The Annals of Statistics
1985, Vol. 13, No. 3, 998–1010

A SIEVE METHOD FOR THE SPECTRAL DENSITY[1]

By Yun-shyong Chow and Ulf Grenander

Academia Sinica and Brown University

We suggest a sieve for the estimation of the spectral density of a Gaussian stationary stochastic process. In contrast to the standard periodogram-based estimates this one aims at exploiting the full Gaussian nature of the process.

1. Introduction. Consider a real stationary Gaussian stochastic process $\{x_t; t \in \mathbb{Z}\}$ on $(\Omega, \mathscr{F}, P)$ with mean value zero and covariances

$$r_t = E(x_s x_{s+t}) = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda)\, d\lambda \tag{1.1}$$

corresponding to an absolutely continuous spectrum with the spectral density f. Given a realization $x = (x_1, x_2, \cdots, x_n)$, we shall study the estimation of f.

Introducing the well-known *periodogram*

$$I_n(\lambda) = (1/2\pi n)\left| \textstyle\sum_{t=1}^{n} x_t e^{-it\lambda} \right|^2, \quad \lambda \in (-\pi, \pi), \tag{1.2}$$

consider an estimate of standard form

$$f_n^*(\lambda) = \int_{-\pi}^{\pi} w_n(\lambda - \mu) I_n(\mu)\, d\mu$$

where w_n is a nonnegative weight function. To make f_n^* consistent, one must let w_n contract to a δ-function at 0 as $n \to \infty$, and this must be done slowly enough. Such estimates are in wide use and their properties are well understood; in particular, one knows much about their asymptotic behavior as $n \to \infty$ (see Parzen, 1967).

As a function of x such an estimate is a quadratic form. To study its variance, for example to establish consistency, it is enough to know the moments up to order 4 for the process. This knowledge is of course available in the Gaussian case as soon as f is specified. When we use an estimate of this type, we do not take full advantage of the assumed Gaussianness, but we shall try to do this now by a very different method.

A natural approach would be to derive an estimate by appealing to the method of maximum likelihood, but this leads to two serious difficulties. Indeed, the

Received July 1984; revised January 1985.

[1] This work was partially supported by the U.S. Army Research Office, Contract DAAG-29-83-K-0116.

AMS 1980 *subject classifications.* Primary 62M15; secondary 62G05.

Key words and phrases. Spectral density estimation, stationary Gaussian process, maximum likelihood method, strong consistency.

likelihood can be written as

$$L_n = L_n(x, f) = \frac{1}{(2\pi)^{n/2}\sqrt{\det R_n(f)}} \exp[-\tfrac{1}{2}x^T R_n^{-1}(f)x]$$

in terms of the covariance matrix

$$R_n(f) = (r_{s-t};\ s, t = 1, 2, \cdots, n). \tag{1.3}$$

Note the complicated dependence of L_n in terms of the unknown parameter f due to the appearance both of the determinant of $R_n(f)$ and of its inverse. Therefore, the analytic treatment of L_n appears formidable.

To overcome this first of two obstacles, we shall appeal to *Toeplitz theory* (see Grenander and Szegö, 1984, Section 7.4) and use the approximation

$$\log L_n \sim -\frac{n}{4\pi} \int_{-\pi}^{\pi} [\log(2\pi)^2 f(\lambda) + I_n(\lambda)/f(\lambda)]\, d\lambda = \log \tilde{L}_n.$$

It is known (see Dzhaparidze and Yaglom, 1983) that for moving average processes of finite order

$$\lim_n (1/n)(\log L_n - \log \tilde{L}_n) = 0, \quad n \to \infty, \text{ in probability}$$

But this is not enough, since if we just solve the maximum problem *for any* f *in* $L^1(-\pi, \pi)$

$$\max_f \tilde{L}_n(x; f),$$

we get the estimate $f(\lambda) = I_n(\lambda)$. It is well known that this is not even consistent and cannot be accepted. We must exercise more care in choosing the function space over which we maximize.

We want to maximize the functional

$$F_n(g) = \int_{-\pi}^{\pi} [\log g(\lambda) - I_n(\lambda) g(\lambda)]\, d\lambda \tag{1.4}$$

where we have introduced $g = 1/f$, and shall do this over a *sieve* of the form

$$S_\mu = \left\{ g \geq 0: \int_{-\pi}^{\pi} (g^{(p)}(\lambda))^2\, d\lambda \leq \frac{1}{\mu} \right\}.$$

Here μ is the mesh size which will be made to tend to zero as $n \to \infty$ at some rate that will be determined later. The values of p that we have in mind are $p = 1$ and 2, but our method of analysis is designed to be applicable to other values and to other variations of the sieve.

To put the above in a general perspective, one can view the periodogram-based estimator as an instance of *regularization in abstract parameter space* (see Grenander, 1981, Section 7.3), while the type of estimate to be studied in this paper belongs to the *sieve type estimate* (ibid., Chapter 8). We are not aware of any study of our estimate in the time-series literature; the closest one that we have seen is in Wahba (1980).

Since our functional $F_n(g)$ is not of linear-quadratic type, we shall encounter analytical difficulties that do not arise otherwise due to the fact that the Euler equations will not be linear. We shall see that they can be dealt with successfully, which may open the way for other applications of the same idea to other estimates.

2. Results. We shall assume throughout this paper that the spectral density f of the real stationary Gaussian process $(x_t: t \in \mathbb{Z})$ satisfies

$$\sum_{u=-\infty}^{\infty} |ur_u| < \infty, \tag{2.1}$$

where the covariances r_u of the process are defined in (1.1).

$$f(\lambda) > 0 \quad \text{for all} \quad \lambda \in (-\pi, \pi). \tag{2.2}$$

Note that (2.1) and the realness of (x_t) imply that f can be regarded as a symmetric, continuously differentiable function on the unit circle. Therefore, there exist by (2.2) constants m, M such that

$$0 < m \le f(\lambda) \le M, \quad \text{for all} \quad \lambda \in [-\pi, \pi]. \tag{2.3}$$

It is probable that (2.1) and (2.2) can be weakened, but we shall not discuss this here.

Under the assumptions (2.1) and (2.2), it is known (Hannan and Nicholls, 1977; Taniguchi, 1979, 1980) that

THEOREM 2.1. *Let $h(\lambda)$ be a continuous symmetric function on $[-\pi, \pi]$. Then with probability* 1,

$$\int_{-\pi}^{\pi} h(\lambda)(I_n(\lambda) - f(\lambda))\, d\lambda \to_n 0 \tag{2.4}$$

$$\int_{-\pi}^{\pi} h(\lambda)(I_n^2(\lambda) - 2f^2(\lambda))\, d\lambda \to_n 0 \tag{2.5}$$

$$\int_{-\pi}^{\pi} h(\lambda)[\log(I_n(\lambda)e^{\gamma}) - \log f(\lambda)]\, d\lambda \to_n 0 \tag{2.6}$$

where γ is the Euler's constant. Moreover $\sqrt{n}\int_{-\pi}^{\pi} h(\lambda)(I_n(\lambda) - f(\lambda))\, d\lambda$ converges weakly to the normal distribution $N(0, 4\pi \int_{-\pi}^{\pi} h^2(\lambda) f^2(\lambda)\, d\lambda)$.

(2.5) implies that with probability 1 the sequence $\int_{-\pi}^{\pi} I_n^2(\lambda)\, d\lambda$ is bounded. Since $L^1(-\pi, \pi)$ is separable, we easily have the following

COROLLARY 2.2. *Let W be the set of all $\omega \in \Omega$ satisfying* (2.5), (2.6) *with $h(\lambda) \equiv 1$, and satisfying* (2.4) *for every continuous symmetric function h on $[-\pi, \pi]$. Then $P(W) = 1$.*

From now on we shall consider, which is also sufficient, only $\omega \in W$.

By using the theory of Lagrange multipliers, we are led to the following equivalent form of problem (1.4)

$$\text{(2.7)} \quad \text{maximize } F_{n,\varepsilon}(g) = \int_{-\pi}^{\pi} (\log g(\lambda) - I_n(\lambda) g(\lambda))\, d\lambda - \varepsilon \int_{-\pi}^{\pi} (g^{(p)}(\lambda))^2\, d\lambda$$

where $p \geq 1$, $\varepsilon > 0$ are constants and the maximum is taken over

$$A = \{g \geq 0: g^{(i)} \in L^2(-\pi, \pi),\ 0 \leq i \leq p\}.$$

Note that each $g \in A$ is continuous.

The right-hand side of (2.7) can be rewritten as

$$\text{(2.7)}' \quad \begin{aligned} F_{n,\varepsilon}(g) = & \int_{-\pi}^{\pi} [\log(I_n(\lambda) g(\lambda)) - I_n(\lambda) g(\lambda)]\, d\lambda \\ & - \varepsilon \int_{-\pi}^{\pi} (g^{(p)}(\lambda))^2\, d\lambda - \int_{-\pi}^{\pi} \log I_n(\lambda)\, d\lambda. \end{aligned}$$

The following lemma is easy.

LEMMA 2.3. *The function* $y(x) = \log x - x$ *is strictly concave on* $(0, \infty)$ *and attains its unique maximum at* $x = 1$.

Hence, by Corollary 2.2, $F_{n,\varepsilon}(g)$ is bounded above with probability 1. Then by using convexity and the calculus of variations it is not difficult to show the following.

PROPOSITION 2.4. *Assume* $\varepsilon > 0$. *Then with probability* 1 *problem* (2.7) *has a unique solution* $g_{n,\varepsilon}$ *which satisfies*

(i) $g_{n,\varepsilon}$ *is positive, symmetric and* C^{2p}*-continuous.*

(ii) $g_{n,\varepsilon}$ *is the unique positive solution to the following differential equation*

$$(g(\lambda))^{-1} - I_n(\lambda) - (-1)^p 2\varepsilon g^{(2p)}(\lambda) = 0$$

with $g^{(p+i)}(\pm\pi) = 0,\ 0 \leq i \leq p - 1$.

(iii) $\int_{-\pi}^{\pi} (g_{n,\varepsilon}(\lambda))^{-1}\, d\lambda = \int_{-\pi}^{\pi} I_n(\lambda)\, d\lambda$.

(iv) $2\pi = \int_{-\pi}^{\pi} I_n(\lambda) g_{n,\varepsilon}(\lambda)\, d\lambda + 2\varepsilon \int_{-\pi}^{\pi} (g_{n,\varepsilon}^{(p)}(\lambda))^2\, d\lambda$. *In particular,* $\int_{-\pi}^{\pi} (g_{n,\varepsilon}^{(p)}(\lambda))^2\, d\lambda \leq \pi/\varepsilon$.

(v) *Mean value property*: $\min_\lambda I_n(\lambda) \leq (g_{n,\varepsilon}(\lambda))^{-1} \leq \max_\lambda I_n(\lambda)$.

(vi) *For any approximating sequence* g_m *to the problem* (2.7), *i.e.,* $\lim_m F_{n,\varepsilon}(g_m) = \max F_{n,\varepsilon}(g)$, *we have* $\lim_m \| g_m - g_{n,\varepsilon} \|_\infty = 0$.

The second integral on the right-hand side of (2.7) can be regarded as a penalty term aiming at making $g_{n,\varepsilon}$ smooth. The constant ε will converge as $n \to \infty$ to 0, so that $(g_{n,\varepsilon}(\lambda))^{-1}$ will be close to $I_n(\lambda)$ and thus become consistent in some sense. The convergence rate to 0 of ε will be determined later.

Since $I_n(\lambda)$ converges weakly to $f(\lambda)$ for each $\omega \in W$, the following problem is apparently related to problem (2.7)

$$\text{(2.8)} \quad \text{maximize } F_\varepsilon(g) = \int_{-\pi}^{\pi} (\log g(\lambda) - f(\lambda)g(\lambda))\, d\lambda - \varepsilon \int_{-\pi}^{\pi} (g^{(p)}(\lambda))^2\, d\lambda$$

where the maximum is taken over A if $\varepsilon > 0$ and over

$$A' = \{g \geq 0: g \in L^1(-\pi, \pi)\}$$

in case $\varepsilon = 0$. By the same argument we have

PROPOSITION 2.5. *For $\varepsilon > 0$ the problem* (2.8) *has a unique solution g_ε which satisfies*

(i) *g_ε is positive, symmetric and C^{2p}-continuous.*
(ii) *g_ε is the unique positive solution to the following differential equation*

$$(g(\lambda))^{-1} - f(\lambda) - (-1)^p\, 2\varepsilon g^{(2p)}(\lambda) = 0$$

with $g^{(p+i)}(\pm\pi) = 0,\ 0 \leq i \leq p - 1$.
(iii) $\int_{-\pi}^{\pi} (g_\varepsilon(\lambda))^{-1}\, d\lambda = \int_{-\pi}^{\pi} f(\lambda)\, d\lambda$.
(iv) $2\pi = \int_{-\pi}^{\pi} f(\lambda)g_\varepsilon(\lambda)\, d\lambda + 2\varepsilon \int_{-\pi}^{\pi} (g_\varepsilon^{(p)}(\lambda))^2\, d\lambda$. *In particular*, $\int_{-\pi}^{\pi} (g_\varepsilon^{(p)}(\lambda))^2\, d\lambda \leq \pi/\varepsilon$.
(v) *Mean value property*: $m \leq (g_\varepsilon(\lambda))^{-1} \leq M$.
(vi) *For any approximating sequence g_m to the problem* (2.8), *we have* $\lim_m \| g_m - g_\varepsilon \|_\infty = 0$.

In case $\varepsilon = 0$, the problem (2.8) *has a unique solution $g_0(\lambda) = (f(\lambda))^{-1}$ and any approximating sequence g_m satisfies*

$$\lim_m \int_{-\pi}^{\pi} | g_m(\lambda)f(\lambda) - 1 |\, d\lambda = 0.$$

Note that the equation above is still true if we assume only that f, $\log f \in L^1(-\pi, \pi)$. In some sense $F_0(g)$ measures the distance between $1/g$ and f, i.e., $F_0(g)$ is sort of an entropy function.

In view of the previous two propositions, we would expect that in some sense $1/g_{n,\varepsilon}$ is close to $1/g_\varepsilon$, and then close to f if $\varepsilon = \varepsilon(n)$ is properly chosen. This vague idea will be implemented in the following.

Because of Corollary 2.2 and Proposition 2.4(iii),(iv), with probability 1, $\{g_{n,\varepsilon}: n = 1, 2, \cdots\}$ is relatively compact in $L^2(-\pi, \pi)$ and in $C[-\pi, \pi]$ also. Then by applying Fatou's lemma it is not hard to see

$$\text{(2.9)} \qquad \liminf_n F_\varepsilon(g_{n,\varepsilon}) \geq \max F_\varepsilon(g) \quad \text{with probability 1.}$$

Hence by Proposition 2.5(vi), for each fixed $\varepsilon > 0$

$$\text{(2.10)} \qquad P(\| g_{n,\varepsilon} - g_\varepsilon \|_\infty \to_n 0) = 1.$$

On the other hand, since A is a dense subset of A' with respect to L^1 norm and for each $g \in A$, $F_\varepsilon(g_\varepsilon) \geq F_\varepsilon(g) \to_\varepsilon F_0(g)$; therefore

$$\liminf_\varepsilon F_0(g_\varepsilon) \geq \liminf_\varepsilon F_\varepsilon(g_\varepsilon) \geq \max_{A'} F_0(g) \tag{2.11}$$

and then by Proposition 2.5

$$\lim_\varepsilon \int_{-\pi}^{\pi} | g_\varepsilon(\lambda) f(\lambda) - 1 | \, d\lambda = 0.$$

From (2.10) and Borel-Cantelli's lemma, it is theoretically possible to choose $\varepsilon(n)$ such that

$$P\left(\int_{-\pi}^{\pi} | g_{n,\varepsilon(n)}(\lambda) f(\lambda) - 1 | \, d\lambda \to_n 0\right) = 1. \tag{2.12}$$

In particular, $(g_{n,\varepsilon(n)}(\lambda))^{-1}$ converges in measure to $f(\lambda)$ even in the product space $[-\pi, \pi] \times \Omega$.

Since with probability 1

$$\int_{-\pi}^{\pi} (g_{n,\varepsilon}(\lambda))^{-1} \, d\lambda = \int_{-\pi}^{\pi} I_n(\lambda) \, d\lambda \to_n \int_{-\pi}^{\pi} f(\lambda) \, d\lambda,$$

and

$$E \int_{-\pi}^{\pi} I_n(\lambda) \, d\lambda = \int_{-\pi}^{\pi} f(\lambda) \, d\lambda$$

by Scheffé's theorem

$$\begin{cases} P\left(\int_{-\pi}^{\pi} | (g_{n,\varepsilon(n)}(\lambda))^{-1} - f(\lambda) | \, d\lambda \to_n 0\right) = 1 \quad \text{and} \\ E\left(\int_{-\pi}^{\pi} | (g_{n,\varepsilon(n)}(\lambda))^{-1} - f(\lambda) | \, d\lambda\right) \to_n 0, \end{cases} \tag{2.13}$$

so $(g_{n,\varepsilon(n)})^{-1}$ will be strongly consistent in L^1 sense.

Because Corollary 2.2 shows that $I_n(\lambda)$ converges only weakly to $f(\lambda)$ with probability 1, it seems difficult to estimate the rate of convergence in (2.10). Namely, $P(\| g_{n,\varepsilon} - g_\varepsilon \|_\infty > \delta)$ in terms of ε, n, δ. So we shall use another approach.

If we can show that there exists a sequence $\varepsilon(n)$ such that

$$P(\lim_n F_0(g_{n,\varepsilon(n)}) = \max_{A'} F_0(g)) = 1, \tag{2.14}$$

then, by Proposition 2.5, (2.12) holds, and so does (2.13). That is, $(g_{n,\varepsilon(n)})^{-1}$ is a strongly L^1-consistent estimator to f.

In order to verify (2.14), let us introduce

$$D_{n,\varepsilon} = \max | F_{n,\varepsilon}(g) - F_\varepsilon(g) |$$

where the maximum is taken over

$$B_\varepsilon = \left\{ g \geq 0: \ g \text{ symmetric and } \int_{-\pi}^{\pi} (g'(\lambda))^2 \, d\lambda \leq \pi/\varepsilon, \ \| g \|_\infty \leq 8/\sqrt{\varepsilon} \right\}.$$

By using the Fourier series expansion of $g_{n,\varepsilon}(\lambda) = \sum a_k \exp(ik\lambda)$ and Proposition 2.4(iii),(iv), it is not hard to show that

$$\int_{-\pi}^{\pi} (g'_{n,\varepsilon}(\lambda))^2 \, d\lambda = 2\pi \sum |ka_k|^2 \le 2\pi \sum |k^p a_k|^2$$

$$= \int_{-\pi}^{\pi} (g^{(p)}_{n,\varepsilon}(\lambda))^2 \, d\lambda \le \pi/\varepsilon,$$

$$\textstyle\sum_{k\neq 0} |a_k| \le (\sum |ka_k|^2 \cdot \sum_{k\neq 0} k^{-2})^{1/2} \le 2/\sqrt{\varepsilon},$$

$$2\pi(a_0 - \textstyle\sum_{k\neq 0} |a_k|)^{-1} \ge \displaystyle\int_{-\pi}^{\pi} (g_{n,\varepsilon}(\lambda))^{-1} \, d\lambda \to_n \int_{-\pi}^{\pi} f(\lambda) \, d\lambda.$$

Hence for n large

$$a_0 \le 2/\sqrt{\varepsilon} + 4\pi\left(\int_{-\pi}^{\pi} f(\lambda) \, d\lambda\right)^{-1},$$

$$\| g_{n,\varepsilon} \|_\infty \le \sum |a_k| \le 4/\sqrt{\varepsilon} + 4\pi\left(\int_{-\pi}^{\pi} f(\lambda) \, d\lambda\right)^{-1}.$$

Thus for $\varepsilon > 0$ and small

$$P(g_{n,\varepsilon} \in B_\varepsilon \text{ for } n \text{ large}) = 1. \tag{2.15}$$

Similarly $g_\varepsilon \in B_\varepsilon$. Therefore by the definition of $g_{n,\varepsilon}$, g_ε

$$\begin{aligned} F_0(g_{n,\varepsilon}) &\ge F_\varepsilon(g_{n,\varepsilon}) \ge F_{n,\varepsilon}(g_{n,\varepsilon}) - D_{n,\varepsilon} \\ &\ge F_{n,\varepsilon}(g_\varepsilon) - D_{n,\varepsilon} \ge F_\varepsilon(g_\varepsilon) - 2D_{n,\varepsilon}. \end{aligned} \tag{2.16}$$

Define $\xi_n(\lambda) = \int_0^\lambda [I_n(t) - E(I_n(t))] \, dt$. Then integrating by parts

$$\begin{aligned} &F_\varepsilon(g) - F_{n,\varepsilon}(g) \\ &= \int_{-\pi}^{\pi} (I_n(\lambda) - f(\lambda))g(\lambda) \, d\lambda \\ &= 2\int_0^{\pi} (I_n(\lambda) - EI_n(\lambda))g(\lambda) \, d\lambda + \int_{-\pi}^{\pi} (EI_n(\lambda) - f(\lambda))g(\lambda) \, d\lambda \\ &= 2\xi_n(\pi)g(\pi) - 2\int_0^{\pi} \xi_n(\lambda)g'(\lambda) \, d\lambda + \int_{-\pi}^{\pi} (EI_n(\lambda) - f(\lambda))g(\lambda) \, d\lambda. \end{aligned} \tag{2.17}$$

By computing the moments of $\xi_n(\lambda)$, it can be shown that

LEMMA 2.6. *For $s = 1, 2, \cdots$, $E(D_{n,\varepsilon}^{2s}) \le C_s M^{2s}/(n\varepsilon)^s$, where C_s is a constant depending only on s and M is defined in* (2.3).

Let c, δ be two constants with $0 < \delta < 1$, and

$$\varepsilon(n) = cn^{\delta-1}. \tag{2.18}$$

If s is large enough, $E(\sum_n D^{2s}_{n,\varepsilon(n)}) < \infty$. Then

$$P(D_{n,\varepsilon(n)} \to_n 0) = 1$$

and (2.14) follows from (2.16) and (2.11). Hence we have

THEOREM 2.7. *Let $\varepsilon(n)$ be given in* (2.18). *Then* (2.13) *is true. That is $(g_{n,\varepsilon(n)})^{-1}$ is a strongly L^1-consistent estimator to the spectral density f.*

It can also be shown that

THEOREM 2.8. *Let $\varepsilon(n) = cn^{\delta-1/3}$, where c, δ are constants with $0 < \delta < 1/3$. Then*

$$P(\| (g_{n,\varepsilon(n)})^{-1} - f \|_\infty \to_n 0) = 1$$

if we assume besides (2.1) *and* (2.2) *that $1/f \in A$.*

Note that in case $p = 1$ the additional condition $1/f \in A$ automatically holds under the assumptions (2.1) and (2.2).

Finally, as a consequence of Theorem 2.7, we have the following result for the original maximization problem (1.4).

THEOREM 2.9. (i) *Problem* (1.4) *has a unique solution $h_{n,\mu}$ which is positive, continuous, symmetric and satisfies*

$$\int_{-\pi}^{\pi} (h_{n,\varepsilon}(\lambda))^{-1}\, d\lambda = \int_{-\pi}^{\pi} I_n(\lambda)\, d\lambda$$

(ii) *Let $\mu(n) = cn^{\delta-1}$, where c, δ are two constants with $0 < \delta < 1$. Then* (2.13) *holds with $g_{n,\varepsilon(n)}$ replaced by $h_{n,\mu(n)}$. That is $(h_{n,\mu(n)})^{-1}$ is a strongly L^1-consistent estimator to the spectral density f.*

REMARK. To study the computational aspects of the suggested estimator, a preliminary computer experiment was performed. The algorithm was based on direct maximization of the functional using a gradient method. This requires only modest computing effort. The reason why we did not instead base the algorithm on the differential equation in Proposition 2.4(ii) was that the chaotic behavior of the periodogram would make quadrature for ODE difficult.

In retrospect, this seems less convincing and we plan to supplement the earlier algorithm by solving the stochastic differential equation directly and apply a shooting method to satisfy the constraints on the solution.

3. Proof of Proposition 2.4. First we shall prove the existence and uniqueness of the solution $g_{n,\varepsilon}$. Since $F_{n,\varepsilon}(g)$ is bounded above it is easy to see from (2.7)′ that if g_m is a maximizing sequence, i.e., $F_{n,\varepsilon}(g_m) \to_m \sup F_{n,\varepsilon}(g)$, then

$$\sup_m \int_{-\pi}^{\pi} (g_m^{(p)}(\lambda))^2 \, d\lambda < \infty \quad \text{and} \quad \sup_m \| g_m \|_\infty < \infty.$$

The second inequality is due to the fact that $\lim_{x\to\infty}(\log x - x) = -\infty$ and can be verified by considering the Fourier coefficients of g_m. Otherwise, we may assume for brevity that $\lim_m \| g_m \|_\infty = \infty$. Let $g_m(\lambda) = \sum_k a_{m,k}\exp(i\lambda k)$. From the proof of (2.14), we can see that $\sum_{k\neq 0} | a_{m,k} |$ is bounded in m. Because $g_m \geq 0$, $\lim a_{m,0} = \infty$ and thus $\lim g_m(\lambda) = \infty$ uniformly in λ. But then by using (2.7)′, $\lim_m F_{n,\varepsilon}(g_m) = -\infty$ and this is a contradiction. By Ascoli-Arzelà's theorem $\{g_m\}$ is relatively compact with respect to supnorm. For brevity we may assume g_m converges in supnorm to a certain function $g_{n,\varepsilon}$, which belongs to A by Fatou's lemma.

Write $\log g_m = \log^+ g_m - \log^- g_m$. Since $| \log^+ g_m - \log^+ g_{n,\varepsilon} | \leq | g_m - g_{n,\varepsilon} |$, it is then fairly easy to show by using Fatou's lemma that

$$F_{n,\varepsilon}(g_{n,\varepsilon}) \geq \lim \sup_m F_{n,\varepsilon}(g_m) = \sup F_{n,\varepsilon}(g)$$

so that $g_{n,\varepsilon}$ is a solution to problem (2.7). The uniqueness of the solution follows from Lemma 2.3 and the fact that $y(x) = x^2$ is a strictly convex function. Note that because of the uniqueness of the solution any maximizing sequence converges in supnorm to $g_{n,\varepsilon}$, not merely having a convergent subsequence. This proves (vi).

(i),(ii),(v). That $g_{n,\varepsilon}$ is symmetric follows from the uniqueness of the solution, because $I_n(\lambda)$ is a symmetric function. Now we shall show that $g_{n,\varepsilon}(\lambda) > 0$ for all $\lambda \in [-\pi, \pi]$.

From Lemma 2.3 and (2.7)′ we know that, pointwise, $I_n(\lambda)g_{n,\varepsilon}(\lambda)$ should be as close as possible to 1. Suppose $g_{n,\varepsilon}(\lambda_0) = 0$ for some λ_0. Then there exists a neighborhood $[\alpha, \beta]$ of λ_0 in which $g_{n,\varepsilon}(\lambda)$ is so small that $I_n(\lambda)h(\lambda)$ is more close to 1 than $I_n(\lambda)g_{n,\varepsilon}(\lambda)$, where the function h is defined by

$$h(\lambda) = \begin{cases} g_{n,\varepsilon}(\lambda), & \text{if } \lambda \notin [\alpha, \beta] \\ 2g_{n,\varepsilon}(\alpha) - g_{n,\varepsilon}(\lambda), & \text{if } \lambda \in [\alpha, \beta](g_{n,\varepsilon}(\alpha) = g_{n,\varepsilon}(\beta)). \end{cases}$$

Since $| h^{(p)}(\lambda) | = | g_{n,\varepsilon}^{(p)}(\lambda) |$, $F_{n,\varepsilon}(h) > F_{n,\varepsilon}(g_{n,\varepsilon})$. This is a contradiction and thus $g_{n,\varepsilon}$ is positive. Note that the same argument can be used to prove (v).

By applying Euler's equation in the calculus of variations (Gelfand and Fomin, 1963), it remains only to show that the differential equation in (ii) has only one solution. Suppose there are two solutions g_1 and g_2. By using the boundary conditions and integrating by parts, we can easily get

$$\int_{-\pi}^{\pi} (g_1 - g_2)^2(\lambda)(g_1(\lambda)g_2(\lambda))^{-1} \, d\lambda + 2\varepsilon \int_{-\pi}^{\pi} ((g_1 - g_2)^{(p)}(\lambda))^2 \, d\lambda = 0.$$

Thus $g_1 = g_2$ and the solution is unique.

(iii) and (iv) follow from integrating from $-\pi$ to π the differential equation in (ii), after multiplying it first by 1 and $g_{n,\varepsilon}$ respectively.

This completes the proof of the proposition.

4. Proof of Proposition 2.5. Only the case $\varepsilon = 0$ needs to be considered. But this is easy. Because in general the convergence to 0 of $\int_{-\pi}^{\pi} (h_m(x) - \log h_m(x) - 1)\, dx$ implies that of $\int_{-\pi}^{\pi} |\, h_m(x) - 1 \,|\, dx$. Note that the integrand of the former integral is nonnegative by Lemma 2.3.

5. Proof of Lemma 2.6. The proof is based on Ibragimov (1963). Denote by I_1, I_2 and I_3, respectively, those three terms on the right-hand side of (2.17). Since

$$\int_0^{\lambda} I_n(t)\, dt = (2\pi n)^{-1} \sum_{u,v=1}^{n} x_u x_v \int_0^{\lambda} (\cos(u - v)t)\, dt$$

is a quadratic form of normal variables, it is known from Lemma 8.4 of (Ibragimov, 1963) that for $s = 1, 2, \cdots$

$$E\,|\, \xi_n(\lambda) \,|^{2s} \le C_s M^{2s}/n^s,$$

where C_s is a constant depending only on s. Hence for each $g \in B_\varepsilon$,

$$E\,|\, I_1 \,|^{2s} \le C_s (16M)^{2s}/(n\varepsilon)^s, \tag{5.1}$$

and by the generalized Hölder's inequality

$$\begin{aligned} E\,|\, I_2 \,|^{2s} &\le 2^{2s}\pi^{s-1}\left(\int_0^{\pi} (g'(\lambda))^2\, d\lambda\right)^s \cdot E\left(\int_0^{\pi} (\xi_n(\lambda))^{2s}\, d\lambda\right) \\ &\le C_s (2\pi M)^{2s}/(n\varepsilon)^s. \end{aligned} \tag{5.2}$$

As to the nonrandom term I_3, it follows from (1.7) of Ibragimov (1963) that there exists a constant C such that

$$\begin{aligned} |\, I_3 \,| &\le (CM \log n) \int_{-\pi}^{\pi} |\, g'(\lambda) \,|\, d\lambda/n \le (2\pi\, CM \log n)/(n\sqrt{\varepsilon}) \\ &\le 4\pi CM/\sqrt{n\varepsilon}. \end{aligned} \tag{5.3}$$

Now combine (5.1), (5.2) and (5.3) together and the lemma follows with a different constant C_s.

6. Proof of Theorem 2.8. Under the assumption $1/f \in A$, $F_\varepsilon(g_\varepsilon) \ge F_\varepsilon(1/f)$. A simple computation shows

$$\int_{-\pi}^{\pi} (g_\varepsilon^{(p)}(\lambda))^2\, d\lambda \le \int_{-\pi}^{\pi} ((1/f)^{(p)}(\lambda))^2\, d\lambda$$

so that $\{ g_\varepsilon \}$ is relatively compact in the supnorm by Proposition 2.5(v) and Ascoli-Arzelà's theorem. Then

$$\|\, g_\varepsilon - 1/f \,\|_\infty \to_\varepsilon 0, \tag{6.1}$$

because we know already that $\|\, g_\varepsilon f - 1 \,\|_1 \to_\varepsilon 0$.

By Propositions 2.4 and 2.5

$$(g_{n,\varepsilon}(\lambda))^{-1} - I_n(\lambda) - (-1)^p 2\varepsilon g_{n,\varepsilon}^{(2p)}(\lambda) = 0$$

$$(g_\varepsilon(\lambda))^{-1} - f(\lambda) - (-1)^p 2\varepsilon g_\varepsilon^{(2p)}(\lambda) = 0.$$

Multiplying the difference of the above two equations by $(g_{n,\varepsilon} - g_\varepsilon)(\lambda)$ and then integrating, we shall have with probability 1

$$\int_{-\pi}^{\pi} (g_{n,\varepsilon} - g_\varepsilon)^2(\lambda)/(g_{n,\varepsilon}(\lambda) g_\varepsilon(\lambda))\, d\lambda + 2\varepsilon \int_{-\pi}^{\pi} ((g_{n,\varepsilon} - g_\varepsilon)^{(p)}(\lambda))^2\, d\lambda$$

$$(6.2) \qquad = \int_{-\pi}^{\pi} (g_{n,\varepsilon}(\lambda) - g_\varepsilon(\lambda))(f(\lambda) - I_n(\lambda))\, d\lambda$$

$$= (F_{n,\varepsilon}(g_{n,\varepsilon}) - F_\varepsilon(g_{n,\varepsilon})) - (F_{n,\varepsilon}(g_\varepsilon) - F_\varepsilon(g_\varepsilon)) \le 2D_{n,\varepsilon}.$$

Since Lemma 2.6 and Borel-Cantelli's lemma imply

$$P(D_{n,\varepsilon(n)}/\varepsilon(n) \to_n 0) = 1,$$

it follows from (2.15) and (6.2) that

$$P(\| g_{n,\varepsilon(n)} - g_{\varepsilon(n)} \|_2 \to_n 0, \quad \| g_{n,\varepsilon(n)}^{(p)} - g_{\varepsilon(n)}^{(p)} \|_2 \to_n 0) = 1.$$

Then by considering the Fourier coefficients of $g_{n,\varepsilon(n)}$, $g_{\varepsilon(n)}$

$$P(\| g_{n,\varepsilon(n)} - g_{\varepsilon(n)} \|_\infty \to_n 0) = 1.$$

Now use Proposition 2.5(v) and (6.1). The conclusion follows easily.

7. Proof of Theorem 2.9. Part (i) can be proved by using the same method as in Proposition 2.4. For part (ii) let us first note that similar to (2.15)

$$P(h_{n,\mu} \in B_\mu \text{ for } n \text{ large}) = 1.$$

Define $\varepsilon(n) = \pi\mu(n)$. Since $g_{n,\varepsilon(n)} \in S_{\mu(n)}$ and $P(D_{n,\mu(n)} \to_n 0) = 1$, Theorem 2.7 implies

$$F_0(h_{n,\mu(n)}) \ge F_n(h_{n,\mu(n)}) - D_{n,\mu(n)}$$

$$\ge F_n(g_{n,\varepsilon(n)}) - D_{n,\mu(n)}$$

$$\ge F_0(g_{n,\varepsilon(n)}) - 2D_{n,\mu(n)} \to_n \max_{g\in A'} F_0(g)$$

Now the conclusion follows from Proposition 2.5 and Scheffé's theorem.

Acknowledgements. The authors appreciate helpful discussions with S. Geman and D. McClure.

REFERENCES

DZHAPARIDZE, K. O. and YAGLOM, A. M. (1983). Spectral parameter estimation in time series analysis. In *Developments in Statistics* 4 (P. R. Krishnaiah, ed.). Academic, New York.

GELFAND, I. M. and FOMIN, S. V. (1963). *Calculus of Variations*. Prentice-Hall, Englewood Cliffs, New Jersey.

GRENANDER, U. (1981). *Abstract Inference.* Wiley, New York.
GRENANDER, U. and SZEGÖ, G. (1984). *Toeplitz Forms and their Applications*, 2nd ed. Chelsea, New York.
HANNAN, E. J. and NICHOLLS, D. F. (1977). The estimation of the prediction error. *J. Amer. Statist. Assoc.* **72** 834–840.
IBRAGIMOV, I. A. (1963). On estimation of the spectral function of a stationary Gaussian process. *Theory Probab. Appl.* **8** 366–401.
PARZEN, E. (1967). *Time Series Analysis Papers.* Holden-Day, San Francisco.
TANIGUCHI, M. (1979). On estimation of parameters of Gaussian stationary processes. *J. Appl. Probab.* **16** 575–591.
TANIGUCHI, M. (1980). On estimation of the integrals of certain functions of spectral density. *J. Appl. Probab.* **17** 73–83.
WAHBA, G. (1980). Automatic smoothing of the log periodogram. *J. Amer. Statist. Assoc.* **75** 122–132.

Appendix to the proof of Lemma 2.6

By Johnson and Kotz (1970)

$$E\exp(it\xi_n(\lambda)) = \textstyle\prod_1^n (1 - 2it\lambda_j^{(n)})^{-1/2}, \tag{1}$$

where $\lambda_j^{(n)}$ are the eigenvalues of the product matrix R_nA_n. Here r_n is the $n \times n$ covariance matrix defined in (1.3) and $A_n = (a_{uv})_{n\times n}$ with

$$a_{uv} = (2\pi n)^{-1}\int_0^\lambda \cos(u-v)t\,dt.$$

Since (Ibragimov, 1963)

$$\|R_n\| = \sup_{\|y\|=1}\textstyle\sum y_uy_vr_{u-v} = \sup_{\|y\|=1}\displaystyle\int_{-\pi}^{\pi} |\textstyle\sum_{k=1}^n y_ke^{ik\lambda}|^2 f(\lambda)\,d\lambda \le M$$

$$\|A_n\| = \sup_{\|y\|=1}\textstyle\sum y_uy_va_{uv} = (2\pi n)^{-1}\sup_{\|y\|=1}\displaystyle\int_0^{\lambda} |\textstyle\sum_{k=1}^n y_ke^{ik\lambda}|^2\,d\lambda$$

$$\le (2\pi n)^{-1}\sup_{\|y\|=1}\int_{-\pi}^{\pi} |\textstyle\sum_1^n y_ke^{ik\lambda}|^2\,d\lambda = n^{-1}.$$

we have

$$\max_j|\lambda_j^{(n)}| \le \|R_n\|\|A_n\| \le M/n.$$

Let χ_s be the cumulants of $\xi_n(\lambda)$, i.e.,

$$\log E\exp(it\xi_n(\lambda)) = \textstyle\sum_0^\infty \chi_s(it)^s/s!.$$

Then by (1)

$$\begin{aligned}|\chi_s| &= |2^{s-1}(s-1)!\textstyle\sum_{j=1}^n(\lambda_j^{(n)})^s| \\ &\le 2^{s-1}(s-1)!\textstyle\sum_{j=1}^n(\lambda_j^{(n)})^2\cdot(\max_j|\lambda_j^{(n)}|)^{s-2} \\ &\le 2^{s-1}(s-1)!\chi_2(M/n)^{s-2}.\end{aligned}$$

Since $\chi_2 = E\xi_n(\lambda)$, by Lemma 2.1 of (Ibragimov, 1963)

$$E\xi_n^2(\lambda) = \int_0^\lambda \int_0^\lambda G_n(u, v)\, du\, dv/(2\pi n)^2,$$

where

$$G_n(u, v) = \left[\int_{-\pi}^{\pi} \frac{\sin(n(\ell - u)/2)}{\sin((\ell - u)/2)} \cdot \frac{\sin(n(\ell - v)/2)}{\sin((\ell - v)/2)}\, d\ell\right]^2 + \left[\int_{-\pi}^{\pi} \frac{\sin(n(\ell - u)/2)}{\sin((\ell - u)/2)} \cdot \frac{\sin(n(\ell + v)/2)}{\sin((\ell + v)/2)}\, d\ell\right]^2.$$

Therefore,

$$E\xi_n^2(\pi) \le E\xi_n^2(\pi) = 4^{-1}\mathrm{Var}\left(\int_{-\pi}^{\pi} I_n(\lambda)\, d\lambda\right)$$

$$\le 4^{-1}\mathrm{Var}(\textstyle\sum_1^n x_t^2/2\pi n) = (4\pi n)^{-2} \cdot 2 \sum_{uv=1}^n r_{u-v}^2$$

$$\le \textstyle\sum_{-\infty}^{\infty} r_u^2/(2\pi n) \le \displaystyle\int_{-\pi}^{\pi} f^2(\lambda)\, d\lambda/n \le M^2/n$$

and then

$$|\chi_s| = 2^{s-1}(s-1)!M^s/n^{s-1}.$$

By using the relation between cumulants and moments and noting that $\chi_1 = 0$, we shall obtain

$$E(\xi_n(\lambda))^{2s} \le C_s M^{2s}/n^s$$

where the constant C_s depends only on s.

REFERENCE

JOHNSON, N. L. and KOTZ, S. (1970). *Continuous Univariate Distributions* 2. Wiley, New York.

INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI, TAIWAN
REPUBLIC OF CHINA

DIVISION OF APPLIED MATHEMATICS
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02912