Comment

Pranab Kumar Sen

In this review of multivariate analysis, Mark Schervish has indeed provided a detailed account of some of the relevant developments with particular reference to their coverages in the second edition of T. W. Anderson's classic text and a somewhat unrelated book by W. R. Dillon and M. Goldstein, both published in 1984 (almost simultaneously) by John Wiley.

It is undoubtedly true that a generation of (mathematical) statisticians (specially, in the North American continent) has been raised on the classical (1958) textbook of T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. It contains a (totally) likelihood principle-motivated, intermediate level, sound theoretical treatment of standard multivariate statistical analysis, mostly in the setup of multivariate normal distributional models. I was not at all surprised to observe that Professor Anderson had decided to retain the same structure in the second edition, albeit with some updating and the addition of some new materials in some specific sections. These new materials, of course, relate to developments taking place after 1958. Some of these specific additions and alterations have been discussed in detail in this review by Professor Schervish and also in a recent review of mine (1986) where a comparatively wider scenario of a battery of some contemporary textbooks in multivariate statistical analysis has been considered.

Multivariate analysis has been one of the most prolific areas of fruitful research (enriched with genuine applications) during the past fifty years or so. The impact of anthropology, biological sciences, genetics, psychometry, econometrics and physical and social sciences on the early developments in the theory part of multivariate analysis has been so overwhelming that a greater part of this basic theory leading to potentially applicable methodology actually resides in the broader realm of these applied sciences. It would not be proper to treat multivariate analysis in isolation without paying due emphasis on this variety of applications it has encompassed over time. At the same time, multivariate statistical analysis has also been a challenging area of research for mathematicians and mathematical statisticians as well, although the end point has often been quite dehydrated and obscured in the realm of abstractions. A balance between sound theory and applicable methodology is therefore a key factor in defining the proper domain of multivariate statistical analysis. The past twenty five years have witnessed a phenomenal growth in the fruitful applications of multivariate statistical analysis in medical studies and clinical trials, although the standard multinormal distributional models may not be very appropriate in this context. In a variety of other problems (for example, in sociological studies), often, the data come in the form of categorical responses, and the utility of discrete multivariate analysis, to cope with such problems, does not need any further elaboration. Graphical techniques, diverse computational methods and a variety of other developments (mostly taking place during the past two decades) have led to the development of multivariate analysis on a far greater basis than envisaged in either of these books.

Faced with this wider spectrum of theory and methods in multivariate statistical analysis and their potential applications in diverse disciplines, the basic issue for the statisticians is to choose a right track for the dissemination of the theory for potentially applicable methodology, and this issue is indeed very intricate in nature. Is it possible to include "A to Z" of multivariate statistical analysis into a single volume that could be covered in a semester or two in a graduate program in (mathematical) statistics? My flat answer to this query is no. Has either of the two books under review been written as a primary reference book for the entire spectrum of multivariate statistical analysis? Again, the answer is flatly no. Keeping in mind these basic difficulties, a variety of textbooks has emerged (mostly, during the past ten years or so) in the general area of multivariate statistical analysis with more intensive treatments of some specific aspects rather than the whole. I believe that this diversity of the treatments of the competing books adds more strength to the total fabrication of multivariate statistical analysis. As such, while reviewing multivariate (statistical) analysis, one needs to pay some attention to the coverage as well as level of theoretical sophistication of any book intended for inclusion in the review.

T. W. Anderson's second edition is certainly a strong contender in this domain. It deals with the main stream of developments in the theory (without ever crossing the fence to the mathematical abstractions that are mostly of academic interest only), and it builds up a coherent methodology that has already shown the stamina for good applications. The (minor)

Pranab Kumar Sen is Cary C. Boshamer Professor of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514.

shortcomings of this book (in coverage and presentations), as have been discussed in this review by Mark Schervish (and also mentioned in Sen (1986)) should not be overemphasized. I have found it difficult to cover (even in a two-semester graduate level course) more than what has been treated in Anderson's 675 pages, and time permitting, I would rather provide additional references to other contemporary textbooks where some of these additional topics (not included in Anderson's) have been dealt with in more depth. A more effective purpose may not be served by aiming at a text of more than 800 pages to cover these additional topics in greater detail. Moreover, for research workers in these active areas, a more natural course would be to look at some of the more specialized books dealing with these topics in depth; in any case, for a proper understanding of these materials, the groundwork provided in Anderson's second edition looks ideal.

I found it little bit out of the way to bring in the other multivariate book by Dillon and Goldstein, even, on a complementary stand to the second edition of Anderson! Are we comparing oranges and apples in a way? Anderson's text is rightfully in the "probability and mathematical statistics" series, whereas the other one is in the "applied" series; this basic distinction is also revealed clearly in the basic treatment of the two books. One has a sound theoretical (but not abstract) structure, whereas the other is mostly deprived of the basic theory and deals exclusively with the application-oriented methodology. I pondered through some 500 pages of Dillon and Goldstein and wonder: Where is the beef! Although this book may qualify as a text for an applied multivariate analysis course (in psychometry, for example), it fails to meet the need for the (mathematical) statisticians who would like to grasp the theory before incorporating the relevant methodology in pertinent applications! Moreover, I may not agree with Mark Schervish that the material in the Dillon and Goldstein text, albeit complementary to Anderson's second edition, covers a greater domain of modern statistical theory and methods in multivariate analysis not treated in Anderson. A closer look into the diversity of the theory and applications in this area may reveal an altogether different picture.

Multivariate analysis has been, fortunately, an area of active interaction between theory and methods leading to genuinely applicable methodologies. From the theoretical perspective, there have been some interesting developments that are not reported adequately in Anderson's second edition, and, of course, not in the least in Dillon and Goldstein (in fact, these developments are mostly outside their domain). Anderson has been rather brief in the treatment of Bayesian, empirical Bayesian and related Stein-Rule estimation procedures. He has also not updated the treatment of simultaneous confidence sets in multivariate analysis of variance models. Further, his treatment of variance component models is most inadequate. Well, let us look at the Dillon and Goldstein, and examine how have they fared in either of these aspects? To my surprise, I found very little mention of these topics in Dillon and Goldstein. This clearly shows that the union of the two domains in the two books under review by Mark Schervish does not cover the entire domain of modern multivariate analysis. I have read with considerable interest Wijsman's (1984) review of the two books by Eaton (1983) and Muirhead (1982), and I have no doubt in my mind that these two books (along with Giri (1977) to some extent) have done an excellent service to theoretical statisticians by providing in detail some of the basic theory having a fundamental role in multivariate analysis. The treatment of topological groups and invariant measures and their applications in Eaton (1983) as well as the treatment of zonal polynomials and their applications (in the distributional problems relating to commonly used multivariate statistics) in Muirhead (1982) both deserve a lot of appreciation from the theoretical statisticians. I would be surprised if the domain of modern multivariate analysis were defined without embracing these important developments. From the applied perspective, we may go to the other extreme: multivariate data analysis. There are some novel developments reported in Gnanadesikan (1977) and in a more updated fashion in Lebert, Morineau and Warwick (1984). Dillon and Goldstein have spent nearly fifty pages on this aspect. However, their treatment lacks the flavor of completeness as well as updatedness! Pattern recognition and cluster analysis form another area of active research interest having an immense scope for fruitful applications. Multiple correspondence analysis and automatic classification-clustering techniques are becoming more adaptable through the development of their relevant theory. I would naturally place emphasis on some of these aspects in an applied multivariate analysis course, and I am not that clear about the coverage and treatment of these in Dillon and Goldstein. The psychometricians may find it more useful to look at the text by Dillon and Goldstein where an up-to-date picture has been neatly presented; however, the omission of the relevant theory part may seriously hamper a proper comprehension of the subject matter. In this respect, I have found a more balanced treatment of theory and methodology (deemphasizing the classical multinormal models) in Takeuchi, Yanai and Mukherjee (1982). I understand that these authors are working toward a vastly revised second edition that would eliminate some of the drawbacks of the first edition and would make it even more attractive.

Discrete multivariate analysis has also witnessed a phenomenal growth of literature in the past two decades and there are by now several texts covering the main developments in this area in an uptodate fashion. However, there appears to be a different picture with the development of mixed multivariate models involving partly continuous and partly discrete variables. These models arise often in applications and it would be a natural expectation to see an adequate treatment of the theory in a text. This may be a glaring omission in Dillon and Goldstein! Another important area with an outstanding growth of literature in the past fifteen years is the so-called variance component models. Frankly, I expected a more detailed treatment of this important topic in Anderson's second edition, and I am to a greater extent disappointed to see an inadequate treatment of this topic in either of the two books reviewed by Mark Schervish. I would like to make a specific reference to the forthcoming book of Rao and Kleffe (1987) for an in-depth coverage of this important area. I expect a significant amount of applications of these models in various applied areas.

As I tend to draw an overview of modern multivariate statistical analysis, more and more, I feel the need for robust (if not nonparametric) methods. Although some of these methods (mostly, in the context of simple MANOVA models) have been treated adequately in some contemporary textbooks, I have no doubt in my mind that in the coming years, there will be a far reaching impact of this vital area in multivariate analysis.

To summarize, let me congratulate Mark Schervish for a job well done. In principle, I would have argued in favor of a modified title "A Review of Two Texts in Multivariate Analysis." The area of multivariate statistical analysis is indeed too vast to be covered entirely by these two (or, as a matter of fact, by any two) texts. However, Anderson's second edition will naturally help us in identifying the other pockets where an equally sound and lucid treatment of the theory (and methodology) should be developed in the form of a text, and once this has been accomplished, we are all set to close the whole area in the form of two texts. Until then, the second edition is a major step in the right direction.

ADDITIONAL REFERENCES

EATON, M. L. (1983). *Multivariate Statistics: A Vector Space Approach.* Wiley, New York.

GIRI, N. C. (1977). *Multivariate Statistical Inference.* Academic, New York.

LEBART, L., MORINEAU, A. and WARWICK, K. M. (1984). *Multivariate Descriptive Statistical Analysis.* Wiley, New York.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory.* Wiley, New York.

RAO, C. R. and KLEFFE, J. (1987). *Estimation of Variance Components and Their Applications.* North-Holland, Amsterdam.

TAKEUCHI, K., YANAI, H. and MUKHERJEE, B. N. (1982). *The Foundations of Multivariate Analysis.* Halsted, New York.

Comment

R. Gnanadesikan and J. R. Kettenring

In our experience, most statistical problems that arise in practice are genuinely multivariate in character. This is almost surely as true in other settings as it is in the telecommunications business that we work in. A recent literature search (Gnanadesikan and Kettenring, 1984) covering seven disciplines over the period 1965 to 1982 turned up 15,000 articles that involved multivariate methods.

It is natural, therefore, to expect that new books on the subject, such as those by Anderson and by Dillon and Goldstein, as well as comprehensive reviews, such as that of Schervish, will have a wide audience. However, our intention in this commentary is not so much to critique either the books or the review as it is to bring out some of our own views on multivariate data analysis.

In outlook, if not detail, these overlap with views of Schervish who makes many telling points about the state of multivariate analysis. The best known and most frequently used of the classical methods have not always served well and often leave the user with the question "What have I really learned about my data and how sure can I be about it?" Much of the elegant theory is of little practical value. Standard multivariate hypothesis tests, which have been so extensively developed (see Schervish's comments in

R. Gnanadesikan is Assistant Vice President of the Mathematical, Communications, and Computer Sciences Research Laboratory and J. R. Kettenring is Division Manager of Statistics and Economics Research at Bell Communications Research, 435 South Street, Morristown, New Jersey 07960.