〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 13, No. 2, 1984

Rationale of the Maximum Entropy Probability Density

B.S. Choi*

ABSTRACT

If $\{X_t\}$ is a sequence of independent identically distributed normal random variables, then the conditional probability density of $X_1, X_2, \cdots, X_n$ given the first $p+1$ sample autocovariances converges to the maximum entropy probability density satisfying the corresponding covariance constraints as the length of the sample sequence tends to infinity. This establishes that the maximum entropy probability density and the associated Gaussian autoregressive process arise naturally as the answers of conditional limit problems.

1. Introduction.

Jaynes (1957, 1958, 1968, 1978, 1982) introduced the concept of the maximum entropy of statistical mechanics into statistics, and proposed the Principle of Maximum Entropy as follows. When we choose a p.d.f. under some constraints, we select one which has the maximum entropy among p.d.f.'s satisfying the constraints. In other words, the p.d.f. which maximizes the entropy is identical to one which can be realized in the greatest number of ways (Jaynes [1968, p. 231]). Jaynes' principle has been primarily used to estimate spectral densities in time series analysis. For a brief review we assume that a sequence of random variables $\{X_t, t=0, \pm1, \pm2, \cdots\}$ is a real-valued, regular, stationary stochastic process which is measured at unit intervals of time. Without loss of generality, we assume the process to have zero-mean. We define the autocovariance function $\sigma(k)=E(X_t X_{t+k})$ for $k=0, \pm1, \pm2, \cdots$. Since the process is real valued,

*Department of Applied Statistics, Yonsei University, Seoul 120, Korea

Key Words: Maximum Entropy, Jaynes' Principle, Spectral Density, Multivariate Normal, Autoregressive Process, Kullback-Leibler Information, Diagnostic Checking, Model Modification.

$\sigma(-k)=\sigma(k)$ for all k. The Fourier transform of the autocovariance function is defined as the spectral density $S_x(\lambda)$:

$$S_x(\lambda)=\frac{1}{2\pi}\sum_{l=-\infty}^{\infty}\sigma(l)e^{-i\lambda l}.$$

If we know the whole autcovariance function, we obtain the spectral density through the definition. However, in practice, we usually do not know the entire autocovariance function $\sigma(l)$, $l=0, 1, 2, \cdots$. If we have only a finite number of reliable autocovariances, $\sigma(k)=\alpha_k$ for $|k|\leq p$, we may be interested in procedures that estimate the spectral density in a "natural" manner using only this partial information. The traditional methods of estimating the spectral densities are to assume that $\sigma(k)=0$ for $|k|>p$, to give weights $w(k)$ to $\sigma(k)$ for $|k|\leq p$, and to take the Fourier transform of $w(k)\sigma(k)$. We call the weight function $w(k)$ the lag window. However, using a window is a violation of a basic rule called the Principle of Data Reduction by Ables (1974), which says, "The result of any transformation imposed on the experimental data shall incorporate and be consistent with all relevant data and be maximally noncommittal with regard to unavailable data." The traditional methods in spectral analysis violate this principle in two ways: the unavailable data are assumed and the available data are distorted by windows. Thus, some scientists turned their attention to estimating spectral densities in other ways. One of new methods, which was proposed by Burg(1967, 1968), is based on the principle of maximum entropy as follows. Let $\{X_t\}$ be a stationary (Gaussian) time series with the first $p+1$ autocovariances $\sigma(0)=\alpha_0$, $\sigma(1)=\alpha_1$, $\cdots$, $\sigma(p)=\alpha_p$. Then the spectral density that maximizes the entropy (rate) of the process is the same as that of an autoregressive process of order p with the same first $p+1$ autocovariances α_0, α_1, $\cdots$, α_p. Since there are an infinite number of spectral densities which satisfy the partial information $\sigma(k)=\alpha_k$ for $|k|\leq p$, there must be an additional criterion to choose one spectral density which is the representative of the class of all possible spectral densities. Burg used it as the criterion that the maximum entropy estimate is the most random, the most unpredictable among all possible estimates which satisfy the given constraints. Ables (1974) has asserted that Burg's method satisfies the principle of data reduction, and so it is theoretically more reasonable than any of the traditional methods. Also, Lacoss(1971) and Kaveh et. al. (1976) have shown that the maximum entropy method gives a much higher resolution spectral estimate than any of the traditional methods, particularly, when the data record is of short duration.

Van Campenhout and Cover(1981) have presented a concrete meaning to the maximum entropy p.d.f. by proposing a relation between maximum entropy p.d.f.'s and conditional ones. The relation is that the conditional p.d.f. of X_1 given $\frac{1}{n}\sum_{i=1}^{n} X_i^2=\sigma^2$, where X_1, X_2, $\cdots$, X_n are i.i.d. (independent identically distributed) random variables with p.d.f. $g(x)$ satisfying certain regularity conditions, converges to a p.d.f. of the form $f_\lambda(x)=c(\lambda)\exp(-\lambda x^2)\,g(x)$, where λ and $c(\lambda)$ are determined by the equations $\int f_\lambda(x)dx=1$ and $\int x^2 f_\lambda(x)\,dx=\sigma^2$, as $n\to\infty$. Since a random variable X having the maximum entropy subject to $E(X^2)=1/(2\lambda)$ is the normal random variable with p.d.f. $(\lambda/\pi)^{1/2}\exp(-\lambda x^2)$ (see Kagan et. al. [1973, p. 410]), the limiting p.d.f. $f_\lambda(x)$ is the normalized product of a maximum entropy p.d.f. and the initial p.d.f. $g(x)$. Also, it is known that the p.d.f. $f_\lambda(x)$ minimizes the Kullback-Leibler(1951) information number of f relative to g,

$$D(f\,;g)=\int f(x)\ln\left\{\frac{f(x)}{g(x)}\right\}\,dx,$$

among the p.d.f.'s satisfying $\int x^2 f(x)dx=\sigma^2$, (Vincze [1972]). The Kullback-Leibler number is regarded as a good measure how much the p.d.f. $g(x)$ is different from the p.d.f. $f(x)$ in the sense of statistical distinguishability. This idea is backed up by Stein's lemma (see, e.g., Csiszár and Körner [1981, p. 28]) and Sanov's theorem (1957). Thus the limiting p.d.f $f_\lambda(x)$ is the closest p.d.f., in the Kullback-Leibler sense, to the initial p.d.f. $g(x)$ among the p.d.f.'s $f(x)$ satisfying $\int x^2 f(x)dx=\sigma^2$. This kind of convergence problem has been also studied by Darwin and Fowler (1922), Bartfai (1972), Lanford (1973), Tjur (1974, pp. 306-321), Zabell (1974), and Vasicek (1980).

It may be an interesting question what kind of the initial p.d.f. $g(x)$ results in the maximum entropy p.d.f. itself as the limiting p.d.f. $f_\lambda(x)$, in other words, which is the nearest p.d.f., in the Kullback-Leibler sense, to the maximum entropy p.d.f. subject to $E(X^2)=\sigma^2$, i.e., the normal p.d.f. with mean 0 and variance σ^2. From the previous lemma, we know that there are two candidates for the initial p.d.f. $g(x)$. One is that the initial p.d.f. $g(x)$ itself is a normal p.d.f.. Then, the limiting p.d.f. $f_\lambda(x)$ is the maximum entropy p.d.f.. However, this case does not arouse any interest to statisticians. The other is that $g(x)$ is uniform over $(-\infty,\infty)$. Then the limiting functinn $f_\lambda(x)$ does possibly become the maximum entropy p.d.f.. Unfortunately, the uniform function over $(-\infty,\infty)$ is improper and does not satisfy the regularity conditions. To overcome these deficiencies, we may define a double array of uniform random variables (Choi [1984]). More precisely, for each $n\geq 1$, let $X_{1,n}$, $X_{2,n}$, $\cdots$, $X_{n,n}$ be a sequence of i.i.d.

uniform random variables with p.d.f. $g_n(x)=\frac{1}{2\sigma\sqrt{n}}I_{(-\sigma\sqrt{n},\sigma\sqrt{n})}(x)$, where $\sigma>0$. Then the conditional p.d.f. of $X_{1,n}$ given $\frac{1}{n}\sum_{i=1}^{n}X_{i}^{2},_{n}=\sigma^2$ tends to the maximum entropy p.d.f. satisfying $E(X^2)=\sigma^2$ as $n\to\infty$.

Choi (1983) has generalized this result to a multivariate case as follows. The conditional p.d.f.'s of the random vectors $(X_{1,n}, X_{2,n}\cdots, X_{m,n})$ given sample autocovariances $\frac{1}{n}\sum_{l=1}^{n}X_{l,n}X_{l+j,n}=\alpha_j$ for $0\leq j\leq p$, where $\{X_{l,n}\,;\,1\leq l\leq n\}$ is the sequence of i.i.d. uniform random variables with p.d.f. $g_n(x)$, tend to the p.d.f. of the maximum entropy random vector subject to the corresponding covariance constraints $E(X_iX_j)=\alpha_{|i-j|}$ for $|i-j|\leq p$, which is the random vector whose elements constitute a Gaussian autoregressive (AR) process (Choi and Cover [1984]).

Based on these two results we can conjecture that the conditional p.d.f. of the random vector $(X_1, X_2, \cdots, X_m)$ given the sample covariances $\frac{1}{n}\sum_{l=1}^{n}X_lX_{l+j}=\alpha_j$ for $0\leq j\leq p$, where $\{X_t\,;\,t=0,\pm1,\pm2,\cdots\}$ is the sequence of i.i.d. normal random variables with mean 0 and variance σ^2, tends to the p.d.f. of the maximum entropy random vector subject to the covariance constraints $E(X_iX_j)=\alpha_{|i-j|}$ for $|i-j|\leq p$. This conjecture is backed up by the lemma that the p.d.f. of the maximal entropy random vector satisfying the conditions $E(X_iX_j)=\alpha_{|i-j|}$ for $|i-j|\leq p$ is closest in the Kullback-Leibler sense to the joint p.d.f. of the i.i.d. normal random variables among the p.d.f.'s satisfying $E(X_iX_j)=\alpha_{|i-j|}$ for $|i-j|\leq p$. (Its proof is in Appendix.)

In Section 2 some basic notions and lemmas coming in useful to prove the conjecture will be presented. In Section 3 the conjecture will prove true. It will give a concrete rationale to the maximum entropy probability density and Burg's maximum entropy spectral density. In Section 4 we will finish this paper by proposing a possible use of the main theorem to the model modification in time series analysis.

2. Preliminaries.

We assume that $\{X_1, X_2, \cdots, X_n\}$ is a sequence of random variables with mean 0 and covariances $\mathrm{Cov}(X_i, X_j)=\sigma(i-j)$ for $1\leq i, j\leq n$. If we denote the covariance matrix of the random vector $\underset{\sim}{X}_n=(X_1, X_2, \cdots, X_n)'$ by

$$\Sigma_n=\begin{bmatrix} \sigma(0) & \sigma(1) & \sigma(2)\cdots\cdots\sigma(n-1) \\ \sigma(-1) & \sigma(0) & \sigma(1)\cdots\cdots\sigma(n-2) \\ \sigma(-2) & \sigma(-1) & \sigma(0)\cdots\cdots\sigma(n-3) \\ \vdots & \vdots & \vdots \quad\quad \vdots \\ \sigma(1-n) & \sigma(2-n) & \sigma(3-n)\cdots\sigma(0) \end{bmatrix},$$

then $\sigma(j)=\sigma(-j)$ for all j and Σ_n is positive definite (see, e.g., Box and Jenkins [1970, p. 28]).

Using Hawkins-Simon's lemma that a matrix is positive definite if and only if all the leading principal minors are positive (see, e.g., Gantmacher [1959, p. 306]), we construct an autoregressive process from given autocovariance terms.

Lemma 2.1. If the random vector $\underset{\sim}{X}_n$ has the first $p+1$ known covariances $\sigma(0)=\alpha_0$, $\sigma(1)=\alpha_1$, $\cdots$, $\sigma(p-1)=\alpha_{p-1}$, $\sigma(p)=\alpha_p$, and if $\alpha_j=\alpha_{-j}$ for all j, then the followings are satisfied.

(i) The system of simultaneous equations $\sum_{j=1}^{p} a_j\alpha_{l-j}=-\alpha_l$ for $l=1, 2, \cdots, p-1, p$ has a unique solution $a_1, a_2, \cdots, a_p$.

(ii) If $\alpha_{p+1}, \alpha_{p+2}, \alpha_{p+3}, \cdots$ are extrapolated from $\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_p$ according to the Yule-Walker equations : $\alpha_l=-\sum_{j=1}^{p} a_j\alpha_{l-j}$ for $l=p+1, p+2, \cdots, n-1$, and if a Toeplitz matrix A_n and a positive number σ^2 are defined by

$$A_n=\begin{bmatrix} \alpha_0 & \alpha_1 & \alpha_2\cdots\cdots\alpha_{n-2} & \alpha_{n-1} \\ \alpha_{-1} & \alpha_0 & \alpha_1\cdots\cdots\alpha_{n-3} & \alpha_{n-2} \\ \alpha_{-2} & \alpha_{-1} & \alpha_0\cdots\cdots\alpha_{n-4} & \alpha_{n-3} \\ \vdots & \vdots & \vdots \quad\quad \vdots & \vdots \\ \alpha_{2-n} & \alpha_{3-n} & \alpha_{4-n}\cdots\cdots\alpha_0 & \alpha_1 \\ \alpha_{1-n} & \alpha_{2-n} & \alpha_{3-n}\cdots\cdots\alpha_{-1} & \alpha_0 \end{bmatrix}$$

and $\quad \sigma^2=\sum_{j=0}^{p} a_j\alpha_{-j}$, where $a_0=1$,

then $|A_n|=(\sigma^2)^{n-p}|A_p|$ for $n\geq p$ and A_n is positive definite for all n.

Proof. (i) Since Σ_p (i.e., A_p) is positive definite, there exists a unique set of solutions $\{a_1, a_2, \cdots, a_p\}$.

(ii) For $r\geq p$, we multiply the $(r-p+1)$-th row of $|A_{r+1}|$ by a_p, its $(r-p+2)$-th row by a_{p-1}, $\cdots$, its r-th row by a_1, and then add them to the last row. We then obtain a formula

$$|A_{r+1}|=|A_r|\times\left(\sum_{j=0}^{p} a_j\alpha_{-j}\right), \text{ where } a_0=1.$$

In case of $r=p$, $|A_{p+1}|$ is positive and so is $|A_p|$ by Hawkins' lemma. Consequently,

$\sum_{j=0}^{p} a_j \alpha_{-j}$ is positive. If we define $\sigma^2=\sum_{j=0}^{p} a_j \alpha_{-j}$, then $|A_n|=(\sigma^2)^{n-p}|A_p|$ for $n\geq p$. This relation implies that $|A_r|$ is positive for all r, i.e., A_n is positive definite. Q.E.D.

Henceforth, we regard $\alpha_0, \alpha_1, \cdots, \alpha_p$ as given numbers and $\alpha_{p+1}, \alpha_{p+2}, \cdots$ as numbers extrapolated from $\alpha_0, \alpha_1, \cdots, \alpha_p$ by the Yule-Walker equations. From the autocovariance function $\{\alpha_k ; k=0, \pm1, \pm2, \cdots\}$ we can construct a stationary autoregressive process as follows.

Lemma 2.2. Let $\cdots, X_{-n}, X_{1-n}, \cdots, X_{-1}, X_0, X_1, \cdots, X_n, \cdots$, be a stationary normal stochastic process with mean 0 and autocovariance function $\sigma(k)=\alpha_k$ for all k. If we define a sequence $\cdots, V_{-n}, V_{1-n}, \cdots, V_{-1}, V_0, V_1, \cdots, V_n, \cdots$ such that

$$V_t=X_t+a_1 X_{t-1}+a_2 X_{t-2}+\cdots+a_p X_{t-p}, \quad t=0, \pm1, \pm2, \cdots, \tag{2.1}$$

then $\{V_t\}$ is the sequence of i.i.d. normal random variables with mean 0 and variance σ^2, and X_t is independent of V_{t+l} for $l=1, 2, \cdots$. Moreover, all roots of the polynomial equation

$$\sum_{j=0}^{p} a_j z^{p-j}=0 \tag{2.2}$$

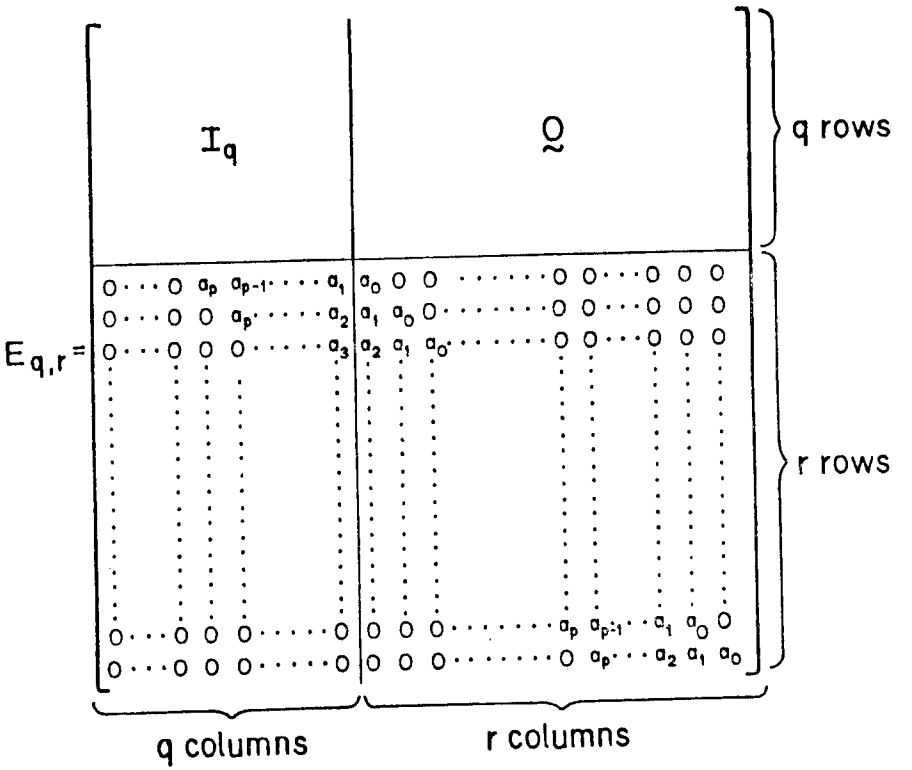
are less than 1 in absolute value.

Proof. The first part of this lemma can be proved in many ways, e.g., by use of the Wold decomposition theorem, or by use of the Toeplitz matrix theory (Grandell et. al. [1980]), or spectrum techniques (Anderson [1971, p. 407]). We shall prove this lemma by the linear transformation method, which will be useful in proving subsequent results.

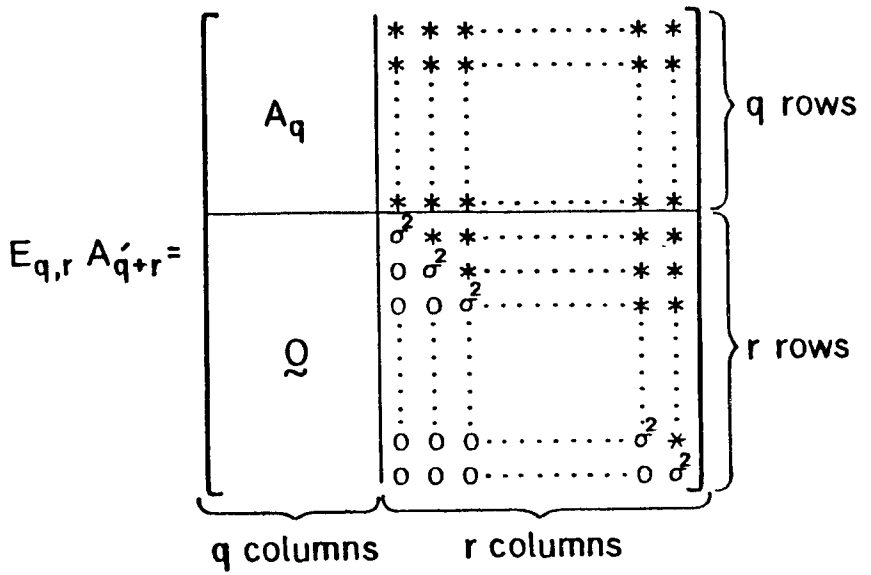
For any r, s, $q(\geq p)$, (2.1) yields that

$$\begin{bmatrix} X_{s+1} \\ X_{s+2} \\ \vdots \\ X_{s+q} \\ V_{s+q+1} \\ V_{s+q+2} \\ \vdots \\ V_{s+q+r-1} \\ V_{s+q+r} \end{bmatrix} = E_{q,r} \begin{bmatrix} X_{s+1} \\ X_{s+2} \\ \vdots \\ X_{s+q} \\ X_{s+q+1} \\ X_{s+q+2} \\ \vdots \\ X_{s+q+r-1} \\ X_{s+q+r} \end{bmatrix},$$

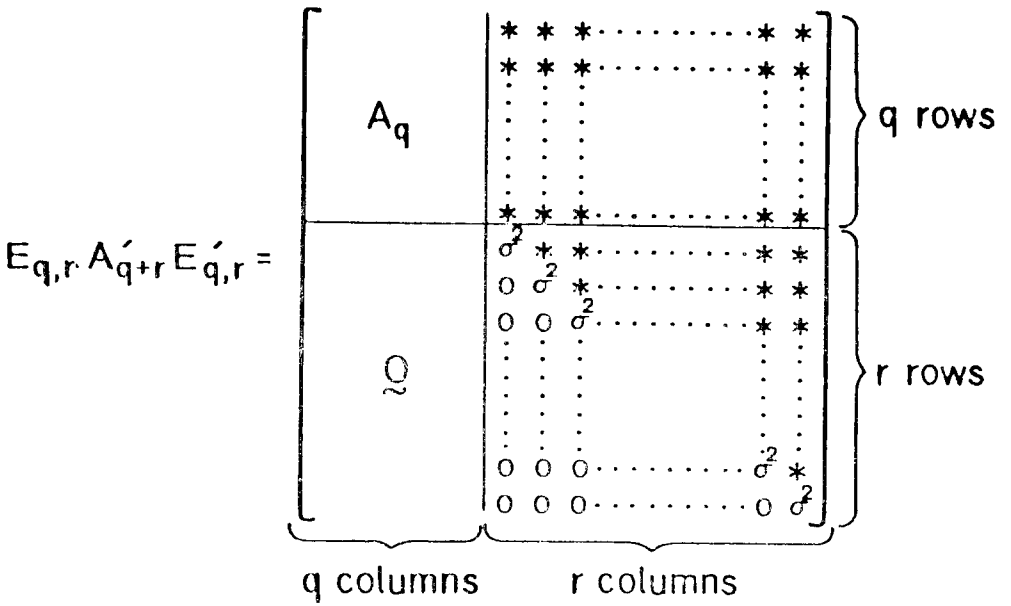
where $E_{q,r}$ is



To calculate the covariance matrix $E_{q,r} A_{q+r} E_{q,r}'$, we first derive the lower triangular part of $E_{q,r} A_{q+r}'$:



where it is not necessary to calculate the values of $*$. Then, we obtain the lower triangular part of $E_{q,r} A_{q+r}' E_{q,r}'$ as



The symmetry of $E_{q,r} A_{q+r} E_{q,r}'$ follows that of A_{q+r}. Thus,

$$E_{q,r} A_{q+r} E_{q,r}' = \begin{bmatrix} A_q & \underset{\sim}{0} \\ \underset{\sim}{0} & \sigma^2 I_r \end{bmatrix}.$$

In summary, the random vector $(X_{s+1}, X_{s+2}, \cdots, X_{s+q}, V_{s+q+1}, V_{s+q+2}, \cdots, V_{s+q+r})'$ has the multivariate normal distribution with mean 0 and covariance matrix

$$\begin{bmatrix} A_q & \underset{\sim}{0} \\ \underset{\sim}{0} & \sigma^2 I_r \end{bmatrix} \qquad \text{for any } s,\ r,\ q(\geq p).$$

Thus, $\{V_t\}$ is the sequence of i.i.d. normal random variables with mean 0 and variance σ^2, and X_t is independent of V_{t+l} for $l=1, 2, \cdots$. It is known that a stochastic process $\{X_t\}$ that satisfies the stochastic difference equation (2.1), where $\{V_t\}$ is a sequence of i.i.d. random variables and X_t is independent of $V_{t+1}, V_{t+2}, \cdots$, is stationary if and only if all the roots of the polynomial equation (2.2) are less than 1 in absolute value (Anderson [1971, p. 256 and pp. 166-173]). Q.E.D.

The next lemma is about the asymptotic normality of the sample autocovariances. The proof is based on Anderson and Rubin (1950).

Lemma 2.3. Let $\{X_t\}$ be a regular stationary Gaussian stochastic process with mean 0 and autocovariance functions $\sigma(k)=\alpha_k$ for all k. If we define $R_n(j)=\frac{1}{n}\sum_{l=1}^{n-j} X_l X_{l+j}$, $j=0, 1, \cdots, p$, then the random vector

$$(\sqrt{n}[R_n(0)-\alpha_0],\ \sqrt{n}[R_n(1)-\alpha_1],\ \cdots,\ \sqrt{n}[R_n(p)-\alpha_p])$$

has a limiting normal distribution with mean $\underset{\sim}{0}$ and covariance matrix $\Omega=(\omega_{gh})$, where

$$\omega_{gh}=\frac{\sigma^4}{\pi}\int_{-\pi}^{\pi}\cos\lambda g\ \cos\lambda h\ |\sum_{j=0}^{p} a_j \exp(-i\lambda j)|^{-4}\, d\lambda.$$

Proof. First, we cite Anderson and Rubin's theorem (1950, also refer to Theorem 8.4.2 of Anderson's book [1971, p. 478]) :

"Let $\{Y_t\}$ be a stationary time series with mean 0, autocovariance function $\sigma(k)$ and spectral density $S_y(\lambda)$. Let $R_n(j)$ be $\frac{1}{n}\sum_{l=1}^{n-j} Y_l Y_{l+j}$ for $j=0, 1, 2, \cdots, p$. If Y_t can be written as $Y_t=\sum_{s=-\infty}^{\infty}\gamma_s V_{t-s}$, where $\{V_t\}$ consists of i.i.d. random variables with $E(V_t)=0$, $E(V_t^2)=\sigma^2$ and $E(V_t^4)=3\sigma^4+\kappa_4<\infty$, and $\sum_{t=-\infty}^{\infty}|\gamma_t|<\infty$, then $\sqrt{n}(R_n(0)-\sigma(0))$, $\sqrt{n}(R_n(1)-\sigma(1))$, $\cdots$, $\sqrt{n}(R_n(p)-\sigma(p))$ have a limiting normal distribution with means 0 and covariances

$$\lim_{n\to\infty}\mathrm{Cov}(\sqrt{n}R_n(g),\ \sqrt{n}R_n(h))=4\pi\int_{-\pi}^{\pi}\cos\lambda g\cos\lambda h\ S_y^2(\lambda)\,d\lambda+\frac{\kappa_4}{\sigma^4}\sigma(h)\sigma(g).$$"

We know that all the roots of Equation (2.2) are less than 1 in absolute value by Lemma 2.2. Then X_t can be represented by

$$X_t=\sum_{l=0}^{\infty}\gamma_l V_{t-l}, \tag{2.3}$$

where the sequence $\{\gamma_l\}$ satisfies that

$$\begin{aligned}
1&=\gamma_0\\
0&=a_0\gamma_1+a_1\gamma_0\\
0&=a_0\gamma_2+a_1\gamma_1+a_2\gamma_0\\
&\vdots\\
0&=a_0\gamma_{p-1}+a_1\gamma_{p-2}+a_2\gamma_{p-3}+\cdots+a_{p-1}\gamma_0
\end{aligned} \tag{2.4}$$

and

$$0=a_0\gamma_t+a_1\gamma_{t-1}+a_2\gamma_{t-2}+\cdots+a_{p-1}\gamma_{t-p+1}+a_p\gamma_{t-p}, \qquad t\geqslant p. \tag{2.5}$$

If we let $n_1, n_2, \cdots, n_r$ be the roots of Equation (2.2) with multiplicities $m_1, m_2, \cdots, m_r$ where $r\leq p$, then the difference equation (2.5) has the solution γ_t :

$$\gamma_t=\sum_{j=1}^{r}(b_1^{(j)}+b_2^{(j)}t+\cdots+b_{m_j}^{(j)}t^{j-1})\eta_j^t,$$

where the coefficients $b_l^{(j)}$ for $l=1, 2, \cdots, m_j$ and $j=1, 2, \cdots, r$ are constants which satisfy the initial conditions (2.4), (see, e.g., Choi [1973]). Therefore, $\sum_{t=0}^{\infty}|\gamma_t|<\infty$, since $|\eta_j|<1$ for $1\leq j\leq r$. Since the sequence $\{X_t\}$ of (2.3) satisfies the conditions of Anderson and Rubin's theorem and $\kappa_4=0$ by the normality of $\{V_t\}$, and since Lemma 2.2 implies that the spectral density of the process $\{X_t\}$ is $S_X(\lambda)=\frac{\sigma^2}{2\pi}|\sum_{j=0}^{p}a_j\exp(-i\lambda j)|^{-2}$, $\sqrt{n}[R_n(0)-\alpha_0]$, $\sqrt{n}[R_n(1)-\alpha_1]$, $\cdots$, $\sqrt{n}[R_n(p)-\alpha_p]$ have a limiting normal distribution with means 0 and covariances $\omega_{g,h}$ for $0\leq g,\ h\leq p$. Q.E.D.

We define the following vectors: $\underset{\sim}{x}_n=(x_1, x_2, \cdots, x_n)'$, $\underset{\sim}{x}_{-n}=(x_{1-n}, x_{2-n}, \cdots, x_{-1}, x_0)'$, $\underset{\sim}{x}^{(2)}{}_n=(x_{m+1}, x_{m+2}, \cdots, x_n)'$, $\underset{\sim}{x}^e{}_n=\underset{\sim}{x}_{n+p}$, $\underset{\sim}{x}^e{}_{-n}=\underset{\sim}{x}_{-n-p}$, $\underset{\sim}{x}_{n_1,n_2}=\begin{bmatrix}\underset{\sim}{x}_{-n_1}\\ \underset{\sim}{x}_{n_2}\end{bmatrix}$, $\underset{\sim}{x}^e{}_{n_1,n_2}=\begin{bmatrix}\underset{\sim}{x}_{-n_1-p}\\ \underset{\sim}{x}_{n_2+p}\end{bmatrix}$ and $\underset{\sim}{x}^{(2)}{}_{n_1,n_2}=\begin{bmatrix}\underset{\sim}{x}_{-n_1}\\ \underset{\sim}{x}^{(2)}{}_n\end{bmatrix}$. Also, we define the sample autocovariance function by

$$R_{n_1,n_2}(j)=\frac{1}{n_1-n_2}\sum_{l=1-n_1}^{n_2-j}X_l X_{l+j} \text{ for } 0\leq j\leq n_1+n_2-1,$$

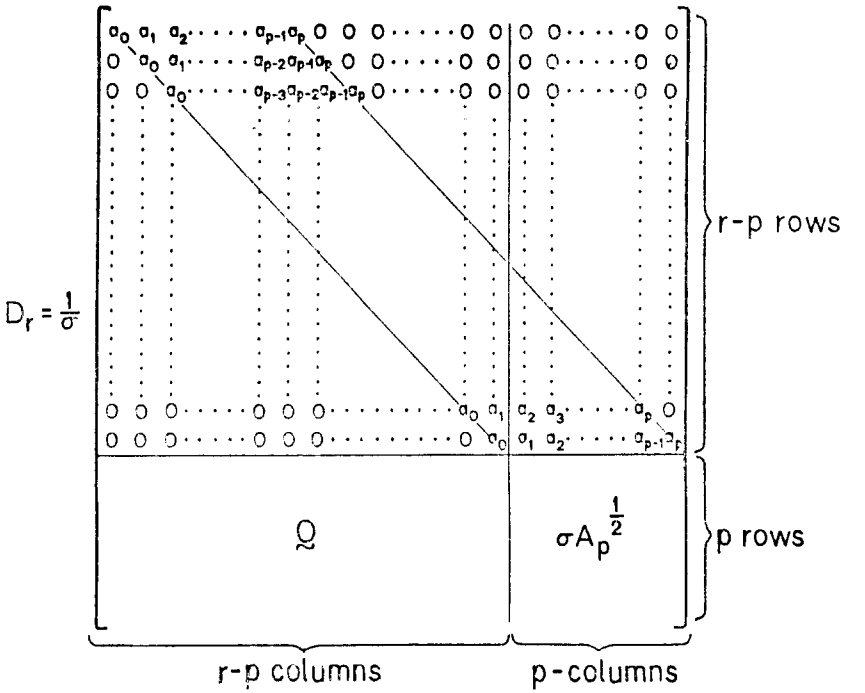
and $R_{n_1,n_2}(-j)=R_{n_1,n_2}(j)$ for all j. The next lemma plays an important role in proving the main theorem.

Lemma 2.4. Define $G(n_1, n_2 ; h)$ by

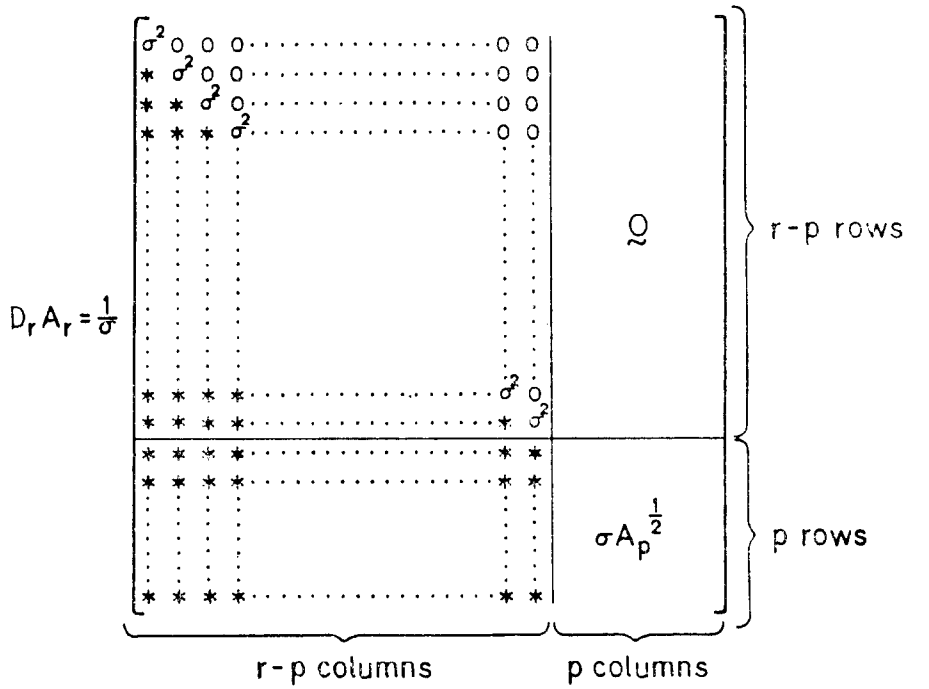
$$\left\{\underset{\sim}{x}_{n_1,n_2}\in R^{n_1+n_2}\ \middle|\ |R_{n_1,n_2}(j)-\alpha_j|<h \text{ for } j=0, 1, 2, \cdots, p\right\}.$$

If $x_{-n_1+1-p}=x_{-n_1+2-p}=\cdots=x_{-n_1-1}=x_{-n_1}=0$, if $x_{n_2+1}=x_{n_2+2}=\cdots=x_{n_2+p}=0$, and if $\underset{\sim}{x}_{n_1,n_2}\in G(n_1, n_2 ; h)$, then $|\underset{\sim}{x}^{e\prime}{}_{n_1,n_2}A_{n_1+n_2+2p}{}^{-1}\underset{\sim}{x}^e{}_{n_1,n_2}-(n_1+n_2)|<(n_1+n_2)Kh$, where K is a constant.

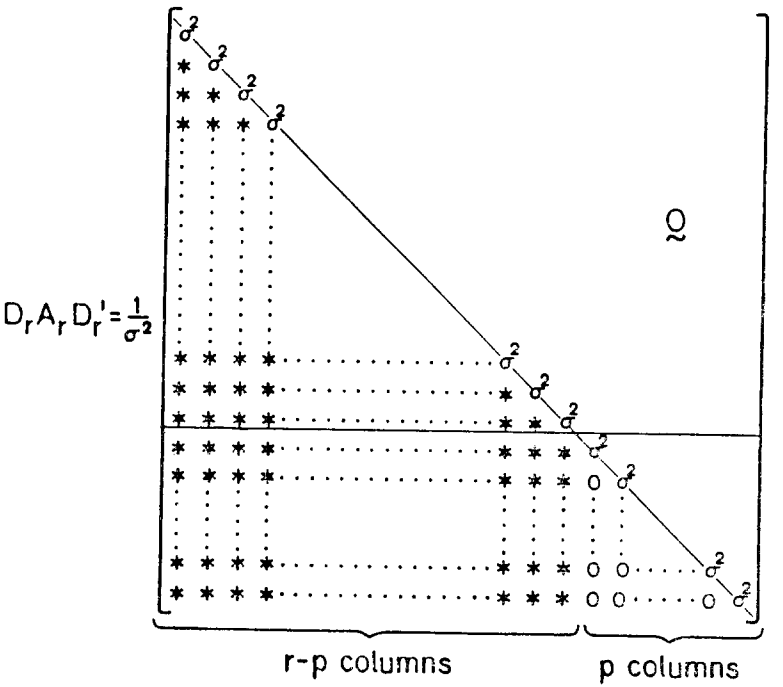
Proof. First, we will obtain $A_r{}^{-1}$ for $r\geq p$. Define the matrix D_r by



A direct product of matrices D_r and A_r yields



where the values of $*$ need not be calculated. By another direct product of matrices $D_r A_r$ and D_r', we obtain the upper triangular part of $D_r A_r D_r'$:



Since A_r is symmetric, so is $D_r A_r D_r'$. Thus, $D_r A_r D_r' = \mathrm{I}$ or $A_r^{-1} = D_r' D_r$.

Second, we calculate $D_{n_1+n_2+2p}\, \underset{\sim}{x}^e{}_{n_1,n_2}$:

$$D_{n_1+n_2},\ x^e{}_{n_1,n_2} = \begin{bmatrix} \frac{1}{\sigma}\sum_{j=0}^{p} a_j x_{j-n_1+1-p} \\ \frac{1}{\sigma}\sum_{j=0}^{p} a_j x_{j-n_1+2-p} \\ \vdots \\ \frac{1}{\sigma}\sum_{j=0}^{p} a_j x_{j+l} \\ \vdots \\ \frac{1}{\sigma}\sum_{j=0}^{p} a_j x_{n_2+j} \\ \hline \underset{\sim}{0} \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \end{matrix} \begin{matrix} n_1+n_2+p \text{ rows} \\ \\ \\ \\ \\ p \text{ rows} \end{matrix}$$

where $x_{-n_1} = x_{-n_1-1} = x_{-n_1-2} = \cdots = x_{-n_1+1-p} = 0$ and $x_{n_2+1} = x_{n_2+2} = \cdots = x_{n_2+p} = 0$.

Then,

$$\underset{\sim}{x}^{e\prime}{}_{n_1,n_2}\, A^{-1}{}_{n_1+n_2+2p}\, \underset{\sim}{x}^e{}_{n_1,n_2}$$
$$= (D_{n_1+n_2+2p}\, \underset{\sim}{x}^e{}_{n_1,n_2})'\ (D_{n_1+n_2+2p}\, \underset{\sim}{x}^e{}_{n_1,n_2})$$

$$= \sum_{l=n_1+1-p}^{n_2} \left(\frac{1}{\sigma} \sum_{j=0}^{p} a_j x_{j+l} \right)^2$$

$$= \frac{1}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j \left(\sum_{l=-n_1+1-p}^{n_2} x_{l+i} x_{l+j} \right)$$

$$= \frac{1}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j \sum_{l=1-n_1}^{n_2-|i-j|} x_l x_{l+|i-j|} + \frac{1}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j \sum_{l=-n_1+1-p+\min(i,j)}^{-n_1} x_l x_{l+|i-j|}$$

$$+ \frac{1}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j \sum_{l=n_2+1}^{n_2+\max(i,j)} x_{l-|i-j|} x_l$$

$$= \frac{1}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j \sum_{l=1-n_1}^{n_2-|i-j|} x_l x_{l+|i-j|}$$

$$= \frac{n_1+n_2}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j R_{n_1,n_2}(i-j).$$

The second to the last equality results from the zero assumptions. Moreover, the Yule-Walker equations yield that

$$\frac{1}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j \alpha_{i-j} = \frac{1}{\sigma^2} \sum_{i=0}^{p} a_i \left(\sum_{j=0}^{p} a_j \alpha_{i-j} \right) = \frac{1}{\sigma^2} \sum_{i=0}^{p} a_i \sigma^2 \delta_{i,0} = 1,$$

where the Kronecker delta function $\delta_{a,b}$ is defined as

$$\delta_{a,b} = \begin{cases} 1 & \text{if } a=b, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, if $\underset{\sim}{x}_{n_1,n_2} \in G(n_1, n_2 ; h)$, then

$$|\underset{\sim}{x}^{e'}{}_{n_1,n_2} A^{-1}{}_{n_1+n_2+2p} \underset{\sim}{x}^{e}{}_{n_1,n_2} - (n_1+n_2)|$$

$$= (n_1+n_2) \left| \frac{1}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j R_{n_1,n_2}(i-j) - \frac{1}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} a_i a_j \alpha_{i-j} \right|$$

$$\leq \frac{n_1+n_2}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} |a_i a_j \{R_{n_1,n_2}(i-j) - \alpha_{i-j}\}|$$

$$< \frac{n_1+n_2}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} |a_i a_j| h.$$

By letting $K = \frac{1}{\sigma^2} \sum_{i=0}^{p} \sum_{j=0}^{p} |a_i a_j|$, the proof is completed. Q.E.D.

We end this section by letting $\phi_l(\underset{\sim}{x} | \underset{\sim}{\mu}, \Sigma)$ be the p.d.f. of the l-variate normal random vector with mean $\underset{\sim}{\mu}$ and covariance matrix Σ.

3. The main theorem and its proof.

Theorem 3.1. Let $\{X_l ; l=0, \pm1, \pm2, \cdots\}$ be the sequence of i.i.d. normal random variables with common p.d.f. $\phi_1(x|0, \delta^2)$. If we assume that the observations x_{-n_1+1-p}, $x_{-n_1+2-p}, \cdots, x_{-n_1-1}, x_{-n_1}$ and $x_{n_2+1}, x_{n_2+2}, \cdots, x_{n_2+p}$ are zeros, then the conditional p.d.f. of the random vector $\underset{\sim}{X}_m=(X_1, X_2, \cdots, X_m)'$ given $R_{n_1,n_2}(0)=\alpha_0$, $R_{n_1,n_2}(1)=\alpha_1, \cdots, R_{n_1,n_2}(p)=\alpha_p$ tends to $\phi_m(\underset{\sim}{x}_m|\underset{\sim}{0}, A_m)$ as $n_1, n_2 \to \infty$.

Proof. If we let $\psi_0=\phi_{n_1+n_2+2p}(\underset{\sim}{x}^e{}_{n_1,n_2}|\underset{\sim}{0}, \delta^2 I)$, then, for any $\underset{\sim}{x}^e{}_{n_1,n_2} \in G(n_1, n_2 ; h)$,

$$K_0 \exp\{-h(n_1+n_2)/(2\delta^2)\} < \psi_0 < K_0 \exp\{h(n_1+n_2)/(2\delta^2)\},$$

where $K_0=(2\pi\delta^2)^{-(n_1+n_2+2p)/2}\exp\{-\alpha_0(n_1+n_2)/(2\delta^2)\}$. If ψ_1 is defined by $\phi_{n_1+n_2+2p}(\underset{\sim}{x}^e{}_{n_1,n_2}|\underset{\sim}{0}, A_{n_1+n_2+2p})$, then Lemma 2.4 implies that, for any $\underset{\sim}{x}^e{}_{n_1,n_2} \in G(n_1, n_2 ; h)$,

$$K_1 \exp\{-(n_1+n_2)\, hK/2\} < \psi_1 < K_1 \exp\{(n_1+n_2)hK/2\},$$

where $K_1=\{(2\pi)^{n_1+n_2+2p}|A_{n_1+n_2+2p}|\}^{-1/2}\exp\{-(n_1+n_2)/2\}$. Combining these inequalities, we obtain that, for any $\underset{\sim}{x}^e{}_{n_1,n_2} \in G(n_1, n_2 ; h)$,

$$C(-h, n_1+n_2)\, K_0/K_1 \le \psi_0/\psi_1 \le C(h, n_1+n_2)\, K_0/K_1,$$

where $C(h, n)=\exp\{hn(K+\delta^{-2})/2\}$. As a matter of convenience, we denote $G(n_1, n_2 ; h)$ by G unless there is confusion. Then these inequalities imply that

$$C(-h, n_1+n_2)K_0/K_1 \le \int_G \psi_0\, d\underset{\sim}{x}_{n_1,n_2} \Big/ \int_G \psi_1\, d\underset{\sim}{x}_{n_1,n_2} \le C(h, n_1+n_2)\, K_0/K_1.$$

A similar procedure yields that

$$C(-h, n_1+n_2)\, K_0/K_1 \le \int_G \psi_0\, d\underset{\sim}{x}^{(2)}{}_{n_1,n_2} \Big/ \int_G \psi_1\, d\underset{\sim}{x}^{(2)}{}_{n_1,n_2} \le C(h, n_1+n_2)\, K_0/K_1.$$

Combining the above inequalities implies that

$$C(-2h, n_1+n_2) \le \frac{\int_G \psi_0\, d\underset{\sim}{x}^{(2)}{}_{n_1,n_2}}{\int_G \psi_0\, d\underset{\sim}{x}_{n_1,n_2}} \cdot \frac{\int_G \psi_1\, d\underset{\sim}{x}_{n_1,n_2}}{\int_G \psi_1\, d\underset{\sim}{x}^{(2)}{}_{n_1,n_2}} \le C(2h, n_1+n_2). \tag{3.1}$$

The conditional p.d.f. of the random vector $\underset{\sim}{X}_m$ given $R_{n_1,n_2}(0)=\alpha_0$, $R_{n_1,n_2}(1)=\alpha_1, \cdots, R_{n_1,n_2}(p)=\alpha_p$ is

$$\lim_{h\to 0} f(\underset{\sim}{x}|G(n_1, n_2 ; h)) = \lim_{h\to 0} \int_G \psi_0\, d\underset{\sim}{x}^{(2)}{}_{n_1,n_2} \Big/ \int_G \psi_0\, d\underset{\sim}{x}_{n_1,n_2}$$

$$= \lim_{h\to 0} \int_G \psi_1\, d\underset{\sim}{x}^{(2)}{}_{n_1,n_2} \Big/ \int_G \psi_1\, d\underset{\sim}{x}_{n_1,n_2}, \tag{3.2}$$

where the last equality holds by (3.1). The R.H.S. of (3.2) is the conditional p.d.f. of $\underset{\sim}{X}_m=(X_1, X_2, \cdots, X_m)'$ given the conditions that $R_{n_1,n_2}(j)=\alpha_j$, $j=0, 1, \cdots, p$ where $\underset{\sim}{X}_{n_1,n_2}=(X_{1-n_1}, X_{2-n1}, \cdots, X_{n_2-1}, X_{n_2})$ is the normal random vector with mean $\underset{\sim}{0}$ and covariance matrix $A_{n_1+n_2}$ under the zero assumptions. We denote the conditional p.d.f. in the R.H.S. of (3.2) by $\phi(\underset{\sim}{x}_m|R_{n_1,n_2}(j)=\alpha_j,\ 0\leq j\leq p)$.

What is left to be shown is that the conditional p.d.f. $\phi(\underset{\sim}{x}_m|R_{n_1,n_2}(j)=\alpha_j,\ 0\leq j\leq p)$ tends to the unconditional p.d.f. $\phi_m(\underset{\sim}{x}_m|\underset{\sim}{0}, A_m)$. Since it is known that

$$\begin{aligned}&\phi(\underset{\sim}{x}_m|R_{n_1,n_2}(j)=\alpha_j,\ 0\leq j\leq p)=\lim_{h\to 0}\phi(\underset{\sim}{x}_m|G(n_1, n_2; h))\\&=\phi(\underset{\sim}{x}_m)\lim_{h\to 0}\Pr(G(n_1, n_2; h)|\underset{\sim}{x}_m)/\Pr(G(n_1, n_2; h)),\end{aligned}\tag{3.3}$$

where $\phi(\underset{\sim}{x}_m)$ is the marginal p.d.f. of $\underset{\sim}{X}_m$, it is necessary to show that the limit in the R.H.S. of (3.3) tends to one as $n_1\to\infty$ and $n_2\to\infty$. If we let $d(j; n_1, n_2)$ be

$$\frac{1}{\sqrt{n_1+n_2}}\{\sum_{l=1}^{j}X_{l-j}(x_l-X_l)+\sum_{l=j+1}^{m}(x_{l-j}x_l-X_{l-j}X_l)+\sum_{l=m+1}^{m+j}(x_{l-j}-X_{l-j})X_l\}$$

for $j=0, 1, \cdots, p$, if we let

$$H=H(\alpha; n_1, n_2)=\{(z_0, z_1, \cdots, z_p)|\ |z_j|<(n_1+n_2)^{\alpha+1/2},\ 0\leq j\leq p\},$$

$$H_*=H_*(\alpha; n_1, n_2)=\{(z_0, z_1, \cdots, z_p)||z_j-d(j; n_1, n_2)|<(n_1+n_2)^{\alpha+1/2},\ 0\leq j\leq p\},$$

$$H_1=\{(z_0, z_1, \cdots, z_p)||z_j|\leq 1,\ 0\leq j\leq p\},$$

and if we let $h=(n_1+n_2)^\alpha$ where $\alpha<0$, then Lemma 2.3 implies that

$$\lim_{n_1,n_2\to\infty}\lim_{h\to 0}\Pr\{G(n_1, n_2; h)\}$$

$$=\begin{cases}\lim\limits_{n_1,n_2\to\infty}\int_H\phi_{p+1}(\underset{\sim}{z}|\underset{\sim}{0}, \Omega)d\underset{\sim}{z} & \text{if } \alpha<-0.5,\\ 1 & \text{if } -0.5<\alpha<0,\\ \int_{H_1}\phi_{p+1}(\underset{\sim}{z}|\underset{\sim}{0}, \Omega)\,d\underset{\sim}{z} & \text{if } \alpha=-0.5,\end{cases}$$

where the case that α is greater than -0.5 is due to problem 3 of Chung (1974, p. 93). A similar procedure yields that

$$\lim_{n_1,n_2\to\infty}\lim_{h\to 0}\Pr\{G(n_1, n_2; h)|\underset{\sim}{x}_m\}$$

$$=\begin{cases}\lim\limits_{n_1,n_2\to\infty}\int_{H*}\phi_{p+1}(\underset{\sim}{z}|\underset{\sim}{0}, \Omega)\,d\underset{\sim}{z} & \text{if } \alpha<-0.5,\\ 1 & \text{if } -0.5<\alpha<0,\\ \int_{H_1}\phi_{p+1}(\underset{\sim}{z}|\underset{\sim}{0}, \Omega)\,d\underset{\sim}{z} & \text{if } \alpha=-0.5,\end{cases}$$

since $d(i; n_1, n_2)$ tends to zero almost everywhere for $i=0, 1, 2, \cdots, p$. Thus, if $-0.5\leq\alpha<0$, then

$$\lim_{n_1,n_2\to\infty}\lim_{h\to 0}\Pr(G(n_1,\ n_2\ ;\ h)\,|\,\underset{\sim}{x}_m)/\Pr(G(n_1,\ n_2\ ;\ h))=1. \tag{3.4}$$

If α is less than -0.5, we obtain that

$$\lim_{n_1,n_2\to\infty}\lim_{h\to 0}\Pr(G(n_1,\ n_2\ ;\ h)\,|\,\underset{\sim}{x}_m)/\Pr(G(n_1,\ n_2\ ;\ h))$$

$$=\lim_{n_1,n_2\to\infty}\frac{\{2(n_1+n_2)^{\alpha+1/2}\}^{-(p+1)}\int_{H*}\phi_{p+1}(\underset{\sim}{z}\,|\,\underset{\sim}{0},\ \Omega)\,d\underset{\sim}{z}}{\{2(n_1+n_2)^{\alpha+1/2}\}^{-(p+1)}\int_{H}\phi_{p+1}(\underset{\sim}{z}\,|\,\underset{\sim}{0},\ \Omega)\,d\underset{\sim}{z}}$$

$$=\phi_{p+1}(\underset{\sim}{0}\,|\,\underset{\sim}{0},\ \Omega)/\phi_{p+1}(\underset{\sim}{0}\,|\,\underset{\sim}{0},\ \Omega)=1, \tag{3.5}$$

where the second to the last equality holds by the fact that $\lim_{n_1,n_2\to\infty} d(i\ ;\ n_1,\ n_2)=0$ a.e. for $i=0,\ 1,\ \cdots,\ p$. From Equations (3.3), (3.4) and (3.5), we conclude that

$$\lim_{n_1,n_2\to\infty}\phi(\underset{\sim}{x}_m\,|\,R_{n_1,n_2}(j)=\alpha_j,\ j=0,\ 1,\ \cdots,\ p)=\phi(\underset{\sim}{x}_m).$$

Since $\underset{\sim}{X}_{n_1,n_2}$ is the normal random vector with mean $\underset{\sim}{0}$ and covariance matrix $A_{n_1+n_2}$, the marginal p.d.f. $\phi(\underset{\sim}{x}_m)$ of the random vector $\underset{\sim}{X}_m$ is $\phi_m(\underset{\sim}{x}_m\,|\,\underset{\sim}{0},\ A_m)$. (See, e.g., Anderson [1958, p. 22]). Q.E.D.

4. Concluding Remarks.

Theorem 3.1 tells us that the conditional p.d.f. of the random vector $\underset{\sim}{X}_m=(X_1,\ X_2,\ \cdots,\ X_m)'$ given the sample autocovariance constraints $R_{n_1,n_2}(j)=\alpha_j$ for $0\leq j\leq p$, where $\{X_t\,|\,t=0,\ \pm1,\ \pm2,\ \cdots\}$ is a sequence of i.i.d. normal random variables with common p.d.f. $\phi_1(x\,|\,0,\ \delta^2)$, tends to $\phi_m(\underset{\sim}{x}_m\,|\,\underset{\sim}{0},\ A_m)$ which is the maximum entropy p.d.f. subject to the corresponding constraints $\mathrm{Cov}(X_l,\ X_{l+j})=\alpha_j$ for $l=1,\ 2,\ \cdots,\ m-j$ and $j=0,\ 1,\ \cdots,\ p$, (Choi and Cover [1984]).

Van Campenhout and Cover (1981) have shown that the conditional p.d.f. of X_1 given $\frac{1}{n}\sum_{i=1}^{m}X_i=\alpha$, where $\{X_i\}$ is a sequence of i.i.d. random variables with common p.d.f. $g(x)$ satisfying some regularity conditions, is the asymptotically closest, in Kullback-Leibler's sense, to the initial p.d.f. $g(x)$ among the p.d.f.'s satisfying $E(X_1)=\alpha$. As shown in Appendix, the p.d.f. $\phi_m(\underset{\sim}{x}_m\,|\,\underset{\sim}{0},\ A_m)$ is closest, in the Kullback-Leibler sense, to $\phi_m(\underset{\sim}{x}_m\,|\,\underset{\sim}{0},\ \delta^2 I)$ among the p.d.f.'s satisfying $\mathrm{Cov}(X_l,\ X_{l+j})=\alpha_j$ for $l=1,\ 2,\ \cdots,\ m-j$ and $=j0,\ 1,\ \cdots,\ p$. Thus Theorem 3.1 is a multivariate extension of Van Campenhout and Cover's theorem.

If we let

$$X_t=-(a_1X_{t-1}+a_2X_{t-2}+\cdots+a_pX_{t-p})+V_t,\ \ t=0,\ \pm1,\ \pm2,\ \cdots, \tag{4.1}$$

where $\{V_t\}$ is a sequence of i.i.d. normal random variables with mean 0 and variance σ^2, then the random vector $\underset{\sim}{X}_m=(X_1, X_2, \cdots, X_m)'$ has the p.d.f. $\phi_m(\underset{\sim}{x}_m|\underset{\sim}{0}, A_m)$. Thus the conditional p.d.f.'s of i.i.d. normal random variables tend to the joint p.d.f. of the Gaussian AR process in (4.1), which has the autocovariance function $\{\alpha_k\ ;\ k=0, \pm 1, \cdots\}$. This result can be related to the maximum entropy spectral density. Burg has proposed that the maximum entropy (rate) spectral density among the stationary (Gaussian) time series $\{X_t\}$ satisfying $\mathrm{Cov}(X_t, X_{t+j})=\alpha_j$ for $0\leq j\leq p$ is

$$S_x(\lambda)=\frac{1}{2\pi}\sum_{l=-\infty}^{\infty}\alpha_l \exp(-i\lambda l)=\frac{\sigma^2}{2\pi}\Big|\sum_{j=0}^{p} a_j \exp(-i\lambda j)\Big|^{-2},$$

which is the spectral density of the AR process defined by (4.1). If $m\to\infty$ so that $m=0((n_1+n_2)^{\alpha})$ for $0<\alpha<\frac{1}{2}$, then Theorem 3.1 still holds. In this case, the theorem implies that the Fourier transform of the autocovariance function of the limiting conditional p.d.f. is the maximum entropy spectral density subject to the constraints $\mathrm{Cov}(X_l, X_{l+j})=\alpha_j$ for $j=0, 1, 2, \cdots, p$ and any integer l. Thus, Theorem 3.1 gives another rationale to the maximum entropy spectral density.

Theorem 3.1 may be used in $ARMA$ model modification. Box-Jenkins' approach to model building consists of three stages called identification, estimation and diagnostic checking. The identification stage is deciding the orders of AR part and MA part, the estimation stage is estimating the AR coefficients and the MA coefficients, and the diagnostic checking stage is testing whether possible lack of fit exists or not. If any inadequacy is found in diagnostic checking, then we should modify our model. Since the diagnostic checking methods, which are proposed by Quenoulle (1947), Box and Pierce (1970), Ljung and Box (1978), etc., are based on the autocovariance function of residuals of the chosen model, it will be very useful to modify the model by use of the autocovariance function of the residuals. If the chosen model is correct, then the residuals $\{x_t\}$ should be uncorrelated. If the residuals $\{x_t\}$ are correlated, we conclude that the model is incorrect and should be modified. More precisely, if we let $R(j)=\frac{1}{n}\sum_{t=1}^{n-j} x_t x_{t+j}$, $j=0, 1, 2, \cdots$, where n is the number of residuals, and if $R(0)=\alpha_0$, $R(1)=\alpha_1, \cdots$, $R(p)=\alpha_p$ where the p is the largest positive integer that $R(p)$ is significantly different from zero, then Theorem 3.1 tells us that the residuals are considered to satisfy the stochastic difference equation (4.1). In other words, an autoregressive moving-average ($ARMA$) process $\{Y_t\}$ is identified and estimated by $\phi(B)Y_t=\theta(B)X_t$ where $\phi(B)=1+\phi_1 B+\phi_2 B^2+\cdots+\phi_r B^r$, $\theta(B)=1+\theta_1 B+\theta_2 B^2+\cdots+\theta_s B^s$ and B is the backward-

shift operator, and if the autocovariances of residuals, $R(0)$, $R(1)$, $\cdots$, $R(p)$, are α_0, α_1, $\cdots$, $\alpha_p(\neq 0)$, then we may replace the $ARMA(r, s)$ model $\phi(B)Y_t=\theta(B)X_t$ into the $ARMA(r+p, s)$ model $\phi(B)\eta(B)Y_t=\theta(B)V_t$, where $\eta(B)=1+a_1B+a_2B^2+\cdots+a_pB^p$ and $\{V_t\}$ is a sequence of white noises.

Appendix

Lemma. Define $g(\underset{\sim}{x})$ by $\phi_m(\underset{\sim}{x}|\underset{\sim}{0}, \delta^2 I)$. The p.d.f. that minimizes the Kullback-Leibler information number $D(f;g)=\int f(\underset{\sim}{x}) \ln\{f(\underset{\sim}{x})/g(\underset{\sim}{x})\}d\underset{\sim}{x}$ subject to $\int x_i x_j f(\underset{\sim}{x})d\underset{\sim}{x}=\alpha_{i-j}$ for $|i-j|\leq p$ is $\phi_m(\underset{\sim}{x}_m|\underset{\sim}{0}, A_m)$.

Proof. $D(f;g)=\int f(\underset{\sim}{x}) \ln f(\underset{\sim}{x})d\underset{\sim}{x}-\int f(\underset{\sim}{x}) \ln g(\underset{\sim}{x})\, d\underset{\sim}{x}$

$$=\int f(\underset{\sim}{x}) \ln f(\underset{\sim}{x})d\underset{\sim}{x}-\int f(\underset{\sim}{x})\left\{-\frac{m}{2}\ln(2\pi\delta^2)-\frac{1}{2}\delta^{-2}\underset{\sim}{x}'\underset{\sim}{x}\right\}d\underset{\sim}{x}$$

$$=\int f(\underset{\sim}{x}) \ln f(\underset{\sim}{x})\, d\underset{\sim}{x}+\frac{m}{2}\ln(2\pi\delta^2)+\frac{1}{2}\delta^{-2}\ \mathrm{tr}\int \underset{\sim}{x}\underset{\sim}{x}' f(\underset{\sim}{x})\, d\underset{\sim}{x}$$

$$=\int f(\underset{\sim}{x}) \ln f(\underset{\sim}{x})\, d\underset{\sim}{x}+\frac{m}{2}\ln(2\pi\delta^2)+\frac{1}{2}\delta^{-2}m\alpha_0$$

$$\geq\int \phi_m(\underset{\sim}{x}|\underset{\sim}{0},\ A_m) \ln \phi_m(\underset{\sim}{x}|\underset{\sim}{0},\ A_m)\, d\underset{\sim}{x}+\frac{m}{2}\ln(2\pi\sigma^2)+\frac{1}{2}\delta^{-2}m\alpha_0,$$

where the last inequality is proved by Choi and Cover (1984). Q.E.D.

REFERENCES

(1) Ables, J.G. (1974). "Maximum Entropy Spectral Analysis," *Astronomy Astrophysics: Supplement Series*, Vol. 15, pp. 383-393. Reprinted in *Modern Spectrum Analysis*, D.G. Childers, Ed. (1978), *IEEE* Press, pp. 23-33.

(2) Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, Inc., New York.

(3) Anderson, T.W. (1971). *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc., New York.

(4) Anderson, T.W. and H. Rubin (1950). "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *The Annals of Mathematical Statistics*, Vol. 21, pp. 570-582.

(5) Bartfai, P. (1972). "On a Conditional Limit Theorem," *Progress in Statistics: Colloquia Mathematica Societatis Janos Bolyai*, Vol. 9-1, pp. 85-91.

(6) Box, G. and G. Jenkins (1970). *Time Series Analysis: Forecasting and Control*, Holden-Day, Inc., San Francisco.

(7) Box, G. and D. Pierce (1970). "Distributions of Residual Autocorrelations in Autoregressive

Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, Vol. 65, pp. 1509-1526.

(8) Burg, J.P. (1967). "Maximum Entropy Spectral Analysis," Proceedings of the 37th Meeting of the Society of Exploration Geophysicists. Reprinted in *Modern Spectrum Analysis*, D.G. Childers, ED. (1978), IEEE Press, pp. 34-41.

(9) Burg, J.P. (1968). "A New Analysis Technique for Time Series Data," presented at the NATO Advanced Study Institute Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands. Reprinted in *Modern Spectrum Analysis*, D.G. Childers, Ed. (1978), IEEE Press, pp. 42-48.

(10) Childers, D.G., Ed. (1978). *Modern Spectrum Analysis*, IEEE Press, New York.

(11) Choi, B.S. (1973). *A Study of Recursive Formulae*, (Korean edition), Moonji Pub., Co., Seoul, Korea.

(12) Choi, B.S. (1983). "A Conditional Limit Characterization of the Maximum Entropy Spectral Density in Time Series Analysis," Ph. D. Dissertation, Stanford University, Stanford, Calif.

(13) Choi, B.S. (1984). "A Conditional Limit Construction of the Normal Probability Density," Discussion Paper No. 831206, The Industrial Management Research Center, Yonsei University, Seoul, Korea. Also, Will be appeared in The Annals of the Institute of Statistical Mathematics.

(14) Choi, B.S. and T.M. Cover (1984). "An Information Theoretic Proof of Burg's Maximum Entropy Spectrum," *Proceedings of the IEEE*, Vol. 72, pp. 1094-1095.

(15) Chung, K.L. (1974). *A Course in Probability Theory*, (2nd edition), Academic Prees, New York.

(16) Csiszár, I. and J. Körner (1981). *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Akademiai Kiado, Budapest, Hungary.

(17) Darwin, C.G. and R.H. Fowler (1922). "On the Partition of Energy," *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, Vol. 44, Sixth Series, pp. 450-479.

(18) Gantmacher, F.R. (1959). *The Theory of Matrices*, Chelsea Pub., Co., New-York.

(19) Grandell, J., M. Hamrud and P. Toll(1980). "A Remark on the Correspondence between the Maximum Entropy Method and the Autoregressive Model," *IEEE Trans. on Information Theory*, Vol.. IT-26, No. 6, pp. 750-751.

(20) Jaynes, E.T. (1957, a). "Information Theory and Statistical Mechanics," *Physical Review*, Vol. 106, No. 4, pp. 620-630.

(21) Jaynes, E.T. (1957, b). "Information Theory and Statistical Mechanics, Ⅱ," *Physical Review*, Vol. 108, No. 2, pp. 171-190.

(22) Jaynes, E.T. (1958). *Probability Theory in Science and Engineering*, Lecture notes issued by the Socony-Mobil Research Labs, Dallas, Texas.

(23) Jaynes, E.T. (1968). "Prior Probabilities," *IEEE Trans. on Systems Science and Cybernetics*, Vol. SSC-4, No. 3, pp. 227-241.

(24) Jaynes, E.T. (1978). "Where Do We Standon Maximum Entropy?" *The Maximum Entropy Formalism*, R.D. Levine and M. Tribus, Ed. (1978), pp. 15-118.

(25) Jaynes, E.T. (1982). "On the Rationale of Maximum-Entropy Methods," *Proceedings of the IEEE*, Vol. 70, No. 9, pp. 939-962.

(26) Kagan, A.M., Y.V. Linnik and C.R. Rao (1973). *Characterization Problems in Mathematical Statistis*, John Wiley & Sons, Inc., New York.

(27) Kaveh, M. and C.R. Cooper (1976). "An Empirical Investigation of the Properties of the Autoregressive Spectral Estimator," *IEEE Trans. on Information Theory*, Vol. IT-22, pp. 313-323. Reprinted in *Modern Spectrum Analysis*, D.G. Childers, Ed. (1978), IEEE Press, pp. 81-91.

(28) Kullback, S. and R.A. Leibler (1951). "On Information and Sufficiency," *The Annals of Mathematical Statistics*, Vol. 22, pp. 79-86.

(29) Lacoss, R.T. (1971). "Data Adaptive Spectral Analysis Methods," *Geophysics*, Vol. 36, pp. 661-675. Reprinted in *Modern Spectrum Analysis*, D.G. Childers, Ed. (1978), IEEE Press, pp. 134-148.

(30) Lanford, D.E. (1973). "Entropy and Equilibrium States in Classical Statistical Mechanics," *Statistical Mechanics and Mathematical Problems*, Lecture Notes in Physics, No. 20, pp. 1-113, Springer-Verlag, Berlin.

(31) Ljung, G.M. and G. Box (1978). "On the Measure of Lack of Fit in Time Series Models," *Biometrika*, Vol. 65, pp. 297-303.

(32) Quenouille, M.H. (1947). "A Large-Sample Test for the Goodness of Fit of Autoregressive Schemes," *Journal of* the *Royal Statistical Society*, Vol. A110, pp. 123-129.

(33) Sanov, I.N. (1957). "On the Probability of Large Deviations of Random Variables," *IMS and AMS Selected Translations in Mathematical Statistics and Probability*, Vol. 1, (1961), pp. 213-244.

(34) Tjur, T. (1974). *Conditional Probability Distributions*, Institute of Mathematical Statistics, University of Copenhagen.

(35) Van Campenhout, J.M. and T.M. Cover (1981). "Maximum Entropy and Conditional Probability," *IEEE Trans. on Information Theory*, Vol. IT-27, No. 4, pp. 483-489.

(36) Vasicek, O.A. (1980). "A Conditional Law of Large Numbers," *The Annals of Probability*, Vol. 8, No. 1, pp. 142-147.

(37) Vincze, I. (1972). "On the Maximum Probability Principle in Statistical Physics, "*Progress in Statistics: Colloquia Mathematico Societatis Janos Bolyai*, Vol. 9-Ⅱ, pp. 869-893.

(38) Zabell, S.L. (1974). "A Limit Theorem for Conditional Expectations with Applications to Probability Theory and Statistical Mechanics," Ph.D. Dissertation Harvard University, Cambridge, Mass.