_{-\infty}^{\infty} |\delta_m(x)|\, dx \leqslant A$

if and only if

$$\int_{-\infty}^{\infty} g(x)\delta_m(x)\, dx \to g(0) \qquad \text{as} \quad m \to \infty$$

for each density g which is continuous at $x = 0$ (i.e., $E(\delta_m(X)) \to g(0)$ *where X is a random variable with density g*).

The sufficiency of condition (iv) has been proved in [17], page 104. In fact, the conditions (i)-(iv) are exactly those hypothesized there.

In order to prove the necessity, we use the Banach-Steinhaus theorem ([15], page 165) and consider the Banach space C_0 of continuous functions with compact support and $\| \ \|_\infty$ norm. We define a sequence $[T_m]$ of bounded linear functionals on C_0 by $T_m(g) = \int \delta_m g$. If (iv) is not satisfied, then there is a sequence of functions $\{g_m\}$ in C_0 such that $\|g_m\|_\infty = 1$, and

$$T_{m_k}(g_{m_k}) \to \infty \qquad \text{as} \quad k \to \infty,$$

for some increasing sequence $\{m_k\}$.

To show that this is the case, we first define g_m^* by

$$g_m^*(x) = \delta_m(x)/|\delta_m(x)|, \qquad \delta_m(x) \neq 0$$
$$= 0 \qquad \delta_m(x) = 0.$$

We then modify g_m^* by setting it equal to 0 for $|x|$ sufficiently large and by smoothing it at points of discontinuity to get g_m. We may do this in such a way that

$$|\int \delta_m g_m^* - \int \delta_m g_m| < \varepsilon$$

since δ_m has only finitely many discontinuities on each compact set. If (iv) is not satisfied there exists a sequence $\{m_k\}$ such that

$$\int \delta_{m_k} g_{m_k}^* = \int |\delta_{m_k}| \to \infty$$

which in turn implies that

$$T_{m_k}(g_{m_k}) = \int \delta_{m_k} g_{m_k} \to \infty.$$

Hence

$$\|T_{m_k}\| = \sup_{\|g\|=1} |T_{m_k} g| \geqslant T_{m_k}(g_{m_k}) \to \infty$$

and by the Banach-Steinhaus theorem, there exists a $g \in C_0$ such that

$$\sup_m |T_m(g)| = \infty.$$

That is, we have $\int \delta_m g \not\to g(0)$ for some continuous functions with compact support. We may suppose that $g(x) \geqslant 0$, by adding a constant to g, since $\int \delta_m = 1$. We then truncate the new g by multiplying by the indicator function of an interval $[-\gamma, \gamma]$ such that $\int_{-\gamma}^{\gamma} g = 1$. This gives us a density, continuous at 0, for which

$$\int \delta_m g \not\to g(0).$$

Hence the conclusion.

REMARK (i). This is a standard type of proof based on the Banach-Steinhaus theorem (uniform boundedness principle). See [19], page 298 for a typical application to Fourier series.

REMARK (ii). The conditions (i), (ii) and (iii) are satisfied for most common delta sequences, in particular all those of Fejér type. Those of Dirichlet type (see [2], page 323) can be modified (if necessary) to make them fit these conditions. The former satisfy condition (iv) as well; the latter usually do not. Hence the estimators based on the former are asymptotically unbiased for all continuous densities while those based on the latter are not. This seems to be a greater advantage than the improved rate associated with the latter.

REMARK (iii). There may exist a delta sequence which satisfies all four conditions and shares the property of the Dirichlet sequence that the $(s, 2)$ rate is $m^{-s+\frac{1}{2}}$ for each positive integer s. We have been unable to find one and conjecture that the two sets of conditions are contradictory.

REFERENCES

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

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

[3] KOREVAAR, J. (1971). *Mathematical Methods*. Academic Press, New York.

[4] KRONMAL, R. and TARTER, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. *J. Amer. Statist. Assoc.* **63** 925–952.

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

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

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

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

[9] SINGH, R. S. (1974). Estimation of derivatives of the average μ-densities and sequence-compound estimation in exponential families. RM-318, Depart. Statist. Probability, Michigan State Univ.

[10] SUSARLA, V. and KUMAR, S. (1975). Estimation of partial derivatives of the average of densities belonging to a family of densities. *J. Austral. Math. Soc.* **20** 230–241.

[11] Van Ryzin, J. (1973). A histogram method of density estimation. *Comm. Statist.* **2** 493–506.
[12] Wahba, G. (1971). A polynomial algorithm for density estimation. *Ann. Math. Statist.* **42** 1870–1886.
[13] Wahba, G. (1975). Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. *Ann. Statist.* **3** 15–29.
[14] Wahba, G. (1975) Interpolating spline methods for density estimation I. Equi-spaced knots. *Ann. Statist.* **3** 30–48.
[15] Walter, G. (1965). Expansions of distributions. *Trans. Amer. Math. Soc.* **116** 492–510.
[16] Walter, G. (1977). Properties of Hermite series estimates of probability densities. *Ann. Statist.* **5** 1258–1264.
[17] Watson, G. W. and Leadbetter, M. R. (1965). Hazard analysis II. *Sankhya* **26** 101–116.
[18] Winter, B. B. (1975). Rate of strong consistency of two nonparametric density estimators. *Ann. Statist.* **3** 759–766.
[19] Zygmund, A. (1959). *Trigonometric Series*. Cambridge Univ. Press.

Department of Mathematics
The University of Wisconsin-Milwaukee
Milwaukee, Wisconsin 53201

Committee on Statistics
The University of Arizona
Tucson, Arizona 85721