BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, Calif.

GNANADESIKAN, R. and KETTENRING, J. R. (1984). A pragmatic review of multivariate methods in applications. In *Statistics: An Appraisal* (H. A. David and H. T. David, eds). Iowa State Univ. Press, Ames.

JONES, M. C. and SIBSON, R. (1987). What is projection pursuit (with discussion)? *J. Roy. Statist. Soc. Ser. A* **150** 1–36.

KETTENRING, J. R. (1971). Canonical analysis of several sets of variables. *Biometrika* **58** 433–451.

Comment

S. James Press

I thoroughly enjoyed Mark Schervish's review of multivariate analysis, a subject that has been near and dear to me for many years. The review was written in a very light, free-flowing format that made it interesting and pleasant reading, while at the same time the points made were usually deep and insightful. I will comment generally on the Schervish review by offering my own perspectives on multivariate analysis, and then I will give a few brief specifics on his review. All comments will necessarily be brief but indicative of directions in which the field is moving.

A COMPARISON OF CLASSICAL AND MODERN MULTIVARIATE ANALYSIS

I would like to distinguish "classical" multivariate analysis (CMA) from "modern" multivariate analysis (MMA). I will do so on the basis of how they compare on various (randomly ordered) characteristics.

1. *Distribution theory.* In CMA, the theory derives largely from the multivariate normal and Wishart distributions. It also is concerned with the study of the distribution of latent roots of random matrices.

In MMA there is increasing focus on non-normal inference and distribution theory. It is based upon nonabsolutely continuous distributions, such as the mixed discrete and continuous distributions, or the mixed singular and absolutely continuous distributions, exemplified by the multivariate exponential distribution. Focus has shifted away from the latent root distributions because the models that require them have languished for lack of use.

2. *Estimation.* In CMA, the emphasis was on MLE and moment estimation. In MMA there has been a substantial shift in emphasis to Stein-type estimation, empirical Bayes estimation and Bayes estimation. This shift is natural with the improvements in multidimensional estimation achievable by using higher dimensional shrinkage estimators (for dimension greater than two) and by introducing subjective prior information into a problem in a formal way.

3. *Noncentral distributions.* In CMA, power calculations demanded the development of various noncentral distributions, such as the noncentral Student t and noncentral F distributions, the Hotelling T_0^2 distribution and the noncentral Wishart distribution, which arose in coefficient estimation for simultaneous equation systems.

In MMA a unified theory of noncentral distributions has developed around the theory of hypergeometric functions of matrix arguments, zonal polynomials and generalized distributions.

4. *Distribution theory of sample estimators.* CMA was deeply concerned with the distribution theory of sample estimators, although the introduction of the "bootstrapping" technique (Efron) and the technique of simulating complicated multivariate distributions by simulating functions of known distributions (Kass) have liberated modern multivariate analysts from their former distributional burdens of having to develop the distributional theory of complicated multivariate distributions.

5. *Discrete multivariate analysis.* CMA dealt with discrete data by means of traditional contingency table analysis, i.e., estimating cell probabilities by MLE.

MMA is more concerned with analyzing discrete data by using multivariate log-linear and logistic models; by using models involving ordered categories and by using both dimensions of a contingency table simultaneously to study categorical data, by means of "correspondence analysis."

6. *Factor analysis.* CMA was wary of the factor analysis approach and was concerned with centroid solutions, rotations, maximum likelihood factor analysis and exploratory factor analysis (rather than confirmatory).

MMA has become more accepting of the factor analysis approach. Today the emphasis has shifted to confirmatory factor analysis, Bayesian factor analysis methods and to nonparametric factor analysis

S. James Press is Professor, Department of Statistics, University of California, Riverside, California 92521.

methods, such as MINRES and least squares factor analysis, methods that permit analysis of problems involving discrete variables.

7. *Principal components, canonical correlations, MANOVA, MANOCOVA and other linear models.* Principal components analysis and linear models were treated as separate models in CMA. In MMA these various models are all treated simultaneously, through covariance structures analysis by using LISREL in its various forms. Moreover, analysts are becoming increasingly concerned about lack of independence in the data when they do principal components analysis, and canonical correlations analysis, and are correcting for it. The multicollinearity problem of multivariate regression in CMA is approached as ridge regression in MMA.

8. *Latent variables.* Latent variables played a minor role in CMA, except for their appearance in models of factor analysis and latent structure analysis. In MMA we see a fusion of the latent variable models of psychometrics and econometrics, with MIMIC (Jöreskog) and other models of Bentler, Goldberger and Jöreskog, in simultaneous equation systems, as well as in psychometric factor analysis models. Biometric and sociometric path analysis models have now been enriched with latent variables.

9. *Probit and logit models.* Probit, logit and angular transformation models for analyzing discrete data have been modified to include latent variables in MMA (Muthen). These models permit a richer structure than was possible in CMA.

10. *Data analysis.* The general methodological approach of CMA encouraged very little use of data analysis in its modern sense. MMA encourages the use of graphical methods and computers for looking at high dimensional data. "Biplot analysis" (Gabriel) has been available since 1971 (and higher dimensional versions of it have become available more recently).

Modern multivariate analysts study higher order cumulants; they use multivariate "Box-Cox transformation methods" for transforming data prior to linear modeling; and they may use "projection pursuit" methods (Friedman) to carry out a regression, or to estimate a density, on their personal computers at home. "Box plots" and stem and leaf diagrams (Tukey) are becoming common statistical practice, and such pictures are printed out as an integral part of many software packages. Modern multivariate regression analysis would not be complete without a study of "influential observations." Modern multivariate data analysts think in terms of jackknives, bootstrapping, regression trees and other computer-intensive clustering algorithms. They also use "multidimensional scaling," "conjoint analysis" and "profile analysis."

11. *Multivariate stochastic processes.* Most stochastic processes analyzed by statisticians in CMA were univariate time series, and they were traditionally analyzed by using frequency-based spectral analysis.

MMA focusses much more on Box/Jenkins methods of analyzing multivariate time series in the time, rather than frequency, domain. Interest has shifted to study of ARIMA processes. Interest among statisticians has also developed in the study of spatial stochastic processes. Spatial structure has involved modern multivariate analysts in variogram analysis, kriging, Markov random fields and Gibbsian distributions.

12. *Estimation vs. prediction.* CMA was concerned largely with estimation and hypothesis testing of unobservable parameters. MMA is more concerned with prediction of observables. Methods for implementing the more modern approach that have now come into common usage are sample reuse and cross-validation (Geisser and Stone).

13. *Estimation vs. modeling.* CMA was more concerned with estimation than with modeling. Focus was on improving estimation techniques by finding unbiased estimators, and estimators with smaller variance, or estimators with smaller mean squared error if the estimator was biased. MMA is more concerned with model comparison, and with elimination of poor models in favor of ones that predict observables better.

14. *Analytical vs. numerical solutions.* Largely because of the minimal availability of appropriate computer systems for statistical analysis, CMA was concerned mostly with the development of analytical solutions to problems. Computers modified that attitude of course. With the advent of sophisticated computer software programs, modern multivariate analysts have vigorously pursued numerically based, computer-dependent solutions. Moreover, entirely new multivariate methods have been developed that are computer intensive and nonanalytical in form ("projection pursuit" and its successors).

15. *Bayesian multivariate analysis.* Bayesian multivariate analysis has developed substantially over the last few decades. The procedures that have been developed have involved hierarchical models (Lindley and Smith), computer-assisted assessment of subjective probabilities and utilities (CADA) and the development of a complete armamentarium of Bayesian methods to handle the traditional multivariate models of classification, factor analysis, regression, MANOVA, MANOCOVA, latent structure analysis, canonical correlations analysis, simultaneous equation systems, etc.

Conclusion

Schervish reviewed Anderson's second edition and Dillon and Goldstein's multivariate analysis books. From the point of view of CMA vs. MMA, the former book leans more heavily toward CMA than does the latter, but neither one really exemplifies MMA.

SPECIFIC COMMENTS ON THE SCHERVISH REVIEW

1. Where Schervish discusses the development of power functions in multivariate analysis, it would have helped a bit if power were discussed in terms of how power functions are normally used in multivariate statistical practice, namely, from the viewpoint of someone trying to determine sample size for an experiment involving multiple (correlated) outcomes. How is this sample size selection problem best solved? There is not much discussion of this kind of question in the two books reviewed.

2. The author refers to Anderson's discussion of the Scheffé procedure (it was extended to the multivariate case by Bennett) for dealing with the multivariate Behrens-Fisher problem of testing equality of means in two normal populations with unequal variances, when he says, "Data is discarded with a vengeance." The issue here is that if we have M observations on one population, and N observations on another, and $M < N$, Scheffé suggests that we randomly match M observations from the two populations and discard the remaining $(N - M)$ in the matching process. Actually, all of the observations in each of the populations should be used to estimate each of the variances. If M and N are large there is of course no problem in ignoring $(N - M)$ observations in the testing. The only occasion when a problem arises is in the case of M and N small, and $M \ll N$. From the Bayesian point of view these types of issues never arise, at the tradeoff cost of having to develop prior information for the parameters.

3. Schervish suggests that "one other unfortunate feature of Section 5.5 is ... This test is simply not another example of the type of test proposed for the Behrens-Fisher problem." Here, Anderson suggests that we can test the hypothesis that two normal subvectors have equal means (with unequal covariance matrices, so that it is a Behrens-Fisher type problem) by forming the difference in the sample mean vectors, "and this statistic is most relevant to $(\mu^{(1)} - \mu^{(2)})$" (Anderson, page 178). This is a special case of the Scheffé/Bennett approach discussed in Item 2, above, for the case of $M = N$, where the two-sample problem is reduced to a one-sample problem by subtraction of the sample means.

4. Schervish's suggested alternative to a second principle of classical inference is a bit harsh. Although "unbiasedness" is not a particularly relevant property for situations in which we are going to have to estimate only once, or only a few times, and although unbiasedness is a property that violates the "likelihood principle," I believe that most any reasonable statistician who is in the position of having to recommend an estimator that will be close to the true value on the average, over many estimations of the same quantity, would find unbiasedness a compelling property when taken in conjunction with the requirement of small variance.

Summary

In summary, the review of these important books on multivariate analysis by Mark Schervish not only provides a helpful perspective from which to appreciate these contributions to our field, but also, is refreshing and entertaining.

Rejoinder

Mark J. Schervish

I wish to thank the discussants for taking the time to carefully read the review and offer their own views on the topics covered. They have each made it more informative and useful for the interested reader. Because some of the comments of the authors of the two books reviewed are in the way of rejoinder to my review, I will refrain from offering further commentary on those remarks. Much of the discussion provides the readers with brief overviews of areas that I failed to mention in my review. As my review already is a comment, at great length, on the work of many people, I will keep my comments on the discussion brief.

Because Professor Anderson's comments are almost entirely concerned with my review of his text, I will let him have the last word on the matter. I will thank him, however, for bringing to my attention part (b) of Problem 3 in Chapter 11 of his book, which indeed does suggest the predictive interpretation of principal components. A further suggestion of this interpretation appears in the paper of Kettenring (1971), whom I also thank for the reference.

Some of the discussants mention projection pursuit as a computationally intensive multivariate method that I did not discuss. Professor Goldstein remedies