POWER SPECTRUM ESTIMATION THROUGH AUTOREGRESSIVE MODEL FITTING

HIROTUGU AKAIKE

(Received July 16, 1969)

1. Introduction and summary

Recently the efficiency of the estimates of spectral characteristics of time series obtained by first fitting autoregressive models is attracting increasing attentions [1], [7], [10], [11]. The purpose of the present paper is to introduce a systematic approach to the evaluation of asymptotic sampling variabilities of this type of estimates of power spectra, under the assumption of strict stationarity and mutual independence of the innovations of the original stochastic processes.

Starting from the basic asymptotic distribution of the estimates of autoregression coefficients of an autoregressive process with independent innovations, which is due to Mann and Wald [8] and Anderson and Walker [3], we introduce a linear transformation of variables to get a distribution of a set of mutually uncorrelated variables. The limit distribution of any linear combinations of the original estimates can very easily be obtained by representing them with these uncorrelated variables. The transformation is induced by a process of successive orthogonalization of the variables in the time series, the coefficients of the transformation being obtained by recursive fitting of autoregressive models with increasing orders.

Though the estimate of power spectral density is a quadratic function of the estimates of autoregression coefficients its main variability in the limit is attributed to the linear term and the above stated procedure can be applied for the evaluation of its asymptotic distribution. The limit distribution being Gaussian, the evaluation of the variance matrix of the limit distribution of the estimates at various frequencies is sufficient for our purpose. The result is given in a readily computable form.

Some examples of evaluation of the variances of estimates are given and the relation of the results with the Parzen's guess ([11], p. 33) of the form of complex Wishart limit distribution of the estimates of spectral matrix is noticed.

As to the only difficulty in practically applying the present procedure, the decision of the order of the autoregressive model to be fitted, a practical scheme of decision has been proposed by the present author [2]. By using a result of application of this procedure a numerical example of estimation of a real power spectrum is illustrated and some general comments on practical applications are made.

The results of the present paper clearly indicate various potential advantages of this estimation procedure over the conventional one of Blackman-Tukey type [4].

2. Preliminary considerations

We first assume the time series under observation to be a realization of a purely non-deterministic weakly stationary process $X(n)$. Thus we are assuming $E(X(n))=0$ and the memory of the infinite past history of $X(n)$ is vanishing in the sense of mean square.

Under the present assumption $X(n)$ has a one-sided moving average representation

$$X(n)=\sum_{l=0}^{\infty} c_l \varepsilon(n-l)\,, \tag{2.1}$$

where $c_0=1$ and $\{\varepsilon(n)\}$ is a white noise, i.e., $\varepsilon(n)$'s are mutually uncorrelated, with $\mathfrak{M}\{\varepsilon(m);\, m\leqq n\}=\mathfrak{M}\{X(m);\, m\leqq n\}$, $\mathfrak{M}\{\cdots\}$ representing the closed linear manifold generated by the elements in the braces [6]. Accordingly, $X(n)$ can be approximated by $\varepsilon(n)+\sum_{m=1}^{l} a_{l,m}X(n-m)$ arbitrarily closely with increasing l, where $\sum_{m=1}^{l} a_{l,m}X(n-m)$ is the projection of $X(n)$ on $\mathfrak{M}\{X(n-m);\, m=1, 2, \cdots, l\}$. It is known [12] that the roots of

$$1-\sum_{m=1}^{l} a_{l,m}z^m=0 \tag{2.2}$$

lie outside the unit circle.

These observations provide the rationale of the wide applicability of an autoregressive model for the estimation of power spectra in practical situations.

3. Basic results of the estimation of autoregressive model

Taking into account the observations of the preceding section, we assume, without impairing the practical utility of the following discussions, that the time series under observation is a realization of an autoregressive process defined by

$$X(n)=\sum_{m=1}^{M} a_m X(n-m)+\varepsilon(n)\,, \tag{3.1}$$

where $\{\varepsilon(n)\}$ is a white noise, i.e., the variables are uncorrelated with zero mean and finite variance $E\varepsilon^2(n)=\sigma^2$ (>0), and the roots of the characteristic equation

$$1-\sum_{m=1}^{M} a_m z^m=0 \tag{3.2}$$

are lying outside the unit circle. $\varepsilon(n)$ is sometimes called the innovation of $X(n)$. The power spectral density function $p_{xx}(f)$ of the process $\{X(n)\}$ is given by

$$\begin{aligned} p_{xx}(f)&=\sum_{l=-\infty}^{\infty} R_{xx}(l)\exp(-i2\pi fl) \\ &=\frac{\sigma^2}{\left|1-\sum_{m=1}^{M} a_m \exp(-i2\pi fm)\right|^2}\,, \end{aligned} \tag{3.3}$$

where $R_{xx}(l)=EX(n+l)X(n)$.

We consider the situation where a set of data $\{X(n);\, n=1, 2, \cdots, N\}$ is given. Our estimate $\{\hat{a}_m;\, m=1, 2, \cdots, M\}$ of the autoregressive coefficients is obtained by solving the equation

$$\sum_{m=1}^{M} C_{xx}(l-m)\hat{a}_m=C_{xx}(l) \qquad l=1, 2, \cdots, M\,, \tag{3.4}$$

where $C_{xx}(l)=\frac{1}{N}\sum_{n=1}^{N-|l|} X(|l|+n)X(n)$. It should be noted that we are assuming $EX(n)=0$ and in practical applications $X(n)$ should be replaced by $X(n)-\bar{X}$, where $\bar{X}=\frac{1}{N}\sum_{n=1}^{N} X(n)$, in the definition of $C_{xx}(l)$. We also define an estimate $S^2(M)$ of σ^2 by

$$S^2(M)=C_{xx}(0)-\sum_{m=1}^{M} \hat{a}_m C_{xx}(m)\,. \tag{3.5}$$

This means that we are fitting a model with first $M+1$ covariances equal to $\{C_{xx}(l);\, l=0, 1, \cdots, M\}$. If the process $X(n)$ is strictly stationary and ergodic, $\hat{a}_m$ and $S^2(M)$ converge to a_m and σ^2 with probability one as N tends to infinity.

By using the result of Anderson and Walker [3] of the limit distribution of sample auto-correlation coefficients of a linear process we can get, THEOREM 1. *Under the assumption of strict stationarity and mutual independence of* $\{\varepsilon(n)\}$ *the distribution of* $\sqrt{N}\Delta a_m=\sqrt{N}(\hat{a}_m-a_m)$ $(m=1, 2, \cdots, M)$ *converges, as* N *tends to infinity, to an* M*-dimensional Gaussian distribution with zero mean vector and variance matrix* $\sigma^2 R_{MM}^{-1}$, *where* R_{MM} *is an* $M\times M$ *matrix with* (l, m)*-element equal to* $R_{xx}(m-l)=EX(n-l)\cdot X(n-m)$.

The theorem can be proved by using the fact that the limit distribution of $\{\sqrt{N}\Delta a_m; m=1, 2, \cdots, M\}$ is identical to that of $\{\eta_m; m=1, 2, \cdots, M\}$, where

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_M \end{bmatrix} = \sqrt{N} R_{MM}^{-1} \begin{bmatrix} C_{x\varepsilon}(1) \\ C_{x\varepsilon}(2) \\ \vdots \\ C_{x\varepsilon}(M) \end{bmatrix} \tag{3.6}$$

and

$$C_{x\varepsilon}(l)=\frac{1}{N}\sum_{n=1}^{N} X(n-l)\varepsilon(n) \qquad (l=1, 2, \cdots, M).$$

From the result of Anderson and Walker [3] it can be shown that distribution of $\{\sqrt{N}C_{x\varepsilon}(l); l=1, 2, \cdots, M\}$ tends asymptotically to be Gaussian with zero mean vector and variance matrix $\sigma^2 R_{MM}$. The assertion of the theorem is a direct consequence of these observations.

Now the matrix R_{MM}^{-1} admits the following factorization

$$R_{MM}^{-1}=B_M^T D_M D_M B_M, \tag{3.7}$$

$$B_M=\begin{bmatrix} 1 & 0 & 0 \cdots 0 \\ -a_{1,1} & 1 & 0 \cdots 0 \\ -a_{2,2} & -a_{2,1} & 1 \cdots 0 \\ \vdots & \vdots & \vdots \quad \vdots \\ -a_{M-1,M-1} & -a_{M-1,M-2} & \cdots\ 1 \end{bmatrix} \tag{3.8}$$

$$D_M=\begin{bmatrix} \frac{1}{\sigma(0)} & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma(1)} & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma(2)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\sigma(M-1)} \end{bmatrix}, \tag{3.9}$$

$\sigma(l)$ and $\{a_{l,m};\, m=1, 2, \cdots, l\}$ being given by the relation

$$\sigma^2(l)=E\left|X(n-l)-\sum_{m=1}^{l} a_{l,m}X(n-l+m)\right|^2 \tag{3.10}$$
$$=\underset{\{\beta_m\}}{\mathrm{Min}}\, E\left|X(n-l)-\sum_{m=1}^{l} \beta_m X(n-l+m)\right|^2 ,$$

and T denotes the transpose. From this relation we can see that the random variables $Y(n-l)=X(n-l)-\sum_{m=1}^{l} a_{l,m}X(n-l+m)$ $(l=0, 1, \cdots, M)$ are mutually uncorrelated and with variances equal to $\sigma^2(l)$, respectively. We are assuming $a_{M,k}=a_k$ $(k=1, 2, \cdots, M)$, $\sigma^2(M)=\sigma^2$, $Y(n)=X(n)$ and $\sigma^2(0)=EX^2(n)$. Thus by applying the transformation $D_M B_M$ to the vector $X_M=(X(n), X(n-1), \cdots, X(n-M+1))^T$ we get $D_M B_M X_M=(z_0(n), z_1(n), \cdots, z_{M-1}(n))^T$ with $Ez_l(n)z_m(n)=0$ $(l\neq m)$ and $=1$ $(l=m)$. As the variance matrix of the limit distribution of $\{\sqrt{N}C_{x\varepsilon}(l);\, l=1, 2, \cdots, M\}$ is equal to $\sigma^2 R_{MM}$, where $R_{MM}=EX_M X_M^T$, we can see that for $C_{x\varepsilon}=\{C_{x\varepsilon}(1), C_{x\varepsilon}(2), \cdots, C_{x\varepsilon}(M)\}^T$ the variance matrix of the limit distribution of $U_M=\sigma^{-1}\sqrt{N}D_M B_M C_{x\varepsilon}$ is an $M\times M$ identity matrix I_M. Thus we get an asymptotic representation of $\sqrt{N}\Delta a=\{\sqrt{N}\Delta a_m;\, m=1, 2, \cdots, M\}^T$ by variables which in their limit distribution are mutually orthogonal and normalized;

$$\sqrt{N}\Delta a \sim \sigma B_M^T D_M U_M\, , \tag{3.11}$$

where $\sim$ means that the both side members have one and the same limit distribution. By using this relation we can very easily obtain the limit distributions of any linear transforms of $\sqrt{N}\Delta a$.

The foregoing discussions are all valid under the assumption of mutual independence and strict stationarity of $\{\varepsilon(n)\}$. To discuss the overall sampling variability of the estimate of power spectrum we have to add a further assumption of finiteness of the fourth order moment of $\varepsilon(n)$. In this case we have, for $C_{\varepsilon\varepsilon}(l)=\frac{1}{N}\sum_{n=1}^{N}\varepsilon(n+l)\varepsilon(n)$,

THEOREM 2. *If $\{\varepsilon(n)\}$ is strictly stationary and $\varepsilon(n)$'s are mutually independent with finite fourth order moment $m_4=E\varepsilon^4(n)$, the simultaneous distribution of $\sqrt{N}\frac{C_{\varepsilon\varepsilon}(0)-\sigma^2}{\sqrt{m_4-\sigma^4}}$ and $\sqrt{N}\frac{C_{\varepsilon\varepsilon}(l)}{\sigma^2}$ $(l=1, 2, \cdots, L)$ (L; any positive integer) converges to $(L+1)$-dimensional Gaussian distribution with zero mean vector and variance matrix I_{L+1}, where I_{L+1} is an $(L+1)\times(L+1)$ identity matrix.*

The proof of this theorem is readily obtained by using the central limit theorem for finitely dependent stationary process due to Diananda [5].

As $C_{x\varepsilon}(l)$'s are arbitrarily closely approximated, in the sense of mean square, by linear combinations of $C_{\varepsilon\varepsilon}(m)$'s our present theorem shows that $\sqrt{N}(C_{\varepsilon\varepsilon}(0)-\sigma^2)$ and $\sqrt{N}\Delta a$ tend to be independent in their limit distribution. This fact will be used in the following discussion of the sampling variability of the estimate of power spectral density.

It should be mentioned here that the mutual independence assumption of $\{\varepsilon(n)\}$ is quite restrictive. This point would deserve further investigations.

4. Estimation of power spectral density

Our estimate $\hat{p}_{xx}(f)$ of the power spectral density $p_{xx}(f)$ is given by

$$\hat{p}_{xx}(f)=\frac{S^2(M)}{\left|1-\sum_{m=1}^{M}\hat{a}_m \exp(-i2\pi fm)\right|^2}, \tag{4.1}$$

where $\{\hat{a}_m\}$ and $S^2(M)$ are given by (3.4) and (3.5). If we put $A_M(f)=1-\sum_{m=1}^{M} a_m \exp(-i2\pi fm)$ and $\hat{A}_M(f)=1-\sum_{m=1}^{M}\hat{a}_m \exp(-i2\pi fm)$, we have

$$\begin{aligned}\Delta p_{xx}(f)&=\hat{p}_{xx}(f)-p_{xx}(f)\\ &=\frac{\Delta S^2(M)\,|A_M(f)|^2-\Delta|A_M(f)|^2\sigma^2(M)}{|\hat{A}_M(f)|^2|A_M(f)|^2},\end{aligned} \tag{4.2}$$

where $\Delta S^2(M)=S^2(M)-\sigma^2(M)$ and $\Delta|A_M(f)|^2=|\hat{A}_M(f)|^2-|A_M(f)|^2$. We assume the condition of Theorem 2 to hold hereafter. For the following discussion of limit distributions reference should be made to another paper by Mann and Wald [9].

Under the present assumption the process $X(n)$ is strictly stationary and ergodic and, as N tends to infinity, $S^2(M)$ and $\hat{A}_M(f)$ converges with probability one to $\sigma^2(M)$ and $A_M(f)$ which are never equal to zero. Also it can be shown that the limit distribution of $\sqrt{N}\Delta S^2(M)$ and $\sqrt{N}\Delta|A_M(f)|^2$ is identical to that of $\sqrt{N}\nabla S^2(M)$ and $\sqrt{N}\nabla|A_M(f)|^2$, where by definition $\nabla S^2(M)=(\bar{\varepsilon}^2-\sigma^2(M))$, $\nabla|A_M(f)|^2=(A_M(f)\overline{\Delta A_M(f)}+\overline{A_M(f)}\Delta A_M(f))$, $\bar{\varepsilon}^2=C_{\varepsilon\varepsilon}(0)$ $\left(=\frac{1}{N}\sum_{m=1}^{N}\varepsilon^2(n)\right)$ and $\Delta A_M(f)=-\sum_{m=1}^{M}\Delta a_m \exp(-i2\pi fm)$. Thus the limit distribution of $\sqrt{N}\Delta p_{xx}(f)$ is identical to that of $Q_N(f)$ which is defined by

$$Q_N(f)=\sqrt{N}\left\{\left(\frac{\nabla S^2(M)}{\sigma^2(M)}\right)-\left(\frac{\nabla|A_M(f)|^2}{|A_M(f)|^2}\right)\right\}p_{xx}(f). \tag{4.3}$$

As was mentioned in the preceding section $\bar{\varepsilon}^2$ tends to be independent with Δa in the limit distribution and the present result shows that

$\sqrt{N}\Delta p_{xx}(f)/p_{xx}(f)$ has a limit distribution with a variance composed of two components: one due to the relative variation of $S^2(M)$ and the other due to that of $|\hat{A}_M(f)|^2$. For the evaluation of the former we have $NE\left(\frac{\nabla S^2(M)}{\sigma^2(M)}\right)^2 = \frac{m_4}{\sigma^4} - 1$. For the latter, the orthogonal representation of $\sqrt{N}\Delta a$ obtained in the former section is very conveniently used.

Using the notations of the preceding section, we have, from (3.11), for any $M \times L$ matrices V and W

$$E_\infty\{N[V^T \Delta a(\Delta a)^T W]\} = \sigma^2[V^T B_M^T D_M D_M B_M W]\,, \tag{4.4}$$

where $E_\infty\{\cdot\}$ denotes the expectation in the limit distribution of the quantity whithin the brace. Using this result we can evaluate the variances and covariances of our estimates at various frequencies. As $\Delta A_M(f) = -\sum_{m=1}^{M} \Delta a_m \exp(-i2\pi fm)$, we have for $\Delta A_M = (\Delta A_M(f_1), \Delta A_M(f_2), \cdots, \Delta A_M(f_K))^T$

$$\begin{aligned} E_\infty\{N\Delta A_M(\Delta A_M)^T\} &= E_\infty\{NF_K^T \Delta a(\Delta a)^T F_K\} \\ &= \sigma^2[F_K^T B_M^T D_M D_M B_M F_K]\,, \end{aligned} \tag{4.5}$$

where

$$F_K = \begin{bmatrix} \exp(-i2\pi f_1) & \exp(-i2\pi f_2) & \cdots & \exp(-i2\pi f_K) \\ \exp(-i2\pi 2f_1) & \exp(-i2\pi 2f_2) & \cdots & \exp(-i2\pi 2f_K) \\ \vdots & \vdots & & \vdots \\ \exp(-i2\pi Mf_1) & \exp(-i2\pi Mf_2) & \cdots & \exp(-i2\pi Mf_K) \end{bmatrix}. \tag{4.6}$$

From the definition of B_M we have

(4.7) $B_M F_K =$

$$\begin{bmatrix} \overline{A_0(f_1)}\exp(-i2\pi f_1) & \overline{A_0(f_2)}\exp(-i2\pi f_2) & \cdots \overline{A_0(f_K)}\exp(-i2\pi f_K) \\ \overline{A_1(f_1)}\exp(-i2\pi 2f_1) & \overline{A_1(f_2)}\exp(-i2\pi 2f_2) & \cdots \overline{A_1(f_K)}\exp(-i2\pi 2f_K) \\ \vdots & \vdots & \vdots \\ \overline{A_{M-1}(f_1)}\exp(-i2\pi Mf_1) & \overline{A_{M-1}(f_2)}\exp(-i2\pi Mf_2) & \cdots \overline{A_{M-1}(f_K)}\exp(-i2\pi Mf_K) \end{bmatrix},$$

where $A_k(f) = 1 - \sum_{m=1}^{k} a_{k,m}\exp(-i2\pi mf)$ and — denotes the complex conjugate, and accordingly

$$\begin{aligned} &[E_\infty\{N\Delta A_M(\Delta A_M)^T\}](l, m) \\ &= \sigma^2 \sum_{k=0}^{M-1} (\overline{A_k(f_l)}\exp(-i2\pi(k+1)f_l) \\ &\quad \cdot \overline{A_k(f_m)}\exp(-i2\pi(k+1)f_m))\frac{1}{\sigma^2(k)}\,, \end{aligned} \tag{4.8}$$

where [ ](l, m) denotes (l, m)-element of the matrix inside the square brackets. In the same manner we can get

$$[E_\infty\{N\Delta A_M(\overline{\Delta A_M})^T\}](l, m) \\ =\sigma^2 \sum_{k=0}^{M-1} (\overline{A_k(f_l)} \exp(-i2\pi(k+1)f_l) \\ \cdot \overline{A_k(f_m)} \exp(-i2\pi(k+1)f_m))\frac{1}{\sigma^2(k)} . \tag{4.9}$$

By using these results we get

$$E_\infty\left\{N\frac{\nabla|A_M(f_l)|^2}{|A_M(f_l)|^2}\frac{\nabla|A_M(f_m)|^2}{|A_M(f_m)|^2}\right\} \\ =4p_{xx}(f_l)p_{xx}(f_m)\sum_{k=0}^{M-1}\mathrm{Re}\,(C_k(f_l))\,\mathrm{Re}\,(C_k(f_m)) , \tag{4.10}$$

where

$$C_k(f)=\frac{A_M(f)}{\sqrt{\sigma^2(M)}}\frac{A_k(f)}{\sqrt{\sigma^2(k)}}\exp(i2\pi(k+1)f) \tag{4.11}$$

and Re$(*)$ denotes the real part of $*$. Our final result is thus given by the following:

$$E_\infty\left\{N\frac{\Delta p_{xx}(f_l)}{p_{xx}(f_l)}\frac{\Delta p_{xx}(f_m)}{p_{xx}(f_m)}\right\} \\ =\left(\frac{E\varepsilon^4}{(E\varepsilon^2)^2}-1\right)+4p_{xx}(f_l)p_{xx}(f_m)\sum_{k=0}^{M-1}\mathrm{Re}\,(C_k(f_l))\,\mathrm{Re}\,(C_k(f_m)) . \tag{4.12}$$

If we are only interested in the relative shape of the spectrum we should use $\hat{q}_{xx}(f)=\hat{p}_{xx}(f)/S^2(M)$ instead of $\hat{p}_{xx}(f)$. For the evaluation of the variance of the limit distribution of $\hat{q}_{xx}(f)$ we have only to retain the second term in the above expression and for this we need only the assumption of Theorem 1.

It also should be noted that the foregoing results are all valid when M is equal to or greater than the order M_0 of the process, M_0 being the value of M for which the model holds exactly with $a_M\neq 0$.

5. Some examples of evaluation of variances of limit distributions

Here we shall see some of the results of evaluation of variances of asymptotic distributions.

Case I. *White noise.* In this case we have $A_k(f)=1$ and $\sigma^2(k)=\sigma^2$ $(k=0, 1, \cdots, M)$. Accordingly we have $C_k(f)=(1/\sigma^2)\cdot\exp(i2\pi(k+1)f)$ and we get

$$E_\infty=\left\{N\left(\frac{\nabla|A_M(f)|^2}{|A_M(f)|^2}\right)\right\}=4\sum_{k=0}^{M-1}(\cos 2\pi(k+1)f)^2 \tag{5.1}$$
$$=2M\left(1+\frac{\sin 2\pi Mf}{M\sin 2\pi f}\cos 2\pi(M+1)f\right)$$
$$=4M \quad \text{at } f=0,\frac{1}{2},$$
$$=2M \quad \text{at } f=\frac{2k+1}{4(M+1)}$$
$$(k=0,1,\cdots,M) \text{ and at } f=\frac{l}{2M} \quad (l=1,2,\cdots,M-1).$$

This result shows a typical behavior of our estimates at $f=0$ and $1/2$, where the variance is nearly doubled compared with that around $f=1/4$.

Case II. *Markovian noise.* This is the case where the order M_0 of the process is equal to 1, i.e., $X(n)$ is given by the relation $X(n)=aX(n-1)+\varepsilon(n)$ with $|a|<1$. For this case we have, for $M=1$,

$$E_\infty=\left\{N\left(\frac{\nabla|A_M(f)|^2}{|A_M(f)|^2}\right)^2\right\}=4(1-a^2)\frac{(\cos 2\pi f-a)^2}{(1+a^2-2a\cos 2\pi f)^2} \tag{5.2}$$
$$=4\frac{1+a}{1-a} \quad \text{at } f=0$$
$$=4\frac{1-a}{1+a} \quad \text{at } f=\frac{1}{2}.$$

The result shows that when a is close to 1 the variability of our estimate is very large at $f=0$. The effect of increasing M beyond $M_0=1$ can be evaluated as in the following discussion of Case III.

Case III. *General case, but with large* $M(\gg M_0)$. When k is equal to or greater than M_0 we have $A_k(f)=A_M(f)$. If we put $A_M(f)=|A_M(f)|\exp(i\Phi(f))$ we have, for $k\geqq M_0$,

$$C_k(f)=(p_{xx}(f))^{-1}\exp(i(2\Phi(f)+2\pi(k+1)f)). \tag{5.3}$$

When M_0 is negligibly small compared with M, $\sum_{k=0}^{M-1}[\mathrm{Re}\,(C_k(f))]^2$ will then be approximated by

$$(p_{xx}(f))^{-2}(2M)\left(1+\frac{\sin 2\pi Mf}{M\sin 2\pi f}\cos(2\pi(M+1)f+4\Phi(f))\right). \tag{5.4}$$

Thus in this case we can see that if the contribution of the fourth order moment of $\varepsilon(n)$ is negligible the variance of the limit distribution of $\hat{p}_{xx}(f)/p_{xx}(f)$ will be nearly equal to that of the case of a white noise discussed in Case I.

This is just the result which is often assumed to hold for the conventional type of estimates, which are essentially local averages of periodgrams, assuming the averaging operation to be extended over such a narrow bandwidth that the spectrum can be considered to be flat within the band.

In practical applications of our present estimation procedure we try to keep M as close as possible to M_0 and the situation assumed in Case III will never take place. Thus our estimate will often be with smaller variance than that of a conventional estimate. Also by taking into account of the fact that the bias of our estimate is generally considered to be of the order of $1/N$ we can expect that the bias of our estimate will automatically be decreasing as N is increased.

In passing we will mention that the discussion of this section suggests that the Parzen's guess ([11], p. 33) of the complex Wishart limit distribution of an estimate of spectral matrix, obtained by fitting an auto-regressive model to a multiple time series, to be with a variance matrix equal to the spectral density matrix may only be valid under the situation of Case III.

6. Comments on practical applications

The only difficulty left for the practical application of the present procedure will be the decision of M to be fitted to the observed data. The decision must be useful even for the cases where actually the true order of the process is infinity. A practical procedure of this decision has recently been proposed by the present author [2]. Our numerical examples of successive applications of this decision procedure and the present estimation procedure produced reasonable estimates of power spectra without very much amount of help of our subjective judgement. A result of application of this procedure to a real time series of length $N=511$ is illustrated in Fig. 1.

At the upper part of the figure, the estimate obtained by this successive procedure is represented by a solid line. Along with this, there are illustrated two other estimates which were obtained by applying the hanning type windows [4] with truncation points, or the maximum numbers of lags (MAXLAG's) used for estimation, equal to 45 and 90, respectively. Details of this overall procedure and the numerical results will be discussed in a subseqent paper. At the lower part of the figure is illustrated the values of $\left(\frac{1}{N}E_\infty\left\{N\left(\frac{\nabla|A_M(f)|^2}{|A_M(f)|^2}\right)\right\}\right)^{1/2}$, or the asymptotic coefficient of variation of $\hat{q}_{xx}(f)$, which is obtained by assuming the estimated quantities to be the real values. This last procedure is going to be briefly discussed in the following.

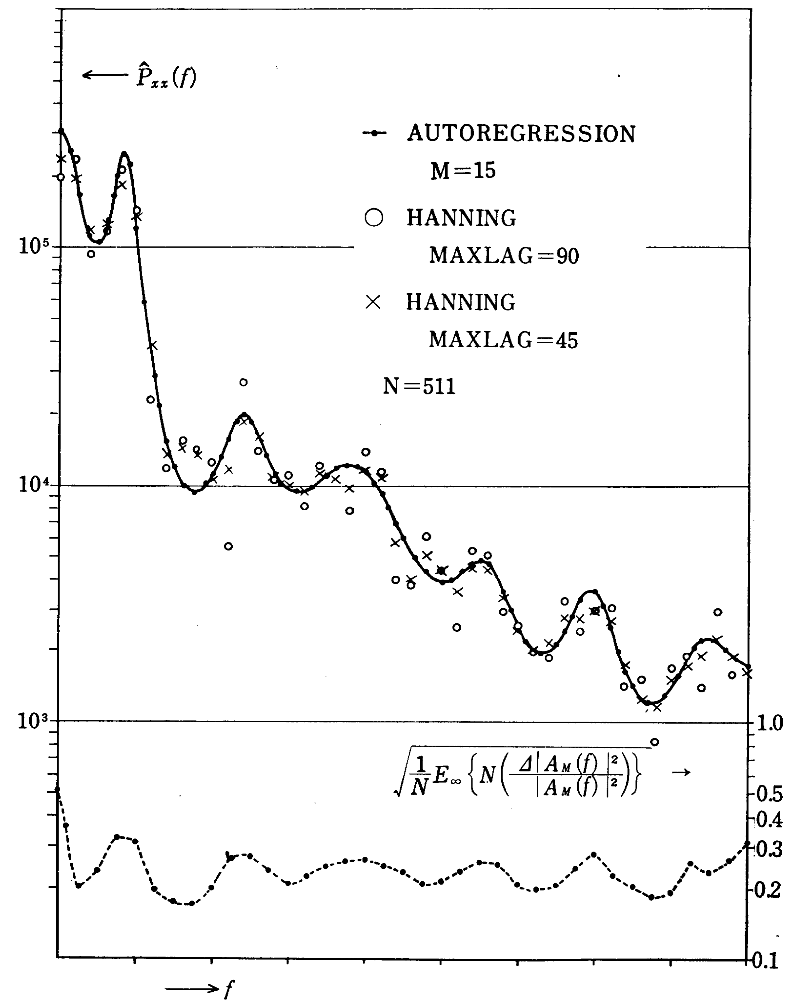

Fig. 1 Estimates of power spectrum of a real time series and the asymtotic variability of autoregressive estimate.

In practical situations, we can get estimates of $C_k(f)$'s by replacing $\{a_{l,m}\}$ and $\sigma^2(l)$ in their definitions by $\{\hat{a}_{l,m}\}$ and $S^2(l)$ $(l=0, 1, 2, \cdots, M)$ which are obtained by successively solving

$$(6.1) \qquad \sum_{m=1}^{l} C_{xx}(k-m)\hat{a}_{l,m}=C_{xx}(k) \qquad k=1, 2, \cdots, l$$

and

$$(6.2) \qquad S^2(l)=C_{xx}(0)-\sum_{m=1}^{l} \hat{a}_{l,m}C_{xx}(m) .$$

The computation is very simply implemented by the recursive procedure. Also an estimate of $E\varepsilon^4(n)$ would be available by using $\{\hat{a}_m\}$ and the original data.

By using the quantities thus obtained we can get an estimate of the variance of the limit distribution, but the overall statistical characteristics of this last estimate is left for further investigation.

Assuming the fitted model to be strict and the process to be Gaussian, we can obtain the equivalent degrees of freedom of our estimate as defined by $2N\left[E_\infty\left\{N\left(\frac{\Delta p_{xx}(f)}{p_{xx}(f)}\right)^2\right\}\right]^{-1}$ with $m_4=3\sigma^2$, which is to be used for the comparison of the present estimate with conventional ones.

Table 1. Equivalent degrees of freedom and equivalent truncation points for hanning and Parzen windows of the estimate of Fig. 1.

f	Equivalent degrees of freedom	Equivalent truncation points	
		HANNING	PARZEN
1/36	36.4	40.1	55.7
2/36	21.0	67.5	93.7
3/36	64.6	23.7	33.0
4/36	50.7	29.5	41.0
5/36	27.4	52.4	72.2
6/36	45.4	32.7	45.4
7/36	32.7	44.3	61.6
8/36	29.1	49.5	68.7
9/36	36.5	40.0	55.5
10/36	43.5	34.0	47.2
11/36	29.9	48.2	67.0
12/36	46.3	32.1	44.5
13/36	46.1	32.2	44.8
14/36	24.9	57.4	79.8
15/36	47.6	31.3	43.4
16/36	53.9	27.9	38.8
17/36	36.3	40.2	55.8

Also by multiplying $E_\infty\left\{N\left(\frac{\Delta p_{xx}(f)}{p_{xx}(f)}\right)^2\right\}$ by proper constants we can convert this quantity into the ones which we will call the equivalent truncation points of respective windows, which give the truncation points, or the maximum numbers of the lags of autocovariances used for estimation, of respective windows which will give one and the same variability as the present one when the spectrum is locally flact and the process is Gaussian. In Table 1 are given these quantities along with the equivalent degrees of freedom for the data of Fig. 1. The result confirms our feeling [2] that the present estimate shows generally smaller variability than the classical estimates which are shown in Fig. 1 and considered to be with reasonable resolvabilities.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] H. Akaike, "A method of statistical identification of discrete time parameter linear systems, *Ann. Inst. Statist. Math.*, 21 (1969), 225-242.

[2] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Statist. Math.*, 21 (1969), 243-247.

[3] T. W. Anderson and A. M. Walker, "On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process," *Ann. Math. Statist.*, 35 (1964), 1296-1303.

[4] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*, New York, Dover, 1959.

[5] P. H. Diananda, "Some probability limit theorems with statistical applications," *Proc. Cambridge Philos. Soc.*, 49 (1953), 239-246.

[6] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley, New York, 1957.

[7] R. H. Jones, D. H. Crowell and L. E. Kapuniai, "Change detection model for serially correlated multivariate data," Unpublished manuscript, 1968.

[8] H. B. Mann and A. Wald, "On the statistical treatment of linear stochastic difference equations," *Econometrica*, 11 (1943), 173-220.

[9] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *Ann. Math. Statist.*, 14 (1943), 217-226.

[10] E. Parzen, "Statistical spectral analysis (single channel case) in 1968," Stanford University Statistics Department Technical Report, 11 (1968).

[11] E. Parzen, "Multiple time series modelling," Stanford University Statistics Department Technical Report, 12 (1968), (A paper presented at the Second International Symposium on Multivariate Analysis, Dayton, Ohio on June 20, 1968).

[12] P. Whittle, "On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix," *Biometrika*, 50 (1963), 129-134.