STATISTICAL PREDICTOR IDENTIFICATION

HIROTUGU AKAIKE

(Received Dec. 26, 1969)

1. Introduction and summary

In a recent paper by the present author [1] a simple practical procedure of predictor identification has been proposed. It is the purpose of this paper to provide a theoretical and empirical basis of the procedure.

Our procedure is based on a figure of merit of a predictor, which is called the final prediction error (FPE) and is defined as the mean square prediction error of the predictor. We consider the application of the least squares method for the identification of the predictor when the stochastic process under observation is an autoregressive process generated from a strictly stationary and mutually independent innovations. The identification is realized by fitting autoregressive models of successive orders within a prescribed range, computing estimates of FPE for the models, and adopting the one with the minimum of the estimates.

The statistical characteristics of these estimates of FPE and the overall procedure are discussed to show the practical utility of the procedure. A modified version of this original procedure is proposed, which shows a consistency, as an estimation procedure of the order of a finite order autoregressive process, which is lacking in the original procedure. The notion of FPE is also applied for the determination of the constants of the decision procedure, which was proposed by T. W. Anderson [2] for the decision of the order of a Gaussian autoregressive process, to provide a third procedure.

Performances of the three types of procedures, the original one, a modified version and that of Anderson’s type, are compared by using various realizations of artificial time series. The results show that for practical applications, where the true orders of autoregressive processes would generally be infinite, the original procedure would be the most useful.

Implication of the present identification procedure on the estimation of power spectra will be discussed in a subsequent paper [3].

We shall use the convention of denoting by \( (u(l)) \) the column vector of \( u(l) \) \((l = 1, 2, \ldots, M)\) and by \( v \) or \( (v(l, m)) \) the matrix with \((l, m)\) ele-
v(l, m) \ (l, m = 1, 2, \ldots, M)$. When the dimension $M$ is of special interest we shall add the subscript $M$ and thus $u_M$ and $v_M$ are used for the above $u$ and $v$. The symbol $'$ will be used to denote the transpose of a matrix or a vector.

2. Definition of FPE of a predictor and the statement of the problem

Here we first introduce a general definition of a figure of merit of a predictor. This is defined simply as the mean square prediction error and is called the FPE (final prediction error) of the predictor, i.e., for a predictor $\hat{X}(n)$ of $X(n)$

$$FPE \ of \ \hat{X}(n) = E(X(n) - \hat{X}(n))^2.$$  

In practical situations $\hat{X}(n)$ is given as a function of the recent values of $X(n)$ and the structure or the parameter of the function is determined, or identified, by using the whole past history of $X(n)$. Assuming the dependency of this identified structure on the recent values of $X(n)$ which are to be used to give $\hat{X}(n)$ to be decreasing as the length of the past history used for the identification is increased, we consider the idealized situation where the dependency is completely vanishing. This is equivalent to the situation where the structure of a predictor is identified by using an observation of a process $X(n)$ and, using the structure, the prediction is made with another process $Y(n)$ which is independent of $X(n)$ but with one and the same statistical property as $X(n)$.

When the process $X(n)$ is stationary and the predictor $\hat{Y}(n)$ of $Y(n)$ is linear and given by

$$\hat{Y}(n) = \sum_{m=1}^{M} \hat{a}_M(m) Y(n-m) + \hat{a}_M(0),$$

where $\hat{a}_M(m)$ is a function of $\{X(n)\}$, we have

$$FPE \ of \ \hat{Y}(n) = \sigma^2(M) + \sum_{l=0}^{M} \sum_{m=0}^{M} E(\Delta a_M(l) \Delta a_M(m)) V_{M+1}(l, m),$$

where

$$\sigma^2(M) = E(Y(n) - \sum_{m=1}^{M} a_M(m) Y(n-m) - a_M(0))^2$$

$$= \min_{\{a_M(m)\}} E(Y(n) - \sum_{m=1}^{M} a(m) Y(n-m) - a(0))^2,$$

$$V_{M+1}(l, m) = E\{Y(n-l)Y(n-m)\} \quad l, m = 1, 2, \ldots, M,$$