FITTING AUTOREGRESSIVE MODELS FOR PREDICTION

HIROTUGU AKAIKE

(Received June 17, 1969)

1. Introduction and summary

This is a preliminary report on a newly developed simple and practical procedure of statistical identification of predictors by using autoregressive models. The use of autoregressive representation of a stationary time series (or the innovations approach) in the analysis of time series has recently been attracting attentions of many research workers and it is expected that this time domain approach will give answers to many problems, such as the identification of noisy feedback systems, which could not be solved by the direct application of frequency domain approach [1], [2], [3], [9].

The main difficulty in fitting an autoregressive model

$$X(n)=\sum_{m=1}^{M} a_m X(n-m)+a_0+\varepsilon(n)\ ,$$

where $X(n)$ is the process being observed and $\varepsilon(n)$ is its innovation which is uncorrelated with $X(l)$ $(l<n)$ and is forming a white noise, lies in the decision of the order M. We assume the mutual independence and strict stationarity of $\{\varepsilon(n)\}$.

There have been extensive investigations of topics which are very closely related with this subject but the definite description of the procedure, which could directly be adopted for practical applications, is quite lacking yet [6], [7]. T. W. Anderson [4] has treated this problem as a multiple decision problem and given a description of a procedure which in some sense has an optimal property. Though the procedure is well described, it contains many constants which are to be determined before its application and the problem of selection of these constants are left open. E. Parzen is advocating the use of autoregressive representation for the estimation of spectra and suggested the use of the maximum likelihood test criterion proposed by P. Whittle for this purpose [8], [9].

The main difficulty in applying this kind of procedures stems from the fact that they are essentially formulated in the form of a successive

test of the whiteness of the series against multiple "alternatives." Actually one of the "alternatives" is just the model we are looking for and thus it is very difficult for us to get the feeling of the possible alternatives to set reasonable "significance levels."

To overcome this difficulty we adopt entirely decision theoretic approach where a figure of merit is defined for each model being fitted and the one with the best figure is chosen as our predictor. This figure of merit which we shall call the final prediction error (FPE) is defined as the expected variance of the prediction error when an autoregressive model fitted to the present series of $X(n)$ is applied to another independent realization of $X(n)$, or to the process with one and the same covariance characteristic as that of $X(n)$ and is independent of the present $X(n)$, to make a one step prediction. The notion of FPE can also be utilized for the decision of the constants in Anderson's procedure.

In the practical application of our procedure we compute an estimate of FPE of each autoregressive model within a prescribed sufficiently wide range of possible orders and select the one which gives the minimum of the estimates. This procedure we shall call the FPE scheme.

In this paper we shall only give a description with a brief discussion of the FPE scheme for practical use and the theoretical and numerical details of our investigation of the scheme will be discussed in separate papers.

2. FPE scheme

We consider the situation where a set of data $\{X(n);\ n=1, 2, \cdots, N\}$ is given.

0) First we replace $X(n)$ by $\tilde{X}(n)=X(n)-\bar{X}$, where $\bar{X}=\frac{1}{N}\sum_{n=1}^{N} X(n)$.

1) We set the upper limit L of the order of autoregressive models to be fitted to the data. L should be chosen large enough not to exclude the efficient model. Also confer the description in the following 2).

2) We calculate the sample autocovariances

$$C_{xx}(l)=\frac{1}{N}\sum_{n=1}^{N-l}\tilde{X}(n+l)\tilde{X}(n) \qquad \text{for} \quad l=0, 1, 2, \cdots, L\ .$$

The value of L here is generally much smaller than the value of L usually considered to be necessary for the estimation of power spectrum by Fourier transforming the windowed sample autocovariance function.

3) Then we try to fit the autoregressive model of order M ($M=1, 2, \cdots, L$) by the least squares method which requires the mean square of

residuals

$$R(a^{(M)})=\frac{1}{N}\sum_{n=1}^{N}\left(\tilde{X}(n)-\sum_{m=1}^{M}a_m^{(M)}\tilde{X}(n-m)\right)^2$$

to be minimized with respect to $\{a_m^{(M)};\ m=1, 2, \cdots, M\}$ assuming $\tilde{X}(0)=\tilde{X}(-1)=\cdots=\tilde{X}(-M+1)=0$. The required set of coefficients $\{\hat{a}_m^{(M)};\ m=1, 2, \cdots, M\}$ is obtained by solving the normal equation

$$\begin{pmatrix} C_{xx}(0), & C_{xx}(1), & \cdots, & C_{xx}(M-1) \\ C_{xx}(1), & C_{xx}(0), & \cdots, & C_{xx}(M-2) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ C_{xx}(M-1), & C_{xx}(M-2), & \cdots, & C_{xx}(0) \end{pmatrix} \begin{pmatrix} \hat{a}_1^{(M)} \\ \hat{a}_2^{(M)} \\ \cdot \\ \cdot \\ \cdot \\ \hat{a}_M^{(M)} \end{pmatrix} = \begin{pmatrix} C_{xx}(1) \\ C_{xx}(2) \\ \cdot \\ \cdot \\ \cdot \\ C_{xx}(M) \end{pmatrix}.$$

We denote the value of $R(a^{(M)})$ corresponding to this solution $\{\hat{a}_m^{(M)};\ m=1, 2, \cdots, M\}$ by R_M. An estimate $(\mathrm{FPE})_M$ of FPE of the autoregressive model of order M is calculated by the definition

$$(\mathrm{FPE})_M=\left(1+\frac{M+1}{N}\right)S_M,$$

where

$$S_M=\frac{N}{N-1-M}R_M.$$

We put $R_0=C_{xx}(0)$, $S_0=\frac{N}{N-1}R_0$ and $(\mathrm{FPE})_0=S_0$.

For the purpose of comparison of the magnitudes of $(\mathrm{FPE})_M$ the relative value $(\mathrm{RFPE})_M$ of $(\mathrm{FPE})_M$ defined by

$$(\mathrm{RFPE})_M=\frac{(\mathrm{FPE})_M}{(\mathrm{FPE})_0}$$

can conveniently be used.

The recursive method of solution of the equation is most conveniently applied for the computation of this step.

4) We adopt the value M_0 of M which gives the minimum of $(\mathrm{FPE})_M$ within $M=0, 1, 2, \cdots, L$ as the order of our model for prediction. $\{\hat{a}_m^{(M)};\ m=0, 1, \cdots, M\}$, where $\hat{a}_0^{(M)}=\left(1-\sum_{M=1}^{M}\hat{a}_m^{(M)}\right)\bar{X}$, and $S_M=\frac{N}{N-1-M}R_M$ with $M=M_0$ are adopted as our estimate of the set of coefficients of predictor and that of the innovation variance, respectively.

5) An estimate $\hat{p}_{xx}(f)$ of the power spectrum density $p_{xx}(f)$ of

$\{X(n)-E(X(n))\}$ at frequency f is obtained by the formula

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

with $M=M_0$.

3. A brief discussion of FPE scheme

As is suggested by the definition of its estimate, the definition of FPE of the autoregressive model of order M is given by the relation

$$\mathrm{FPE}=\left(1+\frac{M+1}{N}\right)r_M,$$

where r_M is the minimum of $E\left(X(n)-\sum_{m=1}^{M}a_m^{(M)}X(n-m)-a_0^{(M)}\right)^2$ with respect to $\{a_m^{(M)};\ m=0,1,\cdots,M\}$. Obviously r_M is equal to the variance of the innovation $\varepsilon(n)$ when $X(n)$ is generated from $\varepsilon(n)$ by a finite autoregression of order equal to or less than M. In this case FPE gives the asymptotic mean square of the prediction error when the least squares estimate $\{\hat{a}_m^{(M)};\ m=0,1,\cdots,M\}$ is applied to another independent observation of the same process, or to the process with one and the same covariance characteristic and is independent of the present process, to make a one-step prediction. We can see that FPE tends to be large when unnecessarily large value of M is adopted. When M is less than the true order of the process, r_M and its estimate include, beside the contribution of the innovation variance, the contribution of the inevitable bias of the model and thus it tends to be significantly large when a too small value of M is adopted. Thus by seeking the minimum of FPE we shall be able to arrive at an autoregressive model of an order which will not be giving a significant bias and at the same time will not be giving a too big mean square prediction error in the above stated sense.

We have simultaneously applied our procedure, a practical version of Anderson's procedure and a modified version of our original procedure to many artificial and practical time series. At present it seems that our original procedure is the simplest and the most satisfactory one giving good results in wide variety of practical situations.

Also we have computed estimates of some of the power spectra by using the Fourier transforms of the estimates of predictors and estimates of the innovation variances obtained by following our FPE scheme and the results have been compared with the estimates obtained by the classical procedure described by Blackmann and Tukey [5]. The comparison has shown that our new procedure is giving extremely good

results suggesting that our estimate is with a good traceability at both high and low power level regions and is well balancing the bias and variance of the estimate, a fact which was scarcely expected by the simple application of the classical procedure.

Experimental applications of FPE scheme with $L=44$ to three real time series with $N=511$, 511 and 524, respectively, have resulted in $M_0=15$, 13 and 5, respectively. Our estimates of the power spectra have shown, though this remains as a subjective judgement at present, a good resolvability which nearly corresponds to that of the classical estimate obtained by applying a hanning type window with truncation point $L=90$ and yet probably with a sampling variability generally smaller than that of the classical estimate obtained with truncation point $L=45$.*) From the series of residuals or estimated innovations obtained by fitting these models we could not find significant trace of deviation from whiteness.

We shall discuss the details of these theoretical and experimental investigations of FPE scheme in subsequent papers.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] H. Akaike, "Some problems in the application of the cross-spectral method," *Spectral Analysis of Time Series* (ed. B. Harris), New York, John Wiley, (1967), 81-107.

[2] H. Akaike, "On the use of a linear model for the identification of feedback systems," *Ann. Inst. Statist. Math.*, 20 (1968), 425-439.

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

[4] T. W. Anderson, "Determination of the order of dependence in normally distributed time series," *Time Series Analysis* (ed. M. Rosenblatt), New York, John Wiley, (1963), 425-446.

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

[6] J. Durbin, "Efficient estimation of parameters in moving-average models," *Biometrika*, 46 (1959), 306-316.

[7] J. Durbin, "The fitting of time series models," *Rev. Int. Statist. Inst.*, 28 (1960), 233-243.

[8] E. Parzen, "Statistical spectral analysis (single channel case) in 1968," Stanford University Statistics Department Technical Report, No. 11, June 10, (1968), 44 pages.

[9] E. Parzen, "Multiple time series modelling," Stanford University Statistics Department Technical Report, No. 12, July 8, (1968), 38 pages.

*) On this point confer the forthcoming paper by the present author entitled "Power spectrum estimation through autoregressive model fitting."