Comment

Hirotugu Akaike

Professor Rao bases his discussion of the prediction of future observations entirely on the cross-validation or leave-one-out technique. In particular, he simply dismisses the possibility of using other model selection criteria for the reason that they require some assumptions on the form of the conditional expectation of the future observation given the past observations and an estimate of the conditional variance.

At this point it might be of some interest to note the past popularity in the engineering literature of the cross-validatory approach to predictor selection. Around 1970, the use of the procedure GMDH (group method of data handling) for the modeling of linear and nonlinear systems was advocated by Ivakhnenko (1971). The basic idea of GMDH was to use polynomial regressions and select system models by the cross-validatory approach where separate parts of the data were used for the estimation and checking of a model, respectively.

In spite of the claim that the method could provide a very powerful general method of data handling, apparently the method gained only a temporal popularity. This fact is somewhat suggestive of the future of the cross-validatory approach to the handling of statistical data in general.

Obviously, the cross-validatory approach can provide practical solutions to the problem of model selection when no other methods are available. However, this apparent cure all type of versatility tends to block scientific thinking about the model construction and evaluation. In that sense I see some danger in uncritical acceptance of the cross-validatory approach to model selection.

In the cross-validatory approach, the forms of the predictor and loss function are assumed to be given. It is the choice of these elements that crucially controls the process of realizing a prediction and the proper choice is highly dependent on the type of the distribution of the data under consideration. The adoption of the linear predictor and the mean square loss function is highly indicative of the Gaussian assumption of the distribution of data.

Once this point is recognized, it can easily be seen that the criticism made by Professor Rao under equation (3.8) against model selection criteria other than the cross-validation assessment error (CVAE) is unfounded. In particular, the numerical decompositions of the squared multiple correlations of the three sets of data given by Professor Rao easily allows the application of the AIC criterion to the evaluation of direct regressions given in Table 5 of the paper. By assuming the Gaussian structure, the necessary AIC is defined by

$$\mathrm{AIC}(k) = n \log_e\{S(k)/S(0)\} + 2(k+1),$$

where n denotes the sample size, k the number of regressors and $S(k)$ the residual sum of squares. We have

$$S(k) = S(0)(1 - \rho^2_{[p+1][p,\cdots,p-k+1]}),$$

where $\rho^2_{[p+1][p,\cdots,p-k+1]}$ denotes the squared multiple correlation coefficient of Y_{p+1} on Y_p, Y_{p-1}, $\cdots$, Y_{p-k+1}. By using Rao's notation we have

$$\rho^2_{[p+1][p,\cdots,p-k+1]} = \rho^2_{[p+1][p]} + \rho^2_{[p+1][(p-1)\cdot p]} + \cdots + \rho^2_{[p+1][(p-k+1)\cdot p-k+2,\cdots,p]}$$

and the necessary numerical values are easily obtained from the decompositions of the squared multiple correlation coefficients in terms of partial correlations. The values of AICs are given in Table 1.

The choices realized by minimizing AICs are identical with those by CVAE for the mice and dental data but differ for the ramus data. Obviously the differences of the CVAE and AIC of $k = 1$ and 2 are very small for the ramus data and the discrepancy between the two choices does not seem to be of any practical significance. The difference between the CVAE of $k = 1$ and 2 for the dental data also looks small. However, the difference of the AIC suggests that it is quite significant.

Taking into account the computational simplicity of AIC, compared with that of CVAE, the present result clearly demonstrates that Rao's rejection of model selection criteria such as C_p and AIC is unjustified.

In the present example of simple regressions, both CVAE and AIC are based on the estimates of the prediction error variances. This circumstance obscured the crucial distinction between the selection of the predictor and the evaluation of the underlying model.

The fitting of Professor Rao's last model, the factor analytic type regression (Method 2), requires the clarification of this point. The method fits a multivariate Gaussian distribution to the data by using a proper parametrization and uses the resulting conditional

Hirotugu Akaike is Director General of the Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan.

TABLE 1
AIC values of simple regression predictor (direct regression)

	Previous measurements used						
	Y_1–Y_6	Y_2–Y_6	Y_3–Y_6	Y_4–Y_6	Y_5–Y_6	Y_6	None
	a. Mice data (prediction of Y_7, $n = 13$)						
k	6[a]	5	4	3	2	1	0
AIC	−17.0	−18.0	−19.8	−21.3	−23.1	−24.4[b]	2.0

	Previous measurements used			
	Y_1–Y_3	Y_2–Y_3	Y_3	None
	b. Ramus data (prediction of Y_4, $n = 20$)			
k	3	2	1	0
AIC	−47.3	−49.3[b]	−48.3	2.0

	Previous measurements used			
	Y_1–Y_3	Y_2–Y_3	Y_3	None
	c. Dental Data (prediction of Y_4, $n = 27$)			
k	3	2	1	0
AIC	−28.9	−30.5[b]	−23.0	2.0

[a] Might be too large for the application of AIC for $n = 13$.
[b] Denotes the minimum.

distribution for prediction. The fitting is realized by using the method of maximum likelihood and thus AIC can be applied for the evaluation of estimated models.

In this case, the AIC is not simply defined by the estimated prediction error variance. A model with small estimated prediction error variance may be judged to be a poor fit to the data. In such a situation, by using exp(−0.5 AIC) as the likelihood of an estimated model, we may find a reasonable choice of the predictor. This idea could have been applied even to the example of the simple linear regression predictor of the ramus data. This kind of scientific investigation of the structure of data by models is not possible if we pay attention only to the cross-validatory assessment of the prediction error variance.

I admit that the cross-validatory approach taken by Professor Rao can be useful to provide pragmatic solutions in certain situations. Nevertheless, my conviction is that only through the systematic application of the scientific approach of statistical modeling and evaluation can we expect the future development of statistics as a science.

ADDITIONAL REFERENCE

IVAKHNENKO, A. G. (1971). Polynomial theory of complex systems. *IEEE Trans. Systems Man Cybernet.* **SMC-1** 364–378.

Comment

Seymour Geisser

For approximately the past third of a century, one of a multiplicity of C. R. Rao's intermittent interests has been the development of the theory and methods involved in growth curves. His earlier work mainly reflected his concern with estimation, testing and various covariational structures. Recently he has become more interested in the predictive aspects of this subject.

In this regard there are basically three prediction problems of interest. Assuming we have observed n individuals (vectors) with complete data (over the same components), we may be interested in predicting,

Seymour Geisser is Professor and Director, School of Statistics, University of Minnesota, 270 Vincent Hall, 206 Church Street, S.E., Minneapolis, Minnesota 55455.