Statistical Science
1992, Vol. 7, No. 1, 34–48

R. A. Fisher: The Founder of Modern Statistics

C. Radhakrishna Rao

Abstract. Before the beginning of this century, statistics meant observed data and descriptive summary figures, such as means, variances, indices, etc., computed from data. With the introduction of the χ^2 test for goodness of fit (specification) by Karl Pearson (1900) and the t test by Gosset (Student, 1908) for drawing inference on the mean of a normal population, statistics started acquiring new meaning as a method of processing data to determine the amount of uncertainty in various generalizations we may make from observed data (sample) to the source of the data (population).

The major steps that led to the establishment and recognition of statistics as a separate scientific discipline and an inevitable tool in improving natural knowledge were made by R. A. Fisher during the decade 1915–1925. Most of the concepts and methods introduced by Fisher are fundamental and continue to provide the framework for the discussion of statistical theory. Fisher's work is monumental, both in richness and variety of ideas, and provided the inspiration for phenomenal developments in statistical methodology for applications in all areas of human endeavor during the last 75 years.

Some of Fisher's pioneering works have raised bitter controversies that still continue. These controversies have indeed helped in highlighting the intrinsic difficulties in inductive reasoning and seeking refinements in statistical methodology.

Key words and phrases: Ancillary statistics, Bayes theorem, confounding, consistency, efficiency, F-test, factorial experiments, fiducial probability, Fisher information, Fisher optimal scores, likelihood, local control, maximum likelihood, nonparametric tests, randomization, regression, replication, roots of determinantal equation, sufficiency.

1. INTRODUCTION

In his stimulating address to the members of the Institute of Mathematical Statistics, the American Statistical Association and the Biometric Society, entitled *Rereading of R. A. Fisher*, the late L. J. Savage (1981, pages 678–720) expressed his admiration at the deep and diverse nature of Fisher's contributions and the richness of his ideas, but he was critical of some of the claims made by Fisher as lacking in mathematical rigor and/or logical content. He concluded:

> Fisher is at once very near and very far from modern statistical thought generally.

Savage referred to the bitter controversies on some of Fisher's contributions and said:

> Of course, Fisher was by no means without friends and admirers too. Indeed, we are all his admirers. Yet he has a few articulate partisans in his controversies on the foundations of statistical inference, the closest, perhaps being Barnard (e.g., 1963) and Rao (e.g., 1961).

I must emphasize what Savage did not record was that whatever Fisher did was strongly motivated by practical applications. Fisher's research papers look quite different from those we find in current statistical journals. Fisher started off with a description of some live data presented to him for analysis, formulated the questions to be answered and developed the appropriate statistical methodology for the analysis of data. His style of writing was aphoristic and cryptic; often, intermediate mathematical steps are skipped, which may be

C. Radhakrishna Rao is Eberly Professor of Statistics and Director of the Center for Multivariate Analysis at Pennsylvania State University, 123 Pond Laboratory, University Park, Pennsylvania 16802.

annoying to the reader. This way, he introduced new ideas and expanded the scope of statistical methodology instead of dissipating his energy and talents in pursuing narrow theoretical concepts. Commenting on Fisher's contributions, Neyman (1951) says:

> A very able mathematician, Fisher enjoys a real mastery in evaluating multiple integrals. In addition, he has remarkable talent in the most difficult field of approaching problems of empirical research.

I have been associated with Fisher from 1940 until his death in 1962. I was one of the very few students (perhaps the only student) who did Ph.D. work in statistics under his guidance. I tried to understand his contributions by reading and rereading his writings, and writings on his writings, and above all by numerous personal discussions with him. My own research work is influenced largely by Fisher's ideas. I shall try to comment on different aspects of Fisher's work in the following sections.

We are indebted to Savage for his critical rereading of Fisher. It is important for us to know that some of Fisher's results specially in estimation are not valid in the wide generality claimed by him, some of his conjectures need slight modifications for their validity, and there are minor slips here and there in the mathematical treatment of some of the problems. These are matters of details that need to be examined and the findings put on record. But Savage's criticism does not detract Fisher's pioneering contributions from their usefulness and the motivation they provided for research in statistical theory and applications. They only highlight, what we are aware of today, that there are difficulties in building up a coherent structure for statistical inference, and the search for a monolithic structure for extracting information from data is bound to fail. Fisher himself expressed this view in his last book on *Statistical Methods and Scientific Inference*, a view also held by Savage (1981, page 734):

> The foundations of statistics are shifting, not only in the sense that they have been, and will continue to be changing, but also in the idiomatic sense that no known system is quite solid.

Section 2 gives a broad survey of Fisher's contributions toward the development of statistics over a period of 50 years, starting from 1912, which may be called the Fisherian era of statistics. Some comments on Fisher's book on *Statistical Methods for Research Workers* are given in Section 3.

In the remainder of the paper, references to Fisher's papers will be indicated by the year of publication followed by the volume and paper number, and sometimes the page number in the collected papers edited by J. H. Bennett and published by The University of Adelaide. The titles of books, *Statistical Methods for Research Workers, Statistical Methods and Scientific Inference* and *Calcutta University Lectures* will be abbreviated as SMRW, SMSI and CUL, respectively.

2. FISHERIAN ERA OF STATISTICS: 1912 – 1962

The earliest contribution to statistics by Fisher was (1912, 1–1), where the method of maximum likelihood was first used to estimate the unknown parameters. This was followed by a series of papers in the next 12 years, which laid the foundations of statistics as a separate, full-fledged scientific discipline with a great potential for applications in all sciences. He pointed out that a distinction should be drawn between a sample and the population that gave rise to the sample and defined statistics as inductive reasoning of generalizing from a sample to the population. Inductive reasoning has baffled philosophers for a long time, and although its codification started with the writings of Fisher, controversies are bound to raise and continue. It is now generally recognized that there is no monolithic structure for statistical inference, and Fisher's contributions have been of great help in our search for refinements and in introducing new ideas on the subject. In order to understand and evaluate Fisher's work, let us look at the status of statistics before 1912.

2.1 Statistics before 1912

Descriptive statistics. There was what is now described as descriptive statistics dealing with presentation of data through histograms, bar charts etc., and computation of measures of location, dispersion and association (such as correlation, partial correlation and regression in the case of continuous variables and coefficient of contingency in the case of discrete variables). No distinction was drawn between a sample and the population, and what was calculated from the sample was attributed to the population.

Curve fitting. Adolph Quetelet popularized the normal curve of error introduced by Gauss and Laplace by fitting normal curves to all sorts of biological data. Karl Pearson introduced a system of frequency curves to accommodate curves differing in shape from the normal in symmetry and kurtosis. He developed a method of choosing an appropriate curve for given data and also subjecting the choice to an objective test. This test, called the χ^2 test, is the forerunner of all methods of statistical inference and has been rightly hailed as one of the top 20 discoveries of this century, considering all fields of science and technology.

Testing of hypotheses. There have been a few attempts at testing of hypotheses, but a major breakthrough was the introduction of the t test, with an exact distribution in finite samples, by Gosset in 1908.

Estimation. The method of least squares due to Gauss and the method of least absolute deviations due to Laplace were generally used to estimate parameters in a linear model. Later, Karl Pearson introduced the method of moments to estimate parameters of frequency curves. There was also the concept of standard error (asymptotic standard deviation) to express the precision of an estimate.

Bayes' theorem. This was the first major attempt to quantify uncertainty based on observed data. Bayes recommended the use of a uniform prior distribution for the unknown parameters in the absence of any other knowledge and expressing the conclusions in the form of a posterior distribution. Bayes' theorem had engaged the attention of mathematicians like Boole, Venn, Chrystal, Laplace and Poisson, but was not actively pursued.

2.2 Fisher's Contributions: 1912 – 1962

Exact sampling distributions. In the words of Savage (1981, page 686), "In the art of calculating explicit sampling distributions, Fisher led statistics out of its infancy, and he may never have been excelled in this skill." Fisher obtained the exact null and nonnull distributions of the correlation (1915, 1–4), partial correlation (1924, 1–35) and multiple correlation (1928, 2–61) coefficients, and the F statistic (z statistic) that arises in tests of hypotheses in regression theory (1922, 1–20; 1924, 1–36).

Correct use of χ^2. Although Karl Pearson (1900) introduced the χ^2 statistic and found its asymptotic distribution when the parameters are known, he thought that the same distribution will hold when estimates are substituted for the unknown parameters in computing χ^2. Fisher (1922, 1–19) introduced the concept of degrees of freedom, which depends on the number of unknown parameters estimated and specifies the appropriate asymptotic distribution to be used. He showed that, for the validity of the χ^2 distribution, the parameters may have to be estimated by a more efficient method than that of moments, like maximum likelihood. He also pointed out the modification needed when an observed frequency in any cell is small or even zero.

Estimation. Fisher made a major contribution to theoretical statistics in (1922, 1–18), where he considered estimation as a method of reduction of data. He discussed the associated problems of specification (mathematical form of the population involving unknown parameters), estimation (of unknown parameters) and distribution of statistics (estimates) computed from the sample. He recommended maximum likelihood as a general method of estimation, which he used earlier in (1912, 1–1).

Regression. Fisher (1922, 1–20) developed the statistical methodology for testing goodness of fit of a regression function and for testing the significance of the individual coefficients. He also considered the problem of selection of variables (1938, 4–157) based on tests of significance on individual coefficients. Unfortunately, this procedure does not seem to be satisfactory (see Rao, 1984). In SMRW (7th edition, page 305), Fisher recommended the regression method in disaggregation of data:

> It often happens that the statistician is provided with data on aggregates which it is required to allocate to different items. Thus, we may have data on the total consumption of different households, without knowing how this consumption is allocated between a man and his wife, or among children of different ages. If the composition of each household is known, the relative importance of each class of consumer may be obtained by minimizing the deviation between the consumption recorded, and that expected, *on assigned scores*, from the composition of the family.

Fisher's suggestion was tried on different sets of data, but the results were very discouraging (like a negative score for the consumption of rice by the housewife in a Calcutta study). The flaw appears to be in the assumption that the *scores* for children, wife and husband remain the same for all compositions of households.

Design of experiments. Fisher introduced a whole new area of research, the design of experiments with a wealth of new ideas on scientific experimentation, analysis and interpretation of data. Design of experiments also gave a new impetus for research in combinatorial mathematics. Some fundamental contributions to combinatorics were made in searching for combinatorial arrangements for design of experiments. [See, e.g., the contributions of Bose and Shrikhande (1959, disproof of Euler's conjecture) and Rao (1949, orthogonal arrays).]

Discriminant function. In my paper (Rao, 1964) on R. A. Fisher—"The Architect of Multivariate Analysis," I have described in some detail Fisher's contributions to multivariate analysis, of which special mention may be made of discriminant functions. The linear discriminant function for the classification of an individual into one of two alternative populations proposed by Fisher (Martin, 1936) has become an important tool in numerical taxonomy and medical diagnosis. The discriminant function for genetic selection suggested by Fisher (Fairfield Smith, 1936) is an important tool in the selection of plants and animals for genetic improvement. This is, perhaps, the first attempt at using an empirical Bayes method, placing a prior distribution on the genetic parameters to be estimated simultaneously for several individuals. This is also an early example where Stein-type shrinkage estimates emerge when several parameters to be estimated are considered as a random sample from a population of parameters (see Rao, 1953). However, the Stein phenomenon of reduction in compound mean square loss for fixed values of the parameters was not noticed.

Roots of determinantal equation. Generalizing ANOVA techniques to multiple variables, Fisher (1939, 4–163) introduced nonlinear statistics obtained as roots of a determinantal equation of the type $|W - \lambda T| = 0$, where W and T are positive definite random matrices. Originally intended to study the dimensionality of the configuration of a set of points in a high-dimensional space, the distribution of the roots has received new applications in solving problems in physics (Mehta, 1967), signal detection, etc.

Dispersion on a sphere. Fisher (1953, 5–249) initiated a new area of research when observations are in the form of direction cosines specifying the orientation of an object, such as the direction of remnant magentism in the lava flows. Fisher considered the direction cosines in three dimensions as points on a sphere and suggested the density function of the form, $\exp(k \cos\theta)$, where k is positive and θ is the angular displacement from the true position, $\theta = 0$, at which the density is a maximum. k is a measure of precision; a large value k implies a high concentration of points around a central point and a small value of k, implies wide dispersion of points on the sphere. Fisher showed how to estimate the mean direction and the parameter k by the method of maximum likelihood. He also worked out the sampling distributions of the estimates. There is now a considerable body of literature on the statistical analysis of directional data (see Mardia, 1972).

Fisher (1953, 5-249) used his model to estimate the true direction (θ) of remnant rock magnetism in lava flow assuming that the observations collected over a geographical area are independent. He did not consider the possibility of spatial correlations which may have some effect on estimation as shown in Rao (1975).

Quantification of categorical variables. In the seventh edition of SMRW (1938, page 299), Fisher gave an example of 12 samples of human blood tested with 12 different sera, giving reactions represented by five symbols, –, ?, w, (+) and +, indicating different strengths of reactions. The data were arranged in a 12×12 two-way table, with the observed reaction given in each cell. Fisher showed how the qualitative reactions can be scored in such a way that the row and column effects are close to additivity, and the estimated scores used in a standard ANOVA of two-way classification to test for the row and column effects. Gower (1990) points out that Fisher's optimal scoring methods come close to methods of multiple correspondence analysis and related methods developed several years later by the French school (see Greenacre, 1984).

In a different context of the analysis of contingency tables, Fisher (see Maung, 1941) suggested a singular value decomposition of the matrix of cell frequencies, which is the basis of modern correspondence analysis (Greenacre, 1984). The singular vectors corresponding to the largest singular value provide scores for the row and column categories in such a way that the product moment correlation is maximized. Fisher (1940, 4–175) discussed what is now known as the reciprocal averaging method by which the scores can be computed. Fisher's approach can be extended to the general case of the construction of test criteria for comparison of treatments, etc., when the response variable is qualitative [see Mathen (1954), where I have suggested such an approach].

Multiplicative model. Fisher (1923, 1–32) introduced a multiplicative model in a two-way table to explain the responses of 12 varieties to 6 manures, as the usual additive model was found to be unsuitable. The method of fitting the multiplicative model given in the paper is the first attempt to approximate a given matrix by a matrix of unit rank, anticipating the matrix-approximation theorem of Eckart and Young (1936). The paper also provides an extended use of the ANOVA technique (based on nonlinear statistics) to test manural and varietal effects and deviations from the product formula. Perhaps, the first example using regular ANOVA is in 1918 (1–9).

What else did Fisher do? In variety and depth of scientific contributions, Fisher had no equal. He touched almost every aspect of current statistical research. Mention may be made of the transformation of statistics to stabilize asymptotic variance and/or to induce a higher rate of approach to asymptotic normality, as in the case of inverse hyperbolic tangent transformation of the correlation coefficient (1921, 1–14); analysis of residuals in fitting a regression function (1921, 1–15; 1922, 1–19; 1924, 1–37); distribution of extreme values (1928, 2–63); branching processes in genetics (1930, 2–87); maximum-likelihood estimation in probit analysis (1935, 3–126); specification based on methods of ascertainment of genetic data (1934, 3–113) that led to the investigation of weighted distributions (Rao, 1965); nonparametric inference, such as the sign test, run test, use of order statistics, Fisher–Yates normal score test (see *Statistical Tables* by Fisher and Yates) and treatment of outliers (1922, 1–18); and so on. The list is almost endless. Amazed at Fisher's scholarly and massive achievements, Savage (1981, page 686) says that it would have been more economical to list a few statistical topics in which Fisher displayed no interest than those in which he did.

3. STATISTICAL METHODS FOR RESEARCH WORKERS

This is the first full-length book on statistical methods; the first edition was published in 1925 and since then it has gone into numerous editions, with little revision but with some additional material or comments in each edition. In her biography of Fisher, Joan Box says, "The objective of the book was shaped on the anvil of Fisher's scientific thought under the hammer of empirical problems. He conceived of statistics as a tool for research workers and shaped it here to their ends." It is a guide to data analysis, lively and thought-provoking in its presentation, cautioning against misinterpretation of data, questioning the validity of given data to test specified hypotheses, providing reasons for the choice of particular statistical methodology and stressing that the aim of statistical analysis is not just answering the questions raised by the client but to extract all the information from it to answer possibly a wider set of questions and to guide future research. The book is based on his own experience of working with biologists and with the intimate knowledge of the data arising in actual research work. The book also demonstrates the need for an interface between statistics and other sciences for the development of meaningful methodology for analysis of data. It is not a textbook on the mathematical aspects of statistical theory. It is not statistical theory illustrated with made-up examples. It is statistics as it should be used and developed. Each section starts with a live data set, takes the reader step-by-step through various stages of statistical analysis and gradually introduces modern statistical concepts.

Fisher himself says the following in the introductory chapter of SMRW:

> ... many will wish to use the book for laboratory reference and not as a connected course of study. ... The great part of the book is occupied by numerical examples. ... The examples have rather been chosen to exemplify a particular process. ... By a study of the processes exemplified, the student should be able to ascertain to what questions, in his material, such processes are able to give a definite answer; and, equally important, what further observations would be necessary to settle other outstanding questions.

It is interesting to note that the SMRW is the only advanced book in statistics that emphasizes the importance of graphical techniques in the analysis of the data. There is a full chapter entitled, *Diagrams*, where Fisher says:

> The preliminary examination of most data is facilitated by the use of diagrams. Diagrams prove nothing, but bring out outstanding features readily to the eye; they are therefore no substitute for such critical tests as may be applied to the data, but are valuable in suggesting such tests and in explaining the conclusions based on them.

The following complaints about the book are, perhaps, not without some justification, but appear to be based on an insufficient appreciation of what it is meant to convey.

> A prerequisite for reading *Statistical Methods for Research Workers* is that you must have a Master's degree in statistics.
>
> Attributed to M. G. Kendall

> In modern mathematical education there is great repugnance to transmitting a mathematical fact without its demonstration. ... Fisher freely pours out mathematical facts in his didactic works without even a bow in the direction of demonstration. I have encountered relatively unmathematical scholars of intelligence and perseverance who are able to learn much from these books, but for most people, time out for some mathematical demonstration seems indispensable to mastery.
>
> Savage (1981, page 685)

Fisher is not against the use of mathematics in teaching statistical methodology. Fisher (1938, 4–159) says:

> I want to insist on the important moral that the responsibility for the teaching of statistical methods in our universities must be entrusted, certainly to highly trained mathematicians, but only to such mathematicians as have had sufficiently prolonged experience of practical research and of responsibility for drawing conclusions from actual data, upon which practical action is to be taken. Mathematical acuteness is not enough.

SMRW was the only text on statistical methodology and inference until the Second World War and has inspired many generations of statisticians, and there are plenty of new ideas in the book from which one can still learn.

4. FOUNDATIONS OF THEORETICAL STATISTICS

Fisher (1922, 1–18), on mathematical foundations of theoretical statistics, is a classic in many ways. For the first time, a clear distinction is drawn between a population and a random sample from it, and the fundamental problems of statistics are stated (see page 280):

(i) Specification (population model specified as a family of probability distributions P_θ indexed by a parameter θ).
(ii) Estimation (choosing a value of θ or a member of P_θ, based on the sample, as the appropriate population distribution).
(iii) Distribution (for expressing the precision of the estimate of θ or the uncertainty in the choice of θ).

The formulation of these problems mark the beginning of the development of statistics as a full-fledged discipline.

Fisher did not give any general discussion of the problem of specification. He was appreciative of Karl Pearson's work on his system of frequency curves and the χ^2 goodness of fit:

> We may instance the development by Pearson of a very extensive system of skew curves, ... Nor is the introduction of the Pearsonian system of frequency curves the only contribution which their author has made to the problem of specification: of greater importance is the introduction of an objective criterion of fit.
> Fisher (1922, 1–18, page 281)

A great part of Fisher's paper is devoted to the introduction of new ideas and concepts to provide a logical framework for discussing problems of estimation. For the first time, a clear distinction is made between a *parameter* and an *estimate*. Criteria of *consistency*, *efficiency* and *sufficiency* were introduced to study properties of estimates and to compare alternative estimates. A measure of *information* on unknown parameters in observed data from an experiment was introduced. The concept of *ancillary statistic* for conditional inference was put forward. *Likelihood*, as a function of the parameters given the data, was defined and the method of *maximum likelihood* was proposed as a general method of estimation.

The terminology and the methods introduced by Fisher form the core of modern estimation theory. No doubt, certain refinements and modifications were made, and will continue to be made, to provide rigor and to meet new situations. We shall examine some of the concepts, results and conjectures made by Fisher and controversies surrounding them, in light of modern developments.

4.1 Consistency

In Fisher (1922, 1–18, page 276), consistency of an estimate is defined as follows:

> A statistic satisfies the criterion of consistency, if, when it is calculated from the whole population, it is equal to the required parameter.

To interpret this definition, we may look at the example on page 283, where Fisher says that

$$\sigma_2 = \left[n^{-1}\Sigma(x_i - \bar{x})^2\right]^{1/2}$$

when calculated from the whole population will lead to the correct value of the standard deviation σ. Perhaps, the implication is that a statistic T_n, as an estimate of a parameter of θ is defined for all sample sizes and that as $n \to \infty$, $T_n \to \theta_0$ (true value of θ) in probability. We call this CP (consistent in probability). All the editions of SMRW up to the eleventh (1950, page 11) and papers on estimation written subsequent to the 1922 paper carry the above definition of consistency. In a series of lectures delivered at Calcutta University in 1938, Fisher (CUL, page 42) says:

> ... if T tends to a limiting value, it is easily recognized by inserting for the frequencies in our sample their mathematical expectations.

In the same publication, Fisher also mentions that an estimate is a function defined on the sample space (presumably of observed relative frequencies) and it is consistent if its value at the expected frequencies is the required parameter. Thus, by consistency, Fisher had in mind both the properties: the estimate tending to the true value in probability and the estimate, considered as a function of observed relative frequencies, taking the true value when the expected values are substituted for the relative frequencies. [The meaning of Fisher's statements is clear when sampling is from a multinomial distribution where we have observed and expected relative frequencies and the estimating function is defined for all vectors $(x_1, \ldots, x_k)$, $x_i \geq 0$, $\Sigma x_i = 1$. Fisher regarded a continuous distribution as an infinite multinomial, and thus included the continuous case in his definition. Kallianpur and Rao (1955) clarified the situation.]

In 1954, I was preparing a paper on asymptotic efficiency for presentation at the International Statistical Conference, to be held in Rio de Janeiro. I wanted to prove Fisher's bound for the asymptotic variance of a consistent estimate rigorously using conditions under which Hodges–Le Cam phenomenon of superefficiency will not hold. Fisher was visiting Calcutta at that time, and I asked him what exactly did he mean by consistency. He referred to the multinomial case, where the expected cell probabilities are $\pi_1(\theta), \ldots, \pi_k(\theta)$ which are functions of, say, a single parameter θ. He said that an estimate $T(p_1, \ldots, p_k)$ of θ based on observed relative frequencies $p_1, \ldots, p_k$ only, is consistent if

$$(1) \qquad T(\pi_1(\theta), \ldots, \pi_k(\theta)) \equiv \theta$$

and no limiting property is involved.

Choosing (1) as the definition of consistency, which I termed as Fisher consistency (FC), and assuming differentiability of T as a function of k variables and of $\pi_i(\theta)$ as functions of θ, I gave a simple proof of Fisher's inequality in Rao (1955) on the following lines. Differentiating both sides of (1) and the relationship $\Sigma \pi_i(\theta) = 1$, we have

$$(2) \qquad \begin{aligned} &\sum_{i=1}^{k} \frac{\partial T}{\partial \pi_i} \frac{d\pi_i}{d\theta} = 1, \qquad \sum \frac{d\pi_i}{d\theta} = 0, \\ &\Rightarrow \sum_{i=1}^{k} \sqrt{\pi_i} \left(\frac{\partial T}{\partial \pi_i} - \sum_{i=1}^{k} \pi_i \frac{\partial T}{\partial \pi_i} \right) \frac{1}{\sqrt{\pi_i}} \frac{d\pi_i}{d\theta} = 1. \end{aligned}$$

Applying the Cauchy–Schwarz inequality to (2) yields

$$(3) \qquad a^2 = \sum \pi_i \left(\frac{\partial T}{\partial \pi_i} \right)^2 - \left(\sum \pi_i \frac{\partial T}{\partial \pi_i} \right)^2 > \frac{1}{I}$$

where

$$I = \sum \frac{1}{\pi_i} \left(\frac{d\pi_i}{d\theta} \right)^2$$

is the Fisher information. Notice that under the assumed differentiability condition it is well known that the asymptotic distribution of $\sqrt{n}\,(T(p_1, \ldots, p_k) - \theta)$ is $N(0, a^2)$ where a^2 is the desired asymptotic variance. Equation (3) provides a simple and rigorous demonstration of Fisher's inequality. The proof is similar to that of Cramér–Rao lower bound. In my paper, I have considered the more general inequalities arising in the multiparameter case.

In a subsequent paper, in collaboration with Kallianpur (Kallianpur and Rao, 1955), the concept of FC was extended to the continuous case by considering an estimating function defined on the space of all distribution functions (d.f.'s). If F_n is the empirical d.f. based on a fixed sample size n and $F(\theta)$ is the corresponding true d.f., then an estimate $T_n = T(F_n)$ is FC if $T(F(\theta)) \equiv \theta$. We were able to establish Fisher's inequality by considering Frechet differentiability of T.

In the 12th edition of SMRW published in 1954, Fisher, for the first time, added the following paragraph (page 12), while retaining the earlier definition:

> The foregoing paragraphs specify the notion of consistency in terms suitable to the theory of large samples, i.e. by means of the properties required as the sample is increased without limit. Logically it is important that consistency can also be defined strictly for small (i.e. finite) samples by the stipulation that if for each frequency observed its expectation were substituted, then consistent statistics would be equal identically to the parameters of which they are estimates.

In SMSI (pages 144–146), Fisher says that his definition of consistency as a limiting property is unsatisfactory, and the alternative definition applicable to finite samples is the appropriate one. However, such a definition of consistency may be too restrictive for application in many problems.

4.2 Maximum Likelihood

Early references to the method of maximum likelihood (m.l.) for estimation can be found in the works of Gauss, Laplace and Edgeworth, but it was Fisher who saw its great potential for universal use and started studying the properties of m.l. estimates (m.l.e.'s). Encouraged by the nice properties of the m.l.e.'s, judged by criteria such as consistency, efficiency and sufficiency, in the numerous examples he examined, Fisher suggested the m.l. method for universal use. He proved some propositions claiming optimum properties for the m.l.e.'s, which we now know are not universally true. [See the counter examples by Bahadur (1958), Basu

(1988) and Savage (1981).] Most of the results can, however, be established rigorously under certain conditions. Fisher was aware of the shortcomings in his mathematical proofs. He suggested the m.l. method only as a "formal solution of problems of estimation" with the following note:

> For my own part I should gladly have withheld publication until a rigorously complete proof could have been formulated; but the number and variety of the new results which the method discloses press for publication.
>
> Fisher (1922, 1–18, page 290)

Fisher has been accused of priesthood in advocating the m.l. method, because it fails to give acceptable estimates in certain situations. Of course, there is no known method in statistics which is universally optimal, and different methods may have to be used for different purposes and in different situations. The m.l. method is no exception, but it remains as the main stay, which everyone tries when confronted with a new situation. When the m.l. method fails to give acceptable estimates, other methods are sought for.

What can be said in support of m.l. estimates? Under some smoothness conditions, the following are true:

(i) They are Fisher efficient in the sense of having minimum asymptotic variance.

(ii) They are second-order efficient in a decision theoretic sense of having a minimum value for the terms up to the order $(1/n^2)$ in the expansion of expected loss under a bowl-shaped loss function (Rao 1961b, 1963; Ghosh and Subramaniam, 1974; Efron, 1975; and Akahira and Takeuchi, 1981).

(iii) Suppose that the unknown density function g does not lie in a specified parametric family $f(\theta)$ and an m.l. estimate $\hat{\theta}$ of θ is obtained under the wrong assumption that it does. In such a case, $\hat{\theta}$ estimates θ_g defined by

$$I(g; f(\theta_g)) = \min_{\theta} I(g; f(\theta))$$

where

$$I(g; f(\theta)) = \int g(x) \log\{g(x)/f(x \mid \theta)\}\, dx$$

is the Kullback–Liebler information number. Or, in other words, the m.l. estimate provides a close approximation to the true density (see Foutz and Srivastava, 1977; White, 1982; and Nishi, 1988).

Can m.l. be used for model selection? Suppose that the class of possible regressions in a given problem is the set

$$\{\beta_0, \beta_0 + \beta_1 x, \beta_0 + \beta_1 x + \beta_2 x^2\}$$

of zero-, first- and second-degree polynomials. Then, the m.l. method will always choose the second-degree polynomial. Fisher did not consider model selections as a problem of estimation, but, perhaps, as a problem in testing of hypothesis. In recent work, the m.l. principle is extended to cover model selection by using m.l. with an appropriate penalty function depending on the number of parameters (see Akaike, 1973; Rissanen, 1978; Schwarz, 1978; Zhao, Krishnaiah and Bai, 1986; Nishi, 1988; and Rao and Wu, 1989.)

Fisher's work on estimation is of a pioneering nature. The basic concepts and the terminology introduced by him are now routinely used in discussing problems of estimation. Some propositions in estimation proved by Fisher may lack in rigor, but this does not detract their value in their logical content. As stated earlier, Fisher himself was aware of the imperfections in his mathematical treatment. Commenting on the early work of Fisher, Mahalanobis (1938) says:

> Mechanical drill in the technique of rigorous statement was abhorrent to him, partly for its pedantry, and partly as an inhibition to the active use of the mind. He felt it was more important to think actively, even at the expense of occasional errors from which an alert intelligence would soon recover, than to proceed with perfect safety at a snail's pace along well-known paths with the aid of the most perfectly designed mechanical crutches. ... Fisher himself thinks that he was merely a very willful and impatient young man. This is no doubt true, but he was impatient not because he was young but because he was a creative genius.

4.3 Estimation as Reduction of Data

Fisher viewed the problem of estimation as that of reduction of data. If

$$\underset{\sim}{x} = (x_1, \ldots, x_n)$$

constitutes the data, the problem may be posed as that of finding a k-vector statistic, with $k \ll n$,

$$\underset{\sim}{T} = (T_1(\underset{\sim}{x}), \ldots, T_k(\underset{\sim}{x}))$$

such that "$\underset{\sim}{T}$ contains as much as possible, ideally the whole, of the relevant information" (1922, 1–18, page 278). Reduction of data ("which is usually by its bulk is incapable of entering the mind") in the form of summary figures may offer some convenience in understanding the data and drawing inferences, in addition to the economy in recording only the summary figures for future use, instead of preserving the entire mass of data, much of which may be irrelevant. These considerations are important, especially in scientific research. If, therefore, we define the purpose of estimation as condensation of data, what are the appropriate criteria for choosing the statistic $\underset{\sim}{T}$ to replace the whole sample $\underset{\sim}{x}$? These criteria are developed in Rao (1961a), of which a brief outline is given below. Under these wider criteria, the m.l. estimates seem to provide a good summary of given data provided some regularity conditions hold.

The log-likelihood ratios

$$l(\theta_1, \theta_2 \mid \underset{\sim}{x}) = \log L(\theta_1 \mid \underset{\sim}{x}) - \log L(\theta_2 \mid \underset{\sim}{x}) = l(\theta_1 \mid \underset{\sim}{x}) - l(\theta_2 \mid \underset{\sim}{x})$$

play an important role in statistical inference, whether it is Bayesian or frequentist. If there exists $\underset{\sim}{T}$ such that

$$l(\theta_1, \theta_2 \mid \underset{\sim}{x}) = l(\theta_1, \theta_2 \mid \underset{\sim}{T}),$$

then no information is lost, in which case $\underset{\sim}{T}$ is said to be sufficient. If no such $\underset{\sim}{T}$ exists, then we need some criteria for choosing $\underset{\sim}{T}$.

(i) *Wider consistency:* In general, as $n \to \infty$, $l(\theta_1, \theta_2 \mid \underset{\sim}{x}) \to \infty$ or $-\infty$ according as θ_1 or θ_2 is the true value so that complete discrimination is possible in large samples between any two alternative parameter values θ_1 and θ_2. If the same holds when $\underset{\sim}{x}$ is replaced by $\underset{\sim}{T}$, i.e.,

$$l(\theta_1, \theta_2 \mid \underset{\sim}{T}) \to \infty \text{ or } -\infty,$$

then $\underset{\sim}{T}$ is said to be consistent in the wider sense.

(ii) *First-order efficiency:* Let us consider discrimination between two close alternatives θ and $\theta + \delta\theta$. Then expanding $l(\theta + \delta\theta, \theta \mid \underset{\sim}{x})$, we have up to the first-order term

$$l(\theta + \delta\theta, \theta \mid \underset{\sim}{x}) = \frac{L'(\theta \mid \underset{\sim}{x})}{L(\theta \mid \underset{\sim}{x})}\delta\theta = S_{\underset{\sim}{x}}(\theta)\delta\theta.$$

Similar expansion with $\underset{\sim}{T}$ gives

$$l(\theta + \delta\theta, \theta \mid \underset{\sim}{T}) = \frac{L'(\theta \mid \underset{\sim}{T})}{L(\theta \mid \underset{\sim}{T})}\delta\theta = S_{\underset{\sim}{T}}(\theta)\delta\theta,$$

where $S_{\underset{\sim}{x}}(\theta)$ and $S_{\underset{\sim}{T}}(\theta)$ are the score functions based on $\underset{\sim}{x}$ and $\underset{\sim}{T}$ respectively. If $S_{\underset{\sim}{x}}(\theta)$ and $S_{\underset{\sim}{T}}(\theta)$ are equivalent in some sense, then $\underset{\sim}{T}$ is a good summary of $\underset{\sim}{x}$. Let $U = S_{\underset{\sim}{x}}(\theta) - S_{\underset{\sim}{T}}(\theta)$. We define $\underset{\sim}{T}$ to be first-order efficient if $n^{-1/2}U \to 0$ in probability or in terms of the variance of $n^{-1/2}U$

$$V(n^{-1/2}U) = \frac{1}{n}\left(I_{\underset{\sim}{x}}(\theta) - I_{\underset{\sim}{T}}(\theta)\right) \to 0 \tag{4}$$

as $n \to \infty$, where $I_{\underset{\sim}{x}}(\theta)$ is Fisher information based on the whole sample and $I_{\underset{\sim}{T}}(\theta)$, that based on $\underset{\sim}{T}$. Equation (4) means that the information per observation tends to be the same for $\underset{\sim}{x}$ and $\underset{\sim}{T}$ as $n \to \infty$.

(iii) *Second-order efficiency:* Let us consider

$$V(U) = I_x(\theta) - I_T(\theta). \tag{5}$$

Generally, both the terms on the right-hand side diverge, but the difference may tend to a constant as $n \to \infty$. The limiting constant is defined as second-order efficiency and $\underset{\sim}{T}$ for which this constant is a minimum has the maximum second-order efficiency.

We see that Fisher information comes in a natural way as a criterion for distinguishing between estimates.

Fisher (1925, 2–42) found a general expression for the limit of (5) as $n \to \infty$. However, the proof is not rigorous. Rao (1961b) and Efron (1975) have given a different interpretation of Fisher's expression.

5. CONDITIONAL INFERENCE

Fisher maintained that statistical methodology developed for data analysis in one area may not be directly applicable in another, since the objectives in different areas may be different. For a business concern, profit or loss in the long run, aggregated over all of its diversified activities, is relevant. In such a case, it is appropriate that decisions taken over a period of time and activities are such that the expected compound loss is minimized. In this process, it may happen that losses in some activities are heavy, but are compensated by larger gain in others. But the situation is different in scientific research. The concept of aggregate loss over a number of different scientific projects or loss in the long run in taking scientific decisions is not meaningful. The data arising from each experiment to estimate an unknown parameter or to test a hypothesis have to be considered separately, and the amount of uncertainty in the best possible decision taken on each has to be specified. Further, in each case, we need a summary of the data to communicate to others or to place on record for future use.

For statistical analysis of sample data, we need to know how the data are generated, which is specified by a model $(\mathscr{X}, \mathscr{B}, \Theta)$, with $\mathscr{X}$ as the space of all samples, $\mathscr{B}$ as Borel sets and Θ as identifying the family of probability distributions indexed by a parameter $\theta \in \Theta$. In the terminology of Dawid

(1991), this is called the production model. In the Bayesian set-up, a hypothetical probability model, called the prior, is imposed on the space Θ in the form $(\Theta, \mathscr{B}^*, p)$, where p is a known probability measure, in which case, statistical analysis can be done in the standard Bayesian way. No other complications arise, except for the pertinent question, what p? [Bayesians may not agree, but I believe, that Bayesian decision theory is essentially based on minimizing loss in the long run (expected) with respect to a chosen prior distribution.]

In the frequentist approach, some new methodological problems arise. Suppose that a certain parameter θ is estimated by $T(X)$ when $X \in \mathscr{X}$ is observed. In the frequentist approach, the distribution of $T(X)$ is used for drawing inference on θ. Fisher suggested that the distribution of $T(X)$ should be obtained not with respect to the production model, but with reference to a restricted model $(\mathscr{X}_A, \mathscr{B}_A, \Theta)$ where

$$\mathscr{X}_A = \{X \in \mathscr{X} : A(X) = A(x)\}$$

and x is the observed value of X, and $A(X)$ is an ancillary statistic (i.e., whose distribution is independent of θ). Fisher considered a number of examples using particular ancillary statistics but did not lay down rules for choosing them.

There has been considerable debate on conditional inference of the type previously described. There are examples like mixtures of experiments (Cox, 1958) and simple random sampling with replacement (Basu, 1988), where there is broad agreement on the need and choice of an ancillary for conditioning. But the general recommendation of Fisher, which is logically of the same status as in the previously described problems, has run into some difficulties, mainly because there can be many choices of an ancillary statistic, each leading to a different kind of inference (Basu, 1988). Some further research appears to be necessary on the choice of an ancillary.

I shall give some examples to show how conditioning on certain features of the observed data can be of help in refining statistical inference.

Example 1. The first is a finite sample version of Fisher's example (1925, 2-42, page 26). Suppose that we have two independent samples X and Y, giving information on the same parameter θ, from which the estimates $T_1(X)$ and $T_2(Y)$ obtained are such that

$$E[T_1(X)] = E[T_2(Y)] = \theta, \tag{6}$$

$$V[T_1(x)] = v_1, \qquad V[T_2(Y)] = v_2, \tag{7}$$

where v_1 and v_2 are independent of θ. Further, suppose that there exist statistics $A_1(X)$ and $A_2(Y)$ such that

$$\begin{aligned} E[T_1 \mid A_1(X) = A_1(x)] &= \theta, \\ E[T_2 \mid A_2(Y) = A_2(y)] &= \theta, \end{aligned} \tag{8}$$

$$\begin{aligned} V[T_1 \mid A_1(X) = A_1(x)] &= \nu_1(x), \\ V[T_2 \mid A_2(Y) = A_2(y)] &= \nu_2(y), \end{aligned} \tag{9}$$

where x and y are observed values of X and Y, respectively, and $\nu_1(x)$ and $\nu_2(y)$ are independent of θ. Then, we might consider the conditional distributions of T_1 and T_2 given A_1 and A_2 at the observed values and report the variances of T_1 and T_2 as $\nu_1(x)$ and $\nu_2(y)$, respectively, as an alternative to (7). What is the right thing to do?

Now, consider the problem of combining the estimates T_1 and T_2 using the reciprocals of v_1, v_2 and $\nu_1(x)$, $\nu_2(y)$ as alternative sets of weights:

$$t_1 = \left(\frac{T_1}{v_1} + \frac{T_2}{v_2}\right) \bigg/ \left(\frac{1}{v_1} + \frac{1}{v_2}\right), \tag{10}$$

$$t_2 = \left(\frac{T_1}{\nu_1(x)} + \frac{T_2}{\nu_2(y)}\right) \bigg/ \left(\frac{1}{\nu_1(x)} + \frac{1}{\nu_2(y)}\right). \tag{11}$$

It is easy to see that the unconditional variances of t_1 and t_2 satisfy the relation

$$V(t_1) \geq V(t_2) \tag{12}$$

so that t_1 is inadmissible. Does this not mean that $\nu_1(x)$ and $\nu_2(y)$ are more appropriate measures of precision than v_1 and v_2 of T_1 and T_2?

Note that in the previous example, A_1 and A_2 need not be strictly ancillaries. We need only the conditions in equations (6) to (9) to be satisfied.

Example 2. Suppose that a random sample of size 3 has been taken from a row of plants to estimate their average height, and we observe:

height	h_1	h_2	h_3
position of the plant	10	30	31.

We note that in the observed sample, two values are from two contiguous plants, and there is likely to be a high correlation between the heights of successive plants. In such a case, we could refer the observed sample to the set of samples, where two units out of the three chosen are contiguous and estimate the average height as

$$\frac{2^{-1}(h_2 + h_3) + h_1}{2}$$

which is better than the traditional estimate

$$\frac{h_1 + h_2 + h_3}{3}$$

if, in fact, there is high correlation between contiguous plants.

EXAMPLE 3. Even after a half century of debate, there still seems to be no consensus on when to treat the margins of a 2×2 contingency table as fixed when conducting a significance test. In a recent article, Greenland (1991) provides a logical justification of conditional tests without appealing to ancillarity, conditionality or marginal information.

EXAMPLE 4. Suppose that we want to estimate the population of the state of West Bengal (India) by random sampling from the complete list of N towns and cities, on which we do not have any prior information on the relative sizes. Theory says that if $x_1, \ldots, x_k$ are the population sizes of k observed units, then an estimate of the population of West Bengal is $N\bar{x}$. Rao (1971) argued that if, among the observed $x_1, \ldots, x_k$, there is one value say, x_3 (perhaps the population of Calcutta), which is much larger than the rest, then a better estimate would be $x_3 + (N - 1)\bar{x}'$ where $\bar{x}'$ is the average excluding x_3. Here, the sample is referred to the set where x_3 is observed.

These examples show that the configuration of the observed sample provides some information on appropriate analysis of the data. We owe it to Fisher for introducing this useful concept and the few examples I have given show the need for further discussion and research on conditional inference or choice of a frame of reference. It appears that different samples of the same size from the same population have different information on the unknown parameters, depending on the configuration of the observations in the sample.

6. DESIGN OF EXPERIMENTS

In their biographical account of Fisher, Yates and Mather (1963) say the following about Fisher's contributions to design of experiments:

> . . . the new ideas of experimental design and analysis soon came to be accepted by research workers. . . . The recent spectacular advances in agricultural production owe much to their consistent use. They certainly rank as one of Fisher's greatest contributions to practical statistics.

The subject of experimental designs was developed by Fisher during the years 1991–1923, while he was working at the Rothamsted Experimental Station. He saw the need to collect data in such a way that differences between effects of treatments (or yields of varieties) can be estimated unbiasedly and in the most efficient way under the given constraints on resources. He laid down three fundamental principles, *randomization, replication* and *local control* to be followed in designing an experiment to ensure validity of statistical analysis, to provide an estimate of error for estimated treatment comparisons and to minimize the variance of estimates. Fisher (1931) expressed the roles played by these principles in the form of a diagram (see Figure 1). Describing the importance of experimental design in the collection of data, Fisher (1938, 4–159, page 163) said:

> A competent overhauling of the process of collection, or of experimental design may often increase the yield (precision of results) ten or twelve fold, for the same cost in time and labour. To consult a statistician after an experiment is finished is often merely to ask him to conduct a post-mortem examination. He can perhaps say what the experiment died of.

Fisher introduced the concept of factorial designs where each treatment is formed by combining a number of factors at different levels. The aim of an experiment is such a case is to study the effects of individual factors and the interactions which would be of help in determining the optimum mix of factors. Such designs are now routinely used in agriculture and industrial experimentation.

Design of experiments is the most outstanding contribution of Fisher to statistics. G. E. P. Box says, "It is, perhaps, the only tool in statistics which had the greatest impact on analytic and investigative studies in all scientific disciplines and given a status to the statistician as a valued

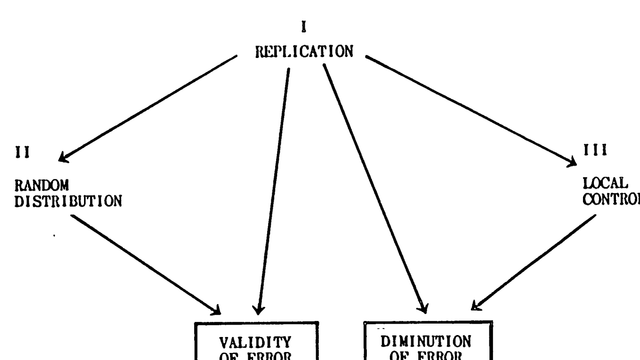

FIG. 1. *Fisher's diagram: "Principles of Field Experiments."*

member of a scientific team." It should also be mentioned that Fisher's ideas of incomplete block designs, confounding and the single replicate factorial designs have inspired a considerable volume of research in combinatorial mathematics. As examples, reference may be made to contributions by the Indian School of Statisticians led by R. C. Bose on finite geometries and mutually orthogonal latin squares, partially balanced and intra- and intergroup balanced incomplete block designs, disproof of Euler's conjecture on the nonexistence of orthogonal latin squares of the order $4n + 2$ (Bose and Shrikande, 1959), coding theory, Kirkman's school girl problem, orthogonal arrays (Rao, 1949), etc.

The importance of replication and local control is well understood, but the principle of randomization and its role in providing valid tests of significance and estimates of error have been questioned from time to time and the controversy still continues. At one end, we have outright rejection of the randomization principle by Basu (1988), and on the other the contributions by Neyman (1935), Rao (1959, 1960), Yates (1964), Youden (1972), Preece (1990), Bailey and Rowley (1987), Bailey (1991) and others who see the need for randomization and examine its role with reference to "validity of the estimation of error." Fisher did not elaborate on the question of validity, but from his writings and comments on papers by others, it is clear, that by "no treatment difference" (null hypothesis), he meant that "each treatment has the same effect on each experimental unit" and by "validity," that "under the null hypothesis, the expected mean square for treatments is equal to that for error" in the appropriate analysis of variance table. A randomization scheme that achieves Fisher's requirement is called *weakly valid*. For instance, it is known (Rao, 1959) that in the case of incomplete block designs such as BIBD, linked block designs and special cases of PBIBD, weak validity is achieved by the randomization rules:

(R_1) The subsets of treatments are assigned to blocks at random.

(R_2) Within each block, the varieties of a subset are assigned to the plots at random.

It is also true that, under the nonnull hypothesis, the expected mean square for treatments (say M_1) exceeds that for the error (say M_2) if an additive model holds (i.e., when an observed yield can be expressed as the sum of treatment effect and plot effect). But, under a nonadditive model, M_1 can be less than M_2 under the nonnull hypothesis and also under the wider null hypothesis that on the total, over all the experimental units, there are no treatment differences (Neyman, 1935; Rao, 1959).

Weak validity does not ensure the more desirable property that the expected mean square for testing a null hypothesis on subsets of treatments or subsets of contrasts is equal to that for the error. A randomization that ensures this is said to be *strongly valid*, a concept introduced by Grundy and Healy (1950). In some cases, the further randomization rule:

(R_3) Label the treatments randomly

may introduce strong validity. For a general discussion on randomization schemes for strong validity and recent developments, the reader is referred to Bailey and Rowley (1987) and Bailey (1991).

One of the early arguments against randomization is that an experimenter who knew his material could choose arrangements that were more accurate than some of the arrangements that would be arrived at by random chance. Fisher thought that the experimenter's knowledge could be better utilized in stratifying the material into homogeneous clusters that could be used as blocks, but the act of randomization is necessary at some stage to produce a valid estimate of error which is of great importance, and worth a small reduction in accuracy.

However, a few points remain to be resolved in the practice and theory of randomization. What should one do if a design arrived at by random choice exhibits systematic features? Should one reject this and make another random choice? Any design of experiment must specify the set of designs from which one may be chosen at random. The only condition the set has to satisfy is that the act of randomization provides unbiased estimates of treatment comparisons and valid estimates of error. If there is more than one set with these properties, what further criteria should be used in choosing an appropriate set? To what extent randomization can be sacrificed when some random assignments of treatments to experimental units are difficult to implement in practice as in the case of Youden's (1972) example? Some discussion clarifying these issues will be useful.

7. FORMS OF QUANTITATIVE INFERENCE

In Chapter 3 of his book, SMSI, Fisher describes at some length different forms of making inference from observational data. He thinks that a monolithic structure of statistical inference, requiring statements of probability about alternative hypotheses given the observed data, may not always be possible or necessary in taking decisions in experimental sciences. He discusses different types of inference depending on the nature of problems and available data. Rereading this book, one gets the impression that Fisher was reviewing his own work, modifying some statements he had made earlier, answering the criticisms leveled against his contributions and introducing some new ideas. We shall briefly review the contents of this chapter and comment on some of Fisher's statements.

7.1 Tests of Significance

In his early writings (SMRW and DOE), Fisher laid great emphasis on tests of significance. Given a null hypothesis, H_0, a test statistic is chosen and its 95% or 99% percentile point is computed. If the observed value of the test statistic exceeds this value, the H_0 is rejected. The decision is based on the logical disjunction: Either an intrinsically improbable event has occurred, or the offered hypothesis is not correct. "The level of significance in such a case fulfills the conditions of a measure of the rational grounds for the disbelief it engenders" (SMSI, second edition, page 43). Such a prescription was, perhaps, necessary at a time when statistical concepts were not fully understood and the exact level of significance attained by a test statistic could not be calculated due to lack of computational power. Fisher clarified his views in SMSI both in respect to the level of significance to be used and the situations in which a test of significance is relevant and useful:

> ... for in fact no scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas.... Further, the calculation is based solely on a hypothesis, which, in the light of the evidence, if often not believed to be true at all, so that the actual probability of erroneous decision, supposing such a phrase to have any meaning, may be much less than the frequency specifying the level of significance.... It (the level of significance) is more primitive or elemental than, and does not justify, any exact probability statement about the proposition.
>
> (SMSI, 2nd edition, pages 42–43)

Fisher thinks that tests of significance have a role to play in scientific research, although they result in a weak form of inference. Thus, when one wants to know whether a normal distribution fits a given data, a general test like the χ^2 goodness of fit is appropriate. If the hypothesis is rejected, an alternative model is sought. The test, by itself, does not indicate what the alternative is. However, once the specification such as the normal family is accepted, then the problem is that of discriminating between alternative values of the parameters of a normal distribution, which falls within the realm of estimation.

It appears from reading SMSI that Fisher gives a limited role to tests of significance in statistical inference, only useful in situations where alternative hypotheses are not specified. He does not recommend any fixed level of significance, but suggests that the observed level of significance has to be used with other evidence that the experimenter may have in making a decision.

However, Fisher's emphasis on testing of null hypotheses in his earlier writings has probably misled the statistical practitioners in the interpretation of significance tests in research work and motivated much theoretical research and publication of text books on a statistical methodology of "limited utility and applicability." [See Wolfowitz (1967) for further remarks.]

7.2 Mathematical Likelihood

In the beginning, Fisher introduced likelihood as a quantity "to designate the state of our information with respect to the parameters of the hypothetical population" (Fisher 1922, 1–18, page 334), and more specifically as "measuring our order of preference among different possible populations" (SMRW, 12th edition, page 30). He used these concepts to introduce the m.l. estimation, but found the need to derive the distribution of the m.l. estimates, sometimes together with some ancillary statistics, for making inferential statements. He did not state the *pure likelihood principle* as later discussed by Barnard and Birnbaum. However, in SMSI, he referred to likelihood as a "measure of rational belief" in some well-defined sense and proceeded to make inferential statements based on the likelihood function only, keeping the observations fixed. He measured the plausibility of a given value of the parameter by the ratio of its likelihood to the m.l. It is not clear how such a measure can be of help in guiding research investigations, apart from the fact that the likelihood function cannot always be defined, and there are certain other difficulties in dealing with m.l. estimates when there are nuisance parameters (see Cox, 1978).

7.3 Fiducial Distribution and Bayes Theorem

This was an attempt by Fisher to make probability statements about the unknown parameters of the Bayesian type without using a prior distribution. In the words of L. J. Savage (1981):

> Fisher's fiducial argument is a gallant but unsuccessful attempt to make the Bayesian omlette without breaking the Bayesian egg.

Fisher was aware of the usefulness of prior probabilities in statistical inference, only when they are inherent and can be ascertained by prior knowledge or introduced through a random mechanism for the choice of a population to be sampled, or estimable from data (empirical Bayes). In fact he uses these concepts in genetic work, as in the construction of discriminant function for genetic selection (Fairfield Smith, 1936; Rao, 1953). But he thought that, in some situations, prior probabilities do not exist (e.g., when the atomic weight of a chemical has to be determined). He describes the axiomatic and personalistic approaches to choose a prior distribution as "bogus" (see 1959, 5-273, for further discussion). However, Fisher's approach for making probability statements about unknown parameters by using the information provided by pivotal statistics is strewn with logical difficulties.

7.4 The Final Thought

In a series of lectures given at the Indian Statistical Institute in 1954-1955, Fisher wrote on the blackboard different forms of quantitative inference in the form of an incomplete list (the explanation within parenthesis is mine)

1. Tests of significance (logical disjunction)
2. Mathematical likelihood (measure of rational belief)
3. Fiducial probability (inversion of a pivotal quantity)
4. Bayes theorem (when there is an inherent prior)
5. ...
6. ⋯

When I asked him what he meant by 5 and 6, he said, these represent other ways which have yet to be discovered, and it is up to the younger generation like you to think about it.

8. CONCLUDING REMARKS

Fisher is the author of about 300 research publications (reproduced in 5 volumes of his collected papers) and six books, of which four are on statistics and two on genetics. The originality of his papers, their thought-provoking contents, and many suggestions for further development should, in spite of the lack of mathematical rigor of some of his contributions, provide a stimulus and challenge to research workers for many years to come.

The recognition of statistics as a separate scientific discipline came only after the theoretical foundations of the subject were laid and its applications to scientific research was demonstrated by Fisher.

The basis of statistics is inductive logic, which remained uncodified until the beginning of the present century because of inherent difficulties in generalizing from the particular. Attempts at quantifying uncertainty in hypothesis testing through levels of significance provided the initial breakthrough, but there are bound to be controversies in the development of the subject. Fisher may have been wrong in some of the statistical methods he advocated. I say, "may," because there are inherent difficulties in judging the merits of any rule of procedure in inductive logic. Some of Fisher's ideas are still being debated. But, undoubtedly, Fisher was the founder of modern statistical theory and an innovator of various topics that are in the main stream of current research.

I would like to point out that, in the ultimate analysis, statistics as practiced by Fisher and some of his predecessors (see 1947, 4-214, 1938, 4-159; 1953, 5-251) is a way of thinking. Statistical methodology is a process by which we analyze data to provide insight into the phenomenon under investigation rather than a prescription for final decision. There is no fixed rule for answering all questions. Search for new methods will continue. I recall what Fisher said in his preface to his 1950 volume of papers, *Contributions to Mathematical Statistics*:

> In each of these fields there is still much to be done. I am still too often confronted by problems, even in my own research, to which I cannot confidently offer a solution, even to be tempted to imply that finality has been reached (or to take very seriously this claim when made by others)!

REFERENCES

Akahira, M. and Takeuchi, K. (1981). *Asymptotic Efficiency of Statistical Estimators: Concepts of Higher Order Asymptotic Efficiency*. Springer, New York.

Akaike, H. (1973). Information theory and extension of the maximum likelihood principle. In *Proceedings, 2nd International Symposium on Information Theory* (B. N. Petrov et al., eds.) 267-281. Akademiai Kiado, Budapest.

Bahadur, R. R. (1958). Examples of inconsistency of maximum likelihood estimates. *Sankhyā* **20** 207-210.

Bailey, R. A. (1991). Strata for randomized experiments (with discussion). *J. Roy. Statist. Soc. Ser. B* **53** 27-78.

Bailey, R. A. and Rowley, C. A. (1987). Valid randomization. *Proc. Roy. Soc. London Ser. A* **410** 105-124.

Barnard, G. A. (1963). Fisher's contributions to mathematical statistics. *J. Roy. Statist. Soc. Ser. A* **126** 162-166.

Basu, D. (1988). *Statistical Information and Likelihood: A Collection of Critical Essays. Lecture Notes in Statist.* **45**. Springer, New York.

Bose, R. C. and Shrikande, S. S. (1959). On the falsity of Euler conjecture about the non-existence of two orthogonal Latin squares of order $4t + 2$. *Proc. Nat. Acad. Sci. U.S.A.* **45** 734-737.

COX, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29** 357–372.

COX, D. R. (1978). Foundations of statistical inference: The case for eclecticism. *Austral. J. Statist.* **20** 43–59.

DAWID, A. P. (1991). Fisherian inference in likelihood and prequential frames of reference (with discussion). *J. Roy. Statist. Soc. Ser. B* **53** 79–110.

ECKART, C. and YOUNG, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika* **1** 211–218.

EFRON, B. (1975). Defining the curvature of a statistical problem, with applications to second order efficiency (with discussion). *Ann. Statist.* **3** 1189–1242.

FAIRFIELD SMITH, H. (1936). A discriminant function for plant selection. *Annals of Eugenics* **7** 240–250.

FISHER, R. A. (1931). Principles of plot experimentation in relation to the statistical interpretation of the results. In *Rothamsted Conferences* **13** 11–13.

FOUTZ, R. V. and SRIVASTAVA, R. C. (1977). The performance of the likelihood ratio test when the model is incorrect. *Ann. Statist.* **5** 1183–1194.

GOSH, J. K. and SUBRAMANIAM, K. (1974). Second order efficiency of maximum likelihood estimators. *Sankhyā Ser. A* **36** 325–358.

GOWER, J. C. (1990). Fisher's optimal scores and multiple correspondence analysis. *Biometrics* **46** 947–961.

GREELAND, S. (1991). On the logical justification of conditional tests for the two-by-two contingency tables. *Amer. Statist.* **45** 248–251.

GREENACRE, M. J. (1984). *Theory and Applications of Multiple Correspondence Analysis*. Academic, London.

GRUNDY, P. M. and HEALY, M. J. R. (1950). Restricted randomization and quasi-latin squares. *J. Roy. Statist. Soc. Ser. B* **12** 286–291.

KALLIANPUR, G. and RAO, C. R. (1955). On Fisher's lower bound to asymptotic variance of a consistent estimate. *Sankhyā* **16** 331–342.

MAHALANOBIS, P. C. (1938). Professor Ronald Aylmer Fisher. *Sankhyā* **4** 265–272.

MARDIA, K. V. (1972). *Statistics of Directional Data*. Academic, London.

MARTIN, E. S. (1936). A study of an Egyptian series of mandibles with special reference to mathematical methods of sexing. *Biometrika* **28** 149.

MATHEN, K. K. (1954). Note on design of experiments and testing the efficiency of drugs having local healing power. *Sankhyā* **14** 175–180.

MAUNG, K. (1941). Measurement of association in a contingency table with special reference to pigmentation of hair and eye colour in Scottish school children. *Annals of Eugenics* **11** 189–223.

MEHTA, M. L. (1967). *Random Matrices and Statistical Theory of Energy Levels*. Academic, New York.

NEYMAN, J. (1935). Statistical problems in agricultural experimentation. *J. Roy. Statist. Soc. Suppl.* **2** 107–154.

NEYMAN, J. (1951). Review of Fisher's Collected Papers. *Scientific Monthly* **LXii** 406–408.

NISHI, R. (1988). Maximum likelihood principle and model selection when the true model is unspecified. *J. Multivariate Anal.* **27** 392–403.

PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series* **V** 157–175.

PREECE, D. A. (1990). R. A. Fisher and experimental design: A review. *Biometrics* **46** 925–935.

RAO, C. R. (1949). On a class of arrangements. *Proc. Edinburgh Math. Soc. (2)* **8** 119–125.

RAO, C. R. (1953). Discriminant function for genetic differentiation and selection. *Sankhyā* **12** 229–246.

RAO, C. R. (1955). Theory of the method of estimation by minimum chi-square. *Bull. Inst. Internat. Statist.* **35** 25–32.

RAO, C. R. (1959). Expected values of mean squares in the analysis of incomplete block experiments and some comments based on them. *Sankhyā* **21** 327–336.

RAO, C. R. (1960). Experimental designs with restricted randomization. *Bull. Inst. Internat. Statist.* **37** 394–404.

RAO, C. R. (1961a). Apparent anomolies and irregularities in MLE. *Sankhyā Ser. A* **24** 73–101.

RAO, C. R. (1961b). Asymptotic efficiency and limiting information. In *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 531–546. Univ. California Press, Berkeley.

RAO, C. R. (1963). Criteria of estimation in large samples. *Sankhyā Ser. A* **25** 189–206.

RAO, C. R. (1964). R. A. Fisher—The architect of multivariate analysis. *Biometrics* **20** 286–300.

RAO, C. R. (1965). On discrete distributions arising out of methods of ascertainment. *Sankhyā Ser. A* **27** 311–324.

RAO, C. R. (1971). Some aspects of statistical inference in problems of sampling from finite populations. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) 177–202. Holt, Rinehart and Winston, Toronto.

RAO, C. R. (1975). Some problems of sample surveys. *Adv. in Appl. Probab.* **7** 50–61.

RAO, C. R. (1984). Discussion on "Present position and potential developments: Some personal views, multivariate analysis" by R. Sibson. *J. Roy. Statist. Soc. Ser. A* **147** 205–207.

RAO, C. R. and WU, Y. (1989). A strongly consistent procedure for model selection in regression problem. *Biometrika* **76** 369–374.

RISSANEN, J. (1978). Modeling by shortest data description. *Automatica* **14** 465–471.

SAVAGE, L. J. (1981). *The Writings of Leonard Jimmie Savage. A Memorial Selection*. Amer. Statist. Assoc. and IMS, Hayward, Calif.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.

STUDENT (1908). The probable error of a mean. *Biometrika* **6** 1–25.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrika* **50** 1–25.

WOLFOWITZ, J. (1967). Remarks on the theory of testing of hypotheses. *New York Statistician* **18** 1–3.

YATES, F. (1964). Sir Ronald Fisher and the design of experiments. *Biometrics* **20** 307–321.

YATES, F. and MATHER, K. (1963). Ronald Aylmer Fisher, 1890–1962. *Bibliographic Memoirs of Fellows of the Royal Society of London* **9** 91–129.

YOUDEN, W. J. (1972). Randomization and experimentation. *Technometrics* **14** 13–22.

ZHAO, L. C., KRISHNAIAH, P. R. and BAI, Z. D. (1986). On the detection of the number of signals in the presence of white noise. *J. Multivariate Anal.* **20** 1–25.