The Annals of Statistics
1979, Vol. 7, No. 3, 516–530

HIGH-ORDER EFFICIENCY IN THE ESTIMATION OF LINEAR PROCESSES[1]

By Yuzo Hosoya
Tohoku University, Japan

Estimation as a reduction of data is usually accompanied by some loss of information. This paper theoretically compares asymptotically efficient estimation methods for parameters in Gaussian linear processes. By means of the concept of "asymptotic information loss" suitably defined, estimates equivalent to the order of $N^{-\frac{1}{2}}$ are differentiated. This problem was studied by C. R. Rao for multinomial distributions and by K. Takeuchi for the exponential family of distributions. They showed that for the i.i.d. case the maximum likelihood estimate is superior to other efficient estimates. This paper extends their results to the Whittle-Walker model of Gaussian linear processes, demonstrating the optimality of the maximum likelihood estimate for that model. In addition, the paper contains a lemma of independent interest. The Craig-Aitken theorem is concerned with the independence of two quadratic forms of a finite-dimensional Gaussian random vector; the theorem is extended to infinite-dimensional Gaussian random vectors.

0. Introduction. Suppose that observations are generated by the general linear process $X_t = \sum_{i=0}^{\infty} \mu_i(\theta)\varepsilon_{t-i}$, $t = 0, \pm 1, \pm 2, \cdots$, where the ε_t are independent random variables which are identically normally distributed with mean 0, and the coefficients μ_i depend on an unknown parameter θ. The maximum-likelihood estimate of θ is efficient in many cases of linear processes. However, except for Box and Jenkins (1970), who advocate the likelihood principle, the usual practice so far is not to calculate the maximum-likelihood estimate, but to search for other efficient and computationally simpler estimates. This can be seen, for instance, in Hannan's inference for rational spectra (1970), Parzen's method (1971) and Anderson's method of scoring (1975).

Though usual investigations seem to stop when an efficient estimate is discovered for a particular model, the concept of efficiency still requires further examination. An estimate θ^* is ordinarily called efficient if the asymptotic distribution of $N^{\frac{1}{2}}(\theta^* - \theta^0)$ is normal with minimal variance. Accordingly, if the maximum-likelihood estimate $\hat{\theta}$ is known to be efficient, θ^* is also efficient if $N^{\frac{1}{2}}(\theta^* - \hat{\theta})$ goes to 0 in probability. In other words, efficient estimates form an equivalence class of probability order $N^{-\frac{1}{2}}$; i.e., for any two efficient estimates θ_1^* and θ_2^*, $\theta_1^* - \theta_2^*$ converges to 0 in probability order of $N^{-\frac{1}{2}}$. Then, as a next step, it is natural to ask whether it is possible to differentiate this equivalence class by means of a higher probability order, say of $1/N$. This problem was originally investigated

Received September 1976; revised October 1977.

[1]This is a revised version of chapter two of the author's Ph.D. thesis (Yale University, 1974).

AMS 1970 *subject classifications.* Primary 62M10; secondary 62F20.

Key words and phrases. Linear processes, the maximum-likelihood estimate, high-order efficiency, information amount, asymptotic information loss, asymptotic theory, Whittle-Walker model.

by Fisher (1922, 1925), who conjectured the superiority of the maximum-likelihood estimate over other efficient estimates by comparing their amounts of information. Later Rao (1962) and Takeuchi (1965) affirmed this fact for estimates from independent observations of the multinomial and the exponential family of distributions respectively; i.e., by means of a measure which suitably expresses the information loss suffered through the reduction of original data to an estimate, a comparison becomes possible between the maximum-likelihood estimate and other efficient estimates. Rao and Takeuchi concluded that in their models the information loss suffered by $\hat{\theta}$, the maximum-likelihood estimate, is less than that of θ^*, another estimate, and that the difference in information losses between them is proportional to the asymptotic variance of $N(\hat{\theta} - \theta^*)$.

The objective of this paper is to extend these results to the Gaussian linear process, and to establish the optimality of the maximum-likelihood estimate in that situation. In a sense this is a defense of the likelihood principle from the sampling-theoretic standpoint.

1. Asymptotic information loss. Estimation can be regarded as a reduction of data (or reduction of evidence) as well as "point" estimation of parameters. From the former point of view, any desirable estimation procedure must be such that given data are reduced to an estimate without much loss of their information. The amount of information contained in a statistic is usually measured by Fisher's information: namely, let $P(\theta^*|\theta)$ be the probability density of a statistic θ^*, then Fisher's information is defined as $E(d \log P(\theta^*|\theta)/d\theta)^2$. (We assume that $\theta \in R^1$ throughout this paper for simplicity of exposition. The following arguments can be easily extended to the vector valued θ so long as θ is finite-dimensional). Therefore the information loss suffered by the reduction of data $X = (X_1, X_2, \cdots, X_N)$ to an estimate θ^* can be expressed as

$$E\left(\frac{d \log L(\theta|X)}{d\theta}\right)^2 - E\left(\frac{d \log P(\theta^*|\theta)}{d\theta}\right)^2$$

(the likelihood function of θ given X is denoted as $L(\theta|X)$) which reduces to $E \operatorname{Var}(d \log L(\theta|X)/d\theta|\theta^*)$ if $dP(\theta^*|\theta)/d\theta$ is dominated by an integrable function (the $\operatorname{Var}(\cdot|\cdot)$ is the conditional variance). However, if θ^* is consistent, this information loss $E \operatorname{Var}((d \log L(\theta|X)/d\theta)|\theta^*)$ is meaningless asymptotically. Thus, in order to investigate the asymptotic behaviour of θ^*, it is more convenient to consider the quantity $l_{\theta^*}(\theta^0) = \underline{E}\ \underline{\operatorname{Var}}((d \log L(\theta^0|X)/d\theta)|N^{\frac{1}{2}}(\theta^* - \theta^0))$, where the underlined expectation and conditional variance are understood to be taken according to the joint asymptotic distribution of $d \log L(\theta^0|X)/d\theta$ and $N^{\frac{1}{2}}(\theta^* - \theta^0)$. (This formulation of asymptotic information loss was given implicitly by Rao (1962) and explicitly by Takeuchi (1965)). Let us call this quantity the *asymptotic information loss of* θ^*.

REMARK. In the above definition, the conditional variance is taken with respect to the asymptotic distribution of $d \log L(\theta|X)/d\theta$ given $N^{\frac{1}{2}}(\theta^* - \theta^0)$. Namely, let $\mathfrak{L}_N(\cdot|N^{\frac{1}{2}}(\theta^* - \theta^0))$ be the conditional distribution of $d \log L(\theta|X)/d\theta$ given

$N^{\frac{1}{2}}(\theta^* - \theta^0)$, and let $N^{\frac{1}{2}}(\theta^* - \theta^0)$ converge to η^* in distribution. Then, if $\mathcal{L}_N(\cdot|N^{\frac{1}{2}}(\theta^* - \theta^0)) \to \mathcal{L}(\cdot|\eta^*)$ as $N \to \infty$, the notation $\underline{\mathrm{Var}}(\cdot|N^{\frac{1}{2}}(\theta^* - \theta^0))$ is used to express the variance of $\mathcal{L}(\cdot|\eta^*)$. Thus $\underline{\mathrm{Var}}(\cdot|N^{\frac{1}{2}}(\theta^* - \theta^0))$ is a function of η^*, and the first expectation in the above definition is taken with respect to the distribution of η^*. Throughout this paper, the notations $\underline{E}$, $\underline{\mathrm{Var}}$ and $\underline{\mathrm{Cov}}$ will be used to signify respectively the expectation, variance and covariance of the asymptotic distribution of the argument statistics.

Now return to our model of the linear process $X_t = \Sigma_{i=0}^{\infty}\mu_i(\theta)\varepsilon_{t-i}$. The theory of second-order stationary processes says that the nondeterministic stationary process X_t is a process in which the X_t can be expressed as $X_t = \Sigma_{i=0}^{\infty}\mu_i\varepsilon_{t-i}$ where ε_t is an orthogonal process with mean 0 and $\{\mu_i\}$ satisfies $\Sigma_{i=0}^{\infty}\mu_i^2 < \infty$. In this case the spectral density of X_t is given by $f(\omega) = \sigma^2/2\pi|\Sigma_{j=0}^{\infty}\mu_j e^{i\omega j}|^2$. Now suppose further that X_t is a Gaussian process with mean zero and the auto-covariance γ_s. Let V_N be the matrix whose (t, s) element is γ_{t-s}, $(t, s = 1, 2, \cdots, N)$. Then the log-likelihood function, given observations $X_1, X_2, \cdots, X_N$, can be expressed as

$$\text{(1)} \qquad \log L_N(V_N|X_1, \cdots, X_N) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\log|V_N| - \frac{1}{2}Q_N(X, V_N)$$

where X is the vector $(X_1, \cdots, X_N)'$ and $Q_N(X, V_N) = X'V_N^{-1}X$. Whittle (1952) showed that when X_t is nondeterministic, $Q_N(X, V_N)$ can be approximated by $U_N(X, \mu)$ $(= 1/(2\pi)^2\int_{-\pi}^{\pi}|\Sigma_{t=1}^N X_t e^{i\omega t}|^2/f(\omega)d\omega)$ in the sense that

$$\left(\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}|\Sigma_{t=1}^N X_t e^{i\omega t}|^2/f(\omega)d\omega\right)\Big/\left(X'V_N^{-1}X\right) \to 1 \text{ a.e.} \qquad \text{as} \quad N \to \infty.$$

He also showed that if $\Sigma_{k=0}^{\infty}\mu_k Z^k$ is analytic and nonzero on $\{Z : |Z| < 1 + \delta\}$ for some $\delta > 0$, $1/N \log|V_N| \to \log \sigma^2$ as $N \to \infty$. Accordingly the above log-likelihood function may be approximated by

$$\text{(2)} \qquad \log L_N^*(\mu|X) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\log \sigma^2 - \frac{1}{2}U_N(X, \mu);$$

equation (2) gives the function which is the basis of our study from now on. In the following $L(\theta|X)$ will be used to designate $L_N^*(\theta|X)$. This assumption will evidently limit the generality of the following argument to a certain extent.

In order for any statistical investigation to have meaningful results, further restriction of the model is required: as above, observations X_t are generated by a linear process $X_t = \Sigma_{i=0}^{\infty}\mu_i(\theta)\varepsilon_{t-i}$, where, however, the ε_t are independent normal random variables with $E(\varepsilon_t^2) = \sigma^2$, and the μ_i are functions solely of θ. Thus the spectral density is a function of θ. Denoting by $f(\omega|\theta)$ the function $\sigma^2|\Sigma\mu_j(\theta)^{i\omega j}|^2/2\pi$, assume

(A-1) θ^0, the true value of θ, is in Θ, a compact subset of R^1,

(A-2) $f(\omega|\theta^1)$ cannot be equal to $f(\omega|\theta^2)$ a.e., for $\theta^1 \neq \theta^2$;

(A-3) $h(\omega|\theta) = 1/f(\omega|\theta)$ and $f(\omega|\theta)$ has continuous first derivatives in ω for $|\omega| < \pi$, $\theta \in \Theta$.

(A-4) The second order and the third order derivatives of $h(\omega|\theta)$ with respect to θ

exist and are continuous in (ω, θ) for $|\omega| \leqslant \pi$, and for θ in $N_{\delta_1}(\theta^0)$, a neighborhood of θ^0; namely $N_{\delta_1}(\theta^0) = \{\theta : |\theta - \theta^0| < \delta_1\}$;

(A-5) $\sum_{i=0}^{\infty} i|\mu_i(\theta^0)| < \infty$.

Under conditions A-1 to A-5, it is known that for $\hat{\theta}$, the maximum-likelihood estimate of θ, $N^{\frac{1}{2}}(\hat{\theta} - \theta^0)$ is asymptotically normal with mean 0 and with variance $\{1/4\pi \int (dh(\omega|\theta^0)/d\theta/h)^2\, d\omega\}^{-1}$ (cf. Whittle (1952), Walker (1964) and Hosoya (1974)).

The asymptotic information loss defined above can be calculated for the maximum-likelihood estimate $\hat{\theta}$ as follows.

LEMMA 1. *The asymptotic information loss $l_{\hat{\theta}}(\theta^0)$ of $\hat{\theta}$ at θ^0 is given by*

$$(3)\quad \int \left(\frac{d^2h/d\theta^2}{h}\right)^2 d\omega \left\{\int \left(\frac{dh/d\theta}{h}\right)^2 d\omega\right\}^{-1} - \left\{\int \left(\frac{d^2h/d\theta^2}{h}\right)\left(\frac{dh/d\theta}{h}\right) d\omega\right\}^2 \left\{\int \left(\frac{dh/d\theta}{h}\right)^2 d\omega\right\}^{-2}.$$

PROOF. By the Taylor expansion of $d \log L(\hat{\theta}|X)/d\theta$ around θ^0,

$$-\frac{d}{d\theta}\log L(\theta^0|X) = (\hat{\theta} - \theta^0)\frac{d^2 \log L(\theta^0|X)}{d\theta^2} + \frac{1}{2}(\hat{\theta} - \theta^0)^2 \frac{d^3 \log L(\tilde{\theta}|X)}{d\theta^3}$$

where $\hat{\theta} \gtrless \tilde{\theta} \gtrless \theta^0$. Then the second term in the right-hand side above converges to 0 in probability; thus for large N,

$$\frac{d}{d\theta}\log L(\theta^0|X) \doteq -N^{\frac{1}{2}}(\hat{\theta} - \theta^0)\frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}.$$

Now

$$\begin{aligned} l_{\hat{\theta}}(\theta^0) &= \underline{E}\ \underline{\operatorname{Var}}\left(d \log L(\theta^0|X)/d\theta\, |N^{\frac{1}{2}}(\hat{\theta} - \theta^0)\right) \\ &= \underline{E}\left[\left\{N^{\frac{1}{2}}(\hat{\theta} - \theta^0)\right\}^2 \underline{\operatorname{Var}}\left\{d^2 \log L(\theta^0|X)/N^{\frac{1}{2}}\, d\theta^2 |N^{\frac{1}{2}}(\hat{\theta} - \theta^0)\right\}\right]. \end{aligned}$$

Suppose $N^{\frac{1}{2}}(\hat{\theta} - \theta^0)$ and $d^2 \log L(\theta^0|X)/N^{\frac{1}{2}}\, d\theta^2$ are asymptotically jointly normally distributed (this fact will be established immediately below). Then, denoting the asymptotic variance of $N^{\frac{1}{2}}(\hat{\theta} - \theta^0)$ by $V(\theta^0)$, and using the fact that generally if U and V are jointly normally distributed, the conditional variance $\operatorname{Var}(U|V)$ is equal to $\operatorname{Var}(U)(1 - (\operatorname{Cov}(U, V)^2/\operatorname{Var}(U)\operatorname{Var}(V)))$, we have

$$\begin{aligned} (4)\quad l_{\hat{\theta}}(\theta^0) &= \underline{E}\left[\left\{N^{\frac{1}{2}}(\hat{\theta} - \theta^0)\right\}^2 \underline{\operatorname{Var}}\left\{\frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}\right\}\right. \\ &\quad \times \left.\left[1 - \underline{\operatorname{Cov}}\left\{N^{\frac{1}{2}}(\hat{\theta} - \theta^0), \frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}\right\}^2 \Big/ V(\theta^0)\, \underline{\operatorname{Var}}\left\{\frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}\right\}\right]\right] \\ &= V(\theta^0)\, \underline{\operatorname{Var}}\left\{\frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}\right\} - \left[\underline{\operatorname{Cov}}\left\{N^{\frac{1}{2}}(\hat{\theta} - \theta^0), d^2 \log L(\theta^0|X)/N^{\frac{1}{2}}\, d\theta^2\right\}\right]^2. \end{aligned}$$

Now $N^{\frac{1}{2}}(\hat{\theta} - \theta^0)$ is asymptotically distributed as

$$\left\{\frac{1}{4\pi}\int\left(\frac{dh(\omega|\theta^0)/d\theta}{h(\omega|\theta^0)}\right)^2 d\omega\right\}^{-1}\frac{1}{2\sigma_0^2}\Sigma_{|s|\leqslant N-1}\frac{d\alpha_s(\theta^0)}{d\theta}N^{\frac{1}{2}}\{C_s - E(C_s)\},$$

as was shown by Walker (1964), and also $d^2 \log L(\theta^0|X)/N^{\frac{1}{2}}\, d\theta^2 - E(d^2 \log L(\theta^0|X)/N^{\frac{1}{2}}\, d\theta^2)$ is distributed as

$$\frac{1}{2\sigma_0^2}\Sigma_{|s|\leqslant N-1}\frac{d^2\alpha_s(\theta^0)}{d\theta^2}N^{\frac{1}{2}}\{C_s - E(C_s)\},$$

where $C_s = \Sigma_{t=1}^{N-|s|}X_t X_{t+|s|}/N$ and $\alpha_s(\theta) = \sigma^2/4\pi^2\int_{-\pi}^{\pi}e^{i\omega s}h(\omega|\theta)\,d\omega$. Accordingly, owing to the fact that any finite set of $N^{\frac{1}{2}}\{C_s - E(C_s)\}$ is asymptotically jointly normal, $N^{\frac{1}{2}}(\hat{\theta} - \theta^0)$ and $d^2 \log L(\theta^0|X)/N^{\frac{1}{2}}\, d\theta^2$ are jointly normal with $\underline{\mathrm{Var}}(N^{\frac{1}{2}}(\hat{\theta} - \theta^0)) = [(1/4\pi)\int(dh(\omega|\theta^0)/d\theta/h(\omega|\theta^0))^2\, d\omega]^{-1}$,

$$\underline{\mathrm{Cov}}\left(N^{\frac{1}{2}}(\hat{\theta} - \theta^0), \frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}\right)$$

$$= \lim_{N\to\infty}\Sigma_{|s|\leqslant N}\Sigma_{|t|\leqslant N}\left\{\frac{1}{4\pi}\int\left(\frac{dh(\omega|\theta^0)/d\theta}{h(\omega|\theta^0)}\right)^2 d\omega\right\}^{-1}\frac{1}{4\sigma_0^4}\frac{d^2\alpha_t(\theta^0)}{d\theta^2}\frac{d\alpha_s(\theta^0)}{d\theta}$$

$$\times E\left\{N^{\frac{1}{2}}(C_s - E(C_s))N^{\frac{1}{2}}(C_t - E(C_t))\right\}$$

$$= \left[\int\left(\frac{dh/d\theta}{h}\right)\left(\frac{d^2h/d\theta^2}{h}\right)d\omega\right]\Big/\left[\int\left(\frac{dh/d\theta}{h}\right)^2 d\omega\right]$$

and

$$\underline{\mathrm{Var}}\left(\frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}\right) = \lim_{N\to\infty}\frac{1}{4\sigma_0^4}\Sigma_{|s|\leqslant N}\Sigma_{|t|<N}\frac{d^2\alpha_t(\theta^0)}{d\theta^2}\frac{d^2\alpha_s(\theta^0)}{d\theta^2}$$

$$\times E\left\{N^{\frac{1}{2}}(C_t - E(C_t))N^{\frac{1}{2}}(C_s - E(C_s))\right\}$$

$$= \frac{1}{4\pi}\int\left(\frac{d^2h/d\theta^2}{h}\right)^2 d\omega.$$

Therefore substituting these values to those in (4), the lemma is obtained. □

The asymptotic information loss of other estimates is very difficult to calculate directly. But it is possible to derive a relation which compares the asymptotic information loss of the maximum-likelihood estimate and other efficient estimates. Takeuchi (1965) establishes the following relation for i.i.d. observations of the exponential family of distributions: that is, denoting an efficient estimate by θ^*,

$$(5) \qquad l_{\theta^*}(\theta^0) = \underline{\mathrm{Var}}\left\{N(\hat{\theta} - \theta^*)\right\}/V(\theta^0)^2 + l_{\hat{\theta}}(\theta^0),$$

where the variance is taken with respect to the asymptotic distribution of $N(\hat{\theta} - \theta^*)$ and $V(\theta^0)$ is the asymptotic variance of $N^{\frac{1}{2}}(\hat{\theta} - \theta^0)$.

The purpose of the following sections is to show that the relation (5) holds also for the observations from the Gaussian linear process satisfying the conditions A-1 through A-5. In order to prove this, however, certain limitation on the class of efficient estimates is necessary. In the following discussion, the class of estimates of θ is assumed to satisfy these conditions: if θ_N^* is in this class, then

(B-1) θ_N^* is a function of $C_{0,N}, C_{1,N}, \cdots, C_{N-1,N}$ where $C_{i,N} = \Sigma_{j=1}^{N-i} X_j X_{j+i}/N$;

(B-2) θ_N^* is second-order differentiable with respect to $C_{i,N}$, $i = 0, 1, \cdots, N-1$. Let

$$\left.\frac{\partial \theta_N^*}{\partial C_{i,N}}\right|_{\{C_{i,N} = \gamma_i(\theta)\}} = \alpha_{iN}^*$$

$$\left.\frac{\partial^2 \theta_N^*}{\partial C_{i,N} \partial C_{j,N}}\right|_{\left\{\begin{array}{ll} C_{i,N} = \gamma_i(\theta), & i = 0, 1, \cdots, N-1 \\ C_{j,N} = \gamma_j(\theta), & j = 0, 1, \cdots, N-1 \end{array}\right\}} = \beta_{ijN}^*$$

where $\gamma_i(\theta) = E(X_j X_{j+i} | \theta)$;

(B-3) $\theta_N^* = \theta + \Sigma_{i=0}^{N-1} \alpha_{iN}^* (C_{i,N} - \gamma_i(\theta)) + \Sigma_i \Sigma_j \beta_{ijN}^* (C_{i,N} - \gamma_i(\theta))(C_{j,N} - \gamma_j(\theta)) + R_N$, where $R_N = o_p(1/N)$ if the true parameter is θ, for θ in a neighborhood of θ^0. $\theta_N^*(\gamma_0(\theta), \gamma_1(\theta), \cdots, \gamma_{N-1}(\theta)) \to \theta$ in a neighborhood of θ^0 as $N \to \infty$ (a weak version of the Fisher consistency);

(B-4) there are α_i and β_{ij} such that

$$\Sigma_{i=0}^{N-1} |\alpha_{iN}^* - \alpha_i| i \to 0 \quad \text{and} \quad \Sigma_{i=0}^{N-1} \Sigma_{j=0}^{N-1} |\beta_{ijN}^* - \beta_{ij}| ij \to 0,$$

as $N \to \infty$, and besides, these convergences are uniform in a certain neighborhood of θ^0;

(B-5) for the α in 4 above,

$$N^{\frac{1}{2}} \Sigma_{i=0}^{N-1} |\alpha_{iN}^* - \alpha_i| \to 0 \quad \text{as} \quad N \to \infty.$$

Among the above conditions, the conditions B-4 and B-5 are the strong ones. The condition B-4 might be fairly reasonable if we take into consideration the fact that, practically, an estimate should not depend on the high-order sampling autocovariances, and it can be proved that in the rational spectrum model the maximum-likelihood estimate satisfies this condition. The condition B-5 implies that, for any two estimates θ_1^* and θ_2^*, with $\delta < 1/2$, $N^{\frac{1}{2}+\delta}(\theta_1^* - \theta_2^*) \to 0$ in probability as $N \to \infty$ (this is shown at end of Section 3).

To summarize, we assume that

(A) the X_t is a Gaussian linear process, and the regularity conditions A-1 through A-5 hold;

(B) efficient estimates are those which satisfy the conditions B-1 through B-5.

Under A and B, the relation (5) will be proved. But the proof requires rather complicated steps; therefore, for the purpose of comprehensive reading, we shall give a rough heuristic sketch of the proof here and thus show the interrelation of the following three sections that are necessary steps to complete the proof.

Let θ^* be an efficient estimate of θ and $\hat{\theta}$ be the maximum-likelihood estimate, then for sufficiently large N,

$$0 = \frac{d \log L(\hat{\theta}|X)}{d\theta} \doteq \frac{d \log L(\theta^0|X)}{d\theta} + \frac{d^2 \log L(\theta^0|X)}{d\theta^2}(\hat{\theta} - \theta^0)$$

$$= \frac{d \log L(\theta^0|X)}{d\theta} + \frac{d^2 \log L(\theta^0|X)}{d\theta^2}(\hat{\theta} - \theta^* + \theta^* - \theta^0).$$

Accordingly,

$$\frac{d \log L(\theta^0|X)}{d\theta} \doteq N(\theta^* - \hat{\theta})\frac{d^2 \log L(\theta^0|X)}{N\, d\theta^2} - N^{\frac{1}{2}}(\theta^* - \theta^0)\frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}$$

in which $-d^2 \log L(\theta^0|X)/N\, d\theta^2$ can be replaced by $V(\theta^0)^{-1}$. Now, in view of the definition of $l_{\theta^*}(\theta^0)$,

$$l_{\theta^*}(\theta^0)$$

$$= \underline{E}\left[\underline{\text{Var}}\left\{\frac{d \log L(\theta^0|X)}{d\theta}\middle| N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}\right]$$

$$= \underline{E}\left[\underline{\text{Var}}\left\{N(\hat{\theta} - \theta^*)/V(\theta^0) - N^{\frac{1}{2}}(\theta^* - \theta^0)\frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}\middle| N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}\right]$$

$$= \underline{E}\left[\underline{\text{Var}}\left\{N(\hat{\theta} - \theta^*)|N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}\right]/V(\theta^0)^2$$

$$+ \underline{E}\left[\underline{\text{Var}}\left\{N^{\frac{1}{2}}(\theta^* - \theta^0)\frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}\middle| N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}\right]$$

$$-2\,\underline{E}\left[\underline{\text{Cov}}\left\{N(\hat{\theta} - \theta^*)/V(\theta^0), N^{\frac{1}{2}}(\theta^* - \theta^0)\frac{d^2 \log L(\theta^0|X)}{N^{\frac{1}{2}}\, d\theta^2}\middle| N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}\right]$$

where the second term in the last expression is equal to $l_{\hat{\theta}}(\theta^0)$, since $N^{\frac{1}{2}}(\theta^* - \theta^0)$ is asymptotically distributed in the same way as $N^{\frac{1}{2}}(\hat{\theta} - \theta^0)$. If $N(\theta^* - \hat{\theta})$ and $N^{\frac{1}{2}}(\theta^* - \theta^0)$ are asymptotically independent, the first term is equal to $\text{Var}\{N(\theta^* - \hat{\theta})\}/V(\theta^0)^2$. Section 4 will show that the third term in the above is equal to 0 under the same condition. Thus it is necessary to prove the asymptotic independence. In Section 3 an important property of efficient estimates will be shown: namely if θ_1^* and θ_2^* are efficient, $N^{\frac{1}{2}}(\theta_1^* - \theta^0)$ and $N(\theta_1^* - \theta_2^*)$ are asymptotically independent. That section will show that $N^{\frac{1}{2}}(\theta_1^* - \theta^0)$ and $N(\theta_1^* - \theta_2^*)$ are reduced to a linear form and a quadratic form such as $\Sigma_{i=0}^{\infty}\alpha_i T_i$ and $\Sigma_{i=0}^{\infty}\Sigma_{j=0}^{\infty}(\beta_{ij}^* - \beta_{ij})T_i T_j$ respectively, and the asymptotic independence between them will be proved by means of an extension of the Craig-Aitken theorem. Section 2 will construct this extended version of the Craig-Aitken theorem.

For practical applications of the relation (5), it is straightforward to evaluate $V(\theta^0)$ and $l_{\hat{\theta}}(\theta^0)$ for a specific model by means of Lemma 1, whereas the evaluation of the quantity $\underline{\text{Var}}(N(\hat{\theta} - \theta^*))$ will necessitate computer simulation since it is not usually expressed by simple analytic formula.

2. Extension of Craig-Aitken theorem. Craig (1943) showed that when $X = (X_1, X_2, \cdots, X_N)'$ is a vector of independent normal variates with unit variance, the two quadratic forms $X'AX$ and $X'BX$ are independent if and only if $AB = 0$. Aitken (1950) extended this result to the case where X has a nonsingular variance-covariance matrix Σ and proved that $X'AX$ and $X'BX$ are independent if and only if $B\Sigma A = 0$.

The above Craig-Aitken results can be further extended to an infinite-dimensional X.

Assume

(B-6) $\{X_t : t = \cdots, -1, 0, 1, \cdots\}$ is a real-valued Gaussian (but not necessarily stationary) process with mean 0 and covariances $\kappa_{t,s} = \mathrm{Cov}(X_t, X_s)$, $t, s = \cdots, -1, 0, 1, \cdots$, which are harmonizable; i.e., $E(X_t X_s)$ has the representation

$$E(X_t X_s) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i(t\omega - s\lambda)}\, dF(\omega, \lambda), \tag{6}$$

and $F(\omega, \lambda)$ is such that $F(\pi, \pi) < \infty$.

(B-7) $\{\alpha_i\}_{i=0}^{\infty}$ and $\{\beta_{ij}\}_{i,j=0}^{\infty}$ are sequences such that $\beta_{ij} = \beta_{ji}$ for all i and j, and $\Sigma_i|\alpha_i| < \infty$ and $\Sigma_i\Sigma_j|\beta_{ij}| < \infty$.

We then have

THEOREM 1. *Under conditions* B-6 *and* B-7, $\Sigma_{i=0}^{\infty}\alpha_i X_i$ *and* $\Sigma_i\Sigma_j\beta_{ij}X_iX_j$ *converge in mean square, and are independent if and only if* $\Sigma_j\Sigma_k\beta_{ij}\kappa_{j,k}\alpha_k = 0$ *for all* $i = 0, 1, 2, \cdots$

PROOF. Let $a_M = \Sigma_{i=0}^{M-1}\alpha_i X_i$, $M = 1, 2, \cdots$ Observe that $\{a_M\}$ forms a Cauchy sequence in the L^2 space generated by linear combinations of the $\{X_t, t = 0, 1, \cdots\}$. For if $N > M$,

$$E|A_N - A_M|^2 \leqslant \int|\Sigma_{n=M}^{N-1}|\alpha_n|\,|^2 dF(\omega, \lambda) \leqslant K\{\Sigma_{n=M}^{N-1}|\alpha_n|\}^2,$$

with $K = F(\pi, \pi)$. Thus by taking N, M large enough, $E|A_N - A_M|^2 < \varepsilon$ for any given $\varepsilon > 0$. Similar arguments hold also for the convergence of $\Sigma\Sigma\beta_{ij}X_iX_j$. Let $A = \Sigma_{i=0}^{\infty}\alpha_i X_i$ and $B = \Sigma_{i=0}^{\infty}\Sigma_{j=0}^{\infty}\beta_{ij}X_iX_j$. In order to continue the proof, the following lemma due to Loève (1963, page 476) is necessary.

LEMMA 2. *When the covariance function of a Gaussian process* $\{X_t\}$ *is harmonizable, there exists a family of complex-valued random variables* $\{Z(\lambda)\}$ *indexed by* $\lambda \in [-\pi, \pi)$ *such that*

(i) *for any* $\lambda_1, \cdots, \lambda_m \in [-\pi, \pi)$; $Z(\lambda_1), \cdots, Z(\lambda_m)$ *have a joint normal distribution;*

(ii) $E(Z(\lambda_1)\overline{Z(\lambda_2)}) = F(\lambda_1, \lambda_2)$ *for any* $\lambda_1, \lambda_2 \in [-\pi, \pi)$; *and*

(iii) $X_t = \int_{-\pi}^{\pi} e^{it\omega}Z(d\omega)$ *with probability* 1.

(The integral in (iii) is defined as the L^2-limit(F) of $\Sigma_{j=1}^{M} e^{it\omega_j}Z(I_j)$ as $M \to \infty$, where the intervals $I_j, j = 1, 2, \cdots, M$, are a partition of $[-\pi, \pi)$ and $\omega_j \in I_j$.) To return to the proof, in view of (iii) above, $A_M = \int \Sigma_{j=0}^{M-1} \alpha_j e^{ij\omega} Z(d\omega)$. Let $f(\omega) = \Sigma_{m=0}^{\infty} \alpha_m e^{im\omega}$; then, since $\Sigma_{i=0}^{\infty}|\alpha_i| < \infty$, $\int_{-\pi}^{\pi} f(\omega)Z(d\omega)$ is defined, and $A_M \to \int_{-\pi}^{\pi} f(\omega)Z(d\omega)$ as $M \to \infty$ in $L^2(dF)$. The latter statement comes from the fact that for any $\varepsilon > 0$, by taking M large enough, $E|A_M - \int_{-\pi}^{\pi} f(\omega)Z(d\omega)|^2 \leqslant K\{\Sigma_{j=M}^{\infty}|\alpha_j|\}^2 < \varepsilon$.

Thus,

$$A = \int f(\omega)Z(d\omega). \tag{7}$$

In the same way,

$$B = \int\int g(\omega, \lambda)Z(d\omega)Z(d\lambda) \tag{8}$$

where obviously $\int|f(\omega)|d\omega < \infty$ and $\int|g(\omega, \lambda)|d\omega\, d\lambda < \infty$.

The relation $\Sigma_j \Sigma_k \beta_{ij} \kappa_{j,k} \alpha_k = 0$, for all i, can be written as

$$\int\int g(\nu, \omega)\overline{f(\lambda)}\ dF(\omega, \lambda) = 0 \qquad \text{for all} \quad \nu. \tag{9}$$

In general, a symmetric L^2-function $h(\omega, \lambda)$ has the representation $h(\omega, \lambda) = \Sigma_{i=0}^{\infty} \mu_i \phi_i(\omega)\phi_i(\lambda)$ in terms of an orthonormal system $\{\phi_i : i = 0, 1, 2, \cdots\}$, where the limit of the right-hand side in the above equation is in the sense of L^2 (cf. Tricomi (1957)). Thus $g(\nu, \omega)$ may be written as

$$g(\nu, \omega) = \Sigma_{i=0}^{\infty} \eta_i \phi_i(\nu)\phi_i(\omega). \tag{10}$$

Therefore, from (9), $\Sigma_i \eta_i \phi_i(\nu) \int\int \phi_i(\omega)\overline{f(\lambda)}\ dF(\omega, \lambda) = 0$ for all ν. But this is possible if and only if $\int\int \phi_i(\omega)\overline{f(\lambda)}\ dF(\omega, \lambda) = 0$ for all i such that $\eta_i \neq 0$. Now in view of (10), $B = \Sigma_i \eta_i |\int \phi_i(\omega)Z(d\omega)|^2$. However, by (9)

$$E\{\int \phi_i(\omega)Z(d\omega) \cdot \int \overline{f(\lambda)Z(d\lambda)}\} = \int\int \phi_i(\omega)\overline{f(\lambda)}\ dF(\omega, \lambda) = 0.$$

Since $\int \phi_i(\omega)Z(d\omega)$ and $\int f(\lambda)Z(d\lambda)$ are normal, they are independent. Accordingly, $|\int \phi_i(\omega)Z(d\omega)|^2$ and $\int f(\lambda)Z(d\lambda)$ are independent. Thus, in view of (7), A and B are independent. □

COROLLARY 1. *Assume a Gaussian process $\{X_t\}$ has the same properties as in Lemma 2, and $\{a_{ij}\}_{i,j=0}^{\infty}$ and $\{b_{ij}\}_{i,j=0}^{\infty}$ are symmetric and $\Sigma_i\Sigma_j|a_{ij}| < \infty$ and $\Sigma_i\Sigma_j|b_{ij}| < \infty$. Then $\Sigma\Sigma a_{ij}X_iX_j$ and $\Sigma\Sigma b_{ij}X_iX_j$ are independent if and only if*

$$\Sigma_j \Sigma_h a_{ij} \kappa_{j,h} b_{hl} = 0 \qquad \textit{for all} \quad i \quad \textit{and} \quad l. \tag{11}$$

PROOF. As in Lemma 1, a_{ij} and b_{ij} have corresponding symmetrical L^2-kernels $f(\omega, \lambda)$ and $g(\omega, \lambda)$ which are respectively represented as

$$\begin{aligned} f(\omega, \lambda) &= \Sigma \nu_i \phi_i(\omega)\phi_i(\lambda) \\ g(\omega, \lambda) &= \Sigma \mu_i \psi_i(\omega)\psi_i(\lambda) \end{aligned} \tag{12}$$

Also (11) implies that

$$\int\int \phi_i(\omega)\overline{\psi_j(\lambda)}\ dF(\omega, \lambda) = 0 \text{ for all } i \text{ and } j \text{ such that } \nu_i \neq 0, \mu_j \neq 0. \tag{13}$$

Then the result is straightforward from (12) and (13). □

3. Asymptotic independence between $N^{\frac{1}{2}}(\theta_1^* - \theta^0)$ and $N(\theta_1^* - \theta_2^*)$. By use of the extended Craig-Aitken theorem of the previous section, the asymptotic independence between $N^{\frac{1}{2}}(\theta_1^* - \theta^0)$ and $N(\theta_2^* - \theta_1^*)$ is proved in this section. This is a crucial property for the proof of Theorem 2 given in the next section.

Let T_i be such that $N^{\frac{1}{2}}(C_{i,N} - \gamma_i) \to T_i$ in distribution. Then,

LEMMA 3. *$\Sigma_{i=0}^{N-1}\alpha_{iN}^* N^{\frac{1}{2}}(C_{i,N} - \gamma_i)$ and $\Sigma_i\Sigma_j\beta_{ijN}^* N^{\frac{1}{2}}(C_{i,N} - \gamma_i)N^{\frac{1}{2}}(C_{j,N} - \gamma_j)$ are asymptotically distributed as $\Sigma\alpha_i T_i$ and $\Sigma\Sigma\beta_{ij}T_iT_j$ respectively, where the α_i and the β_{ij} are such limits that $\alpha_{iN}^* \to \alpha_i$ and $\beta_{ijN}^* \to \beta_{ij}$, as $N \to \infty$.*

PROOF. First, take M large enough so that, for a given $\varepsilon > 0$, $|\Sigma_{i=M}^{\infty} i\alpha_i| < \varepsilon$, and $|\Sigma_{i=M}^{\infty}\alpha_i^2| < \varepsilon$. This follows from condition (B-4). For this M fixed, obviously $\Sigma_{i=0}^{M-1}\alpha_{iN}^* N^{\frac{1}{2}}(C_{i,N} - \gamma_i)$ converges in distribution to $\Sigma_{i=0}^{M-1}\alpha_i T_i$ as $N \to \infty$. Next consider the remaining terms.

$$E\left\{\Sigma_{j=M}^{N-1} N^{\frac{1}{2}}\alpha_{jN}^*(C_{j,N} - \gamma_j)\right\}^2$$
$$= \Sigma_j\Sigma_k\alpha_{jN}^*\alpha_{kN}^* N\left\{E(C_{j,N}C_{k,N}) - \gamma_j(EC_{k,N}) - \gamma_k(EC_{j,N}) + \gamma_j\gamma_k\right\}$$
$$= \Sigma_j\Sigma_k\alpha_{jN}^*\alpha_{kN}^* N\left\{\Sigma_{n=1}^{N-j}\Sigma_{m=1}^{N-k}(\gamma_j\gamma_k + \gamma_{n-m}\gamma_{n-m+j-k} + \gamma_{n-m+j}\gamma_{n-m-k})/N^2 - \frac{N-k}{N}\gamma_j\gamma_k - \frac{N-j}{N}\gamma_j\gamma_k + \gamma_j\gamma_k\right\}$$

where

(i) $\Sigma_j\Sigma_k\alpha_{jN}^*\alpha_{kN}^* N\{\Sigma_n\Sigma_m\gamma_j\gamma_k/N^2 - (N - k/N)\gamma_j\gamma_k - (N - j/N)\gamma_j\gamma_k + \gamma_j\gamma_k\}$ $= \Sigma_j\Sigma_k jk\alpha_{jN}^*\alpha_{kN}^*\gamma_j\gamma_k/N < \eta$, for any given $\eta > 0$, by taking N large enough, since, with K such that $|\gamma_j| < K, j = 0, 1, 2, \cdots$, $|\Sigma_j\Sigma_k jk\alpha_{jN}^*\alpha_{kN}^*\gamma_j\gamma_k/N| \leqslant K^2|\Sigma_j j\alpha_{jN}^*|^2/N$;

(ii) (14)
$$\Sigma_j\Sigma_k\alpha_{jN}^*\alpha_{kN}^*\frac{1}{N}\Sigma_n\Sigma_m\gamma_{n-m}\gamma_{n-m+j-k} = \frac{1}{N}\int\int|\Sigma_j\Sigma_n\alpha_{jN}^* e^{i(\omega-\lambda)n - ij\lambda}|^2 f(\omega|\theta^0)f(\lambda|\theta^0)\,d\omega\,d\lambda$$

where $f(\omega|\theta^0)$ is the spectral density of X_t, and is bounded by (A-1) and (A-3). Let L be the essential supremum of $f(\omega|\theta^0)$. We have

$$|\Sigma_j\alpha_{jN}^{*2} - \Sigma_j\alpha_j^2| < \eta'/2 \qquad \text{and} \qquad |\Sigma_j j\alpha_{jN}^{*2} - \Sigma_j j\alpha_j^2| < \eta'/2$$

by taking N large enough. Therefore the right-hand side of (14) is less than or equal to

$$\frac{L^2}{N}\int\int|\Sigma_j\Sigma_n\alpha_{jN}^* e^{i(\omega-\lambda)n + j\lambda}|^2\,d\omega\,d\lambda = \frac{4\pi^2L^2}{N}\Sigma_j\alpha_{jN}^{*2}(N - j) < 2\varepsilon + \eta';$$

(iii) in the same way as in (ii), for any $\eta'' > 0$,

$$\frac{1}{N}\Sigma_j\Sigma_k\alpha_{jN}^*\alpha_{kN}^*\Sigma_n\Sigma_m\gamma_{n-m+j}\gamma_{n-m-k} \leqslant 2\varepsilon + \eta'',$$

by taking N large enough. Therefore, we conclude that, for sufficiently large

N and M,

$$E\left\{\Sigma_{j=M}^{N-1} N^{\frac{1}{2}}\alpha_{jN}^{*}(C_{j,N}-\gamma_j)\right\}^2 < 4\varepsilon+\eta+\eta'+\eta''.$$

Therefore $\Sigma_{j=0}^{N-1}\alpha_{jN}^{*}N^{\frac{1}{2}}(C_{j,N}-\gamma_j)$ converges to $\Sigma_{j=0}^{\infty}\alpha_j T_j$ in distribution.

The proof of the convergence of $\Sigma_i\Sigma_j\beta_{ijN}^{*}N^{\frac{1}{2}}(C_{i,N}-\gamma_i)\times N^{\frac{1}{2}}(C_{j,N}-\gamma_j)$ to $\Sigma_i\Sigma_j\beta_{ij}T_iT_j$ is computationally a little more complicated, but a similar argument as above holds. □

Condition (B-3) says that $\theta_N^{*}(\gamma_0(\theta), \gamma_1(\theta), \cdots, \gamma_{N-1}(\theta))\to\theta$ in a neighborhood of $\theta=\theta^0$. On the other hand, $d\theta_N^{*}(\gamma_0,\cdots,\gamma_{N-1})/d\theta=\Sigma\alpha_{iN}^{*}\,d\gamma_i(\theta)/d\theta$ which converges uniformly to $\Sigma\alpha_i\,d\gamma_i/d\theta$ in a certain neighborhood of θ^0 by condition (B-4). Thus,

$$\Sigma\alpha_i\,d\gamma_i/d\theta=1. \tag{15}$$

(The above statement follows from the fact that if $f_n\in C^1[a,b]$, $n=1,2,\cdots$, and f_n converges at least at $x_0\in[a,b]$, and if df_n/dx converges uniformly to g on $[a,b]$, then f_n converges uniformly to f, and $df/dx=g$. The uniform convergence of $\Sigma\alpha_{iN}^{*}\,d\gamma_i(\theta)/d\theta$ to $\Sigma\alpha_i\,d\gamma_i(\theta)/d\theta$ is obvious because of condition (B-4), since the $d\gamma_i/d\theta$ are bounded.)

For θ^{*} to be efficient, the asymptotic variance of $N^{\frac{1}{2}}(\theta^{*}-\theta^0)$ must be minimal under the restriction (15); let $\sigma_{ij}=\mathrm{Cov}(T_i,T_j)$, then the asymptotic variance of $N^{\frac{1}{2}}(\theta^{*}-\theta^0)$ is equal to $\Sigma_i\Sigma_j\alpha_i\alpha_j\sigma_{ij}$. Then α must be such that it attains $\min_\alpha\Sigma\Sigma\alpha_i\alpha_j\sigma_{ij}$ under $\Sigma\alpha_j\,d\gamma_j/d\theta=1$. Using the Lagrange multiplier λ, put $\phi=\Sigma\Sigma\alpha_i\alpha_j\sigma_{ij}-\lambda(\Sigma\alpha_i d\gamma_i/d\theta-1)$. Differentiate ϕ with respect to α_i, then

$$\Sigma\alpha_j\sigma_{ij}-\lambda\frac{d\gamma_i(\theta)}{d\theta}=0, \qquad \text{from which} \tag{16}$$

$$\Sigma\Sigma\alpha_i\alpha_j\sigma_{ij}-\lambda\Sigma\alpha_i\frac{d\gamma_i(\theta)}{d\theta}=0;$$

thus $\lambda=\Sigma\Sigma\alpha_i\alpha_j\sigma_{ij}=V(\theta)$. Therefore it follows from (15) and (16) that

$$\Sigma\alpha_i\sigma_{ij}=V(\theta)\frac{d\gamma_j(\theta)}{d\theta}, \qquad j=0,1,2,\cdots. \tag{17}$$

LEMMA 4. *The* $\{\alpha_i\}$ *which satisfies* (17) *is unique.*

PROOF. σ_{nm} has the representation $\sigma_{nm}=2\pi\int\{e^{i(n-m)\omega}+e^{i(n+m)\omega}\}(f(\omega|\theta))^2\,d\omega$. This is immediately derived by the Fourier transformation of covariances σ_{nm} where $f(\omega|\theta)$ is the spectral density of the process X_t (see Walker (1964), pages 374–5). Then (17) can be written as

$$2\pi\int g(\omega)e^{-ij\omega}(f(\omega|\theta))^2\,d\omega=V(\theta)\frac{d\gamma_j(\theta)}{d\theta}, \qquad j=0,1,2,\cdots \tag{18}$$

where $g(\omega)=2\Sigma_j\alpha_j\cos j\omega$. But $d\gamma_j(\theta)/d\theta=\int e^{ij\omega}(df(\omega|\theta)/d\theta)\,d\omega$, $j=\cdots,-1,0,1,\cdots$, and $df(\omega|\theta)/d\theta\in L_1$ and also $g(\omega)(f(\omega|\theta))^2$ is in L_1. Then, by the uniqueness of the Fourier coefficients of L_1-functions, (c.f., for example, Hewitt and Stromberg (1965)), $2\pi g(\omega)(f(\omega|\theta))^2 = V(\theta)(df(\omega|\theta))/d\theta$ a.e. where $f(\omega|\theta)$ is bounded away from 0, so that $g(\omega)$ is unique; thus α is unique. □

Now we can prove

LEMMA 5. *Let $\theta^*_{1,N}$ and $\theta^*_{2,N}$ be two efficient estimates, then $N^{\frac{1}{2}}(\theta^*_{1,N} - \theta^0)$ and $N(\theta^*_{1,N} - \theta^*_{2,N})$ are asymptotically independent.*

PROOF. Let $\theta^*_{1,N}$ and $\theta^*_{2,N}$ have expansions such that

$$\theta^*_{1,N} = \theta + \Sigma_{i=0}^{N-1}\alpha_{iN}(C_{i,N} - \gamma_i) + \Sigma_{i,j=0}^{N-1}\beta_{ijN}(C_{i,N} - \gamma_i)(C_{j,N} - \gamma_j) + o_p\left(\frac{1}{N}\right)$$

$$\theta^*_{2,N} = \theta + \Sigma_{i=0}^{N-1}\alpha^*_{iN}(C_{i,N} - \gamma_i) + \Sigma_{i,j=0}^{N-1}\beta^*_{ij,N}(C_{i,N} - \gamma_i)(C_{j,N} - \gamma_j) + o_p\left(\frac{1}{N}\right).$$

Let $\alpha_{iN} \to \alpha_i$ and $\alpha^*_{iN} \to \alpha^*_i$, and let $\beta_{ij,N}$ and $\beta^*_{ij,N}$ converge to β_{ij} and β^*_{ij} respectively.

The derivative of α_{KN} satisfies $d\alpha_{kN}/d\theta = \Sigma_j\beta_{kjN}\, d\gamma_j(\theta)/d\theta$, $k = 0, 1, 2, \cdots$, where the right-hand side uniformly converges to $\Sigma_j\beta_{kj}\, d\gamma_j(\theta)/d\theta$ in a neighborhood of θ^0 by (B-4). Therefore $d\alpha_k/d\theta$ exists and $d\alpha_k/d\theta = \Sigma_j\beta_{kj}\, d\gamma_j(\theta)/d\theta$. In view of Lemma 4, $\alpha_k = \alpha^*_k$, so that

$$\Sigma_j\beta_{kj}\frac{d\gamma_j(\theta)}{d\theta} = \Sigma_j\beta^*_{kj}\frac{d\gamma_j(\theta)}{d\theta}. \tag{19}$$

From (17), $d\gamma_j(\theta)/d\theta$ in the above can be replaced by $(1/V(\theta))\Sigma_l\alpha_l\sigma_{lj}$; hence

$$\Sigma_j\Sigma_l(\beta_{kj} - \beta^*_{kj})\sigma_{jl}\alpha_l = 0, \qquad k = 0, 1, 2, \cdots. \tag{20}$$

Now it follows from Lemma 3 that $N^{\frac{1}{2}}(\theta^*_{1,N} - \theta^0)$ is asymptotically distributed as $\Sigma\alpha_i T_i$. As for $N(\theta^*_{1,N} - \theta^*_{2,N})$, it can be shown, as follows, that the first term in its asymptotic expansion converges in probability to 0. There exists a positive constant M such that for all N

$$E|N\Sigma_i(\alpha^*_{iN} - \alpha_{iN})(C_{i,N} - \gamma_i)| \leqslant \Sigma N^{\frac{1}{2}}|\alpha^*_{iN} - \alpha_{iN}|\left[E\left\{N^{\frac{1}{2}}(C_{i,N} - \gamma_i)\right\}^2\right]^{\frac{1}{2}}$$
$$\leqslant M\Sigma_{i=0}^{N-1}N^{\frac{1}{2}}|\alpha^*_{iN} - \alpha_{iN}|,$$

since $E\{N^{\frac{1}{2}}(C_{i,N} - \gamma_i)\}^2 < M^2$ uniformly in i and N for a certain constant M (assumption (A-5) guarantees this fact). In view of assumption (B-5), it follows that $E|N\Sigma(\alpha^*_{iN} - \alpha_{iN})(C_{i,N} - \gamma_i)|$ converges to 0. Then the Chebyshev inequality concerning first-order absolute moment implies that $N\Sigma_{i=0}^{N-1}(\alpha^*_{iN} - \alpha_{iN})(C_{i,N} - \gamma_i)$ converges in probability to 0. Consequently, $N(\theta^*_{1,N} - \theta^*_{2,n})$ is asymptotically distributed as $\Sigma\Sigma(\beta_{ij} - \beta^*_{ij})T_iT_j$. Thus, in view of Theorem 1 in Section 2, $N^{\frac{1}{2}}(\theta^*_{1,N} - \theta^0)$ and $N(\theta^*_{1,N} - \theta^*_{2,N})$ are asymptotically independent. □

4. Main theorem. By making use of the results obtained in the previous section, we can prove the following.

THEOREM 2. *If the observations are generated by a Gaussian linear process satisfying the conditions* (A-1) *through* (A-5), *and further, if the maximum-likelihood* *estimate* $\hat{\theta}$ *and any efficient* θ^* *satisfy the conditions* (B-1) *through* (B-5), *then* $l_{\theta^*}(\theta^0) = \underline{\text{Var}}\{N(\theta^* - \hat{\theta})\}/V(\theta^0)^2 + l_{\hat{\theta}}(\theta^0)$.

PROOF. It follows from the definition of l_θ^*,

(21) $l_\theta^*(\theta^0)$

$$= \underline{E}\ \underline{\text{Var}}\left\{\frac{d\log L(\theta^0|X)}{d\theta}\,\Big|\,N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}$$

$$= \underline{E}\ \underline{\text{Var}}\left\{N(\hat{\theta} - \theta^*)/V(\theta^0) - N^{\frac{1}{2}}(\theta^* - \theta^0)\frac{d^2\log L(\theta^0|X)}{N^{\frac{1}{2}}\,d\theta^2}\,\Big|\,N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}$$

$$= \frac{\underline{E}\ \underline{\text{Var}}\{N(\hat{\theta} - \theta^*)|N^{\frac{1}{2}}(\theta^* - \theta^0)\}}{V(\theta^0)^2}$$

$$+ \underline{E}\ \underline{\text{Var}}\left\{N^{\frac{1}{2}}(\theta^* - \theta^0)\frac{d^2\log L(\theta^0|X)}{N^{\frac{1}{2}}\,d\theta^2}\,\Big|\,N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}$$

$$-2\,\underline{E}\ \underline{\text{Cov}}\left\{N(\hat{\theta} - \theta^*)/V(\theta^0), N^{\frac{1}{2}}(\theta^* - \theta^0)\frac{d^2\log L(\theta^0|X)}{N^{\frac{1}{2}}\,d\theta^2}\,\Big|\,N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}.$$

The first term in the last expression is equal to $\underline{\text{Var}}\{N(\hat{\theta} - \theta^*)\}/V(\theta^0)^2$ since, by Lemma 5, $N(\hat{\theta} - \theta^*)$ and $N^{\frac{1}{2}}(\theta^* - \theta^0)$ are asymptotically independent. The second is nothing but $l_{\hat{\theta}}(\theta^0)$, the information loss of the maximum-likelihood estimate $\hat{\theta}$, because $|N^{\frac{1}{2}}(\theta^* - \theta^0) - N^{\frac{1}{2}}(\hat{\theta} - \theta^0)| = |N^{\frac{1}{2}}(\theta^* - \hat{\theta})| \to 0$ in probability and so

$$\underline{E}\ \underline{\text{Var}}\left\{N^{\frac{1}{2}}(\theta^* - \theta^0)\frac{d^2\log L(\theta^0|X)}{N^{\frac{1}{2}}\,d\theta^2}\,\Big|\,N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}$$

$$= \underline{E}\ \underline{\text{Var}}\left\{N^{\frac{1}{2}}(\hat{\theta} - \theta^0)\frac{d^2\log L(\theta^0|X)}{N^{\frac{1}{2}}\,d\theta^2}\,\Big|\,N^{\frac{1}{2}}(\hat{\theta} - \theta^0)\right\} = l_{\hat{\theta}}(\theta^0).$$

Lastly the remaining problem is to show that the covariance above converges to 0. This can be proved as follows. (In the following, $L(\theta^0|X)$ is denoted as L.)

$$\underline{E}\left[\underline{\text{Cov}}\left\{N(\hat{\theta} - \theta^*), N^{\frac{1}{2}}(\theta^* - \theta^0)\frac{d^2\log L}{N^{\frac{1}{2}}\,d\theta^2}\,\Big|\,N^{\frac{1}{2}}(\theta^* - \theta^0)\right\}\right]$$

$$= \underline{E}\left[N^{\frac{1}{2}}(\theta^* - \theta^0)\underline{E}\left\{(N(\hat{\theta} - \theta^*) - EN(\hat{\theta} - \theta^*))\right.\right.$$

$$\left.\left.\times\left(\frac{d^2\log L}{N^{\frac{1}{2}}\,d\theta^2} - \underline{E}\left(\frac{d^2\log L}{N^{\frac{1}{2}}\,d\theta^2}\,\Big|\,N^{\frac{1}{2}}(\theta^* - \theta^0)\right)\right)\right|N^{\frac{1}{2}}(\theta^* - \theta^0)\right]$$

$$= \underline{E}\Bigg[N^{\frac{1}{2}}(\theta^* - \theta^0) \underline{E}\Bigg\{ \big(N(\hat{\theta} - \theta^*) - \underline{E}N(\hat{\theta} - \theta^*)\big) \bigg(\frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2} - \underline{E}\bigg(\frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2} \bigg)$$

$$- N^{\frac{1}{2}}(\theta^* - \theta^0)\, \underline{\mathrm{Cov}} \bigg\{ \frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2}, N^{\frac{1}{2}}(\theta^* - \theta^0) \bigg\}$$

$$\Big/ \underline{\mathrm{Var}}\, \{ N^{\frac{1}{2}}(\theta^* - \theta^0) \} \bigg) \bigg| N^{\frac{1}{2}}(\theta^* - \theta^0) \Bigg\} \Bigg]$$

$$= \underline{E}\Bigg[N^{\frac{1}{2}}(\theta^* - \theta^0) \{ N(\hat{\theta} - \theta^*) - \underline{E}N(\hat{\theta} - \theta^*) \} \bigg\{ \frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2} - \underline{E}\bigg(\frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2} \bigg) \bigg\} \Bigg]$$

$$- \underline{\mathrm{Cov}} \bigg\{ \frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2}, N^{\frac{1}{2}}(\theta^* - \theta^0) \bigg\} \Big/ \underline{\mathrm{Var}}\, \{ N^{\frac{1}{2}}(\theta^* - \theta^0) \}$$

$$\times \underline{E}\Big\{ \big(N^{\frac{1}{2}}(\theta^* - \theta^0) \big)^2 \big(N(\hat{\theta} - \theta^*) - \underline{E}N(\hat{\theta} - \theta^*) \big) \Big\}$$

where the second term in the last expression vanishes because of the asymptotic independence between $N^{\frac{1}{2}}(\theta^* - \theta^0)$ and $N(\hat{\theta} - \theta^*)$. In order to consider the first term, let $\underline{\mathrm{Cov}}\{d^2 \log L / N^{\frac{1}{2}}\, d\theta^2, T_i\} = \lambda_i$. Then

$$\underline{E}\Bigg[N^{\frac{1}{2}}(\theta^* - \theta^0) \{ N(\hat{\theta} - \theta^*) - \underline{E}N(\hat{\theta} - \theta^*) \} \bigg\{ \frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2} - \underline{E}\bigg(\frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2} \bigg) \bigg\} \Bigg]$$

$$= \underline{E}\Bigg[(\Sigma_i \alpha_i^* T_i)(\Sigma_j \Sigma_k (\beta_{jk} - \beta_{jk}^*) T_j T_k) \bigg\{ \frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2} - \underline{E}\bigg(\frac{d^2 \log L}{N^{\frac{1}{2}}\, d\theta^2} \bigg) \bigg\} \Bigg]$$

$$- (\Sigma_j \Sigma_k (\beta_{jk} - \beta_{jk}^*) \sigma_{jk})(\Sigma_i \alpha_i^* \lambda_i)$$

$$= \Sigma_i \Sigma_j \Sigma_k \alpha_i^* (\beta_{jk} - \beta_{jk}^*) \{ \sigma_{i,j} \lambda_k + \sigma_{i,k} \lambda_j + \sigma_{jk} \lambda_i \}$$

$$- (\Sigma_j \Sigma_k (\beta_{jk} - \beta_{jk}^*) \sigma_{jk})(\Sigma_i \alpha_i^* \lambda_i)$$

$$= \Sigma_i \Sigma_j \Sigma_k \alpha_i^* (\beta_{jk} - \beta_{jk}^*) \{ \sigma_{i,j} \lambda_k + \sigma_{i,k} \lambda_j \}$$

$$= \Sigma_k \lambda_k (\Sigma_i \Sigma_j \alpha_i^* \sigma_{i,j} (\beta_{j,k} - \beta_{j,k}^*)) + \Sigma_j \lambda_j (\Sigma_i \Sigma_k \alpha_i^* \sigma_{i,k} (\beta_{j,k} - \beta_{j,k}^*)) = 0,$$

by (20). Then, from (21),

$$l_{\theta^*}(\theta^0) = \underline{\mathrm{Var}}\, N(\theta^* - \hat{\theta}) / V(\theta^0)^2 + l_{\hat{\theta}}(\theta^0). \qquad \square$$

Acknowledgment. The author would like to thank Professors B. H. Margolin, E. J. Hannan, E. Parzen and K. Takeuchi for the comments. Thanks are also due to the referees and the associate editor, especially to one of the referees who has gone to an extraordinary amount of work in improving the earlier version of the paper.

REFERENCES

AITKEN, A. C. (1950). Statistical independence of quadratic forms in normal variates. *Biometrika* **37** 93–96.

ANDERSON, T. W. (1975). Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure. *Ann. Statist.* **3** 1283–1304.

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

CRAIG, A. T. (1943). Note on the independence of certain quadratic forms, *Ann. Math. Statist.* **14** 195–197.

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368.

FISHER, R. A. (1925). Theory of statistical estimation. *Proc. Cambridge Philos. Soc.* **22** 700–725.

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

HEWITT, E. and STROMBERG, K. (1969). *Real and Abstract Analysis*. Springer-Verlag, New York.

HOSOYA, Y. (1974). Estimation problems on stationary time-series models. Ph.D. thesis, Yale University.

LOÈVE, M. (1963). *Probability Theory, 3rd ed.* D. Van Nostrand Co., Princeton, N.J.

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

RAO, C. R. (1962). Efficient estimates and optimum inference procedures in large samples. *J. Roy. Statist. Soc. B* **24** 46–72.

TAKEUCHI, K. (1965). Unpublished technical report, Tokyo University.

TRICOMI, F. G. (1957). *Integral Equations*. Interscience Publishers, New York.

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

WHITTLE, P. (1952). Some results in time series analysis. *Skand. Aktuarietidskr.* **Haft 1-2** 48–60.

FACULTY OF ECONOMICS
TOHOKU UNIVERSITY
KAWAUCHI, SENDAI 980
JAPAN