Statistical Science
1987, Vol. 2, No. 1, 1–2

In This Issue

The field of artificial intelligence and expert systems, as a branch of computer science, is currently in a state of intense development and rapid expansion. Indeed, perhaps no other aspect of the "computer revolution" will have a greater impact on science and education than this field. Expert systems are being developed to aid decision makers in virtually all areas of human endeavor, with especially heavy emphasis in public policy, science, the environment, medicine, business, engineering, and military affairs. Although it might be expected that probability and statistics would play a vital part in this field, which is fundamentally concerned with the way in which people learn and make decisions, such has not proved to be the case, and the field has remained largely within the domain of computer scientists and subject-matter specialists in the problem areas for which systems are being developed. It is true that there are expert systems to help students learn statistics and expert systems to help users carry out statistical analyses, but they are mainly incidental to the basic research being carried out in the field. Of much more fundamental importance, and much greater potential impact, is the necessity to develop appropriate methods for representing and handling the multitude of uncertainties about relationships and conclusions that are present in every large-scale expert system. It is here that the door is open for the involvement of probability and statistics.

A conference on "The Calculus of Uncertainty in Artificial Intelligence and Expert Systems" was held at the Institute for Reliability and Risk Analysis of the School of Engineering and Applied Science, George Washington University, Washington, D. C., on December 28–29, 1984. Seymour M. Selig served as Coordinating Editor for the Conference and for the subsequent *Proceedings*, which were issued as a technical report by that Institute. In his Foreword to the *Proceedings*, Selig writes as follows:

"Despite the remarkable progress in the use and application of artificial intelligence and expert systems techniques in the past ten years, several fundamental issues remain unsolved.

"One of these is how best to deal with uncertainty in the conditions of interest involving expert systems. Even with the increased pace of discovery and innovation in the mathematical and information sciences, there still remain to be resolved issues pertaining to methods adequate for the treatment of uncertainty which are acceptable to all practitioners. Obviously many philosophical and methodological questions need to be addressed."

Nozer D. Singpurwalla, who organized the conference, "was the key to accomplishing the transformation from idea to reality" and its "driving force." The program was carefully constructed and balanced. There were four invited speakers: Glenn Shafer, the major developer of the use of belief functions for the representation of uncertainty; Lotfi A. Zadeh, the major developer of the use of fuzzy sets and possibility theory for the representation of uncertainty; Dennis V. Lindley, a leading proponent and developer of the use of subjective probability and Bayesian theory for the representation of uncertainty; and David J. Spiegelhalter, a leading researcher on the use of expert systems in medicine. In addition, there were two invited discussants: Arthur P. Dempster of Harvard University and Stephen R. Watson of Cambridge University, both of whom have themselves done important pioneering work related to expert systems. The moderator of the Conference was the Executive Editor of *Statistical Science*.

The oral presentations of the four invited speakers were eloquent, clear, and strong expressions of their contrasting views. These presentations, together with the comments of the invited discussants and other attendees, all of which are included in the technical report of the *Proceedings*, made the Conference an unusually lively one.

In this issue, we are pleased to present the papers that Shafer, Lindley, and Spiegelhalter contributed to the Conference. (Shafer and Spiegelhalter have revised and updated their contributions for *Statistical Science*.) Regretfully, Professor Zadeh did not prepare a manuscript for the Conference or for publication in *Statistical Science*. Professors Dempster and Watson were then invited to prepare discussion of these three papers. (Dempster's contribution is joint with Augustine Kong.) In a further attempt to recreate the atmosphere of the Conference itself, each of the three authors was also invited to submit discussion of the papers written by the other two. Finally, the three authors were invited to prepare their rejoinders to all of the comments.

The three papers by the invited authors are sharp and easy to read. Shafer writes: "After developing a constructive understanding of the Bayesian theory, I introduce another constructive theory, the theory of belief functions. I argue that both theories should be thought of as languages for expressing probability judgments and constructing probability arguments." Lindley writes: "Our thesis is simply stated: *the only satisfactory description of uncertainty is probability*. . . . In particular, alternative descriptions are unnecessary. These include ... possibility statements in fuzzy logic ... and belief functions." Spiegelhalter writes: "The development of expert systems in medicine has generally been accompanied by a rejection of formal probabilistic methods for handling uncertainty. We argue that a coherent probabilistic approach can ... meet many of the practical demands being made."

We hope that you will find the entire round robin, with its differing points of view, to be stimulating and enjoyable.

* * *

Joan Fisher Box is perhaps known best to statisticians for her excellent biography of her father, *R. A. Fisher, the Life of a Scientist*, John Wiley and Sons, New York, 1978 (paperback edition, 1986). Her article in this issue describes the environment at Guinness's Brewery in Dublin around the turn of the century, when W. S. Gosset (Student) worked there as a brewer. We learn of the high regard in which the brewers were held and of the importance of research to the brewery. She writes that "The life of the beer was important because Guinness is a naturally conditioned beer—it has no additives or preservatives nor, of course, is it pasteurized—and it has to remain potable while it is exported to Africa or the Far East, or stored in the barrel at varying temperatures before reaching the consumer." (This is not a commercial advertisement for Guinness.) She describes how Gosset came to be interested in the statistical analysis of small samples, and R. A. Fisher's introduction to Gosset in 1912 when Fisher was an undergraduate at Cambridge. Her story addresses the following questions: "Gosset was a brewer, Fisher a mathematics student when he started. Why should they have invented statistical methods for experimenters?" And "Why did they persist while experts in the field ignored or belittled their work?"

* * *

In his article on "Collinearity and Least Squares Regression," G. W. Stewart works on the important interface between numerical analysis and statistics. He writes that "Statisticians and numerical analysts share a concern about the effects of near collinearities on regression models—and with good reason.... It is not surprising then that both groups have devoted a great deal of effort to issues related to collinearity. In spite of this the subject has a certain vagueness about it, and it is instructive to ask why." After surveying several measures of collinearity that have appeared in the literature in various contexts, he introduces his "collinearity indices" which "indicate the presence of near collinearity in a precisely quantifiable manner" and provide "simple diagnostics, suitable for incorporation into regression packages."

In their discussion of this article, Donald W. Marquardt, David A. Belsley, and Ronald A. Thisted acknowledge the important contribution that Stewart has made in bringing the results and the perspective of numerical analysis to the attention of statisticians. Marquardt goes on to discuss why he had originally chosen the term "Variance Inflation Factors" in his own work, and the nomenclature chosen by Stewart. Belsley feels that Stewart does not give proper attention to "one of the more important notions that applied statistics has to teach the numerical analyst, namely, the necessity of a context for application: the fact that the data are not just a given set of numbers and the model is not just a linear combination of these data." He then discusses the relation between model and data, and the confusion that can occur. Thisted states that "As is true of most important papers, this one raises as many questions as it answers," and he further expands on measures of the relative importance of individual variables in a regression and the way in which collinearity diagnostics can be useful. Ali S. Hadi and Paul F. Velleman discuss several aspects of the collinearity and relative error measures proposed by Stewart. They pay particular attention to the effects of errors in variables, high leverage points, and collinearity-influential points. They, along with Marquardt and Belsley, also consider the issue of centering the data. They then write: "We commend Stewart for providing specific advice to developers of statistics packages and hope they adopt these methods. We think that they should take this opportunity and extend their packages to keep information about the precision of the data."

* * *

A highlight of this issue is an interview with C. R. Rao, perhaps the most well-known statistician in the world today, in which he discusses his distinguished career and his interests outside of statistics. As an update, shortly before this issue went to press he was appointed a National Professor by the government of India in recognition of his outstanding contributions to science. He now plans to divide his time between that position and his position as University Professor in the Department of Mathematics and Statistics at the University of Pittsburgh.