The Annals of Statistics,
1983, Vol. 11, No. 4, 1206–1217

LIMITING BEHAVIOR OF FUNCTIONALS OF THE SAMPLE SPECTRAL DISTRIBUTION

BY DANIEL MACRAE KEENAN

Carnegie-Mellon University

The parameters of a stationary process can be viewed as functions of the spectral distribution function. This work concerns (estimators) parameters defined as integrals of m (≥ 1)-dimensional kernel functions with respect to the (sample) spectral distribution function. Conditions for asymptotic normality, almost sure convergence, and probability one bounds are derived for such estimators. The approach taken is based upon the reduction of an m-dimensional problem to one-dimension via consideration of a Frechet differential and its linearity. The probability one bound for the estimators is obtained by first establishing it ($O((n^{-1}\log n)^{1/2})$) for the difference of the sample and true spectral distribution functions in the supnorm and then showing that this rate is transferred to the estimators through integration.

1. Introduction. Let $\{X_i, -\infty < i < \infty\}$ be a strictly stationary stochastic process defined on a probability space $(\Omega, \mathscr{A}, P)$. If we are concerned with the marginal distribution of X_i, then estimators can be developed for parameters of the marginal distribution when the dependency dies out quickly. Sen (1972) and Yoshihara (1976) derive asymptotic results in this situation for U and V statistics under the assumptions of *-mixing and absolute regularity, respectively. However, if it is the dependency in which we are interested, then the marginal distribution is not our concern, but rather the joint distributions. For the covariance structure, we are interested in the second moment structure of the bivariate distributions. The (2nd order) frequency domain essentially considers simultaneously the second moment structure of all the bivariate distributions of (X_i, X_j), $i, j \in \mathbb{Z}$. Higher-order cumulant spectra consider more structure (and more variables, jointly). This paper concerns parameters and estimators defined as functionals of the spectral distribution function and of its estimates, respectively. A possible alternative to this is to assume that the process is a given ARMA model and using Akaike's Markovian representation (Akaike, 1974a), represent the univariate process as a multivariate $AR(p)$, for some $p \in N^+$, and apply the results of Sen (1972) and Yoshihara (1976), since the parameters of interest are associated with the marginal distribution of the new process. The order of the process could be determined by Akaike's AIC criterion (1974b) or Parzen's CAT criterion (1974). This is approximately the procedure which is at the foundation of robust estimation for time series models (see Martin, 1978). The approach of this paper does not require the parameter of interest to be determined by some finite dimensional joint distribution (e.g. a restriction to ARMA models nor a prior specification of the order) in order to obtain asymptotic results for estimators of parameters of the joint structure. In fact, the parameter may depend upon the entire process, e.g. parameters of the spectrum of an arbitrary linear process. Asymptotic distributions, probability one bounds, and almost sure convergence will be shown for certain functionals of the sample spectral distribution. A probability one bound for the sample spectral distribution function is also established.

2. Let $\{X_i, -\infty < i < \infty\}$ be a strictly stationary stochastic process with mean zero for which the kth order cumulant spectra, $f(\cdot)$, is finite, $k = 2, 3, \cdots$. We will assume

Received August 1982; revised May 1983.

AMS 1980 *subject classifications.* Primary 60G10; secondary, 62M15.

Key words and phrases. Spectral distribution function, probability one bounds, statistical differential approach.

throughout that

$$\sum_{v_1,v_2,\cdots,v_{k-1}=-\infty}^{\infty} |v_j| \, |c(v_1, v_2, \cdots, v_{k-1})| < \infty$$

for $j = 1, 2, \cdots, k-1$, $k = 2, 3, \cdots$, where $c(v_1, v_2, \cdots, v_{k-1})$ is the kth order cumulant of $\{X(0), X(v_1), X(v_2), \cdots, X(v_{k-1})\}$ (see Brillinger, 1975, Section 2.6). In the case of a Gaussian process this condition is satisfied if

$$\sum_{v=-\infty}^{\infty} |v| \, |c(v)| < \infty.$$

The absolutely continuous spectral distribution function of the X_n process will be denoted by $F(\lambda)$, $\lambda \in [0, 2\pi]$. Consider the functional

$$\theta(F) = \int_0^{2\pi} \cdots \int_0^{2\pi} h(\lambda_1, \lambda_2, \cdots, \lambda_m) \, dF(\lambda_1) \, dF(\lambda_2) \cdots dF(\lambda_m) \tag{2.1}$$

where h is real valued, symmetric in its m (≥ 1) arguments, and of bounded variation (whose definition is given in Section 3). If h were not symmetric, we could replace it by a sum over all permutations. For a sample $\{X_1, X_2, \cdots, X_n\}$, the sample spectral distribution is defined as

$$F_n(\lambda) = \frac{2\pi}{n} \sum_{0<(2\pi s/n)\leq\lambda} I_{xx}\left(\frac{2\pi s}{n}\right) \tag{2.2}$$

where $I_{xx}^{(n)}(\lambda)$ is the sample periodogram

$$I_{xx}^{(n)}(\lambda) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t \exp(-it^{\lambda}) \right|^2, \; \lambda \in [0, 2\pi]. \tag{2.3}$$

One estimator which we will consider is

$$\begin{aligned} \theta(F_n) &= \int_0^{2\pi} \cdots \int_0^{2\pi} h(\lambda_1, \lambda_2, \cdots, \lambda_m) \, dF_n(\lambda_1) \cdots dF_n(\lambda_m) \\ &= \left(\frac{2\pi}{n}\right)^m \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_m}) \prod_{j=1}^{m} I_{xx}^{(n)}(\xi_{i_j}). \end{aligned} \tag{2.4}$$

If $m = 1$, then $\theta(F_n)$ is of the general form of an estimator considered by Parzen (1957). Another empirical spectral distribution function is $F_n^*(\cdot)$, given by

$$F_n^*(\lambda) = \frac{\{\#\lambda_{(j)}^{(n)}{}^{,s}, \leq \lambda\} F_n(2\pi)}{n} \tag{2.5}$$

where $\lambda_{(j)}^{(n)} = \tilde{F}_n^{-1}(jF_n(2\pi)/n)$, $j = 1, 2, \cdots, n$, and $\tilde{F}_n(\cdot)$ is the piecewise linear version of $F_n(\cdot)$. Two additional estimators of $\theta(F)$ are

$$\begin{aligned} \theta(F_n^*) &= \int_0^{2\pi} \cdots \int_0^{2\pi} h(\lambda_1, \cdots, \lambda_m) \, dF_n^*(\lambda_1) \cdots dF_n^*(\lambda_m) \\ &= \frac{F_n(2\pi)}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} h(\lambda_{(i_1)}^{(n)}, \cdots, \lambda_{(i_m)}^{(n)}) \end{aligned} \tag{2.6}$$

and

$$U_n = \binom{n}{m}^{-1} F_n(2\pi) \sum h(\lambda_{(i_1)}^{(n)}, \cdots, \lambda_{(i_m)}^{(n)}). \quad 1 \leq i_1 < i_2 < \cdots < i_m \leq n. \tag{2.7}$$

The values $\{\lambda_{(j)}^{(n)}, j = 1, \cdots, n\}$ can be viewed as approximate order statistics from the spectral distribution F. It may be that estimators defined w.r.t. $F_n^*(\cdot)$ are computationally easier to work with. The last two estimators are of the form of the traditional V and U

statistics. Asymptotic normality, probability 1 bounds and almost sure convergence will be established for these three estimators.

3. Let K be $\{G \mid \theta(G) < \infty, G$ spectral distribution function on $[0, 2\pi]$, $G(2\pi) < \infty\}$ and $D[0, 2\pi]$ be the set of functions on $[0, 2\pi]$ with discontinuities of the first kind. Let $\|\cdot\|_\infty$ be the sup norm

$$\| H \|_\infty = \sup_{0 \le \lambda \le 2\pi} | H(\lambda) |$$

on $D[0, 2\pi]$. This norm does not induce the usual Skorokhod topology (see Billingsley, 1968, Chapter 3), however, that will not prove important to us. We could just as well work with the interval $[0, \pi]$, although there are notational conveniences in using $[0, 2\pi]$.

Let E be the linear space in $D[0, 2\pi]$ generated by $\{\Delta \mid \Delta = c(G - H), c \in R, G, H \in K\}$. The Gateaux differential (see Serfling, 1980, Section 2) of $\theta(F)$ at $F \in K$ in the direction of G is defined as

$$D_1(\theta(F), F, G)$$

$$= \lim_{\lambda \downarrow 0} \frac{\theta[F + \lambda(G - F)] - \theta(F)}{\lambda} \tag{3.1}$$

$$= m \int_0^{2\pi} \int_0^{2\pi} \cdots \int_0^{2\pi} h(\lambda_1, \lambda_2, \cdots, \lambda_m) \prod_{j=2}^{m} dF(\lambda_j)\, d[G(\lambda) - F(\lambda_1)]$$

$$= m\left[\int_0^{2\pi} h_1(\lambda_1)\, dG(\lambda_1) - \theta(F)\right] \tag{3.2}$$

where for $1 \le c \le m$

$$h_c(\lambda_1, \cdots, \lambda_c) = \int_0^{2\pi} \cdots \int_0^{2\pi} h(\lambda_1, \lambda_2, \cdots, \lambda_m) \prod_{j=c+1}^{m} dF(\lambda_j).$$

By the above definition of the Gateaux differential, $D_1(\theta(F), F, G)$ is a Frechet differential at F if

$$\theta(G_n) - \theta(F) - D_1(\theta(F), F, G_n) = o(\| G_n - F \|_\infty) \tag{3.3}$$

all sequences $\{G_n\}_{n=1}^{\infty}$ in K which converge to F in the sup norm. We will be assuming throughout that $D_1(\theta(F), F, G)$ is a Frechet differential w.r.t. the sup norm. Lemma 3.2 establishes (3.3) when in addition to bounded variation, h is assumed to be continuous. The generalization of bounded variation to more than one dimension which we will use is that h is of bounded variation in each component with the total variation norm for that component being integrable with respect to the other components, e.g.

$$\int_0^{2\pi} \cdots \int_0^{2\pi} \| h(\cdot, \lambda_2, \cdots, \lambda_m) \|_v \, d\lambda_2 \cdots d\lambda_m < \infty \quad \text{and}$$

$$T(h_c, k) = \sup_{\{\xi_{i_1}\}, \cdots, \{\xi_{i_k}\}} \sum_{i_1} \cdots \sum_{i_k} | \Delta_k h_c(\xi_{i_1}, \cdots, \xi_{i_k}, \lambda_{k+1}, \cdots, \lambda_c) | < \infty \tag{3.4}$$

for $1 \le k \le c \le m$, $(\lambda_{k+1}, \cdots, \lambda_c) \in [0, 2\pi]^{c-k}$ where Δ_k is defined recursively by

$$\Delta h_c(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_c}) = h_c(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_c}) - h_c(\xi_{i_1 - 1}, \xi_{i_2}, \cdots, \xi_{i_c})$$

and

$$\Delta_k h_c(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_c}) = \Delta_{k-1} h_c(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k}, \cdots, \xi_{i_c}) - \Delta_{k-1} h_c(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_k - 1}, \cdots, \xi_{i_c})$$

where $\{\xi_{i_1}\}, \cdots, \{\xi_{i_c}\}$ are finite partitions of $[0, 2\pi]$ (see Hobson, 1927, page 343–346). Note that by the symmetry of h it was not important that we went from left to right. The following well known result will be of assistance in proving Lemma 3.2.

LEMMA 3.1. *If l is a real (or complex) valued, continuous function of bounded variation on $[0, 2\pi]$ and K is a right continuous, bounded function on $[0, 2\pi]$ with $K(0) = 0$ then*

$$\left| \int_0^{2\pi} l(x)\, dK(x) \right| \le | l(2\pi)K(2\pi) | + \| l \|_v \| K \|_\infty$$

where $\| \cdot \|_v$ is the total variation norm.

PROOF. See Lemma 7.2.2B, page 254, Serfling (1980).

LEMMA 3.2. *For $\theta(F)$ defined by expression (2.1) with h being symmetric, continuous, and of bounded variation, (as defined above), $D_1(\theta(F), F, \Delta)$ given above is a Frechet differential w.r.t. the sup norm.*

PROOF. Let $\{G_n\}_{n=1}^\infty$ be a sequence in K which converges to F. Expression (3.3) can be written as

$$\theta(G_n) - \theta(F) - D_1(\theta(F), F, G_n)$$
$$= \sum_{c=2}^m \binom{m}{c} \left\{ \int_{[0,2\Pi]^c} h_c(\lambda_1, \cdots, \lambda_c) \prod_{j=1}^c d[G_n(\lambda_j) - F(\lambda_j)] \right\}.$$

By Dominated Convergence and bounded variation, h_c is also continuous and of bounded variation and using expression (3.4) we have that for $1 \le l \le c - 1$

$$\int_{[0,2\pi]^{c-l}} h_c(\lambda_1, \cdots, \lambda_c) \prod_{j=l+1}^c d[G_n(\lambda_j) - F(\lambda_j)]$$

is continuous and of bounded variation. By Lemma 3.1 and the definition of bounded variation

$$\left| \int_{[0,2\pi]^c} h_c(\lambda_1, \cdots, \lambda_c) \prod_{j=1}^c d[G_n(\lambda_j) - F(\lambda_j)] \right|$$
$$\le \left\| \int_{[0,2\pi]^{c-1}} h_c(\cdot, \lambda_2, \cdots, \lambda_c) \prod_{j=2}^c d[G_n(\lambda_j) - F(\lambda_j)] \right\|_v \| G_n - F \|_\infty$$
$$+ \left| \int_{[0,2\pi]^{c-1}} h_c(2\pi, \lambda_2, \cdots, \lambda_c) \prod_{j=2}^c d[G_n(\lambda_j) - F(\lambda_j)] \right| | G_n(2\pi) - F(2\pi) |$$
$$= \| G_n - F \|_\infty \sup_{\{\xi_{i_1}\}} \sum_{i_1}$$
$$\cdot \left| \int_{[0,2\pi]^{c-1}} [h_c(\xi_{i_1}, \lambda_2, \cdots, \lambda_c) - h_c(\xi_{i_1-1}, \lambda_2, \cdots, \lambda_c)] \prod_{j=2}^c d[G_n(\lambda_j) - F(\lambda_j)] \right|$$
$$+ \left| \int_{[0,2\pi]^{c-1}} h_c(2\pi, \lambda_2, \cdots, \lambda_c) \prod_{j=2}^c d[G_n(\lambda_j) - F(\lambda_j)] \right| | G_n(2\pi) - F(2\pi) |$$

where the sup is being taken over all partitions $\{\xi_{i_1}\}$ of $[0, 2\pi]$. Applying Lemma 3.1 recursively to the terms of the summand and the latter term we obtain

$$\left| \int_{[0,2\pi]^c} h_c(\lambda_1, \cdots, \lambda_c) \prod_{j=1}^c d[G_n(\lambda_j) - F(\lambda_j)] \right|$$
$$\le (\| G_n - F \|_\infty)^c \sup_{\{\xi_{i_1}\}, \cdots, \{\xi_{i_c}\}} \sum_{i_1, \cdots, i_c} | \Delta_c h_c(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_c}) | + o(\| G_n - F \|_\infty)$$

(the latter term being, due to the recursive evaluation of integrals at 2π). Therefore we have that

$$
\begin{aligned}
&|\,\theta(G_n) - \theta(F) - D_1\theta(F), F, G_n)\,| \\
&\qquad \le \textstyle\sum_{c=2}^{m} \binom{m}{c} T(h_c, c)(\|G_n - F\|_\infty)^c + o(\|G_n - F\|_\infty) \\
&\qquad = o(\|G_n - F\|_\infty).
\end{aligned}
$$

□

Note: In proving Lemma 3.2, the time series nature of the observations is not used. That is, this result holds for the general statistical differential approach, although an assumption of bounded variation in the general case (i.e. unbounded support on R) would be quite limiting since all the (nonconstant) polynomials and therefore, all moments and cumulants, are eliminated. For the case of compact support (e.g., $[0, 2\pi]$), such an assumption is quite reasonable. Also, it should be noted that h is not a function of the data but rather generates the nonrandom weights of the periodogram products (see expression (2.4)).

Let R_{1n} and R_{1n}^* be defined as the remainder terms

$$n^{1/2}R_{1n} = n^{1/2}[\theta(F_n) - \theta(F)] - n^{1/2}D_1(\theta(F), F, F_n) \tag{3.5}$$

$$n^{1/2}R_{1n}^* = n^{1/2}[\theta(F_n^*) - \theta(F)] - n^{1/2}D_1(\theta(F), F, F_n^*). \tag{3.6}$$

We will show that both $n^{1/2}R_{1n}$ and $n^{1/2}R_{1n}^*$ converge in probability to zero. We first need the following lemmas. The first is the statement of a well-known time series analogue of a Kolmogorov-Smirnov result. (See Grenander and Rosenblatt, 1957, Chapter 6, for an historical account)

LEMMA 3.3. *If $F_n(\cdot)$ is the sample spectral distribution (i.e., expression (2.2)) based upon a sample $\{X_1, X_2, \cdots, X_n\}$ from a strictly stationary process satisfying the assumption of Section 2, then $n^{1/2}\|F_n - F\|_\infty$ converges in distribution to $\|Y\|_\infty$ where $\{Y(\lambda), 0 \le \lambda \le 2\pi\}$ is a zero-mean Gaussian process whose sample paths are continuous*, w.p.1.

PROOF. See Theorem 7.63, page 258–259, Brillinger (1975) (or Theorem 4.3, Brillinger, 1969).

LEMMA 3.4. *For F_n^* and F_n defined by expressions* (2.5) *and* (2.2), *respectively, $n^{1/2}\|F_n^* - F_n\|_\infty = o(1)$ with probability one.*

PROOF. $\tilde{F}_n(\cdot)$ was defined to be the linear interpolated version of $F_n(\cdot)$. By the construction of $\tilde{F}_n$ we have

$$n^{1/2}\sup_\lambda |\tilde{F}_n(\lambda) - F_n(\lambda)| = n^{1/2}\sup_{1\le j\le n}\left|\frac{2\pi}{n} I_{xx}^{(n)}(\lambda_j)\right| = 2\pi \sup_{1\le j\le n}\left|\frac{I_{xx}^{(n)}(\lambda_j)}{n^{1/2}}\right|$$

for each $\omega \in \Omega$ where $\lambda_j = 2\pi j/n$. By Brillinger (1975), Theorem 5.3.2, we have that

$$\lim\sup_{n\to\infty}[\sup_\lambda I_{xx}^{(n)}(\lambda)/\log n] \le 2\sup_\lambda f_{xx}(\lambda)$$

with probability one. Since f is continuous and $\log n/n^{1/2} = o(1)$, we have

$$n^{1/2}\|\tilde{F}_n - F_n\| = o(1) \quad \text{w.p.1} \tag{3.7}$$

By the construction of F_n^* from $\tilde{F}_n$, we have

$$\sup_\lambda |F_n^*(\lambda) - \tilde{F}_n(\lambda)| \le F_n(2\pi)/n \quad \text{for all} \quad \omega \in \Omega.$$

Since $F_n(2\pi) - F(2\pi)$ is $o_p(1)$ and

$$P\{|F_n(2\pi) - F(2\pi)| > n^{1/2}\varepsilon\} \le \frac{1}{n\varepsilon^2} E[F_n(2\pi) - F(2\pi)]^2 \le \frac{1}{n^2\varepsilon^2}[\psi + O(n^{-1})]$$

(where ψ is given by expression (4.5) in the Lemma 4.1) we have that $(F_n(2\pi) - F(2\pi))/n^{1/2}$ converges completely to zero and thus $F_n(2\pi)/n^{1/2}$ converges to zero, w.p.1. Consequently,

$$n^{1/2}\| F_n^* - \tilde{F}_n \|_\infty = o(1) \quad \text{w.p.1} \tag{3.8}$$

and the result follows from expressions (3.7) and (3.8). □

Note: Again, it is assumed throughout the paper that h is such that $D_1(\theta(F), F, G)$ is a Frechet differential at $G = F$; Lemma 3.2 shows that it is sufficient to have h continuous and of bounded variation. For other h's the condition would need to be individually verified.

THEOREM 3.5. *For $F_n(\cdot)$ and $F_n^*(\cdot)$ given by expressions* (2.2) *and* (2.5) *we have*

$$n^{1/2}(\theta(F_n) - \theta(F)), \quad n^{1/2}(\theta(F_n^*) - \theta(F)) \quad \text{and} \quad n^{1/2}(U_n - \theta(F))$$

are all asymptotically normal with zero and variance $m\Lambda$ where Λ is given by

$$2\pi\left[\int_0^{2\pi} [h_1(2\pi - \alpha)\overline{h_1(\alpha)} + |h_1(\alpha)|^2] f_{xx}^2(\alpha)\, d\alpha + \int_0^{2\pi}\int_0^{2\pi} h_1(\alpha)\overline{h_1(\beta)} f_{xxxx}(\alpha, \beta, -\alpha)\, d\alpha\, d\beta\right], \tag{3.9}$$

assuming $0 < \Lambda < \infty$.

Note: If Λ were equal to zero, then a second order approximation, as with traditional U and V statistics, could be employed throughout.

PROOF. By Lemmas 6.2.2 A and B, Serfling (1980), page 218, $n^{1/2}R_{1n}$ and $n^{1/2}R_{1n}^*$ are each $o_p(1)$ if $n^{1/2}\| F_n - F\|_\infty$ and $n^{1/2}\| F_n^* - F\|_\infty$ are each $O_p(1)$, which was shown to be true by Lemmas 3.3 and 3.4 above. Therefore $n^{1/2}(\theta(F_n) - \theta(F))$ and $n^{1/2}(\theta(F_n^*) - \theta(F))$ have the same asymptotic distributions as $n^{1/2}D_1(\theta(F_n), F, F_n)$ and $n^{1/2}D_1(\theta(F_n^*), F, F_n^*)$, respectively (if ones exist). By expressions (3.2) and (2.2),

$$\begin{aligned} n^{1/2}D_1(\theta(F), F, F_n) &= n^{1/2}m \int_0^{2\pi} h_1(\lambda_1)\, dF_n(\lambda_1) - \theta(F) \\ &= n^{1/2}m\left[\left(\frac{2\pi}{n}\right) \sum_{j=1}^n h_1(\xi_j) I_{xx}^{(n)}(\xi_j) - \theta(F)\right] \end{aligned} \tag{3.10}$$

where $\xi_j = 2\pi j/n$. Since h is of bounded variation, so is h_1 and by Theorem 5.10.1, page 168, Brillinger (1975) the result follows for $n^{1/2}(\theta(F_n) - \theta(F))$. For each $n \ge 1$ and $\omega \in \Omega$, there is a Borel signed measure $\mu_{n,\omega}$ corresponding to

$$n^{1/2}[F_n^*(\cdot;\omega) - F_n(\cdot;\omega)],$$

the difference of two bounded, non-decreasing functions and therefore a function of nomalized bounded variation. By Lemma 3.4 $\mu_{n,\omega}$ converges weakly to the zero measure on $[0, 2\pi]$, for each $\omega \in \Omega$ (except for a set of measure zero) and by Billingsley ((1979), Problem 25.14, page 195, extended to signed measures).

$$n^{1/2}[D_1(\theta(F), F, F_n^*) - D_1(\theta(F), F, F_n)] \tag{3.11}$$
$$= m \int_0^{2\pi} h_1(\lambda_1)\, d[n^{1/2}(F_n^*(\lambda_1) - F_n(\lambda_1))] = o(1)$$

with probability one. The statistic U_n was defined by expression (2.7). Since h is bounded, the proof of Lemma 5.7.3, Serfling (1980), will still go through even though $\{\{\lambda_{(j)}^{(n)}\}_{j=1}^{n}\}_{n=1}^{\infty}$ are dependent. By the lemma we have

$$E\,|\,U_n - \theta(F_n^*)\,|^3 \le (E\,|\,F_n(2\pi)\,|^6)^{1/2} O(n^{-3}) = O(n^{-3}) \tag{3.12}$$

(first term on r.h.s. is $O(1)$; see proof of Theorem 5.10.1, Brillinger (1975), page 418, for cumulants of $F_n(2\pi)$), and by the Bonferroni Inequality, for every $\varepsilon > 0$

$$P\{n^{1/2}\,|\,U_n - \theta(F_n^*)\,| > \varepsilon \text{ for at least one } n \ge n_0\}$$
$$\le \textstyle\sum_{n \ge n_0} P\{n^{1/2}\,|\,U_n - \theta(F_n^*)\,| > \varepsilon\}$$
$$\le O(1)\varepsilon^{-3} \textstyle\sum_{n \ge n_0} n^{-2} \to 0 \quad \text{as} \quad n_0 \to \infty.$$

Thus $n^{1/2}\,|\,U_n - O(F_n^*)\,| \to 0$ *w.p.*1 $n \to \infty$. □

4. Probability 1 bounds and almost sure convergence. For the following lemmas an additional assumption will be made concerning the kth order cumulants of the X_n process. The assumption is assumption 7.7.2 of Brillinger (1975), page 264. We will assume that C_k is finite for all $k \in N$ where C_k is defined as

$$C_k = \textstyle\sum_{v_1, v_2, \cdots, v_{k-1}} |\,c(v_1, v_2, \cdots, v_{k-1})\,|$$

where $c(v_1, v_2, \cdots, v_{k-1})$ is the kth order cumulant of $(X(0), X(v_1), \cdots, X(v_{k-1}))$. We will also assume that

$$\textstyle\sum_{L=1}^{\infty} (\sum_{\nu} C_{n_1} C_{n_2} \cdots C_{n_p})(Z^L/L!) < \infty \tag{4.1}$$

for Z in a neighborhood of zero, where the inner summation is over all indecomposable partitions $\nu = (\nu_1, \cdots, \nu_p)$ of the table

$$\begin{matrix} 1 & 2 \\ 3 & 4 \\ & \\ 2L-1 & 2L \end{matrix}$$

with ν_j having $n_j > 1$ elements, $j = 1, \cdots, p$. In the case of a Gaussian process this condition is satisfied if

$$\textstyle\sum_{v=-\infty}^{\infty} |\,c(v)\,| < \infty.$$

The next lemma uses an approach of Brillinger (1975) for obtaining probability one bounds. The periodogram $I_{xx}^{(n)}(\cdot)$ at a single frequency is $O(\log\log n)$ w.p.1 (see Parthasarthy, 1960) whereas Brillinger (1975, Theorem 7.7.2, page 263) shows that $\sup_\lambda I_{xx}^{(n)}(\lambda)$ is $O(\log n)$ w.p.1. The next lemma shows that the $\sup_\lambda |\,F_n(\lambda) - F(\lambda)\,|$ is $O(n^{-1/2}(\log n)^{1/2})$ w.p.1, the improvement being due to the averaging of the periodogram. It can be viewed as a law of the "uniterated" logarithm result for the sample spectral distribution function. The slower rate must be connected with the fact that in Lemma 3.3, $n^{1/2}[F_n(\xi) - F(\xi)]$, $0 \le \xi \le 2\pi$, convergences to a Gaussian process, which is neither (necessarily) a Wiener process nor can be viewed as a Wiener process under transformed time (i.e. intrinsic time).

LEMMA 4.1. *If assumption* (4.1) *is satisfied then* $\| F_n - F \|_\infty = O(n^{-1/2}(\log n)^{1/2})$ *almost surely.*

PROOF. Let $G_n(\lambda) = [F_n(\lambda) - E(F_n(\lambda))]$, $\lambda \in [0, 2\pi]$. Using Brillinger (1978), Theorem 5.10.1, page 418 and Lemma 4.2, page 402, we have that the joint cumulant of $(G_n(\lambda_1), G_n(\lambda_2), \cdots, G_n(\lambda_L))$, $\lambda_j \in [0, 2\pi]$, $j = 1, \cdots, L$, $L \in N^+$, is bounded by $n(\sum_\nu C_{n_1} \cdots C_{n_p})$, defined above, and thus for an arbitrary $\varepsilon > 0$ and $\alpha > 0$ sufficiently small

$$\left| \log E(\exp\{\alpha G_n(\lambda)\}) - \frac{\alpha 2}{2} \operatorname{Var} G_n(\lambda) \right| \le \textstyle\sum_{L=3}^{\infty} (\sum_\nu C_{n_1}, \cdots, C_{n_p}) n \dfrac{|\alpha|^L}{L!}$$

$$E(\exp\{\alpha G_n(\lambda)\}) \le \exp\{\alpha^2/2 \operatorname{Var} G_n(\lambda)(1 + \varepsilon)\}$$

and consequently

$$E(\exp\{\alpha | G_n(\lambda) |\}) \le 2 \exp\{\alpha^2/2 \operatorname{Var} G_n(\lambda)(1 + \varepsilon)\}. \tag{4.2}$$

By the construction of $F_n(\lambda)$ and $E(F_n(\lambda))$,

$$\sup_{0 \le \lambda \le 2\pi} | G_n(\lambda) | = \sup \lambda_j | G_n(\lambda_j) |, \quad \lambda_j = (2\pi j/n), \quad j = 1, 2, \cdots, n \tag{4.3}$$

and thus

$$\begin{aligned} E(\exp\{\alpha \sup_\lambda | G_n(\lambda) | \}) &= E(\exp\{\alpha \sup \lambda_j | G_n(\lambda_j) | \}) \le \textstyle\sum_{j=1}^{n} E(\exp\{\alpha | G_n(\lambda_j) | \}) \\ &\le 2n \exp\{(\alpha^2/2)[\sup_\lambda \operatorname{Var} G_n(\lambda)](1 + \varepsilon)\}. \end{aligned} \tag{4.4}$$

The variance of $G_n(\lambda)$ is

$$n\left[2\pi \int_0^\lambda f_{xx}^2(\alpha)\, d\alpha + \int_0^\lambda \int_0^\lambda f_{xxxx}(\alpha, \beta, -\alpha)\, d\alpha\, d\beta\right] + O(1).$$

Let ψ be

$$\sup_{0 \le \lambda \le 2\pi} 2\pi \left[\int_0^\lambda f_{xx}^2(\alpha)\, d\alpha + \int_0^\lambda \int_0^\lambda f_{xxxx}(\alpha, \beta, -\alpha)\, d\alpha\, d\beta\right] \tag{4.5}$$

so that

$$\sup_{0 \le \lambda \le 2\pi} \operatorname{Var} G_n(\lambda) = n[\psi + O(n^{-1})].$$

Choose $\delta > 0$ and let

$$\begin{aligned} c(n) &= (n \log n)^{1/2} \\ \alpha(n) &= 2^{1/2}(2 + \delta)^{1/2}(\log n)^{1/2}/(n^{1/2}(\psi + O(n^{-1}))^{1/2}(1 + \varepsilon)^{1/2}) \\ a(n) &= 2^{1/2}(\psi + O(n^{-1}))^{1/2}(2 + \delta)^{1/2}(1 + \varepsilon)^{1/2} c(n). \end{aligned} \tag{4.6}$$

Therefore, by expression (4.4) and the Markov Inequality

$$\begin{aligned} &P\{n \sup_\lambda | F_n(\lambda) - E(F_n(\lambda)) | \ge a(n)\} \\ &\quad \le 2 \exp\{-\alpha(n) a(n)\} \exp\{\log n + \frac{\alpha^2(n)}{2} n(\psi + O(n^{-1}))(1 + \varepsilon)\} \\ &\quad \le 2 \exp\{-(1 + \delta) \log n\} \le K n^{-(1+\delta)}. \end{aligned}$$

By the Borel-Cantelli Lemma we have

$$\limsup_{n \to \infty} \frac{n^{1/2} \sup_\lambda | F_n(\lambda - E(F_n(\lambda))) |}{(4\psi \log n)^{1/2}} \le 1 \quad \text{w.p.1} \tag{4.7}$$

and since $| E(F_n(\lambda)) - F(\lambda) | = O(n^{-1})$ (see Brillinger, 1978, page 168), the result follows. □

COROLLARY 4.2. *If assumption* (4.1) *is satisfied, then* $\| F_n^* - F \|_\infty = O(n^{-1/2}(\log n)^{1/2})$ *almost surely.*

PROOF. This follows from Lemma 3.4 and 4.1. □

LEMMA 4.3. *If assumption* (4.1) *is satisfied then the remainder terms*

$$R_{1n} = \theta(F_n) - \theta(F) - D_1(\theta(F), F, F_n) \tag{4.8}$$

$$R_{1n}^* = \theta(F_n^*) - \theta(F^*) - D_1(\theta(F), F, F_n^*)$$

are each $o(n^{-1/2}(\log n)^{1/2})$ *almost surely.*

PROOF. The proof is for R_{1n} with an analogous proof holding for R_{1n}^*. If $\varepsilon > 0$ is arbitrarily chosen, then by expression (3.3) we have

$$| R_{1n} | < \varepsilon \| F_n - F \|_\infty$$

for n sufficiently large. Therefore,

$$\limsup_{n\to\infty} \frac{n^{1/2} | R_{1n} |}{(\log n)^{1/2}} \le \varepsilon \limsup_{n\to\infty} \frac{n^{1/2} \| F_n - F \|_\infty}{(\log n)^{1/2}}$$

and the conclusion follows from the preceding lemma. □

Using the above lemmas we obtain the following probability 1 bounds on our estimators of interest.

THEOREM 4.4. *If assumption* (4.1) *is satisfied, then*

$$\limsup_{n\to\infty} \frac{n^{1/2} |\theta(F_n) - \theta(F)|}{(2\Lambda \log n)^{1/2}} \le 1 \quad \textit{almost surely} \tag{4.9}$$

where Λ *is given by expression* (3.9)

PROOF. By Lemma 4.3

$$\limsup_{n\to\infty} \frac{n^{1/2} |\theta(F_n) - \theta(F)|}{(\log n)^{1/2}} = \limsup_{n\to\infty} \frac{n^{1/2} |D_1(\theta(F), F, F_n)|}{(\log n)^{1/2}} \tag{4.10}$$

which by expression (3.10) is

$$m \limsup_{n\to\infty} \frac{n}{(n \log n)^{1/2}} \left| \frac{2\pi}{n} \sum_{r=1}^n h_1\left(\frac{2\pi r}{n}\right) I_x^{(n)}\left(\frac{2\pi r}{n}\right) - \theta(F) \right|.$$

The proof now is parallel to that of Lemma 4.1 except that

$$G_n(\lambda) = n \left\{ \frac{2\pi}{n} \sum_{(2\pi r/n) \le \lambda} \left[I_{xx}^{(n)}\left(\frac{2\pi r}{n}\right) - f_{xx}\left(\frac{2\pi r}{n}\right) \right] \right\} + O(1)$$

is replaced by

$$H_n = n[D_1(\theta(F), F, F_n) - E(D_1(\theta(F), F, F_n))]$$

$$= n \left\{ \frac{2\pi}{n} \sum_{r=1}^n h_1\left(\frac{2\pi r}{n}\right) \left[I_{xx}^{(n)}\left(\frac{2\pi r}{n}\right) - f_{xx}\left(\frac{2\pi r}{n}\right) \right] \right\} - O(1)$$

where h_1 is of bounded variation (and therefore, bounded) and the difficulty with the supremum over all frequencies is avoided since we are averaging over all frequencies. By the same argument as in Lemma 4.1, we obtain

$$E \exp\{\alpha(n) \mid H_n \mid\} \le 2 \exp\{(\alpha^2(n)/2)(1 + \varepsilon)\mathrm{Var}\, H_n\}$$

with Var H_n equal to $n(\Lambda + O(n^{-1}))$. The $\alpha(n)$, $c(n)$, and $a(n)$ given by expression (4.6) are again appropriate with $(2 + \delta)$ replaced by $(1 + \delta)$ and ψ replaced by Λ. □

COROLLARY 4.5. *If assumption* (4.1) *is satisfied then Theorem* 4.4 *holds with* $\theta(F_n^*)$ *and* U_n *substituted for* $\theta(F_n)$ *in expression* (4.9).

PROOF. By Lemma 4.3, expression (4.10) holds with F_n replaced by F_n^* and the result for $\theta(F_n^*)$ follows from expression (3.11) and Theorem 4.4. For U_n, the result follows from the fact that $n^{1/2} \mid U_n - \theta(F_n^*) \mid = o(1)$ almost surely (proven in Theorem 3.5). □

COROLLARY 4.6. *If assumption* (4.1) *is satisfied then* $\theta(F_n)$, $\theta(F_n^*)$ *and* U_n *all converge to* $\theta(F)$ *almost surely.*

5. A few applications. For $h(\lambda) = e^{is\lambda}$, $s \in N$, the sth autocovariance,

$$\theta(F) = \gamma(s) = \int_0^{2\pi} h(\lambda)\, dF(\lambda) \tag{5.1}$$

and under the conditions of the above theorems, the results apply. The probability one bound in Theorem 4.4 is a general bound. Since $\xi_j = X_j X_{j+s}$, $-\infty < j < \infty$, defines a new stationary process, if the X_j process is ϕ-mixing and ξ_j satisfies weak dependency conditions for a function of a ϕ-mixing process (see Billingsley, 1968, Section 21), then a law of the iterated logarithm for $\gamma(s)$ follows from a result of Philipp (1967). One way to view the weaker result of Theorem 4.4 is that $\theta(F)$ may depend upon all of the autocovariances.

A general form of estimators of the spectral density at λ(i.e., $f_{xx}(\lambda)$) is a quadratic form in the data (see Grenander and Rosenblatt, 1957, Chapter 4).

$$(2\pi n)^{-1} \textstyle\sum_{j,k=1}^{n} b_{j,k} x_j x_k.$$

If W_λ is a symmetric function of bounded variation on $[0, 2\pi]^2$ whose first n^2 Fourier coefficients are the $b_{j,k}$'s:

$$W_\lambda(\xi, \eta) = W_\lambda(\eta, \xi) = W_\lambda(2\pi - \xi, 2\pi - \eta), \quad \eta, \xi \in [0, 2\pi]$$

$$b_{j,k} = \int_0^{2\pi} \int_0^{2\pi} \exp(ij\lambda_1 - ik\lambda_2) W_\lambda(\lambda_1, \lambda_2)\, d\lambda_1\, d\lambda_2, \quad j, k = 1, \cdots, n$$

then the variance of the estimator is

$$\theta(F) = \frac{8\pi^2}{n^2} \int_0^{2\pi} \int_0^{2\pi} \mid W_\lambda(\lambda_1, \lambda_2) \mid^2 dF(\lambda_1)\, dF(\lambda_2)$$

and an estimator of the variance is $\theta(F_n)$:

$$\theta(F_n) = \left(\frac{8\pi^2}{n^2}\right)\left(\frac{2\pi}{n}\right)^2 \textstyle\sum_{j=1}^{n} \sum_{k=1}^{n} \mid W_\lambda(\xi_j, \xi_k) \mid^2 I_{xx}^{(n)}(\xi_j) I_{xx}^{(n)}(\xi_k)$$

when $\xi_l = 2\pi l/n$, $l = 1, 2, \cdots, n$. The most general form of an estimator of the spectral density at λ replaces W_λ by a sequence $W_\lambda^{(n)}$ which becomes concentrated at λ as n increases.

The sample moments of a distribution are both U and V statistics. The moments of the spectral distribution function are of interest since the bandwidth, the expected number of zeroes and the expected number of local maxima of stationary Gaussian process in a given interval all depend upon certain moments of the spectral distribution. Under the conditions of the above theorems, the results apply to these estimators. The central moments are defined for a kernel, h, of dimensional $m > 1$.

Let J_n be a sequence of positive integers which are $o((n/\log n)^{1/2})$. For data $X_1, \cdots, X_n$, if we define $c_n(s)$ as:

$$c_n(s) = \int_0^{2\pi} e^{i\lambda s}\, dF_n(\lambda) \quad s \in \mathbb{N}$$

then $c_n(s)$, $0 \le s < \infty$, are the circular autocovariances. Since $[F_n(\cdot) - F(\cdot)]$ is right continuous and of bounded variation (along each realization), integrating by parts we have (by Theorem 4.4):

$$\begin{aligned}\sup_{0\le s\le J_n} | c_n(s) - \gamma(s) | &= \sup_{0\le s\le J_n} \left| \int_0^{2\pi} e^{i\lambda s}\, d[F_n(\lambda) - F(\lambda)] \right| \\ &= \sup_{0\le s\le J_n} \left| [F_n(2\pi) - F(2\pi)] + is \int_0^{2\pi} [F_n(\lambda) - F(\lambda)] e^{i\lambda s}\, d\lambda \right| \\ &\le o(1) + J_n \| F_n - F \|_\infty = o(1) \quad \text{w.p.1.}\end{aligned}$$

For linear processes An, Chen, and Hannan (1982) have given a variety of results which are similar in nature to this.

Lastly, the general form of the estimators of this paper serves as a first step in the establishment of estimators of time series parameters of the spectrum of a stationary process defined as a solution of an integral minimization (analogues of M-estimators). For even if $\theta(F)$ is not defined with respect to a kernel, the differential approach to statistical functions is still applicable with $D_1(\theta(F), F, F_n)$ often still having a representation with respect to a kernel.

Whittle (1953) and Walker (1964) have shown that (under weak assumptions) maximum likelihood estimation of time series parameters is asymptotically equivalent to minimization of a certain integral defined with respect to the sample spectral distribution. Ibragimov (1967) established consistency results for "maximum likelihood " type estimation performed with respect to the spectral density. Minimization estimators can be established which generalize these results.

REFERENCES

AKAIKE, H. (1974a). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Ann. Inst. Statist. Math.* **26** 363–387.

AKAIKE, H. (1974b). A new look at the statistical model identification. *IEEE Trans. Autom. Control* **AC-19** 716–722.

AN, H.-Z., CHEN, Z.-G., and HANNAN, E. J. (1982). Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.* **10** 926–936.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BILLINGSLEY, P. (1979). *Probability and Measure.* Wiley, New York.

BRILLINGER, D. R. (1969). Asymptotic properties of spectral estimates of second order. *Biometrika* **56** 375–390.

BRILLINGER, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart, and Winston, New York.

GRENANDER, U., and ROSENBLATT, M. R. (1957). *Statistical Analysis of Stationary Time Series.* Wiley, New York.

HOBSON, E. W. (1927). *The Theory of Functions of a Real Variable and the Theory of Fourier's Series (3rd ed.) Vol. 1,* Cambridge.

IBRAGIMOV, I. A. (1967). On maximum likelihood estimation of parameters of the spectral density of stationary time series. *Theory Probab. Appl.* **120** 115–119.

MARTIN, R. D. (1979). Robust estimation of autoregressive models. In *Directions in Time Series* (D. R. Brillinger and G. C. Tiao, eds.). IMS Special Topics Meeting on Time Series Analysis, Iowa State University, Ames.

PARTHASARTHY, K. R. (1960). On the estimation of the spectrum of a stationary stochastic process. *Ann. Math. Statist.* **31** 568–573.

PARZEN, E. (1957). On consistent estimates of the spectrum of a stationary time seris. *Ann. Math. Statist.* **28** 329–348.

PARZEN, E. (1974). Some recent advances in time series modelling. *IEEE Trans. Autom. Control* **AC-19** 723–730.

PHILIPP, W. (1967). Das Gesetz vom iterierten Logarithmus für stark mishende stationare prozesse. *Z. Wahrsch. verw. Gebiete* **8** 204–209.

SEN, P. K. (1972). Limiting behavior of regular functionals of empirical distributions for stationary *-mixing processes. *Z. Wahrsch. verw. Gebiete* **25** 71–82.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

WALKER, A. M. (1964). Asymptotic properties of least square estimates of parameters of the spectrum of a stationary nondeterministic time series. *J. Austral. Math. Soc.* **4** 363–384.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat. Astr. Fys.* **2** 423–434.

YOSHIHARA, K. (1976). Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrsch. verw. Gebiete* **35** 237–252.

DEPARTMENT OF STATISTICS
SCHENLEY PARK
CARNEGIE-MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213