Sankhyā : The Indian Journal of Statistics
San Antonio Conference: selected articles
2002, Volume 64, Series A, Pt.2, pp 183-192

A VIEW OF THE PAST, PRESENT AND FUTURE

By HERMAN CHERNOFF
Harvard University, USA

KEYNOTE ADDRESS GIVEN AT THE SAN ANTONIO CONFERENCE

1. Global View

Dean Hammond, Professor Baird, Professor Keating, Professor Rao and organizers of this conference, it is a pleasure and honor to be selected to be the keynote speaker at this conference at the University of Texas at San Antonio, in honour of Professor C.R. Rao in anticipation of his 80th birthday. The theme of the conference is Reflections on the Past, Visions for the Future. As keynote speaker I feel compelled to address this theme although this global view is not my usual style. The presentation which follows will be largely influenced by my own experiences. It will represent a *tour de force* in containing no reference to the work of Professor Rao, which is *everywhere dense* among the accomplishments of statisticians in the last half century, some of which will be explained in the following presentations.

2. Concept of Time

References to the past will be mainly to the twentieth century. Time will be divided into

1. the classical period, roughly from 1900 to 1940,
2. the recent past, from 1940 to 1985,
3. the very recent past, from 1985 to the present,
4. the past future, from 1974 to the present, and
5. the future, from 2000 on, and about which I have little to offer.

Paper received October 2000.
AMS (1991) *subject classification.* 62A01, 62-03, 01-02.
Key words and phrases. Decision theory, Neyman-Pearson theory, empirical Bayes, Fisher, Bayesian approach, subjectivity, inadmissibility, computer intensive methods.

The rationale for some of this terminology is that I took a couple of college courses in Statistics around 1941, and decided to go into statistics seriously around 1945. At that time, almost all of text book statistics seemed very classical with two exceptions on which we shall elaborate later. Everything that happened during my career seemed to be recent. During one lecture around 1980, my class had a big laugh when I referred to a recent result in 1953. Most of the students had not yet been born then.

Old age has contributed to a narrowing view of what was going on outside of Cambridge in the world of Statistics. One example of the effect of old age was expressed in a conversation I had with Harold Hotelling in 1960 in Tokyo on the occasion of a meeting of the ISI. I had gone to the hotel restaurant for breakfast where I saw Professor Hotelling and asked to join him. During the ensuing conversation I asked whether he had previously been to Japan. It seems that he had been there on the occasion of his honeymoon in 1939 with his second wife, which they had celebrated with a round the world trip. That was when he fell behind. What did he mean that he fell behind? He said that until then, he knew all of Statistics. Perhaps not every little bit, but he was basically well acquainted and current with every publication of consequence. After that trip he never was able to catch up. To be sure, he did not feel that he could have maintained this status for long, but he thought that he still had a few more good years to go.

I must confess that it never occurred to me to know more than a few slices of the field. But my excuse is that I am a poor reader, and I expect most of the audience to be better informed that I.

The very recent past will overlap with what I will sometimes call the past future, referring to some forecasts made from 1974 to the present.

3. My Stages

Some perspective will derive from the history of my interactions with Statistics. As an undergraduate Mathematics major at CCNY around 1941, I took a couple of Statistics courses which influenced me in opposite ways. The first was an interdepartmental course in which the TA required me to repeat a drawing of a histogram in India ink because the original was too sloppy. He was so pleased with my painfully redrawn version, that he exclaimed that with this in my portfolio, I could easily get a job. I swore that if this was the intellectual content of Statistics, that was one field I would never enter. Another consequence was that for many years I tended to neglect discussing graphics with introductory students. As an aside, I might mention that one of the difficulties of teaching introductory courses is that it is hard to appreciate how difficult statistical reasoning is for introductory students, and the merits of graphical methods seemed too obvious to warrant discussion.

The other course was more mathematical, but it did not seem to be going anywhere until one week, the instructor had to take a trip, and he asked us to read some papers while he was gone. My task was a paper by Neyman and Pearson (1933). The main idea, which we now explain in 15 minutes in elementary classes, was so revolutionary that I struggled over it for a whole weekend until I finally decided "Yes, it is as simple as it seems". The mathematics was no obstacle, but the idea was a revelation that involved rewiring my brain to appreciate. Current students, told about balancing the probabilities of rejecting a true hypothesis against that of accepting a false one, can not appreciate how difficult it was for one trained in the classical paradigm to shift ideas.

I was basically a mathematician who was too immature to distinguish good from silly mathematics, and felt, as I still do, that solving a nontrivial applied problem must have merit. This was a major factor in my decision to study applied mathematics in graduate school. My mathematical talent seemed adequate, but I felt a lack of the physical insights which enabled first rate applied mathematicians to set up important formulations such as that of free boundary problems in fluid dynamics to explain jet flows. On the other hand the philosophical questions of how one learns from experience fascinated me, and I enjoy looking at data. I decided to study Statistics. At that time Henry Mann, a number theorist who was also working in Statistics, gave me Wald's first paper on Decision Theory, Wald (1938) to read. The effect of this paper on me was very similar to that of the Neyman-Pearson paper.

4. Classical Period

The first half of the 20th century was dominated by the giant, Sir Ronald A. Fisher. Time magazine collected a list of the 100 most important people of the century, and it is my conviction that he should have been on this list, for his work made it possible for the world to support the current population of 6 billion and more. One charming and possibly apocryphal story is about the geneticist who commented to a statistician that he had heard that Fisher also had a good reputation among statisticians. His contributions revolutionized or introduced Experimental Design, Analysis of Variance, Randomization, Maximum-Likelihood, Information, Sufficiency, and Principles of Inference. His early work converted Statistics to a subject attractive to mathematicians. The Neyman-Pearson (NP) theory of testing hypotheses and decision theory were natural extensions of Fisher's concepts of inference, but when NP theory appeared, Fisher disinherited it.

Another possibly apocryphal story that I heard, was that on a visit to India, Fisher was in a row boat with S.N. Roy, and in the middle of the lake he asked Roy whether there really was something to the NP theory. The following tales come from me as a first hand witness at the Summer session in 1946 in North Carolina when the Research Triangle was being started. At Fisher's first lecture, Bliss, a noted biostatistician, asked whether he would discuss errors of the first and second kind. The class became absolutely quiet. Even I, a rank beginner, knew that one did not ask the great man such a question. Unfortunately I don't recall the details of his negative reply. Later in the course, a young and irrepressible Monroe Norden, asked "How can you select an appropriate test statistic without NP theory". Fisher's reply was "I never have any difficulty".

Another development which converted probability theory into a respectable field for mathematicians, was that of the measure theoretic version where events correspond to sets in a *sample-space*. I once asked what took so long for this obvious set theoretic development. The answer was that the treatment of conditional probability and expectation, conditioned on events of probability zero, required the measure theoretic Radon-Nikodym theorem for Kolmogorov (1933) to write his fundamental book. One consequence of this result was that for many years, the most mathematically talented students in Statistics programs tended to specialize in probability theory, although with maturity, some of them drifted back into Statistics. The start of the Annals of Mathematical Statistics in 1930 and the subsequent formation of the Institute of Mathematical Statistics in 1933 were, to a large extent, a consequence of Fisher's work.

5. Distant Past

In the distant past, many important innovations and insights were contributed by scientists who may have had mathematical talent, but were relatively untrained. Factor analysis and quality control are examples. But in the recent past, the paradigms of Fisher, NP, and Wald provided enough questions to occupy theoretical statisticians without much applied experience. This led to some separation between theoretical and applied statisticians with theoreticians having the major influence in the growing departments of statistics in the United States. Major advances were made in decision theory, sequential analysis, multivariate analysis, experimental design, nonparametric testing, robustness and time series analysis.

By the 1950's the British statisticians began to regard the Annals of Statistics as unreadable. I found this strange because I always had trouble reading the British publications since it seemed to me that they would present a result without the detailed calculus based argument justifying it. Although they had accepted NP theory over Fisher's objections, they seemed very reluctant to take decision theory seriously. The arguments raised against the decision theory approach failed to enunciate clearly the deeper reasons for their misgivings. These arguments tended to raise superficial issues, some of which we shall now discuss.

To review briefly, the main fields of statistical inference were regarded as estimation and testing of hypotheses. Testing was mainly devoted to the testing of a null hypothesis which was usually a target for rejection by the investigator. The method was that of significance testing. Here the statistician defines a suitable test statistic, T, based on the data, to measure how poorly the data support the null hypothesis. He computes the P-value. Where P is less than a prescribed significance level α, typically 0.05 or 0.01, the hypothesis is rejected on the grounds that an outcome so unlikely is practically inconsistent with the hypothesis. The data are then said to be *significant*.

This approach keeps two important considerations implicit. What is the appropriate choice of T, and how should we decide on α? Like Fisher, the clever statistician, using common sense, seldom had great difficulty in deciding on an appropriate statistic, even though the issue of why one choice is more appropriate than another was never clearly articulated, much to the frustration of students and less experienced applied statisticians. NP theory attacked this issue by introducing the notion of the alternative hypothesis. The test statistic should be designed to tend to be large under the plausible alternative.

But NP theory still leaves implicit the issue of how one should select α. By introducing the cost of making the wrong decision, decision theory confronted this issue. The choice of alpha should depend on the relative costs of the errors, rejecting a true hypothesis and accepting a false one. Decision theory clarified the problem of inference by making explicit the issues that had been implicit in the past. However, it raised a new problem. A *strategy* for decision making, s has as a consequence a *risk* R, the expected cost, which depends not only on s, but also on the true unknown *state of* *nature*, which determines the actual probability distribution of the observed data. If the risk R_1 for one strategy s_1 is less than R_2 for another strategy s_2 for all possible states of nature, then it is fair to consider s_1 to be better than s_2, and to say that s_1 *dominates* s_2. But it is rarely, if ever, the case that there is a single best strategy which dominates all others. How should one compare s_1 with s_2 if $R_1 < R_2$ for a state of nature, but $R_1 > R_2$ for another?

Wald tentatively offered, as one possible resolution of this problem, the use of the *minimax* criterion. He suggested that one may use the strategy s which minimizes the worst possible R for that strategy. This idea was basically enunciated in ancient times by Thucididees. I recall a conversation with fellow students at Columbia University in 1947 where some expressed disappointment at how tentative Wald seemed in proposing minimax. As decision theory developed, the relationship between that theory and *game theory* became clear. While minimax makes sense in a *zero-sum two-person* game against an opponent, its use in decision theory corresponds to treating the unknown state of nature as the strategy of an opponent, and seems unreasonably conservative. In some problems this criterion produced reasonable results but it is easy to construct situations where this approach would lead to unacceptable strategies.

After a failed attempt to rescue minimax, L.J. Savage (1954), began to promote the *Bayesian* approach supported by de Finetti (1937) and previously pioneered by Ramsay (1931). Proposing a series of reasonable criteria that would be expected of a statistician's *coherent* strategy, Savage demonstrated that such a coherent strategy must be a Bayesian strategy where one acts as though the unknown state of nature is chosen at random according to some *a priori* probability law π.

We are now left with a formally complete picture of a statistical problem, which can be attacked by an applied mathematician to yield a good strategy. But there are disturbing aspects to this picture. While scientists like to think of objective scientific inference, our picture has two subjective elements. The cost considerations may be different for different statisticians. Thus the result of the data analysis will depend on the choice of the cost of the statistician or of his client, and is subjective. The *a priori* probability distribution is also subjective, and difficult to describe. In fact, Bayes theorem describes how the current *a priori* may be regarded as an *a posteriori* distribution based on an original *a priori* and past experience. But where did the original *a priori* come from?

It is interesting that Fisher (1956), devoted much of his discussion of inductive inference to attacking Bayesian analysis (not Bayes) on the grounds that the Laplacian and Jeffrey priors were arbitrary and unacceptable. He saw clearly that any reasonable Bayesian analysis must invoke a subjective prior.

What was the cause of the hesitation of British statisticians to accept decision theory even before the Bayesian controversy heated up on both sides of the Atlantic? Why did Fisher reject Neyman-Pearson theory?

It can't be claimed that the Bayesian decision theoretic picture is in itself a completely satisfactory resolution of the problem of statistical inference. My own position has been that if the statistician cannot interpret his problem from a Bayesian decision theoretic point of view, then he doesn't understand the problem, and an attempt at a solution is comparable to shooting blindly at a target. On the other hand, once the problem is understood, it is no longer necessary to attack it from a formal decision theoretic point of view.

I can only conjecture, based on the experience of some informal discussions in the years from 1950-1960, on why the British found it difficult to accept Decision Theory. Fisher's own work and its natural consequences of Neyman-Pearson and decision theory, opened up the field of theoretical statistics to a world of mathematically oriented statisticians who did not have much experience with applied statistics. Much of the work by decision theorists seemed far removed from the harsh realities of dealing with applied problems and was considered naive.

I believe that there were several apparent and real, but unarticulated, problems that led to the slow acceptance. One real problem was *robustness*. Until that was clearly framed it seemed too easy for an inexperienced statistician to take too seriously the results of his mathematical model without considering how it might lead to procedures which are poor if the model is oversimplified. This difficulty applies also to the use of Maximum Likelihood estimation, but Fisher usually could tell when his applied results were not be trusted.

However most of the objections centered about two apparent issues. One was subjectivity. One of the rocks on which science supposedly stands, is in the objective nature of inferences to be drawn from the data. In my opinion, decision theory has destroyed that belief. Subjectivity appears in several aspects. These are the costs and the prior beliefs, based at least in part on prior experience, and the analytic model formulated with which to analyze the problem, and which itself depends on prior beliefs and knowledge.

A second apparent issue often raised was the difference between inference for science and actions for commerce. Even Fisher, in his later years, accepted the fact that decision theory may make sense for applications in sampling inspection, but he insisted on objective inferences for problems of science.

In the sampling inspection problems, decisions were being made. In scientific problems, conclusions did not necessarily lead immediately to ultimate decisions. A great deal of the distinction between scientific and decision problems would be erased if we thought of the scientific problem as a long term problem in the sequential design of experiments, where important conclusions or terminal decisions were not to be made immediately on the basis of the information currently available. On the other hand that information could be very helpful in deciding what additional information to cumulate. In such considerations cost and efficiency are important.

I doubt that we will soon find an ultimate answer to the philosophical problems of scientific inference. However, an attempt to omit decision theory and prior knowledge issues from consideration can only encourage confusion.

My past powers of prediction were not very bad, but on one item, I was seriously wrong. I thought that Decision theory had pretty much reached a dead end by the middle of the 1950's. Around then there were two developments that Neyman (1962) regarded as major breakthroughs. These were Robbins' empirical Bayes (1956) and Stein's inadmissibility proof (1956) for the mean of a sample for estimating the mean of a normal distribution, and the ensuing James-Stein (1961) estimate. It turned out that both of these results were essentially closely related and constitute one major breakthrough, one which points out that it is sometimes possible to get very useful information on a problem by looking at data from unrelated problems.

All in all, this period, which I have called the recent past, was a wonderful period for mathematical statisticians in which to build and elaborate on the developments of the first half century in exciting, important and insightful ways. I will not elaborate on these here, but I will say a few words about the role of computers during this period.

The early innovators in computer science had dreams of glory about the potential of the computer. But, for a variety of reasons, the main applications of the computer during this period, were confined pretty much to fast calculations of old procedures. Thus insurance companies and banks could eliminate jobs where individuals toiled with electric calculators, but novel uses of the speed of calculation were not very prominent. Statisticians were able to do elaborate regressions quickly. It was only near the end of this period, that the potential of the computer to change what we should calculate began to be widely appreciated. For statisticians this was an opportunity to develop a healthy interface where statisticians employed computer techniques and had some useful input into the development of efficient computer methods such as speech recognition. I recall a comment by

Adrian Smith who claimed that the most important journal in statistics was the proceedings of IEEE.

The development of statistical packages was both a blessing and curse. The ability of people, completely devoid of statistical sense, to press a button to obtain a regression analysis, is a problem, we can't avoid. On the other hand the power of the computer led to a resurgence of interest in the graphical description of statistical information.

At the end of this period it seemed that mathematical statistics was becoming too much a matter of refining certain mathematical and probabilistic issues and too little a matter of new insights into learning from experience. What was needed was the experience of solving nontrivial applied problems to get the new insights which would provide new paradigms of importance.

6. Recent Past

The computer has begun to play a more creative role in this period. Some of its methods employ variations of statistical techniques in pattern recognition and cluster analysis to attack issues which had not been seriously considered before. It has also had a large influence on statistical analysis. Multivariate analysis, which was very difficult to deal with mathematically without the normal distribution, could be attacked with the computer. New methods such as projection pursuit and classification and regression trees depended heavily on the computer. Efron (1979) pioneered the principle that it would be worth making millions of computations to squeeze a little more information out of a small sample. In particular, he developed the bootstrap. Bayesian statistics, which seemed to be winning a partial victory in its long controversy, was energized by the presence of computer intensive methods of estimating posterior probability distributions in nontrivial problems. Monte Carlo Markoff Chain methods are used extensively. In short, the computer has allowed us to break free from the simple minded models in which we had been pleasantly trapped for many years, and to experiment with powerful new data analytic methods, many of which need theoretical analysis.

Another development has been the recognition in several fields of science that statistical analysis is important. In the past, many physicists claimed that if the data required statistical analysis, the wrong experiment has been performed. Now it looks like physicists and astronomers know that they have to use statistical analysis. The genome project makes it obvious that we will have to deal with large data sets in creative ways to make use of the list of 3 billion nucleotides which make up the human genome.

References

DE FINETTI, B. (1937). La prévision; ses lois logiques, ses sources subjectives, *Ann. Inst. H. Poincaré*, **7**, 1-68.

EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife, *Ann. of Statist.* **7**, 1-26.

FISHER, R.A. (1956). *Statistical Methods and Scientific Inference*, Oliver and Boyd, Edinburgh.

HUBER, P.J. (1975). Applications vs. abstraction: The selling out of mathematical statistics? *Proc. of the Conf. on Directions for Mathematical Statistics*, (S.G. Ghurye ed.). Applied Probability Trust, 84-89.

JAMES, W. and STEIN, C. (1961). Estimation with Quadratic Loss, *Proc. of the Fourth Berkeley Symp. on Math. Statist. and Prob.* Univ. of California Press, Berkeley, **1**, 361-380.

KOLMOGOROV, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin.

KUHN, T.S. (1962). *The Structure of Scientific Revolutions*, Univ. of Chicago Press, Chicago.

LINDLEY, D.V. (1975). The Future of Statistics – A Bayesian 21st Century, *Proc. of the Conf. on Directions for Mathematical Statistics*, (S.G. Ghurye ed.) Applied Probability Trust, 106-115.

NEYMAN, J. (1962). Two Breakthroughs in the Theory of Statistical Decision Making, *Rev. Inst. Internat. Statist.*, **30**, 11-27.

NEYMAN, J and PEARSON, E. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Philos. Trans. Royal Soc. of London, Seriea A.* **231**, 289-337.

RAMSEY, F.P. (1931). Truth and Probability, *The Foundations of Mathematics and other Essays* Kegan Paul, London.

ROBBINS, H. (1956). An Empirical Bayes Aapproach to Statistics, *Proc. 3rd Berkeley Symp. on Math. Statist. and Prob.*, Univ. of California Press, Berkeley, **1**, 157-163.

ROBBINS, H. (1975). Wither mathematical statistics? *Proc. of the Conf. on Directions for Mathematical Statistics*, (S.G. Ghurye ed.), Applied Probability Trust, 116-121.

SAVAGE, L.J. (1954). *The Foundations of Statistics*, Wiley, New York.

STEIN, C. (1956). Inadmissibility of the Usual Estimates for the Mean of a Multivariate Normal Distribution, *Proc. 3rd Berkeley Symp. on Math. Statist. and Prob.*, Univ. of California Press, Berkeley, **1**, 157-163.

WALD, A. (1939). Contributions to the Theory of Statistical Estimation and Testing Hypotheses, *Ann. of Math. Statist.*, **10**, 299-326.

HERMAN CHERNOFF
DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
CAMBRIDGE, MA 02138, USA
E-mail: chernoff@hustat.harvard.edu