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.

for a new vector, values for

(i) all of the components;

(ii) the unobserved subset of components, assuming the rest were observed;

(iii) an extended forecast, i.e., extended to components of the vector for which none of the previous n individuals have been observed (this is applicable to any of the $n + 1$ individuals involved).

The last, and obviously most general situation, requires a more stringent modeling than the first two, precluding the introduction of any new parameters. Our attention here will be restricted to (ii) except to indicate that in the multivariate normal case, Bayesian solutions for an arbitrary covariance structure and for a particular covariance structure introduced by Rao (1967) and termed Rao Simple Structure, were obtained by Geisser (1970) for (i). Bayesian results for (ii) in the framework of a partially observed matrix of vectors were obtained by Lee and Geisser (1972), again in the arbitrary and Rao Simple Structure cases. They also obtained approximate Bayesian solutions when explicit solutions were overly complex or computationally intractable. Rao's paper deals with various methods for providing predictions for (ii). Among other things he discusses using empirical predictive densities, i.e., the conditional sampling density of the unobserved values given the observed values and the parameters where the latter have been replaced by estimates. Lee and Geisser (1975) tended to call procedures like these approximations or approximate Bayes. There is, as Rao concedes, in the frequentist framework, "no appropriate theory for taking the estimation errors of the parameters into consideration especially when using the same parametric estimates repeatedly." Actually this is even true for a single prediction as can readily be seen in the following example. Let $X_1, \cdots, X_n, X_{n+1}$ be $N(\mu, \sigma^2)$ and the first n are observed while X_{n+1} is to be predicted. For example, using the (empred) normal distribution, $N(\bar{x}, s^2)$, where $\bar{x} = n^{-1}(x_1 + \cdots + x_n)$ and $(n-1)s^2 = (x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2$ will yield confidence intervals for X_{n+1} even for moderate sample sizes that are far too narrow for the stated confidence level. Considerable improvement can be made by using the approximate $N(\bar{x}, s^2(1 + n^{-1}))$ and of course exact confidence (and Bayesian predictive) intervals are obtainable from Student's t distribution with $n - 1$ degrees of freedom, where

$$t_{n-1} = \frac{X_{n+1} - \bar{X}}{s\sqrt{1 + n^{-1}}}.$$

Hence, if $N(\bar{x}, s^2)$ is the empirical (Bayes) predictive distribution a la Rao, I would prefer the approximate Bayesian predictive distribution $N(\bar{x}, s^2(1 + n^{-1}))$ at least for predictive regions. There is obviously no difference between the two for point predictors, which is mostly what Rao deals with here, because the density has negligible relevance as long as the "centers" and general shapes are similar.

Section 2.3 deals with situations where the form of the actual sampling densities are unknown. Rao here proposes the sample reuse technique. Now, given the many articles that appear in a variety of journals that publish statistical research on growth curves, it is not easy to be aware of all of them. Hence, I must bring to Professor Rao's attention that the sample reuse approach has already been proposed for growth curves with solutions for two different kinds of predictive functions. The first involves a combination of predictors, one from the n vectors and one from the partially observed vector (Geisser, 1980a), and the second a general regression predictor with various special cases involving Markovian constraints (Geisser, 1981). In fact the methods were illustrated on the ramus data in the latter paper for predicting the last observation. In this case the "best" linear predictor turned out to be the constrained decreasing Markovian regression predictor which gave about the same discrepancy or (CVAE) as Rao's empirical Bayes predictor using estimates of the covariational parameters. Similar discrepancies for this data set were obtained by Lee and Geisser (1975) using a serial correlation structure and by Fearn (1975) using a two stage hierarchical Bayesian approach somewhat similar to Rao's factor analytic model.

When the predictive sample reuse discrepancy or (CVAE) was introduced (Geisser, 1974, 1975a; Lee and Geisser, 1975), its principal motivation was its use as a comparative measure for various predictors and only with extreme caution as an actual estimate of error. A discussion of this point appears in Geisser (1975a, page 322) indicating among other things, what error it is we are trying to assess, and the problems in using various forms of the discrepancy as the error estimate, particularly because of algebraic constraints in repeated use of the same data. Hence, the claim that the discrepancy is a reliable error estimate should be treated with considerable reserve.

Further discrepancies other than squared error are also useful. Absolute error and the number of times one predictor is closer to the actual value than another as well as the empirical distribution of the actual differences are also valuable (Lee and Geisser, 1975; Geisser, 1975b).

One additional point that is worth mentioning is that sample reuse procedures are also of value for estimating hyperparameters in empirical Bayes procedures when maximum likelihood and/or method of moment estimators are computationally infeasible or yield poor results (Geisser, 1980b).

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation Grant DMS-8601314 and National Institutes of Health Grant GM-25271.

ADDITIONAL REFERENCES

FEARN, T. (1975). A Bayesian approach to growth curves. *Biometrika* **62** 89–100.

GEISSER, S. (1970). Bayesian analysis of growth curves. *Sankhyā Ser. A* **32** 53–64.

GEISSER, S. (1974). A predictive approach to the random effect model. *Biometrika* **61** 101–107.

GEISSER, S. (1975b). A new approach to the fundamental problem of applied statistics. *Sankhyā Ser. B* **37** 385–397.

GEISSER, S. (1980a). Growth curve analysis. In *Multivariate Analysis Handbook. I. Analysis of Variance* (P. R. Krishnaiah, ed.) 89–115. North-Holland, Amsterdam.

GEISSER, S. (1980b). Predictive sample reuse techniques for censored data (with discussion). In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 430–468. University Press, Valencia.

GEISSER, S. (1981). Sample reuse procedures for prediction of the unobserved portion of a partially observed vector. *Biometrika* **68** 243–250.

Rejoinder

C. Radhakrishna Rao

For ready reference, the problem considered in the paper is the following. We have observations (U_i, W_i), where U_i is a p vector of measurements taken at p time points and W_i is the measurement taken at a future $(p+1)$th time point, on $i = 1, \cdots, n$ individuals drawn from a population S. Another individual drawn from S provides the first p measurements U_c, and the problem is to predict the $(p+1)$th measurement W_c on the individual.

What is relevant in a problem of this kind is the conditional (predictive) distribution of W_c given U_c,

$$P_{\text{pred}}(W_c \mid U_c, \psi), \tag{1}$$

with respect to some *reference* population, where ψ is a parameter specific to the reference population. One choice of the reference population is S itself. However, when ψ is unknown, we have two possibilities. We may estimate ψ by $\hat{\psi}$ from the available data

$$(U_i, W_i), \quad i = 1, \cdots, n, \quad \text{and} \quad U_c \tag{2}$$

and consider an estimate of (1),

$$P_{\text{empred}}(W_c \mid U_c, \hat{\psi}), \tag{3}$$

as the basic conditional distribution. An alternative is to consider S as a member of a super population generated by a prior distribution on ψ, in which case the relevant distribution is

$$P_{\text{Baypred}}(W_c \mid U_c) \tag{4}$$

obtained by integrating (1) with respect to the posterior distribution of ψ given the observed data (2). On the other hand, we may wish to consider the current individual's observations (U_c, W_c) as arising from a stochastic process *specific* to the individual. In such a case the empred (3) is defined in terms of $\hat{\psi}$ estimated from U_c alone and the Baypred (4) is obtained by choosing a prior on ψ and computing the posterior distribution based on U_c alone. The second possibility of considering an individual separately is specially recommended when on the basis of an initial examination of data, the measurements U_c are found to have an unusual pattern different from those of $U_1, \cdots, U_n$.

The theory as developed in Section 2 of the paper and outlined above is complete in itself although its practical applications involves various issues that I would like to discuss on the basis of the comments made by the discussants of my paper.

DATA AND CROSS-EXAMINATION OF DATA

For illustrative purposes I have chosen three real data sets, which are well documented and which have been studied by a number of authors for predictive purposes. I thank Izenman for giving some details about the mice data that will be helpful to future investigators. I have made the necessary corrections regarding the original source of the dental data based on his comments. In my analysis of the mice data, I omitted the measurements on one mouse (not reported in Table 2, but can be found in Izenman's comments), which looked different from the others and whose weight actually decreased at the end. Izenman asks what effect it would have had on my results if this mouse had been retained in the data set. I have deliberately chosen my reference population as the set of mice that generally exhibit an increase in growth at all time points and derived the appropriate prediction