DISCUSSION

Sadanori Konishi

Kyushu University

I would like to begin by congratulating Drs. Rao and Wu for very concise review and clear exposition of model selection. The paper will stimulate future research in statistical model selection and evaluation problems.

Needless to say, model selection and evaluation are essential and of great importance in modelling process in various fields of natural and social sciences. Akaike (1973) introduced an information criterion as an estimator of the Kullback-Leibler measure of discriminatory information between two probability distributions, and a number of successful applications of AIC in statistical data analysis have been reported. Schwarz (1978) proposed a model selection criterion called BIC (Bayesian information criterion) from a Bayesian viewpoint. AIC and BIC are the most widely used model selection criteria in practical applications.

Now by taking advantage of fast computers, we may construct complicated nonlinear models for analyzing data with complex structure. Nonlinear models are generally characterized by a large number of parameters. We know that the maximum likelihood methods yield unstable parameter estimates and lead to overfitting. In such cases the adopted model is estimated by the maximum penalized likelihood method, Bayes approach, etc.

It might be noticed that the criteria AIC and BIC, theoretically, cover only models estimated by the maximum likelihood methods. The problem is: "Can AIC and BIC be applied to a wider class of statistical models?" Konishi and Kitagawa (1996) proposed an information-theoretic criterion GIC which enables us to evaluate various types of statistical models. By extending Schwarz's basic ideas, I will introduce a criterion to evaluate models estimated by the maximum penalized likelihood method.

Suppose we are interested in selecting a model from a set of candidate models M_1, $\cdots$, M_r for a given observation vector $\boldsymbol{y}$ of dimension n. It is assumed that each model M_k is characterized by the probability density $f_k(\boldsymbol{y}|\boldsymbol{\theta}_k)$, where $\boldsymbol{\theta}_k \in \Theta_k \subset R^k$. Let $\pi_k(\boldsymbol{\theta}_k|\lambda)$ be the prior distribution for parameter vector $\boldsymbol{\theta}_k$ under model M_k, where λ is a hyperparameter. Then the posterior probability of the model M_k for a particular data

Sadanori Konishi is Professor, Graduate School of Mathematics, Kyushu University, 6-10-1 Hakozaki, Higashi-Ku, Fukuoka 812-8581, Japan; email: konishi@math.kyushu-u.ac.jp.