The Annals of Statistics
1981, Vol. 9, No. 3, 615–627

ESTIMATION OF THE SPECTRAL PARAMETERS OF A STATIONARY POINT PROCESS.

By Pham Dinh Tuan

Mathématique Appliquées

This paper considers the two approaches for estimating the parameters specifying the spectral density of the counting process of a stationary point process, namely the frequency domain and the time domain approaches. The relation between the two is clarified; consistency and asymptotic normality of the estimates are established. Finally the special case of a rational spectral density is considered in some detail.

1. Introduction. Let $N(t)$ be the counting process of a stationary point process, that is $N(t)$ denotes the number of point events occurring in $[0, t]$. Suppose we assume a finite parameter model for $N(t)$, then it is usually possible to write down the spectral density of this process and as far as its second order properties are concerned, only those parameters involved in the spectral density are of interest. Examples of models we have in mind are the simple self exciting processes introduced in Hawkes (1972) which can be characterized by

$$\lim_{\Delta t \to 0} P\{N(t+\Delta t) - N(t) \geq 1 \mid N(u), u \leq t\} \div \Delta t$$
$$= \alpha + \int_{-\infty}^{t} A(t-s)\, dN(s) \tag{1.1}$$

for which the rate and the spectral density are respectively

$$\mu = \frac{\alpha}{1 - a(0)}, \; f(\lambda) = \frac{\mu}{2\pi \mid 1 - a(\lambda) \mid ^2},$$

where

$$a(\lambda) = \int_0^\infty e^{i\lambda t} A(t)\, dt.$$

Thus we are led to consider point processes with spectral density of the form $(\mu/2\pi)g_\theta(\cdot)$ where $\mu > 0$ is the rate of the process and $\theta \in \Theta \subset R^m$ is an unknown parameter, such that $g_\theta(\lambda) \to 1$ as $\lambda \to \pm\infty$. The last condition is no great restriction since the spectral density, being the Fourier transform of a measure with mass μ at the origin (Cox & Lewis, 1968, page 74), tends to $\mu/2\pi$ at infinity in most cases.

Another reason for considering the above model is the prediction of point processes from the past. If we are interested in linear least squares prediction, all we need is to know the rate and the spectral density f of the process (Jowett and Vere-Jones, 1972). Now in the absence of any a priori information, one might try to approximate f by some member of a parametric family $\{\mu g_\theta, \mu > 0, \theta \in \Theta\}$, sufficiently rich. An important example is the family of rational functions with numerators and denominators having some prescribed

Received October, 1978; revised May, 1980.

AMS 1970 *subject classifications.* Primary, 60G10, 62F10; secondary, 62M10.

Key words and phrases. Cumulant, least square estimation, maximum likelihood estimation, periodogram, point process, spectral density.

degree. From this point of view, f need not belong to the above family and our purpose is to construct an estimate of the value $\mu^* > 0$ and $\theta^* \in \Theta$ such that $\mu^* g_{\theta^*}$ comes closest to f in some sense to be made precise, hoping that $\mu^* g_{\theta^*}$ would be a good approximation to f.

This estimation problem has been considered by Brillinger (1975b), who suggested a procedure based on the periodogram and derived the asymptotic distribution of the estimate. His proof is however somewhat rapid and he has overlooked the effect of estimating μ on the distribution of the estimate of θ. Here, we shall consider two modified versions of the above and a procedure in the time domain approach similar to that of the Box and Jenkins (1970) method of estimation of parameters in time series models. The relation between frequency and time domain approaches is clarified and the asymptotic distribution of the estimate is obtained. Finally, we consider the special case of a rational spectral density and derive an iterative procedure to compute the time domain estimates.

The motivation for the above procedures is quite standard. In the frequency domain approach, let

$$I^T(\lambda) = \frac{1}{2\pi T} \mid d^T(\lambda)|^2, \; d^T(\lambda) = \int_0^T e^{i\lambda t}\, dN(t),$$

where $[0, T]$ is the interval of observation. Then we observe that (Brillinger, 1972, 1975b), under regularity conditions, the periodogram ordinates $I^T(2\pi s/T)$, $s = 1, \cdots, M$ are jointly asymptotically independent exponential variates with means $f(2\pi s/T)$, $s = 1, \cdots, M$. Thus, let

$$\hat{\mu} = T^{-1}\{N(T) - N(0)\} = T^{-1}\, d^T(0)$$

by the consistent estimate of μ. The random variables $(2\pi/\hat{\mu})I^T(2\pi s/T)$ are jointly asymptotically independent exponential variates with means $(2\pi/\mu)f(2\pi s/T)$, which suggests writing down the "pseudo" log likelihood function

$$-\frac{1}{T}\sum_{s=1}^{M}\left\{\log g_\theta\left(\frac{2\pi s}{T}\right) + g_\theta^{-1}\left(\frac{2\pi s}{T}\right)\frac{2\pi}{\hat{\mu}}\, I^T\left(\frac{2\pi s}{T}\right)\right\}$$

and estimating θ by maximizing this expression.

Now, it is desirable that M should be large in order to make full use of the available information. However the above sum is not convergent as $M \to \infty$, so we modify it by subtracting a term independent of θ to get

$$\text{(1.2)} \qquad \Lambda_{T,M}(\theta) = -\frac{1}{T}\sum_{s=1}^{M}\left[\log g_\theta\left(\frac{2\pi s}{T}\right) + \left\{g_\theta^{-1}\left(\frac{2\pi s}{T}\right) - 1\right\}\frac{2\pi}{\hat{\mu}}\, I^T\left(\frac{2\pi s}{T}\right)\right].$$

Since $I^T(\cdot)$ is bounded for fixed T, if $g_\theta(\lambda) \to 1$ sufficiently fast as $\lambda \to \pm\infty$, then $\Lambda_{T,M}(\theta)$ converges almost surely to $\Lambda_T(\theta)$ say, as $M \to \infty$. To obtain estimates of θ, one might try to maximize $\Lambda_T(\theta)$ or $\Lambda_{T,M_T}(\theta)$ where $M_T \to \infty$ with a prescribed rate, as $T \to \infty$. One might also consider the continuous version of Λ_T:

$$\text{(1.3)} \qquad \tilde{\Lambda}_T(\theta) = -\int_0^\infty \left[\frac{1}{2\pi}\log g_\theta(\lambda) + \{g_\theta^{-1}(\lambda) - 1\}\hat{\mu}^{-1}\hat{I}^T(\lambda)\right] d\lambda$$

where $\hat{I}^T(\lambda)$ is the mean-corrected periodogram defined by

$$\hat{I}^T(\lambda) = \frac{1}{2\pi T} \mid \hat{d}^T(\lambda)|^2, \qquad \hat{d}^T(\lambda) = \int_0^T e^{i\lambda t}\{dN(t) - \hat{\mu}\, dt\}.$$

Note that $\hat{I}^T(2\pi s/T) = I^T(2\pi s/T)$ for $s \neq 0$ and $\hat{I}^T(0) = 0$.

In the time domain approach, consider the problem of linear least squares prediction of

$dN(t)$ by the past up to time t. This is equivalent to constructing a random measure $dm(t)$ with $m(t) - m(s)$, $t > s$ belonging to the closure in mean square of the linear space spanned by $N(u) - N(v)$, $v < u < t$, such that $\epsilon(t) = N(t) - m(t)$ is a process of orthogonal increments. In general, $dm(t) = \phi(t)\,dt + \mu\,dt$, where

$$\phi(t) = \int_{-\infty}^{t} A(t-s)\{dN(s) - \mu(ds)\} \tag{1.4}$$

with A being determined by f as follows. Let $f(x) = (\mu/2\pi)\,|\,h(\lambda)|^2$ be the factorization of f (Doob, 1953, page 586) and suppose that $h - 1$ is square integrable. Then from the spectral representation of $N(t)$ (Doob, 1953, page 552)

$$N(t) = \int_{-\infty}^{\infty} \frac{e^{i\lambda t} - 1}{i\lambda} h(\lambda)\,dZ(\lambda) + \mu t,$$

where $Z(\lambda)$ is of orthogonal increments, we get $N(t) = \int_0^t \phi(s)\,ds + \epsilon(t) + \mu t$ where

$$\phi(t) = \int_{-\infty}^{\infty} e^{i\lambda t}\{h(\lambda) - 1\}\,dZ(\lambda)$$

and $\epsilon(t)$ is of orthogonal increments. Let $a = 1 - h^{-1}$ be the Fourier transform of some A, then, noting that $A(t) = 0$ for $t < 0$ because of the analyticity of a, it can be shown that $\phi(t)$ is also given by (1.4). Note that for the self-exciting process, A is precisely the one involved in (1.2).

We will write h_θ, a_θ, A_θ, $\phi_{\theta,\mu}$ instead of h, a, A, ϕ to indicate their dependence on θ, μ. By analogy with the least squares method, we estimate θ, μ by minimizing the formal "sum of squares of the residuals"

$$T^{-1} \int_0^T \{dN/dt - \mu - \hat{\phi}_{\theta,\mu}(t)\}^2\,dt$$

where

$$\hat{\phi}_{\theta,\mu}(t) = \int_0^t A_\theta(t-s)\{dN(s) - \mu\,ds\} \tag{1.5}$$

is an approximation to $\phi_{\theta,\mu}(t)$ involving only values of $N(s)$ on $[0, T]$. This is justified if $A_\theta(t) \to 0$ sufficiently fast as $t \to \infty$. The above sum of squares is divergent but by ignoring the constant term $\int (dN/dt)^2\,dt$, we get

$$S_T(\theta, \mu) = -\frac{2}{T}\int_0^T \{\phi_{\theta,\mu}(t) + \mu\}\,dN(t) + \frac{1}{T}\int_0^T \{\phi_{\theta,\mu}(t) + \mu\}^2\,dt. \tag{1.6}$$

The integrals are understood as integrals along sample paths and, to avoid ambiguities when $\phi_{\theta,\mu}(t)$ has discontinuities at the points of increase of $N(t)$, and to be consistent with the fact that $\phi_{\theta,\mu}(t)$ depends only on $N(u) - N(v)$, $u < v < t$, we suppose that $\phi_{\theta,\mu}(t)$ has left continuous sample paths.

Now, $S_T(\theta, \mu)$ is quadratic in μ and hence can be easily minimized with respect to μ. We shall see in Section 2 that under some regularity conditions, the value of μ achieving this minimum differs very little from $\hat{\mu}$. Thus one can estimate μ by $\hat{\mu}$ and θ by minimizing

$$S_T(\theta) = S_T(\theta, \hat{\mu}) - \hat{\mu}^2 = -\frac{2}{T}\int_0^T \hat{\phi}_{\theta,\hat{\mu}}(t)\{dN(t) - \hat{\mu}\,dt\} + \frac{1}{T}\int_0^T \hat{\phi}^2_{\theta,\hat{\mu}}(t)\,dt. \tag{1.7}$$

The main results of the paper are summarized in Section 2 where we establish the asymptotic equivalence between frequency and time domain approaches and obtain the consistency and asymptotic normality of the estimates. Because of length and of technicalities involved, the proofs are deferred to Section 4. Section 3 is devoted to the case of a rational spectral density.

2. Main results. We first show that $\tilde{\Lambda}_T(\theta)$ differs very little from $-S_T(\theta)/(2\hat{\mu})$ which shows the connection between time and frequency domain approaches. We shall need the assumptions:

A1. The function h_θ has no zero in the closed upper half plane and $a_\theta = 1 - h_\theta^{-1}$ is continuous there and admits the asymptotic expansion

$$a_\theta(z) = ic(\theta)z^{-1} + O(z^{-2}), \qquad |z| \to \infty. \tag{2.1}$$

A2. For any compact C of Θ, $A_\theta(t)$ is bounded in $C \times R^+$ and satisfies

$$\int_0^\infty t\{\sup_{\theta\in C} | A_\theta(t)|\}\, dt < +\infty.$$

A3. The function $A_\theta(t)$ admits continuous first and second derivatives with respect to θ, satisfying the same condition as $A_\theta(t)$ as stated in A1.

The above assumptions A1–A3 are not very restrictive. They can be verified, for example, in the case where h_θ is a rational function with coefficients being twice continuously differentiable functions of θ.

THEOREM 1. *Let $\hat{\mu}_\theta$ realize the minimum of $S_T(\theta, \mu)$, given by* (1.6), *for fixed θ. Then under* A2

$$\hat{\mu}_\theta = \hat{\mu} + O(T^{-1}), \qquad S_T(\theta, \hat{\mu}_\theta) = S_T(\theta, \hat{\mu}) + O(T^{-2}),$$

where the error terms are uniform in the L^1 norm in any compact C of Θ. If A3 *also holds then these two error terms admit derivatives with respect to θ up to second order, respectively $O(T^{-1})$ and $O(T^{-2})$ in L^1 norm, uniformly in C.*

THEOREM 2. *Let $\tilde{\Lambda}_T$, S_T, $\hat{\phi}$ be given by* (1.3), (1.7), (1.5), *then under* A1

$$\tilde{\Lambda}_T(\theta) = -\frac{1}{2\mu}\left\{S_T(\theta) + \frac{1}{T}\int_T^\infty \hat{\phi}^2_{\theta,\mu}(t)\, dt\right\}.$$

If moreover A2 *holds, then the last term of the above expression is $O(T^{-1})$ in L^1 norm, uniformly in any compact C of Θ, and if* A3 *also holds then this term admits first and second derivatives with respect to θ, with the same properties as above.*

We now turn to the consistency and asymptotic normality of the estimate. Following Brillinger (1972), we shall make the assumption

A4. The $N(t)$ process admits cumulant measures of order k for all k, satisfying

$$\int (1 + | u_j |) \, | dC^{(k)}(u_1, \cdots, u_{k-1}) | < +\infty, \qquad 1 \le j < k, \quad k \ge 2.$$

Recall that the cumulant measure $C^{(k)}$ is defined by

$$\text{cum}\left\{\int h_j(t)\,dN(t),\quad j=1,\cdots,k\right\} = \int h_1(t)h_2(t+u_1)\cdots h_k(t+u_{k-1})\,dC^{(k)}(u_1,\cdots,u_{k-1})$$

where $h_1,\cdots,h_k$ are bounded measurable functions with compact support. The Fourier transform of $dC^{(k)}$ is the k-cumulant spectral density $f^{(k)}$. For $k=2$, we omit the superscript k.

From A4, we can obtain a result similar to that of Brillinger (1975a, page 168) which is of independent interest and is essential for our Theorem 5.

THEOREM 3. *Let $p_1,\cdots,p_m$ be complex valued functions on R, such that for some $\delta>1$, $(1+|\lambda|^{\delta})p_j(\lambda)$, $\lambda\in R$ are bounded. Let*

$$J^T(p_j)=\frac{2\pi}{T}\sum_{1\le|s|\le M_T}p_j\left(\frac{2\pi s}{T}\right)I^T\left(\frac{2\pi s}{T}\right)$$

$$\hat{J}^T(p_j)=\int_{-l_T}^{l_T}p_j(\lambda)\hat{I}^T(\lambda)\,d\lambda.$$

Then as $T\to\infty$, $M_T/T\to\infty$ or $M_T=\infty$, $l_T\to\infty$ or $l_T=\infty$ we have

$$EJ^T(p_j)=\frac{2\pi}{T}\sum_{1\le|s|\le M_T}p_j\left(\frac{2\pi s}{T}\right)f\left(\frac{2\pi s}{T}\right)+O(T^{-1})$$

$$E\hat{J}^T(p_j)=\int_{-l_T}^{l_T}p_j(\lambda)f(\lambda)\,d\lambda+O(T^{-1})$$

and the random variables $\hat{\mu}-\mu$, $J^T(p_j)-EJ^T(p_j)$, $j=1,\cdots,n$ are jointly asymptotically normal with zero mean and covariance structure

$$\lim_{T\to\infty}T\,\text{Var}(\hat{\mu})=2\pi f(0)$$

$$\lim_{T\to\infty}T\,\text{Cov}\{\hat{\mu},J^T(p_j)\}=2\pi\int p_j(\alpha)f^{(3)}(0,\alpha)\,d\alpha$$

$$\lim_{T\to\infty}T\,\text{Cov}\{J^T(p_j),J^T(p_k)\}=2\pi\left[\iint p_j(\alpha)p_k(\beta)f^{(4)}(\alpha,\beta,-\alpha)\,d\alpha\,d\beta + \int\{p_j(-\alpha)\overline{p_k(\alpha)}+p_j(\alpha)\overline{p_k(\alpha)}\}f^2(\alpha)\,d\alpha\right]$$

and similar results hold when $J^T(p_j)$ are replaced by $\hat{J}^T(p_j)$.

To obtain the consistency and asymptotic normality of our estimate, we shall use the following result, which is also of independent interest.

THEOREM 4. 1. *Let L_T be a random function on $\Theta\subset R^m$, satisfying*
(i) *$L_T(\theta)\to L(\theta)$ in probability as $T\to\infty$, with L being upper semicontinuous, admitting a unique maximum at θ^**
(ii) *for every compact C of Θ and $\epsilon>0$*

$$\liminf_{T\to\infty}P\{|L_T(\theta)-L_T(\theta')|\le\epsilon,\quad\forall\theta,\theta'\in C,\quad\|\theta-\theta'\|\le 1/n\}\to 1$$

as $n\to\infty$. Then any θ_T realizing the maximum of L_T in Θ and satisfying $P[\hat{\theta}_T\in C]\to 1$ for some compact C of Θ, tends to θ^* in probability as $T\to\infty$.

2. *Suppose that θ^* is an interior point of Θ and L_T admits continuous first and second derivatives with respect to θ in a neighborhood of θ^*, denoted by the vector $L_T^{(1)}$ and the matrix $L_T^{(2)}$, satisfying*:

(iii) $L_T^{(1)}(\theta^*) \to 0$ *and* $L_T^{(2)}(\theta^*) \to -(J + H)$ *in probability, as* $T \to \infty$, *and* $L_T^{(1)}(\theta^*)$ *is asymptotically normal with zero mean and covariance matrix* $T^{-1}(J + K)$

(iv) *for every* $\epsilon > 0$

$$\liminf_{T\to\infty} P\{\| L_T^{(2)}(\theta) - L_T^{(2)}(\theta^*)\| > \epsilon, \quad \forall \theta \in C, \quad \|\theta - \theta^*\| < 1/n\}$$

tends to 1 *as* $n \to \infty$. *Then* $\hat\theta_T$ *is asymptotically normal with mean* θ^* *and covariance matrix* $T^{-1}(J + H)^{-1}(J + K)(J + H)^{-1}$. *Moreover if* θ_T *is* $T^{1/2}$ *consistent, that is the distributions of* $\sqrt{T}(\theta_T - \theta^*)$ *are tight, then* $\hat\theta_T - \{\theta_T - L_T^{(2)}(\theta_T)^{-1} L_T^{(1)}(\theta_T) \to 0$ *in probability as* $T \to \infty$.

Remark. Our theorem is similar to that of Brillinger. However, he has used a weaker condition than our condition (ii), namely that for every $\epsilon > 0$

$$\liminf_{T\to\infty} P\{L_T(\theta') - L_T(\theta) \le \epsilon, \qquad \forall \theta' \in V(\theta)\} \to 1$$

as the neighborhood $V(\theta)$ of θ shrinks to θ. This condition is inadequate for his proof since when one tries to cover the compact C of Θ by a finite number of the $V(\theta)$, one does not know in advance how many of them are needed. Therefore the stronger condition (ii) is needed. The proof of our Theorem is rather standard and similar to that of Brillinger, so we omit it.

To apply Theorem 4, we have to check its conditions. This is done in

Theorem 5. 1. *Suppose that* g_θ^{-1} *is of bounded variation on finite intervals, jointly continuous in* (θ, λ) *and for some* $\delta > 1$, $|\lambda|^\delta \{g_\theta^{-1}(\lambda) - 1\} \to 0$ *as* $(\theta, \lambda) \to (\theta', \pm\infty)$ *for any* θ'. *Then as* $T \to \infty$, $M_T/T \to \infty$ *or* $M_T = \infty$, $\Lambda_{T,M_T}(\theta)$ *and* $\tilde\Lambda_T(\theta)$ *converge in probability to*

$$\Lambda(\theta) = -\int_0^\infty \left[\frac{1}{2\pi}\log g_\theta(\lambda) + \{g_\theta^{-1}(\lambda) - 1\}\mu^{-1} f(\lambda)\right] d\lambda$$

and they satisfy condition (ii) *of Theorem* 4.

2. *Suppose moreover that* $g_\theta^{-1}(\lambda)$ *admits first and second derivatives with respect to* θ, *continuous in* (θ, λ) *such that for some* $\delta > 1$, $|\lambda|^\delta(\partial/\partial\theta_j) g_\theta^{-1}(\lambda)$ *and* $|\lambda|^\delta(\partial^2/\partial\theta_j\partial\theta_k) g_\theta^{-1}(\lambda)$ *tend to* 0 *as* $(\theta, \lambda) \to (\theta', \pm\infty)$ *for any* θ'. *Then* $\tilde\Lambda_T$ *satisfies conditions* (iii), (iv) *of Theorem* 4 *with*:

$$J_{jk} = \frac{1}{2\pi}\left[\int_0^\infty \left\{\frac{\partial}{\partial\theta_j}\log g_\theta(\lambda)\frac{\partial}{\partial\theta_k}\log g_\theta(\lambda)\right\}\left\{g_\theta^{-1}(\lambda)\frac{2\pi}{\mu}f(\lambda)\right\}^2 d\lambda\right]_{\theta=\theta^*}$$

$$H_{jk} = \frac{1}{2\pi}\left[\int_0^\infty \left\{\frac{\partial^2}{\partial\theta_j\partial\theta_k} g_\theta(\lambda)\right\} g_\theta^{-1}(\lambda)\left\{1 - g_\theta^{-1}(\lambda)\frac{2\pi}{\mu}f(\lambda)\right\} d\lambda \right.$$
$$\left. - \int_0^\infty \left\{\frac{\partial}{\partial\theta_j}\log g_\theta(\lambda)\frac{\partial}{\partial\theta_k}\log g_\theta(\lambda)\right\}\left\{1 - g_\theta^{-1}(\lambda)\frac{2\pi}{\mu}f(\lambda)\right\}^2 d\lambda\right]_{\theta=\theta^*}$$

$$K_{jk} = \frac{2\pi}{\mu^2}\left[\int_0^\infty\int_0^\infty \left\{\frac{\partial}{\partial\theta_j} g_\theta^{-1}(\alpha)\frac{\partial}{\partial\theta_k} g_\theta^{-1}(\beta)\right\} f^{(4)}(\alpha, \beta, -\alpha)\, d\alpha\, d\beta\right.$$

$$+\frac{1}{4}\left\{\frac{\partial}{\partial\theta_j}c(\theta)\frac{\partial}{\partial\theta_k}c(\theta)\right\}f(0)$$

$$-\frac{1}{2}\frac{\partial}{\partial\theta_j}c(\theta)\int_0^\infty\left\{\frac{\partial}{\partial\theta_k}g_\theta^{-1}(\lambda)\right\}f^{(3)}(0,\lambda)\,d\lambda$$

$$\left.-\frac{1}{2}\frac{\partial}{\partial\theta_k}c(\theta)\int_0^\infty\frac{\partial}{\partial\theta_j}\{g_\theta^{-1}(\lambda)\}f^{(3)}(0,\lambda)\,d\lambda\right]_{\theta=\theta^*}$$

where $c(\theta)$ *is given by* (2.1). *The same result holds for* $\Lambda_{T,M_T}(\cdot)$ *provided that the functions*

$$\beta_j(\lambda)=\frac{\partial}{\partial\theta_j}\left\{\log g_\theta(\lambda)+\frac{2\pi}{\mu}g_\theta^{-1}(\lambda)f(\lambda)\right\}_{\theta=\theta^*}$$

are of bounded variation on $(0,\infty)$ *and*

$$\sqrt{T}\int_{2\pi M_T/T}^\infty\beta_j(\lambda)\,d\lambda\to 0,\qquad T\to\infty. \tag{2.2}$$

REMARK. Using the fact that $\log x\le x-1$ with equality if and only if $x=1$, it can be shown that if $f=(\mu/2\pi)g_{\theta^*}$ and g_θ not equal to g_{θ^*} a.e., $\forall\theta\ne\theta^*$, then θ^* is the unique maximum of $\Lambda(\theta)$. In this case $H=0$ and J, K can be simplified. In general g_{θ^*} is near $(2\pi/\mu)f$ so that H can be neglected.

3. Case of rational spectral density. In the case where $g_\theta(\lambda)$ is a rational function of λ, then (Doob, 1953, page 452),

$$g_\theta(\lambda)=\frac{p(i\lambda)}{q(i\lambda)} \tag{3.1}$$

where $p(z)=z^n+a_1z^{n-1}+\cdots+a_n$, $q(z)=z^n+bz^{n-1}+\cdots+b_n$. Here $\theta=(a_1,\cdots,a_n,b_1,\cdots,b_n)$ is the unknown parameter. We shall suppose that p, q have no zeroes in common and no zeroes in the closed right half plane. The motivation for using model (3.1) is that it allows a simple expression for the linear predictor and thus makes the computation of the time domain estimates feasible. Also the family of functions $(\mu/2\pi)g_\theta$ of (3.1) is quite rich so it is hoped that for a well chosen n, this family contains the true spectral density or at least an element very close to it.

The estimating procedure consists of minimizing the sum of the squares $S_T(\theta)$ given by (1.7). We might replace the upper bound T of the last integral in (1.7) by ∞, which amounts to replacing S_T by a negative multiple of $\tilde{\Lambda}_T$ (see Theorem 2). Now, to minimize S_T one usually solves the equations

$$\frac{\partial}{\partial\theta_j}S_T(\theta)=-\frac{2}{T}\int_0^T\frac{\partial}{\partial\theta_j}\hat{\phi}_{\theta,\mu}(t)\{dN(t)-\hat{\mu}\,dt-\hat{\phi}_{\theta,\hat{\mu}}(t)\,dt\}=0,$$

for $j=1,2,\cdots,2n$. Since a closed form solution does not exist, one has to solve these equations numerically, using for example a Newton-Rhapson algorithm:

$$\theta_{T,n+1}=\theta_{T,n}-\{S_T^{(2)}(\theta_{T,n})\}^{-1}S_T^{(1)}(\theta_{T,n})$$

where $S_T^{(1)}$ and $S_T^{(2)}$ are the vector and the matrix of first and second derivatives of $S_T(\cdot)$. By Theorem 4, if the initial estimate $\theta_{T,0}$ is $T^{1/2}$-consistent then the one step estimate $\theta_{T,1}$ is asymptotically as good as the least squares estimate, the one that minimizes S_T. Now

$$\frac{\partial^2}{\partial\theta_j\partial\theta_k} S_T(\theta) = \frac{2}{T}\int_0^T \frac{\partial^2}{\partial\theta_j\partial\theta_k}\hat{\phi}_{\theta,\hat{\mu}}(t)\{dN(t) - \hat{\phi}_{\theta,\hat{\mu}}(t)\,dt - \hat{\mu}\,dt\} - \frac{2}{T}\int_0^T \frac{\partial}{\partial\theta_j}\hat{\phi}_{\theta,\hat{\mu}}(t)\frac{\partial}{\partial\theta_k}\hat{\phi}_{\theta,\hat{\mu}}(t)\,dt,$$

where the first term can be neglected. Indeed by the same argument as in the proof of Theorem 2, this term can be shown to be

$$\int_{-\infty}^{\infty} [\{1 - a_\theta(\lambda)\}\, a^{(2)}_{\theta,jk}(\lambda) + \{1 - a_\theta(-\lambda)\} a^{(2)}_{\theta,jk}(-\lambda)]\hat{I}^T(\lambda)\,d\lambda - \frac{\partial^2}{\partial\theta_j\partial\theta} c(\theta)\hat{\mu} + \frac{2}{T}\int_T^\infty \left\{\frac{\partial}{\partial\theta_j\partial\theta_k}\hat{\phi}_{\theta,\hat{\mu}}(t)\right\}\hat{\phi}_{\theta,\hat{\mu}}(t)\,dt$$

where $a^{(2)}_{\theta,jk}(\lambda) = (\partial^2/\partial\theta_j\partial\theta_k)a_\theta(\lambda)$. The same argument in the first part of the proof of Theorem 1 shows that the above second term is the limit as $|z| \to \infty$ of $za^{(2)}_{\theta jk}(z)$ and hence of $zh_\theta(z)a^{(2)}_{\theta,jk}(z)$; therefore the first two terms above can be rewritten as

$$\int_{-\infty}^{\infty} [h_\theta^{-1}(\lambda)a^{(2)}_{\theta,jk}(\lambda) + h_\theta^{-1}(-\lambda)a^{(2)}_{\theta,jk}(-\lambda)]\left\{\hat{I}^T(\lambda) - \frac{\hat{\mu}}{2\pi} g_\theta(\lambda)\right\} d\lambda,$$

which, when θ is replaced by a consistent estimate of θ^*, converges as $T \to \infty$ to $\int \psi(\lambda)\{f(\lambda) - (\mu/2\pi)g_{\theta^*}(\lambda)\}\,d\lambda$, where ψ is a certain function (see Theorem 3, the proof of Theorem 5 and condition (ii) of Theorem 4). Since f should be very close to $(\mu/2\pi)g_{\theta^*}$, the last integral is negligible.

Thus we shall replace $S_T^{(2)}$ by the simpler expressions

$$-\frac{2}{T}\int_0^T \frac{\partial}{\partial\theta_j}\hat{\phi}_{\theta,\hat{\mu}}(t)\frac{\partial}{\partial\theta_k}\hat{\phi}_{\theta,\hat{\mu}}(t)\,dt, \qquad j, k = 1, \cdots, 2n.$$

To compute $\hat{\phi}_{\theta,\hat{\mu}}(t)$ and its derivatives, let $y_\theta(t)$ be the solution of

$$dy_\theta^{(n-1)}(t) + \textstyle\sum_{j=1}^m a_j y_\theta^{(n-j)}(t) = 1_{[0,\,T]}(t)\{dN(t) - \hat{\mu}\,dt\} \tag{3.2}$$

with initial conditions $y_\theta^{(j)}(0) = 0, j = 1, \cdots, n-1$. The term $dN(t)$ has the effect of adding a jump of magnitude 1 to $y_\theta^{(n-1)}(t)$ at each point event occurring in $[0, T]$. Otherwise (3.2) is an ordinary differential equation. One can then check that

$$\hat{\phi}_{\theta,\hat{\mu}}(t) = \textstyle\sum_{j=1}^n (a_j - b_j)y_\theta^{(n-j)}(t)$$

$$dN(t) = \hat{\phi}_{\theta,\hat{\mu}}(t)\,dt - \hat{\mu}\,dt = dy_\theta^{(n-1)}(t) + \textstyle\sum_{j=1}^n b_j y_\theta^{(n-j)}(t)\,dt$$

$$\frac{\partial}{\partial b_j}\hat{\phi}_{\theta,\hat{\mu}}(t) = y_\theta^{(n-j)}(t)$$

$$\frac{\partial}{\partial a_j}\hat{\phi}_{\theta,\hat{\mu}}(t) = y_\theta^{(n-j)}(t) + \textstyle\sum_{k=1}^n (a_k - b_k)\frac{\partial}{\partial a_j}y_\theta^{(n-k)}(t).$$

But from (3.2), $(\partial/\partial a_j)y_\theta(t)$ is the solution of the differential equation $p(d/dt)\{\partial y_\theta(t)/\partial a_j\} = -y_\theta^{(n-j)}(t)$, with initial conditions $(\partial/\partial a_j)y_\theta^{(k)}(0) = 0$, $k = 0, \cdots, n-1$, and hence it is equal to $-w_\theta^{(n-j)}(t)$ where $w_\theta(t)$ is the solution of $p(d/dt)w_\theta(t) = y_\theta(t)$ with initial condition $w_\theta^{(k)}(0) = 0$, $k = 0, \cdots, n-1$. Thus

$$\frac{\partial}{\partial a_j}\hat{\phi}_{\theta,\hat{\mu}}(t) = q\left(\frac{d}{dt}\right)w_\theta^{(n-j)}(t) = z_\theta^{(n-j)}(t)$$

where $z_\theta(t) = q(d/dt)w_\theta(t)$ can be seen to be the solution of

$$dz_\theta^{(n-1)}(t) + \sum_{j=1}^n a_j z_\theta^{(n-j)}(t)\,dt = dy_\theta^{(n-1)}(t) = \sum_{j=1}^n b_j y_\theta^{(n-j)}.$$

REMARK. In computing the estimate by Newton-Rhapson algorithm, we need an initial estimate, preferably $T^{1/2}$-consistent. So far we do not know a simple procedure for constructing one such. However one can proceed as follows. Consider the mapping $\omega \to \lambda_0\ \tau g(\omega/2)$ with inverse $\lambda \to e^{i\omega} = (1 + i\lambda/\lambda_0)/(1 - i\lambda/\lambda_0)$. Then the function $\tilde{f} : \tilde{f}(\omega) - f(\lambda)$ is rational in $e^{j\omega}$ and hence is the spectral density of an autoregressive moving average process. Now if the covariance of this process

$$c(k) = \int_{-\pi}^{\pi} e^{ik\omega}\tilde{f}(w)\,d\lambda = \int_{-\infty}^{\infty} \left(\frac{1 + i\lambda/\lambda_0}{1 - i\lambda/\lambda_0}\right)^k f(\lambda)\frac{d\lambda/\lambda_0}{1 + \lambda^2/\lambda_0^2}$$

of lag $k = 1, \cdots, 2n$ are known, then the corresponding autoregressive-moving average coefficients $\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_n$, say, can be computed (Hannan, 1970; Box & Jenkins, 1970). We do not know the $c(k)$, but we can estimate them by

$$\hat{c}(k) = \frac{2\pi}{T}\sum_{s=1}^{M_T} 2\,\mathrm{Re}\left\{\left(1 + \frac{2\pi s}{\lambda_0 T}\right)^k\left(1 - \frac{2\pi s}{\lambda_0 T}\right)^{-k}\right\} I^T\left(\frac{2\pi s}{T}\right)\left\{1 + \left(\frac{2\pi s}{\lambda_0 T}\right)^2\right\}^{-1}$$

which are $T^{1/2}$-consistent if $M_T/T \to \infty$ sufficiently fast (Theorem 3). From these, we obtain $T^{1/2}$-consistent estimates of $\hat{a}_j$, $\hat{b}_j$ of the a_j, b_j by the relations:

$$z^n + \sum_{j=1}^n \lambda_0^{-j}\,\hat{a}_j z^{n-j} = \{1 + \sum_j (-1)^j\hat{\alpha}_j\}^{-1}(1 + z)^n + \sum_{j=1}^n \hat{\alpha}_j(1 + z)^{n-j}(1 + z)^j$$

$$z^n + \sum_{j=1}^n \lambda_0^{-1}\,\hat{b}_j z^{n-j} = \{1 + \sum_j(-1)^j\hat{\beta}_j\}^{-1}(1 + z)^n + \sum_{j=1}^n \hat{\beta}_j(1 + z)^{n-j}(1 + z)^j.$$

The choice of λ_0 is arbitrary. One might choose λ_0 small so that the factor $(1 + \lambda^2/\lambda_0^2)^{-1}$ tends to zero rapidly, but this would imply that the polynomials $z^n + \sum a_j z^{n-1}$ and $z^n + \sum \beta_j z^{n-j}$ have zeroes near -1 which can cause difficulty in factoring $\tilde{f}$.

4. Proofs of theorems.

PROOF OF THEOREM 1. From (1.5) and (1.6), we get

$$S_T(\theta, \mu) - S_T(\theta, \hat{\mu}) = \frac{1}{T}(\mu - \hat{\mu})^2 \int_0^T \{1 - \alpha_\theta(t)\}^2\,dt$$
$$- \frac{2}{T}(\mu - \hat{\mu}) \int_0^T \{1 - \alpha_\theta(t)\}[dN(t) - \{\hat{\phi}_{\theta,\hat{\mu}}(t) + \hat{\mu}\}\,dt]$$

where

$$\alpha_\theta(t) = \int_0^T A_\theta(t - s)\,ds = \int_0^t A_\theta(s)\,ds.$$

Hence

$$\hat{\mu}_\theta = \hat{\mu} + \left[\int_0^T \{1 - \alpha_\theta(t)\}\,dt\right]^{-1} \int_0^T \{1 - \alpha_\theta(t)\}[dN(t) - \{\hat{\phi}_{\theta,\hat{\mu}}(t) + \hat{\mu}\}\,dt],$$

$$S_T(\theta, \hat{\mu}_\theta) = S_T(\theta, \hat{\mu}) = \frac{1}{T}(\hat{\mu}_\theta - \hat{\mu})^2 \int_0^T \{1 - \alpha_\theta(t)\}^2\,dt.$$

Now, we observe that $\alpha_\theta(t) \to a_\theta(0)$ as $t \to \infty$, and by A1:

$$\int_0^\infty \mathrm{Sup}_{\theta\in C}\,|\,a_\theta(0) - \alpha_\theta(t)\,|\,dt = \int_0^\infty \mathrm{Sup}_{\theta\in C}\left|\int_t^\infty A_\theta(s)\,ds\right| dt$$
$$\leq \int_0^\infty t\,\mathrm{Sup}_{\theta\in C}|A_\theta(t)|\,dt < +\infty.$$

$$\mathrm{Sup}_{\theta\in C}|\alpha_\theta(t)| \le \int_0^t \mathrm{Sup}_{\theta\in C}|A_\theta(s)|\,ds = O(1),$$

$$E\{\mathrm{Sup}_{\theta\in C}|\hat{\phi}_{\theta,\hat{\mu}}(t)|\} \le E\int_0^t \{\sup_{\theta\in C}|A_\theta(t-s)|\}\{dN(s)+\hat{\mu}\,ds\}$$

$$\le \int_0^t \{\sup_{\theta\in C}|A_\theta(t-s)|\}^2 \mu\,ds = O(1).$$

So, using this result, as $T\to\infty$

$$\int_0^T \{1-\alpha_\theta(t)\}^2\,dt = \{1-a_\theta(0)\}^2 T + O(1), \int_0^T \{1-\alpha_\theta(t)\}[dN(t) + \{\hat{\phi}_{\theta,\hat{\mu}}(t)+\hat{\mu}\}\,dt]$$

$$= \{a_\theta(0)-1\}\int_0^T \left[\int_0^T A_\theta(t-s)\,\{dN(s)-\hat{\mu}\,ds\}\right] dt + O(1)$$

$$= \{a_\theta(0)-1\}\int_0^T \alpha_\theta(T-s)\,\{dN(s)-\hat{\mu}\,ds\} + O(1)$$

$$= \{a_\theta(0)-1\}a_\theta(0)\{N(T)-\hat{\mu}T\} + O(1) = O(1)$$

with the $O(1)$ terms being uniform in any compact C of Θ, and in the sense of L^1 norm when they are random. The result follows. The case when A2 holds is proved similarly.

PROOF OF THEOREM 2. We begin by showing the following results

$$c(\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \{a_\theta(\lambda)+a_\theta(-\lambda)\}\,d\lambda \tag{4.1}$$

$$\int_{-\infty}^{\infty} \log\{g_\theta(\lambda)\}\,d\lambda = 2\pi c(\theta). \tag{4.2}$$

We observe that the function $a_\theta(\lambda) + a_\theta(-\lambda)$ is integrable and $a_\theta(z)$ is analytic in the upper half plane, so by (2.1)

$$\int_{-\infty}^{\infty} \{a_\theta(\lambda)+a_\theta(-\lambda)\}\,d\lambda = \lim_{\rho\to\infty} -2\int_0^\pi a_\theta(\rho e^{-\iota\alpha})\rho\, d(e^{i\alpha})$$

$$= -2ic(\theta)i\pi = 2\pi c(\theta).$$

Result (4.2) can be proved similarly, noting that log $g_\theta(\lambda)$ is the real part of $-2\log\{1-a_\theta(\lambda)\}$ which is analytic in the upper half plane and equals $2ic(\theta)\lambda^{-1}+O(\lambda^{-2})$ as $|\lambda|\to\infty$.

Now, $g_\theta^{-1}(\lambda) - 1 = -a_\theta(-\lambda) - a_\theta(-\lambda) + |a_\theta(\lambda)|^2$, therefore

$$\int_{-\infty}^{\infty} \{g_\theta^{-1}(\lambda)-1\}\hat{I}^T(\lambda)\,d = -\int_{-\infty}^{\infty}\{a_\theta(-\lambda)+a_\theta(-\lambda)\}\hat{I}^T(\lambda)\,d\lambda + \int_{\infty}^{\infty} |a_\theta(\lambda)|^2 \hat{I}^T(\lambda)\,d\lambda. \tag{4.3}$$

By the Fubini Theorem, the first integral in the above right hand side equals

$$\frac{1}{2\pi T}\int_0^T\int_0^T\int_{-\infty}^{\infty} \{a_\theta(\lambda)+a_\theta(-\lambda)\}e^{\iota\lambda(t-s)}\,d\lambda\{dN(t)-\hat{\mu}\,dt\}\{dN(s)-\hat{\mu}\,ds\}$$

$$= \frac{1}{T} \int_0^T \int_0^T \{A_\theta(t-s) + A_\theta(s-t)\}\{dN(t) - \hat{\mu}\, dt\}\{dN(s) - \hat{\mu}\, ds\}.$$

Note that $A_\theta(t) + A_\theta(-t)$ is continuous since it is the Fourier transform of an integrable function. Since $A(t) = 0$ for $t < 0$, we take $A_\theta(0) = 0$ to make $\hat{\phi}_{\theta,\mu}(t)$ left continuous. Let τ_ι be the points of increase of $N(t)$ in $[0, T]$, and observe that the function $A_\theta(t) + A_\theta(-t)$ takes value $c(\theta)$ at $t = 0$ by (4.1), the last double integral is

$$c(\theta)N(T) + 2 \textstyle\sum_{\tau_k < \tau_j} A_\theta(\tau_j - \tau_k)$$

$$+ \int_0^T \int_0^T \{A_\theta(t-s) + A_\theta(s-t)\}\hat{\mu}\{\hat{\mu}\, dt\, ds - dN(t)\, ds - dN(s)\, dt\}$$

$$= c(\theta)\hat{\mu}T + 2 \int_0^T \int_0^T A_\theta(t-s)\{dN(t) - \hat{\mu}\, dt\}\{dN(s) - \hat{\mu}\, ds\}.$$

Therefore the first term in the right side of (4.3) equals

$$-\hat{\mu}c(\theta) = \frac{2}{T} \int_0^T \hat{\phi}_{\theta,\hat{\mu}}(t)\ \{dN(t) - \hat{\mu}\, dt\}.$$

Now observe that the function $a_\theta(\cdot)\ \hat{d}^T(\cdot)$ is the Fourier transform of $\hat{\phi}_{\theta,\mu}(\cdot)$. This can be seen by considering a sequence $A_n \to A_\theta$ pointwise and in mean squares for which the Fubini Theorem is applicable, yielding

$$\int_{-\infty}^{\infty} \left[\int_0^T A_n(t-s)\{dN(s) - \hat{\mu}\, ds\} \right] e^{i\lambda t}\, dt = \left\{ \int_{-\infty}^{\infty} A_n(u) e^{i\lambda u}\, du \right\} \hat{d}^T(\lambda).$$

Thus the second term of (4.3) equals $\int \hat{\phi}^2_{\theta,\mu}(t)\, dt$. The first part of the theorem then follows from (1.3), (1.7), (4.2).

To prove the second part of the theorem, note that

$$\sup_{\theta \in C} |\hat{\phi}_{\theta,\hat{\mu}}(t)| \le \int_0^T \sup_{\theta \in C} |A_\theta(t-s)|\ \{dN(s) + \hat{\mu}\, ds\}$$

(4.4)

$$= \int_0^T k(t-s)\{dN(s) + \hat{\mu}\, ds\}, \qquad \text{say.}$$

Denote by $\|\cdot\|_2$ the L^2 norm, from the triangular inequality the L^2 norm of the above left hand side is bounded by

$$\left\| \int_0^T k(t-s)\{dN(s) - \mu\, ds\} \right\|_2 + \|\mu - \hat{\mu}\|_2 \int_0^T k\,(t-s)\, ds.$$

By A1, k is square integrable. Let $\tilde{k}_{T,t}$ be the Fourier transform of the function $s \to 1_{[0,\,T]}(s)\ k(t-s)$, then

$$E\left[\int_0^T k(t-s)\{dN(t) - \mu\, dt\} \right]^2 = \int |\tilde{k}_{T,t}(\lambda)|^2\, f(\lambda)\, d\lambda$$

$$\le K \int |\tilde{k}_{T,t}(\lambda)|^2\, d\lambda = 2\pi K \int_0^T k^2(t-s)\, ds$$

where $K = \operatorname{Sup} f(\lambda)$. Hence the L^2 norm of the left hand side of (4.4) is bounded when $t \ge T$ by

$$K_1 \left\{ \int_{t-T}^{\infty} k^2(s)\, ds \right\}^{1/2} + K_2 \int_{t-T}^{\infty} k(s)\, ds \le K_3 \left\{ \int_{t-T}^{\infty} k(s)\, ds \right\}^{1/2}$$

where K_1, K_2, K_3 are constants. Thus, using A1

$$E\left\{\sup_{\theta\in C}\int_T^\infty \hat{\phi}^2_{\theta,\hat{\mu}}(t)\,dt\right\} \le \int_T^\infty E\{\sup_{\theta\in C}|\hat{\phi}_{\theta,\hat{\mu}}(t)|\}^2\,dt$$

$$\le K_3^2\int_T^\infty\left\{\int_{t-T}^\infty k(s)\,ds\right\}dt = K_3^2\int_0^\infty sk(s)\,ds.$$

The last part of the theorem is proved similarly.

PROOF OF THEOREM 3. We first replace $I^T(\lambda)$, $\hat{I}^T(\lambda)$ by

$$\tilde{I}^T(\lambda) = \frac{1}{2\pi T}|\tilde{d}^T(\lambda)|^2, \qquad \tilde{d}^T(\lambda) = \int_0^T e^{i\lambda t}\{dN(t) - \mu\,dt\}.$$

Since $I^T(2\pi s/T) = \tilde{I}^T(2\pi s/T)$, $s \neq 0$, $J^T(p_j)$ will be unchanged whereas $\hat{J}^T(p_j)$ will become:

$$\tilde{J}^T(p_j) = \hat{J}^T(p_j) - T^{-2}\int p_j(\lambda)|\Delta^T(\lambda)|^2\tilde{I}(0)\,d\lambda$$
$$+ \frac{1}{2\pi T^2}\int p_j(\lambda)\{\Delta^T(\lambda)\tilde{d}^T(-\lambda) + \Delta^T(-\lambda)\tilde{d}^T(\lambda)\}\tilde{d}^T(0)\,d\lambda$$

where $\Delta^T(\lambda) = \int_0^T e^{i\lambda t}\,dt$. Now from Brillinger's (1972) result

$$\mathrm{Cum}\{d^T(\lambda_1), \cdots, d^T(\lambda_k) = \Delta^T(\lambda_1 + \cdots + \lambda_k)f^{(k)}(\lambda_1, \cdots, \lambda_{k-1}) + O(1)$$

it can be checked that $E\tilde{J}^T(p_j) = E\hat{J}^T(p_j) + O(T^{-1})$ and deduce the expressions for $E\hat{J}^T(P_j)$, $EJ^T(p_j)$ as in the theorem.

To establish the asymptotic normality of the μ, $J^T(p_j)$, we first suppose that the p_j have compact support; then, as in Brillinger (1975a, page 547) the various joint cumulants of these random variables can be shown to converge to the appropriate limits, as $T \to \infty$. Now, from the above results, it can be seen that

$$Var\{\tilde{J}^T(p_j) - J^T(p_j)\} \le \text{const}\; T^{-2}\left\{\int|\Delta^T(\lambda)|\,d\lambda\right\}^2 = 0((T^{-1}\log T)^2).$$

Thus to obtain the asymptotic normality of $\hat{J}_T(p_j)$ we need only to show that $\mathrm{Var}\{\tilde{J}^T(p_j) - J^T(p_j)\} = o(T^{-1})$ or equivalently $\mathrm{Var}\{\tilde{J}^T(p_j)$, $\mathrm{Var}\{J^T(p_j)\}$, $\mathrm{Cov}\{\tilde{J}^T(j), J^T(p_j)\}$ differ from each other by a term $o(T^{-1})$. This can be obtained from

$$\mathrm{Cov}\{\tilde{I}^T(\alpha), \tilde{I}^T(\beta)\} = T^{-2}\{|\Delta^T(\alpha-\beta)|^2 + |\Delta^T(\alpha+\beta)|^2\}f^2(\alpha)$$
$$+ 2\pi T^{-1}f^{(4)}(\alpha, \beta, -\alpha) + \{\Delta^T(\alpha-\beta) + \Delta^T(\alpha+\beta) + 1\}O(T^{-2}).$$

In case the p_j do not have compact support, we use Berstein's lemma (Hannan, 1970, page 242). Write

$$\tilde{J}^T(p_j) = \int_{-l}^{l} p_j(\lambda)\tilde{I}^T(\lambda)\,d\lambda + \epsilon_{T,l}, \qquad l > 0.$$

All we need is to show that $T\,\mathrm{Var}(\epsilon_{T,l}) \to 0$ as $l \to \infty$, uniformly in T, using the above expression for $\mathrm{Cov}\{\tilde{I}^T(\alpha), \tilde{I}^T(\beta)\}$. A similar argument applies for the $J^T(p_j)$.

PROOF OF THEOREM 5. The first result of part 1 of the theorem follows directly from Theorem 3 and the assumptions on g_θ. To obtain the second result, we shall show that $J^T(p_\theta)$ and $\hat{J}^T(p_\theta)$ satisfy the condition (ii) of Theorem 4 if $p_\theta(\lambda)$ is jointly continuous in (θ, λ) and for some $\delta > 1$, $|\lambda|^\delta p_\theta(\lambda) \to 0$ as $(\theta, \lambda) \to (\theta', \pm\infty)$ for any θ'.

Let $\mu > 0$, $\theta' \in \Theta$, $\lambda' \in R$, by assumption there exist $\lambda_0 = \lambda_0(\theta')$ and neighborhoods

$V(\theta')$, $V(\theta'; \lambda')$ of θ' and $U(\lambda') = U(\lambda'; \theta')$ of λ' such that

$$\theta \in V(\theta'), \lambda > \lambda_0 : (1 + |\lambda|^\delta)|p_\theta(\lambda)| < \eta/2$$

$$\theta \in V(\theta'; \lambda'), \lambda \in U(\lambda') : (1 + |\lambda|^\delta)|p_{\theta'}(\lambda) - p_{\theta'}(\lambda)| < \eta$$

The set $\{\lambda \geq \lambda_0\}$, being compact, can be covered by $U(\lambda_1), \cdots, U(\lambda_k)$ for some $\lambda_1, \cdots, \lambda_k$. Let $W(\theta')$ be the intersection of $V(\theta')$ and $V(\theta', \lambda_i)$, $i = 1, \cdots, k$. Then

$$(1 + |\lambda|^\delta)|p_{\theta'}(\lambda) - p_\theta(\lambda)| < \eta, \qquad \forall\lambda.$$

Cover the compact set C by $W(\theta_1), \cdots, W(\theta_r)$. One can assume they are open sets, so for $\theta \in C$, $\gamma_j(\theta) = \inf\{\|\theta' - \theta\|, \theta' \notin W(\theta_j)\} > 0$ for some j. Hence max γ_j admits a minimum $\gamma > 0$ on C. Now $\|\theta' - \theta\| < \gamma$, $\theta, \theta' \in C$ implies θ, θ' both belong to some $W(\theta_j)$, hence $|p_{\theta'}(\lambda) - p_\theta(\lambda)| \leq 2\eta/(1 + |\lambda|^\delta)$. We deduce that $|J^T(p_{\theta'}) - J^T(p_\theta)| \leq \eta R_T$, $|\hat{J}^T(p_{\theta'}) - J^T(p_\theta)| \leq \eta\hat{R}_T$ where ER_T, $E\hat{R}_T$ are bounded. The result follows.

Clearly, the above results also show that $\tilde{\Lambda}_T$ and $\tilde{L}_T$ satisfy the condition (iv) of theorem 4. To show that $\tilde{\Lambda}_T$ satisfies the condition (iii) of this theorem with J, H, K as given in Theorem 5, we write $(\partial/\partial\theta_j)\tilde{\Lambda}_T$ in the form

$$\frac{\partial}{\partial\theta_j}\Lambda(\theta) - \frac{1}{\mu}\int_0^\infty \frac{\partial}{\partial\theta_j} g_\theta^{-1}(\hat{I} - f)\, d\lambda + \left(\frac{1}{\mu} - \frac{1}{\hat{\mu}}\right)\int_0^\infty \frac{\partial}{\partial\theta_j} g_\theta^{-1}\hat{I}\, d\lambda,$$

where the variable λ is omitted. The result then follows from Theorem 3, noting that the derivatives of Λ vanish at θ^* and hence when $\theta = \theta^*$

$$\int_0^\infty \frac{\partial}{\partial\theta_j} g_\theta^{-1} f\, d\lambda = -\frac{\mu}{2\pi}\int_0^\infty \frac{\partial}{\partial\theta_j}\log g_\theta\, d\lambda = -\frac{\mu}{2}\frac{\partial}{\partial\theta_j}c(\theta)$$

by (4.2). The same argument applies for Λ_{T,M_T}, provided

$$\sqrt{T}\left[\frac{1}{T}\sum_{s=1}^{M_T}\frac{\partial}{\partial\theta_j}\log g_\theta\left(\frac{2\pi s}{T}\right) + \frac{2\pi}{\mu}g_\theta^{-1}\left(\frac{2\pi s}{T}\right)f\left(\frac{2\pi s}{T}\right)\right]_{\theta=\theta^*} \to 0$$

as $t \to \infty$. Since the function following the summation sign, β_j say, is of bounded variation, the above expression differs from (2.2) by a term $\sqrt{T}\ O(T^{-1})$ (Brillinger, 1975a, page 415). The result follows.

REFERENCES

BOX, G. P. E., and JENKINS, G. M. (1970). *Time Series Analysis Forecasting and Control.* Holden-Day, San Francisco.

BRILLINGER, D. R. (1972). The spectral analysis of stationary interval function. *Proc. Sixth. Berkeley. Symp. Math. Statist. Prob.* 1 483–513. Univ. of Calif., Berkeley.

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

BRILLINGER, D. R. (1975b). Statistical inference for stationary point processes. In *Stochastic Processes and Related Topics.* 1 55–99. (Madan Lal Puri, ed.) Academic, New York.

COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events.* Methuen, London.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

JOWETT, J. and VERE-JONES, D. (1972). The prediction of stationary point process. In *Stochastic Point Process.* (P.A.W. Lewis, ed.) Wiley, New York.

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

HAWKES, A. G. (1972). Spectra for some mutually exciting point process with associated variable. In *Stochastic Point Process.* (P.A.W. Lewis, ed.) Wiley, New York.

MATHÉMATIQUE APPLIQUÉES
UNIVERSITY OF GRENOBLE
38041 GRENOBLE CÉDEX
FRANCE