The Annals of Statistics
1975, Vol. 3, No. 3, 759–766

RATE OF STRONG CONSISTENCY OF TWO NONPARAMETRIC DENSITY ESTIMATORS

BY B. B. WINTER
University of Ottawa

Two nonparametric density estimators, based on Fourier series and the Fejér kernel, are presented. One of them ($\tilde{f}_N$) is appropriate when the unknown density f vanishes outside a known bounded interval; the other ($f_N^\#$) is applicable without any assumptions about the support of f. The estimator $f_N^\#$ is of the type studied by Watson and Leadbetter (Sankhyā *A* **26**, 1964) and $\tilde{f}_N$ is almost of that type: both may be said to be of the "δ-sequence type". If f satisfies a certain Lipschitz condition at x and the "number of harmonics" used in $\tilde{f}_N$ is asymptotically proportional to $N^{\frac{1}{3}}$, and $\rho_N/\log N \to \infty$, then $(N^{\frac{1}{3}}/\rho_N) \cdot |\tilde{f}_N(x) - f(x)| \to 0$ a.s.; a similar result holds for $f_N^\#$.

1. Introduction. This paper deals with strong consistency of some nonparametric density estimators. Extensive reviews of nonparametric density estimation can be found in [15] and [8]. Pointwise or uniform strong consistency results are available for histogram-like estimators [7, 13], kernel-type estimators [5, 6, 7, 9, 12], and for a modified kernel estimator [16]. Strong consistency results are not available for δ-sequence type estimators which are *not of the kernel type*. We fill this lacuna by establishing the *rate of strong consistency* of two estimators of the δ-sequence type which are *not* of the kernel type. These two closely related estimators, noted $\tilde{f}_N$ and $f_N^\#$, are based on trigonometric series and the Fejér kernel of classical Fourier analysis. The estimator $\tilde{f}_N$ is appropriate when the "unknown" density f vanishes off a known bounded interval, while $f_N^\#$ is applicable without any assumption about the support of f. The estimator $\tilde{f}_N$ is developed in this section, and its asymptotic bias and variance are obtained in Section 2; the main result is given in Section 3, while Section 4 contains the extension from $\tilde{f}_N$ to $f_N^\#$.

Our assumptions and notation are as follows. $(\Omega, \mathscr{F}, \mathbf{P})$ is a probability space on which are defined a rv X and an i.i.d. sequence $(X_k)_1^\infty$; f is the density (with respect to Lebesgue measure) of X and of each X_k. $\mathbf{T} = [-\pi, \pi)$, $\mathbf{R}$ and $\mathbf{C}$ are the real and the complex numbers. If $g : \mathbf{T} \to \mathbf{C}$ then g^e denotes the 2π-periodic extension of g to all of $\mathbf{R}$; if $\mathbf{T}$ is a subset of the domain of g, then g^e is the 2π-periodic extension of the restriction $g|_\mathbf{T}$. In Sections 1, 2 and 3 it is assumed that f vanishes on the complement of a known bounded interval; without loss of generality, suppose that f vanishes off $\mathbf{T}$. In Section 4, nothing is assumed about the support of f.

Received April 1973; revised May 1974.
AMS 1970 *subject classifications.* Primary 60F15, 62G05; Secondary 42A08, 42A20, 65D10.
Key words and phrases. Nonparametric density estimation, strong consistency, rate of convergence, Fourier series.

The estimator f_N is obtained from the *Fejér sums* for f, by which we mean the following. If $g \in L_1(\mathbf{T})$, then the Fourier coefficients of g are $\gamma_j = \int_{-\pi}^{\pi} g(x)e^{-ijx}(2\pi)^{-1}dx$ and the partial sums of the Fourier series of g are $S_m(g, x) = \sum_{j=-m}^{m} \gamma_j e^{ijx}$; the Cesàro means of these partial sums are the *Fejér sums* $\sigma_\nu(g, x)$, i.e.

$$
\begin{aligned}
\sigma_\nu(g, x) &= (\nu + 1)^{-1}(S_0(g, x) + S_1(g, x) + \cdots + S_\nu(g, x)) \\
&= (\nu + 1)^{-1} \textstyle\sum_{m=0}^{\nu} \sum_{j=-m}^{m} \gamma_j e^{ijx} \qquad (1.1)\\
&= \textstyle\sum_{j=-\nu}^{\nu} \left(1 - \dfrac{|j|}{\nu + 1}\right) \gamma_j e^{ijx} .
\end{aligned}
$$

By a well-known result due to Fejér (see, e.g., page 89 of [17]), if g^e is continuous at x then $\sigma_\nu(g, x) \to g^e(x)$ as $\nu \to \infty$. The Fourier coefficients of the density f are

$$a_j = \int_{-\pi}^{\pi} f(x)e^{-ijx}(2\pi)^{-1}dx = \mathbf{E}((2\pi)^{-1}e^{-ijX}) , \qquad j = 0, \pm 1, \pm 2, \cdots$$

and it is plausible to estimate a_j by the corresponding sample average:

$$\hat{a}_{jN} = N^{-1} \textstyle\sum_{k=1}^{N} (2\pi)^{-1}e^{-ijX_k} ; \qquad (1.2)$$

since the dependence on the sample size N is quite obvious, we suppress the second subscript and write $\hat{a}_j$ instead of $\hat{a}_{jN}$. One arrives at a plausible estimator of f if one replaces a_j in $\sigma_\nu(f, x)$ by $\hat{a}_j$; the result is a 2π-periodic function on $\mathbf{R}$, and our estimator is that function multiplied by $I_\mathbf{T}$, the indicator function of $\mathbf{T}$. That is,

$$f_N(x) = \textstyle\sum_{j=-\nu}^{\nu} \left(1 - \dfrac{|j|}{\nu + 1}\right) \hat{a}_j e^{ijx} \qquad (1.3)$$

$$\text{for } -\pi \leqq x < \pi \text{ and } = 0 \text{ elsewhere.}$$

The number of harmonics, ν, should increase with increasing sample size: $\nu = \nu_N \to \infty$ as $N \to \infty$; the dependence of ν on N will be examined more closely in Section 3.

It is convenient to reformulate f_N in terms of the *Fejér kernel*, defined by

$$K_\nu(x) = (2\pi)^{-1} \textstyle\sum_{j=-\nu}^{\nu} \left(1 - \dfrac{|j|}{\nu + 1}\right) e^{ijx} , \qquad x \in \mathbf{R} . \qquad (1.4)$$

(Note that some authors refer to $2\pi K_\nu$ as the Fejér kernel.) Substituting (1.2) in (1.3) one obtains

$$
\begin{aligned}
f_N(x) &= \textstyle\sum_{j=-\nu}^{\nu} \left(1 - \dfrac{|j|}{\nu + 1}\right) e^{ijx} N^{-1} \sum_{k=1}^{N} (2\pi)^{-1}e^{-ijX_k} \\
&= N^{-1} \textstyle\sum_{k=1}^{N} \sum_{j=-\nu}^{\nu} \left(1 - \dfrac{|j|}{\nu + 1}\right) (2\pi)^{-1}e^{ij(x-X_k)} \qquad (1.5)\\
&= N^{-1} \textstyle\sum_{k=1}^{N} K_\nu(x - X_k) , \qquad -\pi \leqq x < \pi .
\end{aligned}
$$

This form is useful in the theoretical study of f_N: standard probabilistic results come into play because now $\tilde{f}_N(x)$ is an average of i.i.d. rv's, and one can exploit various well-known properties of the Fejér kernel (all the properties of K_ν used in this paper can be found, e.g., in Section 18.27 of [1].). For example, K_ν is nonnegative and $\int_{-\pi}^{\pi} K_\nu = 1$; it follows that $\tilde{f}_N$ is a (random) density function.

The estimator $\tilde{f}_N$ is quite similar to an estimator proposed by Kronmal and Tarter ([3], pages 938–940); they use Cesàro means of the density's Fourier cosine series. Furthermore, with

$$k_N(x, y) = \sum_{-\infty}^{\infty} \alpha_j(N)\varphi_j(x)\overline{\varphi_j(y)}\,,$$

$\varphi_j(x) = \exp(ijx)/(2\pi)^{\frac{1}{2}}$, $\alpha_j(N) = 1 - |j|(\nu+1)^{-1}$ if $|j| \leqq \nu$ and 0 otherwise, and $w(x) \equiv 1$, we see from (1.5) that

$$\begin{aligned} N^{-1}\sum_{m=1}^{N} w(X_m)k_N(x, X_m) &= N^{-1}\sum_{m=1}^{N}\sum_{j=-\nu}^{\nu}(1 - |j|(\nu+1)^{-1})(2\pi)^{-1}e^{ij(x-X_m)} \\ &= \tilde{f}_N(x)\,. \end{aligned}$$

This shows that $\tilde{f}_N$ is essentially identical with an estimator considered by Rosenblatt, i.e. the estimator specified by (87), (88), (90), and (99) in [8].

2. Bias and variance. We will study the asymptotic behavior of $\tilde{f}_N$ under a certain Lipschitz-type condition on f. Therefore we introduce the following *terminology*.

If $g: \mathbf{T} \to \mathbf{C}$, we say that g is *π-Lipschitz at* x iff

$$(\exists M)(\forall y \in \mathbf{R})(|x - y| < \pi \Rightarrow |g^e(x) - g^e(y)| \leqq M|x - y|)\,.$$

There is an important relation between Fejér sums, and integrals of Fejér kernels: if $g \in L_1(\mathbf{T})$ then

$$\sigma_\nu(g, x) = \int_{-\pi}^{\pi} g^e(x - t)K_\nu(t)\,dt = \int_{-\pi}^{\pi} g(t)K_\nu(x - t)\,dt\,. \tag{2.1}$$

From this, and (1.5), we see that

$$\mathbf{E}\tilde{f}_N(x) = \mathbf{E}K_\nu(x - X) = \sigma_\nu(f, x)\,. \tag{2.2}$$

By the previously cited Fejér theorem, if x is a continuity point of f^e then

$$N \to \infty \;\Rightarrow\; \nu_N \to \infty \;\Rightarrow\; \sigma_\nu(f, x) \to f(x)\,,$$

i.e. $\tilde{f}_N(x)$ is asymptotically unbiased.

The rate of asymptotic bias can be obtained from a corresponding result for Fourier series:

(2.3) if $g \in L_1(\mathbf{T})$ and g is π-Lipschitz at x then, as $\nu \to \infty$, $\sigma_\nu(g, x) - g(x) = O(\nu^{-1}\log\nu)$.

This can be found in [2], page 21, or in [10], page 442, where it is attributed to S. N. Bernstein. From (2.2) and (2.3) we see that

(2.4) if f is π-Lipschitz at x then there is a constant b such that, for all N, $|f(x) - \mathbf{E}\tilde{f}_N(x)| \leqq b \cdot \dfrac{\log \nu_N}{\nu_N}$.

According to (95) and (100) in [8], there is a constant a such that

$$aN\nu_N^{-1}\mathbf{V}\tilde{f}_N(x) \to f(x) \qquad \text{as } N \to \infty . \tag{2.5}$$

It follows from (1.5) that

$$a\nu_N^{-1}\mathbf{V}K_\nu(x - X) \to f(x) \qquad \text{as } N \to \infty . \tag{2.6}$$

These asymptotic variance results are valid whenever x is a continuity point of f^e. One way to establish this is by relating $\tilde{f}_N$ to so-called δ-function sequences.

If $(\delta_N)_1^\infty$ is a sequence of functions satisfying (a)—(d) on page 102 of [14] then Watson and Leadbetter call it a *δ-function sequence*. Such a sequence can be used to construct a density estimator f_N by putting

$$f_N(x) = N^{-1} \textstyle\sum_{k=1}^N \delta_N(x - X_k) . \tag{2.7}$$

No domain is specified in [14] and integrals are simply written $\int$, but it is clearly implied that $\mathbf{R}$ is the domain of δ_N and that integrals are over $\mathbf{R}$. It is convenient to adapt this concept to the case where the domain of each δ_N is $[-\pi, \pi)$ or, more generally, a bounded interval $[a, b)$ having 0 as an interior point; we say that $(\delta_N)_1^\infty$ is a *δ-function sequence* on $\mathbf{T}$ if $(\delta_N)_1^\infty$ satisfies the indicated conditions, with $\mathbf{R}$ replaced by $\mathbf{T}$. If f is concentrated on $\mathbf{T}$ and $(\delta_N)_1^\infty$ is a δ-function sequence on $\mathbf{T}$, then it can be used to define an estimator of f, in the manner of (2.7). With $-\pi \leqq x < \pi$, $x - X_k$ may be outside $[-\pi, \pi)$; therefore δ_N must be extended. If δ_N as well as f are extended *periodically* (with period 2π) then (i) $\delta_N(x - X_k)$ makes sense and (ii) the principal proofs in [14] work, modulo a few details, when $\mathbf{R}$ is replaced by $[-\pi, \pi)$. Thus the essential properties proved in [14] for δ-function sequences on $\mathbf{R}$ are also valid for δ-function sequences on $\mathbf{T}$.

We will say that an estimator is of the *δ-sequence type* if it is of the form (2.7) and $(\delta_N)_1^\infty$ is a δ-function sequence on $\mathbf{R}$ (i.e., in the sense of [14]), or if $(\delta_N)_1^\infty$ is a δ-function sequence on $\mathbf{T}$, each δ_N is extended 2π-periodically, and the estimator is of the form

$$(N^{-1} \textstyle\sum_{k=1}^N \delta_N(x - X_k))I_{\mathbf{T}}(x) .$$

Standard properties of the Fejér kernel allow us to conclude that if $\nu_N \to \infty$ as $N \to \infty$ then $(K_{\nu_N})_{N=1}^\infty$ is a δ-function sequence on $\mathbf{T}$. It follows that $\tilde{f}_N$ is of the δ-sequence type and the principal results in [14] apply to this estimator; in particular, from Theorem 4 in [14] and the simple identity

$$\int_{-\pi}^{\pi} K_\nu^2(y)\,dy = (2\pi)^{-1} \sum_{j=-\nu}^{\nu} \left(1 - \frac{|j|}{\nu + 1}\right)^2 ,$$

it follows that (2.5) holds whenever x is a continuity point of f^e.

3. Strong consistency. We will prove that if f vanishes off $\mathbf{T}$ and is π-Lipschitz at x then $\tilde{f}_N(x)$ converges a.s. to $f(x)$ at a rate of nearly $N^{-\frac{1}{3}}$, provided the number of harmonics ν_N is chosen appropriately. The rate of convergence is nearly that obtained by Révész in [7] for the almost sure uniform convergence of histograms, though his conditions on f are stronger than ours. The proof uses the following bound.

(3.1) LEMMA. *Suppose* $(Y_k)_1^n$ *are* i.i.d. rv's, *with zero mean and common variance* $\mathbf{V}Y$, *and such that* $|Y_k| \leqq 1$ a.s.; *then, for* $0 \leqq s \leqq 1$ *and* $\varepsilon > 0$,

$$\mathbf{P}[|n^{-1} \sum_1^n Y_k| > \varepsilon] \leqq 2e^{-ns\varepsilon}(1 + s^2\mathbf{V}Y)^n .$$

PROOF. From (6) on page 44 of [4] one sees that $\mathbf{E}(e^{sY_k}) \leqq 1 + s^2\mathbf{V}Y$ when $0 \leqq s \leqq 1$. Now by the "exponential form" of Chebyshev's inequality,

$$\mathbf{P}[\sum_1^n Y_k > n\varepsilon] \leqq e^{-sn\varepsilon}\mathbf{E}\exp(s \sum_1^n Y_k) \leqq e^{-sn\varepsilon}(1 + s^2\mathbf{V}Y)^n$$

and similarly

$$\mathbf{P}[\sum_1^n Y_k < -n\varepsilon] \leqq e^{-sn\varepsilon}(1 + s^2\mathbf{V}Y)^n ,$$

which proves the lemma.

(3.2) THEOREM. *Suppose that, for a particular* $x \in [-\pi, \pi)$,

$$(\exists M)(\forall y \in \mathbf{R})(|x - y| < \pi \Rightarrow |f^e(x) - f^e(y)| \leqq M|x - y|) .$$

If $\nu_N/N^{\frac{1}{3}} \to c > 0$, *and* $(\rho_N)_1^\infty$ *is a sequence of constants such that* $\rho_n/\log n \to \infty$, *then*

$$\frac{N^{\frac{1}{3}}}{\rho_N} |\tilde{f}_N(x) - f(x)| \to 0 \quad \text{a.s.}$$

PROOF. At the outset, we suppose $\nu_N/N^\alpha \to c > 0$ for some yet unspecified $\alpha > 0$ and we consider $\rho_N^{-1} \cdot N^\beta|\tilde{f}_N(x) - f(x)|$ with some as yet unspecified β. The objective is to prove convergence with β as large as possible; ρ_n is included for proper adjustment of the convergence rate. In view of (2.4), and with $c_N = \nu_N/N^\alpha$,

$$\frac{N^\beta}{\rho_N} |f(x) - \mathbf{E}\tilde{f}_N(x)| \leqq \frac{N^\beta}{\rho_N} \cdot b \cdot \frac{\log \nu_N}{\nu_N} = b \cdot \frac{N^\beta}{\rho_N} \cdot \frac{\log c_N + \alpha \log N}{N^\alpha c_N}$$

for some constant b; since $\rho_N/\log N \to \infty$,

$$\frac{N^\beta}{\rho_N} |\mathbf{E}\tilde{f}_N(x) - f(x)| \to 0 \qquad \text{if} \quad \beta \leqq \alpha . \tag{3.3}$$

Fix $\varepsilon > 0$, and let

$$A_N(\varepsilon) = \left[\frac{N^\beta}{\rho_N} |\tilde{f}_N(x) - \mathbf{E}\tilde{f}_N(x)| > \varepsilon\right].$$

Now

$$\begin{aligned}\mathbf{P}(A_N(\varepsilon)) &= \mathbf{P}\left[\frac{N^\beta}{\rho_N}\left|\frac{1}{N} \sum_{k=1}^N (K_\nu(x - X_k) - \mathbf{E}K_\nu(x - X_k))\right| > \varepsilon\right] \\ &= \mathbf{P}\left[\frac{2\pi}{\nu + 1}\left|\frac{1}{N} \sum_{k=1}^N (K_\nu(x - X_k) - \mathbf{E}K_\nu(x - X_k))\right| > \frac{2\pi\varepsilon\rho_N}{(\nu + 1)N^\beta}\right] \\ &= \mathbf{P}\left[\left|\frac{1}{N} \sum_{k=1}^N Y_{Nk}\right| > \varepsilon_N\right]\end{aligned}$$

where

$$Y_{Nk} = \frac{2\pi}{\nu_N + 1}(K_{\nu_N}(x - X_k) - \mathbf{E}K_{\nu_N}(x - X_k)), \qquad \varepsilon_N = \frac{2\pi\varepsilon\rho_N}{(\nu_N + 1)N^\beta}.$$

Each Y_{Nk} has zero mean and, since $0 \leqq K_\nu \leqq (2\pi)^{-1}(\nu + 1)$, $|Y_{Nk}| \leqq 1$. By (3.1)

$$\mathbf{P}(A_N(\varepsilon)) \leqq 2(1 + s^2\mathbf{V}Y_{N1})^N \exp(-Ns\varepsilon_N), \qquad 0 \leqq s \leqq 1.$$

In order to apply the Borel–Cantelli lemma, we would like to arrange matters so that $\sum_N \mathbf{P}(A_N(\varepsilon)) < \infty$, *regardless* of the value of ε. That can be done if we let s depend on N in such a manner that the first factor above remains bounded (as $N \to \infty$) whereas the second becomes dominated by $1/N^2$. Now

$$\begin{aligned}\mathbf{V}Y_{N1} &= \left(\frac{2\pi}{\nu + 1}\right)^2 \mathbf{V}K_\nu(x - X)\\ &\leqq \frac{4\pi^2}{\nu^2} \cdot \frac{\nu}{a} \cdot \left(\frac{a}{\nu}\mathbf{V}K_\nu(x - X)\right) = \frac{a'}{N^\alpha c_N} \cdot \lambda_N\end{aligned}$$

where $\lambda_N \to f(x)$ by (2.6); it follows that

$$(1 + s^2\mathbf{V}Y_{N1})^N \leqq \left(1 + \frac{1}{N} \cdot \frac{Ns^2}{N^\alpha} \cdot \frac{a'\lambda_N}{c_N}\right)^N$$

and that this bound will converge to a finite limit if we take $s = N^{-\sigma}$ ($\sigma > 0$) with $1 - 2\sigma - \alpha \leqq 0$. On the other hand,

$$\exp(-Ns\varepsilon_N) = \exp\left[-\frac{Ns\varepsilon_N}{\log N}\log N\right] = 1/N^{\xi_N}$$

with

$$\begin{aligned}\xi_N = \frac{Ns\varepsilon_N}{\log N} &= \frac{N^{-\sigma}}{\log N} \cdot \frac{N2\pi\varepsilon\rho_N}{(\nu_N + 1)N^\beta}\\ &= \frac{2\pi\varepsilon\rho_N}{\log N} \cdot \frac{N^\alpha}{\nu_N + 1} \cdot N^{1-\sigma-\alpha-\beta}.\end{aligned}$$

Now $\exp(-Ns\varepsilon_N)$ will ultimately be dominated by $1/N^2$ if $\xi_N \to \infty$; since $\rho_N/\log N \to \infty$, we want $1 - \sigma - \alpha - \beta \geqq 0$. The largest β that satisfies this condition, subject to the former conditions $\beta \leqq \alpha$, $1 - 2\sigma - \alpha \leqq 0$, and $\sigma > 0$, is $\beta = \frac{1}{3}$; in order to satisfy all these conditions when β is $\frac{1}{3}$, one must also take $\alpha = \frac{1}{3}$ (and $\sigma = \frac{1}{3}$). So if $\alpha = \beta = \frac{1}{3}$ then, for large enough $m = m(\varepsilon)$, $\sum_m^\infty \mathbf{P}|A_N(\omega)| < \sum_m^\infty 1/N^2$; by a standard use of the Borel–Cantelli lemma,

$$\frac{N^\beta}{\rho_N}|\tilde{f}_N(x) - \mathbf{E}\tilde{f}_N(x)| \to 0 \quad \text{a.s.} \qquad \text{when } \alpha = \beta = \tfrac{1}{3}. \tag{3.4}$$

The theorem follows from (3.3) and (3.4) by the triangle inequality.

4. The estimator $f_N^\#$. In Sections 1–3 it was assumed that f vanishes off $[-\pi, \pi)$; it does not make sense to estimate f by means of $\tilde{f}_N$ if that assumption is not satisfied, since $\tilde{f}_N$ is a probability density which vanishes off $[-\pi, \pi)$. But one can construct an estimator which is quite similar to $\tilde{f}_N$ and which is applicable without any assumptions about the support of f. In this section we outline the construction of one such estimator, denoted by $f_N^\#$, and indicate how the arguments in previous sections apply to $f_N^\#$.

The Fejér kernel K_ν is defined on $\mathbf{R}$ and is 2π-periodic. We shall now use a non-periodic version of K_ν, a function which might be called the "basic pattern" of K_ν. More precisely,

(4.1) The *sharp Fejér kernel* is

$$K_\nu^\#(x) = K_\nu(x)I_{[-\pi,\pi]}(x)\,, \qquad x \in \mathbf{R}\,.$$

Some properties of K_ν extend to $K_\nu^\#$ in an obvious way, and one sees that

(4.2) if $\lim_{N\to\infty} \nu_N = \infty$ then $(K_{\nu_N}^\#)_{N=1}^\infty$ is a δ-function sequence on $\mathbf{R}$.

Recall (2.1): $\sigma_\nu(g, x) = \int_{-\pi}^{\pi} g^e(x-t)K_\nu(t)\,dt$. Supposing that g is in fact defined on $\mathbf{R}$, we replace g^e by g and replace K_ν by $K_\nu^\#$, and call the result $\sigma_\nu^\#$. Thus, for $g \in L_1(\mathbf{R})$ and $x \in \mathbf{R}$,

$$\begin{aligned}(4.3)\qquad \sigma_\nu^\#(g, x) &= \int_{-\pi}^{\pi} g(x-t)K_\nu^\#(t)\,dt = \int_{\mathbf{R}} g(x-t)K_\nu^\#(t)\,dt \\ &= \int_{\mathbf{R}} g(t)K_\nu^\#(x-t)\,dt\,.\end{aligned}$$

Corresponding to (1.5), we define

$$(4.4)\qquad f_N^\#(x) = \frac{1}{N}\sum_{k=1}^{N} K_\nu^\#(x - X_k)\,, \qquad x \in \mathbf{R}\,.$$

It is easily seen that $f_N^\#$ is a density function and that $\mathbf{E}f_N^\#(x) = \sigma_\nu^\#(f, x)$. The rate-of-approximation result (2.3) also applies to $\sigma_\nu^\#$; therefore the rate of asymptotic bias given in (2.4) also applies to $f_N^\#$.

In view of (4.2), $f_N^\#$ is an estimator of the δ-sequence type; therefore the results in [14] apply to $f_N^\#$. Noting that $\int_{\mathbf{R}} (K_n^\#)^2 = \int_{-\pi}^{\pi} (K_n^\#)^2 = \int_{-\pi}^{\pi} (K_n)^2$, we again obtain the asymptotic variance by using Theorem 4 in [14]. With $\tilde{f}_N$ replaced by $f_N^\#$, and K_ν replaced by $K_\nu^\#$, (2.5) and (2.6) are again valid at continuity points of f. Now the proof in Section 3 applies, *mutatis mutandis*, to $f_N^\#$; it suffices to replace K_ν by $K_\nu^\#$ and σ_ν by $\sigma_\nu^\#$. It follows that the rate of strong consistency established in (3.2) also applies to $f_N^\#$, provided $x \in \mathbf{R}$ is a point at which f is π-Lipschitz.

Unlike $\tilde{f}_N$, the derivation of $f_N^\#$ is not based on the assumption that f vanishes off $[-\pi, \pi)$, and the estimator $f_N^\#$ does not necessarily vanish off that interval. Unlike K_ν, the function $K_\nu^\#$ is not periodic and therefore the terms in (4.4) cannot be rearranged into a form corresponding to (1.3). Since (1.3) gives $\tilde{f}_N$ as a sum of $2\nu + 1$ terms, rather than N, this form offers some possible computational advantages (as discussed, e.g., in [3]). Apparently $f_N^\#$ cannot be put in a form which offers such advantages.

Acknowledgment. I am indebted to the referees for several helpful comments. In particular, they have pointed out that the proof herein would also apply to many other δ-sequence and kernel estimators.

REFERENCES

[1] HEWITT, E. and STROMBERG, K. (1965). *Real and Abstract Analysis.* Springer-Verlag, New York.

[2] KATZNELSON, Y. (1968). *An Introduction to Harmonic Analysis.* Wiley, New York.

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

[4] LAMPERTI, J. (1966). *Probability.* W. A. Benjamin, New York.

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

[6] NADARAYA, E. A. (1970). Remarks on non-parametric estimates for density functions and regression curves. *Theor. Probability Appl.* **15** 134-137.

[7] REVESZ, P. (1972). On empirical density function. *Periodica Math. Hungar.* **2** 85-110.

[8] ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815-1842.

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

[10] SZ.-NAGY, B. (1965). *Introduction to Real Functions and Orthogonal Expansions.* Oxford Univ. Press, New York.

[11] TARTER, M. and RAMAN, S. (1971). A systematic approach to graphical methods in biometry. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **4** 199-222.

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

[13] VAN RYZIN, J. (1970). On a histogram method of density estimation. Technical Report No. 226, Dept. of Statistics, Univ. of Wisconsin.

[14] WATSON, G. and LEADBETTER, M. (1964). Hazard analysis II. *Sankhyā Ser. A* **26** 101-116.

[15] WEGMAN, E. J. (1972). Nonparametric probability density estimation: I. A summary of available methods. *Technometrics* **14** 533-546.

[16] WOODROOFE, M. (1967). On the maximum deviation of the sample density. *Ann. Math. Statist.* **38** 475-481.

[17] ZYGMUND, A. (1959). *Trigonometric Series I.* Univ. Press, Cambridge.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OTTAWA
OTTAWA, ONTARIO, CANADA