Adv. Appl. Prob. **5**, 469–497 1973
Printed in Israel

ASYMPTOTIC INFERENCE IN STATIONARY GAUSSIAN TIME-SERIES

ROBERT B. DAVIES, *DSIR, Wellington, N. Z.*

Abstract

Conditions are given for the family of distributions of a stationary, discrete-time, Gaussian, vector-valued time-series with covariance structure given up to a finite number of parameters to satisfy the asymptotic differentiability conditions introduced by Le Cam (1969).

ASYMPTOTIC INFERENCE; CONTIGUITY; FOURIER SERIES; PERIODOGRAM; SPECTRUM; STATIONARY GAUSSIAN TIME-SERIES; TOEPLITZ MATRIX

1. Introduction

We observe n consecutive (vector) observations $x_0, \cdots, x_{n-1}$ from a stationary r-dimensional (real) Gaussian process, where n is "large". We will assume for the moment that the expectations of the x_i are zero. Suppose that the covariance structure is known up to a finite number of parameters, $\theta_1, \cdots, \theta_s$ (denoted collectively by the column vector, θ). In this paper we will show that under certain conditions the problem of making statistical inferences about these parameters falls into a class considered by Le Cam (1960), (1969). Since this theory, concerned with "Asymptotically differentiable families", is not yet well known a brief summary of the main results is given in Appendix I. It will enable us, given a "good" estimator of θ to obtain asymptotically optimal tests and estimators with properties similar to the classical properties of likelihood ratio tests and maximum likelihood estimators in the random sample situation. As is well known if one takes a discrete Fourier transform of the observed series one obtains a sequence of $\frac{1}{2}n$ approximately independent random variables so it is hardly surprising that results similar to the classical ones can be obtained.

Whittle (1953), (1962) and Walker (1964) have investigated the properties of the maximum likelihood and related estimators for the present problem and Mann and Wald (1943) and Whittle (1951) amongst others have considered these estimators for the autoregressive/moving average case. Hannan ((1970), chapter VI) and Parzen (1971) have introduced, for this special case, asymp-

Received in revised form 2 March 1973.

totically optimal estimators which are based on a preliminary non-optimal estimator; estimators not unlike that considered in the present work.

In Section 2 the notation and basic assumptions are introduced, several lemmata are proved in Section 3 and the basic theorems proved in Section 4. The results are summarized in Section 5. A simple example is also given in that section. The matrix notation and some results used are given in Appendix II.

2. Notation and basic assumptions

The matrix notation is given in Appendix II. In particular note that we will use * to denote both the transpose of a real matrix or vector and the conjugate transpose of a complex matrix or vector. Also $\|A\|$ denotes the Hilbert norm of a matrix A and $|A|$ the Euclidean norm. For vectors these two norms coincide and will be written as $|\cdot|$. The indices of the elements of the $n \times n$ matrices or of the $r \times r$ sub-blocks of the $nr \times nr$ matrices introduced in this section will run from 0 to $n-1$ rather than from 1 to n with a corresponding notation for n and nr-dimensional vectors.

The observed sequence of random variables $x_0, \cdots, x_{n-1}$ will be denoted by the nr-dimensional column vector X_n. We will suppose

$$E(X_n) = 0. \tag{2.1}$$

Also let the $r \times r$ matrix

$$c_m(\theta) = E(x_k x_{k+m}^*) \tag{2.2}$$

and the $nr \times nr$ symmetric matrix

$$C_n(\theta) = E(X_n X_n^*)$$
$$= \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{-1} & c_0 & c_1 & \cdots & c_{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{-n+1} & c_{-n+2} & c_{-n+3} & \cdots & c_0 \end{bmatrix}. \tag{2.3}$$

By Ω_n will be meant the $nr \times nr$ unitary matrix

$$\Omega_n = \omega_n \otimes I_r \tag{2.4}$$

where ω_n denotes the $n \times n$ matrix with $\exp(2\pi i jk/n)/n^{\frac{1}{2}}$ for its (j, k)th element $(0 \leqq j, k \leqq n-1)$, I_r the $r \times r$ identity matrix and $\otimes$ the Kronecker product.

Let

$$Z_n = \begin{bmatrix} z_{n,0} \\ \vdots \\ z_{n,n-1} \end{bmatrix} = \Omega_n X_n \tag{2.5}$$

so that

$$z_{n,k} = n^{-\frac{1}{2}} \sum_{j=0}^{n-1} x_j \exp\{2\pi i jk/n\}$$

and let

$$f(\lambda, \theta) = \sum_{-\infty}^{\infty} c_m(\theta) \exp\{2\pi i m\lambda\}. \tag{2.6}$$

Condition A1.2 below will ensure the convergence of (2.6). Note that $z_{n,m}\, z_{n,m}^*$ gives the value of the matrix periodogram at frequency $2\pi m/n$ and $f(\lambda, \theta)$, an $r \times r$ matrix, the spectrum of the process. Finally define $F_n(\theta)$ to be the $nr \times nr$ matrix

$$F_n(\theta) = \operatorname{diag}\{f(0, \theta), f(1/n, \theta), \cdots, f((n-1)/n, \theta)\} \tag{2.7}$$

and

$$G_n(\theta) = \Omega_n C_n(\theta) \Omega_n^* - F_n(\theta). \tag{2.8}$$

We will see that in a certain sense $G_n(\theta)$ is small, a result that is not unexpected in view of the approximate independence of spectral estimates.

The basic assumptions which we will need to make are as follows.

A0. The set of possible values of θ is an open set, Θ in s-dimensional Euclidean space.

A1.1. The $c_m(\theta)$ are differentiable functions of θ.

A1.2.
$$\sum_{m=-\infty}^{\infty} \| c_m(\theta) \| < \infty,$$

$$\lim_{\varepsilon \to 0} \sum_{m=-\infty}^{\infty} \| c_m(\theta + \varepsilon) - c_m(\theta) \| = 0, \qquad \text{for all } \theta \in \Theta.$$

A1.3.
$$\sum_{m=-\infty}^{\infty} \left| \frac{\partial}{\partial \theta_k} c_m(\theta) \right|^2 < \infty,$$

$$\lim_{\varepsilon \to 0} \sum_{m=-\infty}^{\infty} \left| \frac{\partial}{\partial \theta_k} c_m(\theta + \varepsilon) - \frac{\partial}{\partial \theta_k} c_m(\theta) \right|^2 = 0,$$

for all k; $1 \leqq k \leqq s$ and $\theta \in \Theta$.

A1.4. $\det f(\lambda, \theta) > 0$ all λ; $0 \leqq \lambda < 1$ and $\theta \in \Theta$.

Other conditions which will be needed for some of the results are the following.

A2.
$$\sum_{k=1}^{s} t_k \frac{\partial}{\partial \theta_k} f(\lambda, \theta) \neq 0$$

for some non-null set of λ for each r-dimensional vector $t \neq 0$ and $\theta \in \Theta$.

A3.1.
$$\sum_1^\infty m^{\frac{1}{2}} \| c_m(\theta) \| < \infty .$$

A3.2.
$$\sum_1^\infty \left\| \frac{\partial}{\partial \theta_k} c_m(\theta) \right\| < \infty \quad \text{each} \quad k, \theta.$$

A3.3.
$$\sup_{\substack{|\xi| < \delta \\ |\eta| < \delta \\ \xi \neq \eta}} \sum_{-\infty}^{\infty} \left\| \frac{\partial}{\partial \theta_k} c_m(\theta + \xi) - \frac{\partial}{\partial \theta_k} c_m(\theta + \eta) \right\| / |\xi - \eta| < \infty ,$$

some δ, each k, θ.

3. Lemmata

In this section we will consider some properties of Fourier series, covariance matrices, etc. The first lemma is concerned with generalizing some properties of Toeplitz matrices — see Grenander and Szegö (1958), Sections 5.2 and 7.6.

Lemma 3.1. Suppose $a_j : j = 0, \pm 1, \pm 2, \cdots$ is a sequence of $r \times r$ matrices with $\Sigma |a_j|^2 < \infty$, A_n is an $nr \times nr$ matrix composed of n^2 $r \times r$ submatrices, the (j, k)th one being a_{j-k} $(0 \leqq j,\ k < n)$, and

$$\alpha(\lambda) = \sum_{-\infty}^{\infty} a_j e^{2\pi i j \lambda},$$

then the following hold.

(i)
$$\| A_n \| \leqq \sup_\lambda \| \alpha(\lambda) \| \leqq \sum_{-\infty}^{\infty} \| a_j \| .$$

(ii) If $\alpha(\lambda)$ is positive definite hermitian for all λ then

$$\| A_n^{-1} \| \leqq \sup_\lambda \| \alpha^{-1}(\lambda) \| .$$

(iii)
$$|A_n|^2 / n \uparrow \sum_{-\infty}^{\infty} |a_j|^2 = \int_0^1 |\alpha(\lambda)|^2 d\lambda .$$

(iv) If $\Sigma_{-\infty}^{\infty} |a_j| < \infty$ then

$$\lim_{n \to \infty} | \Omega_n A_n \Omega_n^* - \operatorname{diag} \{ \alpha(0), \alpha(1/n), \cdots, \alpha((n-1)/n) \} | / n^{\frac{1}{2}} = 0 .$$

(v) Suppose (B, b, β) and (C, c, γ) are related in the same way as (A, a, α) above and are such that

$$\sup_\lambda \| \alpha(\lambda) \| < \infty, \ \sum_{-\infty}^{\infty} |b_j|^2 < \infty \ \text{ and } \ \gamma(\lambda) = \alpha(\lambda) \beta(\lambda).$$

Then

$$c_k = \sum_{-\infty}^{\infty} a_j b_{k-j},$$

$$\sum_{-\infty}^{\infty} |c_j|^2 < \infty,$$

$$|C_n - A_n B_n| / n^{\frac{1}{2}} \to 0 \text{ as } n \to \infty.$$

(vi) If $\sum_{-\infty}^{\infty} |a_j|^2 < \infty$ then $n^{-\frac{1}{2}} \| A_n \| \to 0$ as $n \to \infty$.

Proof. (i) Suppose $X = (x_0, \cdots, x_{n-1})^*$, $Y = (y_0, \cdots, y_{n-1})$ are nr-dimensional vectors. Then

$$\begin{aligned}
|X^* A_n Y| &= \left| \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} x_j^* a_{j-k} y_k \right| \\
&= \left| \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \int_0^1 x_j^* \alpha(\lambda) y_k \exp\{-2\pi i (j-k)\lambda\} d\lambda \right| \\
&= \left| \int_0^1 \left\{ \sum_{j=0}^{n-1} x_j e^{2\pi i j \lambda} \right\}^* \alpha(\lambda) \sum_{k=0}^{n-1} y_k e^{2\pi i k \lambda} d\lambda \right| \\
&\leqq \int_0^1 \left| \sum_{j=0}^{n-1} x_j e^{2\pi i j \lambda} \right| \| \alpha(\lambda) \| \left| \sum_{k=0}^{n-1} y_k e^{2\pi i k \lambda} \right| d\lambda \\
&\leqq \sup_\lambda \| \alpha(\lambda) \| \left\{ \int_0^1 \left| \sum_{j=0}^{n-1} x_j e^{2\pi i j \lambda} \right|^2 d\lambda \int_0^1 \left| \sum_{k=0}^{n-1} y_k e^{2\pi i k \lambda} \right|^2 d\lambda \right\}^{\frac{1}{2}} \\
&= \sup_\lambda \| \alpha(\lambda) \| \, |X| \, |Y|.
\end{aligned}$$

Thus $\| A_n \| \leqq \sup_\lambda \| \alpha(\lambda) \|$.

(ii) The proof is similar to that of (i) but uses the equality for positive definite matrices

$$\| A^{-1} \| = \sup_X \{ |X|^2 / X^* A X \}.$$

With the notation as in (i)

$$\begin{aligned}
X^* A_n X &\geqq \int_0^1 \left| \sum_{k=0}^{n-1} x_k e^{2\pi i k \lambda} \right|^2 \| \alpha^{-1}(\lambda) \|^{-1} d\lambda \\
&\geqq \inf_\lambda \| \alpha^{-1}(\lambda) \|^{-1} |X|^2
\end{aligned}$$

and the result follows.

(iii) The limit result follows from

$$|A_n|^2 = \sum_{j=1-n}^{n-1} (n - |j|)\, |a_j|^2$$

and the equality from Parseval's theorem.

(iv) Let $G = \Omega_n A_n \Omega_n^* - \operatorname{diag}\{\alpha(0), \alpha(1/n), \cdots, \alpha((n-1)/n)\}$. Then for $j \neq k$ $(0 \leqq j,\ k \leqq n-1)$ the (j,k)th $r \times r$ submatrix of G

$$\begin{aligned} g_{jk} &= \frac{1}{n} \sum_{p=0}^{n-1} \sum_{q=0}^{n-1} \exp\{2\pi i jp/n\} a_{p-q} \exp\{-2\pi i kq/n\} \\ &= -\frac{1}{n} \left\{ \sum_{p=1}^{n-1} a_p \exp\{2\pi i jp/n\} \sum_{q=n-p}^{n-1} \exp\{2\pi i (j-k) q/n\} \right. \\ &\qquad \left. + \sum_{p=1-n}^{-1} a_p \exp\{2\pi i jp/n\} \sum_{q=0}^{-p-1} \exp\{2\pi i (j-k) q/n\} \right\} \end{aligned}$$

and for $j = k$

$$g_{jj} = -\sum_{p=-\infty}^{\infty} \min(1, |p|/n) a_p \exp\{2\pi i jp/n\}.$$

Thus we can represent

$$G = \frac{1}{n} \sum_{p=-\infty}^{\infty} \gamma_p \otimes a_p$$

where γ_p is a matrix with (j,k)th element equal to

$$\begin{aligned} &-\exp\{2\pi i jp/n\} \sum_{q=n-p}^{n-1} \exp\{2\pi i (j-k) q/n\} && \text{if } 0 \leqq p < n, \\ &-\exp\{2\pi i jp/n\} \sum_{q=0}^{-p-1} \exp\{2\pi i (j-k) q/n\} && \text{if } -n < p \leqq 0, \\ &-n \exp\{2\pi i jp/n\} && \text{if } j = k \text{ and } |p| \geqq n \end{aligned}$$

and zero otherwise. For $0 \leqq p < n$

$$\begin{aligned} |\gamma_p|^2 &= \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \left| \sum_{q=n-p}^{n-1} \exp\{2\pi i (j-k) q/n\} \right|^2 \\ &= \sum_{q=n-p}^{n-1} \sum_{r=n-p}^{n-1} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \exp\{2\pi i (j-k)(q-r)/n\} \\ &= n^2 p \end{aligned}$$

and similarly $|\gamma_{-p}|^2 = n^2 p$. Hence

$$n^{-\frac{3}{2}}|G| \leqq n^{-\frac{3}{2}} \sum_{p=-\infty}^{\infty} |\gamma_p|\,|a_p|$$

$$= \sum_{p=-\infty}^{\infty} \min\{1, (|p|/n)^{\frac{1}{2}}\}\,|a_p|$$

and the result follows.

(v) The equality follows from the convolution formula and the inequality from Parseval's equality. The matrix $C_n - A_n B_n$ has (i,j)th submatrix

$$\sum_{k=-\infty}^{\infty} a_k b_{i-j-k} - \sum_{k=i-n+1}^{i} a_k b_{i-j-k}$$

$$= \sum_{k=-\infty}^{i-n} a_k b_{i-j-k} + \sum_{k=i+1}^{\infty} a_k b_{i-j-k}.$$

Thus it is sufficient to show that

$$\frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \left| \sum_{k=-\infty}^{i-n} a_k b_{i-j-k} \right|^2$$

and

$$\frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \left| \sum_{k=i+1}^{\infty} a_k b_{i-j-k} \right|^2$$

tend to zero as n tends to infinity. Consider the first term — the second may be handled similarly. Let $b_{j,i} = b_i$ if $i \geqq n-j$ and 0 otherwise and let $\beta_j(\lambda)$ be the function corresponding to $b_{j,i}$. Then the first term is not greater than

$$\frac{1}{n} \sum_{i=-\infty}^{\infty} \sum_{j=0}^{n-1} \left| \sum_{k=-\infty}^{\infty} a_k b_{j,i-j-k} \right|^2$$

$$= \frac{1}{n} \sum_{j=0}^{n-1} \int_0^1 |\alpha(\lambda)\beta_j(\lambda)|^2 \, d\lambda$$

$$\leqq \frac{1}{n} \sup_{\lambda} \|\alpha(\lambda)\|^2 \sum_{j=0}^{n-1} \int_0^1 |\beta_j(\lambda)|^2 d\lambda$$

$$= \sup \|\alpha(\lambda)\|^2 \frac{1}{n} \sum_{j=0}^{n-1} \sum_{i=n-j}^{\infty} |b_i|^2,$$

the equalities following from Result (iii) of the lemma and the inequality from Appendix II (iii). The final term tends to zero as n tends to infinity and so (v) follows.

(vi) $$n^{-\frac{1}{2}}\|A_n\| \leqq n^{-\frac{1}{2}} \sum_{j=1-n}^{n-1} |a_j|$$

$$\leqq n^{-\frac{1}{2}} \sum_{j=1-k}^{k-1} |a_j| + \left(\sum_{j=k}^{n-1} |a_j|^2\right)^{\frac{1}{2}} + \left(\sum_{j=k}^{n-1} |a_{-j}|^2\right)^{\frac{1}{2}}.$$

The second and third terms can be made arbitrarily small and the first term tends to zero as n tends to infinity. This completes the proof of Lemma 3.1.

The second lemma establishes that $f(\lambda, \theta)$ may be differentiated with respect to θ in an L_2 sense and that the derivative is the L_2 limit of the Fourier series of the derivatives of the covariances.

Lemma 3.2. Suppose Assumptions A1.1, A1.2, A1.3 hold. Then for each k

$$\int_0^1 \left|\frac{1}{\varepsilon}[f(\lambda, \theta + \varepsilon t_k) - f(\lambda, \theta)] - h(\lambda, \theta)\right|^2 d\lambda \to 0 \text{ as } \varepsilon \to 0,$$

where t_k has unit kth component and other components zero and $h(\lambda, \theta)$ is the L_2 limit of

$$\sum_{j=-N}^{N} \frac{\partial}{\partial \theta_k} c_j(\theta) e^{2\pi i j \lambda}$$

as N tends to infinity.

Proof.

$$\sum_{j=-N}^{N} \left|\frac{\partial}{\partial \theta_k} c_j(\theta) - \frac{1}{\varepsilon}\{c_j(\theta + \varepsilon t_k) - c_j(\theta)\}\right|^2$$

$$\leqq \sum_{j=-N}^{N} \left|\frac{\partial}{\partial \theta_k} c_j(\theta) - \frac{\partial}{\partial \theta_k} c_j(\theta + \xi t_k)\right| \cdot \left|\frac{\partial}{\partial \theta_k} c_j(\theta) - \frac{\partial}{\partial \theta_k} c_j(\theta + \eta t_k)\right|$$

for a suitable choice of ξ, η between 0 and ε by Appendix II (viii). This term is bounded by

$$\left\{\sum_{j=-\infty}^{\infty} \left|\frac{\partial}{\partial \theta_k} c_j(\theta) - \frac{\partial}{\partial \theta_k} c_j(\theta + \xi t_k)\right|^2 \cdot \sum_{j=-\infty}^{\infty} \left|\frac{\partial}{\partial \theta_k} c_j(\theta) - \frac{\partial}{\partial \theta_k} c_j(\theta + \eta t_k)\right|^2\right\}^{\frac{1}{2}}$$

which tends to zero uniformly in N as ε tends to zero. Hence

$$\int_0^1 \left|h(\lambda, \theta) - \frac{1}{\varepsilon}\{f(\lambda, \theta + \varepsilon t_k) - f(\lambda, \theta)\}\right|^2 d\lambda$$

$$= \sum_{j=-\infty}^{\infty} \left| \frac{\partial}{\partial\theta_k} c_j(\theta) - \frac{1}{\varepsilon}\{c_j(\theta+\varepsilon t_k) - c_j(\theta)\} \right|^2$$

tends to zero as ε tends to zero. This completes the proof.

Applying these results we obtain Corollary 3.3.

Corollary 3.3. Under Assumptions A1.1, A1.2, A1.3, A1.4:

(i) $$\| C_n(\theta) \| \leqq \sup_\lambda \| f(\lambda, \theta) \| \leqq \sum_{-\infty}^{\infty} \| c_j(\theta) \| < \infty,$$

$$\| C_n(\theta + t) - C_n(\theta) \| \leqq \sup_\lambda \| f(\lambda, \theta + t) - f(\lambda, \theta) \|$$

$$\leqq \sum_{-\infty}^{\infty} \| c_j(\theta + t) - c_j(\theta) \| \to 0 \text{ as } t \to 0;$$

(ii) $$\| C_n^{-1}(\theta) \| \leqq \sup_\lambda \| f^{-1}(\lambda, \theta) \| < \infty,$$

$$\sup_{|t|<\delta} \| C_n^{-1}(\theta + t) \| \leqq \sup_{|t|<\delta} \sup_\lambda \| f^{-1}(\lambda, \theta + t) \|$$

$$< \infty \text{ for some } \delta > 0;$$

(iii) $$\frac{1}{n} \left| \frac{\partial}{\partial\theta_k} C_n(\theta) \right|^2 \leqq \sum_{-\infty}^{\infty} \left| \frac{\partial}{\partial\theta_k} c_j(\theta) \right|^2$$

$$= \int_0^1 \left| \frac{\partial}{\partial\theta_k} f(\lambda, \theta) \right|^2 d\lambda < \infty,$$

$$\frac{1}{n} \left| \frac{\partial}{\partial\theta_k} C_n(\theta + t) - \frac{\partial}{\partial\theta_k} C_n(\theta) \right|^2$$

$$\leqq \sum_{-\infty}^{\infty} \left| \frac{\partial}{\partial\theta_k} c_j(\theta + t) - \frac{\partial}{\partial\theta_k} c_j(\theta) \right|^2$$

$$= \int_0^1 \left| \frac{\partial}{\partial\theta_k} f(\lambda, \theta + t) - \frac{\partial}{\partial\theta_k} f(\lambda, \theta) \right|^2 d\lambda$$

$$\to 0 \text{ as } t \to 0;$$

(iv) the matrix $G_n(\theta)$ defined in Equation 2.8 satisfies

$$n^{-\frac{1}{2}} | G_n(\theta) | \to 0 \text{ as } n \to \infty.$$

(v) $$n^{-\frac{1}{2}} \left\| \frac{\partial}{\partial\theta_k} C_n(\theta) \right\| \to 0 \text{ as } n \to \infty.$$

Proof. Part (i) is derived from Lemma 3.1 (i) and A1.2; (ii) from 3.1 (ii), A1.4, 3.3 (i) and Appendix II (xiii); (iii) from 3.1 (iii) and A1.3; (iv) from 3.1 (iv) and A1.2; and (v) from 3.1 (vi) and A1.3.

A further result following from Corollary 3.3 (iii) and Formula (ix) in Appendix II is Corollary 3.4.

Corollary 3.4. $\| C_n(\theta + n^{-\frac{1}{2}}t) - C_n(\theta) \|$ is a bounded sequence in n for fixed θ, t.

The following theorem is relevant to the discussion of the robustness of the procedures described in this paper.

Theorem 3.5. Suppose u_i, $i = 0, 1, \cdots$, is a sequence of independent identically distributed r-dimensional (not necessarily normal) random variables with zero expectation and variance-covariance matrix σ, and p_i, $i = 0, \pm 1, \cdots$, is a sequence of $r \times r$ matrices with $\sum_{-\infty}^{\infty} \| p_i \|^2 < \infty$, $p_i = p_{-i}^*$ and $p_0 = 0$. Then

$$\mathscr{L}\left\{ n^{-\frac{1}{2}} \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} u_j^* p_{j-k} u_k \right\}$$

$$\to \mathscr{N}\left(0, 2 \sum_{-\infty}^{\infty} \| \sigma^{\frac{1}{2}} p_j \sigma^{\frac{1}{2}} \|^2 \right)$$

as n tends to infinity. ($\mathscr{L}$ denotes the law or distribution of a random variable and $\mathscr{N}(m, v)$ denotes the normal distribution with mean, m, variance, v.)

Proof. Suppose U_n, P_n denote the column vector $(u_0, \cdots, u_{n-1})^*$ and matrix with (j, k)th $r \times r$ submatrix $p_{j-k} (0 \leqq j, k < n)$. Then the theorem may be proved by showing that $U_n^* P_n U_n$ is approximated by $U_n^* Q_n U_n$, where Q_n is obtained from P_n by deleting all elements except a sequence of $m \times m$ blocks down the diagonal, and then applying the central limit theorem when n is increased, m remaining fixed.

4. Asymptotic differentiability

We will now show that, under assumptions introduced in Section 2, Le Cam's asymptotic differentiability conditions hold. See Appendix I for a statement of these conditions and some of their implications. Let $P_n(\theta)$ denote the probability measure corresponding to X_n under θ and let the log likelihood ratio be denoted by

$$\Lambda_n(\theta_1, \theta_2) = \log [dP_n(\theta_1) / dP_n(\theta_2)]. \tag{4.1}$$

In what follows we will denote $C_n(\theta + n^{-\frac{1}{2}}t)$, $F_n(\theta + n^{-\frac{1}{2}}t)$, etc, where t is an s-dimensional vector, by C_t, F_t, etc, and $\Lambda_n(\theta + n^{-\frac{1}{2}}t, \theta)$ by Λ_t. In this notation

$$\Lambda_t = \tfrac{1}{2}(\log \det C_0 - \log \det C_t + X_n^* C_0^{-1} X_n - X_n^* C_t^{-1} X_n). \tag{4.2}$$

Note that if A_n is a symmetric matrix ($rn \times rn$)

$$E(X_n^* A_n X_n) = \operatorname{tr}(C_n A_n), \tag{4.3}$$

$$\operatorname{Var}(X_n^* A_n X_n) = 2\,|C_n^{\frac{1}{2}} A_n C_n^{\frac{1}{2}}|^2, \tag{4.4}$$

and hence to show that $X_n^* A_n X_n + a_n$ is either bounded or tends to zero in probability it is sufficient to show that $\operatorname{tr}(C_n A_n) + a_n$ and $|C_n^{\frac{1}{2}} A_n C_n^{\frac{1}{2}}|$ are correspondingly bounded or tend to zero. This result will be used repeatedly throughout this section. In what follows it will be assumed that Assumptions A0, A1.1–A1.4 hold and that θ belongs to Θ. The reader is reminded to refer to Appendix II for matrix notation and elementary results.

Theorem 4.1. The two sequences of probability measures $\{P_n(\theta)\}_{n=1,2,\ldots}$ and $\{P_n(\theta + n^{-\frac{1}{2}}t)\}_{n=1,2,\ldots}$ are contiguous.

Proof. It is sufficient to show that Λ_t is bounded in P_s probability where $s = 0$ or t.

(i) Considering $E(\Lambda_t)$:

$$E(\Lambda_t) = \tfrac{1}{2}\{\operatorname{tr}[C_s(C_0^{-1} - C_t^{-1})] + \log\det C_0 - \log\det C_t\}$$

$$= \tfrac{1}{2} n^{-\frac{1}{2}} \operatorname{tr}[C_s C_u^{-1}(t\nabla C_u) C_u^{-1} - (t\nabla C_u) C_u^{-1}]$$

where $t\nabla C_u = \sum t_k \partial C_u/\partial\theta_k$, $u = \xi t$ for some $\xi \in (0,1)$, by Appendix II (vi, vii). Thus

$$|E(\Lambda_t)| \leqq \tfrac{1}{2} n^{-\frac{1}{2}} |t\nabla C_u|\, |C_s - C_u|\, \|C_u^{-1}\|^2$$

by Appendix II (ii, iii), and this is bounded by Corollary 3.3 (ii, iii), 3.4.

(ii) Considering the variance of Λ_t

$$[2\operatorname{Var}\Lambda_t]^{\frac{1}{2}} = |C_s^{\frac{1}{2}}(C_t^{-1} - C_0^{-1}) C_s^{\frac{1}{2}}|$$

$$\leqq \|C_s\|\, \|C_t^{-1}\|\, \|C_0^{-1}\|\, |C_0 - C_t|$$

by Appendix II (iii), which is bounded by Corollary 3.3 (i, ii) and 3.4. This completes the proof.

Theorem 4.2.

$$\Lambda_t - \tfrac{1}{2} n^{-\frac{1}{2}}\{X_n^* C_0^{-1}(t\nabla C_0) C_0^{-1} X_n - \operatorname{tr}[C_0^{-1}(t\nabla C_0)]\}$$

$$+ \tfrac{1}{4} n^{-1} |C_0^{-\frac{1}{2}}(t\nabla C_0) C_0^{-\frac{1}{2}}|^2 \to 0 \text{ in } P_0 \text{ probability.}$$

Proof. The absolute value of the expectation is equal to

$$\tfrac{1}{2}|\log\det(C_0 C_t^{-1}) + \operatorname{tr}[I - C_0 C_t^{-1} + \tfrac{1}{2} n^{-1} C_0^{-1}(t\nabla C_0)\, C_0^{-1}(t\nabla C_0)]|.$$

It follows from Appendix II (v) that for n large enough this is less than or equal to

$$\tfrac{1}{6}\|A\|\,|A|^2/(1-\|A\|)^3$$
$$+\ \tfrac{1}{4}\left|\operatorname{tr}\left[n^{-1}C_0^{-1}(t\nabla C_0)C_0^{-1}(t\nabla C_0)-C_t^{-1}(C_0-C_t)C_t^{-1}(C_0-C_t)\right]\right|$$

where $A = C_0C_t^{-1} - I = (C_0 - C_t)C_t^{-1}$. In view of Corollary 3.3 (i, ii), 3.4 the first term tends to zero. The second term is equal to

$$\tfrac{1}{4}n^{-1}\left|\operatorname{tr}\left[C_0^{-1}(t\nabla C_0)C_0^{-1}(t\nabla C_0)-C_t^{-1}(t\nabla C_u)C_t^{-1}(t\nabla C_v)\right]\right|$$

where u, $v = \xi t$, ηt for some $\xi, \eta \in (0, 1)$ by Appendix II (viii) and this tends to zero as n tends to infinity since $\|C_0 - C_t\|$, $n^{-\frac{1}{2}}|t\nabla C_0 - t\nabla C_u| \to 0$ and $n^{-\frac{1}{2}}|t\nabla C_0|$, $n^{-\frac{1}{2}}|t\nabla C_u|$, $\|C_0^{-1}\|$, $\|C_t^{-1}\|$ are bounded (Corollary 3.3, i, ii, iii)—break the expression into the traces of 4 separate differences and apply Result (ii) in Appendix II.

The variance is equal to

$$\tfrac{1}{2}\left|C_0^{\frac{1}{2}}\left[C_0^{-1}-C_t^{-1}-n^{-\frac{1}{2}}C_0^{-1}(t\nabla C_0)C_0^{-1}\right]C_0^{\frac{1}{2}}\right|^2$$

which can be shown to tend to zero by arguments similar to those already considered. This completes the proof.

The next theorem may be proved in a somewhat similar manner and its proof will be omitted.

Theorem 4.3. If $t_n \to t$ then

$$\Lambda_n(\theta + n^{-\frac{1}{2}}t_n, \theta) - \Lambda_n(\theta + n^{-\frac{1}{2}}t, \theta) \to 0 \text{ in } P_n(\theta) \text{ probability.}$$

Theorem 4.4.

$$\tfrac{1}{2}n^{-1}\left|C_0^{-\frac{1}{2}}(t\nabla C_0)C_0^{-\frac{1}{2}}\right|^2 \to \tfrac{1}{2}\int_0^1 \left|f^{-\frac{1}{2}}(\lambda,\theta)\,[t\nabla f(\lambda,\theta)]f^{-\frac{1}{2}}(\lambda,\theta)\right|^2 d\lambda.$$

Proof. Let Q, U be square matrices composed of n^2 $r \times r$ submatrices the (j,k)th being q_{j-k}, u_{j-k} respectively $(j, k = 0, 1, \cdots, n-1)$ where

$$q_j = \int_0^1 f^{-\frac{1}{2}}(\lambda,\theta)[t\nabla f(\lambda,\theta)]f^{-\frac{1}{2}}(\lambda,\theta)e^{-2\pi ij\lambda}d\lambda,$$
$$u_j = \int_0^1 f^{-\frac{1}{2}}(\lambda,\theta)e^{-2\pi ij\lambda}d\lambda.$$

It follows from Lemma 3.1 (iii) that

$$\tfrac{1}{2}n^{-1}|Q|^2 \to \tfrac{1}{2}\int_0^1 \left|f^{-\frac{1}{2}}(\lambda,\theta)[t\nabla f(\lambda,\theta)]f^{-\frac{1}{2}}(\lambda,\theta)\right|^2 d\lambda$$

as n tends to infinity. Consider

$$
\begin{aligned}
n^{-1}\big| |Q|^2 - |C_0^{-\frac{1}{2}}(t\nabla C_0)C_0^{-\frac{1}{2}}|^2 \big|
&\leqq n^{-1}\big| |Q|^2 - |U(t\nabla C_0)U|^2 \big| \\
&\quad + n^{-1}\big| |U(t\nabla C_0)U|^2 - |C_0^{-\frac{1}{2}}(t\nabla C_0)C_0^{-\frac{1}{2}}|^2 \big| \\
&\leqq n^{-\frac{1}{2}}\{|Q| + |U(t\nabla C_0)U|\}\cdot n^{-\frac{1}{2}}|Q - U(t\nabla C_0)U| \\
&\quad + n^{-1}\big|\operatorname{tr}\{(U^2 - C_0^{-1})t\nabla C_0(U^2 + C_0^{-1})t\nabla C_0\}\big| \\
&\leqq n^{-\frac{1}{2}}\{|Q| + |U(t\nabla C_0)U|\}\cdot n^{-\frac{1}{2}}|Q - U(t\nabla C_0)U| \\
&\quad + n^{-\frac{1}{2}}|(C_0U^2 - I)t\nabla C_0|\cdot n^{-\frac{1}{2}}|t\nabla C_0|\,\|C_0^{-1}\|\,\|U^2 + C_0^{-1}\|
\end{aligned}
$$

by Appendix II (ii) and this may be shown to tend to zero by repeated application of Lemma 3.1 (v). This completes the proof.

The next theorem follows from the continuity of the $c_j(\theta)$.

Theorem 4.5. The probability $P(X_n \in S)$ under $P_n(\theta)$ where S is a measurable set in nr-dimensional Euclidean space is a measurable function of θ.

It follows that Le Cam's asymptotic differentiability conditions P0–P4 are satisfied when A0, A1.1–A1.4 hold and $\delta_n = n^{-\frac{1}{2}}$ (P1 follows from Theorem 4.1; P2 from 4.2, 4.4; P3 from 4.3; P4 from 4.5). We can take

$$
\{\Delta_n(\theta)\}_j = \tfrac{1}{2}n^{-\frac{1}{2}}\left\{X_n^* C_n^{-1}(\theta)\frac{\partial}{\partial\theta_j}C_n(\theta)C_n^{-1}(\theta)X_n - \operatorname{tr}\left[C_n^{-1}(\theta)\frac{\partial}{\partial\theta_j}C_n(\theta)\right]\right\}, \tag{4.5}
$$

$$
t^*\Gamma(\theta)t = \tfrac{1}{2}\int_0^1 |f^{-\frac{1}{2}}(\lambda,\theta)\,t\nabla f(\lambda,\theta)\,f^{-\frac{1}{2}}(\lambda,\theta)|^2 d\lambda, \tag{4.6}
$$

that is,

$$
\{\Gamma(\theta)\}_{j,k} = \tfrac{1}{2}\int_0^1 \operatorname{tr}\left\{f^{-1}(\lambda,\theta)\frac{\partial}{\partial\theta_j}f(\lambda,\theta)f^{-1}(\lambda,\theta)\frac{\partial}{\partial\theta_k}f(\lambda,\theta)\right\}d\lambda. \tag{4.7}
$$

The requirement P5 follows from the next theorem.

Theorem 4.6. There exist random variables $\{T_n(\theta)\}$ such that

$$\mathscr{L}\{T_n(\theta)\} = \mathscr{L}_\theta\{\Delta_n(\theta)\}$$

and

$$\sup_n P\{|T_n(\theta+\varepsilon) - T_n(\theta)| > \delta\} \to 0 \text{ as } \varepsilon \to 0.$$

Proof. Suppose that for each θ

$$X_n(\theta) = C_n^{\frac{1}{2}}(\theta)\, U_n$$

where U_n is a vector of nr independent standard normal random variables. Consider

$$T_{n,j}(\theta) = \tfrac{1}{2} n^{-\frac{1}{2}} \left\{ X_n^*(\theta)\, C_n^{-1}(\theta) \left[\frac{\partial}{\partial\theta_j} C_n(\theta)\right] C_n^{-1}(\theta)\, X_n(\theta) - \operatorname{tr}\left[C_n^{-1}(\theta) \frac{\partial}{\partial\theta_j} C_n(\theta)\right]\right\}.$$

Clearly

$$\mathscr{L}\{T_n(\theta)\} = \mathscr{L}_\theta\{\Delta_n(\theta)\}.$$

Now consider $T_{n,j}(\theta+\varepsilon) - T_{n,j}(\theta)$. The expectation is zero and the standard deviation equals

$$(2n)^{-\frac{1}{2}} \left\| C_n^{-\frac{1}{2}}(\theta+\varepsilon) \frac{\partial}{\partial\theta_j} C_n(\theta+\varepsilon)\, C_n^{-\frac{1}{2}}(\theta+\varepsilon) - C_n^{-\frac{1}{2}}(\theta) \frac{\partial}{\partial\theta_j} C_n(\theta)\, C_n^{-\frac{1}{2}}(\theta) \right\|.$$

This tends to zero uniformly in n as ε tends to zero, since

$$\sup_n \| C_n^{-\frac{1}{2}}(\theta+\varepsilon) \|, \quad \sup_n n^{-\frac{1}{2}} \left\| \frac{\partial}{\partial\theta_j} C_n(\theta+\varepsilon) \right\|$$

are bounded,

$$\sup_n \| C_n^{-\frac{1}{2}}(\theta+\varepsilon) - C_n^{-\frac{1}{2}}(\theta) \| \to 0 \text{ (by Appendix II, xii)}$$

and

$$\sup_n n^{-\frac{1}{2}} \left\| \frac{\partial}{\partial\theta_j} C_n(\theta+\varepsilon) - \frac{\partial}{\partial\theta_j} C_n(\theta) \right\| \to 0$$

as ε tends to zero. Hence for $\delta > 0$

$$\sup_n P\{ | T_{n,j}(\theta+\varepsilon) - T_{n,j}(\theta) | > \delta \} \to 0$$

as ε tends to zero. This completes the proof.

It follows (Q2 in Appendix I) that for almost all $\theta \in \Theta$, $\Delta_n(\theta + n^{-\frac{1}{2}}t) - \Delta_n(\theta) \to -\Gamma(\theta)t$ in $P_n(\theta)$ probability. This may be extended to all θ by an argument similar to the preceding one. Part of the theory of asymptotic differentiability requires the non-singularity of the matrix $\Gamma(\theta)$. This may be proved if A2 is satisfied.

Theorem 4.7. If $t\nabla f(\lambda,\theta) \neq 0$ for some non-null set of λ for each non-zero value of t then $\Gamma(\theta)$ is non-singular.

The version of $\Delta_n(\theta)$ defined in Equation 4.5 is hardly a practical form as the numerical manipulation of the $nr \times nr$ matrices would be rather difficult. However as we saw in Corollary 3.3 (iv) $C_n(\theta)$ is approximately diagonalized by the unitary transformation Ω_n. This leads to the following theorem.

Theorem 4.8. Suppose Assumptions A3.1, A3.2 are satisfied. Then

$$n^{-\frac{1}{2}}\{X_n^* C_0^{-1}(t\nabla C_0)C_0^{-1}X_n - \operatorname{tr}[C_0^{-1}(t\nabla C_0)]\} \\ - n^{-\frac{1}{2}}\{Z_n^* F_0^{-1}(t\nabla F_0)F_0^{-1}Z_n - \operatorname{tr}[F_0^{-1}(t\nabla F_0)]\}$$

tends to zero in P_0 probability as n tends to infinity.

Proof. (i) The expectation is equal to

$$n^{-\frac{1}{2}}\operatorname{tr}\{F_0^{-1}(t\nabla F_0) - \Omega_n C_0 \Omega_n^* F_0^{-1}(t\nabla F_0)F_0^{-1}\} \\ = n^{-\frac{1}{2}}\operatorname{tr}\{G_0 F_0^{-1}(t\nabla F_0)F_0^{-1}\}$$

where $G_0 = \Omega_n^* C_0 \Omega_n - F_0$ as in Equation (2.8). The absolute value is equal to

$$n^{-\frac{1}{2}}\left|\sum_0^{n-1} \operatorname{tr}\{g_{jj}f_0^{-1}(j/n)(t\nabla f_0(j/n))f_0^{-1}(j/n)\}\right| \\ \leqq n^{-\frac{1}{2}}\sum_0^{n-1} |g_{jj}|\,|f_0^{-1}(j/n)(t\nabla f_0(j/n))f_0^{-1}(j/n)|.$$

(Appendix II, ii). Now $g_{jj} = -\sum_{p=-\infty}^{\infty}\min(1,|p|/n)\,c_p\exp\{2\pi i\,jp/n\}$ and so

$$n^{-\frac{1}{2}}\sum_0^{n-1}|g_{jj}| \leqq n^{\frac{1}{2}}\sum_{p=-\infty}^{\infty}\min(1,|p|/n)\,|c_p| \to 0$$

(by A3.1) and so the expectation tends to zero.

(ii) Considering the standard deviation

$$(2/n)^{\frac{1}{2}}|\,C_0^{\frac{1}{2}}[C_0^{-1}(t\nabla C_0)C_0^{-1} - \Omega_n^* F_0^{-1}(t\nabla F_0)F_0^{-1}\Omega_n]C_0^{\frac{1}{2}}\,| \to 0$$

since $n^{-\frac{1}{2}}|\,G_0\,|$, $n^{-\frac{1}{2}}|\,t\nabla G_0\,| \to 0$ (Lemma 3.1, iv). This completes the proof.

In view of Theorem 4.8 we may take

$$[\Delta_n(\theta)]_k = \tfrac{1}{2}n^{-\frac{1}{2}}\Big[Z_n^* F_n^{-1}(\theta)\frac{\partial}{\partial\theta_k}F_n(\theta)F_n^{-1}(\theta)Z_n \\ - \operatorname{tr}\Big\{F_n^{-1}(\theta)\frac{\partial}{\partial\theta_k}F_n(\theta)\Big\}\Big] \tag{4.8}$$

if Conditions A3.1, A3.2 hold. Alternatively if only A3.2 is satisfied or we wish to avoid some of the error introduced by this use of the periodogram we might choose

$$[\Delta_n(\theta)]_k = \tfrac{1}{2} n^{-\frac{1}{2}}\Big[Z_n^* F_n^{-1}(\theta) \frac{\partial}{\partial\theta_k} F_n(\theta) F_n^{-1}(\theta) Z_n - \operatorname{tr}\Big\{\Omega_n C_n(\theta)\Omega_n^* F_n^{-1}(\theta)\frac{\partial}{\partial\theta_k} F_n(\theta) F_n^{-1}(\theta)\Big\}\Big]. \tag{4.9}$$

Only the diagonal $r \times r$ submatrices of $\Omega_n C_n(\theta)\Omega_n^*$ enter into this formula and these may be readily calculated from the familiar expression

$$\sum_{1-n}^{n-1} (1 - |p|/n)\, c_p \exp\{2\pi i j p/n\} \qquad (j = 0, 1, \cdots, n-1).$$

In the formulae used by Le Cam the estimator $\hat{\theta}_n$ is replaced by one which takes values on a lattice with mesh size of order $n^{-\frac{1}{2}}$. In practice one would prefer not to have to make this replacement. If Conditions A3.2, A3.3 are satisfied then it follows from the next theorem that the discretization of $\hat{\theta}_n$ is in fact unnecessary.

Theorem 4.9. If A3.2, A3.3 hold and Δ_n is given by Equation (4.5)

$$\sup_{\substack{|\xi| \leq \delta \\ |\xi-\eta| < \varepsilon}} |\Delta_n(\theta + n^{-\frac{1}{2}}\xi) - \Delta_n(\theta + n^{-\frac{1}{2}}\eta)| \to 0$$

in $P_n(\theta)$ probability uniformly in n as $\varepsilon \to 0$ for each $\delta > 0$.

Proof.

$$\begin{aligned} &|[\Delta_n(\theta + n^{-\frac{1}{2}}\xi) - \Delta_n(\theta + n^{-\frac{1}{2}}\eta)]_k| \\ &\quad \leq \tfrac{1}{2} n^{-\frac{1}{2}}[|X_n|^2 \, \|B_{n,\xi} - B_{n,\eta}\| + |\operatorname{tr}\{D_{n,\xi} - D_{n,\eta}\}|] \end{aligned}$$

where

$$D_{n,\xi} = C_n^{-1}(\theta + n^{-\frac{1}{2}}\xi)\frac{\partial}{\partial\theta_k} C_n(\theta + n^{-\frac{1}{2}}\xi),$$

$$B_{n,\xi} = D_{n,\xi} C_n^{-1}(\theta + n^{-\frac{1}{2}}\xi).$$

Now $n^{-1}|X_n|^2$ is bounded in probability (uniformly in n).

$$\begin{aligned} &\|C_n(\theta + n^{-\frac{1}{2}}\xi) - C_n(\theta + n^{-\frac{1}{2}}\eta)\| \\ &\quad \leq \Big\|\frac{\partial}{\partial a} C_n\{\theta + n^{-\frac{1}{2}}[a\xi + (1-a)\eta]\}\Big\| \\ &\qquad\qquad \text{(for some } a \in (0,1) \text{ by Appendix II (x)),} \\ &\quad = O(n^{-\frac{1}{2}}|\xi - \eta|) \text{ by Lemma 3.1 (i), A3.2, A3.3.} \end{aligned}$$

By A3.3

$$\left\| \frac{\partial}{\partial\theta_k} C_n(\theta + n^{-\frac{1}{2}}\xi) - \frac{\partial}{\partial\theta_k} C_n(\theta + n^{-\frac{1}{2}}\eta) \right\| = O(n^{-\frac{1}{2}} | \xi - \eta |)$$

and so the result follows.

Corollary 4.10. The same result holds for the versions (4.8) and (4.9) of Δ_n.

Up to now we have assumed that the x_k have zero expectation. In practice this assumption would generally not be satisfied and one would subtract the sample mean from the x_k before calculating the value of the particular version of Δ_n being used. In versions (4.8) and (4.9) of Δ_n this is equivalent to setting $z_{n,0}$ to zero. In order to show subtraction of the mean permissible it is sufficient to show, still assuming (2.1), that this subtraction makes asymptotically negligible difference to the values of the three versions of the Δ_n. (We are assuming that the expected value of the x_k contains no information about θ.) Let $\bar{X}_n$ be the vector obtained by replacing each component in X_n by its corresponding sample mean.

Theorem 4.11.

(i)
$$n^{-\frac{1}{2}}(X_n - \bar{X}_n)^* C_n^{-1}(\theta) \frac{\partial}{\partial\theta_k} C_n(\theta) C_n^{-1}(\theta)(X_n - \bar{X}_n)$$
$$- n^{-\frac{1}{2}} X_n^* C_n^{-1}(\theta) \frac{\partial}{\partial\theta_k} C_n(\theta) C_n^{-1}(\theta) X_n$$

tends to zero in $P_n(\theta)$ probability as n tends to infinity.

(ii) If A3.1 is satisfied

$$n^{-\frac{1}{2}} z_{n,0}^* f^{-1}(0) \frac{\partial}{\partial\theta_k} f(0) f^{-1}(0) z_{n,0}$$

tends to zero in P_0 probability as n tends to infinity.

The proof of (i) depends on Corollary 3.3 (v) and will be omitted, and the proof of (ii) is clear.

This ends the formal theorems in this section. Conditions for the existence of a $\sqrt{n}$-consistent estimator of θ will not be given in this paper. However it should be noted that global conditions are required, for example, the assumptions introduced so far do not imply that different values of θ correspond to different probabilities. To prove consistency of the maximum likelihood estimator presumably some further condition will be required to prevent spurious maxima in the likelihood function (cf. Whittle (1962), Theorem 3). However, if the estimator, $\hat{\theta}_n$, is found by maximizing the log-likelihood

$$-\tfrac{1}{2}\{nr\log(2\pi) + \log\det C_n(\theta) + X_n^* C_n^{-1}(\theta)X_n\} \tag{4.10}$$

and version (4.5) of Δ_n is used then $\Delta_n(\hat{\theta}_n)$ vanishes and hence $\hat{\theta}_n$ coincides with the statistic, T_n, defined in Appendix I. Thus if A3.2, A3.3 hold (and hence Theorem 4.9 applies) then if the maximum likelihood estimator is $\sqrt{n}$-consistent it possesses the optimality properties attributed to T_n in Appendix I. Similarly if the estimator, $\hat{\theta}_n$, is found by maximizing the approximate log-likelihood

$$-\tfrac{1}{2}\{nr\log(2\pi) + \log\det F_n(\theta) + Z_n^* F_n^{-1}(\theta) Z_n\} \tag{4.11}$$

and version (4.8) of Δ_n is used then $\Delta_n(\hat{\theta}_n)$ vanishes. Thus if A3.1, A3.2, A3.3 hold and $\hat{\theta}_n - \theta = O(n^{-\frac{1}{2}})$ under $P_n(\theta)$ then this version of $\hat{\theta}_n$ also possesses the optimality properties of T_n. We note in passing that the asymptotic variance/covariance matrix, $\{n\Gamma(\theta)\}^{-1}$, of T_n agrees with that given by Whittle (1953) in Theorem 9.

In most of the time-series papers referenced in the introduction it is shown that the asymptotic distributions of the estimators are independent of the normality assumption. We give a rough argument to suggest that, under suitable conditions, a similar result is true for $\Delta_n(\theta)$ and hence presumably for T_n. For suppose x_k can be represented

$$x_k = \xi_k + \sum_{j=1}^{\infty} a_j(\theta)\xi_{k-j}, \qquad k = 0, 1, \cdots, n-1, \tag{4.12}$$

where $\{\xi_k\}$ is a doubly infinite sequence of independent identically distributed r-dimensional random variables with zero expectation and with known variance/covariance matrix

$$E(\xi_k \xi_k^*) = \sigma \tag{4.13}$$

and $a_j(\theta)$, $j = 1, 2, \cdots$, is a sequence of $r \times r$ matrices whose elements are functions of the unknown parameters. We will now give a heuristic argument to suggest that when (4.12) holds under certain (unspecified but reasonably general) conditions

$$t^*\Delta_n(\theta) = \tfrac{1}{2}n^{-\frac{1}{2}}\{X_n^* C_n^{-1}(\theta)\, t\nabla C_n(\theta)\, C_n^{-1}(\theta)\, X_n - \operatorname{tr}[C_n^{-1}(\theta)\, t\nabla C_n(\theta)]\}$$

has, under $P_n(\theta)$, a limiting normal distribution with zero mean and variance $t^*\Gamma(\theta)t$. Suppose $\Xi_n = \{\xi_0, \cdots, \xi_{n-1}\}^*$ and $\Sigma_n = I_n \otimes \sigma = E(\Xi_n \Xi_n^*)$. Then X_n is approximately equal to $A_n \Xi_n$ where A_n has (j,k)th submatrix I if $j = k$, a_{k-j} if $k > j$ and zero otherwise, j, $k = 0, 1, \cdots, n-1$. Hence $t^*\Delta_n(\theta)$ is approximately equal to

$$\tfrac{1}{2}n^{-\frac{1}{2}}\Xi_n^*(\Sigma_n^{-1}A_n^{-1}t\nabla A_n + t\nabla A_n^*(A_n^*)^{-1}\Sigma_n^{-1})\Xi_n. \tag{4.14}$$

The matrix $\Sigma_n^{-1}A_n^{-1}t\nabla A_n + t\nabla A_n^*(A_n^*)^{-1}\Sigma_n^{-1}$ has its diagonal $r \times r$ submatrices zero and so asymptotic normality of $t^*\Delta_n(\theta)$ follows from Theorem 3.5 since in view of Lemma 3.1 (v) this matrix is approximately equal to one whose $r \times r$ submatrices depend only on their distance from the diagonal. This completes the heuristic argument. Also deviations of order $n^{-\frac{1}{2}}$ in σ make no difference to the asymptotic distribution and so presumably σ may be replaced by an estimate. It is not clear how general the model described by (4.12) in fact is or how one could test that a time series can be described by it. However it does suggest that one should not carry out 'unnatural' transformations to ensure normality of the marginal distributions.

Finally in this section we make a few comments about the rate of convergence to the asymptotic forms. The theory of asymptotic differentiability does not give any indication concerning the rate of convergence; however one might suspect that the rate of convergence of Postulate P2 in Appendix I would be of considerable importance on determining this. For version (4.5) of $\Delta_n(\theta)$ this, in turn, is determined by the convergence of the expressions in Theorems 4.2, 4.4., It is not hard to show that this is at rate $n^{-\frac{1}{2}}$ if A3.2, A3.3 are satisfied together with

(4.15) $$\sum_1^\infty j^{\frac{1}{2}}\left\|\frac{\partial}{\partial\theta_k} c_j(\theta)\right\|^2 < \infty.$$

For version (4.9) one would require also

(4.16) $$\sum_1^\infty j^{\frac{1}{2}}\left\|\frac{\partial}{\partial\theta_k} c_j(\theta)\right\| < \infty$$

and for version (4.8) also

(4.17) $$\sum_1^\infty j\|c_j(\theta)\| < \infty.$$

If these additional conditions for the version of $\Delta_n(\theta)$ being used are not satisfied one might expect rather slow convergence to the asymptotic results.

5. Summary

We have seen that Le Cam's asymptotic differentiability postulates, P0–P6 as listed in Appendix I, are satisfied with Δ_n defined as in (4.5) (with the sample mean subtracted from the observations if $E(x_k) \neq 0$), $\delta_n = n^{-\frac{1}{2}}$, and Γ as defined in (4.7) if A0, A1.1–A1.4, A2 in Section 2 are satisfied. If in addition A3.2 is satisfied Δ_n may be as defined in (4.9) and if further A3.1 is satisfied Δ_n may be as in (4.8). These latter two forms are more suitable for numerical work. We have not considered conditions for the existence of a preliminary estimator $\hat{\theta}_n$. However we have noted that the requirement that it take on values only on a lattice with spacing of order $n^{-\frac{1}{2}}$ can be relaxed if A3.3 is satisfied. It follows from the results Q1–Q6 in Appendix I that

$$T_n = \hat{\theta}_n + \delta_n \Gamma^{-1}(\hat{\theta}_n)\Delta_n(\hat{\theta}_n)$$

in a certain sense contains most of the relevant information in the sample for n large and further has a distribution close to normal with mean θ and variance $\delta_n^2 \Gamma^{-1}(\theta)$. In fact, procedures close (in an appropriate sense) to optimal can be obtained, in general, by applying procedures optimal for normal variables to T_n (or in certain instances to $\Delta_n(\theta)$). For instance T_n has minimum asymptotic variance amongst asymptotically (locally uniformly) unbiassed estimators of θ.

Example. The following example is chosen to illustrate the theory for a one-dimensional time-series in which fairly long term dependence is present. Suppose we sample at unit intervals from a continuous Gaussian process with spectral density $\theta_1 \theta_2 \pi \exp(-|2\pi\lambda\theta_2|)$ where $\theta_1,\ \theta_2 \in (0,\infty)$ are unknown parameters. Then

$$c_k(\theta) = \theta_1 \theta_2^2/(k^2 + \theta_2^2) \tag{5.1}$$

and

$$f(\lambda, \theta) = \theta_1 \theta_2 \pi \cosh[\theta_2 \pi(1 - 2\lambda)]/\sinh(\theta_2 \pi). \tag{5.2}$$

Thus

$$\begin{aligned} \frac{\partial}{\partial\theta_1} c_k(\theta) &= \theta_2^2/(k^2 + \theta_2^2), \\ \frac{\partial}{\partial\theta_2} c_k(\theta) &= 2k^2\theta_1\theta_2/(k^2 + \theta_2^2)^2, \end{aligned} \tag{5.3}$$

and so the assumptions listed in Section 2 are satisfied. For a one-dimensional series version (4.8) of $[\Delta_n(\theta)]_j$ simplifies to

$$\tfrac{1}{2} n^{-\frac{1}{2}} \sum_{p=0}^{n-1} \frac{\partial}{\partial\theta_j} \log f(p/n,\theta)[|z_{n,p}|^2 - f(p/n,\theta)]/f(p/n,\theta) \tag{5.4}$$

and $[\Gamma(\theta)]_{j,k}$ to

$$\tfrac{1}{2} \int_0^1 \frac{\partial}{\partial\theta_j} \log f(\lambda,\theta) \frac{\partial}{\partial\theta_k} \log f(\lambda,\theta)\, d\lambda. \tag{5.5}$$

To simplify calculation (especially for a general computer program) one is inclined to approximate $\Gamma(\theta)$ by $\Gamma_n(\theta)$ where

$$[\Gamma_n(\theta)]_{j,k} = \frac{1}{2n} \sum_{p=0}^{n-1} \frac{\partial}{\partial\theta_j} \log f(p/n,\theta) \frac{\partial}{\partial\theta_k} \log f(p/n,\theta) \tag{5.6}$$

since the components of this sum must be calculated in order to evaluate

$\Delta_n(\theta)$. Of course when computing (5.4), (5.6) one need sum only $\frac{1}{2}n$ terms since

$$z_{n,p} = z^*_{n,n-p} \text{ and } f(p/n,\theta) = f(1-p/n,\theta).$$

The preliminary estimate, $\hat{\theta}_n$, of θ is obtained by the method of moments:

$$[\hat{\theta}_n]_1 = \hat{c}_{n,0},\ [\hat{\theta}_n]_2 = [\hat{c}_{n,1}/(\hat{c}_{n,0} - \hat{c}_{n,1})]^{\frac{1}{2}} \tag{5.7}$$

where $\hat{c}_{n,0}$, $\hat{c}_{n,1}$ are the sample variance and first autocovariance. (The method fails if $\hat{c}_{n,1} \leqq 0$ and so this preliminary estimator is unsuitable if θ_2 is small.) Thus

$$T_n = \hat{\theta}_n + n^{-\frac{1}{2}}\,\Gamma_n^{-1}(\hat{\theta}_n)\Delta_n(\hat{\theta}_n) \tag{5.8}$$

may be computed. This provides the asymptotically optimal estimate of θ.

Suppose now that the model (5.1) is generalized to

$$c_k(\theta) = \theta_1/[1 + (k/\theta_2)^{2+\theta_3}]$$

and we wish to test the hypothesis $\theta_3 = 0$ against the alternative $\theta_3 \neq 0$. If we are satisfied with asymptotic optimality (amongst locally uniformly asymptotically similar and unbiassed tests) against alternatives with $\theta_3 = O(n^{-\frac{1}{2}})$ then the calculations may be simplified by choosing a preliminary estimator $\hat{\theta}_n$ that is $\sqrt{n}$-consistent under the hypothesis only — for example, with its first two components as in (5.7) and its third component zero. This estimating of θ under the hypothesis is not unfamiliar (cf. Neyman (1959)) and may be justified in the present context with the aid of Result 2 on p. 29 of Le Cam (1969). With this preliminary estimator the first two components of Δ_n are as above and the third may be calculated without too much difficulty. Similar remarks apply to Γ_n and so T_n may be calculated. The asymptotically optimal, asymptotically level α test is then to reject the hypothesis $\theta_3 = 0$ if

$$|[T_n]_3| > z_{1-\frac{1}{2}\alpha}\cdot[\Gamma_n^{-1}(\hat{\theta}_n)]^{\frac{1}{2}}_{33}$$

where z_α is such that the probability that it is exceeded by a standard normal random variable is α.

The model (5.1) has been simulated with $n = 256$, $\theta_1 = 1$, $\theta_2 = 2$ and one hundred realizations of $[T_n]_2$ are presented as a normal probability plot in Figure 1, together with a line corresponding to a

$$\mathcal{N}(\theta_2, n^{-1}[\Gamma_n^{-1}(\theta)]_{22})$$

distribution. The reader is refered to Daniel and Wood (1971) for numerous simulations of probability plots in order to obtain an idea of the goodness of fit that can be expected. However it appears that simulations of the present type are rather sensitive to the limitations of the random number generator being

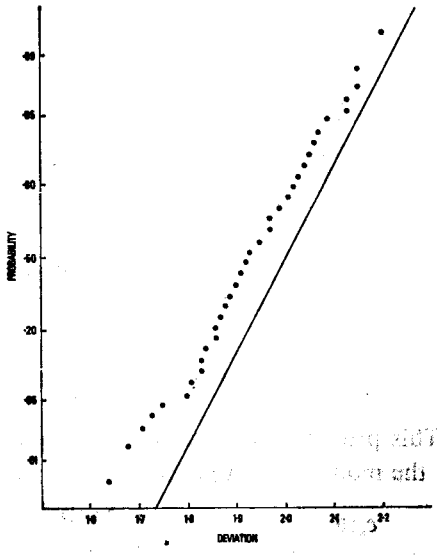

Figure 1

used and so should be treated with a certain amount of caution. Nevertheless this plot suggests that a significant amount of bias is present but that the variance of $[T_n]_2$ is fairly close to that predicted by the asymptotic formula. There is also a suggestion of an extension of the lower tail. The bias is roughly halved if version (4.9) rather than (4.8) of Δ_n is used but is still larger than one would like. With $n = 128$ the fit was much poorer, the lower tail being rather extended with the occasional very low outlier. Possibly the overall fit could be improved by choosing a different parametrization in an effort to make the problem more "linear". However in view of the long term dependence present in this time-series and the need for a large number of observations for accurate estimation of parameters of time-series in general it is perhaps not surprising that a large value of n is required for the asymptotic distributions to be a reasonable approximation to the actual ones.

Acknowledgement

I wish to express my appreciation for the referee's very careful reading of this paper.

Appendix I. Asymptotically differentiable families

In this appendix we will state some of the main results of the theory of asymptotically differentiable families of experiments developed by Le Cam (e.g. (1969), pp. 57–87). We are given a sequence of "experiments", $\mathscr{E}_n = \{\mathscr{X}_n, \mathscr{A}_n, P_{n,\theta}\}$, $\theta \in \Theta$, $n = 1, 2, \cdots$, where $\mathscr{X}_n$ is the set of possible outcomes,

X_n, of the nth experiment, $\mathscr{A}_n$ a sigma-algebra defined on $\mathscr{X}_n$, and $P_{n,\theta}$ a family of distributions defined on $\mathscr{A}_n$ and indexed by $\theta \in \Theta$ where Θ is a set of "theories" about the experiment. We observe $X_n \in \mathscr{X}_n$ drawn according to $P_{n,\theta}$ for some unknown $\theta \in \Theta$ and wish to make inferences concerning θ. The theory of asymptotically differentiable families shows that, under apparently innocuous assumptions, for n large enough, most of the information in X_n concerning θ is summarized in a statistic T_n and further this statistic is approximately normally distributed.

First we require a definition — equivalent to that given by Le Cam ((1969), p. 28).

Definition. Suppose we are given a sequence of experiments with just two probability measures for each n denoted by P_n and Q_n. Suppose ϕ_n is a test of the hypothesis that an observation X_n comes from the probability measure P_n against the alternative that it comes from Q_n. Let α_n be the significance level and β_n the power. We will say that the sequences of probability measures $\{P_n\}$ and $\{Q_n\}$ are *contiguous* if for all possible sequences of tests ϕ_n

(i) $$\limsup_{n\to\infty} \beta_n = 1 \Rightarrow \limsup_{n\to\infty} \alpha_n = 1,$$

(ii) $$\liminf_{n\to\infty} \alpha_n = 0 \Rightarrow \liminf_{n\to\infty} \beta_n = 0.$$

In effect $\{P_n\}$ and $\{Q_n\}$ are contiguous if, for large n, statistical techniques are required to distinguish P_n and Q_n. For more properties of contiguous sequences of probability measures see Le Cam (1960), (1969) and Roussas ((1972), ch. 1). In particular note that a necessary and sufficient condition for $\{P_n\}$ and $\{Q_n\}$ to be contiguous is that the log-likelihood ratio, $\log dQ_n/dP_n$, is bounded in both P_n and Q_n probability as n tends to infinity.

Returning to the original problem let

$$\Lambda_n(\theta, \psi) = \log dP_{n,\theta}/dP_{n,\psi}$$

with $\Lambda_n = -\infty$ when $P_{n,\theta}$ is singular with respect to $P_{n,\psi}$ and ∞ when $P_{n,\psi}$ is singular with respect to $P_{n,\theta}$. Suppose δ_n, $n = 1, 2, \cdots$, is a sequence of positive numbers tending to zero (e.g., $\delta_n = n^{-\frac{1}{2}}$).

Definition. A family of experiments $\{\mathscr{E}_n\}$, $n = 1, 2, \cdots$, is asymptotically differentiable if the following conditions, P0–P4, are satisfied.

P0. Θ is an open set in $\mathscr{R}^s$.

P1. The sequences of probability measures

$$\{P_{n,\theta+\delta_n t}\} \text{ and } \{P_{n,\theta}\}, \qquad n = 1, 2, \cdots,$$

are contiguous for each $t \in \mathscr{R}^s$ and $\theta \in \Theta$.

P2. For each $\theta \in \Theta$ there exists a sequence of s-dimensional random variables $\Delta_n(\theta)$ and an $s \times s$ matrix $\Gamma(\theta)$ such that[1]

$$\Lambda_n(\theta + \delta_n t, \theta) - t^* \Delta_n(\theta) + \tfrac{1}{2} t^* \Gamma(\theta) t \to 0$$

in $P_{n,\theta}$ probability for each $t \in \mathscr{R}^s$ (regarded as an s-dimensional column vector).

P3. If $t_n \to t$ in $\mathscr{R}^s$ and $\theta \in \Theta$ then

$$\Lambda_n(\theta + \delta_n t_n, \theta) - \Lambda_n(\theta + \delta_n t, \theta) \to 0$$

in $P_{n\,\theta}$ probability.

P4. If $A \in \mathscr{A}_n$ the function $\theta \rightsquigarrow P_{n,\theta}(A)$ is a Lebesgue measurable function.

These are the main postulates, P1 and P2 being the major ones; P1 (together with P6, P7 below) ensuring that we are looking at appropriately sized perturbations in θ and P2 stating that the log likelihood can be suitably approximated by a type of Taylor expansion. Note that in many applications Δ_n will be able to be taken as δ_n times the derivative with respect to θ of the logarithm of a probability density corresponding to $P_{n\,\theta}$ and $\Gamma(\theta)$ as the limit as n tends to infinity of $-\delta_n^2$ times the expected value of the second derivative.

Three more conditions are required for some of the results.

P5. If $m(\theta, t, n)$ denotes the Prokhorov distance (e.g., Billingsley (1968), p. 238) between the distribution of $\Delta_n(\theta)$ under $P_{n,\theta}$ and $\Delta_n(\theta + \delta_n t)$ under $P_{n,\theta+\delta_n t}$ then for each θ

$$\limsup_{n\to\infty} m(\theta, t, n) \to 0 \text{ as } t \to 0.$$

P6. $\Gamma(\theta)$ is non-singular for all $\theta \in \Theta$.

P7. For each n there exists an estimator, $\hat{\theta}_n$, (i.e., a measurable function of X_n, not dependent on θ and taking values in Θ) such that for each $\varepsilon > 0$, $\theta \in \Theta$

$$\limsup_{n\to\infty} P_{n,\theta}\{|\hat{\theta}_n - \theta| > \delta_n b\} < \varepsilon \text{ for some } b.$$

It will be supposed that $\hat{\theta}_n$ takes on only values on a lattice of points with spacing δ_n.

Postulate P5 is a convenient way of ensuring that if $\Delta_n(\theta)$ has a limiting distribution under $P_{n,\theta}$ as n tends to infinity then this limiting distribution is a continuous function of θ. P7, required to locate the vicinity of θ, is different from the other assumptions in that it is global rather than local. The requirement that $\hat{\theta}_n$ takes values on a lattice is used for showing that $\theta + \delta_n t$ can be

[1] Le Cam proves this result from an apparently more general one; this is all that is required in the present context.

replaced by $\hat{\theta}_n$ in Result Q2 below. With suitable conditions on Δ_n this requirement can be relaxed.

Le Cam shows that if Conditions P0 to P7 are satisfied then the following six results hold.

Q1. For all $\theta \in \Theta$, $\mathscr{L}\{\Delta_n(\theta) \mid P_{n,\theta}\} \to \mathscr{N}\{0, \Gamma(\theta)\}$.

Q2. For almost all $\theta \in \Theta$

$$\Delta_n(\theta + \delta_n t) - \Delta_n(\theta) \to -\Gamma(\theta)t \text{ in } P_{n,\theta} \text{ probability.}$$

This result is true for all $\theta \in \Theta$ if $\Delta_n(\theta)$ is suitably chosen—we will assume that this has been done.

Q3. For each n, $\theta \in \Theta$ there exists a family of distributions (actually an exponential family), $Q_{n,\theta,t}$ on $(\mathscr{X}_n, \mathscr{A}_n)$ such that for each $b > 0$

$$\sup_{\|t\| < b} \sup_{A \in \mathscr{A}_n} |Q_{n,\theta,t}(A) - P_{n,\theta+\delta_n t}(A)| \to 0 \text{ as } n \to \infty,$$

and such that under $Q_{n\,\theta\cdot t}$ $\Delta_n(\theta)$ is sufficient for t. Thus $\Delta_n(\theta)$ is "asymptotically sufficient" for $\theta + \delta_n t$ for bounded t.

Q4. Suppose that $T_n = \hat{\theta}_n + \delta_n \Gamma^{-1}(\hat{\theta}_n)\Delta_n(\hat{\theta}_n)$. Then for each $\theta \in \Theta$ T_n is asymptotically sufficient for $\theta + \delta_n t$ for bounded t.

Q5. $\mathscr{L}\{\delta_n^{-1}(T_n - \theta) \mid P_{n,\theta}\} \to \mathscr{N}\{0, \Gamma^{-1}(\theta)\}$.

Q6. Let $G_{n,\theta,t}$ denote the $\mathscr{N}\{\theta + \delta_n t, \delta_n^2 \Gamma^{-1}(\theta)\}$ measure on $\mathscr{R}^s$ and $F_{n\,\theta}$ denote the measure corresponding to T_n under $P_{n,\theta}$. Then for each $\theta \in \Theta$ there exists a function $\phi_n : \mathscr{R}^s \to \mathscr{R}^s$ such that as $n \to \infty$

$$\delta_n^{-1} \sup_v \| \phi_n(v) - v \| \to 0$$

and for each $b > 0$

$$\sup_{\|t\| < b} \sup_{A \in \mathscr{A}_n} |G_{n,\theta,t}[\phi_n^{-1}(A)] - F_{n,\theta+\delta_n t}(A)| \to 0.$$

Thus, in general, if a procedure would be optimal when T_n had exactly a normal distribution with unknown mean θ and known variance $\delta_n^2 \Gamma^{-1}(\hat{\theta}_n)$, then it will have a similar asymptotic optimality property for the present problem—provided that it involves only functions of T_n that are discontinuous only on a set of measure zero. For example, for convenience letting $s = 1$, suppose U_n is a function of X_n such that

(i) $$\lim_{b \to \infty} \limsup_{n \to \infty} \sup_{|t| < c} |E_{n,\theta+\delta_n t} \mathscr{S}\{\delta_n^{-1}(U_n - \theta), b\} - t| = 0$$

for some $c > 0$, $\theta \in \Theta$ where $\mathscr{S}(x, b) = x$ if $|x| < b$, 0 otherwise and $E_{n,\theta}$ denotes expectation under $P_{n,\theta}$. Then

(ii)
$$\lim_{b\to\infty} \liminf_{n\to\infty} E_{n,\theta}\mathscr{S}\{\delta_n^{-2}(U_n-\theta)^2, b\}$$
$$\geqq \lim_{b\to\infty} \lim_{n\to\infty} E_{n,\theta}\mathscr{S}\{\delta_n^{-2}(T_n-\theta)^2, b\}$$
$$= \Gamma^{-1}(\theta);$$

i.e., T_n has minimum asymptotic variance amongst asymptotically (locally uniformly) unbiassed estimators of θ. This result may be derived from Q1–Q6 with the following line of reasoning. Suppose U_n satisfies (i) but not (ii). Then:

(a) In view of Q4 there is a family of probability measures $Q_{n,\theta,t}$ arbitrarily close to $P_{n,\theta+\delta_n t}$ for n large enough and under which T_n is sufficient for t. Then under $Q_{n,\theta,t}$, U_n may be replaced by $U_n(T_n)$, a function of T_n and possibly a random variable whose distribution is independent of t. Further under $Q_{n,\theta,t}$, U_n and hence $U_n(T_n)$ satisfies (i) but not (ii). And so under $P_{n,\theta+\delta_n t}$, $U_n(T_n)$ satisfies (i) but not (ii).

(b) Suppose W_n has a $\mathscr{N}\{\theta+\delta_n t, \delta_n^2\Gamma^{-1}(\theta)\}$ distribution. Then in view of Q6 there is a function ϕ_n such that $\phi_n(W_n)$ has a distribution arbitrarily close to that of T_n. Thus $U_n(\phi_n(W_n))$ satisfies (i) but not (ii).

(c) It can be shown using an argument involving the Cramér-Rao inequality that if U_n is a function of a $\mathscr{N}\{\theta+\delta_n t, \delta_n^2\Gamma^{-1}(\theta)\}$ random variable and a random variable whose distribution is independent of t then if (i) holds so does (ii).

Appendix II. Matrix notation and elementary results

Suppose A is an $m \times n$ matrix and x is an n-dimensional column vector. The Hilbert norm of A is denoted by

$$\|A\| = \sup\{|Ax| : |x| = 1\}$$

where $|x|$ denotes the usual L_2 norm of the vector x, and the Euclidean norm of A is denoted by

$$|A| = [\mathrm{tr}(AA^*)]^{\frac{1}{2}} = \left[\sum_i \sum_j |a_{i,j}|^2\right]^{\frac{1}{2}}$$

where the $a_{i,j}$ are the elements of A and A^* denotes the conjugate transpose of A. For matrices of a given size the two norms are topologically equivalent and in this paper in the fixed size situation the norm which leads to the simplest formulae is used. When one considers sequences of matrices of increasing size, however, the two norms have rather different properties and it is necessary to use the appropriate one. Scalar operations applied to matrices will be understood to apply termwise. Thus $\partial A/\partial\theta$ is the matrix of derivatives of the elements of A when these are differentiable functions of θ. By $t\nabla A$ is meant the sum

$$\sum_{1}^{s} t_j \frac{\partial}{\partial \theta_j} A$$

where t is the vector $(t_1, \cdots, t_s)^*$ and A is a function of $\theta = (\theta_1, \cdots, \theta_s)^*$. If A is positive definite hermitian $A^{\frac{1}{2}}$ denotes the (unique) positive definite square root of A.

Some of the matrix results (for $n \times n$ matrices A and B) that we require are:

(i) $$\left| \operatorname{tr}(A) \right| \leqq n^{\frac{1}{2}} | A |,$$

(ii) $$\left| \operatorname{tr}(AB) \right| \leqq | A | \; | B |,$$

(iii) $$| AB | \leqq \| A \| \; | B |,$$

(iv) $$\| A \| \leqq | A | \leqq n^{\frac{1}{2}} \| A \|,$$

(v) $$\left| \log \det(I + A) - \operatorname{tr}(A) + \tfrac{1}{2}\operatorname{tr}(A^2) \right| \leqq \tfrac{1}{3} \| A \| \; | A |^2 (1 - \| A \|)^{-3} \text{ if } \| A \| < 1.$$

If the elements of A and B are continuously differentiable functions of θ then:

(vi) $$\frac{\partial}{\partial \theta} A^{-1} = - A^{-1} \left(\frac{\partial}{\partial \theta} A \right) A^{-1},$$

(vii) $$\frac{\partial}{\partial \theta} \log \det(A) = \operatorname{tr}\left(A^{-1} \frac{\partial}{\partial \theta} A \right),$$

(viii) $$\begin{aligned} &\operatorname{tr}\{A(\theta) B(\psi)\} \\ &= \operatorname{tr}\left\{ \left[A(0) + \theta \frac{\partial}{\partial \theta_1} A(\theta_1) \right] \left[B(0) + \psi \frac{\partial}{\partial \psi_1} B(\psi_1) \right] \right\} \\ &\leqq \left| A(0) + \theta \frac{\partial}{\partial \theta_1} A(\theta_1) \right| \left| B(0) + \psi \frac{\partial}{\partial \psi_1} B(\psi_1) \right| \end{aligned}$$

where $\theta_1 = \xi\theta$, $\psi_1 = \eta\psi$ for some ξ, $\eta \in (0, 1)$.

If $\theta_1 < \theta_2$ then:

(ix) $$| A(\theta_1) - A(\theta_2) | \leqq | \theta_1 - \theta_2 | \left| \frac{\partial}{\partial \theta} A(\theta) \right|$$

for some θ with $\theta_1 \leqq \theta \leqq \theta_2$,

(x) $$\| A(\theta_1) - A(\theta_2) \| \leqq | \theta_1 - \theta_2 | \left\| \frac{\partial}{\partial \theta} A(\theta) \right\|$$

for some θ with $\theta_1 \leqq \theta \leqq \theta_2$.

(xi) If A is positive definite hermitian then

$$\| A^{\frac{1}{2}} \| = \| A \|^{\frac{1}{2}}.$$

(xii) If A and B are positive definite hermitian matrices and

$$\| A - B \| \leqq \varepsilon / \| A^{-1} \|$$

with $\varepsilon < \frac{1}{2}$ then

$$\| A^{\frac{1}{2}} - B^{\frac{1}{2}} \| \leqq \varepsilon / \| A^{-\frac{1}{2}} \|.$$

(xiii) If A is non-singular and $\| A - B \| < 1 / \| A^{-1} \|$ then

$$\| B^{-1} \| \leqq \| A^{-1} \| / [1 - \| A^{-1} \| \, \| A - B \|].$$

Proof. Results (i, ii, iii, iv, vi, vii, xi) are either well known or trivial. The others may be proved as follows.

(v) By making a Taylor expansion of $\log \det (I + \theta A)$ using Results (vi, vii) and putting $\theta = 1$ we find

$$\log \det (I + A) = \operatorname{tr}(A) - \tfrac{1}{2} \operatorname{tr}(A^2) + \tfrac{1}{3} \operatorname{tr} \{(I + \theta A)^{-3} A^3\}$$

for some $\theta \in (0, 1)$. Thus

$$|\log \det (I + A) - \operatorname{tr}(A) + \tfrac{1}{2} \operatorname{tr}(A^2)|$$

$$\leqq \tfrac{1}{3} \| A \| \, |A|^2 \, \| (I + \theta A)^{-3} \|$$

and the result follows.

(viii) First note that if $C(x)$, D are matrices, C being a differentiable function of x, then

$$\begin{aligned} \operatorname{tr} \{C(x) D\} &= \operatorname{tr} \{C(0) D\} + x \frac{\partial}{\partial x_1} \operatorname{tr} \{C(x_1) D\} \\ &= \operatorname{tr} \left\{ \left[C(0) + x \frac{\partial}{\partial x_1} C(x_1) \right] D \right\} \end{aligned}$$

for some x_1 with $0 < x_1 / x < 1$. Note that this equality is valid when D is also a function of x. Thus

$$\begin{aligned} \operatorname{tr} \{A(\theta) B(\psi)\} &= \operatorname{tr} \left\{ \left[A(0) + \theta \frac{\partial}{\partial \theta_1} A(\theta_1) \right] B(\psi) \right\} \\ &= \operatorname{tr} \left\{ \left[A(0) + \theta \frac{\partial}{\partial \theta_1} A(\theta_1) \right] \left[B(0) + \psi \frac{\partial}{\partial \psi_1} B(\psi_1) \right] \right\} \end{aligned}$$

where $\theta_1 = \xi \theta$, $\psi_1 = \eta \psi$ for some $\xi, \eta \in (0, 1)$. The inequality follows from (ii).

(ix) follows from (viii).

(x) $$y^*\{A(\theta_1) - A(\theta_2)\}x = (\theta_1 - \theta_2)y^* \frac{\partial}{\partial\theta} A(\theta)x$$

for some $\theta \in (\theta_1, \theta_2)$. For a suitable choice of x and y the left-hand side is equal to $|x|\,|y|\,\|A(\theta_1) - A(\theta_2)\|$ and the result follows.

(xii) Consider the sequence of matrices, S_k, $k = 0, 1, 2, \cdots$, with $S_0 = B^{\frac{1}{2}}$.

$$S_{k+1} = S_k + \int_0^\infty \exp(-S_k t)(A - S_k^2)\exp(-S_k t)\,dt.$$

Then it can be shown that $S_k^2 \to A$ as $k \to \infty$ and that

$$\Sigma \| S_{k+1} - S_k \| \leqq \varepsilon / \| A^{-\frac{1}{2}} \|$$

(cf. Lancaster (1969), Formula 8.5.1).

(xiii) $B^{-1} = A^{-1} + A^{-1}(A - B)B^{-1}$ and hence

$$\| B^{-1} \| \leqq \| A^{-1} \| + \| A^{-1} \| \; \| A - B \| \; \| B^{-1} \|.$$

References

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

Daniel, C. and Wood, F. S. (1971) *Fitting Equations to Data.* Wiley, New York.

Grenander, U. and Szegö G. (1958) *Toeplitz Forms and their Applications.* University of California Press, Berkeley.

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

Lancaster, P. (1969) *Theory of Matrices.* Academic Press, New York.

Le Cam, L. M. (1960) Locally asymptotically normal families of distributions. *University of California Publications in Statistics* **3**, 37–98.

Le Cam, L. M. (1969) *Théorie Asymptotique de la Décision Statistique.* Montreal University Press.

Mann, H. and Wald, A. (1943) On the statistical treatment of linear stochastic difference equations. *Econometrica* **11**, 173–220.

Neyman, J. (1959) Optimal asymptotic tests of composite statistical hypotheses. *Probability and Statistics, The Harald Cramér Volume.* (Ed. U. Grenander) Wiley, New York.

Parzen, E. (1971) Efficient estimation of stationary time series mixed schemes. *Bull. Inst. Internat. Statist.* **44**, 315–319.

Roussas G. G. (1972) *Contiguity of Probability Measures: Some Applications in Statistics.* Cambridge University Press, Cambridge.

Walker, A. M. (1964) Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Austral. Math. Soc.* **4**, 363–384.

Whittle, P. (1951) *Hypothesis Testing in Time Series Analysis.* Almquist and Wicksell, Uppsala.

Whittle, P. (1953) The analysis of multiple stationary time series. *J. R. Statist. Soc.* B **15**, 125–139.

Whittle, P. (1962) Gaussian estimation in stationary time series. *Bull. Inst. Internat. Statist.* **39**, 105–129.