Ann. Inst. Statist. Math.
32 (1980), Part A, 65-80

# A CLASS OF SPECTRAL DENSITY ESTIMATORS

GILBERT G. WALTER

(Received Feb. 9, 1979; revised Oct. 29, 1979)

## Abstract

A class of spectral windows depending on one parameter is presented and shown to include many of the common windows. The mean square rate of convergence of the associated spectral density estimators are calculated in terms of this parameter for spectral densities which are locally Lipschitz continuous. The class is shown to include certain data tapers and data windows corresponding to missing observations. This is true also for the kernels of $(C-\alpha)$ summability which provide means for estimating the spectral density when the covariance function is periodic.

## 1. Introduction

The problem of estimating the spectral density of a stationary stochastic process has a number of different solutions corresponding to different "spectral windows." Windows for use in such estimation have been proposed by many investigators, in particular Bartlett [2], Blackman-Tukey (see [1]), and Parzen [8]. To this array we should like to add one more, which, however, can be used sometimes when the others cannot. This new window enables one to estimate the spectrum even when the process has a periodic component. Thus the necessity for first removing such components is avoided.

This estimator as well as most others used in spectral density estimation is based on "delta sequences" i.e., sequences of functions that converge to the "delta function." Such sequences were used by Walter and Blum [11] to study probability density estimation. We shall study a one parameter class of such sequences which includes the common windows. The mean square rate of convergence of the associated estimator is found in terms of this parameter for a Lipschitz continuous spectral density. The rate obtained is $O(n^{-2/3+\epsilon})$ for a class of spectral

AMS 1970 subject classifications: Primary 60G15; secondary 60G99.
Key words and phrases: Spectral density, estimation, spectral windows, lag windows, data tapers.

densities for which Farrell* has shown the best possible rate to be $O(n^{-2/3})$.

The problem of estimation when certain observations are missing has been studied by Jones [6] and Brillinger [3] among others. We shall show that under certain conditions on the missing observations, a delta-sequence of this class is obtained and hence rates of convergence of the estimators found. The same is true for data tapers of standard types.

Finally, the kernels of $(C-\alpha)$ summability are shown to belong to this class. In this case, however, stronger results can be obtained. The estimators converge to the spectral density at isolated points of Lipschitz continuity even when it is a generalized function, and the mean square rate can be found.

## 2. Notation and terminology

Let $\{X(t),\ t=0, \pm 1, \pm 2, \cdots\}$ be a discrete parameter stationary random process with zero mean and finite second moments. Its covariance function, $C_k$, is given by

$$C_k = \mathrm{E}\,(X(t)X(t+k))\ , \qquad k=0, \pm 1, \pm 2, \cdots ,$$

while its spectral density, $f(\lambda)$, is given by the finite Fourier transform of $C_k$,

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} C_k e^{-ik\lambda} .$$

This series may not converge pointwise but always will converge weakly $\left(f_n \to f \text{ weakly} \Longleftrightarrow \int f_n \psi \to \int f \psi \text{ for periodic } C^\infty \text{ functions}\right)$.

We assume that the process has been observed from times 1 to $N$. We denote by $Z_N(\lambda)$ the random variable given by the finite Fourier transform

$$(1) \qquad Z_N(\lambda) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} X(t) e^{-i\lambda t} \qquad -\pi \leqq \lambda < \pi$$

and by $H_N(\lambda)$ the function

$$(2) \qquad H_N(\lambda) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} e^{-i\lambda t} \qquad -\pi \leqq \lambda < \pi .$$

We also use the same notation for their periodic extensions to $R^1$.

Then the expected value of $|Z_N(\lambda)|^2$ is given by

* Farrell, R. H. (1979). Asymptotic lower bounds for the risk of estimators of the value of a spectral density function, to appear.

$$\text{(3)} \qquad \mathrm{E}\,|Z_N(\lambda)|^2=\frac{1}{\pi}\int_{-\pi}^{\pi}|H_N(\lambda-\omega)|^2 f(\omega)d\omega$$

where $f(\lambda)$ is the spectral density and $|H_N^2(\lambda)|=F_N(\lambda)$ is the Fejér kernel of Fourier series theory (see Koopmans [7], p. 259). Hence we have that (among other things), at each point $\lambda$ of continuity of $f$,

$$\mathrm{E}\,|Z_N(\lambda)|^2\rightarrow f(\lambda)$$

and therefore that $I_N(\lambda)=|Z_N(\lambda)|^2$, the so-called *periodogram* is an asymptotically unbiased estimator of $f(\lambda)$. It may also be given by the expression

$$\text{(4)} \qquad I_N(\lambda)=\frac{1}{2\pi}\sum_{k=-N+1}^{N-1}\hat{C}_k e^{-i\lambda k} \qquad \text{where}$$

$$\hat{C}_k=\begin{cases}\dfrac{1}{N}\displaystyle\sum_{t=1}^{N-|k|}X(t+|k|)X(t) & |k|<N\\ 0 & N\leqq|k|\,.\end{cases}$$

This periodogram unfortunately, has variance which does not approach zero as $N\rightarrow\infty$. The problem is rectified by introducing another sequence of functions $W_M(\lambda)$, the *spectral window*. The set of its Fourier coefficients $\{w_M(k)\}$, are called a *lag window*;

$$\text{(5)} \qquad W_M(\lambda)=\frac{1}{2\pi}\sum_{k=-M+1}^{M-1}w_M(k)e^{-i\lambda k}\,.$$

The estimator associated with $W_M$ is then given by

$$\text{(6)} \qquad \hat{f}_{MN}(\lambda)=(W_M * I_N)(\lambda)\,.$$

By controlling the relative growth of $M$ and $N$, the variance can be made to approach 0. Thus the estimator (6) can be made into a mean square consistent estimator. This is a well known theory and can be found in more detail in the books by Anderson [1], Grenander-Rosenblatt [5], and Koopmans [7].

We shall be interested mainly in the rate of mean square convergence of these estimators for a certain class of windows. This class generalizes most of the standard windows and includes as well the $(C-\alpha)$ summability kernels. It is given by

DEFINITION 1. Let $\{\delta_m\}$ be a delta sequence of bounded functions on $(-\pi,\pi)$ such that

(i) $\int_{-\pi}^{\pi}\delta_m=1$

(ii) $\int_{-\pi}^{\pi}|\delta_m|=O(1)$

(iii) there exist $0<\beta\leqq 1$, $D_\beta>0$ such that

$$|\delta_m(t)|\leqq D_\beta m/(1+m^{\beta+1}|t|^2) \qquad -\pi<t<\pi;\ m=1,2,\cdots.$$

Then $\{\delta_m\}$ is said to be of *class* $\Gamma_\beta$. Let $\alpha=\sup\{\beta|\text{(iii) holds}\}$; then $\{\delta_m\}$ is said to be of *type* $\alpha$.

It should be observed that (i) if $\beta_1<\beta_2$ then $\Gamma_{\beta_1}\supset\Gamma_{\beta_2}$; (ii) if $\{\delta_m\}$ is of type $\alpha$ then it is in class $\Gamma_\beta$ for each $\beta<\alpha$, but is not in class $\Gamma_\alpha$ necessarily.

*Remark* (i). A delta sequence is one that satisfies $\int\delta_m\phi\to\phi(0)$ for each $C^\infty$ function $\phi$ with compact support in $(-\pi,\pi)$. Those of type $\alpha$ are examples of quasi-positive kernels (see Zygmund [12], I, p. 86) or approximate identities.

PROPOSITION 1. *Let $w(v)$ be a non-negative, concave function in $C'[0,1]$ decreasing from* 1 *to* 0; *let $\{W_M\}$ be given by* (5) *where $w_M(k)=w(|k|/M)$; then $\{W_M\}$ is a delta sequence in class $\Gamma_1$ and hence of type* 1.

*Remark* (ii). A number of common windows satisfy the hypothesis of this proposition. This includes the window of Bartlett ($w(v)=1-v$) and Tukey ($w(v)=1-2a+2a\cos\pi v$) but not of Parzen ($w(v)=1-v^2$) and $\left(w(v)=\begin{cases}1-6v^2+6v^3, & 0\leqq v\leqq 1/2\\ 2(1-v)^3, & 1/2<v\leqq 1\end{cases}\right)$. (See Koopmans [7], p. 279.) However, they can be shown also to be in Class $\Gamma_1$ by a similar argument.

PROOF. Clearly $\int_{-\pi}^{\pi}W_M=1$ since $w(0)=1$. Moreover we have, by summation by parts, that (using the notation $\Delta w(k/M)=w(k/M)-w((k+1)/M)$)

$$\begin{aligned}\pi W_M(\lambda)&=\frac{w(0)}{2}+\sum_{k=1}^{M}w\left(\frac{k}{M}\right)\cos k\lambda\\&=\sum_{k=0}^{M-1}\Delta w\left(\frac{k}{M}\right)D_k(\lambda)+w(1)D_M(\lambda)\\&=\sum_{k=0}^{M-2}\Delta^2 w\left(\frac{k}{M}\right)(k+1)F_k(\lambda)+\Delta w\left(\frac{M-1}{M}\right)MF_{M-1}(\lambda).\end{aligned}$$

Here $D_k$ is the Dirichlet kernel and $F_k$ is the Fejér kernel given in equation (3) which may also be written

$$F_k(\lambda)=\frac{\sin^2((k+1)/2)\lambda}{2(k+1)\sin^2(\lambda/2)}\leqq\frac{\pi^2}{2(k+1)\lambda^2}.$$

Since $w(\Omega)$ was concave, $\Delta_h^2 w\leqq 0$, whence it follows that

$$|\pi W_M(\lambda)| \leqq \frac{\pi^2}{2\lambda^2}\left\{\left|\sum_{k=0}^{M-2} \Delta^2 w\left(\frac{k}{M}\right)\right| + \Delta w\left(1-\frac{1}{M}\right)\right\}$$
$$= \frac{\pi^2}{2\lambda^2}\left\{\left|\Delta w\left(\frac{0}{M}\right) - \Delta w\left(1-\frac{1}{M}\right)\right| + \Delta w\left(1-\frac{1}{M}\right)\right\}.$$

Since $w$ was assumed to be in $C'[0, 1]$, $\Delta_h w(x) = w(x) - w(x+h) = O(h)$, and hence we have

$$(7) \qquad |\pi W_M(\lambda)| \leqq \frac{\pi^2}{\lambda^2 M}\|w'\|_\infty \leqq \bar{C}M/(1+\lambda^2M^2) \quad \text{for } |\lambda| \geqq 1/M,\ \bar{C} \text{ constant}.$$

To show that (ii) holds for $|\lambda| < 1/M$ as well, we observe that

$$\pi \|W_M\|_\infty \leqq \sum_{k=0}^{M} w\left(\frac{k}{M}\right) \approx M\|w\|_1$$

which combined with (7) gives us the condition for $\beta = 1$.

*Remark* (iii). The symbol $\|f\|_p$ denotes the usual $L^p$ norm of the function $f$ on some finite interval with respect to Lebesgue measure: $\|f\|_p = \left\{\int |f(\lambda)|^p d\lambda\right\}^{1/p}$ $1 \leqq p < \infty$, $\|f\|_\infty = \sup_x |f(x)|$. Usually the interval will be $[-\pi, \pi]$ but when there is no danger of confusion may be other intervals.

LEMMA 1. *Let $\{\delta_m\}$ be a delta sequence in class $\Gamma_\beta$, $0 < \beta \leqq 1$, let $\{g_n\}$ be a uniformly bounded family of functions on $(-\pi, \pi)$ which satisfies for some $(a, b) \subset (-\pi, \pi)$, $\gamma > 0$, $D > 0$ the inequality*

$$|g_n(x) - g_n(y)| \leqq D|x-y| + O(n^{-\gamma}), \qquad x, y \in (a, b).$$

*Let $[c, d] \subset (a, b)$; then*

$$(\delta_m * g_n)(x) - g_n(x) = O(n^{-\gamma}) + O(m^{-\beta})$$

*uniformly on $[c, d]$ as $n, m \to \infty$.*

PROOF. Let $\eta = \min(b-d, c-a)$ and let $m^{-1} < \eta$; then we have

$$|(\delta_m * g_n)(x) - g_n(x)| = \left|\int_{-\pi}^{\pi} \delta_m(-t)(g_n(x+t) - g_n(x))dt\right|$$
$$\leqq \left|\int_{-\pi}^{-\eta}\right| + \left|\int_{-\eta}^{-m^{-1}}\right| + \left|\int_{-m^{-1}}^{m^{-1}}\right| + \left|\int_{m^{-1}}^{\eta}\right| + \left|\int_{\eta}^{\pi}\right|.$$

For $x \in [c, d]$ the middle integral satisfies

$$\left|\int_{-m^{-1}}^{m^{-1}}\right| \leqq \int_{-m^{-1}}^{m^{-1}} |\delta_m(-t)|(D|t| + O(n^{-\gamma}))dt$$

$$\leqq D_\beta m D m^{-2} + O(n^{-\gamma}) \int_{-\pi}^{\pi} |\delta_m(t)| dt$$
$$= O(m^{-1}) + O(n^{-\gamma})$$

since $|\delta_m(t)| \leqq D_\beta m/(1+m^{\beta+1}t^2)$. The second integral from the right satisfies

$$\left| \int_{m^{-1}}^{\eta} \right| \leqq \int_{m^{-1}}^{\eta} |\delta_m(-t)| (Dt + O(n^{-\gamma})) dt$$
$$\leqq \int_{m^{-1}}^{\eta} D_\beta Dt/(t^2 m^\beta) dt + O(n^{-\gamma})$$
$$= O(m^{-\beta}) + O(m^{-1} \log m) + O(n^{-\gamma}) .$$

The last integral satisfies

$$\left| \int_{\eta}^{\pi} \right| \leqq \int_{\eta}^{\pi} D_\beta t^{-2} m^{-\beta} 2 \|g_n\|_\infty dt + O(n^{-\gamma}) = O(m^{-\beta}) + O(n^{-\gamma}) .$$

The first two integrals are treated in exactly the same way, and since $\beta < 1$, the conclusion follows.

COROLLARY. *Let $g$ be a bounded function which satisfies a Lipschitz condition on $(a, b)$, and let $[c, d] \subset (a, b)$; then*

$$(\delta_m * g)(x) - g(x) = O(m^{-\beta})$$

*uniformly on $[c, d]$ as $m \to \infty$.*

## 3. Mean square convergence for spectral windows of type $\alpha$

Estimators based on spectral windows of this kind are mean square consistent for spectral densities which are bounded and locally Lipschitz continuous. Their mean square rate can be calculated in terms of the parameter $\alpha$. Lipschitz continuity means only that the slope of the spectral density be bounded, which is always satisfied for smooth spectral densities, but may be considerably weaker, particularly if the continuity is only local.

From a practical point of view, the hypothesis of Lipschitz continuity can always be assumed valid except when the process has a periodic component. It includes all but the most pathological cases and is much weaker than the common hypothesis that the process be linear. In that case the $C_k$ are $O(r^{|k|})$ for some $0 < r < 1$, and hence $f$, the spectral density is in $C^\infty$. Even if the $C_k$ satisfy only the condition $\sum |kC_k| < \infty$, $f$ will be smooth and hence Lipschitz continuous. This is the minimal condition imposed in one common approach (Anderson [1], p. 532).

Of course the data cannot be used to determine whether the spectral density is Lipschitz continuous. However, sometimes it is clear that a periodic component underlies the data. This would happen for example in the case of monthly temperature data in which there is bound to be an annual periodic component. In such cases the hypothesis is not valid and the procedure of Section 5 is more appropriate.

THEOREM 1. *Let* $\hat{f}_{MN}=W_M * I_N$ *where* $\{W_M\}$, *the spectral window, is a delta sequence in class* $\Gamma_\beta$ *of type* $\alpha$ *and* $I_N$ *is the periodogram. Let* $f$, *the spectral density of a stationary Gaussian process with mean* 0 *be bounded and satisfy a uniform Lipschitz condition on* $(a, b)$. *Then the mean square error for* $M=[N^{1/(1+2\alpha)}]$ *is*

$$\mathrm{E}\,[\hat{f}_{MN}(\lambda)-f(\lambda)]^2=O(N^{-2\beta/(1+2\alpha)}) \tag{8}$$

*uniformly for* $\lambda\in[c, d]\subset(a, b)$.

*Remark* (i). The value of $\alpha$ is the supremum of all possible $\beta$'s. Hence the mean square error approaches $O(N^{-2\alpha/(1+2\alpha)})$ as $\beta$ approaches $\alpha$. For some delta sequences, in particular those in $\Gamma_1$, the supremum is attained and hence the mean square error is $O(N^{-2\alpha/(1+2\alpha)})$.

*Remark* (ii). The reason for choosing $M=[N^{1/(1+2\alpha)}]$ is that is balances the mean square error due to the variance with that due to the bias. The former will be shown to be $O(M/N)$ while that due to the square of the bias will approach $O(M^{-2\alpha})$. These two expressions will be equal and hence their sum minimized when $M/N=M^{-2\alpha}$.

PROOF. The variance of $f_{MN}$ is given by the expression (see Jones [6], p. 390)

$$\mathrm{var}\,[\hat{f}_{MN}(\lambda)]=\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}|A(\lambda-\omega, \lambda-\nu)|^2 f(\omega)f(\nu)d\omega d\nu \tag{9}$$

where $A(\lambda, \mu)=A_{NM}(\lambda, \mu)=\int_{-\pi}^{\pi}H_N(\lambda-s)\overline{H_N(\mu-s)}W_M(s)ds.$

We may also express it as a product of series, which can then be used to show that

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}|A(\lambda, \mu)|^2 d\lambda d\mu=O(M/N)\ .$$

Hence the variance of $\hat{f}_{MN}$ is $O(M/N)$ since $f\in L^\infty$. To estimate the bias term $(\mathrm{E}\hat{f}_{MN}-f)^2$ we first break it up into two parts,

$$\mathrm{E}\hat{f}_{MN}-f=W_M * F_N * f-f=(W_M * F_N * f-F_N * f)+(F_N * f-f)\ . \tag{10}$$

The last term is the difference between the Cesaro means of the Fourier series of $f$ and $f$. The rate of convergence of such a difference is $O(N^{-1}\log N)$ for functions satisfying a Lipschitz condition of order 1. (See Zygmund [12], I, p. 91.) The convolution $F_N * f$ satisfies the same Lipschitz condition as $f$, uniformly in $N$, except for an $O(N^{-1})$ term, since

$$\begin{aligned}(11)\qquad &|(F_N * f)(x+\delta)-(F_N * f)(x)| \\ &\leqq \frac{1}{\pi}\int_{-\pi}^{\pi} F_N(t)|f(x+\delta+t)-f(x+t)|dt \\ &=\int_{-\pi}^{a-x}+\int_{a-x}^{b-x}+\int_{b-x}^{\pi} \leqq \int_{-\pi}^{-\gamma}+\int_{\gamma}^{\pi}+\int_{a-x}^{b-x} \\ &=O(N^{-1})+\int_a^b O(\delta)\ .\end{aligned}$$

This is uniform for $x \in [a+\gamma, b-\gamma]$ where $\gamma$ is any positive number such that $\gamma<(b-a)/2$. Now we can use the fact that $\{W_M\}$ is a delta sequence in class $\Gamma_\alpha$ to show that

$$W_M * (F_N * f)-(F_N * f)=O(N^{-1})+O(M^{-\beta})$$

by using Lemma 1 with $g_N=F_N * f$. This enables us to deduce that (10) is dominated by terms of the form $O(M^{-\beta})+O(N^{-1}\log N)$ and that hence the mean square rate will be dominated by

$$O(MN^{-1})+[O(M^{-\beta})+O(N^{-1}\log N)]^2\ .$$

By taking $M=[N^{1/(1+2\alpha)}]$, (8) follows for $[c, d]=[a+\gamma, b-\gamma]$.

*Remark* (iii). Most of the standard examples of windows are in class $\Gamma_1$, as we have seen, and hence this theorem does not distinguish between them. The mean square error in this case will be $O(N^{-2/3})$.

*Remark* (iv). Farrell* has shown the best possible rate to be $O(n^{-2k/(2k+1)})$ for spectral densities in $C^k$. For $k=1$, the spectral density satisfies a Lipschitz condition of order 1, and hence the rate we obtained cannot be improved markedly for the class $\Gamma_1$. Even if higher order derivatives are hypothesized, the usual spectral windows have as their best rate $O(n^{-4/5})$. (See Anderson [1], Section 9.3.4.) The only exception is the truncated (Dirichlet kernel), which however, because of other difficulties, is rarely used.

*Remark* (v). The kernel of $(C-\alpha)$ summability is given by

$$W_M(\lambda)=K_M^\alpha(\lambda)=\frac{\sum_{\nu=0}^{M} A_{M-\nu}^{\alpha-2}F_\nu(\lambda)(\nu+1)}{A_M^\alpha}$$

* See page 66.

where $A_n^\alpha = \binom{n+\alpha}{n}$. It may also be written as (Zygmund [12], I, p. 77)

$$K_M^\alpha(\lambda) = \sum_{\nu=-M}^{M} \frac{A_{M-|\nu|}^\alpha e^{i\nu\lambda}}{A_M^\alpha} \approx \sum_{\nu=-M}^{M} \left(1 - \frac{|\nu|}{M}\right)^\alpha e^{-i\nu\lambda} .$$

It is in class $\Gamma_1$ for $\alpha \geqq 1$ and in $\Gamma_{2\alpha-1}$ for $1/2 < \alpha < 1$. The EBW (equivalent band width, see Koopmans [7], p. 277) is given approximately by $2\pi c/M$ where, in this case,

$$c = 2 \int_0^1 (1-v)^{2\alpha} dv = \frac{2}{2\alpha+1} ,$$

whence the EBW is $\pi(2\alpha+1)/M$. That is, by choosing $\alpha$ sufficiently large, the equivalent band width may be arbitrarily wide, or by choosing it small it may be made to approach that of the truncated window.

This sequence has other properties which we shall exploit later. In part these properties arise because of the inequality (see Zygmund [12], II, p. 60).

(12) $$|D^r K_M^\alpha(\lambda)| \leqq \frac{C'}{\lambda^{\alpha+1} M^{\alpha-r}} ,$$

$$1/M \leqq |\lambda| \leqq \pi, \quad 0 < \alpha \leqq r+1, \quad C' \text{ constant} .$$

*Example* 3.1. Let $C_k$ be given by

$$C_0 = 1 , \quad C_{2k+1} = (-1)^k/(2|k|+1) , \quad C_{2k} = 0 , \qquad k = \pm 1, \pm 2, \cdots .$$

Then the spectral density is given by

$$C_0 + 2 \sum_{k=1}^{\infty} C_k \cos k\lambda = \pi + 1 \quad \text{on} \quad |\lambda| < \pi/2 .$$

Hence it satisfies a Lipschitz condition on this interval but the series $\sum |C_k|$ is not convergent.

*Example* 3.2. Let $C_k$ be given by

$$C_0 = 1 , \quad C_k = (-1)^k/|k| , \qquad k = \pm 1, \pm 2, \cdots .$$

Then the spectral density is

$$C_0 + 2 \sum_{k=1}^{\infty} C_k \cos k\lambda = \ln (e/(\sqrt{2} \cos \lambda/2)) \qquad |\lambda| < \pi .$$

Hence it satisfies a Lipschitz condition on $(-\pi/2, \pi/2)$ but is not bounded. Hence Theorem 1 does not apply but the theory in Section 5 does.

*Example* 3.3. Let $C_k$ be given by

$$C_k = 1 + O(|k|^{-2}) \qquad k = 0, \pm 1, \pm 2, \cdots ,$$

then the spectral density is a generalized function

$$\sum_{k=-\infty}^{\infty} e^{ik\lambda}+\sum_{k=0}^{\infty} O(k^{-2}) \cos k\lambda=\delta(\lambda)+g(\lambda) , \qquad |\lambda|<\pi ,$$

where $g$ is a smooth function. The sum of $\delta$ and $g$ satisfy a modification of the Lipschitz condition to be defined in Section 5, for $\lambda\neq 0$.

## 4. Data windows

Rather than assigning to each data value $X(t)$ an equal weight in the expression (5), it is sometimes desirable (or necessary) to weight them unequally. We suppose in this section that a sequence $\{b_t\}$ of weights is used. Jones [6] studied the estimators of the type we consider here and simulated a process with missing observations. Brillinger [3] estimated the mean and its distribution when the sampling times were events of a point process. We use the approach of Jones and denote by $B_N(\lambda)$ the normalized Fourier transform of these weights,

$$B_N(\lambda)=\frac{1}{\sqrt{2\pi \sum b_t^2}} \sum_{t=1}^{N} b_t e^{-i\lambda t} , \tag{13}$$

which will play the same role as $H_N(\lambda)$ does in the unweighted case.

Then corresponding to equation (3) we have the substitute $J_N(\lambda)$ for the periodogram whose expected value is

$$\mathrm{E}\, J_N(\lambda)=\frac{1}{\pi} \int_{-\pi}^{\pi} |B_N(\lambda-w)|^2 f(w)dw .$$

As in equation (4) $J_N$ may also be given by

$$J_N(\lambda)=\frac{1}{2\pi} \sum_{k=-N+1}^{N-1} \hat{d}_k e^{-i\lambda k} \tag{14}$$

where

$$\hat{d}_k=\begin{cases} \dfrac{1}{\sum_{t=1}^{N} b_t^2} \sum_{t=1}^{N-|k|} b_{t+k}X(t+k)b_t X(t) & \text{for } |k|<N \\ 0 & N\leqq k . \end{cases}$$

If we introduce a spectral window as well we have as the estimator of $f$

$$\hat{f}_{MN}(\lambda)=(W_M * J_N)(\lambda) . \tag{15}$$

In order for this estimator to be asymptotically unbiased, $|B_N(\lambda)|^2$ must be a delta sequence as well as $W_M$. We shall show that, under certain conditions on the data window $\{b_t\}$, it belongs to a class $\Gamma_\alpha$.

THEOREM 2. *Let* $\{b_t\}$, $0 \leqq b_t \leqq 1$ *be a sequence whose variation satisfies*

(i) $V_N = \sum_{k=1}^{N} |\Delta b_k| = O(N^{\delta/2})$, $0 \leqq \delta < 1$, $N = 1, 2, \cdots$,

*and whose* $l^2$ *norm satisfies*

(ii) $\left[\sum_{t=1}^{N} b_t^2\right]^{-1/2} = O(N^{-\varepsilon/2})$, $0 \leqq \delta < \varepsilon \leqq 1$, $N = 1, 2, \cdots$;

*then* $|B_N(\lambda)|^2$ *where* $B_N$ *is given by* (13) *is a delta sequence in class* $\Gamma_{(\varepsilon-\delta)}$.

PROOF. Consider the sum

$$\sum_{k=1}^{N} b_k e^{i\lambda k} = \left\{\sum_{k=1}^{N} b_k^2\right\}^{1/2} \sqrt{2\pi}\, B_N(\lambda) .$$

By the Abel summation formula it equals

$$\sum_{k=1}^{N-1} \Delta b_k P_k(\lambda) + b_N P_N(\lambda)$$

where $P_k(\lambda) = \sum_{t=1}^{k} e^{i\lambda t}$. Since $|P_k(\lambda)| \leqq \pi |\lambda|^{-1}$, $\lambda \neq 0$ it follows that

$$\left|\sum_{k=1}^{N} b_k e^{-i\lambda k}\right| \leqq \frac{\pi}{|\lambda|}\left\{\sum_{k=1}^{N-1} |\Delta b_k| + 1\right\} = \frac{\pi}{|\lambda|}(V_N + 1) .$$

Hence by using (i) and (ii) of the hypothesis, we find that for $\lambda^2 N^{1+\varepsilon-\delta} \geqq 1$,

$$|B_N(\lambda)|^2 \leqq \frac{D_1}{|\lambda|^2} N^{\delta-\varepsilon} \leqq D_2 N/(1 + \lambda^2 N^{1+\varepsilon-\delta}) . \tag{16}$$

We can also obtain the inequality

$$|B_N(\lambda)|^2 \leqq [\sum b_k]^2 / 2\pi \sum b_k^2 \leqq \frac{N}{2\pi} \tag{17}$$

directly. The two may be combined into the form of condition (ii) of Definition 1 by using (16) for $\lambda^2 N^{1+\varepsilon-\delta} \geqq 1$ and (17) for the opposite inequality. Conditions (i) and (ii) follow from the integration of the Fourier series of $|B_N(\lambda)|^2$ termwise.

*Remark* (i). The effect of omitting terms, i.e., having some of the $b_t = 0$ while the others are equal to 1, is to change $\delta$ and $\varepsilon$. If the number of zeros of $\{b_t\}$ are relatively sparse, say $O(N^\delta)$ for $\delta < 1/2$ and for $t \leqq N$ while the other values of $b_t$ are 1 then

$$V_N = O(N^\delta) \qquad \text{and} \qquad \sum_{t=1}^{N} b_t^2 \approx N - N^\delta .$$

Hence the $|B_N(\lambda)|^2$ is in class $\Gamma_{1-\delta}$.

*Remark* (ii). The theorem can also be used to compare various data tapers. Indeed let $\{b_t\}$ be given by $b_t = b(t/N)$ where $b(v)$ is a function with values in $[0, 1]$ which is monotone on $[0, 1/2)$ and on $(1/2, 1]$. Then the total variation $V_N \leqq 2$ while the $l^2$ norm will be greater than $CN^{1/2}$ for some constant $C$. Hence all tapers will be such that $|B_N(\lambda)|^2$ is a delta sequence of class $\Gamma_1$. They will differ only in the constant $C$ and may be compared by it.

The estimator (15) shares may of the properties of the one considered in Theorem 1. Indeed, essentially the same arguments presented there may be modified slightly to prove

THEOREM 3. *Let $\{b_t\}$ be a data window satisfying the hypothesis of Theorem 2; let $\{W_M\}$ be a spectral window and let $f(\lambda)$ be the spectral density satisfying the hypothesis of Theorem 1; let $f_{MN}$ be given by* (15). *Then the mean square error for $M = [N^{\epsilon/(1+2\alpha)}]$ is given by*

$$\mathrm{E}\,[\hat{f}_{MN}(\lambda) - \hat{f}(\lambda)]^2 = O(N^{-\epsilon 2\beta/(1+2\alpha)}) + O(N^{-2(\epsilon-\delta)}) ,$$

*uniformly on* $[c, d] \subset (a, b)$.

The modifications needed in the argument of Theorem 1 are those involving the variance estimation given by equation (9) and the bias given by (10). We revise the definition of $A(\lambda, \mu)$ to correspond to the estimator (15) as

$$A(\lambda, \mu) = \int_{-\pi}^{\pi} B_N(\lambda - s)\overline{B_N(\mu - s)}W_M(s)ds .$$

Then the $L^2$ norm of $A$ satisfies

$$\iint |A(\lambda, \mu)|^2 d\lambda d\mu = \sum_{t=1}^{N} \sum_{k=1}^{N} b_t^2 b_k^2 w_{t-k}^2 (\sum b_t^2)^{-2}$$
$$\leqq (\sum b_t^2)^{-2} \sum_{t=1}^{N} b_t^2 \sum_{k=-\infty}^{\infty} w_k^2 \leqq (\sum b_t^2)^{-1} \|W_M\|_2^2 \leqq MN^{-\epsilon} .$$

The modifications needed in (10) are clear. By using Lemma 1 twice we first conclude that

$$|B_N|^2 * f - f = O(N^{(\delta-\epsilon)})$$

and then that

$$W_M * |B_N|^2 * f - |B_N|^2 * f = O(N^{(\delta-\epsilon)}) + O(M^{-\beta}) .$$

Thus the mean square error is dominated by terms of the form $O(MN^{-\epsilon}) + O(N^{2(\delta-\epsilon)}) + O(M^{-2\beta})$. By taking $M = [N^{\epsilon/1+2\alpha}]$, the conclusion follows.

*Remark* (iii). Again $M$ is chosen as it is to balance the mean square error of the two terms $O(MN^{-\epsilon})$ and $O(M^{-2\alpha})$. The other term $O(N^{-2(\epsilon-\delta)})$

being independent of $M$ will not influence its choice.

*Remark* (iv). If $\{W_M\}$ is in class $\Gamma_1$, the best possible error rate is $O(N^{-2\varepsilon/3})$ and this will be obtained only if $\varepsilon \geqq 3\delta/2$.

## 5. Spectral densities which are generalized functions

We now turn our attention to the case in which the spectral density is no longer a function satisfying a uniform Lipschitz condition. Instead we assume only that the time series $X(t)$ be stationary and hence that its covariance sequence be bounded. Since this sequence constitutes $C_k$, the coefficients of the spectral density, it need not even be continuous since

$$\frac{1}{2\pi} \sum_{k=-\infty}^{\infty} C_k e^{-i\lambda k} = f(\lambda) \tag{18}$$

does not converge pointwise. It does however converge weakly (see Section 2) to a generalized derivative of $F(\lambda)$, the spectral distribution function, in which sense we interpret $f(\lambda)$. (See L. Schwartz [10].)

Such objects may be locally continuous even though the series (18) doesn't converge. If $C_k$ is periodic, then $f$ will be a "delta-function" locally. It is these two types of behavior we are most interested in and which we shall try to isolate.

If $F(\lambda)$ has an ordinary derivative $\gamma$ at $\lambda=\lambda_0$ we shall say that the generalized function $f$ given by (18) has a *value* $\gamma$ at $\lambda_0$. We shall also generalize the concept of Lipschitz condition. If $f$ is an ordinary function in some neighborhood of $\lambda_0$ then $f$ is Lipschitz-Hölder continuous at $\lambda_0$ if there is a number $\mu$ between 0 and 1 such that $f(\lambda)-f(\lambda_0)= O|\lambda-\lambda_0|^{\mu}$. Clearly if $f$ is Lipschitz continuous at $\lambda_0$ it satisfies this condition for all $0<\mu\leqq 1$. If $f$ is merely a generalized function with a value at $\lambda_0$ the analogous condition would be

$$\frac{F(\lambda)}{(\lambda-\lambda_0)} - \gamma = O\,|\lambda-\lambda_0|^{\mu}\,, \qquad 0<\mu<1\,. \tag{19}$$

LEMMA 2. *Let $f$ have a value $\gamma$ at $\lambda_0 \in (-\pi, \pi)$ satisfying* (19) *for some $0<\mu<1$. Then for each $\alpha>1$*

$$(K_n^{\alpha} * f)(\lambda_0) - \gamma = O(n^{-\beta})$$

*where $K_n^{\alpha}$ is the kernel of $(C-\alpha)$ summability and $\beta=\min(\alpha-1, \mu)$.*

The proof is straightforward. We may suppose $1<\alpha\leqq 2$, and obtain, by integration by parts,

$$(20)\qquad K_n^\alpha * f(\lambda_0)-\gamma=\frac{1}{\pi}\int_{-\pi}^{\pi} K_n^\alpha(\lambda_0-\lambda)[f(\lambda)-\gamma]d\lambda$$
$$=\frac{1}{\pi}\int_{-\pi}^{\pi} DK_n^\alpha(\lambda_0-\lambda)[F(\lambda)-(\lambda-\lambda_0)\gamma]d\lambda$$
$$+K_n^\alpha(\lambda_0-\pi)\left[\int_{-\pi}^{\pi} f-2\pi\gamma\right].$$

The integrated term is $O(n^{-1})$ since $K_n^\alpha$ is in class $\Gamma_1$ and hence obeys the inequality

$$|K_n^\alpha(t)|\leqq D_1 n^{-1}t^{-2} \qquad \text{for } 1/n\leqq t,\ D_1 \text{ constant}.$$

The other term in (20) may be written as $2/\pi$ times the integral

$$(21)\qquad \int_0^{\pi}(\lambda_0-\lambda)DK_n^\alpha(\lambda_0-\lambda)[F(\lambda)/(\lambda_0-\lambda)-\gamma]d\lambda=\int_0^{\lambda_0-1/n}+\int_{\lambda_0-1/n}^{\lambda_0+1/n}+\int_{\lambda_0+1/n}^{\pi},$$

provided $\lambda_0\in(0,\pi)$. The middle integral satisfies

$$\left|\int_{\lambda_0-1/n}^{\lambda_0+1/n}\right|\leqq\int_{-1/n}^{1/n}|\lambda|Cn^2\eta(\lambda)d\lambda=O(n^{-\mu})$$

since $\eta(\lambda)=F(\lambda+\lambda_0)/\lambda-\gamma=O(|\lambda|^\mu)$ by (19) and since (Zygmund [12], II, p. 60) $|DK_n^\alpha(t)|\leqq Cn^2$. The other integrals satisfy, e.g.,

$$\left|\int_0^{\lambda_0-1/n}\right|\leqq C'n^{2-\alpha}\int_0^{\lambda_0-1/n}|\lambda-\lambda_0|^{1-\alpha}\eta(\lambda-\lambda_0)d\lambda=O(n^{2-\alpha})+O(n^{-\mu})$$

by (12) and (19).

THEOREM 4. *Let $X(t)$ be a stationary Gaussian process with 0 mean; let $\lambda_0$ be a point where the spectral density $f(\lambda)$ is a generalized function satisfying (19). Then for $\alpha$ such that $1<\alpha\leqq 2$, the estimator*

$$\hat f_{MN}(\lambda)=(K_M^\alpha * I_N)(\lambda)$$

*satisfies, for $M=[N^{1/(2\beta+3)}]$, $\beta=\min(\alpha-1,\mu)$*

$$\mathrm{E}\,[\hat f_{MN}(\lambda_0)-\gamma]^2=O(N^{-2\beta/(2\beta+3)}).$$

*If $f(\lambda)=a\delta(\lambda-\lambda_1)+g(\lambda)$ where $g$ is continuous in a neighborhood of $\lambda_1$ and $a>0$, then*

$$\mathrm{E}\,[\hat f_{MN}(\lambda_1)]\geqq\frac{aM}{\pi(1+\alpha)}+g(\lambda_1).$$

*Moreover $\hat a=\hat f'_{MN}(\lambda_1)((\alpha+1)/(M+\alpha+1))$ is an asymptotic unbiased estimator for the strength $a$ of the periodicity at $\lambda_1$. Its bias is $O(M/N)$.*

The proof involves a number of straightforward modifications to the proof of Theorem 1. The function $A(\lambda,\mu)$ used in calculation of the var $\hat{f}_{MN}$ (9) can be used to obtain a bound on the latter. It is found by integration by parts and Schwarz's inequality to be

$$\text{(22)} \qquad \operatorname{var}[f_{MN}(\lambda)] \leqq 2\pi C_0 A^2(\pi-\lambda, \pi-\lambda) + \|F\|_2^2 \left\| \frac{\partial A}{\partial \lambda}(\lambda, \lambda) \right\|_2^2 .$$

The first term may be shown to be $O(M^{-2}N^{-2})$ while the second, by using a calculation similar to that following (9), may be seen to be $O(M^3N^{-1})$.

The bias term may be estimated by means of Lemma 2, and the decomposition used previously to be

$$\text{(23)} \qquad \mathrm{E}\, \hat{f}_{MN}(\lambda) - f(\lambda) = O(M^2N^{-1} \log N) + O(M^{-\beta})$$

where $\beta = \min(\alpha-1, \mu)$. By choosing $M$ appropriately the first conclusion follows. The second conclusion is straightforward. The third conclusion is a consequence of the result in Zygmund [12], II, p. 63.

*Example* 5.1. Let $F(\lambda) = \lambda^2(\cos 1/\lambda + \operatorname{sgn} \lambda) + \lambda$ in some neighborhood of 0. Then $f(\lambda) = F'(\lambda) = 2\lambda \cos 1/\lambda + 2|\lambda| + \sin 1/\lambda + 1$, $\lambda \neq 0$ and $F(\lambda)/\lambda \to 1$ as $\lambda \to 0$. Hence $f$ has value 1 at 0 and $|F(\lambda)/\lambda - 1| \leqq 2|\lambda|$ so that the hypothesis of Theorem 4 is satisfied at 0 for $\mu = 1$. Yet $f(\lambda)$ is not even continuous at 0.

*Remark*. For another approach to the estimation of the density in the presence of periodic components see Priestly [9]. His approach assumes that the process is of the form

$$X(t) = A \cos(\lambda_0 t + \phi) + Y(t)$$

where $Y(t)$ is a linear process. This implies that $f(\lambda)$ is, except for the delta function at $\lambda_0$, a $C^\infty$ function, and hence is a much more restrictive hypothesis. Moreover the presence of the delta function at $\lambda_0$ must be inferred from another estimator since his estimator is independent of $A$.

UNIVERSITY OF WISCONSIN-MILWAUKEE

## REFERENCES

[1] Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, Wiley, New York.
[2] Bartlett, M. S. (1948). Smoothing periodgrams from time series with continuous spectra, *Nature*, **161**, 686-687.
[3] Brillinger, D. R. (1973). Estimation of the mean of a stationary time series by sampling, *J. Appl. Prob.*, **10**, 419-431.
[4] Brillinger, D. R. (1975). *Time Series, Data Analysis and Theory*, Holt, Rinehart, and Winston, New York.
[5] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*,

Wiley, New York.
[6] Jones, R. H. (1971). Spectrum estimation with missing observations, *Ann. Inst. Statist. Math.*, **23**, 387-398.
[7] Koopmans, L. H. (1974). *The Spectral Analysis of Time Series*, Academic Press, New York and London.
[8] Parzen, E. (1967). *Time Series Analysis Papers*, Holdon-Day, San Francisco.
[9] Priestly, M. B. (1964). Estimation of the spectral density function in the presence of harmonic components, *J. R. Statist. Soc.*, B, **26**, 123-132.
[10] Schwartz, L. (1957). *Théorie des distributions I*, Hermann, Paris.
[11] Walter, G. and Blum, J. (1979). Probability density estimation using delta sequences, *Ann. Statist.*, **7**, 328-340.
[12] Zygmund, A. (1959). *Trigonometric Series*, I, II, Cambridge.