different θ_0 if the underlying process is not of the fitted structure, although they lead to the same θ_0 if the process is of the fitted structure. Suppose we fit an AR(k) model. Consider, for example, the Gaussian likelihood

$$L_T(\theta) = \frac{1}{T}\log\det\Sigma_\theta + \frac{1}{T}Y'\,\Sigma_\theta^{-1}\,Y$$

and alternatively an M estimate

$$L_T^*(\theta) = \frac{1}{T}\sum_{t=1}^{T}\rho\left(\sum_{s=0}^{k}a_sY_{t-s}\right)$$

$$\text{with}\quad Y_t = 0 \quad\text{if}\quad t \le 1$$

(if σ^2 is unknown the estimate has to be modified, cf. Martin and Yohai (1985)). Then,

$$EL_T(\theta) = \frac{1}{T}\log\det\Sigma_\theta + \frac{1}{T}\operatorname{tr}\{\Sigma\Sigma_\theta^{-1}\}$$

and

$$EL_T^*(\theta) \approx E\rho\left(\sum_{s=0}^{k}a_sY_{t-s}\right).$$

If Y_t is also an AR(k) process then both $EL_T(\theta)$ and $EL_T^*(\theta)$ are minimized by the true parameter value, while in the case where Y_t is not an AR(k) process, $EL_T(\theta)$ and $EL_T^*(\theta)$ are minimized by different values. This means that one has not only to consider the quality of the estimation procedure, but also the "quality" of the estimated parameter.

In the formula below (5.8), Hannan should not compare the estimate $\hat{\Phi}_h(j)$ with $\Phi(j)$ but with the estimated parameter $\Phi_h(j)$ (in the above sense), obtained as a solution of the theoretical counterpart of equation (5.8), and then ask in a second step how good the $\Phi_h(j)$ represent the structure of the series (in fact, the finitely many $\Phi_h(j)$, $j = 1, \cdots, h$, describe the structure of the process "better" than the finitely many $\Phi(j)$, $j = 1, \cdots, h$).

It is obvious that the choice of an estimation procedure doesn't only imply an estimated parameter θ_0 but also an optimal order. The results of Shibata (1980) favoring AIC are only for the case where the parameters are estimated by the Yule-Walker equations. It would be interesting to know whether using other estimation procedures (e.g., robust ones) leads to other order criteria.

ADDITIONAL REFERENCES

DAHLHAUS, R. (1986). Small sample effects in time series analysis. I. Preprint. University of Essen.

MARTIN, R. D. and YOHAI, V. J. (1985). Robustness in time series and estimating ARMA models. In *Handbook of Statistics* (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds.) **5** 119–155. North Holland, Amsterdam.

PARZEN, E. (1983). Autoregressive spectral estimation. In *Handbook of Statistics* (D. R. Brillinger and P. R. Krishnaiah, eds.) **3** 221–247. North Holland, Amsterdam.

Comment

Jorma Rissanen

In this exceptionally lucid and comprehensive survey, Professor Hannan covers essentially all the important ideas in the theory of linear dynamic systems, both deterministic and stochastic, developed during the past twenty years or so. In addition, he describes the more recently introduced new statistical ideas for selecting such models for time series. I was particularly impressed by the apparent ease and elegance with which Professor Hannan managed to explain the rather intricate notions without any undue sacrifice in precision.

I would like to comment on two issues of a general nature raised by Professor Hannan. There have been several attempts to apply the beautiful and deep approximation theory of Adamyan, Arov and Krein in a statistical context for the purpose of obtaining an optimal low order model reduction. As explained in the paper, such a procedure begins with a high order dynamic system, arrived at, perhaps, by applying physical or chemical laws to a process, or by other means. This is then, in the second stage, reduced to a desired complexity, optimally in the sense of minimum distance in a certain norm. The point I wish to make is that because the initial system, which necessarily has the status of a model rather than any "true" system, is nonunique, the end result cannot be assigned any meaningful optimality property. Instead, it is just an optimal approximation of an arbitrary model of the data.

My remaining comments aim to amplify and, perhaps, modify some of the concluding remarks in

Jorma Rissanen is a Member of the Research Staff, IBM Almaden Research Center, 650 Harry Road, San Jose, California 95120.

Sections 4 and 5 made by Professor Hannan about my minimum description length (MDL) principle. Although some of the main analytical results of the predictive and the semipredictive versions of the criterion do presently require Gaussian assumptions, the same is not true of the general criterion nor by any means of the applicability of the principle itself. Furthermore, the MDL principle has more recently been expressed in a new and more satisfactory form (Rissanen, 1987), where the several earlier versions appear as computable approximations of the central notion, the stochastic complexity and which certainly is not restricted to Gaussian likelihoods nor any other ad hoc choices. In fact, an application of the principle amounts to searching for a model class among any that we can think of which permits the largest assignment of a density or probability to the actually observed set of data. The classes may, if desired, be restricted by constraints not determined by the observed data, such as "prior knowledge" or considerations involving the intended application of the model. Hence, it represents a sort of "global" maximum likelihood principle, which is free from any choices with the possible exception of the desired extraneous constraints. The principle is equally well applicable to the selection of models, regardless of the number of parameters in them, as to hypothesis testing, and consequently it is difficult for me to imagine a statistical problem which could not be dealt with in such a manner. In this my thinking appears to be bolder than Professor Hannan's more cautious view, according to which the existence of any generally applicable principle is in doubt.

ADDITIONAL REFERENCE

RISSANEN, J. (1987). Stochastic complexity (with discussion). To appear in *J. Roy. Statist. Soc. Ser B* **49**.

Comment

Ritei Shibata

It is my great pleasure to comment on Professor Ted Hannan's excellent review paper. This paper covers a wide range of topics in stationary multiple time series analysis. My comment is only on a part, "order estimation procedure" for the case $n = 1$. I strongly agree with him that there is no means by which it can be established that AIC is always to be preferred to BIC or the reverse. The admissibility result that any choice of C_T implies admissible order estimation (Stone, 1981, 1982; Takada, 1982; Kempthorne, 1984) supports us.

The results by Shibata (1986a, 1986b) suggest that consistency of order estimation and uniform order of consistency, in terms of mean squared error, of the resulting parameter estimates are not compatible. I therefore also agree with the author that the choice of procedure should be related to the purpose of the analysis. In this respect, I could not understand the derivation of BIC by Rissanen, particularly the relevance of quantizing and coding both observations and parameters. I prefer the original derivation of BIC by Schwarz (1978) from the Bayesian point of view. For a Koopman-Darmois family, the log of the marginal likelihood,

$$\log \int e^{T(y'\theta - b(\theta))}\, d\mu(\theta)$$

is approximated by

$$\sup_{\theta\in\Theta} T(y'\theta - b(\theta)) - \frac{d}{2}\log T,$$

for large T, provided that $\mu(\theta)$ has a density with respect to Lebesgue measure, which is bounded and locally bounded away from zero. The penalty term $-d/2 \log T = \log T^{-d/2}$ follows from the boundedness assumption on $\mu(\theta)$ and the fact that the integration of $\exp(-T\|\theta\|^2)$ over d-dimensional Euclidean space Θ is $(2\pi T)^{-d/2}$. However, if $\mu(\theta)$ is chosen as a measure whose density becomes peaky as T increases, then the penalty is not necessarily of the order of $\log T$. For example, if $\mu(\theta)$ is concentrated on a $1/\sqrt{T}$ neighborhood, the penalty is of the order of constant like as in AIC (Takada, 1982).

One significant difference of AIC from other criteria is in the derivation based on a distance, the Kullback-Leibler information number for the model and the

Ritei Shibata is Associate Professor in Statistics, Department of Mathematics, Keio University, 3-14-1 Hiyoshi Kohoku, Yokohama 223, Japan.