Fisher's Contributions to Statistics

T Krishnan

T Krishnan received his Ph.D. from the Indian Statistical Institute. He joined the faculty of ISI in 1965 and has been with the Institute ever since. He is at present a professor in the Applied Statistics, Surveys and Computing Division of the Institute. Krishnan's research interests are in Statistical Pattern Recognition, Biostatistics and Psycho-metry. He has taken an interest in practical applications of Statistics and has helped research workers of various disciplines in designing their studies and in analysing data thereof. He is also called upon to advise organisations like the Indian Council of Medical Research in their work. Above all, he is an excellent teacher of Statistics.

Introduction

The contributions of Sir Ronald Aylmer Fisher to the discipline of statistics are multifarious, profound and long-lasting. In fact, he can be regarded as having laid the foundations of statistics as a science. He is often dubbed the 'father of statistics'. He contributed both to the mathematical theory of statistics and to its applications, especially to agriculture and the design of experiments therein. His contributions to statistics are so many that it is not even possible to mention them all in this short article. We, therefore, confine our attention to discussing what we regard as the more important among them.

Fisher provided a unified and general theory for analysis of data and laid the logical foundations for inductive inference. Fisher regarded statistical methods from the point of view of applications. Since he was always involved in solving biological problems which needed statistical methods, he himself developed a large body of methods. Many of them have become standard tools in a statistician's repertoire. Some of these developments required rather deep mathematical work and it was characteristic of Fisher to use elegant geometrical arguments in the derivation of his results. An excellent example of this type is his derivation of the sampling distribution of the correlation coefficient. Fisher had poor eyesight even at an young age, which prevented private reading and he relied largely on being read to, which in turn involved doing mathematics without pencil, paper and other such visual aids. It is believed that this situation helped him develop a keen geometrical sense.

Fisher's contributions to statistics have also given rise to a number of bitter controversies, due to the nature of the ideas and his personality and idiosyncrasies. Some of the controversies only go to show that it is not possible to build the whole gamut of statistical theory and methodology on a single paradigm and that no single system is quite solid, as Fisher himself realised.

Fisher had poor eyesight even at an young age, which prevented private reading and he relied largely on being read to, which in turn involved doing mathematics without pencil, paper and other such visual aids.

Sampling Distributions

Fisher derived mathematically the sampling distribution of the Student's t statistic which Gosset (pen name: Student) had derived earlier by 'simulation'. Fisher also derived mathematically the sampling distributions of the F statistic, the correlation coefficient and the multiple correlation coefficient and the sampling distributions associated with the general linear model. Fisher's derivation of the sampling distribution of the correlation coefficient from a bivariate normal distribution was the starting point of the modern theory of exact sampling distributions. Another useful and important contribution was the $\tanh^{-1}$ transformation he found for the correlation coefficient to make its sampling distribution close to the normal distribution, so that tables of the standard normal distribution could be used in testing significance of the correlation coefficient. Fisher made a modification in the degree of freedom of the Pearson's χ^2 when parameters are to be estimated.

Maximum Likelihood

Fisher's very first paper published in 1912 (at the age of 22) was on the method of maximum likelihood (although he did not call it so at that time). He developed this in view of his lack of satisfaction with the methods of moment estimators and least squares estimators. At that time the term 'likelihood' as opposed to probability or inverse probability caused some controversies. Although the basic idea of likelihood dates back to Lambert and Bernoulli and the method of estimation can be found in the works of Gauss, Laplace and Edgeworth, it was

Fisher's mathematics was not always rigorous, certainly not by modern-day standards, but even then, his mathematical work provides a great deal of insight.

Fisher to whom the idea is credited, since he developed it and advocated its use. Fisher studied the maximum likelihood estimation in some detail establishing its efficiency. Fisher's mathematics was not always rigorous, certainly not by modern-day standards, but even then, his mathematical work, like in the case of his work on maximum likelihood estimation, provides a great deal of insight. Some of his claims on the properties of maximum likelihood estimators were proved to be false in their generality by Bahadur, Basu and Savage. Subsequent authors developed strong theories based on the likelihood function. Fisher advocated maximum likelihood estimation as a standard procedure and since then it has become the foremost estimation method and has been developed for innumerable problems in many different sciences and contexts. It also has seen enormous ramifications and plays a central role in statistical theory, methodology and applications.

Following his work on the likelihood, Fisher did a lot of work on the theory of estimation and developed the notions of sufficiency, information, consistency, efficiency and ancillary statistic and integrated them into a well-knit theory of estimation. His pioneering work on this is contained in two papers he wrote in 1922 and 1925.

Analysis of Variance

In 1919, Fisher joined the Rothamsted Experimental Station, where one of his tasks was to analyse data from current field trials. It is in this context that he formulated and developed the technique of *analysis of variance*. The analysis of variance is really a convenient way of organising the computation for analysing data in certain situations. Fisher developed the analysis of variance initially for orthogonal designs such as randomised block designs and latin square designs. Later, Frank Yates extended the technique to nonorthogonal designs such as balanced incomplete block designs, designs with a factorial structure of treatments, etc. The technique of analysis of variance developed rapidly and has come to be used in a wide variety of problems formulated in the set-up of the linear model. Although initially developed as a convenient means of testing hypotheses, it also throws light on sources of experimental error and helps set up confidence intervals for means, contrasts, etc.

Fisher evolved design of experiments as a science and enunciated clearly and carefully the basic principles of experiments as randomisation, replication and local control.

Design of Experiments

Fisher's studies on the analysis of variance brought to light certain inadequacies in the schemes being used then for experiments, especially agricultural experiments. It is in an attempt to sort out these inadequacies that Fisher evolved design of experiments as a science and enunciated clearly and carefully the basic principles of experiments as randomisation, replication and local control (blocking, confounding, etc.). The theory of design of experiments he formulated was intended to provide adequate techniques for collecting primary data and for drawing valid inferences from them and extracting efficiently the maximum amount of information from the data collected. Randomisation guarantees validity of estimates and their unbiasedness. Replication helps provide a source of estimate of error, which can be used to compare treatments and other effects, test hypotheses and set up confidence limits. Local control helps to reduce sampling variations in the comparisons by eliminating some sources of such variations. Fisher formulated randomised block designs, latin square designs, factorial arrangements of treatments and other efficient designs and worked out the analysis of variance structures for them. The subject of design of experiments then developed rapidly both in the direction of formulation and use of efficient designs, especially in agricultural experiments, in the direction of statistical theory formulating useful and efficient designs and working out their analyses, and in the direction of interesting and difficult combinatorial mathematics investigating the existence of designs of certain types and their construction. Surely, the formulation of the basics of experimental design should be regarded as Fisher's most important contribution to statistics and science.

Though statistical and other pattern recognition methods and image processing techniques have made considerable progress in theory and in applications, Fisher's linear discriminant function still has a place in the pattern recognition repertoire.

Discriminant Analysis

From the time Fisher derived the sampling distributions of correlation coefficient and the multiple correlation coefficient, he was interested in the study of relationships between different measurements on the same individual and the use of multiple measurements for the purposes of classification and other problems. Fisher formulated the problem of discriminant analysis (what might be called a statistical pattern recognition problem today) in statistical terms and arrived at what is called the *linear discriminant function* for classifying an object into one of two classes on the basis of measurements on multiple variables. He derived the linear discriminant function as the linear combination of the variables that maximises the between-group to within-group squared distance. Since then the same function has been derived from other considerations such as a Bayes decision rule and has been applied in many fields like biological taxonomy, medical diagnosis, engineering pattern recognition and other classification problems. Statistical and other pattern recognition methods and image processing techniques have made considerable progress in the last two or three decades, in theory and in applications, but Fisher's linear discriminant function still has a place in the pattern recognition repertoire.

Books by Fisher

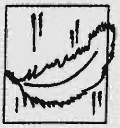
Fisher wrote several books on statistics, many of them containing his original ideas. The most important among them are: *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd, first published in 1925 and which has seen several editions. This unusual book is full of original ideas, written from the point of view of applications; here, each technique is explained starting with an actual scientific problem and a live data set collected to be able to answer certain questions and an enunciation of an appropriate statistical method with illustration on this data set. *The*

Design of Experiments, first published in 1935 has also seen several editions. Besides these, with F Yates, he compiled and published *Statistical Tables for Biological, Agricultural and Medical Research* in 1938 (with several subsequent editions), also by Edinburgh: Oliver & Boyd. These tables, together with those by Pearson and Hartley, were essential tools of a statistician's trade in those days when a statistical laboratory consisted of manually or electrically operated calculating machines and even in the days of electronic desk calculators. Students may not find Fisher's books quite readable and until one has mastered the material from some other source or with the help of a good teacher, his books may not help. However, they make very useful and enjoyable reading for an expert and for a teacher!

Suggested Reading

- June 1964 issue of *Biometrics*. Vol.20. No.2. *in memoriam Ronald Aylmer Fisher*. Dedicated to the memory of Fisher soon after his death, contains many articles on his life and work.
- Box J. *R A Fisher: The Life of a Scientist*. John Wiley & Sons. New York, 1978.
- Savage L J. Rereading of R A Fisher. In L J Savage: *The writings of Leonard Jimmie Savage: A Memorial Selection*. American Statistical Association and the Institute of Mathematical Statistics. Hayward. Calif. pp. 678-720, 1981.
- Fisher-Box J. Fisher, Ronald Aylmer. In Kotz S, Johnson N L and Read C B (Eds.). *Encyclopedia of Statistical Sciences*. New York. Wiley Interscience. Vol.3. pp. 103-111, 1988.
- December 1990 issue of *Biometrics*. Vol. 46. No. 4. published in the year of Fisher's birth centenary, contains a few articles on his life and work.
- Rao C R. R A Fisher: The founder of modern statistics. *Statistical Science*. Vol. 7. pp.34-48, 1992.

Address for Correspondence
T Krishnan
Computer Science Unit
Indian Statistical Institute
203 B T Road
Calcutta 700 035, India
e-mail: krishnan@isical.ernet.in



Science is facts, just as houses are made of stones, so is science made of facts; but a pile of stones is not a house and a collection of facts is not necessarily science.

Henri Poincaire