Comment

George Box

I am heartened by this paper because the author has traveled some distance along a road which I believe might ultimately lead to the salvation of departments of statistics. However, I do not think he has traveled far enough. Our difficulties, I believe, stem from the fact, lost sight of long ago, that the *raison d'être* for statistics is its role as the catalyst of scientific investigation. As is the case for many other subjects, mathematics is one necessary tool for its effective use, but it is only that, and any statistician who actually practices his art must possess many additional resources. Our present lamentable situation and somewhat dubious future arises from the fact that the mathematical tail has been allowed to wag the statistical dog for far too long. This is evident even in this enlightened paper in which people who actually apply statistics are clearly still regarded as poor relations. No doubt they should be properly grateful to be allowed "to communicate these problems along with their own attempts at solutions to their statistical [read mathematical?] colleagues." But will these colleagues be interested in the problems? Will they be prepared to take the time to really understand them? And why would the "joint appointments ... with degrees in statistics" not be able to solve them for themselves? After all appointments of this kind involving, for example, Jerry Friedman, Brad Efron, Don Rubin, John Tukey and George Tiao have been responsible for many of the genuinely new ideas in present day statistics (Box, 1990), an outcome that historically we should expect (Box, 1984).

I am not sure what the originators of the idea of statistics departments intended. But I think that the built-in mathematical bias of many such departments and of much that we are presently teaching is not innocuous; it is in fact antiscientific. I am at one with the author in believing that we need to change what we are presently doing.

There is another substantive issue which I would like to raise. This concerns the fatal fascination of the word "unity." Unity in many things is desirable but we should not be trying to impose "oneness" on a situation where "twoness" is of the essence. For example, a visitor from outer space who was an overzealous believer in the principle of parsimony might have difficulty in understanding human behavior if he could not accept that there were two sexes rather than one on this planet.

Why I should think that this has anything to do with the Bayesian-frequentist deadlock was discussed some time ago (Box, 1980, 1983); but perhaps it is worthwhile to outline the argument again. Progress in scientific investigation occurs as a result of a deductive-inductive iteration which employs *two* distinct types of statistical inference. The first of which may be called *estimation* and the second *criticism.* The first, *estimation*, involves a process in which information from the model is *combined* with data, but which alone can tell us nothing about whether the model and data are consonant. A second type of inference, *criticism*, involves a process in which the information from the model and from the data are *contrasted.* It can spark off the inductive process of appropriately modifying our ideas and models and is particularly important because it is the only point in the iterative cycle at which genuinely new ideas are injected. These two processes are, I believe, as different as addition and subtraction and consequently require different treatment. Estimation is concerned with the possible different values of the set of parameters consistent with the one fixed set of data actually obtained. It is conditional on the assumption that the model is true. No idea of repeated sampling is involved in this formulation and inferences should I believe be made using Bayes' theorem or, for the faint-hearted, the likelihood function. Criticism, however, involves the question, "Is is plausible that data of this kind could have occurred *at all* given the postulated model?" This concept seems to me to call for consideration of the plausibility of data that actually occurred in relation to a reference set of data that might have occurred given that the model was true. This is a frequentist idea requiring the concept of repeated sampling and the consideration of other possible sets of data which did not actually occur.

I am reinforced in my beliefs by the difficulties that frequentists have with estimation and that Bayesians have with criticism. For example, some Bayesians tell us that we can do criticism by listing all the models that might describe the system under study, calculating their relative posterior probabilities and selecting the model(s) that look most probable.

But ideas sparked off during the course of an investigation, but *not thought of initially*, are frequently the key to successful problem-solving. Specifically,

George Box is Director, Center for Quality and Productivity Improvement, 610 Walnut Street, Madison, Wisconsin 53705.

suppose we have, say, a complex chemical system for which k kinetic models are considered, all of which happen to be totally wrong. Suppose that one of these wrong models nevertheless produces a posterior probability say 20 times as large as its nearest competitor. It can still be true that residuals from this best wrong model will be many times their standard deviation and so on a frequentist's argument will indicate lack of fit. Consequent study by a subject matter specialist of the pattern of these residuals and of appropriate diagnostic checking functions could suggest a different model or class of models not previously conceived of. This use of Bayes' theorem for the purpose of criticism would thus seem to abort the scientific process. Arguments of equal force can be made against frequency theory when used for estimation. The scientist and engineer are rightly suspicious of statistical procedures that seem to hamstring their creativity.

Comment

A. P. Dempster

I am an active supporter of the main thrust of Glenn Shafer's remarks, both on the need to restore the subjectivist interpretation of probability to a central position that forms a unified whole with frequentist interpretations and on the need to reform and revitalize departments of statistics by redesigning and strengthening ties to less mathematically oriented disciplines. The near term health and long term survival of statistics as an independent academic discipline depend on departmental policy discussions, for example on curriculum, recruiting and promotion, that place these items high on agendas.

Views may differ on details and strategies. For example, I see the main ideological split lying not between frequentism and subjectivism, where as Glenn says the debate long ago grew stale. Rather it lies between advocates of a nearly exclusive emphasis on methods, and proponents of formal reasoning about uncertainty, whether in the spirit of R. A. Fisher, or in the similar but more recent style of Bayesian or belief function modeling and inference that appears to me to be the obvious and natural way to do statistical science. As with the related but narrower differences between frequentists and subjectivists, there is in fact a fundamental unity between methods and reasoning, in the sense that the former are vehicles for the latter. What does not fly, in my opinion, is the widespread tacit assumption that statistics is mainly about choosing and applying correct or good methods. We need to learn how to understand and teach a more active logic of the processes of doing statistics, including formal probabilistic reasoning about uncertainty.

Glenn is on target when he argues that joint appointments based on the model of statistical technology flowing from core departments to users is rapidly losing viability as user fields become increasingly technical and able to produce their own technologists. In a sense, we have succeeded too well at that game and must use our wits to stay several steps ahead of the competition. As Glenn suggests, emphasis on mathematical statistics, however high its quality, is unlikely to produce the required innovations. I do, however, see a long term market niche for mathematically talented individuals able to match understanding of empirical phenomena with formal mathematical representations of both the phenomena themselves and the scientist's uncertain knowledge of the phenomena. Such work concerns not statistical generalities, but specific problem-solving in many fields and opens the way to a multitude of creative initiatives in the way mathematics is used. The successful statistician will be a generalist drawing on knowledge and experience gained from several fields and will base competitive advantage on having broad understanding and knowledge of statistical methodologies that training in a particular user discipline can rarely provide.

Each institution needs to develop fitting mechanisms that promote and sustain live connections between the statistical generalists that I see occupying the core of our discipline and both mathematicians and substantive researchers. For instance, Glenn points to the need to teach students in the biological and social sciences "not only the logic of the subject [of statistics] but also the decades-long record of its successes and failures in their discipline," and in effect links the excessive mathematization of statistics to our failure to develop teachers sufficiently broad in their knowledge and training to do such teaching as a source of the "growing isolation of the statistics department." Leading departments need to formulate specific plans to turn this situation around. In my view, growth, or in some cases survival, lies that way.

A. P. Dempster is Professor of Theoretical Statistics, Science Center, Harvard University, Cambridge, Massachusetts 02138.